The role of clonal interference across genetic backgrounds and environments

Jorge Moura de Sousa

Dissertation presented to obtain the Ph.D degree in Biology

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Research work coordinated by:

FUNDAÇÃO CALOUSTE GULBENKIAN
Instituto Gulbenkian de Ciência

Oeiras, March, 2016

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER /UNL

Knowledge Creation

Cover image:

Inoculum of fluorescently labelled Escherichia coli *K12 MG1655 (YFP and CFP) dividing and expanding in an agar plate after 24 hours. The same fluorescent channel is shown both at the top and bottom of the image, with the latter mirroring the former for aesthetic purposes.*

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

Financial support from Fundação para a Ciência e a Tecnologia, through grant SFRH/BD/89151/2012 awarded to Jorge Moura de Sousa

Research work developed in coordination with Instituto Gulbenkian de Ciência

Supervisor: Dr. Isabel Gordo

Acknowledgments

6 years is a great portion of one's life. A fifth of my entire existence was spent – no, not spent… enjoyed! – accomplishing what has now been embodied in this mass of pages. So many things happened, and so many people were fundamental in this wild ride. For the first time, I want to use this section to, more than acknowledge, appreciate them properly. The first person who merits these words, for so many reasons, is my mother who did plenty of efforts so that it was possible for me to embark in this adventure. We had difficult times, lost important, loved ones during these years. But we always had each other, and we always will, no matter the distance, or anything else.

The second person I want to mention is my "other mother", the academic one. Everything I could say about the role Isabel has played in this journey is an understatement. Thank you for the all conversations, for opening my eyes to science, for the patience and, most of all, for trusting me. We have done so much, learned so much, in these past years. And yet, I feel there is so much more we could have done, which is, I think, a powerful statement on how fruitful our relationship was. I would like to think that it does not end here. But let's see where our lives lead us. I also want to underline the fundamental role of Margarida throughout this experience. Besides your central and invaluable role in everything we do in the lab, I find it hard to think of someone that has a creativity and passion for science like you do, and it is difficult to properly stress how inspiring that was for me during all these years. Thank you for helping me in the beginning, the middle and the end and for being, more than a colleague, a confidant in times of need. And thank you for all the other many interesting conversations we had outside science as well. That was crucial in keeping me sane during these years. To all my lab colleagues, thank you for making these years so enjoyable and for allowing me to engage in new projects and ideas, whilst helping me along the way. All of you, including those who are now in other places, were so important. From

Ricardo R., with all your crazy and fantastic ideas, to Catarina P. for the patience in teaching me how to do PCRs. From the apples with Daniela Z. to the theoretical discussions with João A. Roberto B., and Paulo D., you were fundamental in the last part of my PhD, and I am thankful that you came along at the right time. Let's get this thing out! And all the others I am not mentioning… I have so many good memories. It was a pleasure to be part of a common group with all of you. And even in our darkest moment, when Henrique left us, I felt part of a united, strong group of friends. I will not forget that. There are also so many people at the IGC and beyond who deserve to be mentioned here. To my colleagues from the first year, for making me, an "adopted biologist", feel welcome in this environment. For Thiago and Élio, and particularly for Manuela, thank you for all the support. For my thesis committee, Jorge and Christen, thank you for the helpful discussions. For the entire IGC PhD community, thank you for the interesting and fruitful conversations. We are not as strong as we could (and should) be, but still what I encountered here during these years made me a better person and a better scientist.

A special word is needed for those special friends that one encounters unexpectedly but who become so important. Sascha W., I will always have a place on my training schedule to run with you, whenever (and wherever) you want. And to those from before, during and after, which played such a fundamental part in what I am today. Catarina M., I am so glad that I can have you in my life, despite everything that happened. Who else is going to comment on the ethology of my bacteria?...

And, finally, a very special mention to Marta. For all the smiles, laughs and silliness that made it so much fun to be in the lab, and that led to… whatever this amazing thing is. For helping me push through these last months with renewed energies and a very clear goal in my mind. For being the perfect full stop to this whole experience, and the wonderful preamble to a new chapter. Thank you for being by my side. You are my love and my future.

Summary

The study of adaptation in microorganisms has led to a significant expansion in knowledge at many biological levels, ranging from biochemistry and genetics, to ecology and demography. Experimental evolution, in particular, has been invaluable at elucidating how complex the adaptive dynamics in microbial populations can be. One of the most fundamental characteristics of these dynamics is the distribution of beneficial mutations driving the adaptive process. How often do microorganisms acquire these mutations? And what are their expected effects? These questions have been at the heart of evolutionary biology from the very beginning, and the studies that have tackled these difficult issues have been tremendously enlightening about adaptive processes. However, the increasing awareness of the complexity of the environment where microorganisms live requires constant development of new approaches to answer these fundamental questions about their evolution. Large population sizes lead to increased levels of clonal interference, and thus to a deviation from the expected outcome in classical regimes of periodic selection. Genetic variation within an evolving population, which is now easily detected by sequencing technologies, can create complex interactions between phenotypes. Environments with antagonistic biotic interactions, pose very different selective pressures from the ones experienced when a species grows alone. All these factors influence adaptation in microorganisms and, importantly, drive the pathogenicity traits that create severe clinical and epidemiological problems.

The main goal of the research presented in this thesis is to address some of these outstanding questions by experimentally evolving the bacterium *Escherichia coli*, in scenarios that are closer to clinically relevant conditions. We study a particular pathogenic trait, antibiotic resistance, by first addressing compensatory adaptation in multiple resistant backgrounds. By performing whole genome sequencing of the evolved populations, we were able to uncover, for the first time, the genetic basis for compensation in bacteria resistant to both

rifampicin and streptomycin. Moreover, we showed that the paths taken for compensating resistance costs in multiple resistant strains differ substantially from the ones observed in the corresponding individual resistant alleles. We also studied compensatory adaptation in populations with standing genetic variation for resistance, a scenario closer to the that found in clinical infections, showing that the relative fitness difference between the different resistances is not predictive of their long-term maintenance in the population. Instead, the fate of resistances is highly dependent on their potential for adaptation. We also investigate the adaptation of a commensal *E. coli* in a biotic environment where antagonistic interactions with cells of the immune system play a crucial role in driving the transition into pathogenicity. We identified potential pathoadaptive targets, observing the existence of antagonistic pleiotropy and clonal interference as main drivers in the transition from commensalism to pathogenicity. Across these studies, we use and develop new theoretical methods to estimate important evolutionary parameters from experimental data and study the dynamics of individual mutations in evolving populations in the context of Fisher Geometric Model.

Using an integrative approach involving experimental evolution, theoretical modelling and next generation sequencing, we studied the adaptive dynamics of evolving microorganisms, in the hope that these sort of studies may contribute to potential avenues to more efficiently predict, prevent and reverse diverse microbial pathologies.

Resumo

O estudo da adaptação em microrganismos levou a importantes desenvolvimentos no conhecimento a vários níveis biológicos, desde bioquímica e genética, até ecologia e demografia. A contribuição da evolução experimental, em particular, tem sido inestimável ao elucidar a complexidade envolvida nas dinâmicas de adaptação em populações microbianas. Uma das questões mais fundamentais destas dinâmicas é a distribuição de mutações que impele o processo adaptativo. Com que frequência são estas mutações adquiridas pelos microrganismos? E quais são os seus efeitos esperados? Estas questões têm sido centrais à biologia evolutiva desde a sua génese, e os vários estudos que abordam estes assuntos complexos têm sido tremendamente instrutivos acerca do processo adaptativo. No entanto, há uma consciência crescente de que o nível de complexidade do ambiente onde habitam os microrganismos requer o desenvolvimento constante de novas abordagens de modo a responder a estas questões fundamentais acerca da sua evolução. Populações de grande tamanho levam à presença de interferência clonal e, portanto, a desvios das observações esperadas em regimes clássicos de seleção periódica. A variação genética gerada nas populações em evolução, que é agora facilmente detectada através de tecnologias de sequenciação, leva a interações complexas entre os fenótipos. Ambientes com interações antagonistas bióticas colocam diferentes pressões selectivas daquelas encontradas quando uma espécie cresce em isolamento. Todos estes factores influenciam a adaptação de microrganismos sendo muito importantes no desenvolvimento de características patogénicas que criam sérios problemas clínicos e epidemiológicos.

O objectivo principal da investigação apresentada nesta tese é responder a algumas destas questões prementes através da evolução experimental de *Escherichia coli*, em cenários que se aproximam de condições clinicamente relevantes. Estudámos uma característica patogénica específica, a resistência antibióticos, investigando a adaptação compensatória em genomas bacterianos

com múltiplas mutações de resistências. Através de sequenciação completa do genoma das populações evoluídas, descobrimos, pela primeira vez, a base genética de compensação em bactérias resistentes a rifampicina e streptomicina. Adicionalmente, mostrámos que os caminhos que levam à compensação de resistências custosas em estirpes com resistências múltiplas diferem substancialmente daquelas observadas com as correspondentes resistências individuais. Estudámos também adaptação compensatória em populações com variação genética para a resistência, um cenário mais próximo daquele encontrado em infecções bacterianas na clínica, mostrando que as diferenças relativas de fitness entre as resistências não prevê a sua manutenção a longo termo na população. Em vez disso, o destino das resistências é altamente dependente do seu potencial adaptativo. Investigámos também a adaptação de uma estirpe comensal de *E. coli* a um ambiente biótico, onde interações antagonistas com células do sistema imunitário são um factor crucial na transição para a patogenicidade. Identificámos potenciais alvos genéticos patoadaptativos, observando a existência de pleiotropia antagonista e interferência clonal como alguns dos principais impulsionadores na transição entre comensalismo e patogenicidade. Ao longo destes estudos, usámos e desenvolvemos novos métodos teóricos para estimar importantes parâmetros evolutivos a partir de dados experimentais, e estudamos as dinâmicas de mutações individuais em populações em evolução no contexto do Modelo Geométrico de Fisher.

Ao usar uma abordagem integrativa, envolvendo evolução experimental, modelação teórica e sequenciação de nova geração, estudámos as dinâmicas de adaptação durante a evolução de organismos, na esperança que este tipo de estudos possa contribuir para potenciais caminhos que levem a meios mais eficientes de prever, prevenir e inverter diversas patologias microbianas.

Table of Contents

Thesis Outline

The main and unifying purpose of this thesis, and the main driver of the work performed during these last years, is to better understand how adaptation occurs across increasingly complex adaptive environments, from simple and monoclonal populations, to polymicrobial populations, and finally in an environment where strong antagonistic interactions occur. We start by developing theoretical methods and simulations to infer key evolutionary parameters from biological observations. We then use antibiotic resistance as a trait to understand compensatory evolution and the role of epistasis in compensation under strong clonal interference, both in single and multiple antibiotic resistant backgrounds. Next, we explore the evolutionary parameters underlying adaptation in populations initially composed of different antibiotic resistance alleles, as observed in polymicrobial infections. Finally, we use a biotic environment that includes macrophages to understand the adaptive processes in the transition of a commensal into a pathogen, and how clonal interference might generate dynamics that strongly influence this process.

First, **Chapter I** introduces some fundamental concepts in evolutionary biology, with a focus on experimental evolution of microbial organisms, as well as antibiotic resistance and the interaction of bacteria with the innate immune system.

Chapter II presents and discusses a computational method to infer the beneficial mutation rate and the distribution of arising beneficial mutations, from populations undergoing experimental evolution tracked by neutral markers.

Then, in **Chapter III**, we use Fisher's Geometric Model to study the dynamics of individual mutations and their aggregation in cohorts, under scenarios of intense clonal interference.

We move away from purely theoretical scenarios in **Chapter IV**, where we study the compensatory process of antibiotic resistant *Eschericha coli*, and try to understand whether the compensatory process is shaped by epistatic interactions

between different resistances. We characterize this process in terms of adaptive dynamics and fitness increase, both of which we use to infer evolutionary parameters, and we describe the genetic basis of compensation of double resistant bacteria.

In **Chapter V** we study a similar compensation process, but with *E. coli* populations that are composed of different antibiotic resistant clones to begin with. We ask if the initial relative fitness difference between these resistant mutants can predict the outcome of composition of the population, or if their different evolvability determines adaptation.

Chapter VI explores a more complex biotic environment, where we study the initial steps of *E. coli* adapting to the intracellular environment of macrophages. We describe this adaptation in phenotypic terms and provide the genetic basis for this transition.

In **Chapter VII**, we discuss the ability of clonal interference to generate genetic variation that explains morphologic and phenotypic dynamics in the adaptation of *E. coli* coexisting with macrophages.

Finally, **Chapter VIII** summarizes and discusses the investigation presented in the previous chapters, contextualizing the results in the current and updated literature and debating future avenues of research.

CHAPTER I

Introduction

Microbial adaptation and experimental evolution

Evolution, the change in heritable traits over time, is a unifying principle that holds information on why something is or behaves in a certain way. This variation in traits within a population, however, is not random, and it is observed across many biological levels of organization, from molecules to interactions between individuals. Therefore, understanding why certain traits are selected over others – that is, why they are adaptive – is fundamental in order to understand life itself. The primary mechanism of evolution is the generation of mutations, random changes in the genome of an organism. Most importantly, the emergence of these genetic variants means there is variation on the phenotype of an organism (be it how many offspring it can generate or the color of its fur) over which natural selection can then act. Most of these changes are neutral or even detrimental to an individual, but there is also a chance for a mutation to confer a beneficial effect, either by optimizing a given trait or by creating a novel adaptive one. Whenever such an event occurs, this new variant should increase in frequency in the population, replacing the worse variants. This fundamental principle drives adaptation in a population, through the emergence of new adapted phenotypes.

One way to better understand adaptation is by studying microbial evolution. Microbes are incredibly diverse and inhabit a vast multitude of habitats, from the depths of oceans (DeLong 2005) to the mammalian digestive system (Jeffrey I Gordon 2012). This microbial diversity is likely to be due to the tremendous adaptive potential of microorganisms. Their large population sizes and fast generation times, with some microorganisms dividing within minutes (Sezonov et al. 2007), allow adaptation to occur and be observed in a matter of a few days (Lenski et al. 1991; Novella et al. 1995). The rapid generation of diversity from a single microbial genotype leads to the coexistence of multiple clonal variants, which can then either compete for resources (Fredrickson & Stephanopoulos 1981; Orr 1998; Hibbing et al. 2010) or interact cooperatively (Deborah M Gordon 2014). Many complex microbial community behaviours can be observed, ranging

from motility (Zhang et al. 2009) to biofilm formation (Xavier & Foster 2007). Competition between microbes has been, however, the most frequently used scenario for studying adaptive dynamics (Gerrish & Lenski 1998; Sniegowski & Gerrish 2010). A wide and well-established set of molecular, genetic and technological tools facilitates the study of microorganisms. While the genetic manipulation of microbial genomes allows putting forward hypothesis regarding the role and effect of particular mutations on the phenotype of an organism, advances in genomic sequencing technology allow the observation of genetic changes that underlie adaptation (Lee et al. 2012; Chedom et al. 2015).

Another significant advantage in studying microbial adaptation concerns the clinical relevance of microorganisms to human hosts. Microbes can be a vicious pathogenic foe, but also a powerful commensal ally, and often the same organism can display both these characteristics, depending on environmental conditions (Littman & Pamer 2011). Phenotypic changes allowing a commensal to become a pathogen (Ebert & Bull 2008; Leimbach et al. 2013), or a microbial invader to acquire resistance to an antibiotic (Paulsen et al. 2003; O'Neill et al. 2001), can occur very quickly. Moreover, certain microbial genomes have mobile DNA, such as plasmids (Touchon et al. 2012) or other mobile genetic elements (Touchon et al. 2014), which can cause horizontal gene transfer of pathogenic traits to occur (Iwasaki & Takagi 2009; Gluck-Thaler & Slot 2015), enhancing their ability to adapt under strong selective pressures. Thus, investigating how adaptation occurs in microbes can not only lead to useful insights on the key principles of adaptation, but also to increase our knowledge on pathologies and the evolution of infectious diseases. The information gathered in microbial adaptation studies can help to exploit their weaknesses and curb their effects on human health.

A very useful tool for studying the adaptation of microbial populations in real time is experimental evolution, where evolutionary scenarios can be created by starting from a common ancestral genotype and controlling the conditions experienced by the organisms, such as the size of populations, temperature or

resource availability (Kawecki et al. 2012; Adams & Rosenzweig 2014). It usually consists on several independent replicate populations evolving in parallel, which are then analyzed for traits that have changed during adaptation. Phenotypic changes, such as the growth rate or the ability to consume a given resource, as well as genomic changes, can be tracked over time. Asking how these modifications have occurred across replicate populations can assess the spectrum of adaptive mutations arising under specific conditions (Lang & Desai 2014) or the even how predictable is adaptation (Blank et al. 2014; Barroso-Batista et al. 2014). Experimental evolution has therefore been used to estimate evolutionary parameters (Hegreness 2006) and test evolutionary hypothesis (Perfeito et al. 2007; Fitzpatrick et al. 2007). A wide range of organisms have been studied under this framework, including *Drosophila melanogaster* (Simões et al. 2008) or *Arabidopsis thaliana* (Kolodynska & Pigliucci 2003). However, many studies that use this tool focus on microorganisms (for instance, virus (Bull et al. 2003), bacteria (Barrick et al. 2009), yeast (Ratcliff et al. 2015) or nematodes (Denver et al. 2012)), due to their easy manipulation and storage. In what is arguably the most famous foray into an experimental evolution study, Richard Lenski's research group has been propagating several replicate *Escherichia coli* populations in minimal media supplemented with glucose for more than 25 years, having now surpassed 64000 generations (Lenski et al. 1991; Maddamsetti et al. 2015). While the initial response of interest was the change in mean fitness of populations over time, it has now spawned a myriad of interesting observations, from the genetic constraints to evolvability (Woods et al. 2011), to selection in mutation rates (Sniegowski et al. 1997) and adaptation involving the cross-feeding between individuals (Rozen et al. 2009). Because populations are frozen and stored every 500 generations, it is even possible to go back to a previous point and replay evolution to understand its contingency (Woods et al. 2011) or the reasons for the extinction of ecotypes (Turner et al. 2015). This hallmark study from Lenski, as many others that followed, focuses on how microorganisms improve in a given environment, but the reverse question is also being explored using a similar framework. Mutation accumulation experiments, which consist in applying drastic

population bottlenecks (usually a single individual or colony), study how the accumulation of mutations in the absence of selection impacts the fitness of organisms over time (Kibota & Lynch 1996; Perfeito et al. 2014). Clinically, experimental evolution has somewhat played an ancient role in the creation of vaccines (Woo & Reifman 2014), which consists in the serial passage of pathogens until their disease causing traits were severely impaired. Nowadays, however, experimental evolution can be used to assess, for instance, the speed and types of mutations associated with drug resistance to a certain antibiotic, by evolving bacteria in media supplemented with the drug (Perron et al. 2006; Palmer & Kishony 2013).

Although based on simple principles, adaptive processes occurring in evolving populations are complex and can sometimes display cryptic properties. They are also fairly difficult to analyze quantitatively and comparatively. Theoretical models, by making certain explicit evolutionary assumptions, allow the interpretation and quantification of adaptive processes, putting forward testable predictions (Smith & Haigh 1974; Wahl & Gerrish 2001). Furthermore, they allow inferring key parameters from experimental observations. For instance, the rates and effects of mutations can be inferred from marker dynamics in an experimental evolution setup (Hegreness 2006; Barrick et al. 2010). Historically, mathematical modeling provided, for instance, a crucial intellectual convergence by conciliating Mendelian genetics with the effect of natural selection on the change in gene frequencies, mainly through the work of Ronald A. Fisher, Sewall Wright and J.B.S. Haldane (Ronald Aylmer Fisher 1930; S Wright 1932; Haldane 1927). Theoretical models have since then been very useful in untangling adaptive processes and explaining certain evolutionary observations, be it at the level of linkage between genes (Desai & D S Fisher 2007), changes in fitness along time (Wiser et al. 2013), organism complexity (Wang et al. 2010), or social interactions in meta-populations (Bucci & Xavier 2014). Developing a useful model, however, requires careful identification of the relevant processes for the particular question one proposes to tackle. As John H. Holland put it, model building can be an art,

because it stems from a process of induction in choosing the relevant parameters (Holland 1995). Therefore, it is crucial to identify and understand what are the key evolutionary processes that underlie the dynamics observed in bacterial evolution.

Patterns and dynamics in bacterial evolution

Generation of bacterial variants

The spontaneous mutation rate in bacteria is estimated to be in the order of 10^{-3} mutations per cell per generation (Lee et al. 2012). This rate, combined with the large population sizes of bacteria, results in a fast generation of genetic diversity within a population. From all these mutations, the vast majority will have a deleterious or neutral effect on an organism (Drake et al. 1998; Eyre-Walker & Keightley 2007), but beneficial mutations, although rare, are the drivers of long-term adaptation (Sniegowski & Gerrish 2010; Frenkel et al. 2014). The fraction of beneficial mutations is strongly dependent on the initial level of adaptation of an organism to a given environment, and the true distribution of effects of these arising beneficial mutations is very hard to determine experimentally (Hietpas et al. 2011). Once a beneficial mutation appears, because it is initially rare in the population, it has to survive stochastic fluctuations in the production of offspring, a process known as genetic drift (Orr 1998). Indeed, many arising beneficial mutations are lost due to drift. Since the probability of a mutation not being lost due to drift is **~2s** (Haldane 1927), with **s** being its fitness effect (or, more precisely, $\frac{1-e^{\frac{-2sNe}{N}}}{1-e^{-2sNe}}$, where **N** and **Ne** are the total and effect population, respectively) (Kimura 1962), mutations of stronger beneficial effects have a lower likelihood of being lost due to this process. Natural selection is less effective in smaller populations, as these are more susceptible to these stochastic fluctuations (Lanfear et al. 2013). A mutation that is able to overcome drift will start to increase in frequency and, if no other beneficial variant of a similar or higher effect appears, it will sweep to fixation. This occurs in a regime where adaptation is driven by

successional accumulation of mutations, and is typically observed when populations are small and beneficial mutations are rare (Atwood et al. 1951). However, larger populations with a higher influx of beneficial variants will lead to alternative regimes (Desai & D S Fisher 2007).

Competition between beneficial variants – Clonal Interference

Beneficial alleles that survive loss by drift and start increasing in frequency in the populations are considered contending, or competing, mutations (Gerrish & Lenski 1998). In regimes of large population sizes and high frequency of beneficial mutations, it is expectable that, contrary to the successional regime, multiple beneficial variants arise and segregate simultaneously. Because microbes reproduce mainly asexually, the emergence, through recombination, of clones with multiple beneficial mutations (Ronald Aylmer Fisher 1930; Felsenstein 1974) in the same individual is rare. As a consequence, individual mutations compete against each other for increase in frequency and, ultimately, for fixation. This is a process that is known as clonal interference (Gerrish & Lenski 1998).

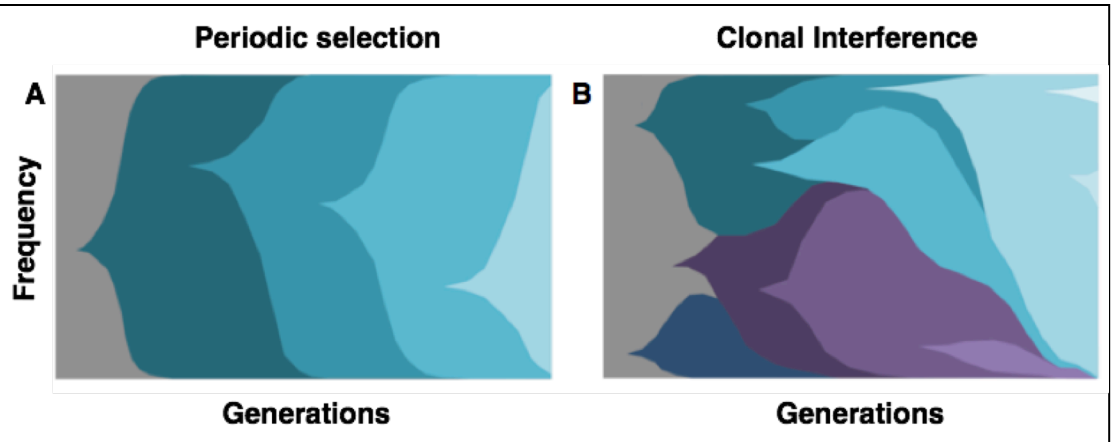


Figure 1. Dynamics in regimes of periodic selection and clonal interference. Different colors represent alleles emerging and segregating through time, with grey being the ancestral genotype. In panel **A**, the successional mutation regime is represented, where the periodic sweep of each beneficial allele. Panel **B** represents a regime of clonal interference, where multiple mutations segregate simultaneously and interfere with the other variants. Adapted from (Marx 2013).

While in the successional regime any mutation that survives drift can reach fixation, regardless of its fitness advantage, in regimes of clonal interference there is an enrichment of mutations with stronger effects, since these are the ones that are able to outcompete the other beneficial variants. Moreover, because mutations interfere with each other's segregation, the fixation of single alleles is less common than expected if such interference did not occur. Fixations take a longer time to occur, and there is higher likelihood for observing genetic diversity within a population (Comeron & Kreitman 2002; de Visser 2005). These predictions have been observed across experimental evolution studies that use large population sizes (de Visser 2005; Lang et al. 2013). Moreover, clonal interference has also been observed in clinical settings, notably in persistent infections of antibiotic resistant pathogens (Navarro et al. 2011) or viral evolution (Chedom et al. 2015; Zanini et al. 2016). For instance, temporal analysis of tuberculosis infections has show that clonal interference is prevalent (Eldholm et al. 2014), and can even occur within the same host (Al-Hajoj et al. 2010). The Influenza virus has also been shown to evolve under strong clonal interference, with selection for viral strains of increasingly stronger beneficial mutations (Strelkowa & Lässig 2012). Other biological processes that rely on clonal reproduction have also been observed to be under the influence of clonal interference. Namely, in cancer development and progression (Ding et al. 2012), where fast division times and high mutation rate of deregulated cells originates genetic diversity and competition between different cancer lineages (Greaves & Maley 2012). Across all these systems, one important consequence of the longer fixation times caused by clonal interference is that there is time, before the fixation of an allele, for a genetic background to acquire subsequent mutations.

Interactions between mutations – Epistasis

If the effect of a mutation depends on whether another mutation is present in the genome, that is, if it depends on the genetic background where it occurs, there is epistasis between these mutations (Bateson 1910; Phillips 2008). The likelihood for any two mutations to interact (and the strength of this interaction)

depends on the connectedness of the processes that are affected. For instance, mutations that affect a similar biosynthetic pathway or cellular process are more likely to conflict than functionally distinct mutations (Phillips 2008). There are two main types of epistasis: magnitude and sign. For magnitude epistasis, two mutations can interact antagonistically, if the combined effect of the mutations is smaller than predicted by their individual effects, or synergistically, if it is bigger than expected. An extreme case of this latter type of epistasis is synthetic lethality, where a specific mutation becomes lethal if it co-occurs with another one (St Onge et al. 2007). Sign epistasis occurs if the sign of the effect of a mutation changes: e.g., a beneficial mutation becomes deleterious in the presence of another mutation. The extreme version of this type of interaction is called reciprocal sign epistasis, and it occurs if two deleterious mutations result in increased fitness when put together (Kvitek & Sherlock 2011).

Evolutionarily, the concept of epistasis is crucial because it can effectively constrain the adaptive paths an organism can follow (Fenster et al. 1997), since a modified allele might have a different spectrum of subsequent mutations that become deleterious in that background. Therefore, it has an important role in the predictability of evolution (Weinreich et al. 2006) and can also constrain the evolution of genetic architecture (Kondrashov 1988; Gros et al. 2009). Epistasis ultimately defines the ruggedness of the adaptive landscape of organisms (Kauffman & S Levin 1987; Martin et al. 2007). Fitness landscapes (S Wright 1932) were once discussed purely in conceptual terms, but this is now changing, as real genetic landscapes, built by constructing genotypes with the different possible mutations are being explored (Jacquier et al. 2013). By performing combinations between the mutations, it is possible to clearly map the landscape with types of combinations that are viable (Khan et al. 2011). The role of epistasis in health is also very important. One particular example is regarding the emergence of multiple drug resistances (Chakrabarti & Gorini 1977; Hall & MacLean 2011), where it is important to understand the likelihood of multiple mutations that confer resistance to different drugs to accumulate in a genetic

background. Understanding an epistatic interaction between two resistance alleles could predict a given cost of resistance, and the correspondent likelihood for certain resistance alleles to co-occur together (Trindade et al. 2009).

Segregation and competition of haplotypes

Once multiple mutations accumulate in a genome, and unless their epistatic interactions are completely known, it is difficult to disentangle the effects of single alleles. Instead of referring to a beneficial variant, the term haplotype is used, denominating the group of polymorphisms that are inherited together. Haplotypes can be composed entirely by beneficial mutations, but might also include neutral, or even slightly deleterious alleles, which hitchhike with the ones that are positively selected (Smith & Haigh 1974; Desai & D S Fisher 2007). Experimental evolution studies, coupled with sequencing technologies, allow the observation of the dynamics of individual mutations in evolving populations (Lang et al. 2013; Zanini et al. 2016). The general observation is that there is an impressive level of polymorphism in populations and that selective sweeps of a single mutation are rare. Instead, multiple mutations are usually seen in aggregates – cohorts – that change in frequency together, competing with other groups of mutations and eventually fixing as a group (Lang et al. 2013). Similar observations come from the analysis of bacterial infections, where a mutation that provides resistance to an antibiotic segregates with other additional mutations that might improve their otherwise costly fitness effect (Comas et al. 2012).

Antibiotic resistance and compensatory adaptation

The history and ecology of antibiotics

One of the most important medical advances of the past century is the discovery and development of drugs to treat bacterial infections. Recognized as a clinical threat in late 19th century, the first therapeutic agent developed against these infections came almost half a century later, with serendipitous discovery of

penicillin by Alexander Fleming in 1929, and the development of its antibiotic in 1942. Fleming left plates cultured with *Staphylococcus aureus* for a long time in the laboratory, and upon his return from summer holidays he realized that the fungi that had contaminated his petri dishes were inhibiting the growth of *S. aureus* (Hare 1982). This marked the beginning of a golden era for clinicians and humans in general. For the first time in history, infections were manageable by therapeutics and even invasive surgery was possible. The discovery of antibiotics provided several clinical breakthroughs and, since then, millions of deaths have been prevented each year, with the consequent increase in the average life expectancy of humans. Perhaps overly optimistic by the success of antibiotic treatments, society started to overuse these drugs, which may have played an important role in directing us back to a state of increasing powerlessness against these pathogens (World Health Organization 2014).

An antibiotic treatment imposes a very strong selection for bacteria that survive exposure to the drugs. This strong selective pressure, together with the ability of microbes to quickly adapt, lead to the emergence of bacteria that are resistant against the treatment of antibiotics (Abraham & Chain 1940; J Davies & D Davies 2010). The full extent of this ability was only fairly recently realized, since microbiologists thought that the frequency of resistance mutations was too low to be an actual threat (J Davies 1994). However, the potential dangers that come with the misuse of antibiotics were identified as soon as after the discovery of penicillin, when Fleming himself realized that some bacteria had the ability to become resistant to the treatment. In his own words, “the time may come when penicillin can be bought by anyone in the shops. Then there is the danger that the ignorant man may easily underdose himself and by exposing his microbes to non-lethal quantities of the drug make them resistant” (Fleming 1945). The first identified acquisition of resistance to an antibiotic treatment marked the beginning of an arms race between scientists and the ever-adaptable microbes, where the discovery of new treatments was – and is – soon counteracted by the evolution of resistance mechanisms (Hede 2014).

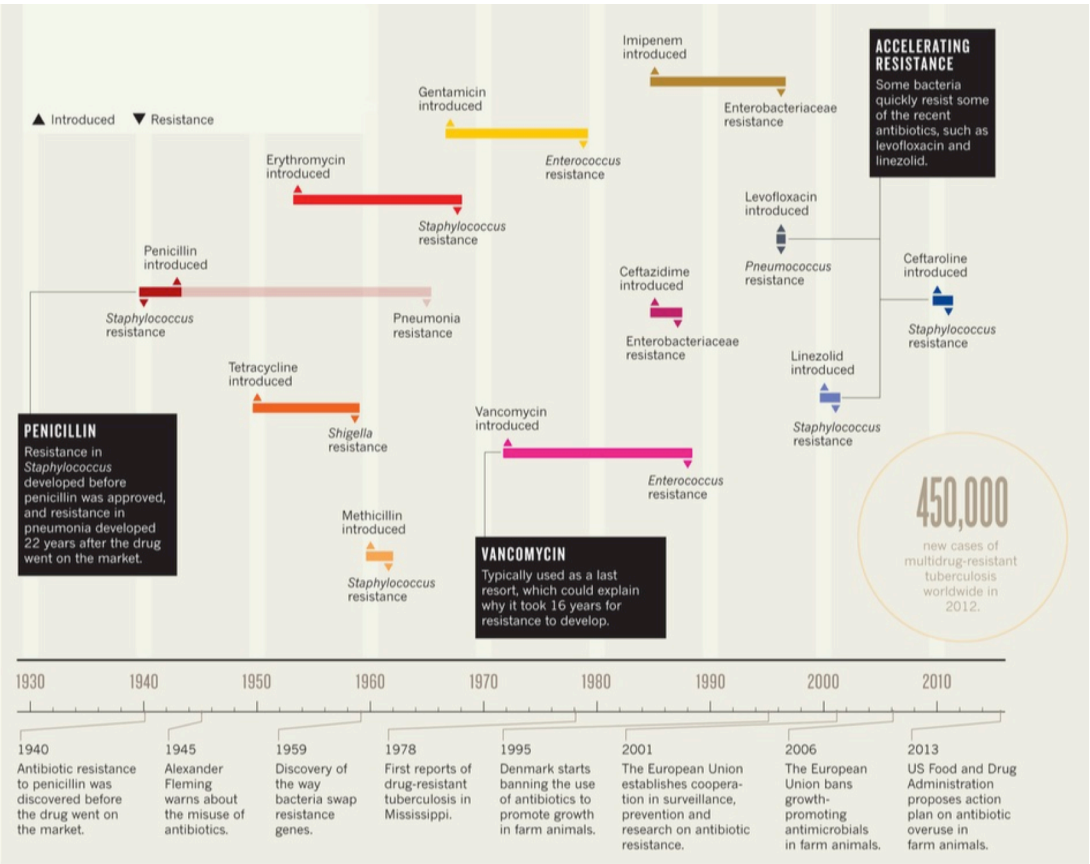


Figure 2. Timeline of antibiotic clinical introduction and emergence of resistance. Upward arrows indicate when the drug was introduced as therapeutic treatment; downward arrows indicate when resistance was first detected. Retrieved from (Hede 2014).

Despite this recent (from a human standpoint) competition between drug development and resistance, one has to keep in mind that this process has been occurring for far longer than humans even existed (Brown & Gerard D Wright 2016). Both antibiotic production and mechanisms of resistance are ancient tools that microorganisms have been evolving for millions of years, using these secreted weapons competitively against other strains and species, but also as means of communication (Romero et al. 2011). In fact, many of the antibiotics used

clinically, like the case of penicillin, are identified and isolated from microorganisms themselves (e.g., vancomycin (Levine 2006)). When surveying microorganisms in various environments, either from the soil (Forsberg et al. 2012) or isolated bacterial communities (Bhullar et al. 2012), several studies have realized that the resistome (Gerard D Wright 2010), a global collection of genes conferring resistance, is genetically diverse, environmentally widespread and predates by far the clinical use of antibiotics (Brown & Gerard D Wright 2016). While the evolutionary and ecological aspects of antibiotic production and resistance are far from being completely understood, the unrestrained use of antibiotics by humans applied a tremendous selective pressure for the selection and dissemination of these resistance genes (Gullberg et al. 2011; World Health Organization 2014).

Types of antibiotics and mechanisms of action and resistance

Antibiotics fall into three main categories: natural antibiotics, which are isolated from living organisms (e.g., aminoglycosides); others are semi-synthetic modifications of natural compounds (e.g., beta-lactam or carbapenems); the remaining ones are completely synthetic (e.g., sulfonamides and quinolones). Streptomycin, for instance, is a naturally isolated aminoglycoside that binds to the ribosome and inhibits elongation of protein synthesis (Luzzatto et al. 1968), and the resulting misfolding of proteins produces many toxic and pleiotropic effects (Abad & Amils 1994). It is a drug that has been paramount in controlling infections of tuberculosis, and, according to the WHO, is one essential medicinal compound. Another example of a semi-synthetic antibiotic is rifampicin, which was derived from the naturally isolated rifamycin. This drug inhibits the RNA polymerase, impairing the synthesis of RNA (Wehrli 1983), causing cell death, and it is also one of the most widely used drugs to treat many different bacterial infections, such as tuberculosis and MRSA.

The main function of antibiotics is to perturb important biochemical processes, therefore disrupting the normal functioning and the survival ability of cell.

Antibiotics can have bactericidal properties, meaning that they actively kill bacteria by targeting the cell wall, membranes or essential enzymes and functions; or they can be bacteriostatic, meaning that bacterial reproduction is prevented by targeting protein synthesis (Yamori et al. 1992). Although these are the two classical means of describing an antibiotic, whether they actively kill a cell or simply stop its reproduction is strongly dependent on the concentration of the compound, as protein synthesis are also required for cell survival (Pankey & Sabath 2004).

Resistance to the action of an antibiotic can be of different types (Tenover 2006; Blair et al. 2014): 1) inactivation or modification of drug compound, by the production of additional residues that are added to the antibiotic molecule and prevent binding its target; 2) modification of the target site, by conformational changes; 3) reduction of levels of the drugs, for instance through efflux pumps located in the cell wall that export antibiotics out of the bacterial cell; and 4) modification of the metabolic pathway affected by a drug, by relocating the metabolic functions through other means. Many of these mechanisms are not readily at the dispose of a bacterial cell. Resistance can often be obtained through the exchange of DNA from one bacteria to another, through conjugation or plasmid mediation (Maiden 1998; van Hoek et al. 2011). Genomic islands with entire collections of antibiotic resistance genes can be transmitted through horizontal gene transfer, which has been shown to play a crucial role in the spread of resistance in bacteria populations, particularly for MRSA (Kriegeskorte & Peters 2012). Other means of acquisition of resistance involve modifications in the essential proteins and mechanisms targeted by antibiotics (Jacoby & Archer 1991; Spratt 1994; Blair et al. 2014). For instance, in *Mycobacterium Tuberculosis*, drug resistance arises almost exclusively through chromosomal mutations in genes that are required for antibiotic action, i.e., genes that encode the proteins targeted by the drugs or the enzymes required for drug activation (Sun et al. 2012; Farhat et al. 2013; Zhang et al. 2013). This is the case for mutations in a gene coding for the beta subunit of RNA polymerase, *rpoB*, which are known to confer resistance to

rifampicin (Telenti et al. 1993). Similarly, mutations in genes coding for ribosomal proteins (for instance, *rpsL* (Kurland et al. 1996)), allow bacterial cells to survive exposure to streptomycin. These chromosomal mutations can occur at relatively high rates. Resistance to rifampicin, for instance, can spontaneously arise at frequencies higher than 10^{-8} in *E. coli* (M R Baquero et al. 2004). How these resistances spread in bacterial communities depend also on demographic parameters such as the population size and bottlenecks (Schrag et al. 1997; Andersson & Hughes 2011).

Polymorphism in antibiotic resistant populations

Given both the high rate of generation of resistant mutants and the multiple targets that confer resistance to an antibiotic, the emergence, competition and co-existence between different resistance alleles are to be expected. Clonal interference plays an important role in clinical infections (Miralles et al. 1999; Mariam et al. 2011), since it can select for the most fit resistant clones (Hughes & Andersson 2015). This has been observed in several studies that sampled clones from infections at multiple time points, where diverse segregating resistant backgrounds are observed (Al-Hajoj et al. 2010; Navarro et al. 2011). As an example, in a study where 3 patients infected with *M. tuberculosis* were followed along time, 4 different resistant clones were detected to be segregating simultaneously, while no resistance had been previously detected (Sun et al. 2012). At the last time point sampled, 94% of the population had a single resistant allele, whilst the remaining 6% had other resistance mutations. Another study followed a single tuberculosis patient for 42 months, a period over which 12 drug resistance mutations where detected above a frequency of 25% in the population. Over this time period, only 7 of these mutations reached fixation status (Eldholm et al. 2014). Clonal interference and cohort fixation were also shown to be involved in virus evolution, where each selective sweep detected in influenza A was inferred to be composed of 3 or 4 mutations on average (Strelkowa & Lässig 2012). The observations that there are complex dynamics occurring within an infection point to the importance of detecting these clinical cases early, in order to avoid selection

for stronger resistance mutations. Additionally, they also imply that practices of isolating a single resistant clone from infections might be misleading, since there the level of diversity is high. Recognizing the segregation of cohorts of mutations that hitchhike with the fitter resistant alleles is also fundamental to understand the chains of transmission of infections between patients (Eldholm et al. 2014).

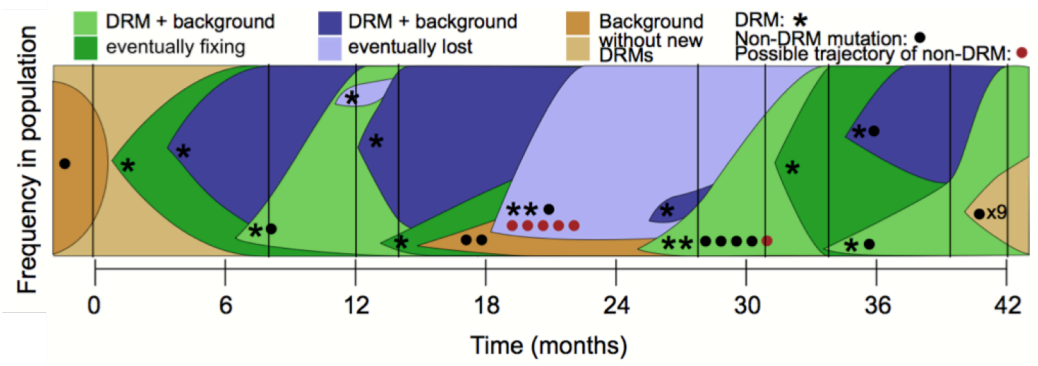


Figure 3. Dynamics of clonal interference in an *M. tuberculosis* infection. Many clonal lineages harbouring different drug resistance mutations (DRM, identified by asterisks) appear along time and compete amongst themselves. Some of these lineages eventually fix in the population (green), whilst others never reach fixation status (blue). Brown background indicates lineages with mutations not known to confer DRM (identified by circles), and mutations for which inferred trajectories differ from clonal dynamics are in red. Based on the data from (Eldholm et al. 2014), figure retrieved from (Wilson et al. 2016).

Compensation to costly resistance mutations

Mutations that confer resistance to antibiotics allow bacteria to survive in the presence of the drugs, but most of them are actually costly to the cell (Schrag & Perrot 1996; B R Levin et al. 1997; Andersson & Hughes 2010) (but see also (Rodríguez-Verdugo et al. 2013) for a study where rifampicin resistance has emerged in the absence of the antibiotic). Because many antibiotic resistance mutations occur in genes that encode essential cellular mechanisms (Spratt 1994; Andersson & Hughes 2010; Melnyk et al. 2015), this cost is usually expressed in

terms of growth rate, virulence and competitive performance (Björkman et al. 1998). For instance, resistance to fluoroquinolones can cause impaired motility (Stickland et al. 2010), while resistance to aminoglycosides can interfere with the structure of the ribosome, reducing the translation rate and impairing growth in rich media (Schrag et al. 1997; Paulander et al. 2007). These deleterious effects of antibiotic resistance mutations are mostly prevalent when the drug is withdrawn, and the resistant cells have a competitive disadvantage to sensitive cells (Andersson & Hughes 2010). Therefore, without the selective pressure of antibiotics, resistant cells should decrease in frequency, ultimately driven to extinction by sensitive cells. Indeed, the fitness cost of resistant cells is the main responsible for the decrease in frequency of resistant variants in bacterial communities, as postulated theoretically (B R Levin et al. 1997; 2000), and assuming no further adaptation would occur.

There are two main strategies for resistant bacteria to avoid extinction: the first is through a mutation that reverts the resistance allele, returning the bacteria into a sensitive state (Gifford & MacLean 2013); the second is by acquiring additional chromosomal modifications, in other sites of the genome, that buffer the deleterious effects of the resistance mutation, conferring a compensatory effect (Schrag et al. 1997; Reynolds 2000; Björkman et al. 2000; Maisnier-Patin et al. 2002; Brandis & Hughes 2013). The latter is arguably the worse outcome for the host, because it allows bacteria to maintain their resistance ability, but at a much lower competitive cost. As a consequence, a resistance allele might be stabilized in the population (Schrag & Perrot 1996; Andersson & B R Levin 1999), because its fitness cost is either reduced or non-existent, and a reversion of the mutation now lies across an adaptive valley (B R Levin et al. 2000). Compensatory adaptation is not limited to antibiotic resistance mutations, potentially occurring whenever there is a deleterious modification in the genome (Szamecz et al. 2014; Bouma & Lenski 1988). A deleterious mutation was estimated to have, on average, 12 possible compensatory mutations (Poon & Chao 2005) and therefore, because the number of sites in a genome that can compensate a costly mutation

is higher than the single allelic location that can lead to a back mutation, compensation is more likely to occur than true reversions (B R Levin et al. 2000). Even if a reversion has a strong beneficial effect, since it restores the original state of an individual, its low rate of appearance explains why compensatory mutations of lower beneficial effects can dominate a population. Once a reversion appears, different compensatory alleles are already segregating, and its relative lower effect increases the likelihood of being lost by drift (Gifford & MacLean 2013).

Discovering the evolutionary dynamics of acquisition and segregation of these compensatory mutations is the subject of intense investigation, because it is paramount to predict stabilization and avoid the long-term maintenance of the resistance mutations and the spread of drug resistance (Andersson & Hughes 2011; Hughes & Andersson 2015). Compensatory adaptation has been observed and studied both in laboratory and clinical settings, and has been shown to be strongly dependent on the demographic conditions of the populations (Maisnier-Patin et al. 2002). When *E. coli* resistant to streptomycin, through mutations in *rpsL*, was serially passaged in streptomycin-free media, the populations maintained the resistance to the drug whilst acquiring compensatory mutations in other loci, restoring the efficacy of translation (Schrag & Perrot 1996). Mutations compensatory for a resistance mutation in the *rpsL* gene were estimated to occur at a rate of at least 10^{-7} per cell per generation (Maisnier-Patin et al. 2002). In another study by Björkman and colleagues, *S. typhimurium* populations resistant to streptomycin, rifampicin or nalidixic acid were serially passed in laboratory mice without antibiotic treatment. The virulence of these strains was initially impaired, but after evolution this ability was restored, and in the majority of cases these was due to mutations in other loci (Björkman et al. 1998). In clinical settings, a study by Comas et al has shown that the genomes of rifampicin resistant isolates of *M. tuberculosis* harboured mutations in other loci, such as *rpoA* or *rpoC*, that compensated the fitness cost without the loss of resistance (Comas et al. 2012).

Importantly, studying compensatory adaptation has also shed light on important cellular functions, including functional interactions between ribosomal

proteins (Maisnier-Patin et al. 2007). This is due to the fact that many compensatory mutations occur either in the same protein where the original resistance mutation emerged, or on functionally connected proteins (Brandis et al. 2012). These secondary mutations that interact epistatically with drug resistance mutations are important factors that need to be incorporated when making predictions about the epidemiological consequences of drug resistance mutations (Wilson et al. 2016).

Multiple antibiotic resistances

Because organisms adapt to the cost of resistance mainly through mutations in other alleles, if an infection is maintained, or re-occurs, the same antibiotic is ineffective and cannot be used again. Occasionally, some resistance mutations acquired in the presence of an antibiotic are known to, by themselves, confer resistance to other drugs. This effect is called collateral resistance, or cross-resistance, and it renders certain alternative treatments less efficient (Lázár et al. 2014). However, the most common scenario is that resistance is specific to a certain drug and, in such cases, alternative antibiotics can be used to curb infections (Kim et al. 2014). This new pressure can, however, lead to the acquisition of yet another resistance, thus increasing the repertoire of antibiotics to which a microorganism can survive, and increasingly constrain the options for the treatment of infections. The emergence and spread of these multiple drug resistances (MDR) are increasing at an alarming rate (World Health Organization 2014). In some case, bacteria seem to acquire resistance faster than new antibiotics are developed and, for antibiotic treatments, technology might be losing the arms race with evolution (Hede 2014). In some extreme cases, namely for tuberculosis, extensively drug resistance (XDR) strains have been clinically detected, meaning that they are resistant to the majority of known therapeutics, rendering clinical treatment highly ineffective (Velayati et al. 2009). A broad research effort is dedicated on how to deal with these MDR pathogens, including how to effectively use antibiotic therapeutics to minimize the emergence of these strains or even revert their multiple resistant state (Kim et al. 2014).

An interesting aspect of MDR is that different resistance mutations are known to interact epistatically (Borrell et al. 2013; Trindade et al. 2009). This is not surprising, since, as discussed, antibiotic resistance mutations often modify proteins that are responsible for essential machinery of the cells, which commonly interact and are part of the same cellular processes. So, for instance, it is likely that a mutation affecting translation interacts epistatically with another that affects transcriptional processes (Hall et al. 2011). The actual mechanistic processes involved in these interactions are, for the most part, unknown. However, research suggests that different allelic mutations in the same genes display a variety of epistatic effects (Durão et al. 2015; Trindade et al. 2009). Of particular concern is the fact that some secondary resistant mutations are themselves compensatory for the fitness deficits imposed by one of the resistance alleles (Hall & MacLean 2011). For instance, a study by Borrell and colleagues indicates that 35% of combinations of *Mycobacterium smegmatis* resistant to both rifampicin and oflaxocin have a higher fitness than at least one of the corresponding single resistant mutant (Borrell et al. 2013). These observations indicate that not only the acquisition of multiple antibiotic resistances has a non-linear effect on the cellular processes, but also that the interactions can be intricate and concealed by other effects. A possible implication from this observation is that the compensatory adaptation of multi-resistant bacteria might differ from what is known regarding single resistant alleles. So far, this question remains unexplored (MacLean & Vogwill 2015).

The immune system as an adaptive pressure

The thin line between pathogenicity and commensalism

Many laboratory studies that follow real-time adaptation in microorganisms are performed in simple abiotic environments, with either single or multiple sugars, and constant, daily bottlenecks (Kawecki et al. 2012). There is a vast array of known and unknown differences between controlled laboratory conditions and the

real world of organisms. Therefore, parameters inferred from such *in vitro* experiments, like mutation rates and selective effects, might differ from the ones where microorganisms normally exist. For instance, the effects of antibiotic resistance mutations are very different when assessed *in vitro* or *in vivo* (Martinez & F Baquero 2000; Luangtongkum et al. 2012). Although recently experimental evolution studies following adaptation in more realistic and complex environments have begun to be explored (Gordo et al. 2014; Hindré et al. 2012), most of the knowledge about evolutionary parameters originates from adaptation to abiotic conditions.

One of the complex environments where microorganisms live is within a mammalian host, where they can play both the role of commensals and pathogens (Leimbach et al. 2013). Often this definition is ambiguous, because it depends strongly on the conditions experienced by a microorganism. Disease or illness symptoms within a host are commonly the result of the wrong cell multiplying in a wrong place at the wrong time. For instance, *Klebsiella Pneumoniae*, an opportunistic pathogen that causes pneumonia and urinary tract infections, is a normal commensal member of the mammalian gut flora (Lau et al. 2008). Another example is found when bacterial cells take advantage of a disrupted gastrointestinal barrier to invade the blood stream and cause septicemia (Rowlands et al. 1999; Macpherson et al. 2005). Therefore, uncontrolled growth of bacteria that would, otherwise, be harmless can spell disaster to its host. This is why the control of bacterial populations by the immune system is crucial.

The antagonism of the innate immune system

Whilst the immune response to the presence of a pathogen is part of a complex cascade, the first responders to an invasion are usually components of the innate immune system (Akira et al. 2006). Macrophages are one of the key elements of this system, providing defense against invading bacteria by the direct and immediate bactericidal response through phagocytosis (Underhill & Ozinsky 2002). This process is complex, and involves several steps, through which a

bacterial cell is recognized, engulfed and destroyed within the macrophages. First, recognition is mediated by pattern-recognition receptors (PRR), which identify molecular structures in the surface of the bacterial cell (Medzhitov 2007). The main structures targeted by PRRs are microbe-associated molecular patterns (MAMPs), which can include lipopolysaccharides, peptidoglycans, carbohydrates and endotoxins (Medzhitov 2007). Different pathways can be activated upon a successful recognition, and depending on the type of activation, macrophages are modified physically and chemically in order to engulf the microbes (Mauel 1982). When this happens, bacteria are confined to phagosomes, which are endocytic vessels that undergo a continuous fusing process (first with early and then with late endosomes, and finally with lysosomes) (Vieira et al. 2002). At the end of this process, bacteria are destroyed through mechanisms such as dramatic increases in pH and the deployment of high concentrations of reactive oxygen species and antimicrobial peptides (Slauch 2011).

Pathoadaptative mutations

With such strong antagonistic interactions with macrophages, bacteria face a tremendous selective pressure. Many bacterial pathogens have evolved mechanisms to escape, resist or even kill phagocytic cells (Rosenberger & Finlay 2003; Baxt et al. 2013). These mechanisms include avoidance of detection by the innate immune system (by modifying the molecules that trigger TLR signaling, for instance (Le Negrate 2012)), inhibition of engulfing by macrophages (through the secretion of proteins that interfere with the phagocytic process (Celli & Finlay 2002)), or survival in the intracellular environment within phagocytes (by increasing its resistance mechanisms against the microbicidal attacks of the cell (Ernst et al. 1999)). This latter strategy is particularly interesting, because it marks the beginning of an adaptive path that may lead to pathogens that were once extracellular to be able to survive (and even replicate (Helaine et al. 2010)) within host cells. Inhabiting a host cell can serve both the purpose of avoiding other elements of the immune system (Li & Yang 2008), or to acquire nutrients from eukaryotic cells (Saka & Valdivia 2010). While some microbes become obligate

intracellular pathogens, as in the case of *Chlamydia* (Fields et al. 2011), many pathogens are facultative intracellular, being able to survive both inside and outside eukaryotic cells (Silva & Pestana 2013). This is the case for *Shigella*, *Salmonella* or *M. tuberculosis*, the last of which inhabits macrophages by arresting phagosome maturation and avoiding delivery to the lysosome (Kelley & Schorey 2003). The exit of the intracellular niche usually implies that the host cells are destroyed in the process, through microbial induction of cell death, for instance (Hybiske & Stephens 2008).

Adaptation to endure encounters with the immune system can occur through the acquisition of pathoadaptive mutations, which are changes in particular traits that increase pathogenicity behaviours in bacteria (Sokurenko et al. 1999). It is then important to understand what are the types of genomic adaptations that can lead to pathoadaptive mutations, and what are the dynamics of clonal interference that give rise to strong pathoadaptations. It is likely that studying these dynamics in complex environments differ from the ones observed in abiotic environments, due to different ecological and demographic processes, such as intracellular replication rates (Helaine et al. 2010) or the existence of spatial structure in the gut (Lu et al. 2014). The mere presence of an additional species in a medium can influence the dynamics of adaptation of bacteria, as suggested by the increased rates of adaptation and diversification of *E. coli* evolving in the presence of phages (Buckling & Rainey 2002). Therefore, it is crucial to understand how biotic environments and antagonistic interactions with the immune system shape the adaptation of bacteria and drive the important transition from commensalism to pathogenicity.

References

Abad JP, Amils R. 1994. Location of the streptomycin ribosomal binding site explains its pleiotropic effects on protein biosynthesis. Journal of Molecular Biology.

Abraham EP, Chain E. 1940. An enzyme from bacteria able to destroy penicillin. Nature.

Adams J, Rosenzweig F. 2014. Experimental microbial evolution: history and conceptual underpinnings. Genomics. 104:393–398. doi: 10.1016/j.ygeno.2014.10.004.

Akira S, Uematsu S, Takeuchi O. 2006. Pathogen recognition and innate immunity. Cell.

Al-Hajoj SAM et al. 2010. Microevolution of Mycobacterium tuberculosis in a Tuberculosis Patient. J. Clin. Microbiol. 48:3813–3816. doi: 10.1128/JCM.00556-10.

Andersson DI, Hughes D. 2010. Antibiotic resistance and its cost: is it possible to reverse resistance? Nature Reviews Microbiology. 8:260–271. doi: 10.1038/nrmicro2319.

Andersson DI, Hughes D. 2011. Persistence of antibiotic resistance in bacterial populations. FEMS Microbiology Reviews.

Andersson DI, Levin BR. 1999. The biological cost of antibiotic resistance. Current Opinion in Microbiology.

Atwood KC, Schneider D, Ryan FJ. 1951. Periodic selection in Escherichia coli. Proceedings of the National Academy of Sciences. 37:146–155.

Baquero MR et al. 2004. Polymorphic Mutation Frequencies in Escherichia coli: Emergence of Weak Mutators in Clinical Isolates. Journal of Bacteriology. 186:5538–5542. doi: 10.1128/JB.186.16.5538-5542.2004.

Barrick JE et al. 2009. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature. 461:1243–1247. doi: 10.1038/nature08480.

Barrick JE, Kauth MR, Strelioff CC, Lenski RE. 2010. Escherichia coli rpoB

Mutants Have Increased Evolvability in Proportion to Their Fitness Defects. Molecular Biology and Evolution. 27:1338–1347. doi: 10.1093/molbev/msq024.

Barroso-Batista J et al. 2014. The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps Coop, G, editor. PLoS Genet. 10:e1004182. doi: 10.1371/journal.pgen.1004182.s016.

Bateson W. 1910. Mendels principles of heredity. Molecular and General Genetics MGG.

Baxt LA, Garza-Mayers AC, Goldberg MB. 2013. Bacterial Subversion of Host Innate Immune Pathways. Science. 340:697–701. doi: 10.1126/science.1235771.

Bhullar K et al. 2012. Antibiotic resistance is prevalent in an isolated cave microbiome. PLoS ONE. 7:e34953. doi: 10.1371/journal.pone.0034953.

Björkman J, Hughes D, Andersson DI. 1998. Virulence of antibiotic-resistant Salmonella typhimurium. Proceedings of the National Academy of Sciences. 95:3949–3953.

Björkman J, Nagaev I, Berg OG, Hughes D, Andersson DI. 2000. Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. Science. 287:1479–1482.

Blair JMA, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJV. 2014. Molecular mechanisms ofantibiotic resistance. Nature Reviews Microbiology. 13:42–51. doi: 10.1038/nrmicro3380.

Blank D, Wolf L, Ackermann M, Silander OK. 2014. The predictability of molecular evolution during functional innovation. Proceedings of the National Academy of Sciences. 111:3044–3049. doi: 10.1073/pnas.1318797111.

Borrell S et al. 2013. Epistasis between antibiotic resistance mutations drives the evolution of extensively drug-resistant tuberculosis. Evolution, Medicine, and Public Health. 2013:65–74. doi: 10.1093/emph/eot003.

Bouma JE, Lenski RE. 1988. Evolution of a bacteria/plasmid association. Nature. 335:351–352. doi: 10.1038/335351a0.

Brandis G, Hughes D. 2013. Genetic characterization of compensatory evolution in strains carrying rpoB Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates. J. Antimicrob. Chemother. 68:2493–2497. doi:

10.1093/jac/dkt224.

Brandis G, Wrande M, Liljas L, Hughes D. 2012. Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. Molecular Microbiology. 85:142–151. doi: 10.1111/j.1365-2958.2012.08099.x.

Brown ED, Wright GD. 2016. Antibacterial drug discovery in the resistance era. Nature. 529:336–343. doi: 10.1038/nature17042.

Bucci V, Xavier JB. 2014. Towards Predictive Models of the Human Gut Microbiome. Journal of Molecular Biology. 1–10. doi: 10.1016/j.jmb.2014.03.017.

Buckling A, Rainey PB. 2002. Antagonistic coevolution between a bacterium and a bacteriophage.

Bull JJ, Badgett MR, Rokyta D, Molineux IJ. 2003. Experimental evolution yields hundreds of mutations in a functional viral genome. J Mol Evol.

Celli J, Finlay BB. 2002. Bacterial avoidance of phagocytosis. Trends in Microbiology.

Chakrabarti SL, Gorini L. 1977. Interaction between mutations of ribosomes and RNA polymerase: a pair of strA and rif mutants individually temperature-insensitive but temperature-sensitive in combination.

Chedom DF, Murcia PR, Greenman CD. 2015. Inferring the Clonal Structure of Viral Populations from Time Series Sequencing Kosakovsky Pond, SL, editor. PLoS Comput Biol. 11:e1004344. doi: 10.1371/journal.pcbi.1004344.s014.

Comas I et al. 2012. Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nature Genetics. 44:106–110. doi: 10.1038/ng.1038.

Comeron JM, Kreitman M. 2002. Population, evolutionary and genomic consequences of interference selection. Genetics. 161:389–410.

Davies J. 1994. Inactivation of antibiotics and the dissemination of resistance genes. Science. 264:375–382.

Davies J, Davies D. 2010. Origins and Evolution of Antibiotic Resistance. Microbiology and Molecular Biology Reviews. 74:417–433. doi: 10.1128/MMBR.00016-10.

de Visser JAGM. 2005. Clonal Interference and the Periodic Selection of New Beneficial Mutations in Escherichia coli. Genetics. 172:2093–2100. doi: 10.1534/genetics.105.052373.

DeLong EF. 2005. Microbial community genomics in the ocean. Nature Reviews Microbiology. 3:459–469. doi: 10.1038/nrmicro1158.

Denver DR et al. 2012. Variation in Base-Substitution Mutation in Experimental and Natural Lineages of Caenorhabditis Nematodes. Genome Biology and Evolution. 4:513–522. doi: 10.1093/gbe/evs028.

Desai MM, Fisher DS. 2007. Beneficial Mutation Selection Balance and the Effect of Linkage on Positive Selection. Genetics. 176:1759–1798. doi: 10.1534/genetics.106.067678.

Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC. 2012. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature.

Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. Genetics. 148:1667–1686.

Durão P, Trindade S, Sousa A, Gordo I. 2015. Multiple Resistance at No Cost: Rifampicin and Streptomycin a Dangerous Liaison in the Spread of Antibiotic Resistance. Molecular Biology and Evolution. 32:2675–2680. doi: 10.1093/molbev/msv143.

Ebert D, Bull JJ. 2008. The evolution and expression of virulence. Evolution in health and disease.

Eldholm V et al. 2014. Evolution of extensively drug-resistant Mycobacterium tuberculosis from a susceptibleancestor in a single patient. 1–11. doi: 10.1186/s13059-014-0490-3.

Ernst RK, Guina T, Miller SI. 1999. How intracellular bacteria survive: surface modifications that promote resistance to host innate immune responses. Journal of Infectious Diseases.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nature Publishing Group. 8:610–618. doi: 10.1038/nrg2146.

Farhat MR et al. 2013. Genomic analysis identifies targets of convergent positive

selection in drug-resistant Mycobacterium tuberculosis. Nature Genetics. 45:1183–1189. doi: 10.1038/ng.2747.

Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics.

Fenster CB, Galloway LF, Chao L. 1997. Epistasis and its consequences for the evolution of natural populations. Trends in Ecology & Evolution.

Fields KA, Heinzen RA, Carabeo R. 2011. The obligate intracellular lifestyle. Front Microbiol.

Fisher RA. 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Fitzpatrick MJ, Feder E, Rowe L, Sokolowski MB. 2007. Maintaining a behaviour polymorphism by frequency-dependent selection on a single gene. Nature.

Fleming A. 1945. *Penicillin. Nobel Lecture, December 11, 1945*. Nobel e-museum.

Forsberg KJ et al. 2012. The shared antibiotic resistome of soil bacteria and human pathogens. Science. 337:1107–1111. doi: 10.1126/science.1220761.

Fredrickson AG, Stephanopoulos G. 1981. Microbial competition. Science.

Frenkel EM, Good BH, Desai MM. 2014. The fates of mutant lineages and the distribution of fitness effects of beneficial mutations in laboratory budding yeast populations. Genetics. doi: 10.1534/genetics.113.160069/-/DC1.

Gerrish PJ, Lenski RE. 1998. The fate of competing beneficial mutations in an asexual population. Genetica. 102-103:127–144.

Gifford DR, MacLean RC. 2013. Evolutionary reversals of antibiotic resistance in experimental populations of Pseudomonas aeruginosa. Evolution. 67:2973–2981. doi: 10.1111/evo.12158.

Gluck-Thaler E, Slot JC. 2015. Dimensions of Horizontal Gene Transfer in Eukaryotic Microbial Pathogens Hogan, DA, editor. PLoS Pathog. 11:e1005156. doi: 10.1371/journal.ppat.1005156.s001.

Gordo I, Demengeot J, Xavier K. 2014. Escherichia coliadaptation to the gut environment: a constant fight for survival. Future Microbiology. 9:1235–1238. doi: 10.2217/fmb.14.86.

Gordon DM. 2014. The ecology of collective behavior. PLoS Biol. 12:e1001805. doi: 10.1371/journal.pbio.1001805.

Gordon JI. 2012. Honor thy gut symbionts redux. Science. 336:1251–1253. doi: 10.1126/science.1224686.

Greaves M, Maley CC. 2012. Clonal evolution in cancer. Nature. 481:306–313. doi: 10.1038/nature10762.

Gros PA, Le Nagard H, Tenaillon O. 2009. The evolution of epistasis and its links with genetic robustness, complexity and drift in a phenotypic model of adaptation. Genetics.

Gullberg E, Cao S, Berg OG, Ilbäck C, Sandegren L. 2011. Selection of resistant bacteria at very low antibiotic concentrations. PLoS Pathog.

Haldane J. 1927. A mathematical theory of natural and artificial selection. Proc. Camb. Philos. Soc.

Hall AR, Iles JC, MacLean RC. 2011. The fitness cost of rifampicin resistance in Pseudomonas aeruginosa depends on demand for RNA polymerase. Genetics.

Hall AR, MacLean RC. 2011. Epistasis buffers the fitness effects of rifampicin-resistance mutations in Pseudomonas aeruginosa. Evolution. 65:2370–2379. doi: 10.1111/j.1558-5646.2011.01302.x.

Hare R. 1982. New light on the history of penicillin. Medical history.

Hede K. 2014. Antibiotic resistance: An infectious arms race. Nature. 509:S2–3. doi: 10.1038/509S2a.

Hegreness M. 2006. An Equivalence Principle for the Incorporation of Favorable Mutations in Asexual Populations. Science. 311:1615–1617. doi: 10.1126/science.1122469.

Helaine S et al. 2010. Dynamics of intracellular bacterial replication at the single cell level. Proc. Natl. Acad. Sci. U.S.A. 107:3746–3751. doi: 10.1073/pnas.1000041107.

Hibbing ME, Fuqua C, Parsek MR, Peterson SB. 2010. Bacterial competition: surviving and thriving in the microbial jungle. Nature Reviews Microbiology. 8:15–25. doi: 10.1038/nrmicro2259.

Hietpas RT, Jensen JD, Bolon DNA. 2011. Experimental illumination of a fitness landscape. Proc. Natl. Acad. Sci. U.S.A. 108:7896–7901. doi: 10.1073/pnas.1016024108.

Hindré T, Knibbe C, Beslon G, Schneider D. 2012. New insights into bacterial adaptationthrough in vivo and in silicoexperimental evolution. Nature Reviews Microbiology. 10:352–365. doi: 10.1038/nrmicro2750.

Holland JH. 1995. *Hidden order: How adaptation builds complexity*.

Hughes D, Andersson DI. 2015. Evolutionary consequences of drug resistance: shared principles across diverse targets and organisms. Nature Publishing Group. 16:459–471. doi: 10.1038/nrg3922.

Hybiske K, Stephens RS. 2008. Exit strategies of intracellular pathogens. Nature Reviews Microbiology. 6:99–110. doi: 10.1038/nrmicro1821.

Iwasaki W, Takagi T. 2009. Rapid Pathway Evolution Facilitated by Horizontal Gene Transfers across Prokaryotic Lineages Matic, I, editor. PLoS Genet. 5:e1000402. doi: 10.1371/journal.pgen.1000402.s001.

Jacoby GA, Archer GL. 1991. New mechanisms of bacterial resistance to antimicrobial agents. N. Engl. J. Med. 324:601–612. doi: 10.1056/NEJM199102283240906.

Jacquier H et al. 2013. Capturing the mutational landscape of the beta-lactamase TEM-1. Proceedings of the National Academy of Sciences. 110:13067–13072. doi: 10.1073/pnas.1215206110/-/DCSupplemental.

Kauffman S, Levin S. 1987. Towards a general theory of adaptive walks on rugged landscapes. Journal of Theoretical Biology.

Kawecki TJ et al. 2012. Experimental evolution. Trends in Ecology & Evolution. 27:547–560. doi: 10.1016/j.tree.2012.06.001.

Kelley VA, Schorey JS. 2003. Mycobacterium's arrest of phagosome maturation in macrophages requires Rab5 activity and accessibility to iron. Molecular biology of the cell.

Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. 2011. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. Science. 332:1193–1196. doi: 10.1126/science.1203801.

Kibota TT, Lynch M. 1996. Estimate of the genomic mutation rate deleterious to overall fitness in E. coli. Nature.

Kim S, Lieberman TD, Kishony R. 2014. Alternating antibiotic treatments constrain evolutionary paths to multidrug resistance. Proc. Natl. Acad. Sci. U.S.A. 111:14494–14499. doi: 10.1073/pnas.1409800111.

Kimura M. 1962. On the probability of fixation of mutant genes in a population. Genetics. 47:713–719.

Kolodynska A, Pigliucci M. 2003. Multivariate responses to flooding in Arabidopsis: an experimental evolutionary investigation. Functional Ecology.

Kondrashov AS. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature.

Kriegeskorte A, Peters G. 2012. Horizontal gene transfer boosts MRSA spreading. Nat Med.

Kurland CG, Hughes D, Ehrenberg M. 1996. *Limitations of translational accuracy*.

Kvitek DJ, Sherlock G. 2011. Reciprocal Sign Epistasis between Frequently Experimentally Evolved Adaptive Mutations Causes a Rugged Fitness Landscape Zhang, J, editor. PLoS Genet. 7:e1002056. doi: 10.1371/journal.pgen.1002056.s010.

Lanfear R, Kokko H, Eyre-Walker A. 2013. Population size and the rate of evolution. Trends in Ecology & Evolution. 1–9. doi: 10.1016/j.tree.2013.09.009.

Lang GI et al. 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature. 1–6. doi: 10.1038/nature12344.

Lang GI, Desai MM. 2014. The spectrum of adaptive mutations in experimental evolution. Genomics. 104:412–416. doi: 10.1016/j.ygeno.2014.09.011.

Lau HY, Huffnagle GB, Moore TA. 2008. Host and microbiota factors that control Klebsiella pneumoniae mucosal colonization in mice. Microbes and Infection.

Lázár V et al. 2014. Genome-wide analysis captures the determinants of the antibiotic cross-resistance interaction network. Nature Communications. 5:4352. doi: 10.1038/ncomms5352.

Le Negrate G. 2012. Subversion of innate immune responses by bacterial hindrance of NF-κB pathway. Cellular Microbiology.

Lee H, Popodi E, Tang H. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing.

Leimbach A, Hacker J, Dobrindt U. 2013. E. coli as an all-rounder: the thin line between commensalism and pathogenicity. Curr. Top. Microbiol. Immunol. 358:3–32. doi: 10.1007/82_2012_303.

Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. American Naturalist.

Levin BR et al. 1997. The population genetics of antibiotic resistance. Clin Infect Dis. 24 Suppl 1:S9–16.

Levin BR, Perrot V, Walker N. 2000. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. Genetics. 154:985–997.

Levine DP. 2006. Vancomycin: a history. Clin Infect Dis.

Li B, Yang R. 2008. Interaction between Yersinia pestis and the host immune system. Infection and Immunity.

Littman DR, Pamer EG. 2011. Role of the commensal microbiota in normal and pathogenic host immune responses. Cell Host & Microbe.

Lu HP, Lai YC, Huang SW, Chen HC, Hsieh C. 2014. Spatial heterogeneity of gut microbiota reveals multiple bacterial communities with distinct characteristics. Sci. Rep.

Luangtongkum T et al. 2012. Impaired fitness and transmission of macrolide-resistant Campylobacter jejuni in its natural host. Antimicrobial Agents and Chemotherapy. 56:1300–1308. doi: 10.1128/AAC.05516-11.

Luzzatto L, Apirion D, Schlessinger D. 1968. Mechanism of action of streptomycin in E. coli: interruption of the ribosome cycle at the initiation of protein synthesis. Proceedings of the National Academy of Sciences. 60:873–880.

MacLean RC, Vogwill T. 2015. Limits to compensatory adaptation and the persistence of antibiotic resistance in pathogenic bacteria. Evolution, Medicine, and Public Health. 2015:4–12. doi: 10.1093/emph/eou032.

Macpherson AJ, Geuking MB, McCoy KD. 2005. Immune responses that adapt the intestinal mucosa to commensal intestinal bacteria. Immunology. 115:153–162. doi: 10.1111/j.1365-2567.2005.02159.x.

Maddamsetti R, Lenski RE, Barrick JE. 2015. Adaptation, Clonal Interference, and Frequency-Dependent Interactions in a Long-Term Evolution Experiment with Escherichia coli. Genetics. 200:619–631. doi: 10.1534/genetics.115.176677.

Maiden M. 1998. Horizontal genetic exchange, evolution, and spread of antibiotic resistance in bacteria. Clin Infect Dis.

Maisnier-Patin S, Berg OG, Liljas L, Andersson DI. 2002. Compensatory adaptation to the deleterious effect of antibiotic resistance in Salmonella typhimurium. Molecular Microbiology. 46:355–366.

Maisnier-Patin S, Paulander W, Pennhag A, Andersson DI. 2007. Compensatory evolution reveals functional interactions between ribosomal proteins S12, L14 and L19. Journal of Molecular Biology. 366:207–215. doi: 10.1016/j.jmb.2006.11.047.

Mariam SH, Werngren J, Aronsson J, Hoffner S, Andersson DI. 2011. Dynamics of Antibiotic Resistant Mycobacterium tuberculosis during Long-Term Infection and Antibiotic Treatment Gagneux, S, editor. PLoS ONE. 6:e21147. doi: 10.1371/journal.pone.0021147.s002.

Martin G, Elena SF, Lenormand T. 2007. Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nature Genetics.

Martinez JL, Baquero F. 2000. Mutation frequencies and antibiotic resistance. Antimicrobial Agents and Chemotherapy.

Mauel J. 1982. Macrophage activation and effector mechanisms against microbes. Macrophages and Natural Killer Cells.

Medzhitov R. 2007. Recognition of microorganisms and activation of the immune response. Nature.

Melnyk AH, Wong A, Kassen R. 2015. The fitness costs of antibiotic resistance mutations. Evol Appl.

Miralles R, Gerrish PJ, Moya A, Elena SF. 1999. Clonal interference and the evolution of RNA viruses. Science.

Navarro Y et al. 2011. Systematic survey of clonal complexity in tuberculosis at a populational level and detailed characterization of the isolates involved. J. Clin. Microbiol. 49:4131–4137. doi: 10.1128/JCM.05203-11.

Novella IS et al. 1995. Exponential increases of RNA virus fitness during large population transmissions. Proceedings of the National Academy of Sciences. 92:5841–5844.

O'Neill AJ, Cove JH, Chopra I. 2001. Mutation frequencies for resistance to fusidic acid and rifampicin in Staphylococcus aureus. Journal of Antimicrobial ….

Orr HA. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution. 935–949.

Palmer AC, Kishony R. 2013. Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. Nature Publishing Group. 14:243–248. doi: 10.1038/nrg3351.

Pankey GA, Sabath LD. 2004. Clinical relevance of bacteriostatic versus bactericidal mechanisms of action in the treatment of Gram-positive bacterial infections. Clin Infect Dis.

Paulander W, Maisnier-Patin S, Andersson DI. 2007. Multiple mechanisms to ameliorate the fitness burden of mupirocin resistance in Salmonella typhimurium. Molecular Microbiology. 64:1038–1048. doi: 10.1111/j.1365-2958.2007.05713.x.

Paulsen IT et al. 2003. Role of mobile DNA in the evolution of vancomycin-resistant Enterococcus faecalis. Science. 299:2071–2074. doi: 10.1126/science.1080613.

Perfeito L, Fernandes L, Mota C, Gordo I. 2007. Adaptive Mutations in Bacteria: High Rate and Small Effects. Science. 317:813–815. doi: 10.1126/science.1142284.

Perfeito L, Sousa A, Bataillon T, Gordo I. 2014. Rates of fitness decline and rebound suggest pervasive epistasis. Evolution.

Perron GG, Zasloff M, Bell G. 2006. Experimental evolution of resistance to an antimicrobial peptide. Proc. Biol. Sci. 273:251–256. doi: 10.1098/rspb.2005.3301.

Phillips PC. 2008. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nature Publishing Group. 9:855–867. doi: 10.1038/nrg2452.

Poon A, Chao L. 2005. The rate of compensatory mutation in the DNA bacteriophage φX174. Genetics.

Ratcliff WC, Fankhauser JD, Rogers DW, Greig D, Travisano M. 2015. Origins of multicellular evolvability in snowflake yeast. Nature Communications. 6:6102. doi: 10.1038/ncomms7102.

Reynolds MG. 2000. Compensatory evolution in rifampin-resistant Escherichia coli. Genetics. 156:1471–1481.

Rodríguez-Verdugo A, Gaut BS, Tenaillon O. 2013. Evolution of Escherichia coli rifampicin resistance in an antibiotic-free environment during thermal stress. BMC Evol Biol. 13:50. doi: 10.1186/1471-2148-13-50.

Romero D, Traxler MF, López D, Kolter R. 2011. Antibiotics as signal molecules. Chemical reviews.

Rosenberger CM, Finlay BB. 2003. Phagocyte sabotage: disruption of macrophage signalling by bacterial pathogens. Nature reviews Molecular cell biology.

Rowlands BJ, Soong CV, Gardiner KR. 1999. The gastrointestinal tract as a barrier in sepsis. British medical bulletin.

Rozen DE, Philippe N, Arjan de Visser J, Lenski RE, Schneider D. 2009. Death and cannibalism in a seasonal environment facilitate bacterial coexistence. Ecol Lett. 12:34–44. doi: 10.1111/j.1461-0248.2008.01257.x.

Saka HA, Valdivia RH. 2010. Acquisition of nutrients by Chlamydiae: unique challenges of living in an intracellular compartment. Current Opinion in Microbiology.

Schrag SJ, Perrot V. 1996. *Reducing antibiotic resistance.* Nature.

Schrag SJ, Perrot V, Levin BR. 1997. Adaptation to the fitness costs of antibiotic resistance in Escherichia coli. Proceedings of the Royal Society B: Biological Sciences. 264:1287–1291. doi: 10.1098/rspb.1997.0178.

Sezonov G, Joseleau-Petit D, D'Ari R. 2007. Escherichia coli Physiology in Luria-Bertani Broth. Journal of Bacteriology. 189:8746–8749. doi: 10.1128/JB.01368-07.

Silva MT, Pestana NTS. 2013. The in vivo extracellular life of facultative intracellular bacterial parasites: role in pathogenesis. Immunobiology. 218:325–337. doi: 10.1016/j.imbio.2012.05.011.

Simões P et al. 2008. How repeatable is adaptive evolution? The role of geographical origin and founder effects in laboratory adaptation. Evolution. 62:1817–1829. doi: 10.1111/j.1558-5646.2008.00423.x.

Slauch JM. 2011. How does the oxidative burst of macrophages kill bacteria? Still an open question. Molecular Microbiology. 80:580–583. doi: 10.1111/j.1365-2958.2011.07612.x.

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23:23–35.

Sniegowski PD, Gerrish PJ. 2010. Beneficial mutations and the dynamics of adaptation in asexual populations. Philosophical Transactions of the Royal Society B: Biological Sciences. 365:1255–1263. doi: 10.1126/science.285.5426.422.

Sniegowski PD, Gerrish PJ, Lenski RE. 1997. Evolution of high mutation rates in experimental populations of E. coli. Nature. 387:703–705. doi: 10.1038/42701.

Sokurenko EV, Hasty DL, Dykhuizen DE. 1999. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. Trends in Microbiology. 7:191–195.

Spratt BG. 1994. Resistance to antibiotics mediated by target alterations. Science.

St Onge RP et al. 2007. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nature Genetics. 39:199–206. doi: 10.1038/ng1948.

Stickland HG, Davenport PW, Lilley KS, Griffin JL, Welch M. 2010. Mutation of nfxB causes global changes in the physiology and metabolism of Pseudomonas aeruginosa. J. Proteome Res. 9:2957–2967. doi: 10.1021/pr9011415.

Strelkowa N, Lässig M. 2012. Clonal interference in the evolution of influenza. Genetics.

Sun G et al. 2012. Dynamic Population Changes in Mycobacterium tuberculosis

During Acquisition and Fixation of Drug Resistance in Patients. Journal of Infectious Diseases. 206:1724–1733. doi: 10.1093/infdis/jis601.

Szamecz B et al. 2014. The Genomic Landscape of Compensatory Evolution Barton, NH, editor. PLoS Biol. 12:e1001935. doi: 10.1371/journal.pbio.1001935.s018.

Telenti A, Imboden P, Marchesi F. 1993. Direct, automated detection of rifampin-resistant Mycobacterium tuberculosis by polymerase chain reaction and single-strand conformation polymorphism analysis. Antimicrobial agents ….

Tenover FC. 2006. Mechanisms of antimicrobial resistance in bacteria. Am. J. Med.

Touchon M et al. 2012. Antibiotic resistance plasmids spread among natural isolates of Escherichia coli in spite of CRISPR elements. Microbiology. 158:2997–3004. doi: 10.1099/mic.0.060814-0.

Touchon M, Bobay L-M, Rocha EP. 2014. The chromosomal accommodation and domestication of mobile genetic elements. Current Opinion in Microbiology. 22:22–29. doi: 10.1016/j.mib.2014.09.010.

Trindade S et al. 2009. Positive Epistasis Drives the Acquisition of Multidrug Resistance Zhang, J, editor. PLoS Genet. 5:e1000578. doi: 10.1371/journal.pgen.1000578.s005.

Turner CB, Blount ZD, Lenski RE. 2015. Replaying Evolution to Test the Cause of Extinction of One Ecotype in an Experimentally Evolved Population. PLoS ONE.

Underhill DM, Ozinsky A. 2002. Phagocytosis of microbes: complexity in action. Annu. Rev. Immunol.

van Hoek AHAM et al. 2011. Acquired antibiotic resistance genes: an overview. Front Microbiol. 2:203. doi: 10.3389/fmicb.2011.00203.

Velayati AA et al. 2009. Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran. Chest. 136:420–425. doi: 10.1378/chest.08-2427.

Vieira OV, Botelho RJ, Grinstein S. 2002. Phagosome maturation: aging gracefully. Biochem. J. 366:689–704. doi: 10.1042/BJ20020691.

Wahl LM, Gerrish PJ. 2001. The probability that beneficial mutations are lost in populations with periodic bottlenecks. Evolution. 55:2606–2610.

Wang Z, Liao B-Y, Zhang J. 2010. Genomic patterns of pleiotropy and the evolution of complexity. Proc. Natl. Acad. Sci. U.S.A. 107:18034–18039. doi: 10.1073/pnas.1004666107.

Wehrli W. 1983. Rifampin: mechanisms of action and resistance. Review of Infectious Diseases.

Weinreich DM, Delaney NF, Depristo MA, Hartl DL. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science. 312:111–114. doi: 10.1126/science.1123539.

Wilson BA, Garud NR, Feder AF, Assaf ZJ, Pennings PS. 2016. The population genetics of drug resistance evolution in natural populations of viral, bacterial and eukaryotic pathogens. Mol Ecol. 25:42–66. doi: 10.1111/mec.13474.

Wiser MJ, Ribeck N, Lenski RE. 2013. Long-term dynamics of adaptation in asexual populations. Science. 342:1364–1367. doi: 10.1126/science.1243357.

Woo HJ, Reifman J. 2014. Quantitative Modeling of Virus Evolutionary Dynamics and Adaptation in Serial Passages Using Empirically Inferred Fitness Landscapes. Journal of Virology.

Woods RJ et al. 2011. Second-Order Selection for Evolvability in a Large Escherichia coli Population. Science. 331:1433–1436. doi: 10.1126/science.1198914.

World Health Organization. 2014. Antimicrobial Resistance - Global Report on Surveillance. http://apps.who.int/iris/bitstream/10665/112642/1/9789241564748_eng.pdf.

Wright GD. 2010. The antibiotic resistome. Expert Opin. Drug Discov. 5:779–788. doi: 10.1517/17460441.2010.497535.

Wright S. 1932. *The roles of mutation, inbreeding, crossbreeding, and selection in evolution*.

Xavier JB, Foster KR. 2007. Cooperation and conflict in microbial biofilms. Proceedings of the National Academy of Sciences. 104:876–881. doi: 10.1073/pnas.0607651104.

Yamori S, Ichiyama S, Shimokata K, Tsukamura M. 1992. Bacteriostatic and bactericidal activity of antituberculosis drugs against Mycobacterium tuberculosis, Mycobacterium avium-Mycobacterium intracellulare complex and Mycobacterium kansasii in different growth phases. Microbiol. Immunol. 36:361–368.

Zanini F et al. 2016. Population genomics of intrapatient HIV-1 evolution. eLife. doi: 10.7554/eLife.11282.001.

Zhang H et al. 2013. Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nature Genetics. 45:1255–1260. doi: 10.1038/ng.2735.

Zhang HP, Be'er A, Smith RS, Florin EL, Swinney HL. 2009. Swarming dynamics in bacterial colonies. Europhys. Lett. 87:48011. doi: 10.1209/0295-5075/87/48011.

CHAPTER II

An ABC Method for Estimating the Rate and Distribution of Effects of Beneficial Mutations

Manuscript published in ***Genome Biology and Evolution***, May 1st 2013

The author of this thesis designed the research, developed the algorithms and performed all the simulations described in this chapter.

An ABC method for estimating the rate and distribution of effects of beneficial mutations

Jorge A. Moura de Sousa[1], Paulo R. A. Campos[2] and Isabel Gordo[1]*

[1] Instituto Gulbenkian de Ciência, Rua da Quinta Grande, PT-2780-156 Oeiras, Portugal

[2] Departamento de Física, Universidade Federal de Pernambuco, Cidade Universitaria, 50670-901, Recife PE, Brazil

*Author for Correspondence: igordo@igc.gulbenkian.pt

Abstract

Determining the distribution of adaptive mutations available to natural selection is a difficult task. These are rare events and most of them are lost by chance. Some theoretical works propose that the distribution of newly arising beneficial mutations should be close to exponential. Empirical data are scarce and do not always support an exponential distribution. Analysis of the dynamics of adaptation in asexual populations of microorganisms has revealed that these can be summarized by two effective parameters, the effective mutation rate U_e and the effective selection coefficient of a beneficial mutation S_e. Here we show that these effective parameters will not always reflect the rate and mean effect of beneficial mutations, especially when the distribution of arising mutations has high variance, and the mutation rate is high. We propose a method to estimate the distribution of arising beneficial mutations, which is motivated by a common experimental setup.

The method, which we call One Bi-allelic Marker ABC, makes use of experimental data consisting of periodic measures of neutral marker frequencies and mean population fitness. Using simulations, we find that this method allows the discrimination of the shape of the distribution of arising mutations and that it provides reasonable estimates of their rates and mean effects in ranges of the parameter space that may be of biological relevance.

Keywords

Experimental Evolution, Mutation Rate, Distribution of Fitness Effects, Parameter Estimation

Introduction

At what rate do beneficial mutations arise and what are their fitness effects? These are two of the most important questions regarding adaptation of organisms to novel environments (Kimura and Ohta 1974; Lang, Botstein, et al. 2011). Reflecting its importance, estimating genomic mutation rates of new beneficial alleles (U) and uncovering the mean effects of those beneficial mutations (E(S)) have been the subject of many studies (Bataillon, Zhang, et al. 2011; Estes, Phillips, et al. 2011; MacLean and Buckling 2009; Perfeito, Fernandes, et al. 2007; Rozen, de Visser, et al. 2002; Sawyer, Parsch, et al. 2007; Sousa, Magalhães, et al. 2012). Experimental evolution in clonal populations presents some advantages in determining these parameters, but some difficulties still arise, even in these controlled and relatively

simple environments. One of these difficulties is being able to assay all the beneficial mutations. Different distributions of fitness effects are important to the adaptive process: the distribution of newly arising mutations, the distribution of contending mutations, which escape initial stochastic loss, and the distribution of mutations that survive competition with other mutations (clonal interference) and are able to actually fix, contributing to long term adaptation (see (Gordo, Perfeito, et al. 2011) for a review). The greatest difficulty is to uncover the distribution of arising mutations, because they may easily be lost before reaching detectable frequencies. Despite this difficulty, determining the distribution that characterizes arising mutations, $f(S)$, is important, because it is this distribution that determines the nature of adaptation (Orr 2010; Perfeito, Fernandes, et al. 2007; Rozen, de Visser, et al. 2002; Sousa, Magalhães, et al. 2012). For this reason, some studies have tried to determine this distribution in viruses (Rokyta, Beisel, et al. 2008; Sanjuán, Moya, et al. 2004), in bacteria (Kassen and Bataillon 2006; Stevens and Sebert 2011) and in other organisms (Burke, Dunham, et al. 2010; Desai, Fisher, et al. 2007; Orozco-terWengel, Kapun, et al. 2012; Schoustra, Bataillon, et al. 2009). Experimental support for an exponential distribution of arising beneficial mutations has been obtained (Kassen and Bataillon 2006; MacLean and Buckling 2009), but this has not always been the case in all organisms and environments (Barrett, MacLean, et al. 2006; Bataillon, Zhang, et al. 2011; Gordo, Perfeito, et al. 2011; Mcdonald, Cooper, et al. 2011; Rokyta, Beisel, et al. 2008). From the mutations that arise, those that end up outcompeting other beneficial mutations will drive long-term adaptation (Gerrish and Lenski 1998; Good, Rouzine, et al. 2012). The difference between the distributions of

arising, contending and fixed mutations is expected to depend on the effective population size (Crow and Kimura 1970), the mutation rate (Charlesworth, Morgan, et al. 1993) and on the level of mal-adaptation, with increasingly adapted organisms having access to increasingly lower amount of beneficial mutations (Fisher 1930),

The biggest challenge in determining *f*(S) lies in the rarity of beneficial mutations. In principle, this distribution can be determined directly by measuring fitness effects of extremely large samples of mutants (Hietpas, Jensen, et al. 2011; Lind, Berg, et al. 2010). It can also be inferred from sequence data collected from natural populations (Eyre-Walker and Keightley 2007; Jensen, Thornton, et al. 2008a; Jensen, Thornton, et al. 2008b; Nielsen 2005; Schneider, Charlesworth, et al. 2011). Indeed, scans for signatures of positive selection across the genome of different species, including our own, have been performed (Biswas and Akey 2006; Cutter and Choi 2010; Enard, Depaulis, et al. 2010; Grossman, Andersen, et al. 2013; Hancock and Di Rienzo 2008). Disentangling the signature of selection from that caused by a complex demography is difficult (Grossman, Shlyakhter, et al. 2010; Sinha, Dincer, et al. 2011), and checking the performance of different methods under departures from model assumptions is therefore an important task (Keightley and Eyre-Walker 2010). Recent advances have been made in developing methods for estimating selection coefficients from time series data of allele frequencies (Bollback, York, et al. 2008; Malaspinas, Malaspinas, et al. 2012; Mathieson and McVean 2013) and also in disentangling alleles under positive selection from passenger mutations (Illingworth and Mustonen 2011). In the context of experimentally evolved populations, where typically the experimenter imposes a particular demographic regime, one method that

has been used proposes to study beneficial mutations through assaying the evolutionary dynamics of neutral markers in asexual populations (Hegreness, Shoresh, et al. 2006; Imhof and Schlotterer 2001). The basic principle underlying this method relies on the "hitchhiking effect" of a neutral allele with mutations that give an advantage to the organism (Maynard-Smith and Haigh 1974). This same principle is at the heart of methods to detect positive selection across the genome of sexually reproducing organisms (Thornton, Jensen, et al. 2007). In experimentally evolved populations, the frequency of a neutral allele can be easily measured (e.g. by using neutral fluorescent markers), and inferring evolutionary parameters from neutral marker dynamics can thus be performed under certain theoretical assumptions (Barrick, Kauth, et al. 2010; Dykhuizen and Hartl 1983; Hegreness, Shoresh, et al. 2006; Illingworth and Mustonen 2012). A simple and quite elegant method was proposed by Hegreness et al. (Hegreness, Shoresh, et al. 2006): using simulations, they showed that a simple population genetics model, where all beneficial mutations have the same effect, is able to reproduce the dynamics of a commonly used marker system involving one locus with two neutral markers. The dynamics can therefore be summarized by 2 parameters that theoretically represent the evolutionary process: the effective mutation rate (U_e) and the effective selection coefficient (S_e). Barrick et al. extended this method and determined the values of U_e and S_e in different strains of *Escherichia coli* (Barrick, Kauth, et al. 2010; Woods, Barrick, et al. 2011). While it may be useful to be able to summarize the process under a single mutational effect, far more realistic distributions of fitness effects can also explain the data. Recently, Illingworth and Mustonen (Illingworth and Mustonen 2012) proposed a new method to

estimate the distribution of haplotype fitnesses in experimentally evolving populations. When tested against simulated data under the assumption of an exponential distribution of arising beneficial mutations, the method is able to retrieve the correct distribution of haplotype fitnesses for values U below 10^{-6}. It is however not known how the method performs for other distributions of arising beneficial mutations and for larger values of U. Moreover, this method estimates the distribution of haplotype fitnesses segregating in populations and not the distribution of beneficial arising mutations. Here we ask two questions: how do the effective parameters compare with the more biologically meaningful parameters U and E(S)?; and, since frequency dynamics appear insufficient to distinguish between different distributions (Hegreness, Shoresh, et al. 2006), is there a reasonable set of data that can be obtained, which allows the determination of the distribution of arising beneficial mutations?

We address both these questions from a theoretical perspective, taking a commonly used experimental setup to study the adaptation of asexual populations in controlled environments as a motivation. This setup simply involves tracking a marker locus with two neutral alleles. We show that the effective evolutionary parameters can provide good estimates of U and of the mean effect of beneficial mutations only when the distribution of effects of arising mutations has limited variance. However, when the variance is increased (for example, if arising mutations follow an exponential distribution), we find that U_e can underestimate the true value U, while S_e can overestimate the true value of E(S). We propose a new method based upon measurements of both the frequency of neutral markers and mean population fitness, at periodic time intervals. This method, which was motivated by typical experimental

setups easily applied to experimental evolution, is theoretically expected to estimate the mutation rate reasonably well, and allows distributions of arising beneficial mutations with different shapes to be distinguished.

Methods

Model of adaptation to simulate evolutionary dynamics

We assume a clonal population reproducing according to the Wright-Fisher model, where periodic bottlenecks occur (with a period of T_{bot}). The population is initially isogenic, with the exception of a neutral marker, which is biallelic and has a frequency f_0=0.5 for one of the alleles. The initial population size is N_0. Generations are discrete and the population doubles each generation for $t < T_{bot}$. With period T_{bot}, the population size is reduced by random sampling to N_0. This assumed demography, involving periodic bottlenecks where the number of individuals is fixed, is typical of most experimental setups, where daily passages of a sample of the population are performed, and population numbers are experimentally controlled. At each generation, mutations occur at a rate U per genome, following a Poisson distribution. All mutations are beneficial and the effects of each mutation (S) are drawn from a continuous distribution f(S). We allow for variation of the selective effects of arising mutations, assuming a Gamma distribution, with shape and scale parameters, α and β, respectively, implying a mean $E(S) = \alpha\beta$. Like other studies that previously proposed to estimate the distribution of arising deleterious mutations (Eyre-Walker and Keightley 2007; Keightley 1998), we have assumed a Gamma distribution

because it can have a wide range of shapes. Multiplicative fitness is assumed, so that the effects of mutations do not depend on the genetic background where they arise. This is obviously an oversimplification, since the distribution can change along the adaptive walk (Martin and Lenormand 2006; Sousa, Magalhães, et al. 2012), but we consider a short-term evolution scenario where U and *f*(S) may be assumed constant. Genetic drift is modeled by sampling, from a multinomial distribution, classes of individuals with the same fitness. The frequency dynamics of the neutral marker (f(t)), as well as the mean population fitness (w(t)), are followed. This model of adaptation is used to produce a set of simulated evolutionary dynamics, from which evolutionary parameters are estimated using different methods: a method developed by Hegreness et al. (and extended by Barrick *et al*), and a new method that we propose here that simultaneously tries to estimate U and f(S) (see below).

The range of parameters chosen to produce simulated data with the described model was made in accordance with current estimates in different systems, but mostly in microorganisms. U is currently estimated to achieve values between 10^{-4} and 10^{-9}, depending on the environment and genetic background (Denver, Wilhelm, et al. 2012; Drake, Charlesworth, et al. 1998; Lang, Botstein, et al. 2011; Perfeito, Fernandes, et al. 2007). An effective population size $N_e=10^5$ was assumed (corresponding to bottlenecks with a period $t_{bot}=5$ generations) for all simulations except when indicated differently.

Generating pseudo-observed data

Pseudo-observed data sets were generated under the model of adaptation described above, with a specific value of the mutation rate U and a specific Gamma distribution (with parameters α and β) with mean E(S). These data sets represent biological data that can be acquired in an experiment. The new method proposed here with the goal of estimating *f*(S) and U, requires the periodic measure of the frequency of the neutral markers and the mean population fitness (every 50 generations for a 300 generation experiment). These appear reasonable to obtain experimentally and require experimental work that is typical in evolution experiments performed in controlled environments: in addition to assaying the frequency of the markers (as already is typically done (Woods, Barrick, et al. 2011)), fitness has to be measured by performing either a direct competitive fitness assay against the ancestral strain or a measurement of the population growth rate at different times along the experiment (Gordo, Perfeito, et al. 2011). Furthermore, the choice of studying 100 replicate populations reflects the 100 or 96 well plate experimental setup that is commonly used (Kvitek and Sherlock 2011; Lemonnier, Levin, et al. 2008). These plates are affordable by most labs, and, with a multichannel pipet, several passages can be performed in little time, space, and at low cost, particularly when studying microbial populations. Regarding the markers, many strains expressing different fluorescent alleles are available, which makes the acquisition of frequency data a relatively easy task. This can be performed using flow cytometry or another fluorescence reader. Competitive fitness measurements (Elena and Lenski 2003) can also be easily performed using a similar setup.

The pseudo-observed data therefore consist of the marker frequencies and the fitnesses at periodic time points of the experiments (ti) for n independently evolved populations. Different pseudo-observed data sets assuming different distributions of S were generated to test the two different methods: Barrick et al. method (which requires the marker frequencies only), to compare U_e with U and S_e with E(S), and the One Bi-allelic Marker ABC (which requires both the marker frequencies and the fitnesses) to assess its ability to estimate U, α and β (see later).

Estimation of U and E(S) by U_e and S_e based upon the dynamics of the first significant deviation of f(t)

For a given set of pseudo-observed data we obtained the effective parameters U_e and S_e and compared them with the biologically meaningful values of U and E(S). To obtain U_e and S_e we followed Barrick at al. A large set of simulated evolutionary dynamics under the assumption that all beneficial mutations have the same value of S was generated. This simulated data consists of sets of 100 replicate populations evolved under different parameter combinations of U and S. The range of $\log_{10}(U)$ was [-8; -3.95], with increments of 0.15, and the range of S was [0.01; 0.18], with increments of 0.01. This simulated data is the input of Barrick et al. method (Barrick, Kauth, et al. 2010) to obtain U_e and S_e. For each simulation it consists of the logarithm of the ratio of the two subpopulation frequencies (Rf=f(ti)/(1-f(ti)) at several time points, ti, (ln(Rf(ti))), where ti=5*i. We then use this input in the program marker_divergence_fit.pl, whose output is fed into the program marker_divergence_significance.pl, both available at http://barricklab.org/twiki/bin/view/Lab/ToolsMarkerDivergence, to obtain U_e

and S_e. The first program summarizes the evolutionary dynamics (both in the simulated data sets and in the pseudo-observed data) in two statistics: τ_e and α_e. The first is the time, τ_e, where a significant deviation of $Ln(Rf(\tau_e))$ from $Ln(Rf(t=0))$ occurs. The second is the rate of change of $Ln(Rf(t))$ with time, i.e. τ_e sets the time of divergence of marker frequency and α_e the rate of divergence. Each replicate population is summarized by a single value of τ_e and α_e., and the n replicate populations (characterized by a given combination (U , S)) result in a distribution of $T(\tau_e)$ and $A(\alpha_e)$. These distributions are then compared, using the second program, to the distributions of τ_e and α_e that summarize the pseudo-observed data $To(\tau_e)$ and $Ao(\alpha_e)$ using a two-dimensional Kolmogorov-Smirnov to test the fit between the simulated data and the pseudo-observed data. The combination (U, S) that gives rise to the highest P-value is taken as U_e and S_e, even when the hypothesis that the distributions are different cannot be rejected.

This procedure was done to obtain the results in **Figures 1**, **S1** and **S4**, where twenty independent replicates of each pseudo-observed data set (under the same U, α and β) were performed and the average of U_e and S_e obtained for each pseudo-observed data set is presented.

New estimation method based upon the dynamics of frequency and fitness

We propose a new method, the One Bi-allelic Marker ABC (**Figure 2**), which aims to infer the distribution of arising mutations. The pseudo-observed data used to infer the performance of the method are generated under the model of adaptation

explained before, but the method now analyses the distributions, along time intervals (ti), of both marker frequency (f(ti)) and fitness (w(ti)), where ti=i*50 generations (i=0 to 6) are measured for 100 replicate populations evolving under a given U, α and β. A large data set with 1 million simulated evolutionary dynamics is produced, each with a set of 100 replicate populations evolving under a specific combination of parameters U, a and b. For each of the simulations, each parameter is randomly chosen from the following distributions: $\log_{10}(U) \sim$ Uniform [-9;-4]; $a \sim$ Uniform [0.5; 15]; and $\log_{10}(b) \sim$ Uniform [-4;-0.08]. Both the pseudo-observed data and the simulated data are summarized as the distribution of the values of |0.5-f(ti)| represented as a histogram, with 5 binned classes, for the marker frequency at different time points (ti), and of the distribution of fitness effects at the same time points, w(ti), represented as a histogram with 6 binned classes. This results in 11 summary statistic values for each of the 6 time points used in the analysis (table in **Figure 2**).

Summary statistics from pseudo-observed and simulated data are compared using an Approximate Bayesian Computation (ABC) method (Beaumont, Zhang, et al. 2002) implemented in R (Csilléry, François, et al. 2012) (package downloaded from and available at http://cran.r-project.org/web/packages/abc/index.html). ABC approaches have previously been used, for example, to determine rates of selective sweeps using sequence data from populations of *Drosophila melanogaster* (Jensen, Thornton, et al. 2008a). The input of the ABC method are the summary statistics of the 100 replicate populations that compose the pseudo-observed data ($S(y_0)$) and the previously described 1 million simulations ($S(y_i)$). The ABC method computes the posterior probability distribution of a multivariate parameter, θ (composed of a

combination of U, α and β). A value for this parameter, θ_i, is sampled from the prior distributions, and the summary statistics computed from simulated data $S(y_i)$ are compared to those of the pseudo-observed data $S(y_0)$ using the Euclidian distance *d*. If *d* is below a given threshold, the parameter value θ_i is accepted. The threshold (tolerance) chosen was 0.5%, which corresponds to the proportion of accepted simulations. The estimation of the posterior probability distribution for θ can be improved by different regression-based methods available in the ABC R package (Csilléry, François, et al. 2012): local linear regression and neural networks. We used the neural network method, which performs a dimensionality reduction in the summary statistics, and is suggested to be appropriate for use with high dimensionality (Csilléry, François, et al. 2012). Through this procedure, we obtain estimates for U, α and β, outputted as posterior distributions for each parameter. For each combination of parameters (U, α and β), twenty independent pseudo-observed data sets were considered to produce the statistics presented in the results. A scheme with the different steps described here is represented in **Figure 2**.

Effect of variation in initial frequency of marker and presence of deleterious mutations

We tested the effect of small variations in the initial frequency of each of the initial subpopulations (**Figure S3**) as they may occur in any experimental setup. We also tested how the estimates would be affected by the occurrence of deleterious mutations (**Figure S2**). For the first scenario, we generated pseudo-observed data sets under the same assumptions of the adaptation model described above except that the initial frequency of the neutral marker f(t=0)=0.5 + e, where e is drawn from a

Uniform distribution, e ~ Uniform [-0.03, 0.03] (**Figure S3**). For testing the effect of deleterious mutations we generated pseudo-observed data sets assuming that, in addition to beneficial mutations, deleterious mutations can also occur at a rate of 10^{-3} and each having a selection coefficient S_{del} = 2%. Multiplicative fitness was also assumed (**Figure S2**). All other assumptions were kept the same. Pseudo-observed data sets for both cases were generated with a Gamma distribution of fitness effects with $\alpha = 1$.

Results

Comparison of effective parameters U_e and S_e with U and the average effect of beneficial mutations

To determine if the effective parameters U_e and S_e are good estimates of U and S, we generated pseudo-observed data for a given value of U and with fitness effects drawn from a Gamma distribution with different shape and scale parameters (**Figure 3**). For populations with $N_e=10^5$, the estimated values of the effective parameters are shown in **Figure 1**, where we also have included results for the case where pseudo-observed data was generated under a model where all beneficial mutations have the same effect, since this is the case where the estimates are expected to perform best. **Figure 1A** shows that U_e provides a good estimate of U for Gamma distributions with shape parameter bigger than 1. This is observed in the cases where U is low ($<10^{-6}$), but when $U=10^{-5}$ and E(S)=0.02, U_e underestimates U by a quarter of its real value. Larger biases can be seen for the Exponential

distribution (α=1), particularly under high mutation rates (>10^{-7}), where clonal interference may be more pronounced. We find that when the distribution of S is Exponential with mean 2% and the mutation rate is 10^{-5} (parameters that have been estimated in some bacteria evolution experiments (Perfeito, Fernandes, et al. 2007)), U_e considerably underestimates the real value of U by an order of magnitude. This also happens when the mean effect of beneficial mutations is 6%. The underestimation becomes smaller when either the mutation rate or the variance in S decreases. As expected, U_e provides an accurate estimate of U when S is constant (except for the case where a high value of the mutation rate is considered). **Figure 1B** shows that S_e overestimates E(S) 2 to 4 times for a mutation rate higher 10^{-6}, with a=1. For α=10 this overestimation is small (less than 1.5-fold). Importantly, however, most of these values for S_e seem to provide an estimate of the order of magnitude of the mean effect of beneficial mutations. To test if the bias in U_e and S_e increases with clonal interference, we also studied populations with increased effective population size (N_e=10^6). Indeed, we find that both U_e (**Figure S1A**) and S_e (albeit to a lesser extent) (**Figure S1B**), show larger deviations from U and E(S), that can be up to a 50 fold underestimates of U and a 6 fold overestimates of E(S). In sum, higher levels of clonal interference (more pronounced in larger populations and with higher values of the mutation rate) lead to larger biases in U_e and S_e. These biases are dependent upon the underlying distribution of beneficial mutations.

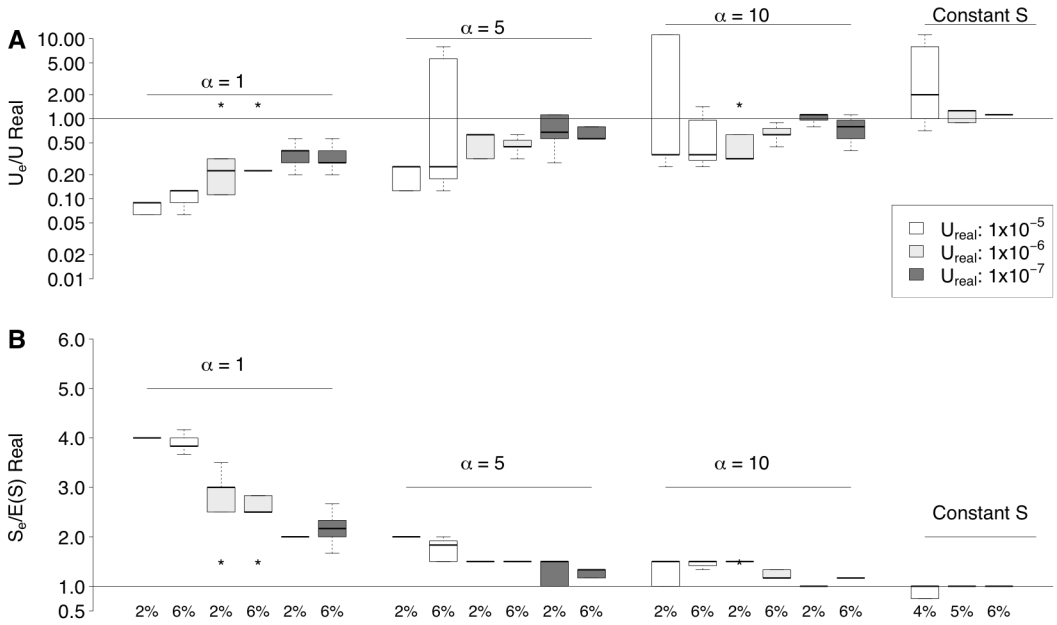
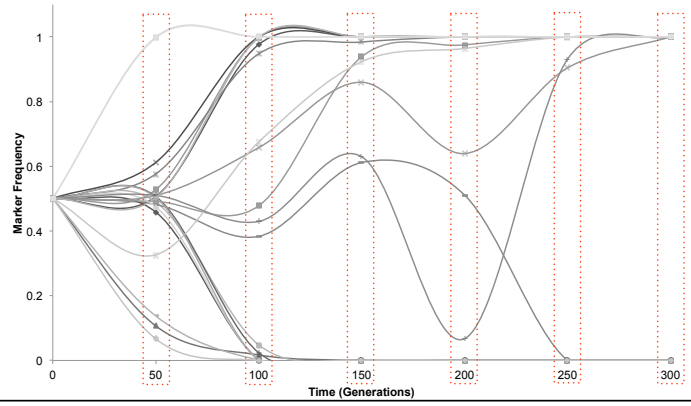


Figure 1: Performance of the single S model, according to the highest scoring estimates for pseudo-observed data with *f*(S) as Gamma distributions of different variances. A) Ratios of estimates of U over real parameter U; B) Ratios of estimates of S over the mean effect of S. The box plots of 20 independent estimation processes are shown, with the median indicated as a bar. Asterisks indicate cases where none of the 20 independent replicates was fitted significantly.

Experimental Data



Summary Statistics

Difference in Frequency = |0.5 - Frequency|

$$BinInterval = \frac{MaxDistance - MinDistance}{5}$$

Minimum Distance ↘ ↙ Maximum Distance

Fraction of populations at

	[0.00-0.10]	[0.10-0.20]	[0.20-0.30]	[0.30-0.40]	[0.40-0.50]
Gen. 50	0.65	0.1	0	0.1	0.15
Gen. 100	[0.02-0.12] 0.15	[0.12-0.21] 0.1	[0.21-0.31] 0	[0.31-0.40] 0	[0.40-0.50] 0.75
Gen. 150	[0.11-0.19] 0.1	[0.19-0.27] 0	[0.27-0.34] 0	[0.34-0.42] 0.05	[0.42-0.50] 0.85
Gen. 200	[0.01-0.11] 0.05	[0.11-0.21] 0.05	[0.21-0.30] 0	[0.30-0.40] 0.05	[0.40-0.50] 0.85
Gen. 250	[0.40-0.42] 0.05	[0.42-0.44] 0.05	[0.44-0.46] 0	[0.46-0.48] 0	[0.48-0.50] 0.9
Gen. 300	[0.50-0.50] 1	[0.50-0.50] 0	[0.50-0.50] 0	[0.50-0.50] 0	[0.50-0.50] 0

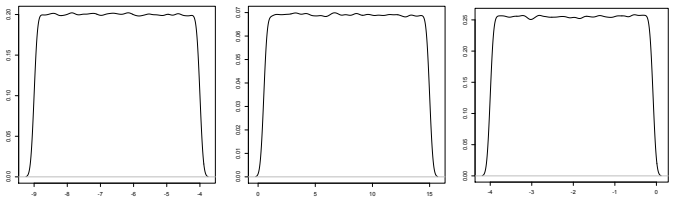
Average Population Fitness

$$BinInterval = \frac{MaxAverFitness - MinAverFitness}{6}$$

[1.00-1.07]	[1.07-1.14]	[1.14-1.21]	[1.21-1.28]	[1.28-1.34]	[1.34-1.41]
0.65	0.1	0.1	0.05	0.05	0.05
[1.15-1.23] 0.35	[1.23-1.30] 0.2	[1.30-1.38] 0.2	[1.38-1.45] 0.1	[1.45-1.53] 0.05	[1.53-1.60] 0.1
[1.39-1.54] 0.35	[1.54-1.68] 0.3	[1.68-1.82] 0.1	[1.82-1.96] 0.1	[1.96-2.10] 0.1	[2.10-2.24] 0.05
[1.58-1.76] 0.2	[1.76-1.94] 0.2	[1.94-2.12] 0.15	[2.12-2.30] 0.15	[2.30-2.48] 0.15	[2.48-2.66] 0.15
[1.87-2.13] 0.2	[2.13-2.39] 0.25	[2.39-2.65] 0.15	[2.65-2.90] 0.15	[2.90-3.16] 0.15	[3.16-3.42] 0.1
[2.29-2.63] 0.15	[2.63-2.98] 0.25	[2.98-3.32] 0.2	[3.32-3.67] 0.25	[3.67-4.02] 0.05	[4.02-4.36] 0.1

Priors for ABC Inference

1 million simulations generated with parameters uniform and randomly chosen from:



Priors for U | Priors for α | Priors for β

Transform simulations into the summary statistics created with experimental data

ABC Inference

Summary Statistics (*Experimental*) vs Summary Statistics (*Priors*)

ABC Posterior Distributions

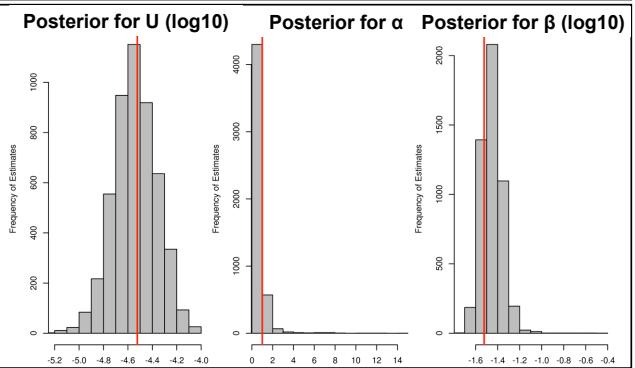


Figure 2: Schematic description of the One Bi-allelic Marker ABC. Data is obtained from an evolution experiment (here called pseudo-observed data), at specific time points, involving replicate adaptations to a common environment (an example of 20 replicate populations is shown). For each time point the data is condensed to summary statistics, for marker frequency and mean population fitness, which are histograms with the frequency of populations that fall in different bins (5 for frequency data, 6 for fitness data) at every 50 generations (Gen.). The choice of the bin for the frequency statistics is dictated by the module of the difference between the initial and current frequency of the subpopulations, so that this value is, at most, 0.5 (for marker frequencies of 1 or 0). A large simulated data set is built against which the experimental data is compared. The priors chosen to produce the simulated data set, which consist in 1 million simulations, are shown. Each simulated data (obtained with a given value of U, α and β) is then classified according to the same summary statistics as calculated for the observed data – called Summary stats (Priors) and Summary stats (Experimental), respectively. Using ABC inference these summary statistics are compared and the ones closest to the experimental data chosen. The 5000 top ranked values (0.5%) of each of the parameters are shown as the posterior distribution where the median value is highlighted in red.

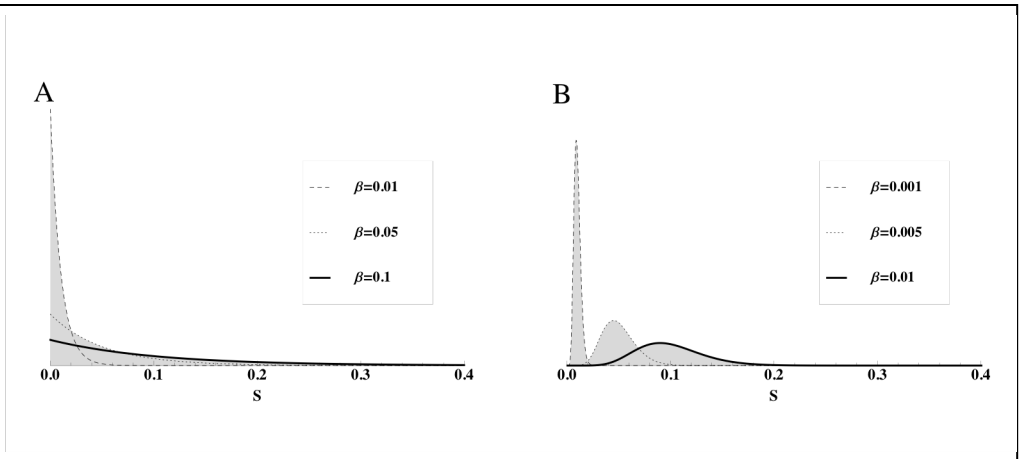


Figure 3: Theoretical distributions of beneficial selective coefficients assumed to produce pseudo-observed data. A) Gamma distribution with shape parameter α=1 (Exponential distribution), for different scale (β) parameters; **B**) Gamma distribution with shape parameter α=10, for different scale (β) parameters.

Estimation of the distribution of arising beneficial effects

To go beyond the mean effect of beneficial mutations and to try to estimate the distribution of arising beneficial mutations, we developed a new method, which we call One Bi-allelic Marker ABC (**Figure 2**). To test its performance in retrieving the evolutionary parameters U, α and β, we explored different sets of pseudo-observed data with combinations of parameter values that seem reasonable given the current literature (Denver, Wilhelm, et al. 2012; Lang, Botstein, et al. 2011; Perfeito, Fernandes, et al. 2007).

In **Figure 4** we show the ability of the One Bi-allelic Marker ABC method to estimate U, α and β, when the distribution of arising mutations is Exponential. This is the most commonly assumed distribution in theoretical studies of the adaptive process (Betancourt & Bollback, 2006; Orr, 2010). **Figure 4A** shows that the One Bi-allelic Marker ABC method provides estimates of U within an order of magnitude, for all cases tested. The worst performance lies in retrieving U for both high values of E(S) (5% and 10%) and high U ($3x10^{-5}$), but even in these cases the estimated value allows for a correct estimate of the order of magnitude of U. For the intermediate value of the mutation rate studied ($U=3x10^{-6}$) the method provides an accurate estimate of U. **Figures 4B** and **4C** provide the results for the estimates of the shape and scale parameters of *f*(S). As shown in **Figure 4B**, the estimate of the shape parameter a is close to 1 or 2, for the majority of the cases considered. Exceptions occur for the high mutation rate and the larger β values, which have a very high

variance. Estimation of β, shown in **Figure 4C**, is remarkably good, across the parameter range studied, being always below 2-fold the real value of β.

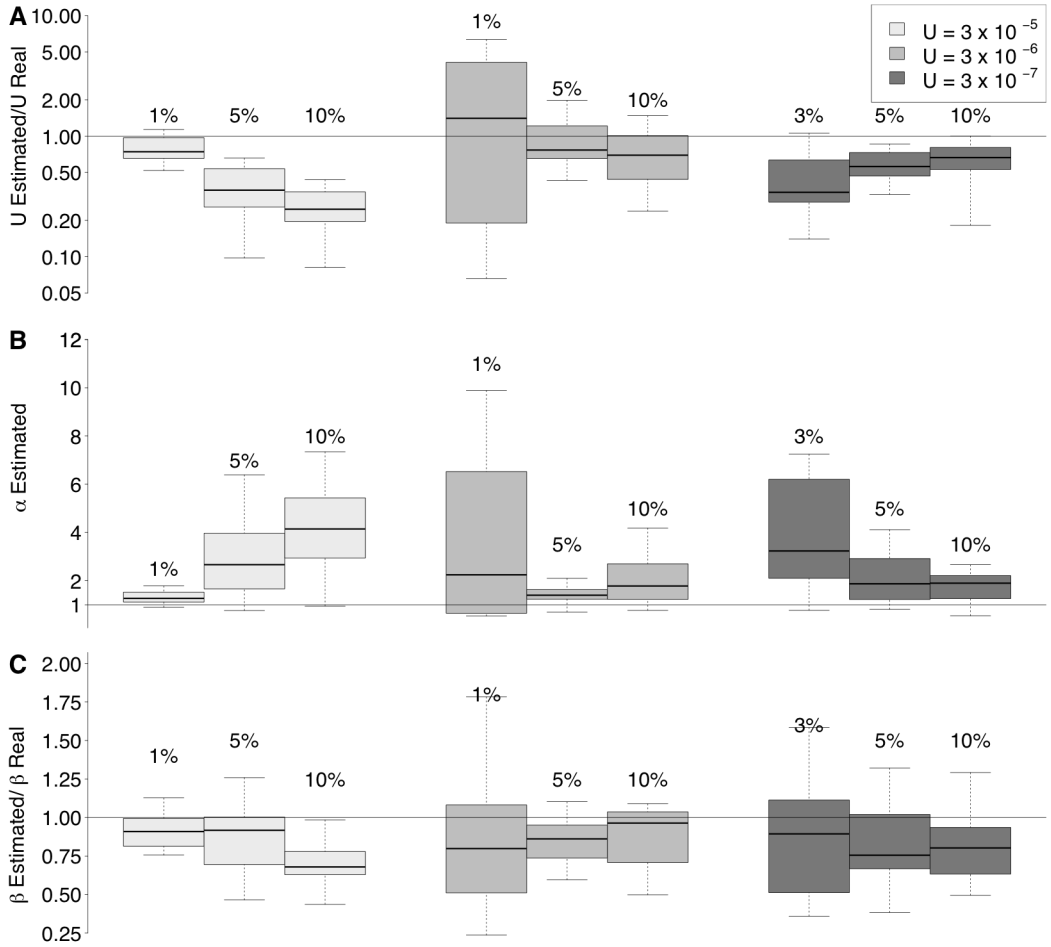


Figure 4: Performance of the One Bi-allelic Marker ABC method, for pseudo-observed data generated with *f*(S) as a Gamma distribution with α=1 and different scales (β). Labels above the bars show the mean selection coefficient for each case, and x-axis shows the value assumed for the mutation rate. A) Ratios of estimates of U over real parameter U; B) Estimates of the shape (α) parameter; C) Ratios of estimates of the scale (β) parameter over the real scale parameter. The box plots of 20 independent estimation processes are shown, with the median indicated as a bar.

We also studied the case where the distribution of arising beneficial mutations has a different shape, specifically α=10 (see **Figure 3**). As shown in **Figure 5A**, the estimated values of U are very close to the real ones in this case, rarely exceeding 2 times the real U values, although it can be either over or underestimated, depending upon the average selective effect. The two parameters characterizing the distribution of arising mutations are also remarkably close to the real values. **Figure 5B** shows that α is always estimated to be close to its true value (between 7 and 12), irrespectively of the value of U. Importantly, this estimate of α allows us to detect that the distribution of arising mutations is not Exponential. The method, therefore, has power to reliably distinguish between distributions of effects with distinct shapes. In **Figure 5C** the performance of the estimates regarding the β parameter of the distribution of effects is shown. β is well estimated, never exceeding twice the real value.

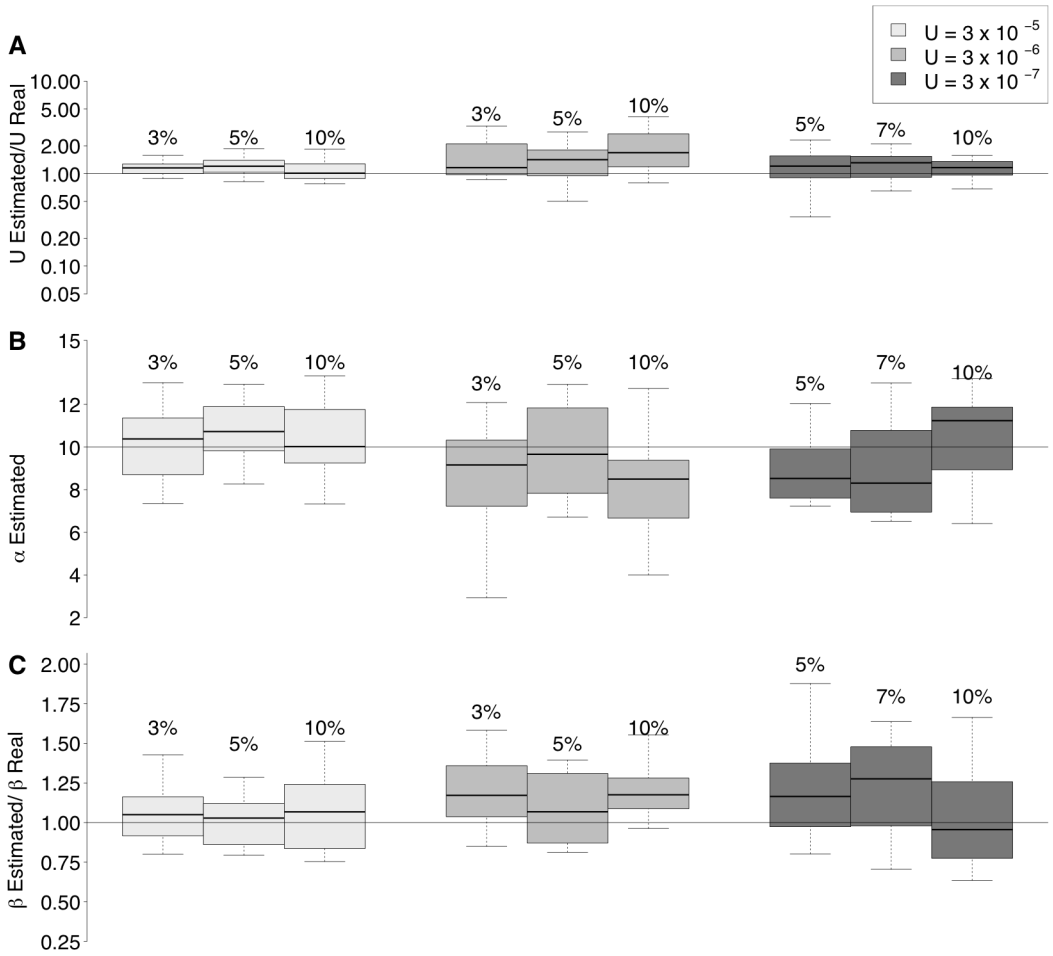


Figure 5: Performance of the One Bi-allelic Marker ABC method, for pseudo-observed data generated with *f*(S) as a Gamma distribution with α=10 and different scales (β). Labels and symbols are as in **Figure 4.**

To further assess the power of the method in distinguishing distributions with different shapes, we studied intermediate values of α, between 0.75 and 10. In **Figure 6**, we show that the One Bi-allelic Marker ABC method is able to discriminate not only between the two limiting cases in our simulations, but also between

intermediate α values. The method fails to distinguish the shape of the distribution of arising mutations when 0.75 < a < 2, especially when U is large. In these cases, α is overestimated (by about 2 fold). When a > 2, estimates of a consistent with the true value are obtained. When a=4, rejection of an exponential distribution is obtained. Overall, the method provides a reliable distinction between different shapes of the distribution of arising mutations, although distinguishing between α values lower than 2 remains difficult.

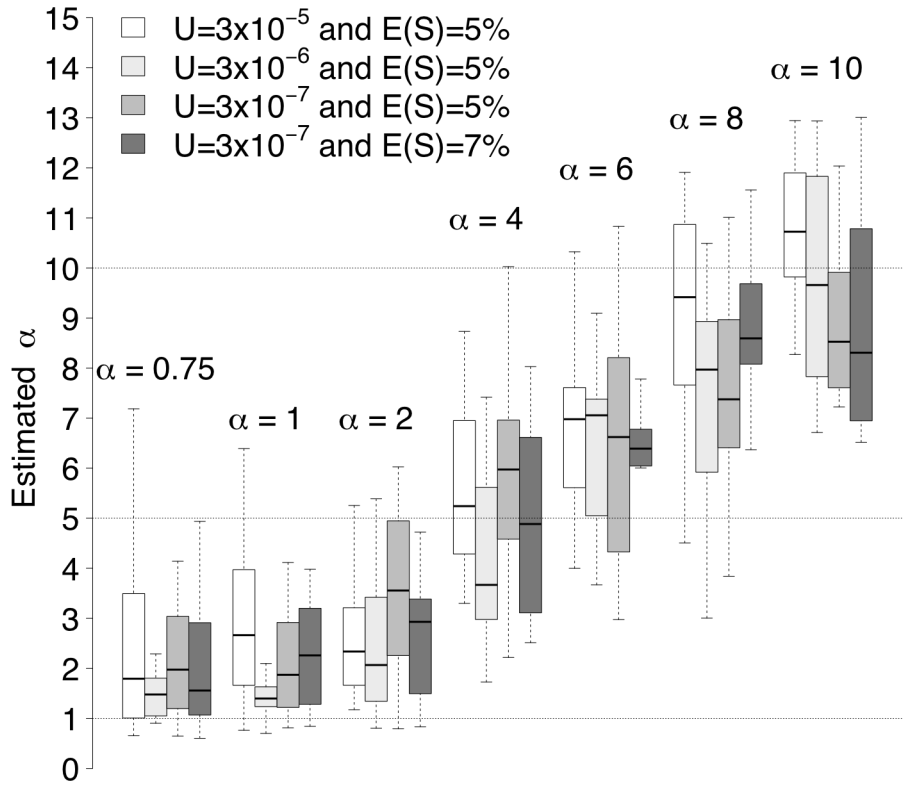


Figure 6: Estimates for the shape (α) parameter, for pseudo-observed data generated with *f*(S) as a Gamma distribution with varying α and scales (β), in order for E(S) to be constant within a mutation rate. Labels show the α parameters used as pseudo-observed data for each case. The box plots of 20 independent estimation processes are shown, with the median indicated as a bar.

Discussion

In order to estimate the parameters that describe the dynamics of adaptation, we need powerful methods. Beneficial mutations are essential in driving adaptation and their statistical properties remain an open question (Orr 2010). Although methods developed to tackle this subject may never perfectly capture the complete nature of the evolutionary process, they can provide reasonable estimates regarding the strength of the forces involved in the process (Keightley and Eyre-Walker 2010; Thornton, Jensen, et al. 2007).

A simple theoretical approach assumes that all mutations have the same fitness effect and has been shown to have predictive power in explaining certain patterns of data obtained in experimentally evolved populations (Hegreness, Shoresh, et al. 2006). Notwithstanding, several direct measurements of mutation effects point to the existence of considerable variation (Kassen and Bataillon 2006) which motivates the development of new methods that try to infer the underlying distribution of arising beneficial effects.

Regarding the estimated effective evolutionary parameters studied here, it seems clear that the relation with the real parameters is dependent on the actual distribution of effects of arising mutations: exponential-like distributions of beneficial effects result in values of U_e below the true mutation rate and values of S_e above the true mean effect of mutations, with the difference being reduced when the distribution of effects decrease in variance. Nevertheless, assuming a fixed value for S has been a commonly used method to infer the evolutionary parameters from

experimental data, for example in studies that address how evolvability is dependent upon the genetic background. In one such study, Barrick et al. (Barrick, Kauth, et al. 2010) isolated 8 clones of *Escherichia coli* with different mutations in the *rpoB* gene, encoding the *β* subunit of RNA polymerase. As these mutations are generally deleterious in environments without antibiotics, and they can cause a wide range of fitness defects (Trindade, Perfeito, et al. 2010), the authors estimated U_e and S_e to determine the evolvability of different (but related) genotypes. The two neutral markers dynamics were used to estimate the evolutionary parameters and, from these dynamics, it was inferred that mutants with a higher fitness defect had a higher evolvability caused by a stronger selective effect of beneficial mutations. Interestingly, the inferred mutation rate (through U_e) appears to be independent of the genetic background. Since we show here that U_e may be below U and S_e above E(S), some caution is to be taken when drawing conclusions regarding the relation between evolvability and fitness effects of such mutations. Similar caveats apply in the study of Woods et al. (Woods, Barrick, et al. 2011). That study involved a long-term evolution experiment, running for more than 50,000 generations, where clones sampled at generation 500 were found to carry mutations in *topA* and *rbs*. These were shown to be beneficial and fixed after generation 1500, and their carriers were called “eventual winners”. Other contemporaneous genotypes (with other mutations) were deemed “eventual losers”. Even though both sets of clones had increased fitness related to the ancestral, the “eventual losers” also had, counter intuitively, increased fitness relative to the “eventual winners”. To understand why the “eventual winners” ultimately won the competition, their evolvability was studied, and U and

E(S) were inferred (through U_e and S_e) by assaying neutral marker dynamics. The authors found that “eventual winners” had, indeed, the ability to generate beneficial mutations with stronger effects, compared to the “eventual losers”.

The approach used in both studies to determine evolvability may provide an overestimate of the mean selective coefficient in the order of two to three times the real values if the mutation rates are in the order estimated by the authors, or even more, if the mutation rates are underestimated (see **Figure 1B**). As a consequence, this could imply that the actual mean selective coefficients are lower than the one estimated, and small differences in evolvability may be difficult to detect.

In general, inferring evolutionary parameters and, more specifically, the distribution of arising mutations, from data of evolving populations is a difficult task. Experimentally, one way to gain further insight into the distribution of effects is to use more than two neutral markers, which can bring more power (Perfeito, Fernandes, et al. 2007). Theoretically, we can expect that new and improved methods are likely to emerge. Recently, Zhang et al. (Zhang, Sehgal, et al. 2012) extended the previous model by Hegreness et al. to incorporate a continuous initial growth phase, dividing it in 50 time intervals, and developing an analytical model to find the distributions of estimators for U and S. Like the previous work, however, only the initial dynamics are considered (the first significant deviation), and the method does not consider the occurrence of clonal interference. Illingworth and Mustonen (Illingworth and Mustonen 2012), on the other hand, developed a maximum likelihood method where the marker dynamics over the total amount of time followed is used. The method determines the

minimum number of mutations that best describe the dynamics and allows inferring the distribution of haplotype fitnesses that are segregating. Although the performance of the method is quite good under certain conditions, it is not clear how it will perform under a wide range of mutation rates.

Here we propose a new theoretical approach that is expected to contribute to improved insight regarding the distribution of arising beneficial mutation effects. Using Approximate Bayesian Computation, we propose a set of summary statistics to be used under a simple experimental setup, where distributions of marker frequencies and the mean fitness of the population are recorded at periodic time intervals. These statistics allow a reasonable estimation of the distribution of arising mutations and of the mutation rate, provided that we accept that such a distribution may be well approximated by a Gamma. Combining the parameters of the Gamma distribution (α and β), it is also possible to estimate the mean effect of arising beneficial mutations (E(S)). **Figure 7** shows the estimates of E(S) given by the method when α=1 or α=10. Under an Exponential distribution of fitness effects (**Figure 7A**), which is commonly assumed, the mean effect can be overestimated up to 5 or 6 fold, for large values of the mutation rate. For α=10, the estimate E(s) is very accurate, reflecting its real value for every condition tested (**Figure 7B**).

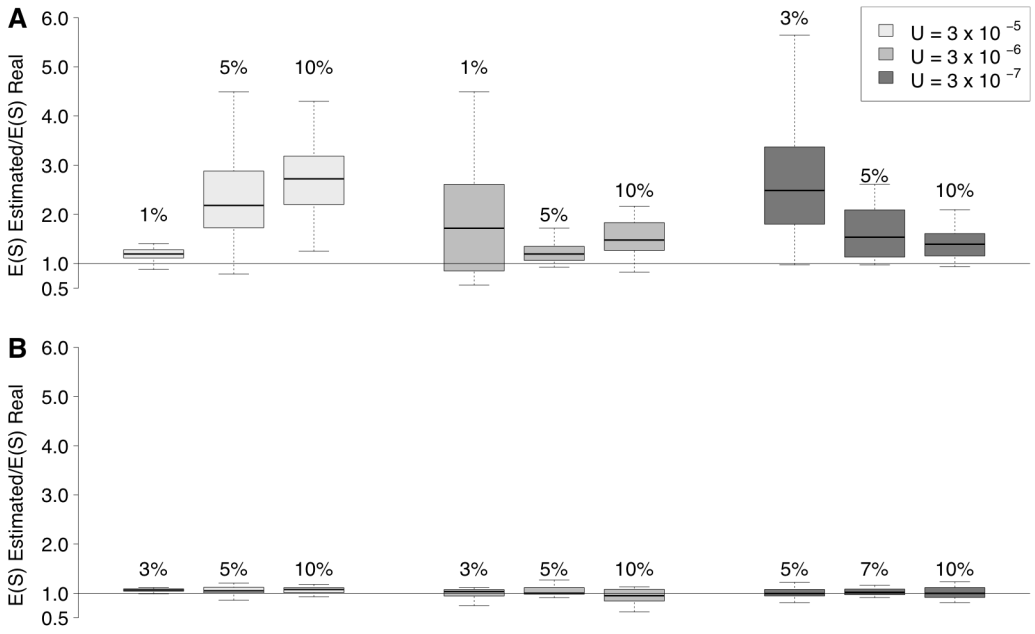


Figure 7: Ratios of estimated average beneficial selective coefficient over the real average beneficial selective coefficient (both calculated as αβ). Labels above the bars show the mean selection coefficient for each case, and x-axis shows the value assumed for the mutation rate. Pseudo-observed data was generated with *f*(S) as a Gamma distribution with α=1 (Exponential distribution of *f*(S)) (A) or α=10 (B). The box plots of 20 independent estimation processes are shown, with the median indicated as a bar.

The One Bi-allelic ABC method seems to allow distinguishing between distributions with different shapes and scales. The underlying model used makes several assumptions, which could be violated in a real experiment. In particular, it assumes that the initial population is composed of two equally sized subpopulations, each with a different marker, and it also assumes that no deleterious mutations occur. To test the robustness of the approach in the face of these assumptions, we performed new simulations where pseudo-observed data was generated. In one

case, the initial marker frequency was allowed to deviate from its expectation of 0.5 (**Figure S3**). In the other case, deleterious mutations were allowed to occur with rates and effects typical of those inferred from mutation accumulation experiments with bacteria (Kibota and Lynch 1996; Trindade, Perfeito, et al. 2010) (**Figure S2**). In both these cases, the inference of the values of U, α and β were similar to those obtained before.

We performed the analysis of a method, which assumes a common experimental setup with only one neutral locus with 2 alleles and fitness measurements at periodic time intervals. In principle this setup can be extended to follow variation of one locus with more alleles or neutral variation at more loci. The method could then be extended and a thorough study of the best summary statistics would be needed to ask what would be the minimal set of data required to reasonably estimate the rate and distribution of arising beneficial mutations.

We have also tested the effect of considering a smaller number of populations in order to determine if the approaches can provide reasonable estimates when applied to data that have been obtained in studies involving experimental evolution with fewer replicates. **Figure S4** shows the comparison of U_e with U and S_e with E(S) when the number of replicate populations is 10, which corresponds to the approximate size of previously published experiments (Barrick, Kauth, et al. 2010; Hegreness, Shoresh, et al. 2006; Woods, Barrick, et al. 2011). We observed similar biases to those found when considering 100 replicate evolved populations. Regarding the One Bi-allelic Marker ABC approach, we can observe that even with these

reduced number of populations, reasonable estimates of U can be obtained; the estimated values of a tend to produce an overestimation which can be up to 15 fold, while the estimates of β are close to the real ones (**Figure S5**).

The One Bi-allelic Marker ABC method, as the alternatives discussed above, displays certain limitations in its performance, which are particularly apparent when dealing with very intense clonal interference, for which a system with more markers would be desirable. It is a method that tries to estimate the distribution without limiting the number of mutations in a given genetic background, and taking into account the dynamics of the entire process of adaptation. For a wide spectrum of mutation rates, we are able to estimate the parameters of the underlying distribution of beneficial mutations. The One Bi-allelic Marker ABC method was tested over a range of distributions of beneficial selective coefficients and beneficial mutation rates, including high mutation rates, which are typically not studied in the analysis of other methods. This gives us a fairly good degree of confidence that, in applying the method to real biological data from adaptation experiments of clonal populations using the two-marker methodology, we are able to gain information on the distribution of beneficial arising mutations.

Acknowledgments

We thank Lilia Perfeito, Ana-Hermina Ghenu, Lindi Wahl, the Gordo's Lab members, two anonymous referees and the editor for their comments and suggestions. The research leading to these results has received funding from the European Research

Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement nº 260421 – ECOADAPT. JAMS acknowledges the scholarship support provided by FCG and FCT. IG acknowledges the salary support of LAO/ITQB & FCT. PC acknowledges financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) and program PRONEX/MCT-CNPq- FACEPE.

References

Barrett RDH, MacLean RC, Bell G 2006. Mutations of intermediate effect are responsible for adaptation in evolving Pseudomonas fluorescens populations. Biology Letters 2: 236-238. doi: 10.1098/rsbl.2006.0439

Barrick JE, Kauth MR, Strelioff CC, Lenski RE 2010. Escherichia coli rpoB Mutants Have Increased Evolvability in Proportion to Their Fitness Defects. Molecular Biology and Evolution 27: 1338-1347. doi: 10.1093/molbev/msq024

Bataillon T, Zhang T, Kassen R 2011. Cost of Adaptation and Fitness Effects of Beneficial Mutations in Pseudomonas fluorescens. Genetics 189: 939-949. doi: 10.1534/genetics.111.130468

Beaumont MA, Zhang W, Balding DJ 2002. Approximate Bayesian computation in population genetics. Genetics 162: 2025-2035.

Biswas S, Akey JM 2006. Genomic insights into positive selection. Trends Genet 22: 437-446. doi: 10.1016/j.tig.2006.06.005

Bollback JP, York TL, Nielsen R 2008. Estimation of 2Nes from temporal allele frequency data. Genetics 179: 497-502. doi: 10.1534/genetics.107.085019

Burke MK, et al. 2010. Genome-wide analysis of a long-term evolution experiment with Drosophila. Nature 467: 587-590. doi: 10.1038/nature09352

Charlesworth B, Morgan MT, Charlesworth D 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289-1303.

Crow JF, Kimura M. 1970. An introduction to population genetics theory. New York,: Harper & Row.

Csilléry K, François O, Blum MGB 2012. abc: an R package for approximate Bayesian computation (ABC). Methods in Ecology and Evolution 3: 475-479. doi: 10.1111/j.2041-210X.2011.00179.x

Cutter AD, Choi JY 2010. Natural selection shapes nucleotide polymorphism across the genome of the nematode Caenorhabditis briggsae. Genome Research 20: 1103-1111. doi: 10.1101/gr.104331.109

Denver DR, et al. 2012. Variation in base-substitution mutation in experimental and natural lineages of Caenorhabditis nematodes. Genome biology and evolution. doi: 10.1093/gbe/evs028

Desai MM, Fisher DS, Murray AW 2007. The speed of evolution and maintenance of variation in asexual populations. Curr Biol 17: 385-394. doi: 10.1016/j.cub.2007.01.072

Drake JW, Charlesworth B, Charlesworth D, Crow JF 1998. Rates of spontaneous mutation. Genetics 148: 1667-1686.

Dykhuizen DE, Hartl DL 1983. Selection in chemostats. Microbiol Rev 47: 150-168.

Elena SF, Lenski RE 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet 4: 457-469. doi: 10.1038/nrg1088

Enard D, Depaulis F, Roest Crollius H 2010. Human and Non-Human Primate Genomes Share Hotspots of Positive Selection. PLoS Genet 6: e1000840. doi: 10.1371/journal.pgen.1000840.t002

Estes S, Phillips PC, Denver DR 2011. Fitness recovery and compensatory evolution in natural mutant lines of C. elegans. Evolution 65: 2335-2344. doi: 10.1111/j.1558-5646.2011.01276.x

Eyre-Walker A, Keightley PD 2007. The distribution of fitness effects of new mutations. Nat Rev Genet 8: 610-618. doi: 10.1038/nrg2146

Fisher RA. 1930. The genetical theory of natural selection. Oxford,: The Clarendon press.

Gerrish PJ, Lenski RE 1998. The fate of competing beneficial mutations in an asexual population. Genetica 102-103: 127-144.

Good BH, et al. 2012. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. Proceedings of the National Academy of Sciences of the United States of America 109: 4950-4955. doi: 10.1073/pnas.1119910109

Gordo I, Perfeito L, Sousa A 2011. Fitness effects of mutations in bacteria. J Mol Microbiol Biotechnol 21: 20-35. doi: 10.1159/000332747

Grossman SR, et al. 2013. Identifying recent adaptations in large-scale genomic data. Cell 152: 703-713. doi: 10.1016/j.cell.2013.01.035

Grossman SR, et al. 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327: 883-886. doi: 10.1126/science.1183863

Hancock AM, Di Rienzo A 2008. Detecting the Genetic Signature of Natural Selection in Human Populations: Models, Methods, and Data. Annu. Rev. Anthropol. 37: 197-217. doi: 10.1146/annurev.anthro.37.081407.085141

Hegreness M, Shoresh N, Hartl D, Kishony R 2006. An equivalence principle for the incorporation of favorable mutations in asexual populations. Science 311: 1615-1617. doi: 10.1126/science.1122469

Hietpas RT, Jensen JD, Bolon DNA 2011. From the Cover: Experimental illumination of a fitness landscape. Proc Natl Acad Sci USA 108: 7896-7901. doi: 10.1073/pnas.1016024108

Illingworth CJ, Mustonen V 2011. Distinguishing driver and passenger mutations in an evolutionary history categorized by interference. Genetics 189: 989-1000. doi: 10.1534/genetics.111.133975

Illingworth CJR, Mustonen V 2012. A method to infer positive selection from marker dynamics in an asexual population. Bioinformatics 28: 831-837. doi: 10.1093/bioinformatics/btr722

Imhof M, Schlotterer C 2001. Fitness effects of advantageous mutations in evolving Escherichia coli populations. Proceedings of the National Academy of Sciences of the United States of America 98: 1113-1117. doi: 10.1073/pnas.98.3.1113

Jensen JD, Thornton KR, Andolfatto P 2008a. An Approximate Bayesian Estimator Suggests Strong, Recurrent Selective Sweeps in Drosophila. PLoS Genet 4: e1000198. doi: 10.1371/journal.pgen.1000198.t003

Jensen JD, Thornton KR, Aquadro CF 2008b. Inferring Selection in Partially Sequenced Regions. Molecular Biology and Evolution 25: 438-446. doi: 10.1093/molbev/msm273

Kassen R, Bataillon T 2006. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nat Genet 38: 484-488. doi: 10.1038/ng1751

Keightley PD 1998. Inference of genome-wide mutation rates and distributions of mutation effects for fitness traits: a simulation study. Genetics 150: 1283-1293.

Keightley PD, Eyre-Walker A 2010. What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? Philos Trans R Soc Lond, B, Biol Sci 365: 1187-1193. doi: 10.1098/rstb.2009.0266

Kibota TT, Lynch M 1996. Estimate of the genomic mutation rate deleterious to overall fitness in E. coli. Nature 381: 694-696. doi: 10.1038/381694a0

Kimura M, Ohta T 1974. On some principles governing molecular evolution. Proceedings of the National Academy of Sciences of the United States of America 71: 2848-2852.

Kvitek DJ, Sherlock G 2011. Reciprocal Sign Epistasis between Frequently Experimentally Evolved Adaptive Mutations Causes a Rugged Fitness Landscape. PLoS Genet 7: e1002056. doi: 10.1371/journal.pgen.1002056.g005

Lang GI, Botstein D, Desai MM 2011. Genetic variation and the fate of beneficial mutations in asexual populations. Genetics 188: 647-661. doi: 10.1534/genetics.111.128942

Lemonnier M, et al. 2008. The evolution of contact-dependent inhibition in non-growing populations of Escherichia coli. Proc Biol Sci 275: 3-10. doi: 10.1098/rspb.2007.1234

Lind PA, Berg OG, Andersson DI 2010. Mutational robustness of ribosomal protein genes. Science 330: 825-827. doi: 10.1126/science.1194617

MacLean RC, Buckling A 2009. The distribution of fitness effects of beneficial mutations in Pseudomonas aeruginosa. PLoS Genet 5: e1000406. doi: 10.1371/journal.pgen.1000406

Malaspinas AS, Malaspinas O, Evans SN, Slatkin M 2012. Estimating allele age and selection coefficient from time-serial data. Genetics 192: 599-607. doi: 10.1534/genetics.112.140939

Martin G, Lenormand T 2006. The fitness effect of mutations across environments: a survey in light of fitness landscape models. Evolution 60: 2413-2427.

Mathieson I, McVean G 2013. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. Genetics 193: 973-984. doi: 10.1534/genetics.112.147611

Maynard-Smith J, Haigh J 1974. The hitch-hiking effect of a favourable gene. Genet Res 23: 23-35.

Mcdonald MJ, Cooper TF, Beaumont HJE, Rainey PB 2011. The distribution of fitness effects of new beneficial mutations in Pseudomonas fluorescens. Biology Letters 7: 98-100. doi: 10.1098/rsbl.2010.0547

Nielsen R 2005. MOLECULAR SIGNATURES OF NATURAL SELECTION. Annu. Rev. Genet. 39: 197-218. doi: 10.1146/annurev.genet.39.073003.112420

Orozco-terWengel P, et al. 2012. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. Mol Ecol 21: 4931-4941. doi: 10.1111/j.1365-294X.2012.05673.x

Orr HA 2010. The population genetics of beneficial mutations. Philos Trans R Soc Lond, B, Biol Sci 365: 1195-1201. doi: 10.1098/rstb.2009.0282

Perfeito L, Fernandes L, Mota C, Gordo I 2007. Adaptive mutations in bacteria: high rate and small effects. Science 317: 813-815. doi: 10.1126/science.1142284

Rokyta DR, et al. 2008. Beneficial fitness effects are not exponential for two viruses. J Mol Evol 67: 368-376. doi: 10.1007/s00239-008-9153-x

Rozen DE, de Visser JAGM, Gerrish PJ 2002. Fitness effects of fixed beneficial mutations in microbial populations. Curr Biol 12: 1040-1045.

Sanjuán R, Moya A, Elena SF 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proceedings of the National Academy of Sciences of the United States of America 101: 8396-8401. doi: 10.1073/pnas.0400146101

Sawyer SA, Parsch J, Zhang Z, Hartl DL 2007. Prevalence of positive selection among nearly neutral amino acid replacements in Drosophila. Proceedings of the National Academy of Sciences of the United States of America 104: 6504-6510. doi: 10.1073/pnas.0701572104

Schneider A, Charlesworth B, Eyre-Walker A, Keightley PD 2011. A method for inferring the rate of occurrence and fitness effects of advantageous mutations. Genetics 189: 1427-1437. doi: 10.1534/genetics.111.131730

Schoustra SE, Bataillon T, Gifford DR, Kassen R 2009. The Properties of Adaptive Walks in Evolving Populations of Fungus. PLoS Biol 7: e1000250. doi: 10.1371/journal.pbio.1000250.t001

Sinha P, et al. 2011. On Detecting Selective Sweeps Using Single Genomes. Front. Gene. 2: 1-5. doi: 10.3389/fgene.2011.00085

Sousa A, Magalhães S, Gordo I 2012. Cost of antibiotic resistance and the geometry of adaptation. Molecular Biology and Evolution 29: 1417-1428. doi: 10.1093/molbev/msr302

Stevens KE, Sebert ME 2011. Frequent beneficial mutations during single-colony serial transfer of Streptococcus pneumoniae. PLoS Genet 7: e1002232. doi: 10.1371/journal.pgen.1002232

Thornton KR, Jensen JD, Becquet C, Andolfatto P 2007. Progress and prospects in mapping recent selection in the genome. Heredity: 1-9. doi: 10.1038/sj.hdy.6800967

Trindade S, Perfeito L, Gordo I 2010. Rate and effects of spontaneous mutations that affect fitness in mutator Escherichia coli. Philos Trans R Soc Lond, B, Biol Sci 365: 1177-1186. doi: 10.1098/rstb.2009.0287

Woods RJ, et al. 2011. Second-Order Selection for Evolvability in a Large Escherichia coli Population. Science 331: 1433-1436. doi: 10.1126/science.1198914

Zhang W, et al. 2012. Estimation of the rate and effect of new beneficial mutations in asexual populations. Theor Popul Biol 81: 168-178. doi: 10.1016/j.tpb.2011.11.005

Supplementary Figures

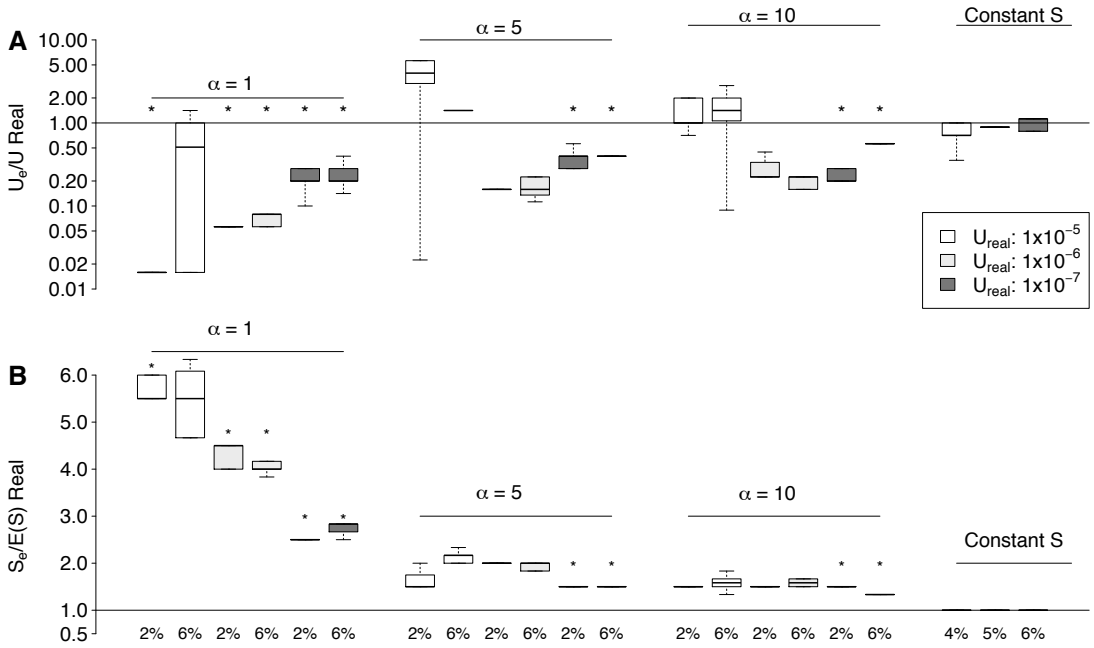


Figure S1: Performance of the single S model, according to the highest scoring estimates for experimental data with *f*(S) as Gamma distributions of different variance and N_e=10^6. A) Ratios of estimates of U over the real parameter U; B) Ratios of estimates of S over the mean effect of S. The box plots of 20 independent estimation processes are shown, with the median indicated as a bar. Asterisks indicate cases where none of the 20 independent replicates was fitted significantly.

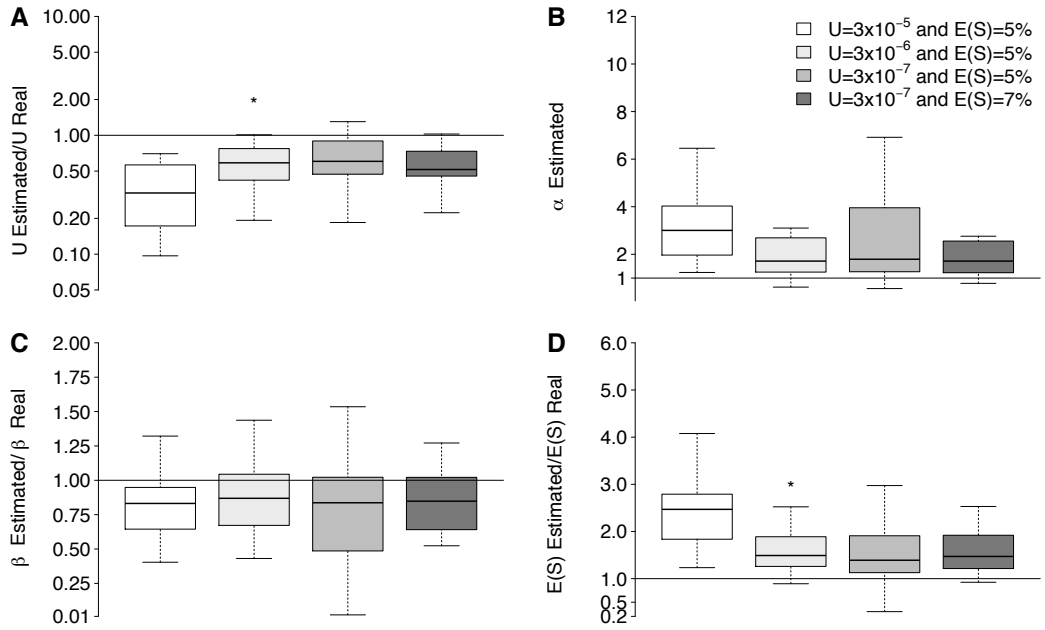


Figure S2: Effect of the presence of deleterious mutations in the estimation of U, a, b and E(S) with the One Bi-allelic Marker ABC method. Pseudo-observed data sets were generated with a model where deleterious mutations occurred at rates U_d=0.001, each having an effect 0.02. Beneficial mutations were sampled from a gamma distribution with α=1. A) Estimation of U; B) Estimation of α; C) Estimation of β; D) Estimation of the mean beneficial effect αβ . Asterisks indicate the cases where the inferences for pseudo-observed data sets with both beneficial and deleterious mutations is significantly different (Mann-Whitney test $P < 0.05$) from the inference from pseudo-observed data sets without deleterious mutations.

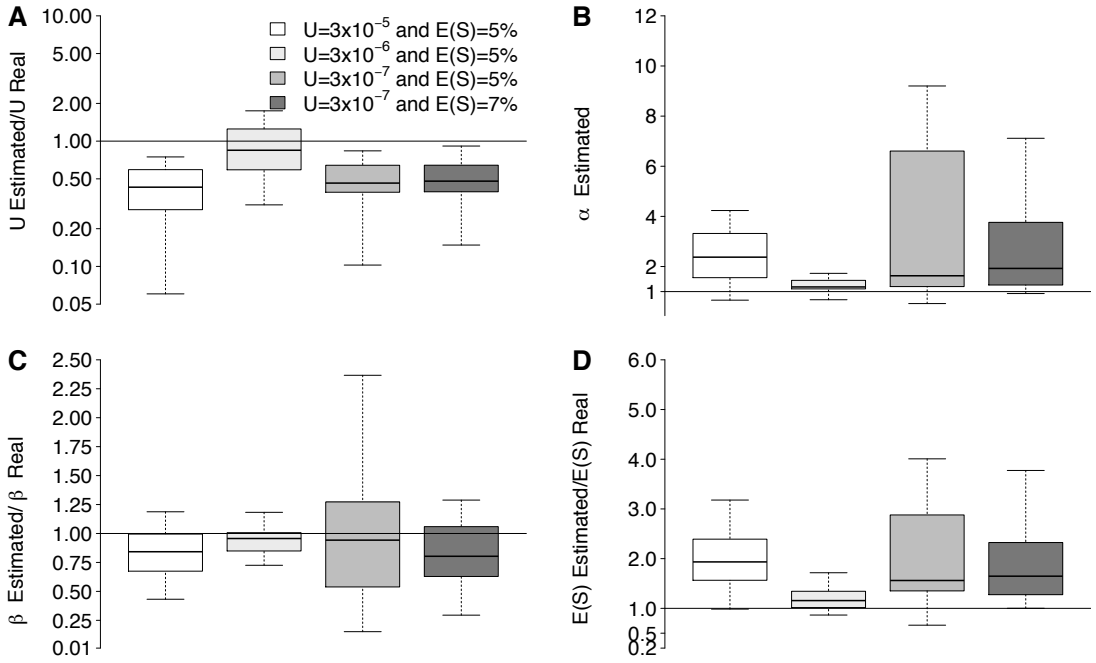


Figure S3: Effect of uncertainty in the value of the initial marker frequency when estimating U, a, b and E(S) under the One Bi-allelic Marker ABC method. Pseudo-observed data sets were generated assuming random variation around the initial marker frequency (0.5) by adding a random error uniformly distributed [-0.03; 0.03]. A) Estimation of U; B) Estimation of α ; C) Estimation of β; D) Estimation of the mean beneficial effect αβ. Asterisks indicate the cases where the inferences for pseudo-observed data sets including error in initial frequencies is significantly different (Mann-Whitney test $P < 0.05$) from the inference without error.

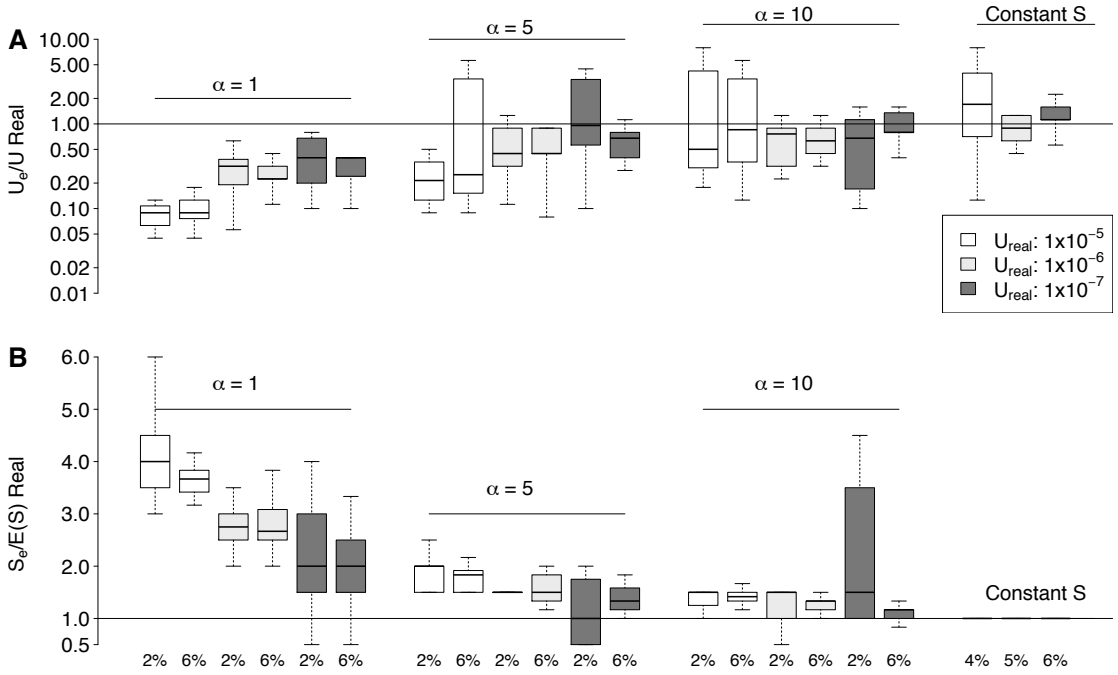


Figure S4 Estimation of U and S through the effective evolutionary parameters (U_e and S_e) with pseudo-observed data from 10 replicate populations with *f*(S) as Gamma distributions of different variances and $N_e=10^5$. A) Ratios of estimates of U over real parameter U; B) Ratios of estimates of S over the mean effect of S. The box plots of 20 independent estimation processes are shown, with the median indicated as a bar. Asterisks indicate cases where none of the 20 independent replicates was fitted significantly.

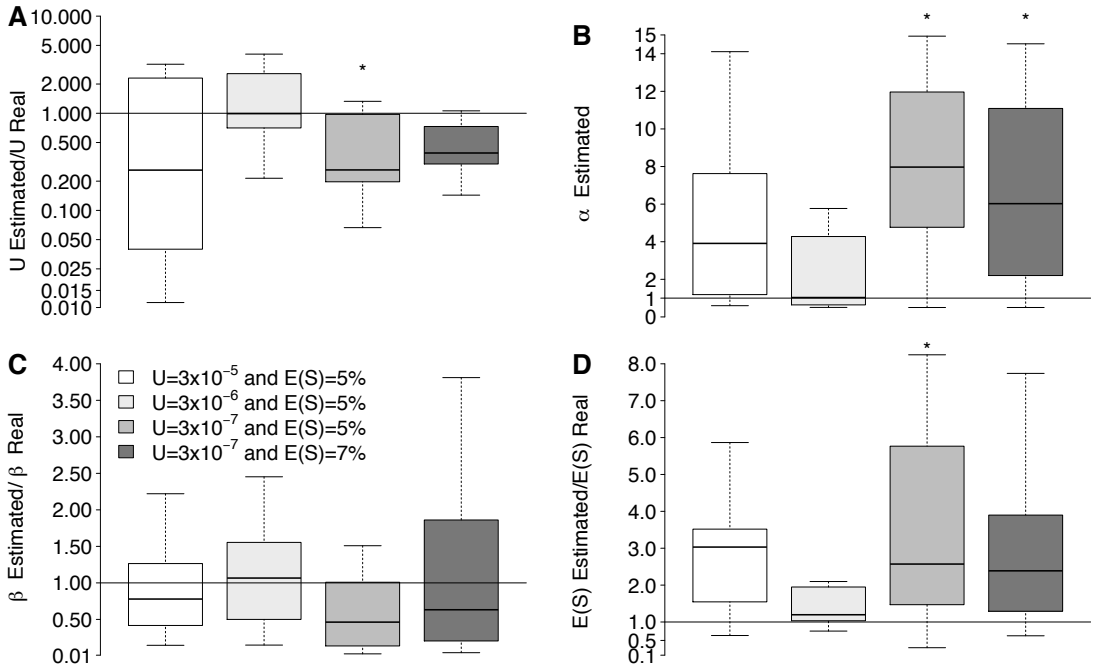


Figure S5 Estimation of U, a, b and E(S) with the One Bi-allelic Marker ABC method with pseudo-observed data from 10 replicate populations. A) Estimation of U; B) Estimation of α; C) Estimation of β; D) Estimation of the mean beneficial effect αβ. Asterisks indicate the cases where the inferences are significantly different (Mann-Whitney test $P < 0.05$) from those derived from 100 replicate populations.

CHAPTER III

Competition and fixation of cohorts of adaptive mutations under Fisher geometrical model

Manuscript under revision in ***PeerJ***

The author of this thesis, João Alpedrinha and Paulo Campos developed the algorithms used in this chapter. The author of this thesis performed all the simulations described in this chapter.

Competition and fixation of cohorts of adaptive mutations under Fisher geometrical model

Jorge A. Moura de Sousa[1], João Alpedrinha[1], Paulo R. A. Campos[2], Isabel Gordo[1]

[1] Instituto Gulbenkian de Ciência. Rua da Quinta Grande, 6. 2780-156 Oeiras, Portugal.

[2] Departamento de Física, Universidade Federal de Pernambuco, Cidade Universitaria, Recife PE, Brazil.

[*] Corresponding author

jasousa@igc.gulbenkian.pt

Telf: +351 214407915

Fax: +351 214407970

Abstract

One of the simplest models of adaptation to a new environment is Fisher's Geometric Model (FGM), in which populations move on a multidimensional landscape defined by the traits under selection. The predictions of this model have been found to be consistent with current observations of patterns of fitness increase in experimentally evolved populations. Recent studies investigated the dynamics of allele frequency change along adaptation of microbes to simple laboratory conditions and unveiled a dramatic pattern of competition between cohorts of mutations, i.e. multiple mutations simultaneously segregating and ultimately reaching fixation. Here, using simulations, we study the dynamics of

phenotypic and genetic change as asexual populations under clonal interference climb a Fisherian landscape, and ask about the conditions under which FGM can display the simultaneous fixation of mutations along the adaptive walk. We find that FGM under clonal interference, and with varying levels of pleiotropy, can reproduce the experimentally observed competition between different cohorts of mutations ("cohort interference"), some of which have a high probability of fixation along the adaptive walk. Overall, our results show that a simple version of FGM is able to capture the dynamics of cohorts of mutations that are observed in experiments involving the adaptation of microbial populations.

Keywords: Fisher Geometric Model, clonal interference, genetic hitchhiking, natural selection.

Introduction

Understanding the mechanisms and dynamics underneath the adaptive process is still a great challenge in evolutionary biology. Even in relatively simple environments, evolution experiments demonstrate that this process often involves complex dynamics such as: (1) competition between clones carrying different adaptive alleles (Maharjan & Ferenci 2015; Desai & D S Fisher 2007; Perfeito et al. 2007) , (2) hitchhiking, along with beneficial alleles, of neutral and even deleterious mutations (Maharjan & Ferenci 2015; Gerrish & Lenski 1998; Desai & D S Fisher 2007; Lang et al. 2013; Perfeito et al. 2007), (3) second-order selection of mutations which lead to increased mutation rates and mutator phenotypes (Maharjan & Ferenci 2015; Sniegowski et al. 1997; Desai et al. 2007; Tenaillon et al. 2001; Perfeito et al. 2007; Barrick et al. 2009; Wielgoss et al. 2013), or (4) the emergence of negative frequency-dependent interactions between genotypes (Desai & D S Fisher 2007; Maharjan & Ferenci 2015; Lang et al. 2013; Maharjan 2006; Gerrish & Lenski 1998; Desai et al. 2007; Herron & Doebeli 2013; Perfeito et al. 2007). It is increasingly evident that not only these dynamics influence the adaptive process but also that they emerge as a result of the adaptive process.

For instance, the fixation of mutator phenotypes has been typically observed in adapting populations, as their higher mutation rate provides them with a higher probability of acquiring and hitchhiking with rare beneficial mutations (Desai & D S Fisher 2007; Chao & Cox 1983; Lang et al. 2013; Taddei et al. 1997; Tanaka et al. 2003; Gentile et al. 2011; Torres-Barceló et al. 2013). More recently, experimental findings from microbial evolution experiments coupled with sequencing analysis unveiled that a dramatic level of polymorphism in populations can occur during adaptation. Interestingly, a pattern of aggregation and hitchhiking of multiple mutations is observed – cohorts – in populations adapting to the same environmental laboratory conditions. Synchronous increase or decrease in frequency of these mutations, competition between distinct cohorts and the simultaneous fixation of the mutations that form the cohorts is pervasive during this laboratory microbial adaptations (Sniegowski et al. 1997; Lang et al. 2013; Tenaillon et al. 2001; Ming-Chun Lee & Marx 2013; Barrick et al. 2009; Maddamsetti et al. 2015; Wielgoss et al. 2013).

A classical model of adaptation to a novel environment is Fisher's Geometrical Model (FGM), where a population adapts towards a fixed optimum (Ronald Aylmer Fisher 1930). FGM considers the process of adaptation assuming that individuals are defined by their traits under selection, which are geometrically represented in a defined multidimensional landscape. In this model, directionality in selection emerges by assuming that fitness is related to the distance of each phenotype to the optimum. Thus, a population moves towards the fitness peak through the gradual accumulation of beneficial mutations. FGM has been extensively studied beyond its original scope to make predictions under different scenarios about the distribution of beneficial mutations during adaptation (Orr 1998; Martin & Lenormand 2008; Bataillon et al. 2011), the level of epistasis between mutations (Martin et al. 2007; Blanquart et al. 2014), the effects of deleterious mutations accumulated under relaxed selection (Martin & Lenormand 2006; Perfeito et al. 2014), the effect of drift load in the fitness at equilibrium (Otto & Orive 1995; Lourenço et al. 2011), sympatric speciation in an environment with

multiple fitness peaks (Barton 2001; Sellis et al. 2011) and the effect of mutation pleiotropy (the number of traits affected by a single mutation) in adaptation (Welch & Waxman 2003; Chevin, Martin, et al. 2010b; Lourenço et al. 2011). Martin (Martin 2014) recently proposed that FGM basic assumptions can emerge from models which consider the nature of complex metabolic networks within a cell. FGM predictions are largely compatible with observations coming from experimental evolution studies, mostly in microorganisms (MacLean et al. 2010; Chou et al. 2011; Khan et al. 2011; Gordo & Campos 2013; Sousa et al. 2012; Weinreich & Knies 2013; Tenaillon 2014).

Here we ask whether the patterns of competition and fixation of simultaneous segregating mutations (cohorts) along an adaptive walk observed experimentally can be reproduced under FGM. Since the simplest version of FGM assumes full mutational pleiotropy, which is a restrictive assumption and thought to bear poor biological realism (Welch & Waxman 2003; Orr 2005; Wang et al. 2010; Chevin, Lande, et al. 2010a; Wagner & Zhang 2011; Lourenço et al. 2011), we also studied a model assuming partial pleiotropy. The degree of mutational pleiotropy is expected to influence the dynamics of adaptation (Wagner & Zhang 2011). In our model of partial pleiotropy, similar to that of (Lourenço et al. 2011), a single mutation can only change a subset of traits (m), taken at random from the full set of traits (n) that contribute to fitness. When populations have small sizes or mutation rates are low, the analytical expressions for predicting the rate of adaptation under this model suggest that mutational pleiotropy can affect the dynamics of adaptation of populations approaching a fitness peak (Lourenço et al. 2011). However, such analytical results rely on a strong simplifying assumption: the populations are monomorphic most of the time. This assumption is quite restrictive given the increasing experimental evidence for high rates of beneficial mutations both in natural (Eyre-Walker & Keightley 2007; Jensen et al. 2008) and in experimental populations (Perfeito et al. 2007; Good et al. 2012), which promptly produce competition between segregating mutations arising in distinct lineages and drive the dynamics of cohorts described above. To address these

more relevant scenarios, we use stochastic simulations of FGM for populations undergoing strong clonal interference. We consider large populations and values of mutation rate and mean effect of mutations that are in reasonable agreement with current estimates for microbial populations (Gordo et al. 2011; Perfeito et al. 2014).

Most of theoretical analysis done so far focused on predicting the equilibrium mean fitness, and did not address the time scale at which such equilibrium is in fact reached. As experiments where evolution is followed for longer and longer periods are emerging (Barrick & Lenski 2013; Lang et al. 2013), it is also important to have theoretical expectations on the full dynamics of the approach to equilibrium under classical models of adaptation, both at the phenotypic and genotypic level, as we do here. By tracking each individual mutation during the adaptive walk as the populations approach the optimum, we find that the simplest version of FGM can generate the complex cohort dynamics observed in microbial adaptation experiments, under specific evolutionary parameters within a biological realistic range.

Methods

Simulation Methods of Fisher Geometrical Model

FGM considers each individual as a point in a *n*-dimensional space, where *n* is the number of traits under selection. Each individual is characterized by a vector of coordinates (z_1, z_2,…, z_n) that gives the position of the individual in the fitness landscape. This vector represents the phenotypic values for each trait. Without loss of generality, we define the optimum as the origin of the *n*-dimensional space. As commonly done, we assume that fitness is given by a Gaussian function of the distance to the optimum, $w = \exp\left(-\sum_{i=1}^{n} z_i^2\right)$. We assume that mutations follow a Poisson distribution with a genomic mutation rate *U*, per

individual, per generation. Each mutation changes *m* traits chosen at random from the total *n* traits, and the effect it causes in each affected trait follows a normal distribution with mean 0 and variance s^2. We consider a Wright-Fisher model and assume multinomial sampling with fixed population size *N*. The contribution of each individual to the next generation is proportional to its fitness. We assume large population sizes and consider values of genomic mutation rates (*U* ~0.001) that are reasonable estimates for the genome-wide mutation rate in DNA microbes (Drake et al. 1998; H Lee et al. 2012). The code for the simulations is provided as supplementary material.

Results

Dynamics of approach to equilibrium mean fitness

We start by studying the dynamics of fitness increase along tens of thousands of generations for different levels of phenotypic complexity, pleiotropy and mean effect mutations. **Figure 1A** shows that the initial rate of fitness increase is lower under low pleiotropy across all values of the mean fitness effect of mutations (E(S)) studied. The effect is particularly strong for large values of E(S) (E(S)>0.01). However, in the long run populations with lower pleiotropy reach higher levels of mean fitness (see also **Supplementary Figure S1**). Increasing genetic complexity (n), while maintaining a similar level of pleiotropy, shows a similar pattern for the fitness plateau, where populations with fewer traits reach higher fitness values (**Figure 1B**).

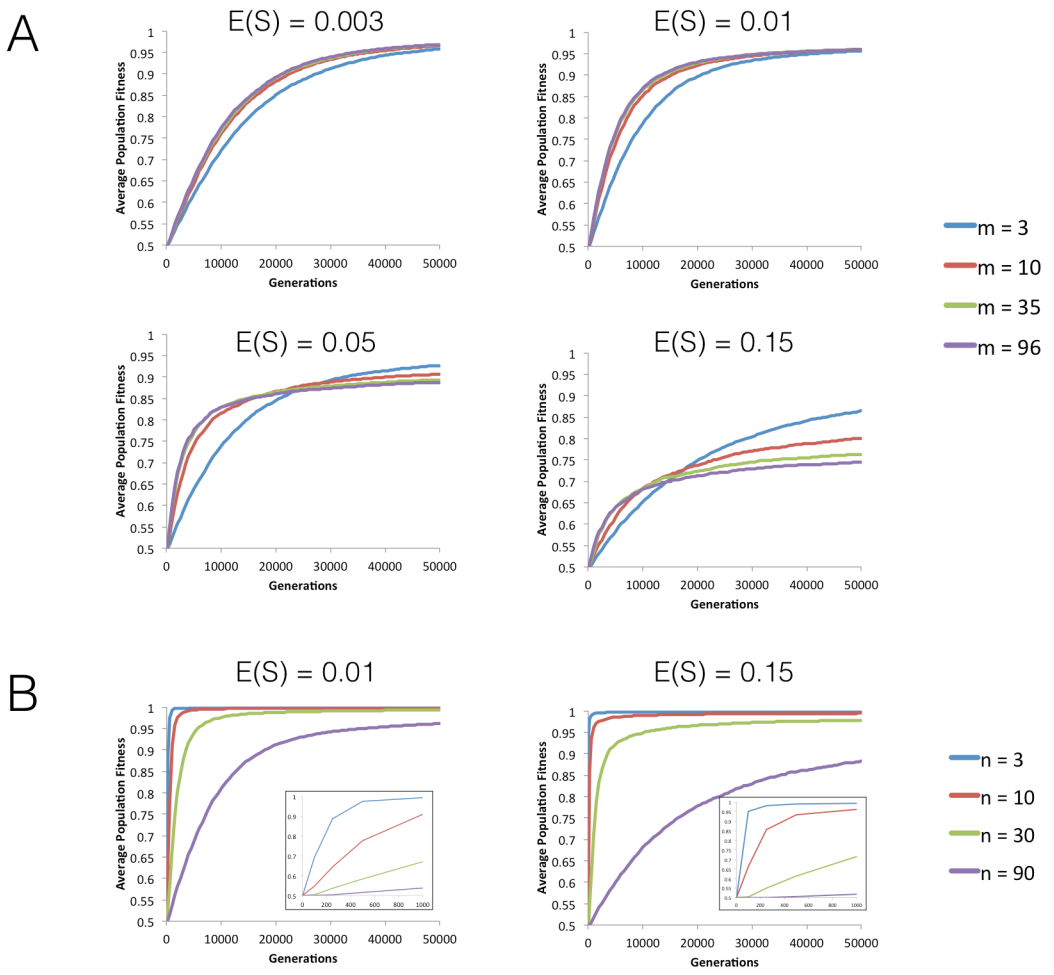


Figure 1: Dynamics of mean fitness increase under FGM with partial pleiotropy

A) Dynamics of mean fitness of asexual populations with varying degrees of pleiotropy. All populations have high complexity (n=96) and distributions of fitness effects (DFEs) with different means are studied. Other parameters are N=10^4, U=0.001 and the initial fitness w_0=0.5. B) The effect of increasing phenotypic complexity (n) on the dynamics of fitness increase. Other population parameters are: N=10^4, U=0.001, s^2=0.004, m=3 and the initial fitness w_0=0.5. Short term dynamics are highlighted as an inset.

Cohort of mutations fixed in the initial steps of adaptation

Next, we studied the dynamics of mutation fixation along the adaptive walk**.** We first studied populations with maximum pleiotropy and various degrees of complexity across different E(S) and mutation rates, and ask how many mutations fix simultaneously in the first step. **Figure 2** shows that fixations of cohorts of mutations can be very common, reflecting the degree of clonal interference occurring in these large populations. Across all parameters, the major determinant of the number of mutations fixing in cohorts is the mean effect of mutations (E(S)), with lower effect mutations promoting fixation of cohorts of larger size. The other relevant parameter to the size of the fixed cohorts, as expected, is the mutation rate, with an increased mutation rate showing the largest cohorts of mutations fixed. Therefore, the combination of small effect mutations generated at a high rate leads to the fixation of larger cohorts. We performed the same analysis on simulations where we relax the assumption of full pleiotropy. Populations with partial pleiotropy (m=3, 10 or 20) for the highest level of complexity previously tested (n=30) show patterns that are qualitatively similar (**Supplementary Figure S2**). The main difference detected occurs in simulations with a high mutation rate, where the likelihood of observing large cohorts of stronger effect mutations increases relative to the case of full pleiotropy. Additionally, both in the cases of full or partial pleiotropy, the complexity of evolving populations shows a minimal effect on the size of the fixed cohorts of mutations. Therefore, the number of mutations observed fixing simultaneously in the first step of adaptation is mainly determined by the mutation rate and the mean selective effect of mutations.

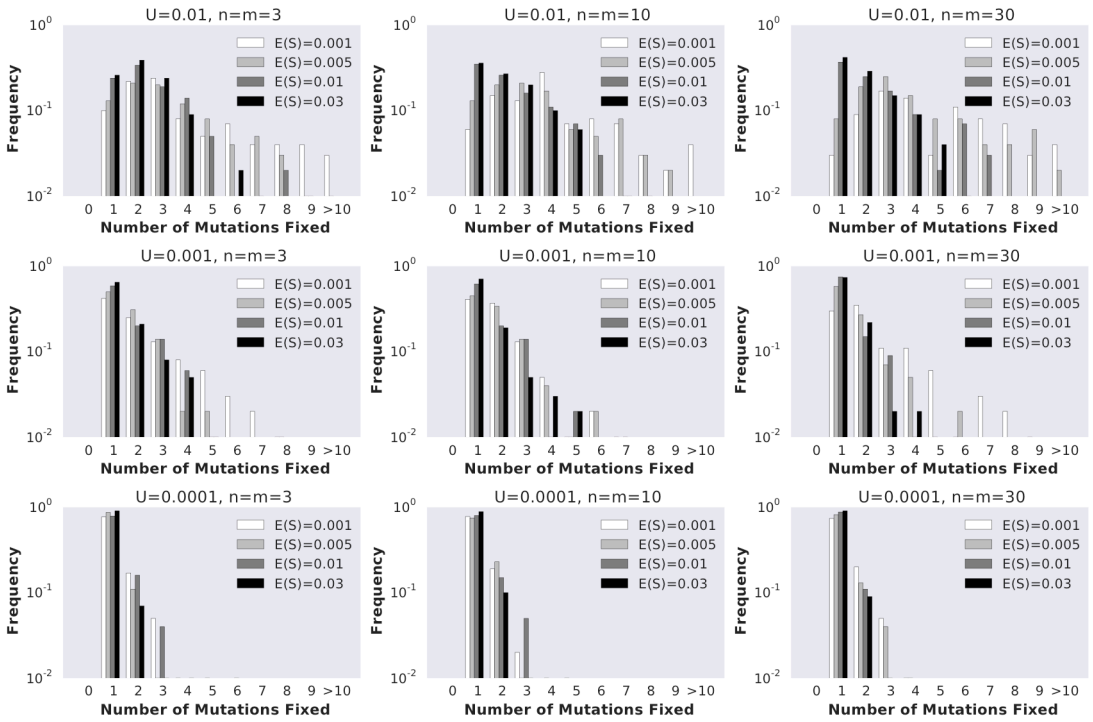


Figure 2: FGM can lead to simultaneous fixation of cohorts of mutations. The probability distribution of the number of mutations fixed during the first fixation event in the adaptive walk. Parameter values are $N=10^4$, the initial fitness $w_0=0.5$ and other parameters as indicated in each panel. Data is shown for 100 simulations per combination of parameters.

Cohort of mutations fixed along the adaptive walk

In order to understand how the probability of observing the fixation of cohorts changes along the adaptive walk, we next study the distribution of mutations fixed beyond the first step of adaptation. Figure 3 shows the pattern of cohorts fixed along an adaptive walk lasting 30000 generations. Each point in the panels of **Figure 3** corresponds to a fixation event occurring during this time period, with the number of mutations (i.e., the size of the cohorts) that compose each of these fixations represented in the y-axis. The probability of observing cohorts consisting of a large number of mutations later in the adaptive walk is strongly dependent on the average selective effect of these mutations. Lower effect mutations lead to the fixation of cohorts of larger sizes not only in the first steps, but also as populations approach the equilibrium fitness. Interestingly, we observe that, for the lower values of E(S), the likelihood of fixing cohorts of larger sizes (from 4 to 8 mutations) increases for populations with a higher complexity, throughout the adaptive walk. For high values of E(S) and populations with a lower number of traits, fixation of large cohorts becomes an increasingly rare event once they approach the fitness equilibrium. Overall, along the adaptive walk, the size of fixed cohorts tends to shrink, at a faster pace for large E(S). Therefore, for long-term adaptation in populations approaching a fixed optimum, fixation of single mutations is expected to become the dominant pattern. However, when E(S) is small (right panel in **Figure 3**) that regime may take a substantial time to be reached.

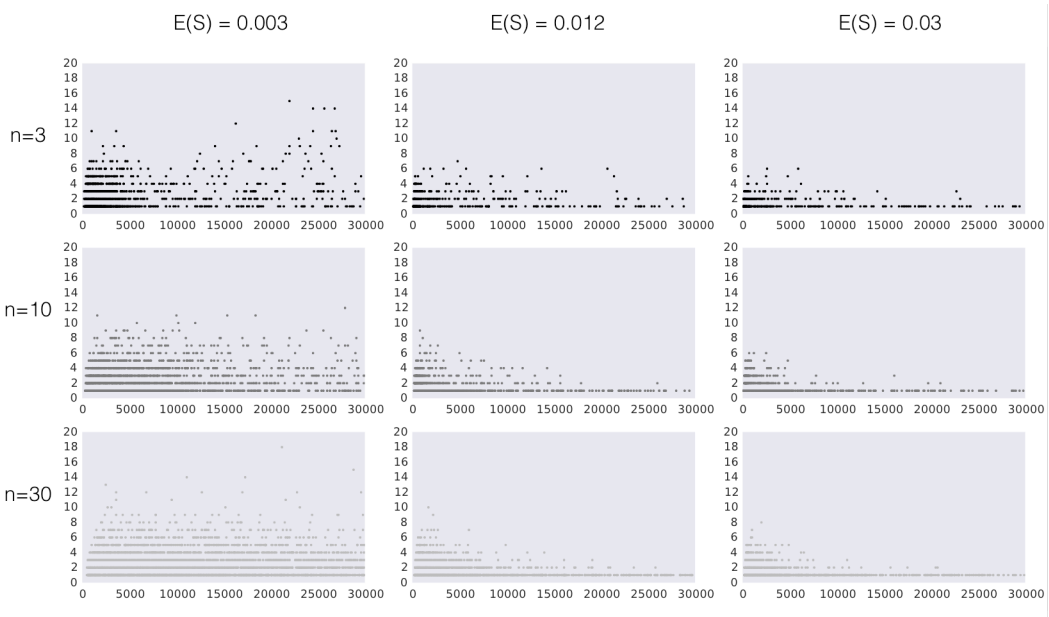


Figure 3: Number of mutations fixed (cohort size) along the adaptive walk. Parameter values are $N=10^4$, the initial fitness $w_0=0.5$ and other parameters as indicated in each panel. In all panels full pleiotropy is assumed (m=n). Data is shown for 100 simulations per combination of parameters.

Dynamics of cohorts of mutations

Finally, we study the dynamics of polymorphism expected in populations climbing the Fisherian landscape. We focus our simulations on short-term evolution, a time scale for which polymorphism data has been obtained recently (Lang et al. 2013). **Figure 4** shows the dynamics of frequency change of each individual mutation segregating in populations evolving for 1000 generations. Aggregation of cohorts of mutations can be clearly observed across the different replicate populations, all with the same evolutionary parameters. The parameter set shown was chosen to be one where we could find a pattern similar to that observed in the evolution experiments done in yeast (Lang et al. 2013). In the replicate simulated populations, cohorts of different sizes emerge and compete against each other, with some achieving fixation and others being outcompeted.

Although this phenomenon of "cohort interference" is more likely for cohorts competing at lower frequencies (where many mutations are segregating), it can also be observed when mutations reach high frequencies (e.g., first and third panels in the first row). Even under the same parameter values different patterns can be observed among the replicates: sequential fixation of cohorts of low size in some populations (e.g., fourth and seventh panels) and fixation of cohorts of large size in other populations (e.g., fifth, sixth and ninth panels). The same qualitative behavior is observed when simulating a higher number of replicate populations adapting under FGM (see **Supplementary Figure S3**).

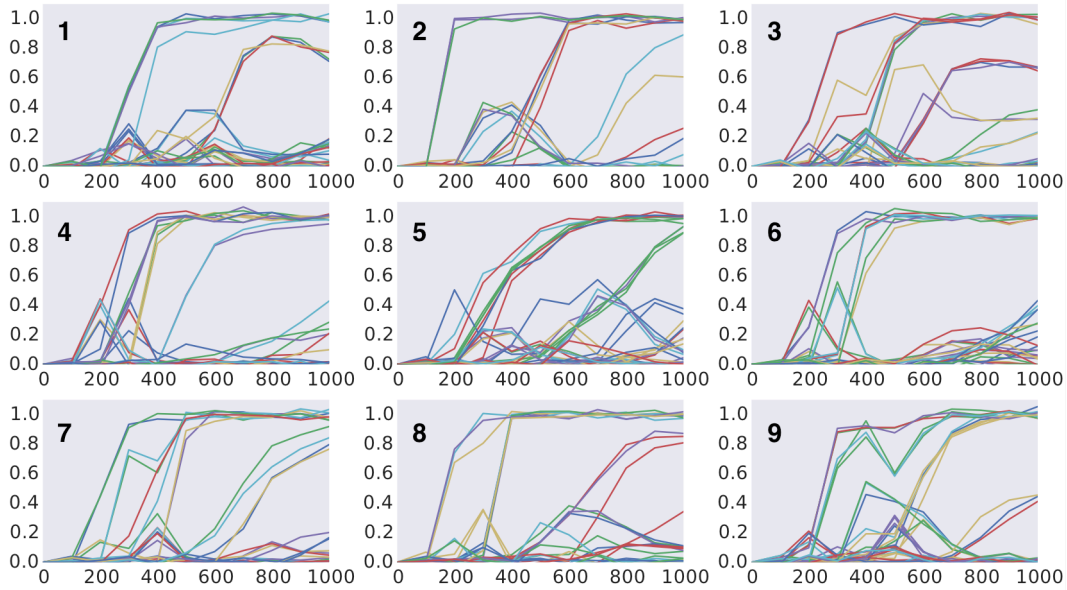


Figure 4: Dynamics of frequency change of individual mutations along time across independently evolving populations. Aggregation of multiple mutations in cohorts can be easily detected by the simultaneous increase in frequency of different mutations (in different colors). Competition between cohorts can be commonly observed during the first 1000 generations of adaptation. Parameter values used are N=10^5, the initial fitness w_0=0.5, n=10, m=3 and E(S)=0.012. Gaussian noise (mean 0, variance 0.02), mimicking experimental error, was added to the dynamics for increasing visibility and comparison with experimental data.

Discussion

With the advances of next generation sequencing, increased power to observe the dynamics of adaptation at the resolution of individual mutations has emerged. The data recently gathered indicates that adaptation of microbial populations adapting in laboratory environments exhibit patterns very distinct from the classic single selective sweep model of periodic selection. Instead the dynamics of molecular evolution in these microbes evolving in real time shows that aggregates of beneficial mutations segregate and fix simultaneously (Lang et al. 2013; Maddamsetti et al. 2015; Zanini et al. 2016). Given the easiness of FGM to produce dynamics of fitness change similar to those observed in such experiments, we asked if such dynamics of molecular change could be expected under this model. The simulations performed, show that Fisher's Geometric Model, in its simplest version, can reproduce dynamics of cohort interference such as the ones observed in experimental settings. Cohort interference can be common during the initial steps of adaptation, but it is more likely if the mean effect of mutations is small and mutation rates are not too small. In these scenarios many small effect mutations simultaneously segregate, each taking a long time to reach fixation, which likely results in the acquisition of additional mutations (either beneficial, neutral or slightly deleterious) in the same genetic background. In contrast, when E(S) is large, beneficial mutations sweep to fixation faster, and the likelihood of acquiring additional mutations in their background diminishes. As expected the size of the interfering cohorts increases as the mutation rate increases, since an increased amount of mutations segregating in these high U populations prevents the fast fixation of a single mutation. Remarkably when simulating the dynamics of individual mutations produced under FGM, we could find patterns very similar to the ones that are increasingly being assayed through whole genome sequencing of evolving microbial populations (e.g. compare **Figure 4** with Figure 1 in (Lang et al. 2013)). Although such pattern is dependent on the parameters used, it could be observed in simulated populations assuming a set of parameters within a biological plausible range: a mean effect of mutations around

1% (Eyre-Walker & Keightley 2007; Gordo et al. 2011) and a genomic mutation rate of 10^{-3} (H Lee et al. 2012).

The relationship between the size of cohorts and both the mean effect of mutations and the mutation rate is also detected when we study adaptation over longer periods (**Figure 3**). The size of fixed cohorts tends to shrink along the adaptive walk, and does so at a faster pace for large values of E(S). Therefore, for populations approaching a fixed optimum the pattern of long-term adaptation is expected to become dominated by fixation of single mutations. However if E(S) is small such pattern may take many thousands of generations to be detected (right panels in **Figure 3**), a time scale that is out of reach for most laboratory experiments so far studied. The famous LTEE in *Escherichia coli* constitutes an important exception, where patterns of adaptation can be studied over periods as long as 60000 generations (Lenski et al. 1991; Maddamsetti et al. 2015). The access to samples frozen every 500 generations allows the tracking of individual mutations and the reconstruction of the evolutionary genetic history of an individual population. Maddamsetti et al tracked the emergence of 42 mutations in one of the evolving populations and showed competition and interference between lineages carrying several mutations, including the simultaneous fixation of these sets (Maddamsetti et al. 2015). In this population however, not only clonal interference was observed but also frequency-dependent selection was important in driving the dynamics of mutation cohorts. On a shorter-term experiment, also with *E. coli* but now evolving in a chemostat, Maharjan and colleagues detected synchronous sweep of multiple mutations but the levels of polymorphism were also driven by frequency dependent interactions between clones (Maharjan et al. 2015). As we show here clonal interference alone can lead to dynamics of cohort interference, but given the emergence of frequency dependent selection even in the simplest environments it will be important in future work to model other fitness landscapes which can allow for the simultaneous occurrence of both processes.

Conclusions

In the current work, we study a simple version of the Fisher's Geometrical Model that assumes partial or full pleiotropy. Despite its simplicity, FGM has been successfully used to reproduce patterns of the dynamics of the adaptive process (Chevin, Martin, et al. 2010b; Martin 2014). A common pattern emerging from the short-term dynamics of populations of microorganisms evolving in laboratory conditions is the finding that mutants carrying multiple segregating mutations can go to fixation (Lang et al. 2013; Maddamsetti et al. 2015). Before resorting to more complex models of fitness landscapes, we inquired whether a simple and less parameterized model, such as FGM, could capture the essence of this sort of observation under reasonable parameters. Assuming large population sizes close to those in the experiments, and mutation rates typical of microbes, thus naturally driving population to a clonal interference regime, we show that FGM, both under full and partial pleiotropy, generates patterns of segregation and competition of cohorts of mutations that are consistent with experimental observations.

Acknowledgements

IG thanks Olivier Tenaillon for pointing out the potential importance of partially pleiotropy leading to different patterns of mutation accumulation.

References

Barrick JE et al. 2009. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature. 461:1243–1247. doi: 10.1038/nature08480.

Barrick JE, Lenski RE. 2013. Genome dynamics during experimental evolution.

Nature Publishing Group. 14:827–839. doi: 10.1038/nrg3564.

Barton NH. 2001. The role of hybridization in evolution. Mol Ecol.

Bataillon T, Zhang T, Kassen R. 2011. Cost of Adaptation and Fitness Effects of Beneficial Mutations in Pseudomonas fluorescens. Genetics. 189:939–949. doi: 10.1534/genetics.111.130468.

Blanquart F, Achaz G, Bataillon T, Tenaillon O. 2014. Properties of selected mutations and genotypic landscapes under Fisher's geometric model. Evolution.

Chao L, Cox EC. 1983. Competition between high and low mutating strains of Escherichia coli. Evolution.

Chevin L-M, Lande R, Mace GM. 2010a. Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory Kingsolver, JG, editor. PLoS Biol. 8:e1000357. doi: 10.1371/journal.pbio.1000357.s001.

Chevin L-M, Martin G, Lenormand T. 2010b. FISHER'S MODEL AND THE GENOMICS OF ADAPTATION: RESTRICTED PLEIOTROPY, HETEROGENOUS MUTATION, AND PARALLEL EVOLUTION. Evolution. 64:3213–3231. doi: 10.1111/j.1558-5646.2010.01058.x.

Chou HH, Chiu HC, Delaney NF, Segre D, Marx CJ. 2011. Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. Science. 332:1190–1192. doi: 10.1126/science.1203799.

Desai MM, Fisher DS. 2007. Beneficial Mutation Selection Balance and the Effect of Linkage on Positive Selection. Genetics. 176:1759–1798. doi: 10.1534/genetics.106.067678.

Desai MM, Fisher DS, Murray AW. 2007. The Speed of Evolution and Maintenance of Variation in Asexual Populations. Current Biology. 17:385–394. doi: 10.1016/j.cub.2007.01.072.

Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. Genetics. 148:1667–1686.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nature Publishing Group. 8:610–618. doi: 10.1038/nrg2146.

Fisher RA. 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Gentile CF, Yu S-C, Serrano SA, Gerrish PJ, Sniegowski PD. 2011. Competition between high- and higher-mutating strains of Escherichia coli. Biology Letters. 7:422–424. doi: 10.1098/rsbl.2010.1036.

Gerrish PJ, Lenski RE. 1998. The fate of competing beneficial mutations in an asexual population. Genetica. 102-103:127–144.

Good BH, Rouzine IM, Balick DJ. 2012. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. doi: 10.1073/pnas.1119910109/-/DCSupplemental.

Gordo I, Campos PRA. 2013. Evolution of clonal populations approaching a fitness peak. Biology Letters. 9:20120239. doi: 10.1098/rsbl.2012.0239.

Gordo I, Perfeito L, Sousa A. 2011. Fitness Effects of Mutations in Bacteria. J Mol Microbiol Biotechnol. 21:20–35. doi: 10.1159/000332747.

Herron MD, Doebeli M. 2013. Parallel evolutionary dynamics of adaptive diversification in Escherichia coli. PLoS Biol. 11:e1001490. doi: 10.1371/journal.pbio.1001490.

Jensen JD, Thornton KR, Andolfatto P. 2008. An Approximate Bayesian Estimator Suggests Strong, Recurrent Selective Sweeps in Drosophila McVean, G, editor. PLoS Genet. 4:e1000198. doi: 10.1371/journal.pgen.1000198.s005.

Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. 2011. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. Science. 332:1193–1196. doi: 10.1126/science.1203801.

Lang GI et al. 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature. 1–6. doi: 10.1038/nature12344.

Lee H, Popodi E, Tang H. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing.

Lee M-C, Marx CJ. 2013. Synchronous waves of failed soft sweeps in the laboratory: remarkably rampant clonal interference of alleles at a single locus. Genetics. 193:943–952. doi: 10.1534/genetics.112.148502.

Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. American Naturalist.

Lourenço J, Galtier N, Glémin S. 2011. Complexity, pleiotropy, and the fitness effect of mutations. Evolution. 65:1559–1571. doi: 10.1111/j.1558-5646.2011.01237.x.

MacLean RC, Perron GG, Gardner A. 2010. Diminishing Returns From Beneficial Mutations and Pervasive Epistasis Shape the Fitness Landscape for Rifampicin Resistance in Pseudomonas aeruginosa. Genetics. 186:1345–1354. doi: 10.1534/genetics.110.123083.

Maddamsetti R, Lenski RE, Barrick JE. 2015. Adaptation, Clonal Interference, and Frequency-Dependent Interactions in a Long-Term Evolution Experiment with Escherichia coli. Genetics. 200:619–631. doi: 10.1534/genetics.115.176677.

Maharjan R. 2006. Clonal Adaptive Radiation in a Constant Environment. Science. 313:514–517. doi: 10.1126/science.1129865.

Maharjan R, Ferenci T. 2015. Mutational signatures indicative of environmental stress in bacteria. Molecular Biology and Evolution. 32:380–391. doi: 10.1093/molbev/msu306.

Maharjan RP, Liu B, Feng L, Ferenci T, Wang L. 2015. Simple phenotypic sweeps hide complex genetic changes in populations. Genome Biology and Evolution. 7:531–544. doi: 10.1093/gbe/evv004.

Martin G. 2014. Fisher's geometrical model emerges as a property of complex integrated phenotypic networks. Genetics. 197:237–255. doi: 10.1534/genetics.113.160325.

Martin G, Elena SF, Lenormand T. 2007. Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nature Genetics.

Martin G, Lenormand T. 2006. A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. Evolution. 60:893–907.

Martin G, Lenormand T. 2008. The Distribution of Beneficial and Fixed Mutation Fitness Effects Close to an Optimum. Genetics. 179:907–916. doi: 10.1534/genetics.108.087122.

Orr HA. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution. 935–949.

Orr HA. 2005. Theories of adaptation: what they do and don’t say. Genetica. 123:3–13.

Otto SP, Orive ME. 1995. Evolutionary consequences of mutation and selection within an individual. Genetics.

Perfeito L, Fernandes L, Mota C, Gordo I. 2007. Adaptive Mutations in Bacteria: High Rate and Small Effects. Science. 317:813–815. doi:

10.1126/science.1142284.

Perfeito L, Sousa A, Bataillon T, Gordo I. 2014. Rates of fitness decline and rebound suggest pervasive epistasis. Evolution.

Sellis D, Callahan BJ, Petrov DA, Messer PW. 2011. Heterozygote advantage as a natural consequence of adaptation in diploids. Proc. Natl. Acad. Sci. U.S.A. 108:20666–20671. doi: 10.1073/pnas.1114573108.

Sniegowski PD, Gerrish PJ, Lenski RE. 1997. Evolution of high mutation rates in experimental populations of E. coli. Nature. 387:703–705. doi: 10.1038/42701.

Sousa A, Magalhaes S, Gordo I. 2012. Cost of Antibiotic Resistance and the Geometry of Adaptation. Molecular Biology and Evolution. 29:1417–1428. doi: 10.1093/molbev/msr302.

Taddei F et al. 1997. Role of mutator alleles in adaptive evolution. Nature. 387:700–702. doi: 10.1038/42696.

Tanaka MM, Bergstrom CT, Levin BR. 2003. The evolution of mutator genes in bacterial populations: the roles of environmental change and timing. Genetics. 164:843–854.

Tenaillon O. 2014. The utility of Fisher's geometric model in evolutionary genetics. Annual Review of Ecology.

Tenaillon O, Taddei F, Radmian M, Matic I. 2001. Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation. Research in Microbiologoy. 152:11–16.

Torres-Barceló C, Cabot G, Oliver A, Buckling A, MacLean RC. 2013. A trade-off between oxidative stress resistance and DNA repair plays a role in the evolution of elevated mutation rates in bacteria. Proceedings of the Royal Society B: Biological Sciences. 280:20130007. doi: 10.1098/rspb.2013.0007.

Wagner GP, Zhang J. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nature Publishing Group. 12:204–213. doi: 10.1038/nrg2949.

Wang Z, Liao B-Y, Zhang J. 2010. Genomic patterns of pleiotropy and the evolution of complexity. Proc. Natl. Acad. Sci. U.S.A. 107:18034–18039. doi: 10.1073/pnas.1004666107.

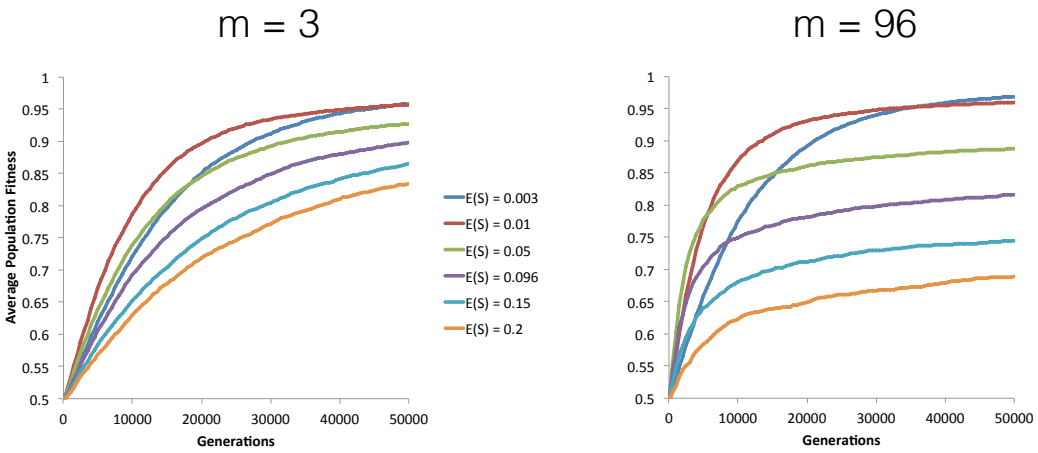
Weinreich DM, Knies JL. 2013. Fisher's geometric model of adaptation meets the functional synthesis: data on pairwise epistasis for fitness yields insights into the shape and size of phenotype space. Evolution. 67:2957–2972. doi: 10.1111/evo.12156.

Welch JJ, Waxman D. 2003. Modularity and the cost of complexity. Evolution. 57:1723–1734.

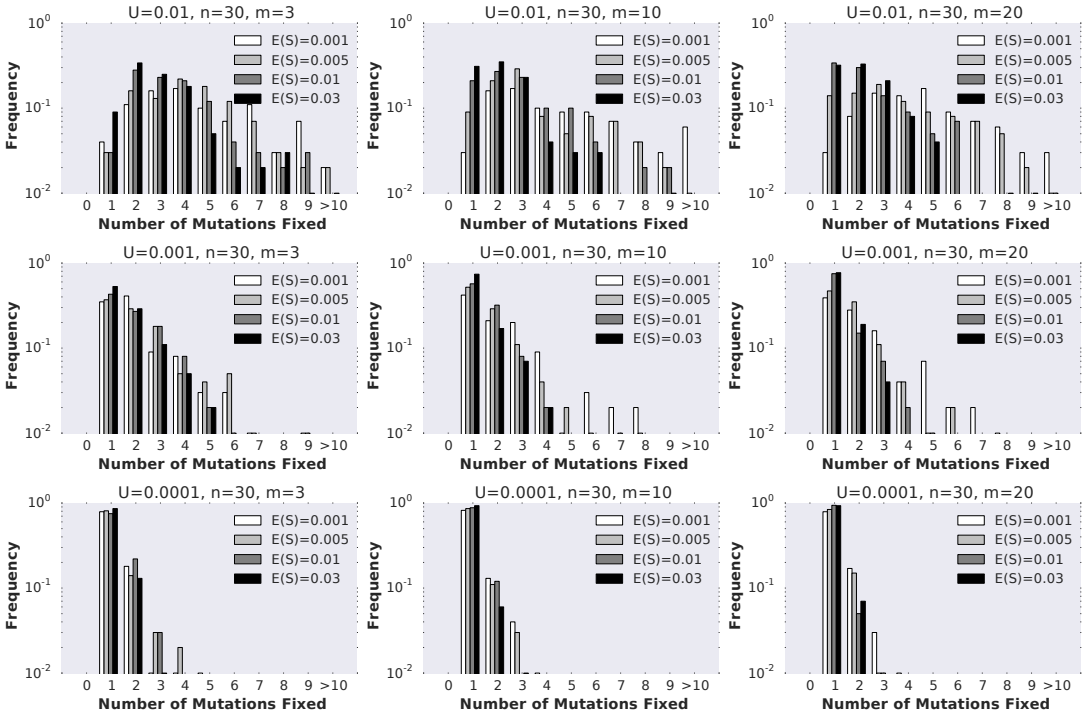
Wielgoss S et al. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. Proceedings of the National Academy of Sciences. 110:222–227. doi: 10.1073/pnas.1219574110.

Zanini F et al. 2016. Population genomics of intrapatient HIV-1 evolution. eLife. doi: 10.7554/eLife.11282.001.

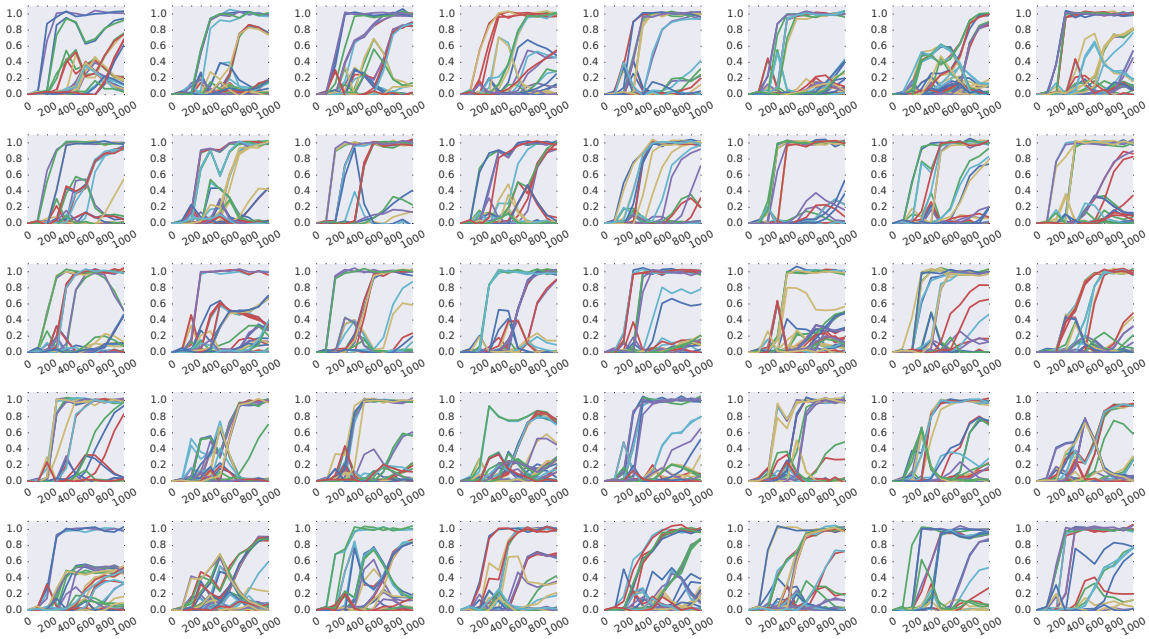
Supplementary Figures



Supplementary Figure S1: Long term dynamics of mean fitness of populations as a function of the mean effect of mutations (E(S)). Dynamics are shown for high and low levels of pleiotropy (m=96 and m=3). Other parameters are as in **Figure 1**.



Supplementary Figure S2: Effect of pleiotropy in the simultaneous fixation of cohorts of mutations during the first step of adaptation. The complexity, n, is fixed at 30 while the pleiotropy, m, varies between 3, 10 and 30 (full pleiotropy). Remaining parameters are as in **Figure 3**.



Supplementary Figure S3: Additional simulations showing the dynamics of frequency change of individual mutations along time. The diverse patterns of cohorts observed emerge from the dynamics of adaptation under FGM. Parameters are as in **Figure 4**.

CHAPTER IV

Epistatic interactions shape compensatory adaptation of multiple antibiotic resistances

Manuscript in preparation

The author of this thesis performed the competitive and sensitivity assays, theoretical simulations and estimation of parameters and the analysis of the whole genome sequencing of the evolved populations. Hajra-bibi Ali provided technical help in the propagations and competitive assays, as well as the DNA extraction for sequencing. Roberto Balbontín performed the allelic reconstructions.

Epistatic interactions shape compensatory adaptation of multiple antibiotic resistances

Jorge A. Moura de Sousa[1], R. Balbontín[1], P. Durão[1], H. Ali[1], I. Gordo[1]

[1] Instituto Gulbenkian de Ciência, Oeiras, Portugal

Abstract

The rate and effects of compensatory mutations are key determinants for the maintenance of costly antibiotic resistance mutations. While compensation of single resistance mutations has been well studied, this process may differ in multi-resistant strains, due to epistasis between resistance alleles. Using experimental evolution and whole genome sequencing, we compared the pace and genetic basis of compensation of a streptomycin and rifampicin double resistant *Escherichia coli* with those of each of the respective single mutants. We found minimal overlap between the compensatory mutations acquired by double resistant bacteria and those found in single resistance strains, and provide direct evidence for mutations that compensate specifically for the interaction between the resistant alleles. Furthermore, the pace of compensatory evolution is higher in the double resistant strain due to rapid accumulation of adaptive mutations with stronger fitness benefits. This is the first study describing the compensatory landscape of double resistant bacteria and demonstrating that the process of compensation is shaped by the epistatic interaction between mutations for different resistances.

Introduction

Antibiotics are one of the most important medical advancements in history, preventing millions of deaths from bacterial infections. Nevertheless, bacteria engage in a constant arms race, through the acquisition of genes or chromosomal mutations that confer resistance to the drugs (J Davies & D Davies 2010; Hughes & Andersson 2015; van Hoek et al. 2011). Resistance mutations are widely spread in bacterial populations, both in clinical (Gullberg et al. 2011; MacLean & Vogwill 2015; Hughes & Andersson 2015) and environmental settings (Forsberg et al. 2012; Bhullar et al. 2012), providing a reservoir of genetically encoded antibiotic resistance that can be transmitted by horizontal gene transfer (van Hoek et al. 2011; Kriegeskorte & Peters 2012). Moreover, as microbes become resistant to a specific drug, the subsequent use of alternative antibiotics might create further resistance, thus leading to the increasing threat of multi-drug resistant strains (Hughes & Andersson 2015; Mwangi et al. 2007; Velayati et al. 2009; Zhang et al. 2013). This is a prevalent case in *Staphylococcus aureus*, *Escherichia coli* or *Mycobacterium tuberculosis*, for instance, where multiple resistances pose a serious threat to human health (Hede 2014; Petty et al. 2014; Mwangi et al. 2007; Warner et al. 2014). Determining the evolutionary dynamics governing the emergence of these mutations and the mechanisms behind resistance is paramount to design more effective treatments. Although the effects vary across environments, mutations conferring resistance are typically deleterious in the absence of antibiotics (Schrag et al. 1997; Sousa et al. 2012), where resistant strains should be outcompeted by sensitive bacteria. Yet, the widespread prevalence of resistance mutations in bacteria indicates that populations can curb this deleterious effect. In order to understand how antibiotic resistant bacteria survive, spread in populations and acquire further resistances, it is crucial to realize how this fitness cost is overcome.

Mutations conferring antibiotic resistance typically alter bacterial targets involved in key processes, such as the ribosome, for the case of resistance to stremptomycin (Luzzatto & Apirion 1968; Couturier et al. 1964) or the RNA

polymerase, for resistance to rifampicin (Wehrli 1983; Goldstein 2014). The changes in essential machinery caused by resistance mutations usually lead to deleterious effects and, in the absence of antibiotics, resistant strains face different possible outcomes: extinction, reversion of the resistance mutation, or acquisition of further mutations that compensate for the fitness cost (Schrag et al. 1997; Reynolds 2000; Andersson & Hughes 2010; Brandis et al. 2012). The latter is, arguably, the worse outcome for the host, since antibiotic resistance alleles might become stabilized in the population while paying little or even no cost (Levin et al. 2000). Acquiring additional mutations to overcome the fitness cost of resistance is also more likely to occur than the reversion of the resistance mutation, since the genomic target range for compensation is much broader (Levin et al. 2000; Poon 2005). Therefore, understanding the key evolutionary dynamics of compensatory adaptation is necessary to predict and counteract the survival and spread of antibiotic resistant strains.

Compensation for the cost of resistance by further genetic changes has been widely described both in clinical (Comas et al. 2012; Zhang et al. 2013) and laboratory (Reynolds 2000; Maisnier-Patin et al. 2002; Qi et al. 2016) settings. The rate of compensation depends on the resistance alleles, whose compensatory targets might differ in nature and size, and might have different distributions of effects (Moura de Sousa et al. 2015). The dynamics of compensatory adaptation depend also on populational parameters, such as bottleneck size (Maisnier-Patin et al. 2002) and mutation rate (Levin et al. 2000). Over the last decades, numerous targets of compensation for single resistance alleles have been described (e.g. (Maisnier-Patin et al. 2002; Brandis & Hughes 2013; Qi et al. 2016)). However, compensation in bacteria harbouring more than one resistance mutation is likely to be different, because these alleles are often known to interact (Borrell & Gagneux 2011). This epistasis might originate from the fact that resistance mutations affect functionally interconnected cell machinery (Chakrabarti & Gorini 1977; Chao 1978). Thus, different, co-occuring, genetic modifications (for instance, in proteins involved in transcription and translation) might lead to

functional conflicts that diverge their cumulative effects from linearity (Trindade et al. 2009; Hall & MacLean 2011; Borrell et al. 2013). We hypothesize that, for cases of strong negative epistasis, genetic reversions in one of the original resistance alleles are amongst the strongest beneficial mutations, since it compensates both the individual effect of the reverted allele and the synergistic cost. Thus, reversions could be more frequently observed in these cases, compared to single resistance backgrounds, where their effect is limited to the cost of the resistance itself. Moreover, certain targets of compensation can be specific to these epistatic interactions, and thus the effects of compensatory mutations may be different in single or double resistance strains. In this scenario, the rate, targets and effects of compensatory mutations in multi-resistant bacteria could deviate considerably from the ones observed during compensation of single resistances.

In order to investigate this question, we performed experimental evolution, in antibiotic-free media, of three different *Escherichia coli* resistant genotypes: $RpoB^{H526Y}$ (conferring resistance to rifampicin), $RpsL^{K43T}$ (conferring resistance to streptomycin), and a double resistant mutant ($RpoB^{H526Y}RpsL^{K43T}$) that shows negative epistasis for fitness costs of double resistance. Neutral marker dynamics indicated a faster pace of adaptation in double resistant bacteria, while whole genome sequencing of clones of evolved populations revealed minor overlap between the genetic targets of compensation in double resistant strains and those in each of the single resistant alleles. Unexpectedly, genetic reversions of the $RpoB^{H526Y}$ allele were detected both in the single Rif^R and the double resistant background at non-significant different frequencies (2 out of 12 and 1 out of 24, respectively). While no reversion for Str^R was detected, overlap at the level of single nucleotide polymorphisms occurred in three of the putative compensatory targets for the Str^R mutation. By testing the effects of a specific allele in the original resistance backgrounds, we showed it could compensate specifically for the epistatic interaction between the resistances, being deleterious in the single resistant background where it was expected to provide a benefit. Together, these

results indicate that the pattern of compensation can be specific to the interactions between different resistance alleles in multi-drug resistant bacterial strains.

Faster pace of adaptation in double resistant populations

E. coli independent populations were founded from one of three different antibiotic resistant genetic backgrounds: 12 populations which are Rif^R ($RpoB^{H526Y}$), 12 populations which are Str^R ($RpsL^{K43T}$) and 24 double mutant populations Rif^RStr^R ($RpoB^{H526Y}RpsL^{K43T}$). These genetic backgrounds have different fitness costs (see **Extended Data Fig. 1**), and the double mutant shows negative epistasis (i.e. its fitness cost is bigger than the sum of the costs of the single resistances). The resistant backgrounds were tagged with a fluorescent marker (either CFP or YFP) and the replicate populations were serially passed for 22 days in rich media without antibiotics (see Methods). The fluorescent markers hitchhike with beneficial mutations and allow tracking adaptive events in evolving populations (Hegreness 2006). Therefore, the frequencies of the markers were followed during the adaptive process (**Fig. 1**). We observed that, for double resistant Rif^RStr^R *E. coli* (**Fig. 1c**), markers deviate from their initial frequency faster and with steeper slopes, compared to the populations in either one of the corresponding single resistant bacteria, Rif^R (**Fig. 1a**) or Str^R (**Fig. 1b**). In double resistant populations, the markers either sweep to near fixation within the first 6 days or show strong signs of clonal interference, which is indicative of multiple adaptive clones competing against each other. In single resistant bacterial populations, near fixation (>0.95) of a marker is only observed in a minority (3 out of 12 Rif^R and 1 out of 12 Str^R populations) of the independently evolving populations, and only after 14 days of evolution. Moreover, the dynamics of all Rif^R alone show no obvious sign of clonal interference, whereas in Str^R some populations are likely adapting by acquiring compensatory mutations in both CFP and YFP backgrounds (e.g., populations 3 and 11). As expected, given its much lower fitness, the double resistant populations show more signs of interference. This suggests that the pace of adaptation is faster for the bacteria harbouring two resistance alleles.

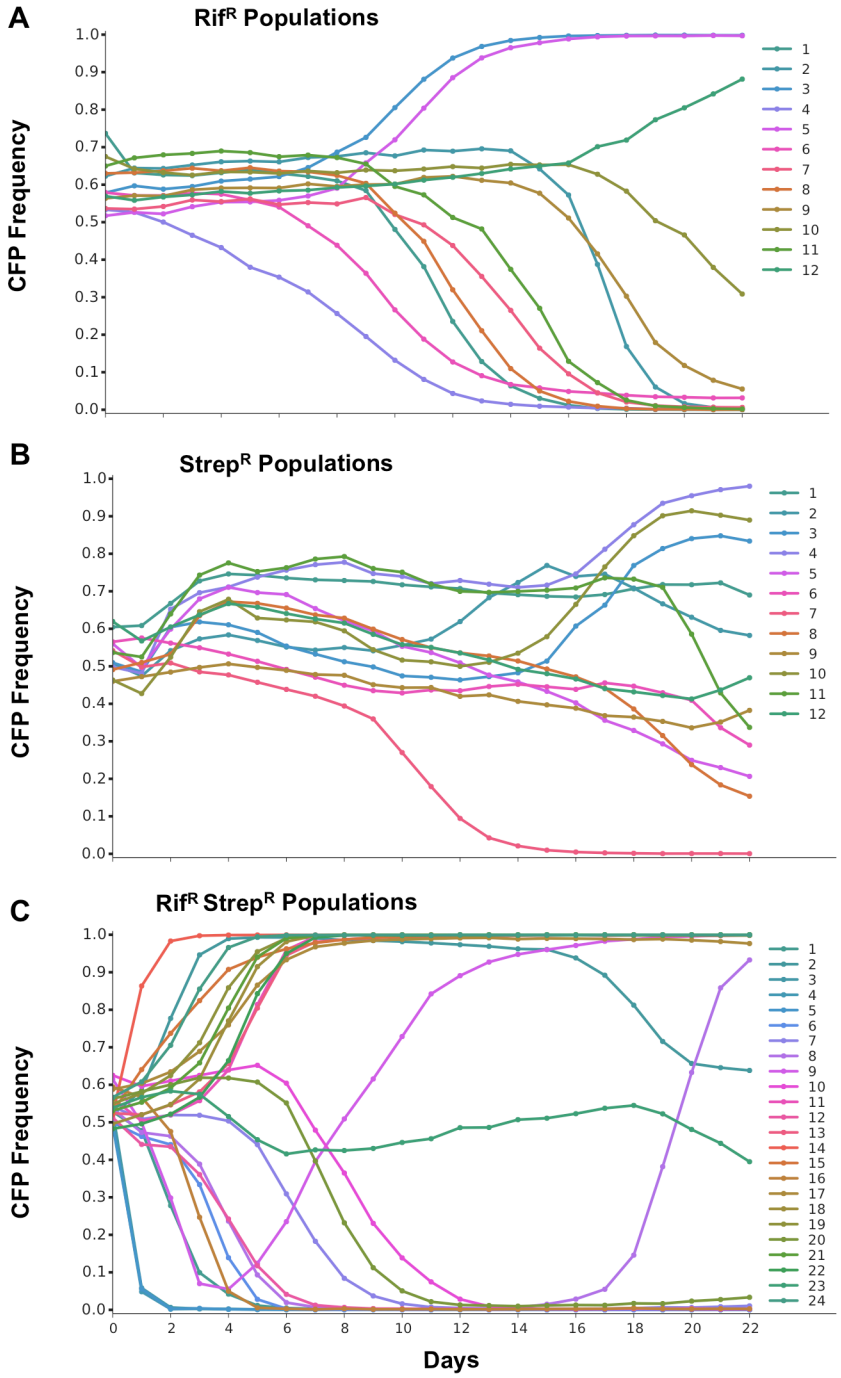


Figure 1: Faster compensatory evolution in double antibiotic resistant bacteria. Dynamics of a fluorescent neutral marker during adaptation to rich media without antibiotics in (A) 12 independent *Escherichia coli* populations resistant to rifampicin (RpoB^{H526Y}); in (B) 12 populations resistant to streptomycin (RpsL^{K43T}) and in (C) 24 populations resistant to both rifampicin and streptomycin (RpoB^{H526Y}RpsL^{K43T}).

Faster adaptation of $Rif^{R}Str^{R}$ is driven by beneficial mutations with stronger effect

To understand why double resistant populations adapt faster than single resistant ones, we measured the average competitive fitness of the evolving populations (see Methods) at days 5, 9, 15 and 22 along the adaptive process (**Fig. 2a**). Despite variation between the populations, the general pattern of fitness increase differs between the resistances. While the Rif^{R} populations on average increase in fitness linearly over time (**Fig. 2a**, leftmost panel), both the Str^{R} populations and the $Rif^{R}Str^{R}$ populations show a substantial increase in fitness at day 5 and subsequently reach a plateau (**Fig. 2a**, middle and rightmost panel, respectively). Moreover, the fitness increase in $Rif^{R}Str^{R}$ bacteria is both remarkably strong and fast, with the high initial cost of the double resistance (red dashed line in rightmost panel of **Fig. 2a**) being on average completely mitigated at day 5. Interestingly, while all populations show increased competitive fitness compared to the ancestral resistances, we see stabilization at different fitness peaks (**Fig. 2a** and **Extended Data Figure 2**), which suggests multiple pathways for compensation.

Rate and effects of compensatory mutations

The higher fitness increase in double resistant bacteria can be due to either a higher rate of acquisition of beneficial mutations (due to a large target size, for instance) or to the acquisition of mutations with stronger beneficial effects. Using a modified version of the TwoMarkerABC algorithm (Moura de Sousa et al. 2013) that considers both the marker dynamics and the fitness along the adaptive walk (see **Methods**), we inferred the rate of acquisition of beneficial mutations and the distribution of effects of arising beneficial mutations for each of the backgrounds. We found that the inferred rate of acquisition of beneficial mutations is not significantly different between the double resistance and the Rif^{R} populations (around $3x10^{-6}$), although both are significantly lower compared to the Str^{R} genetic background, which is in the order of 10^{-5} per cell per generation (**Fig. 2b**).

However, the mean selective effect of beneficial arising mutations is significantly higher for the $Rif^{R}Str^{R}$ background, compared to either of the individual single resistances (around 0.18, versus 0.1 for the Rif^{R} and 0.05 for the Str^{R}) (**Fig. 2c**). This suggests that beneficial mutations with higher mean selective effect are available to the double resistant bacteria. These observations point to the existence of different compensatory paths, which might entail the acquisition of different mutations in each background.

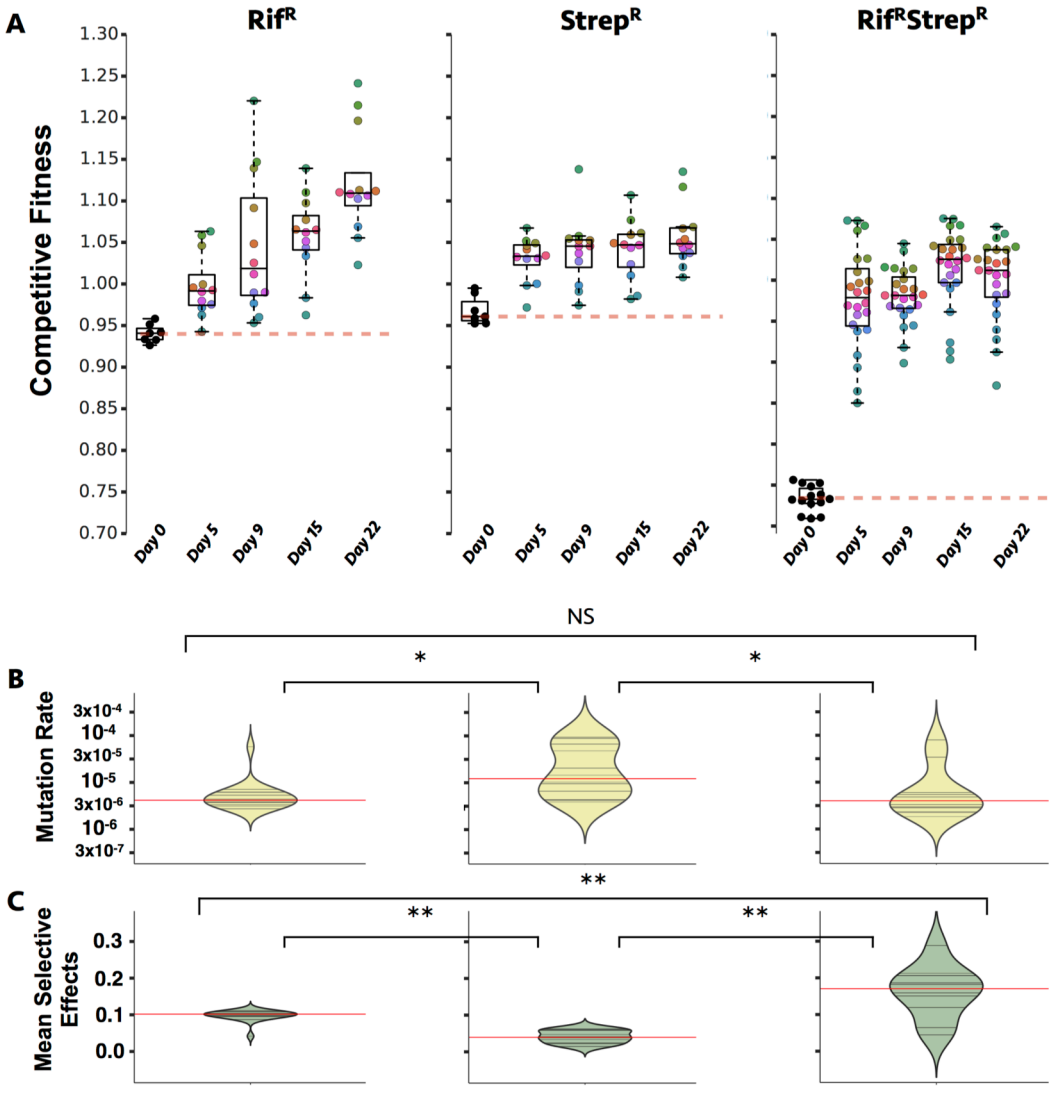


Figure 2: Compensatory mutations have stronger effects in double resistant mutants. A) Competitive fitness of each evolving population at different days during adaptation. Each circle corresponds to a population with similar color shown in figure 1. Red dashed lines correspond to the fitness of the ancestral resistance mutant strain. B) Estimation of the rate of acquisition of beneficial mutations and (C) mean selective effects of arising beneficial mutations are shown in the violin plots, and correspond to the distribution of the top 20 simulations that best explain the experimental data. In panel B Red line indicates the median of the distributions.

The genetic basis of compensation

Selective effects between double and single resistances suggests that the mutations underlying compensation could either be diverse and specific for each background or, alternatively, that similar targets have distinct effects depending on the resistant background. In order to analyse the genetic basis of compensatory adaptation, we performed whole genome sequencing of one clone, from the fluorescence at the highest frequency, from each evolved population at the end of the experiment. For each resistance background, the clones from all populations were pooled and sequenced together (see **Methods**), in order to capture a broad spectrum of the compensatory landscape (**Fig. 3** and **Extended Data Table 1**). We found that mutations are, indeed, diverse and strikingly more so for the double resistance background (37 different gene targets) (**Fig. 3c**) compared to either of the single resistances (9 and 16 genes targeted in Rif^R and Str^R evolved populations, respectively) (**Fig. 3a** and **b**). Several mutations in Rif^R (10 out of 16 allelic changes) and Rif^RStr^R (16 out of 52) were detected in known compensatory targets for the single resistance mutations in either *rpoB* (*rpoB* itself, *rpoA* and *rpoC*) or *rpsL* (*rpsE*, *rpsD* and *tufA*) (see expanded arrows in **Fig. 3**). Other novel putative compensatory targets were identified, which point to an overlap in functional targets of compensation, namely in genes encoding membrane proteins (i.e., *ybjO* and *nanC* in Rif^R; *ydhK* in Rif^R and Str^R*; yeaI* and *yojI* in Str^R*; ompF* and *ydiY* in Rif^RStr^R) and ribosomal proteins (*rplL* in Rif^RStr^R and *rplI* in Str^R). On the other hand, mutations affecting other genes involved in DNA replication, transcription and translation (i.e., *dnaG*, *multL*, *rpiR*, *secE (nusG)* and *rpsJ (nusE)*) seem to occur preferentially in the double resistant evolved populations. At the allelic level, we found that a genetic reversion in *rpoB* and three Str^R mutations overlap between the double resistance and either of the two single resistances. This minimal parallelism indicates that either numerous targets of compensation exist or that there is an abundance of targets specific to the double resistant background.

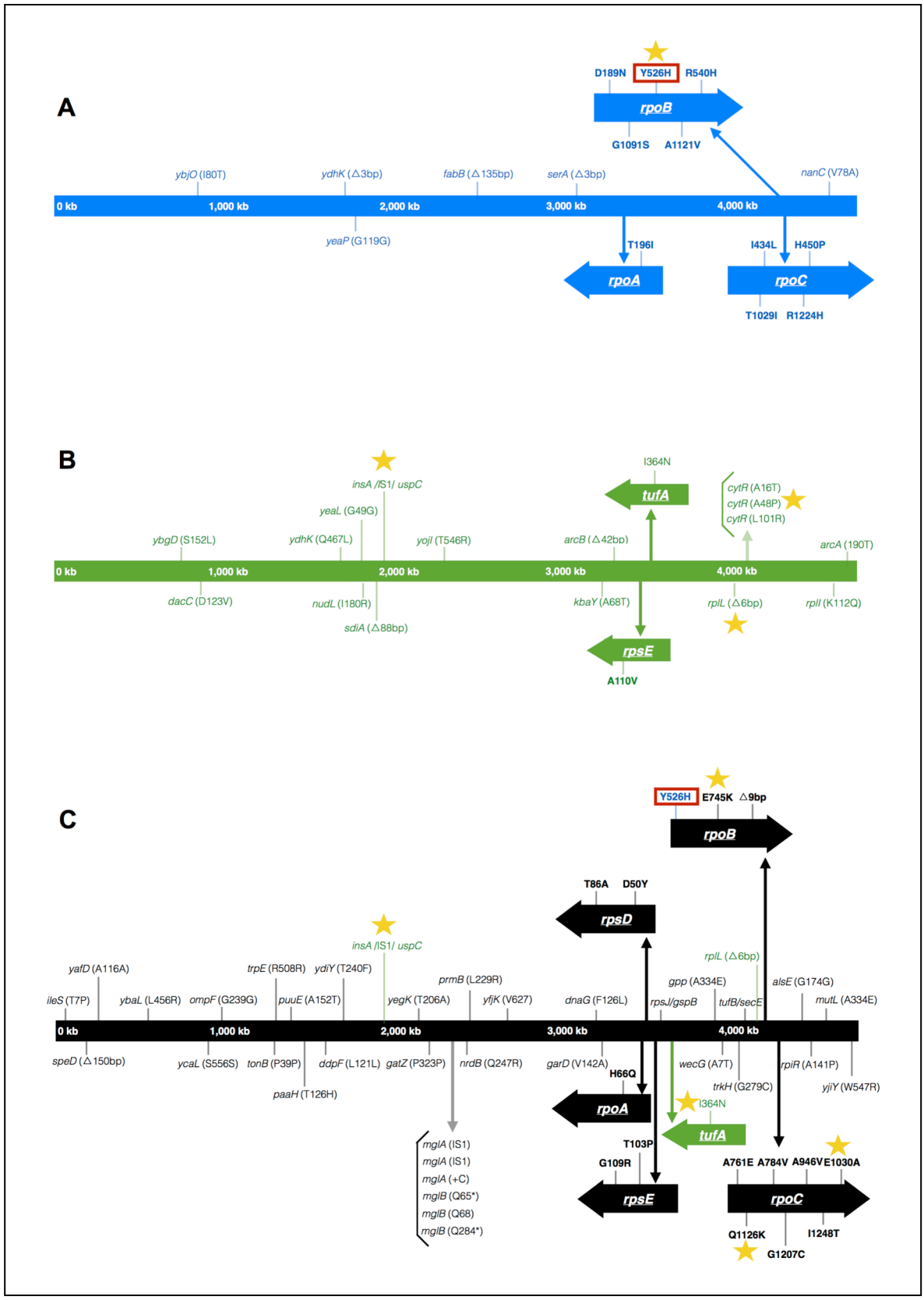

A
D189N
Y526H
R540H
rpoB
G1091S
A1121V
ybjO (I80T)
ydhK (Δ3bp)
fabB (Δ135bp)
serA (Δ3bp)
nanC (V78A)
0 kb
1,000 kb
2,000 kb
3,000 kb
4,000 kb
yeaP (G119G)
T196I
I434L
H450P
rpoA
rpoC
T1029I
R1224H
B
I364N
tufA
insA /IS1/ uspC
yeaL (G49G)
cytR (A16T)
cytR (A48P)
cytR (L101R)
ybgD (S152L)
ydhK (Q467L)
yojI (T546R)
arcB (Δ42bp)
arcA (190T)
dacC (D123V)
nudL (I180R)
sdiA (Δ88bp)
kbaY (A68T)
rplL (Δ6bp)
rplI (K112Q)
rpsE
A110V
C
Y526H
E745K
Δ9bp
rpoB
T86A
D50Y
rpsD
insA /IS1/ uspC
rplL (Δ6bp)
yafD (A116A)
trpE (R508R)
ydiY (T240F)
prmB (L229R)
gpp (A334E)
alsE (G174G)
ileS (T7P)
ybaL (L456R)
ompF (G239G)
puuE (A152T)
yegK (T206A)
yfjK (V627)
dnaG (F126L)
rpsJ/gspB
tufB/secE
mutL (A334E)
0 kb
1,000 kb
2,000 kb
3,000 kb
4,000 kb
speD (Δ150bp)
ycaL (S556S)
tonB (P39P)
ddpF (L121L)
gatZ (P323P)
nrdB (Q247R)
garD (V142A)
wecG (A7T)
rpiR (A141P)
H66Q
trkH (G279C)
yjiY (W547R)
paaH (T126H)
I364N
rpoA
tufA
mglA (IS1)
mglA (IS1)
mglA (+C)
mglB (Q65*)
mglB (Q68)
mglB (Q284*)
T103P
G109R
A761E
A784V
A946V
E1030A
rpsE
rpoC
Q1126K
I1248T
G1207C

Figure 3: Distinct allelic targets of compensation in double resistant bacteria. Genetic targets associated with compensatory adaptation for each antibiotic resistant background were identified through metapopulation sequencing. Mutations detected at least in one population map (approximately) at the position indicated (see **Extended Table 1**). Zoomed genes indicate known targets of compensatory mutations for the corresponding resistant background. A) In blue, mutations found in the 12 independently evolved Rif^R populations. B) In green, mutations found in the 12 independently evolved $Strep^R$ populations. C) In black, mutations found in the 24 independently evolved Rif^RStrep^R, with mutations found also in single resistant populations indicated in the correspondent color. In A) and C) a red box indicates a reversion of the original resistance mutation ($RpoB^{Y526H}$). In all panels, stars indicate mutations that were likely in more than one population (i.e., frequency of reads was above the one expected for one clone in the population pool).

Genetic reversion of Rif^R mutation

The only allelic mutation that was found to be shared between the Rif^R and the Rif^RStr^R evolved populations is a back mutation in *rpoB* (red squares in **Fig. 3a** and **c**). Reversions are rare in single resistance backgrounds (Levin et al. 2000; Poon 2005), but clinically relevant, since bacteria regain sensitivity to the antibiotic. In either the single or double resistant background, reversion here was also rare, being detected in 2 out of 12 Rif^R populations and 1 out 24 Rif^RStr^R populations (see **Extended Data Table 1**), all of which were confirmed to be sensitive to rifampicin (**Extended Data Table 2**). The difference in the number of reversions is non-significant between the single and double resistant mutations (P=0.243, Fisher exact test). Given the epistatic interaction in the double resistance, this mutation should have a very strong beneficial effect (~23%), accounting for a substantial portion of the fitness increase (see **Fig. 2a**). However, its low frequency across double resistant evolved populations is indicative that

there are many other strong effect beneficial mutations that compensate the fitness cost of a double resistant background without the loss of resistance. Notably, a reversion of the RpsL^{K43T} allele, which would also confer a 20% fitness benefit, was not detected through sequencing, although there are one Str^{R} and two Rif^{R}Str^{R} evolved lines that have subpopulations of clones with increased sensitivity to Streptomycin (**Extended Data Table 2**). The fact that, contrary to expectations, reversions are seldom seen in compensated double resistant populations is in agreement with our inference of a distribution with high mean selective effects for this background.

RpoC^{Q1126K} shows sign epistasis between Rif^{R} and Rif^{R}Str^{R}

The limited parallelism for compensatory mutations between the Rif^{R}Str^{R} background and the corresponding single resistances can be expected under two, non mutually exclusive, hypothesis: i) the number of targets for compensation can be so large that our assay was insufficient to capture its full spectrum; and ii) many mutations are only beneficial when both resistances are present. Because *rpoC* is a common compensatory target for rifampicin resistance mutations in the *rpoB* gene (De Vos et al. 2013; Brandis et al. 2012; Comas et al. 2012), we selected two mutations appearing in this gene, RpoC^{H450P} and RpoC^{Q1126K}. The first one was detected in the evolved single Rif^{R} populations and the second was observed in the evolved double Rif^{R}Str^{R} background. We reconstructed new strains (see **Methods**) harbouring one of the original resistance allele(s) (RpoB^{H526Y} or RpoB^{H526Y}RpsL^{K43T}) and either one of these putative compensatory mutations. We then competed these reconstructed strains against their corresponding ancestral (i.e., only the resistance allele(s)) in order to assert its fitness advantage, if any (**Fig. 4**). As expected, the reconstructed double mutant RpoB^{H526Y}RpoC^{H450P} has a competitive fitness advantage (0.05±0.01) over the ancestral strain (RpoB^{H526Y}) but the beneficial effect of this mutation is higher (0.10±0.013) in the double resistance background (RpoB^{H526Y}RpsL^{K43T}RpoC^{H450P}) (**Fig. 4a**). Given the strong advantage

conferred by this allele, the fact that it was not observed in any of the 24 evolved $Rif^{R}Str^{R}$ populations is indicative of the existence of many strong compensatory mutations competing, supporting the first hypothesis (the compensatory targets were not exhausted) and in concordance with our inference (**Fig. 3c**). To test for the second hypothesis, similar competitive assays were performed with the second mutation in *rpoC* ($RpoC^{Q1126K}$), which arose in the $Rif^{R}Str^{R}$ populations. As expected, the triple mutant $RpoB^{H526Y}RpsL^{K43T}RpoC^{Q1126K}$ shows a high fitness advantage over its ancestral ($RpoB^{H526Y}RpsL^{K43T}$) but, strikingly, this mutation shows decreased fitness in the single resistant strain ($RpoB^{H526Y}RpoC^{Q1126K}$) with respect to its ancestral ($RpoB^{H526Y}$) (**Fig. 4b**). Therefore, the beneficial effect of the $RpoC^{Q1126K}$ mutation is conditional to the presence of the second resistance allele $RpsL^{K43T}$, being deleterious in the context of the $RpoB^{H526Y}$ mutation alone. This provides evidence for the second hypothesis, suggesting that this mutation might be specifically compensating for the epistatic interaction between the two resistant alleles.

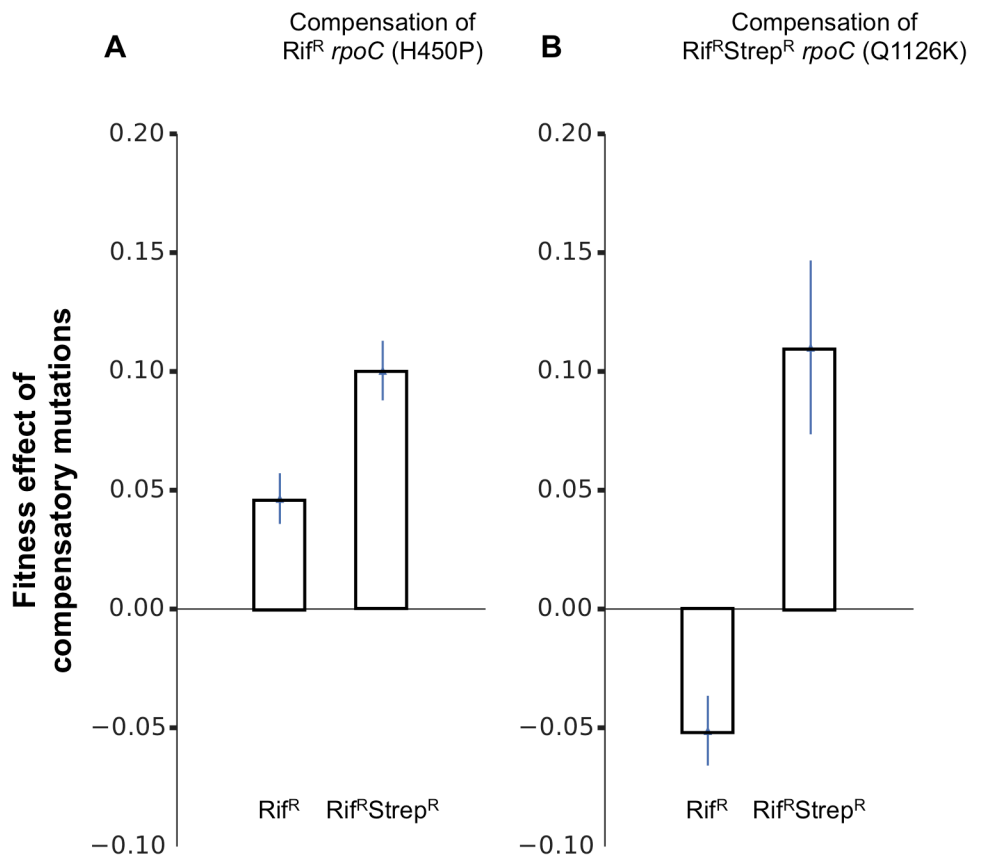


Figure 4: A compensatory mutation buffers the epistasis between Rif^R and Strep^R. Strains reconstructed to incorporate compensatory mutations are competed against their corresponding ancestral (i.e., the resistant strain without the compensatory mutation). A) A compensatory mutation of Rif^R populations, mapping in the rpoC gene (RpoC^{H450P}), is beneficial also in the Rif^{R}Strep^R background, with an even stronger compensatory effect in the latter, demonstrating a remarkably broad range for compensatory targets. B) The allele RpoC^{Q1126K}, found in adapted Rif^{R}Strep^R populations, is beneficial in that background but deleterious in the single Rif^R mutant, indicating that it compensates specifically for the epistasis between Rif^R and Strep^R alleles. Mean values and standard error bars correspond to 3 replicate competitions for each of the 4 independently constructed clones for each allelic replacement.

Discussion

As concerns increase regarding the generation and spread of multi-drug resistant bacteria, it is of crucial importance to understand how these resistant alleles are maintained in populations (Hughes & Andersson 2015). Investigating the process of acquisition of compensatory mutations is essential and, although compensation for single resistance mutations has been well described, our findings suggest that these observations are not entirely transferrable into the compensatory process of multi-drug resistant strains. The results presented here demonstrate that $Rif^{R}Str^{R}$ bacteria, when compensating for their fitness cost, can follow an evolutionary path that is distinct from either one of its corresponding single resistances. Not only the pace of adaptation is faster, but the double resistant background also has access to stronger effect mutations. The results from our theoretical inference provide support for this observation, with estimates of the rate of acquisition of mutations that range from 10^{-6} in RifR and $Rif^{R}Str^{R}$ populations to 10^{-5} in Str^{R} (per cell, per generation). These are slightly above the ones estimated in compensation of the K43N allele in *Salmonella* Thypimurium, which is in the order of 10^{-7} (Maisnier-Patin et al. 2002). Since this rate was estimated from the rate of fixation of mutants, in scenarios of intense clonal interference this value could be an underestimation of the true compensatory rate. It is important to notice here that the underlying model of adaptation used for the estimation of these parameters in our work does not take into account epistasis (Moura de Sousa et al. 2013), which is very likely affecting mutations acquired during adaptation. Therefore, both the rate of acquisition of beneficial mutations and the average mean effects of beneficial mutations are probably underestimated, particularly for the Str^{R} and $Rif^{R}Str^{R}$, where fitness increases shows more prevalent diminishing returns. However, our inference of higher selective effects in the double resistant populations is in agreement with their measured competitive fitness, which also show a higher increase. In principle, this could be due to the fact that the $Rif^{R}Str^{R}$ is inherently more costly, and thus similar mutations could have a stronger effect in this less adapted background. However,

the genetics of evolved populations indicate a different story. The minimal overlap between the putative compensatory mutations identified in the evolved populations is indeed suggestive of alternative evolutionary paths that are shaped by the epistatic interaction between the two resistant alleles. We had *a priori* indications that the we would not be able to completely exhaust the targets of compensation, first because the known targets of compensation for streptomycin, for instance, can be very diverse, of at least 35 allelic changes in the genes coding for ribosomal proteins alone (Maisnier-Patin et al. 2002). Moreover, we sequenced only one clone per population, and our sequencing protocol does not have a resolution that would enable the detection of every single compensatory mutation. Nevertheless, and even with this limited sampling procedure, we found that 3 mutations co-occur between Str^R and double resistant evolved populations, suggesting that some of compensatory mutations can be strongly selected in both the single and double resistant backgrounds. The fixation of a reversion in the $RpoB^{H526Y}$ allele in two out of 12 Rif^R evolved populations indicates that the rate of reversion for this particular allele might be higher than previously thought. However, the fact that the reversion was observed at a comparable frequency in both the single Rif^R and double resistant populations was surprising, since this mutation should have a very strong beneficial effect on the latter background. We reasoned this might indicate competition with several other compensatory mutations that have high selective effects. Moreover, the fact that we do not detect genetic reversions for the Str^R evolved populations is indicative that the compensatory spectrums differ between the resistance mechanisms, possibly both in size and effects. Of notice, however, is the fact that some evolved Str^R (and some Rif^RStr^R) populations show increased sensitivity to streptomycin. No reversion of the original $RpsL^{K43T}$ was found in our sequences, meaning that an eventual reversion, if it emerged, has not swept to fixation as in the Rif^R populations. It is also possible that this is a phenotypic reversion caused by mutations in other genes, which we are currently investigating. For instance, some of the alleles where we detected mutations in *rpsD* and *rpsE* are known to confer a ribosomal ambiguity phenotype (Maisnier-Patin & Andersson 2004; Ogle &

Ramakrishnan 2005), which might functionally revert the effect of the original Str^R $RpsL^{K43T}$ mutation, rendering the cells sensitive to the antibiotic.

Several mutations observed in the ribosomal genes in the evolved Str^R ($RpsE^{A110V}$) in Rif^RStr^R ($RpsD^{T86A}$, $RpsD^{D50Y}$, $RpsE^{T103P}$ and $RpsE^{G109R}$) backgrounds affected residues previously described as compensatory for the $RpsL^{K43N}$ resistance allele in *S.* Typhimurium, at least at the level of the residue (Maisnier-Patin et al. 2002). Some of the SNPs in *rpoA* ($RpoA^{T196I}$) and in *rpoC* ($RpoC^{H450P}$) where also previously identified as compensatory for the $RpoB^{H526Y}$ allele (Moura de Sousa et al. 2015). Nevertheless, many of the remaining putative compensatory targets identified here are, to the best of our knowledge, previously unidentified in the literature. The minimal allelic overlap between the mutations acquired in the three different backgrounds suggests that these mutations confer specific compensatory effects instead of general adaptation to the selected growth conditions. Additionally, the fact that some of the targets are directly related to the cellular mechanisms of replication, transcription and translation, provides further indication of their compensatory role. The double resistant background in particular accumulates a staggering amount and diversity of mutations, which span multiple cellular characteristics and mechanisms. Importantly, functional mechanisms specifically targeted by compensatory mutations in the Rif^RStr^R background include DNA transcription and translation, and might modulate the interaction between the resistances themselves. A particular mutation, $RpoC^{Q1126K}$, even though it maps in a gene known to harbour compensatory mutations for Rif^R, becomes deleterious when associated with the original $RpoB^{H526Y}$ alone, while being beneficial in the double resistance background ($RpoB^{H526Y}RpsL^{K43T}$). Thus, this mutation could specifically compensate for the epistatic interaction between the resistance alleles. If this is true, it would be extremely interesting to disclose the molecular mechanisms underlying this compensatory effect. Moreover, other identified mutations provide an interesting indication of compensation for the interaction between resistances. For instance, mutations affecting the regulatory regions of *rpsJ (nusE)* or the *secE-nusG* operon target proteins that constitute the

physical link between transcription and translation machineries (Burmann et al. 2010), suggesting that they could compensate for the epistasis by affecting the coordination between these two processes. Furthermore, to the best of our knowledge, these targets have not been identified in studies of compensation in single resistance alleles. We expect that our current analyses of the specific effects of these mutations will provide new insights on the molecular mechanisms involved. Since many resistance mutations affect essential mechanisms that commonly interact (Borrell et al. 2013; Durão et al. 2015), there are plenty of opportunities for a spectrum of mutations that supress or correct these epistatic interactions, which will be missed by studying the individual resistance alleles alone (MacLean & Vogwill 2015). The ability to identify them and determine how they affect the interactions between the resistant alleles is determinant for understanding how multi-resistances emerge and persist in bacterial populations. This may provide a theoretical framework for the development of new antimicrobial strategies that exploit potential weaknesses of multi-drug resistant bacteria.

Material and Methods

Strains and growth conditions

E. coli K12 MG1655 marked with constitutive expression of yellow (YFP) or cyan (CFP) fluorescent proteins (ΔlacIZYA galK::CmR YFP/CFP) and with chromosomal resistance to either streptomycin ($RpsL^{K43T}$ allele), rifampicin ($RpoB^{H526Y}$ allele) or both, was used in the long term propagation experiment in to Luria-Bertani broth (LB) without the presence of antibiotics. Cultures were grown in a 96-well plate incubator at 37º C with shaking (700rpm). Non-fluorescent wild-type *E. coli* K12 MG1655, was used as reference strain for the competition fitness assays.

Experimental evolution for compensation

In order to acclimatize bacteria to the environment, strains were grown separately from frozen stocks in LB media (150ul per well) in 96-well plates at 37º C with shaking (12 replicates per strain were inoculated in a checkered format to avoid cross-contaminations). After 24h, 10ul of bacteria culture diluted by a factor of 10-2 was transferred into 140ul fresh LB media and let grow for additional 24h. Isogenic strains differing only in the marker were diluted again by a factor of 10-2 and then mixed based on their cell numbers given by the Flow Cytometer (LSR Fortessa) in order to obtain an initial ratio of 1:1. A total of 48 competitions were initiated by inoculating 140ul of LB media with 10ul of each mixed population which were allowed to grow for 24h, reaching a concentration of, approximately, 109 CFU/ml. After every 24h of growth, and for 22 days, these cultures were propagated by serial passage with a constant dilution factor of 10-2 (10 ul of diluted culture was transferred into 140 ul of fresh medium). In parallel, cell numbers were counted using the Flow Cytometer in order to measure the frequency of each strain in the mixed population during the experiment, by collecting a sample (10ul) from the spent culture each day. Samples were frozen at days 5, 9, 12, 15, 18 and 22.

Competitive fitness assays

The relative fitness of each evolved population at the end of the propagation experiment, at day 22, was measured by competitive growth against the reference strain *E. coli* K12 MG1655. The competitor (each evolved populations, potentially composed of both YFP and CFP) and the reference (unmarked) strains were first unfrozen and acclimatized separately for 48h (with two growth periods of 24h) and then mixed in a proportion of 1:1 using a method similar to the one preveiously described. To assess the cost of the resistances themselves before any compensation, control competitions were performed between the ancestral of each resistant mutant and the reference strain by mixing

25% of YFP + 25% of CFP with 50% of the unmarked strain. Thus, a total of 52 competitions were initiated by inoculating 140ul of LB media with 10ul of each mixed population and allowed to compete for 48h. The initial and final frequency of the strains was obtained by counting their cell numbers in the Flow Cytometer. Generation time was estimated from the doubling time of the reference strain (approx. 8 generations) and the fitness was determined as the average of three independent replicates for each competition.

Sensitivity assay of evolved populations

Evolved cultures were grown in 96 well plates with LB for 24h and subsequently plated in antibiotic-free solid media. From each evolved population clones were picked with a pipette tip, which then was used to inoculate sequentially both a spot in antibiotic (plates containing media with 100ug/ml of rifampicin, streptomycin or both drugs) and antibiotic-free plates (LB agar only). Clones were classified as sensitive if they grew on an antibiotic-free media but not on media with antibiotic(s). 50 clones were initially tested for each population. Populations where sensitive clones were found were subsequently tested for another 150 clones.

Estimation of evolutionary parameters

The method used to estimate the rate of acquisition of beneficial mutations and the distribution of selective effects was modified from the one published in (Moura de Sousa et al. 2013). Briefly, the method simulates 1 million marker dynamics with parameters (mutation rate, shape and mean of a gamma distribution for the selective effects of new mutations) chosen from a uniform distribution. Afterwards, marker frequency and average population fitness is used to summarize both the experimental data and the simulated dynamics. This data is used as summary statistics to be compared by Approximate Bayesian

Computation (ABC). The difference in the method used here is that instead of using the neural network function in ABC, a simple ranking and rejection method was used, where simulations are ranked by the Euclidean distance between summary statistics form the simulated and experimental data, and the top 20 simulations were chosen for the distributions of parameters.

DNA extractions

The evolved populations were plated onto LB agar plates and grown for 24h at 37º C. One colony, with the most frequent marker, was picked from each population was grown overnight into 10 mL of liquid LB, at 37º C with shaking. The bacterial DNA was then extracted and its concentration and purity were quantified using Qubit and NanoDrop, respectively. Following quantification, equal concentration of each DNA sample was taken and pooled together with the clones from the populations of similar resistant background, resulting in 3 pools of DNA (Rif^R pool with 12 clones, Str^R pool with 12 clones and Rif^RStr^R pool with 24 clones).

Whole Genome Sequencing Analysis

The DNA extracted was sequenced using the Miseq Illumina platform. Coverage of the different pools was as following: 521x for 12 clones in the Rif^R pool; 640x for 12 clones in the Str^R pool; 631x for 24 clones in the Rif^RStr^R pool. The resulting sequences were analysed in Breseq version 2.3, using *E. coli* K12 genome NC_000913.2 as a reference, and with the polymorphism option selected and following parameters: a) rejection of polymorphisms in homopolymers of length greater than 3; b) rejection of polymorphisms that are not present in at least 3 reads in each strand; c) rejection of polymorphisms that don not have a p-value for quality bigger than 0.05. All other Breseq parameters were used as default.

Allelic reconstructions and competitions

The putative compensatory mutations located in *rpoC* were co-transduced together with *rpoB* H526Y to wild-type (Δ*lacIZYA galK::CmR YFP/CFP*) and *rpsL* K43T backgrounds by P1vir transduction, using the propagated line harbouring each *rpoC* mutation as donor and either wild-type (Δ*lacIZYA galK::CmR YFP/CFP*) or the *rpsL* K43T ancestral as recipient strains, and selecting for rifampicin resistance. The presence of the desired mutations in the transductant isolates was assessed by amplification of the *rpoC* gene and sequencing. The resulting strains, which were stored at -80ºC, harbour either the *rpoC* mutation linked to the *rpoB* H526Y allele (reconstruction of the *rpoC* mutation in the single resistant background) or these two mutations and the *rpsL* K43T allele (reconstruction of the *rpoC* mutation in the double resistant background). Both reconstructed strains and ancestral resistant strains were unfrozen and acclimatized during 24h, to avoid compensatory mutations to appear in the more costly backgrounds. Reconstructed mutants were competed against the respective ancestral strain to assess their competitive advantage per generation, assuming the competition lasts 8 generations, or 24h.

References

Andersson DI, Hughes D. 2010. Antibiotic resistance and its cost: is it possible to reverse resistance? Nature Reviews Microbiology. 8:260–271. doi: 10.1038/nrmicro2319.

Bhullar K et al. 2012. Antibiotic resistance is prevalent in an isolated cave microbiome. PLoS ONE. 7:e34953. doi: 10.1371/journal.pone.0034953.

Borrell S et al. 2013. Epistasis between antibiotic resistance mutations drives the evolution of extensively drug-resistant tuberculosis. Evolution, Medicine, and Public Health. 2013:65–74. doi: 10.1093/emph/eot003.

Borrell S, Gagneux S. 2011. Strain diversity, epistasis and the evolution of drug resistance in Mycobacterium tuberculosis. Clin. Microbiol. Infect. 17:815–820. doi: 10.1111/j.1469-0691.2011.03556.x.

Brandis G, Hughes D. 2013. Genetic characterization of compensatory evolution in strains carrying rpoB Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates. J. Antimicrob. Chemother. 68:2493–2497. doi: 10.1093/jac/dkt224.

Brandis G, Wrande M, Liljas L, Hughes D. 2012. Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. Molecular Microbiology. 85:142–151. doi: 10.1111/j.1365-2958.2012.08099.x.

Burmann BM et al. 2010. A NusE:NusG Complex Links Transcription and Translation. Science. 328:501–504. doi: 10.1126/science.1184953.

Chakrabarti SL, Gorini L. 1977. Interaction between mutations of ribosomes and RNA polymerase: a pair of strA and rif mutants individually temperature-insensitive but temperature-sensitive in combination.

Chao L. 1978. An unusual interaction between the target of nalidixic acid and novobiocin. Nature.

Comas I et al. 2012. Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nature Genetics. 44:106–110. doi: 10.1038/ng.1038.

Couturier M, Desmet L, Thomas R. 1964. High pleiotropy of streptomycin mutations in Escherichia coli. Biochemical and Biophysical Research Communications. 16:244–248.

Davies J, Davies D. 2010. Origins and Evolution of Antibiotic Resistance. Microbiology and Molecular Biology Reviews. 74:417–433. doi: 10.1128/MMBR.00016-10.

De Vos M et al. 2013. Putative compensatory mutations in the rpoC gene of rifampin-resistant Mycobacterium tuberculosis are associated with ongoing transmission. Antimicrobial Agents and Chemotherapy. 57:827–832. doi: 10.1128/AAC.01541-12.

Durão P, Trindade S, Sousa A, Gordo I. 2015. Multiple Resistance at No Cost: Rifampicin and Streptomycin a Dangerous Liaison in the Spread of Antibiotic

Resistance. Molecular Biology and Evolution. 32:2675–2680. doi: 10.1093/molbev/msv143.

Forsberg KJ et al. 2012. The shared antibiotic resistome of soil bacteria and human pathogens. Science. 337:1107–1111. doi: 10.1126/science.1220761.

Goldstein BP. 2014. Resistance to rifampicin: a review. 1–6. doi: 10.1038/ja.2014.107.

Gullberg E, Cao S, Berg OG, Ilbäck C, Sandegren L. 2011. Selection of resistant bacteria at very low antibiotic concentrations. PLoS Pathog.

Hall AR, MacLean RC. 2011. Epistasis buffers the fitness effects of rifampicin-resistance mutations in Pseudomonas aeruginosa. Evolution. 65:2370–2379. doi: 10.1111/j.1558-5646.2011.01302.x.

Hede K. 2014. Antibiotic resistance: An infectious arms race. Nature. 509:S2–3. doi: 10.1038/509S2a.

Hegreness M. 2006. An Equivalence Principle for the Incorporation of Favorable Mutations in Asexual Populations. Science. 311:1615–1617. doi: 10.1126/science.1122469.

Hughes D, Andersson DI. 2015. Evolutionary consequences of drug resistance: shared principles across diverse targets and organisms. Nature Publishing Group. 16:459–471. doi: 10.1038/nrg3922.

Kriegeskorte A, Peters G. 2012. Horizontal gene transfer boosts MRSA spreading. Nat Med.

Levin BR, Perrot V, Walker N. 2000. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. Genetics. 154:985–997.

Luzzatto L, Apirion D. 1968. Mechanism of action of streptomycin in E. coli: interruption of the ribosome cycle at the initiation of protein synthesis.

MacLean RC, Vogwill T. 2015. Limits to compensatory adaptation and the persistence of antibiotic resistance in pathogenic bacteria. Evolution, Medicine, and Public Health. 2015:4–12. doi: 10.1093/emph/eou032.

Maisnier-Patin S, Andersson DI. 2004. Adaptation to the deleterious effects of

antimicrobial drug resistance mutations by compensatory evolution. Research in Microbiology. 155:360–369. doi: 10.1016/j.resmic.2004.01.019.

Maisnier-Patin S, Berg OG, Liljas L, Andersson DI. 2002. Compensatory adaptation to the deleterious effect of antibiotic resistance in Salmonella typhimurium. Molecular Microbiology. 46:355–366.

Moura de Sousa J, Sousa A, Bourgard C, Gordo I. 2015. Potential for adaptation overrides cost of resistance. Future Microbiology. 10:1415–1431. doi: 10.2217/fmb.15.61.

Moura de Sousa JA, Campos PRA, Gordo I. 2013. An ABC Method for Estimating the Rate and Distribution of Effects of Beneficial Mutations. Genome Biology and Evolution. 5:794–806. doi: 10.1093/gbe/evt045.

Mwangi MM et al. 2007. Tracking the in vivo evolution of multidrug resistance in Staphylococcus aureus by whole-genome sequencing. Proceedings of the National Academy of Sciences. 104:9451–9456. doi: 10.1073/pnas.0609839104.

Ogle JM, Ramakrishnan V. 2005. Structural insights into translational fidelity. Annu. Rev. Biochem. 74:129–177. doi: 10.1146/annurev.biochem.74.061903.155440.

Petty NK et al. 2014. Global dissemination of a multidrug resistant Escherichia coli clone. Proc. Natl. Acad. Sci. U.S.A. 111:5694–5699. doi: 10.1073/pnas.1322678111.

Poon A. 2005. The Coupon Collector and the Suppressor Mutation: Estimating the Number of Compensatory Mutations by Maximum Likelihood. Genetics. 170:1323–1332. doi: 10.1534/genetics.104.037259.

Qi Q, Toll-Riera M, Heilbron K, Preston GM, MacLean RC. 2016. The genomic basis of adaptation to the fitness cost of rifampicin resistance in Pseudomonas aeruginosa. Proceedings of the Royal Society B: Biological Sciences. 283:20152452. doi: 10.1038/ncomms6208.

Reynolds MG. 2000. Compensatory evolution in rifampin-resistant Escherichia coli. Genetics. 156:1471–1481.

Schrag SJ, Perrot V, Levin BR. 1997. Adaptation to the fitness costs of antibiotic resistance in Escherichia coli. Proceedings of the Royal Society B: Biological Sciences. 264:1287–1291. doi: 10.1098/rspb.1997.0178.

Sousa A, Magalhaes S, Gordo I. 2012. Cost of Antibiotic Resistance and the Geometry of Adaptation. Molecular Biology and Evolution. 29:1417–1428. doi: 10.1093/molbev/msr302.

Trindade S et al. 2009. Positive Epistasis Drives the Acquisition of Multidrug Resistance Zhang, J, editor. PLoS Genet. 5:e1000578. doi: 10.1371/journal.pgen.1000578.s005.

van Hoek AHAM et al. 2011. Acquired antibiotic resistance genes: an overview. Front Microbiol. 2:203. doi: 10.3389/fmicb.2011.00203.

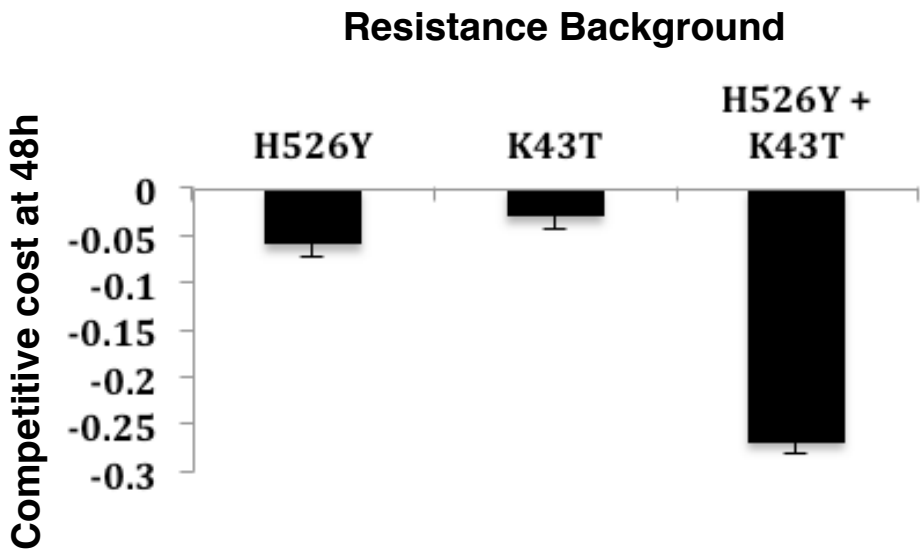
Velayati AA et al. 2009. Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran. Chest. 136:420–425. doi: 10.1378/chest.08-2427.

Warner DF, Koch A, Mizrahi V. 2014. Diversity and disease pathogenesis in Mycobacterium tuberculosis. Trends in Microbiology. 1–8. doi: 10.1016/j.tim.2014.10.005.

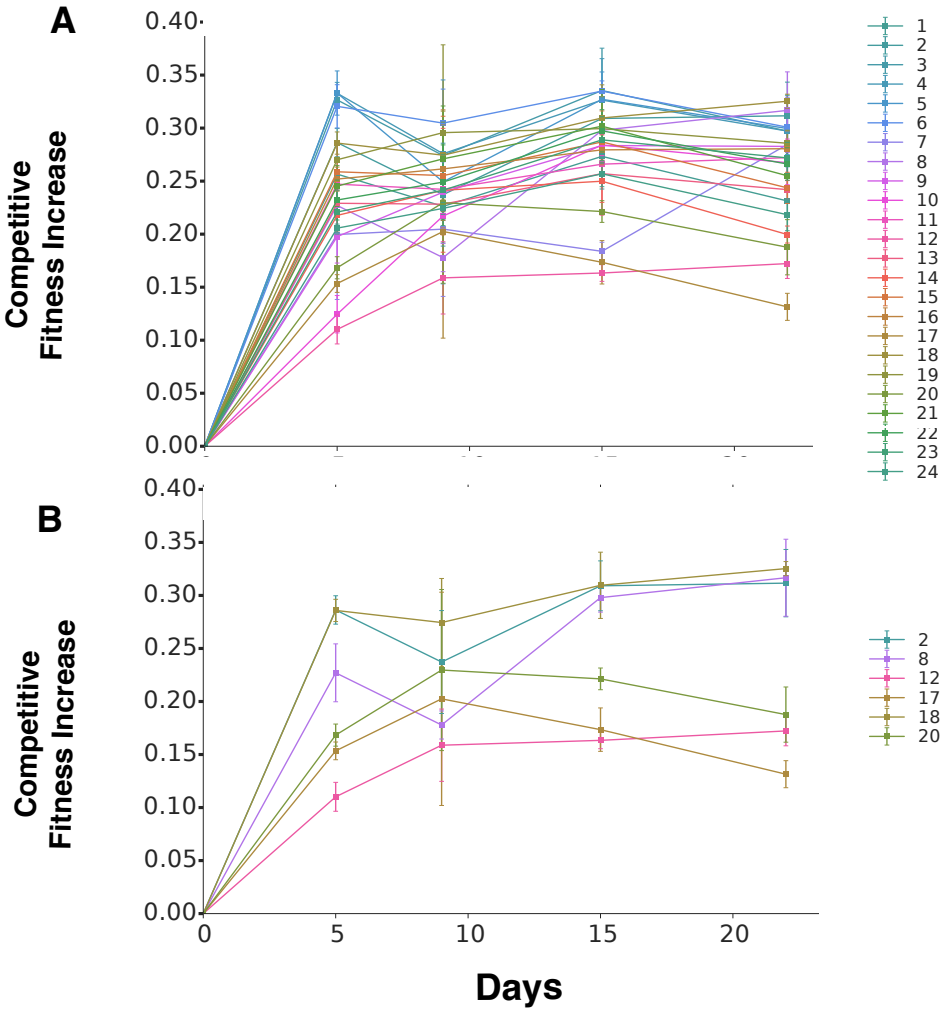
Wehrli W. 1983. Rifampin: mechanisms of action and resistance. Review of Infectious Diseases.

Zhang H et al. 2013. Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nature Genetics. 45:1255–1260. doi: 10.1038/ng.2735.

Supplementary Figures and Tables



Extended Data Figure 1: Competitive fitness at 48h of the resistant strains used in this study. The $Rif^{R}Strep^{R}$ background shows strong negative synergistic epistasis.



Extended Data Figure 2: Dynamics of competitive fitness increase for the $Rif^{R}Strep^{R}$ populations show stabilization at different levels of fitness. Error bars indicate the standard error across 3 independent replicates. In **A** are the dynamics of all the 24 populations, and in **B** are shown 6 indicative populations, for clarity.

Mutations in H526Y Evolved Populations (12)

Genome Position	*Gene(s)*	*Mutation*	*Region*	*Frequency*
897,450	ybjO	T→C	I80T (ATC→ACC)	6.10%
1,720,637	ydhK	Δ3 bp	coding	1%
1,876,095	yeaP	A→G	G119G (GGA→GGG)	8.10%
2,438,935	fabB	Δ135 bp	coding	1%
3,056,065	serA	Δ3 bp	coding	7%
3,438,465	**rpoA**	**G→A**	**T196I (ACC→ATC)**	**16.40%**
4,179,832	**rpoB**	**G→A**	**D189N (GAT→AAT)**	**10.50%**
4,180,843		**T→C**	**Y526H (TAC→CAC)**	**19.30%**
4,180,886		**G→A**	**R540H (CGT→CAT)**	**6.00%**
4,182,538		**G→A**	**G1091S (GGT→AGT)**	**6.40%**
4,182,629		**C→T**	**A1121V (GCG→GTG)**	**7.70%**
4,184,672	**rpoC**	**A→C**	**I434L (ATC→CTC)**	**8.40%**
4,184,721		**A→C**	**H450P (CAC→CCC)**	**6.70%**
4,186,458		**C→T**	**T1029I (ACC→ATC)**	**8.00%**
4,187,043		**G→A**	**R1224H (CGT→CAT)**	**6.10%**
4,537,292	nanC	A→G	V78A (GTT→GCT)	9.50%

Mutations in K43T Evolved Populations (12)

Genome Position	*Gene(s)*	*Mutation*	*Region*	*Frequency*
751,564	ybgD	G→A	S152L (TCA→TTA)	6.80%
880,317	dacC	A→T	D123V (GAC→GTC)	0.70%
1,721,544	ydhK	A→T	Q467L (CAA→CTA)	0.90%
1,868,555	yeaI	T→C	G49G (GGT→GGC)	7.70%
1,894,732	nudL	T→G	I180R (ATA→AGA)	6.30%
1,994,442	sdiA	Δ88 bp	coding (328–415/723 nt)	0.80%
1,977,514	insA/uspC	IS*1* (+) +9 bp	intergenic (–275/–255)	49.80%
2,305,001	yojI	G→C	T546R (ACG→AGG)	8.50%
3,281,366	kbaY	G→A	A68T (GCC→ACC)	9.10%
3,350,771	arcB	Δ42 bp	coding (236–277/2337 nt)	6.3%/8.7%
3,442,923	**rpsE**	**G→A**	**A110V (GCA→GTA)**	**7.50%**
3,468,261	**tufA**	**A→T**	**I364N (ATC→AAC)**	**18.50%**
4,122,178	cytR	A→C	L101R (CTG→CGG)	6.70%
4,122,338		C→G	A48P (GCC→CCC)	6.30%
4,122,434		C→T	A16T (GCC→ACC)	8.80%
4,178,694	rplL	Δ6 bp	coding (112–117/366 nt)	17%/14.8%
4,424,464	rplI	A→C	K112Q (AAG→CAG)	8.10%
4,638,061	arcA	A→G	I90T (ATT→ACT)	8.30%

Mutations in H526Y-K43T Evolved Populations (24)

Genome Position	*Gene(s)*	*Mutation*	*Region*	*Frequency*
22,409	ileS	A→C	T7P (ACC→CCC)	8.40%
134,997	speD	Δ150 bp	coding (437–586/795 nt)	3.10%
231,469	yafD	A→G	A116A (GCA→GCG)	3.60%
501,096	ybaL	A→C	L456R (CTG→CGG)	2.30%
965,210	ycaI	C→T	S556S (AGC→AGT)	5.70%
985,489	ompF	A→G	G239G (GGT→GGC)	4.70%
1,319,447	trpE	G→A	R508R (CGC→CGT)	5.10%
1,364,027	puuE	G→A	A152T (GCG→ACG)	4.40%
1,457,453	paaH	A→G	T126A (ACC→GCC)	6.30%
1,555,702	ddpF	A→G	L121L (TTG→CTG)	6.90%
1,803,390	ydiY	T→C	T240A (ACC→GCC)	6.60%
1,977,418	insA/uspC	IS*1* (+) +8 bp	intergenic (–179/–352)	14.10%
2,149,881	yegK	T→C	T206A (ACC→GCC)	7.80%
2,173,375	gatZ	C→T	P323P (CCG→CCA)	6.40%
2,236,388	mglA	IS*1* (+) +9 bp	coding (916–924/1521 nt)	6.10%
2,236,609		IS*1* (+) +9 bp	coding (695–703/1521 nt)	3.00%
2,237,521	mglB	G→A	Q284* (CAG→TAG)	3.00%
2,238,169		G→A	Q68* (CAG→TAG)	5.00%
2,238,178		G→A	Q65* (CAG→TAG)	3.50%
2,346,145	nrdB	A→G	Q247R (CAG→CGG)	4.40%
2,445,777	prmB	A→C	L229R (CTG→CGG)	4.40%
2,759,683	yfjK	A→G	V627A (GTG→GCG)	3.90%
3,209,504	dnaG	T→C	F126L (TTT→CTT)	4.00%
3,273,728	garD	T→C	V142A (GTC→GCC)	4.60%
3,438,854	**rpoA**	**A→C**	**H66Q (CAT→CAG)**	**4.70%**
3,439,442	**rpsD**	**T→C**	**T86A (ACC→GCC)**	**5.80%**
3,439,550		**C→A**	**D50Y (GAC→TAC)**	**3.10%**
3,442,927	**rpsE**	**C→G**	**G109R (GGT→CGT)**	**5.20%**
3,442,945		**T→G**	**T103P (ACC→CCC)**	**2.90%**
3,451,341	rpsJ / gspB	A→G	intergenic (–49/+189)	3.90%
3,468,261	**tufA**	**A→T**	**I364N (ATC→AAC)**	**10.40%**
3,961,252	gpp	G→T	A334E (GCG→GAG)	4.80%
3,977,997	wecG	G→A	A7T (GCA→ACA)	4.00%
4,032,002	trkH	G→T	G279C (GGC→TGC)	4.40%
4,175,194	tufB / secE	T→C	intergenic (+43/–187)	4.40%
4,178,694	**rplL**	**Δ6 bp**	**coding (112–117/366 nt)**	**4.90%**
4,180,843	**rpoB**	**T→C**	**Y526H (TAC→CAC)**	**6.30%**
4,181,500	**rpoB**	**G→A**	**E745K (GAA→AAA)**	**10.90%**
4,182,295		**Δ9 bp**	**coding (3028–3036/4029 nt)**	**3%**
4,185,654	**rpoC**	**C→A**	**A761E (GCG→GAG)**	**3.50%**
4,185,723		**C→T**	**A784V (GCG→GTG)**	**4.00%**
4,186,209		**C→T**	**A946V (GCT→GTT)**	**2.30%**
4,186,461		**A→C**	**E1030A (GAA→GCA)**	**26.10%**
4,186,748		**C→A**	**Q1126K (CAG→AAG)**	**26.70%**
4,186,991		**G→T**	**G1207C (GGT→TGT)**	**2.30%**
4,187,115		**T→C**	**I1248T (ATT→ACT)**	**3.80%**
4,305,980	alsE	A→G	G174G (GGT→GGC)	6.40%
4,310,594	rpiR	C→G	A141P (GCC→CCC)	3.70%
4,397,222	mutL	Δ1 bp	coding (1788/1848 nt)	4.20%
4,587,664	yjiY	A→G	W547R (TGG→CGG)	5.70%

Extended Data Table 1: Mutations acquired during the 22 days adaptation to antibiotic free media. Mutations are shown for the 3 different resistant backgrounds, along with the frequency at which they were detected. Mutations in known compensatory targets are marked in bold.

		Rif	*Strep*	*Rif + Strep*
Single Resistant	**H2**	Full Reversion		
	H11	Full Reversion		
	K5		Revertant Population	

Double Resistant	**HK2**		Revertant Population	Revertant Population
	HK9	Revertant Population		Revertant Population
	HK11	Full Reversion		Full Reversion
	KH4		Revertant Population	Revertant Population

Extended Data Table 2: Sensitivity assays in the evolved populations. Shown are the populations that show sensitivity (see Methods) in all (orange) or just a fraction (yellow) of the clones tested.

CHAPTER V

Potential for adaptation overrides cost of resistance

Manuscript published in ***Future Microbiology***, in October 2015

The author of this thesis designed the theoretical method and performed the simulations, the analysis of the whole genome sequencing of the evolved populations, the genotyping of evolved populations, and the competitive assays. Ana Sousa designed the research and performed the propagations with technical help from Catarina Bourgard. Ana Sousa collaborated in the analysis of the whole genome sequencing.

Potential for Adaptation Overrides Cost of Resistance

Jorge Moura de Sousa*, Ana Sousa, Catarina Bourgard and Isabel Gordo

Instituto Gulbenkian de Ciência

Rua da Quinta Grande, 6

Oeiras, Portugal

*correspondence to: jasousa@igc.gulbenkian.pt

ABSTRACT Aim: To investigate the cost of antibiotic resistance versus the potential for resistant clones to adapt in maintaining polymorphism for resistance. **Materials & methods:** Experimental evolution of *Escherichia coli* carrying different resistance alleles was performed under an environment devoid of antibiotics and evolutionary parameters estimated from their frequencies along time. **Results & conclusion:** Costly resistance mutations were found to coexist with lower cost resistances for hundreds of generations, contrary to the hypothesis that the cost of a resistance dictates its extinction. Estimated evolutionary parameters for the different resistance backgrounds suggest a higher adaptive potential of clones with costly antibiotic resistance mutations, overriding their initial cost of resistance and allowing their maintenance in the absence of drugs.

KEYWORDS: Evolvability; Antibiotic Resistance; Experimental Evolution; Epistasis; Clonal Interference;

INTRODUCTION

Bacterial populations can acquire antibiotic resistance (AR) as a result of transfer and acquisition of new genetic material between individuals of the same or different species but also by chromosomal DNA mutations, which alter existing bacterial proteins. One landmark example of this second process is provided by

Mycobacterium tuberculosis (the etiologic agent of tuberculosis in humans), which albeit incapable of horizontal gene transfer can display “total drug resistance” (Velayati et al. 2009). This kind of resistance is typically acquired by the sequential accumulation of mutations that alter the cellular target for the drug action. During this process extensive clonal competition has also been observed (Solomon H Mariam et al. 2011). Understanding how these AR determinants disseminate and are maintained in bacterial populations is therefore of paramount importance.

Mutations that confer spontaneous AR can occur at relatively high rates (Mwangi et al. 2013). For instance, rifampicin (Rif) resistance occurs spontaneously at frequencies that can be higher than 10^{-8} in wild isolates of *E. coli* (Baquero et al. 2004). In fact, the levels of resistance in pathogenic populations continue to rise at an alarming rate (Walsh 2000; Laxminarayan et al. 2013) having reached the same significance as any other virulence factor (J Davies & D Davies 2010). However, in the absence of the drug, AR mutations typically bear a cost (Andersson & Levin 1999; Lenski 1998; Andersson & Hughes 2010). This cost depends on the specific resistance allele (Deneke H Mariam et al. 2004), on the environment (Björkman et al. 2000; Chait et al. 2007) and on the genomic background where the mutation happens to arise (Trindade et al. 2009; Angst & Hall 2013). Nevertheless, suppressive mutations that mitigate this cost can occur either in the presence or absence of antibiotics. In the absence of drugs, one can expect that sensitive bacteria, will sweep through the population, driving the AR mutant to extinction. However, more often than not, resistant strains have been observed to acquire additional beneficial mutations that reduce the costs of resistance without loss of resistance, thus preventing the elimination of resistance alleles (Maisnier-Patin & Andersson 2004). These compensatory mutations are common in clinical isolates (Comas et al. 2012; De Vos et al. 2013) and hinder the possibility of reverting the resistance mutation, due to their epistatic nature (Davis et al. 2009). Since the probability of a compensatory mutation tends to be much higher than that of reversion for several resistance mutations (Andersson & Hughes 2010; Levin et al. 2000; Poon 2005; Sousa et al. 2012; Gifford & MacLean

2013), it is likely that resistance alleles may take long periods of time to be eliminated from populations.

Many mutations that confer AR occur in essential genes and are likely to result in pleiotropic effects altering several bacterial traits. Important examples include resistance to Rif and streptomycin (Str) (Couturier et al. 1964; Romero et al. 1973). Rif is one of the frontline anti-tuberculosis drugs (Koch et al. 2014). The co-occurrence of resistance to this antibiotic and isoniazid typically classifies *M. tuberculosis* as multi-drug-resistant. These are the two most commonly used and effective drugs for the treatment of tuberculosis. The main genetic target for Rif resistance is *rpoB,* which codes for the β subunit of the RNA polymerase. Besides typically decreasing the rate of transcription and consequently the growth rate, *rpoB* resistance mutations are probably some of the most pleiotropic among AR mutations. Their effects can range from regulation of competence, sporulation and germination in *Bacillus subtilis* (Maughan et al. 2004), temperature and phage sensitivity in *E. coli* (Jin et al. 1988), growth advantage in stationary phase in *E. coli* and *Salmonella enterica* (Wrande et al. 2008), increased antibiotic resistance in *Staphylococcus aureus* (Cui et al. 2010; Watanabe et al. 2011), amongst others (see (Koch et al. 2014) for a review). *rpoA* and *rpoC*, as well as additional mutations in *rpoB,* have been pointed as targets for compensatory mutation relieving the deleterious effects of *rpoB* Rif resistance mutations in *M. tuberculosis*, *S. enterica* and *E. coli* (Comas et al. 2012; Brandis et al. 2012; Brandis & Hughes 2013; Reynolds 2000). Another frequent resistance is Str, occurring through mutations in the *rpsL* gene, which codes for the S12 subunit of the ribosome compromising translational speed and accuracy. Compensatory mutations for Str resistance have been observed in *Salmonella thyphimurium*, and some of the target genes include *rpsL*, *rpsD*, *rpsE* and *rplS* (Schrag et al. 1997; Björkman et al. 2000; Maisnier-Patin et al. 2002), which encode the ribosomal subunits S12, S4, S5 and L19. Recent analysis of 161 genomes of *M. tuberculosis* with a broad range of resistance profiles, revealed seven possible additional targets for compensatory mutations to Str resistance (Hongtai Zhang et al. 2013).

Given such broad targets for increasing the fitness of AR mutants we may expect that these resistances change the evolvability of bacteria, through altering the range of beneficial mutations that can be accessed by a given resistance genotype. Here we use the term evolvability as “the genome’s ability to produce adaptive variants” (Wagner & Altenberg 1996). In this sense, a costly AR mutant may be able to effectively compete with a less costly resistance allele (*i.e.*, with a higher selective coefficient when in direct competition) if the evolvability of the former is higher. The conditions for this to happen will depend on the distribution of effects of beneficial mutations (DEBM), as well as on the rate of beneficial mutations available to each genotype (Handel et al. 2006; Sousa et al. 2012; Barrick et al. 2010).

Most studies addressing the competitive fitness effects associated with AR, consider resistant strains competing with the ancestral sensitive strain. However, in many natural environments the frequency of resistance can be very high and competition between different resistant alleles may be a common event (Farhat et al. 2013; Hongtai Zhang et al. 2013; Forsberg et al. 2012; Nolan et al. 1995; Mwangi et al. 2007; Solomon H Mariam et al. 2011; Sun et al. 2012). For instance in (Solomon H Mariam et al. 2011), a monoclonal *Mycobacterium tuberculosis* infection was followed and both clonal sweeps and the coexistence of different resistant mutants were observed in the dynamics of the population. Competition between different resistant strains is also likely to occur whenever there is spatial heterogeneity, with different areas posing different selective pressures (Hermsen et al. 2012; Q Zhang et al. 2011; Solomon H Mariam et al. 2011). Furthermore, bacteria with multiple resistance alleles are also commonly segregating in natural populations (Wright 2010; Borrell et al. 2013; Merker et al. 2013). This competitive context might therefore play a crucial role in the maintenance of antibiotic resistance.

Here we study the process of fitness recovery mimicking an environment with different resistance backgrounds competing at high frequency. We use an experimental evolution approach (Jansen et al. 2014) to test the ability of clones

with costly AR alleles to coexist or even outcompete clones with less costly resistances. In a drug free environment, the differences in the fitness costs of resistance alleles should determine their probability of extinction. Contrary to this simple expectation, we observe the maintenance of costly resistance alleles over hundreds of generations even when their fitness impairment should predict fast extinction. We infer that differences in adaptive potential for each AR mutant exist and suggest that these can explain the observed outcomes in the evolution of resistance.

MATERIALS AND METHODS

Bacterial strains and growth conditions

All bacterial strains used in this study were derived from *Escherichia coli* K12 MG1655 and have in common the following genotype: *galK::yfp/cfp* cm^R (pKD3), Δ*lacIZYA*. The yellow (*yfp*) and cyan (*cfp*) alleles were integrated at the *galK* locus under the control of the *lac* promoter and were constructed by P1 transduction(Silhavy et al. 1984) of *yfp/cfp* inserts from previously constructed strains (Hegreness 2006). In these strains the fluorescence marker is constitutively expressed. In this common backbone different antibiotic resistance mutations were introduced by P1 transduction. The donor strains were spontaneous antibiotic resistant mutants previously obtained by plating the sensitive bacteria in Luria-Bertani (LB) supplemented with agar and 100 mg/mL of either streptomycin or rifampicin (Trindade et al. 2009). A total of eight strains were constructed: *K43T*^{Str}-YFP, *S531F*^{Rif}-YFP, *H526Y*^{Rif}-YFP, *H526D*^{Rif}-YFP and *K43T*^{Str}-CFP, *S531F*^{Rif}-CFP, *H526Y*^{Rif}-CFP, *H526D*^{Rif}-CFP, such that the same resistance allele was introduced in the two fluorescence backgrounds. *K43T*^{Str} confers resistance to streptomycin and all the other aminoacid changes confer resistance to rifampicin. In order to confirm the identity of the mutations transferred by P1 transduction, the antibiotic resistance target gene (*rpoB* for rifampicin or *rpsL* for streptomycin) was amplified and sequenced using the following primers:

for the relevant fragment of the *rpoB* gene, 5'-CGTCGTATCCGTTCCGTTGG-3' and 5'-TTCACCCGGATAACATCTCGTC-3' and for the *rpsL* gene, 5'-ATGATGGCGGGATCGTTG-3' and 5'-CTTCCAGTTCAGATTTACC-3'. Each resistant clone was grown from a single colony in LB medium supplemented with the respective antibiotic at 37ºC with aeration and stored in 15% glycerol at - 80°C.

Fitness assays and test for frequency dependent selection

The fitness costs of antibiotic resistance mutations were first measured in competition against a sensitive reference strain (Table S1). The reference strain carried a *yfp* allele if the resistant strain carried the *cfp* allele (and vice versa). Competitions were done after acclimatization, where each bacterial strain was grown in the same environment of the competition: in a 96 well plate with 150µL of LB per well at 37°C with aeration. Acclimatization consisted of two consecutive passages where 5µL from the first 24 hour grown culture were used to inoculate a new plate for another 24h. Competitions were performed by inoculating ~10^5 cells of both competitor and reference strain in LB medium and allowed to grow for 24 hours (~9 generations). The initial and final ratios of both strains were determined by Flow Cytometry. Fitness effects of the resistance mutations were estimated as the slope between 0 and 24h of the $\ln(f(N_R)/(1-f(N_R)))$, where $f(N_R)$ is the frequency of resistant bacteria in the population or one of the reference resistance backgrounds, in the case of the competitions of the evolved clones.

H526Y^{Rif} and *H526D*^{Rif} strains were tested for negative frequency dependent selection in the same conditions as described above, and the number of cells was measured using the Flow Cytometer BD LSR Fortessa (BD Biosciences), at different initial ratios of the two strains: 100:1, 10:1, 1:1, 1:10 and 1:100 (*H526Y*^{Rif}:*H526D*^{Rif}).

Long-term propagation of resistant populations

Prior to the start of the long-term competitions, acclimatization of the bacterial strains was performed in plates with a checkered arrangement (one plate for YFP strains, another for CFP strains). Each well was inoculated with an

independent starting sample, from a frozen culture, into 150ml of LB medium, to maximize the probability of sampling a large pool of beneficial mutations. We note that it is possible that beneficial mutations occur during the acclimatization period, since the rate of mutations which compensate for the cost of resistance can be very high. After 24 hours, 5ml of the grown cultures where inoculated into fresh LB media and after 48h the numbers of bacteria were measured. Appropriate dilutions were done to achieve the required initial ratio (1:1) of YFP and CFP strains for the long-term evolution of competing clones. This evolution was performed in the same conditions as the fitness assays, with daily passages of about ~10^5 bacteria for ~280 generations (30 days). Samples of the evolving populations were frozen every day, from which the relative abundance of each resistance was followed by measuring the frequencies of their linked fluorescent alleles by Flow Cytometry. The following three pairs of mutants were studied: *K43T*^{Str} *vs S531F*^{Rif}, *H526Y*^{Rif} *vs H526D*^{Rif} and *H526Y*^{Rif} *vs S531F*^{Rif}. For each pairwise competition of mutants, 16 replicas were performed: half where one of the resistances, say R1, was linked to the *yfp* background and the other half where it was linked with the *cfp* background. Pairwise combinations of mutants (R1-YPF vs R2-CFP and R1-CFP vs R2-YFP) were settled in a checkerboard arrangement (**Figure 1**), where half of the wells were filled solely with LB to control for external contamination.

Estimation of relative parameters of beneficial mutations

The dynamics of a given costly resistance allele (R1) when competing with another resistant clone (R2), with a different fitness cost, were analyzed under a simple model of positive selection, that assumes the occurrence of new beneficial alleles which are sweeping towards fixation. We first fitted the simplest possible model (Model 1), which can allow for a costly resistance allele to be maintained for hundreds of generations despite its initial cost. In this model we assume an initial population composed of two distinct genotypes, with different resistances and initial fitnesses, w_{R1} and w_{R2}. The initial frequency of the more costly genotype, $R2_0$, is taken from a uniform distribution within the interval $R2_{Experimental0} \pm 0.1$, while

that of the other genotype is $R1_0$=1-$R2_0$. At generation T_N a new genotype, with fitness w_N, is assumed to have arisen and reached a frequency N_0=0.001. From that time onwards it is assumed to change in frequency deterministically towards fixation. Therefore, we modeled selection deterministically in discrete time with two genotypes for $t<T_N$, and three for $t>T_N$. We estimated, by maximum likelihood, the parameters $R2_0$, T_N and w_N that best fit the observed values of ln(M/(1-M)), where M is the measured frequency of a fluorescent marker, at several time points, assuming a normal distribution for measurement error (with average 0 and standard deviation 0.2). The search for the set of parameters that maximize the likelihood of the data in the space of the possible parameters was performed using the Nelder-Mead Method, as implemented in Mathematica 8.0 (http://mathworld.wolfram.com/Nelder-MeadMethod.html), with 100 iterations and repeated for 100 realizations with different initial starting combinations of parameter values. While Model 1 could provide a reasonable fit for some of the replicate experimental lines, for others it did not. We therefore fitted the next simplest model (Model 2) which assumes that two beneficial mutants (N1 and N2), one for each background and fluorescent marker (with fitnesses W_{N1} and W_{N2}), emerged at times (T_{N1} and T_{N2}). We selected the model with the lower AIC (Akaike Information Criteria) to test if the different resistances would have distinct evolvabilities. Specifically, for each pair of competing resistances and each experimental evolved population, we asked if the times of appearance of beneficial mutations or the effects of accumulated beneficial mutations were significantly different. If the genetic target for acquiring beneficial mutations is larger for resistance background R2 than for R1, but the effects of the compensatory mutations are similar, then we expect the time of appearance of mutations to be smaller in R2 than in R1. If the target is similar but the effects depend on the background we expect to observe a significant difference between the effects accumulated in one resistance background versus the other. The method used here is a discrete adaptation of existing methodologies to infer evolutionary parameters (Illingworth & Mustonen 2012), with the added difference of allowing for different initial fitness values in competing genetic backgrounds. Code for these

simulations is available upon request. The inference process and summary relative effects are shown for a single line, as an example, in **Figure S1**.

Whole genome sequencing of H526Y^{Rif} *and S531F*^{Rif} *evolved clones*

A single clone of each resistance (*H526Y*^{Rif} and *S531F*^{Rif}) from each evolved population was isolated in LB agar plates with rifampicin at the end of the evolution experiment. In six populations the frequency of clones from the *S531F*^{Rif} background was very low and no resistant clone was sampled. DNA was extracted from each sampled clone, and then pooled according to resistance background. Paired-end sequencing using Illumina MiSeq Benchtop Sequencer, with mean coverage of 207x was performed. The resulting reads were trimmed at a Phred quality score of 99.9%, and were then aligned using *Escherichia coli* K12 MG1655 (NC_000913.2) as the reference genome. Mutation prediction was done using version 0.23 of the BRESEQ analysis pipeline available for download at http://barricklab.org/twiki/bin/view/Lab/ToolsBacterialGenomeResequencing, with polymorphism detection on. Other settings were used as default except for: a) requirement of a minimum coverage of 3 reads on each strand per polymorphism; b) polymorphism predictions occurring in homopolymers of length greater than 3 were discarded; c) polymorphism predictions with significant (p-value<0.05) strand or base quality score bias were discarded.

RESULTS

Pairwise competitions between different AR alleles were performed in an antibiotic free environment for around 280 generations. Each strain carries a neutral fluorescence marker. Three pairs of resistance mutations were studied: 2 pairs involving 3 different Rif resistance alleles, and one pair involving a Str and a Rif resistance allele (**Figure 1**). The costs of each resistance allele measured against a sensitive reference strain of *E. coli* are shown in **Table S1**. The cost of resistance in relation to the sensitive is higher for the *S531F*^{Rif} mutation (0.1 ± 0.01) and the *K43T*^{Rif} mutation (0.09 ± 0.03), followed by *H526Y*^{Rif} (0.07 ± 0.01)

and *H526D*^{Rif} (0.06 ± 0.02) mutations. We then inferred the fitness cost of the same resistance mutations but this time in the presence of another resistance. Whether a resistance allele was more or less costly was estimated from the change in frequency of the fluorescent markers, linked to the resistances, during the first three days of evolution where the pairs of AR mutants were competing (**Figure 2**, **3** and **4**, and also **Figure S2** for each of the replicates identified by its neutral marker). We assume that during this short period new beneficial mutations have not yet arisen, or are at a frequency too low to interfere with the relative fitness differences that both clones may have.

For each pairwise evolution experiment we query whether different AR mutants show distinct adaptive potential. For that, we inferred the fitness effects of new beneficial mutations that could explain the observed long-term frequency dynamics of the AR alleles. We allowed for one or two beneficial mutations of different effects to emerge at different times, in either of the backgrounds (see Methods). We sought to infer the combination of parameters (time of emergence, effect of the beneficial mutation, and the initial frequency of one of the resistance alleles) that provided the best fit to the observed evolutionary dynamics. We started by fitting a model where a single new beneficial mutation (with effect S_N, at time T_N) escapes stochastic loss in the background with lower initial fitness (Model 1). We estimated the value of S_N and T_N that best fits the dynamics of each resistance, under this model. We then used a model assuming that two beneficial mutations had increased in frequency, one in each background, and fitted the parameters of this second model (T_{N1}, S_{N1}, and T_{N2}, S_{N2}) to the dynamics observed. The lines shown in **Figure S3** represent the model that best fits the data, and the inferred parameters are presented in **Table 1**, **2** and **3**.

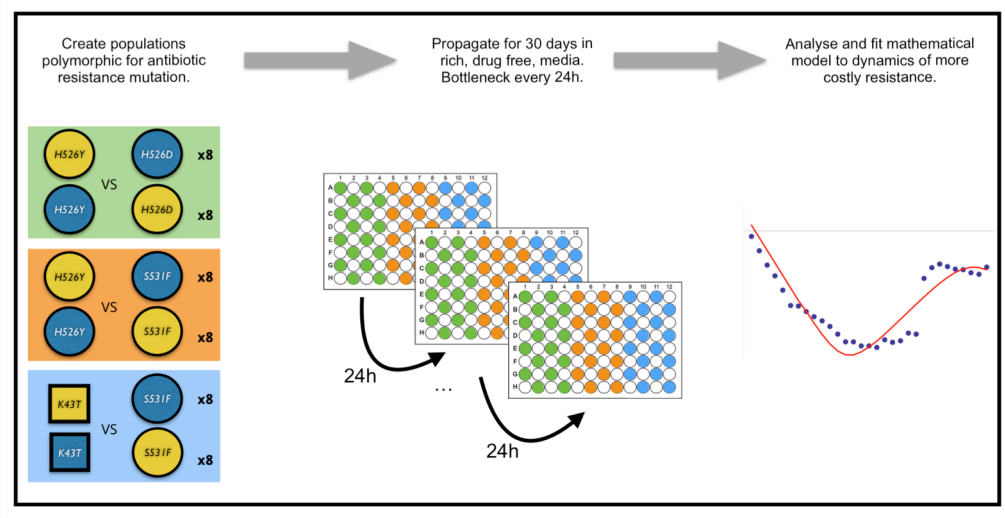


Figure 1. Testing the evolvability of different antibiotic resistances. First, polymorphic populations are created by mixing, in a 1:1 ratio, clones with two different antibiotic resistance mutations. All clones are isogenic, except for this different resistance mutation and a neutral fluorescent marker. 16 replicate populations are followed for each competition, where half of these replicates have one of the resistances linked with one of the neutral markers, while the other to the other marker. 3 different competitive scenarios were studied, between Rifampicin mutations (circles) and between Rifampicin and Streptomycin (squares). Next, populations are passaged in rich media in 96 well plates, organized in a checkered layout and separated by control wells (with media but without bacteria). Every 24 hours, all populations are passaged serially into a new 96 well plate with fresh media, during 30 days. The frequencies of the neutral markers (and hence the resistance alleles) are analyzed to unravel possible differences in the evolutionary parameters according to the resistance background.

Potential for adaptation causes the maintenance of high cost resistance

We first studied the fate of two AR mutations, conferring resistance to the same antibiotic – Rif – and altering the same aminoacid in the β subunit of RNA polymerase: *H526Y*^{Rif} and *H526D*^{Rif}. While the costs of these resistances are not statistically different when measured against a sensitive strain, their relative fitness

cost differs when they compete. From the initial generations of competition between these two strains, a fitness difference of 0.074 was inferred, with strain *H526Y*^{Rif} being highly detrimental in this competitive environment (**Figure 2A**, black solid line). The cost carried by strain *H562Y*^{Rif} leads to the rapid decline in its frequency, detected in the first tens of generations. With such a relative fitness cost, it should go extinct in 100 generations (as indicated by the black line). Strikingly though, *H526Y*^{Rif} is clearly able to resist extinction in the majority of the replicate lines that were evolved. Four of the lines show a frequency too low to be reliably detected by Flow Cytometry. However, plating these populations at generation 280 allows observing the presence of fluorescent colonies of the *H526Y*^{Rif} background at extremely low frequencies. This lack of extinction can be expected under two different scenarios: negative frequency-dependent selection and/or an ability to access higher effect beneficial mutations. We therefore tested for a possible signature of negative frequency-dependent selection, i.e., advantage from rarity of the *H526Y*^{Rif} mutation when competing against *H526D*^{Rif}. In competitions between these two backgrounds, starting with different frequencies of the *H526Y*^{Rif} allele, no advantage from rarity can be detected (**Figure S4**), so frequency-dependent selection is unlikely to be responsible for the observed lack of extinction of this mutation. On the other hand, analysis of the long-term frequency dynamics revealed a significant difference between the fitness effects of the mutations inferred to have emerged, with its median difference (s*H526Y*^{Rif} – s*H526D*^{Rif}) deviating from 0 ($P<0.001$, Wilcoxon Signed Rank Test, **Figure 2B**). The mean difference in fitness effects between these two genetic backgrounds was 0.07 (**Table 1**, **Figure 2B**), with *H526Y*^{Rif} background accumulating stronger effect mutations. Regarding the times of appearance of new beneficial mutations, no significant difference was detected between the backgrounds ($P=0.41$, Wilcoxon Signed Rank Test, **Figure 2C**). When considering these results together, there is a strong indication that the mean effect of beneficial or compensatory mutations for *H526Y*^{Rif} is higher, thus qualifying the strain with the *H526Y*^{Rif} mutation as more evolvable than the strain with the *H526D*^{Rif} mutation. Such

higher evolvability can explain the avoidance of extinction of background *H526Y*^{Rif}, contrary to the *a priori* expectation based on its lower initial fitness.

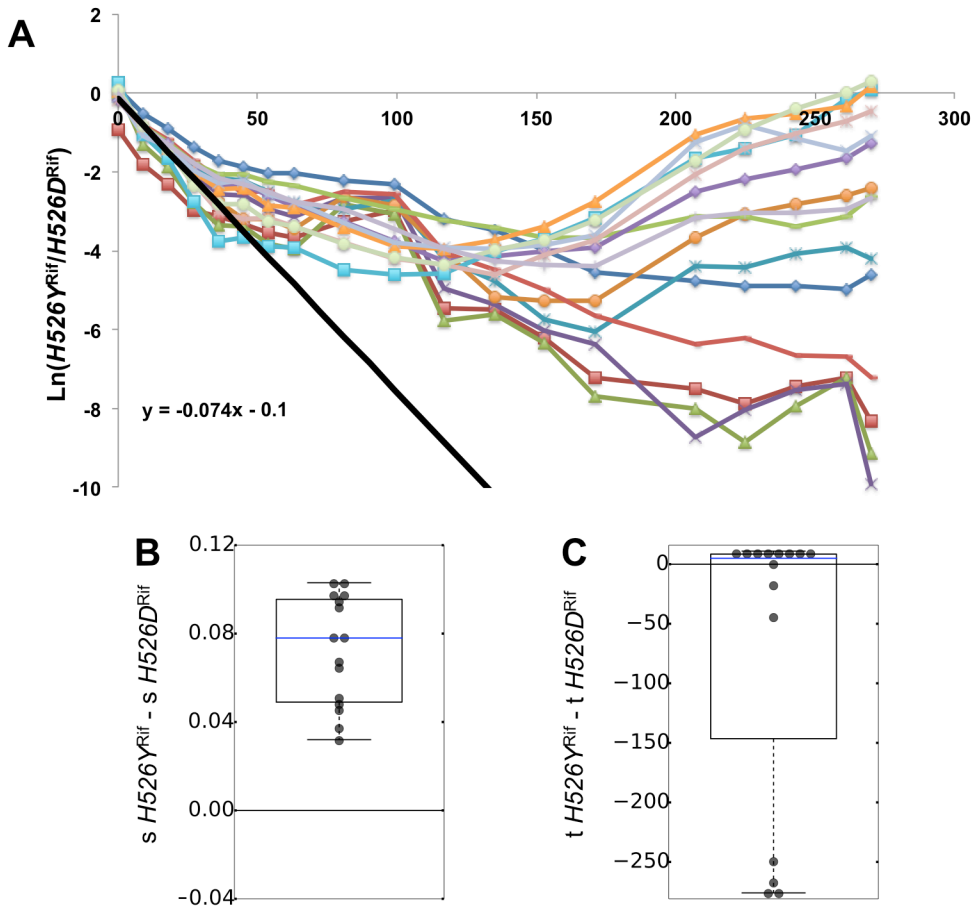


Figure 2. High evolvability of costly *H526Y* allows its long-term maintenance A) Long-term dynamics, for 30 days (280 generations) of evolution in 15 replicates of a population composed of resistance strain H562Y and resistance strain *H526D*. Shown are the dynamics for the *H526Y* background. The slope of the black line represents the initial difference in fitness between the resistances. **B)** Whisker-box shows the relative fitness differences inferred for new beneficial mutations between the two resistant backgrounds, with *H526Y* background as a reference. **C)** Whisker-box shows the relative differences in time of appearance inferred for new beneficial mutations between the two resistance backgrounds.

		H526D (R1)		H526Y (R2)	
	Initial H526Y Frequency	W	T	W	T
Population 1	0.538	1.576	11	1.552	15
Population 2	0.382	1.293	11	1.263	22
Population 3	0.488	1.387	41	1.348	51
Population 4	0.468	1.483	26	1.44	32
Population 5	0.413	1.421	21	1.409	30
Population 6	0.476	1.474	31	1.478	42
Population 7	0.513	1.218	11	1.19	19
Population 8	0.493	1.776	18	1.767	23
Population 9	0.557			1.017	3
Population 10	0.461			1.028	29
Population 11	0.581			1.023	3
Population 12	0.58	1.002	46	1.02	3
Population 13	0.523			1.023	14
Population 14	0.55	1.001	38	1.028	23
Population 15	0.59	1.668	37	1.671	45

Table 1. Evolutionary Parameters estimated for the competition between strains *H526Y* and *H526D*. W stands for the fitness of the emerging haplotype and T its time of appearance. Initial Freq stands for the inferred initial frequency of the *H526Y* background. In the cases where only one of the backgrounds has acquired a mutation, the inferred parameters are in bold for the background where no mutation was inferred.

Higher evolvability of H526Y^{Rif} upheld in competition with an alternative resistance

We then studied the ability of the *H562Y*^{Rif} background to outcompete a mutant bearing a resistance mutation in a different aminoacid, *S531F*^{Rif}. Resistance to the antibiotic rifampicin is now provided by mutations causing different aminoacid substitutions in the β subunit of RNA polymerase. . Relative to

the sensitive strain, *H526Y*^{Rif} is estimated to impose a smaller mean fitness cost than *S531F*^{Rif} (see **Table S1**), although the difference is not significant. The relative fitness difference between strains *H562Y*^{Rif} and *S531F*^{Rif} inferred from their direct competition is 0.015 (**Figure 3A**, black solid line). All lines initially decrease in frequency, in accordance with a cost of *H526Y*^{Rif}, but in all replicates this tendency is inverted. After generation 30, the frequency of each of the resistance background stabilizes at 50% in all populations. After generation 125, additional changes in frequency can be detected. This is expected if other arising beneficial mutations increase in frequency in one or both backgrounds. In some replicate lines one of the resistance alleles starts to increase in frequency but later decreases. This is likely the result of clonal interference (Sniegowski & Gerrish 2010), with multiple beneficial mutations competing amongst them. There were more replicates in which *H526Y*^{Rif} increased in frequency than expected, even under the assumption that the initial fitness difference between the backgrounds would be negligible. At generation 280, 14 out of 16 lines have a frequency of this allele above that expected by chance (p=0.004, Binomial two-sided test), which is even more striking considering that its initial frequency is below 50% in most of the lines. This deviation suggests again differences in the adaptive potential between the resistance backgrounds.

We tested for possible differences in the adaptive potential of each strain by inferring the time of appearance and fitness effects of new beneficial mutations that could explain the changes in frequency of the neutral markers (**Figures 3B** and **3C**). The mean value for differences in fitness between the two genetic backgrounds was 0.034 (see **Figure 3B** and **Table 2**). The median of the distribution of relative effects deviates significantly from 0 (P<0.001, Wilcoxon Signed Rank test, **Figure 3B**). The times of appearance of the new inferred beneficial mutations, however, did not appear to be significantly different between the genetic backgrounds (P=0.3, Wilcoxon Signed Rank test, **Figure 3C**). The overrepresentation of the *H526Y*^{Rif} resistance background in the evolving populations, together with this analysis, indicates that *H526Y*^{Rif} has a higher

evolvability across different competitive contexts and therefore can be easily maintained.

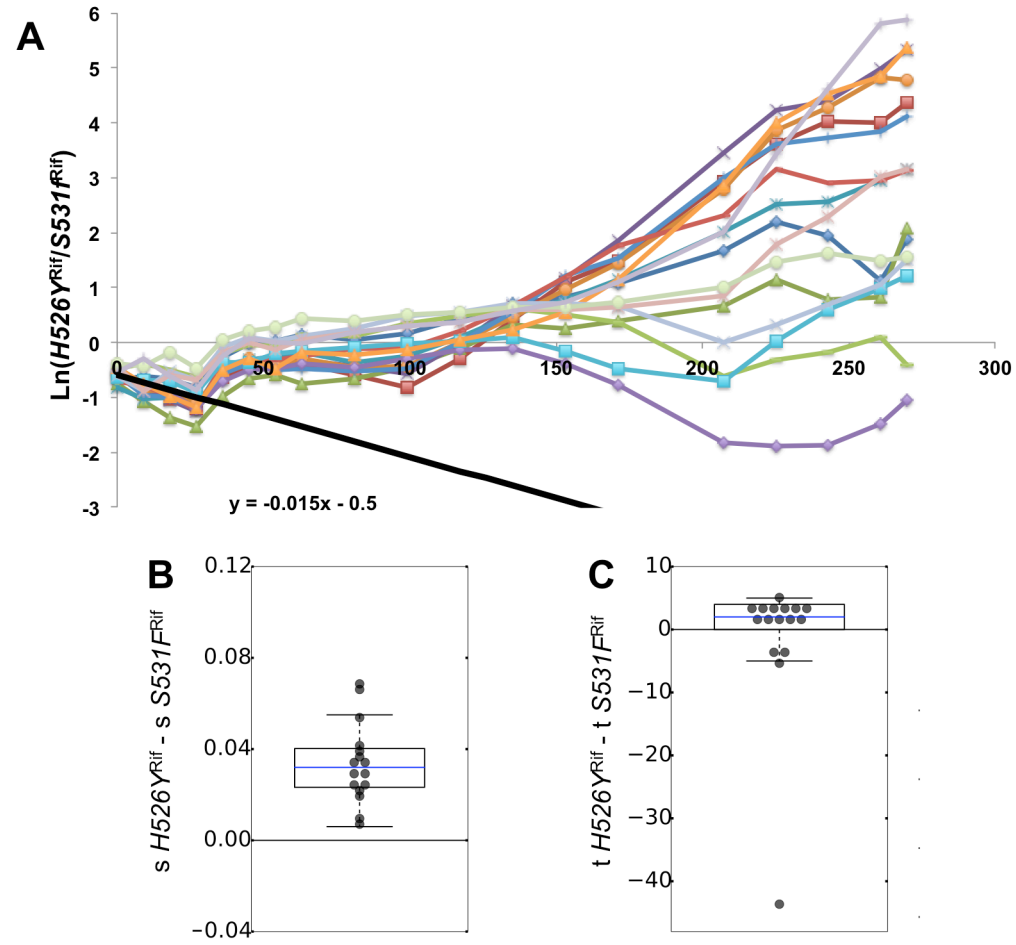


Figure 3. Different evolvabilities between Rifampicin resistance alleles A) Long-term dynamics, for 30 days (280 generations) of evolution in 16 replicates of a population composed of resistance strain *H562Y* and resistance strain *S531F*. Shown are the dynamics for the *H526Y* background. The slope of the black line represents the initial difference in fitness between the resistances. **B)** Whisker-box shows the relative fitness differences inferred for new beneficial mutations between the two resistant backgrounds, with *H526Y* background as a reference. **C)** Whisker-box shows the relative differences in time of appearance inferred for new beneficial mutations between the two resistance backgrounds.

		S531F (R1)		H526Y (R2)	
	Initial H526Y Frequency	W	T	W	T
Population 1	0.369	1.459	15	1.472	17
Population 2	0.458	1.041	46	1.051	3
Population 3	0.277	1.436	15	1.45	19
Population 4	0.472	1.034	7	1.054	4
Population 5	0.286	1.529	8	1.554	12
Population 6	0.454	1.585	55	1.637	60
Population 7	0.381	1.615	36	1.655	40
Population 8	0.325	1.465	11	1.491	15
Population 9	0.473	1.199	19	1.194	14
Population 10	0.378	1.163	14	1.154	10
Population 11	0.356	1.595	17	1.601	19
Population 12	0.483	1.028	3	1.049	4
Population 13	0.37	1.344	11	1.349	12
Population 14	0.444	1.542	9	1.562	12
Population 15	0.446	1.505	16	1.514	17
Population 16	0.468	1.647	55	1.7	59

Table 2. Evolutionary Parameters estimated for the competition between strains *H526Y* and *S531F*. The meaning of the parameters is as in **Table 1**.

Potential for adaptation drives the fate of resistances to different antibiotics

Differences in evolvability are expected to be larger amongst resistances affecting different genes than between alleles from the same gene. This is so because the target for beneficial mutations is expected to be more similar between mutations affecting the same function than between mutations impairing different traits. To query if the costs or the evolvabilities are determinant to the competition of resistances to distinct drugs, we studied the fate of the Str resistance allele

(*K43T*^{Str}) when in competition with the Rif resistance allele (*S531F*^{Rif}). *K43T*^{Str} is estimated to impose a cost to the sensitive strain of ~0.09, which is not statistically different from the cost imposed by *S531F*^{Rif}. However, upon competition between these two strains, *K43T*^{Str} shows a disadvantage of around 0.03 relative to *S531F*^{Rif} (**Figure 4A**, black solid line). This is observed in the initial 25 generations, where *K43T*^{Str} decreases in frequency in most replicate competitions. The long-term evolutionary dynamics, however, depart from those observed in the previous studied cases. In **Figure 4A**, a higher variation in the outcome of which resistance wins the competition emerges in this pair of competing resistances. In the vast majority of the replicate lines the *K43T*^{Str} mutation survived extinction for the duration of the experiment, contrary to what would have been expected from its initial relative fitness cost. In most of the lines, *K43T*^{Str} is kept at a stable frequency of around 30% and in three lines it rises in frequency, sweeping to majority status (above 99% frequency) by generation 280. Contrary to expectations, extinction (frequency below 1%) of the more costly allele was only observed in two of the populations. Most of the lines where the *K43T*^{Str} resistance was kept at a stable but low frequency at the middle time points had different outcomes, with some rising and others decreasing in frequency. Overall, at the end of the 280 generations, the background with the *K43T*^{Str} allele reached a frequency higher than 50% in 6 out of 15 evolved replicate lines. The inference of evolutionary parameters performed for the long-term dynamics indicates that there is a significant difference in the strength of mutations acquired by each background (P=0.014, Wilcoxon Signed Rank Test, **Figure 4B**). *K43T*^{Str} acquires beneficial mutations of a stronger effect than does the background *S531F*^{Rif} (mean difference 0.03, **Table 3**). No significant difference was detected for the times of appearance of beneficial mutations between both backgrounds (P=0.6, Wilcoxon Signed Rank Test, **Figure 4C**). The differences inferred from the dynamics in **Figure 4A** therefore depart from the expectation of a non-epistatic model of beneficial mutations, where both backgrounds would access mutations of similar effect. Extinction might have been avoided for this Str resistance mutation by its ability to accumulate stronger effect mutations when competing with the Rif resistant strain.

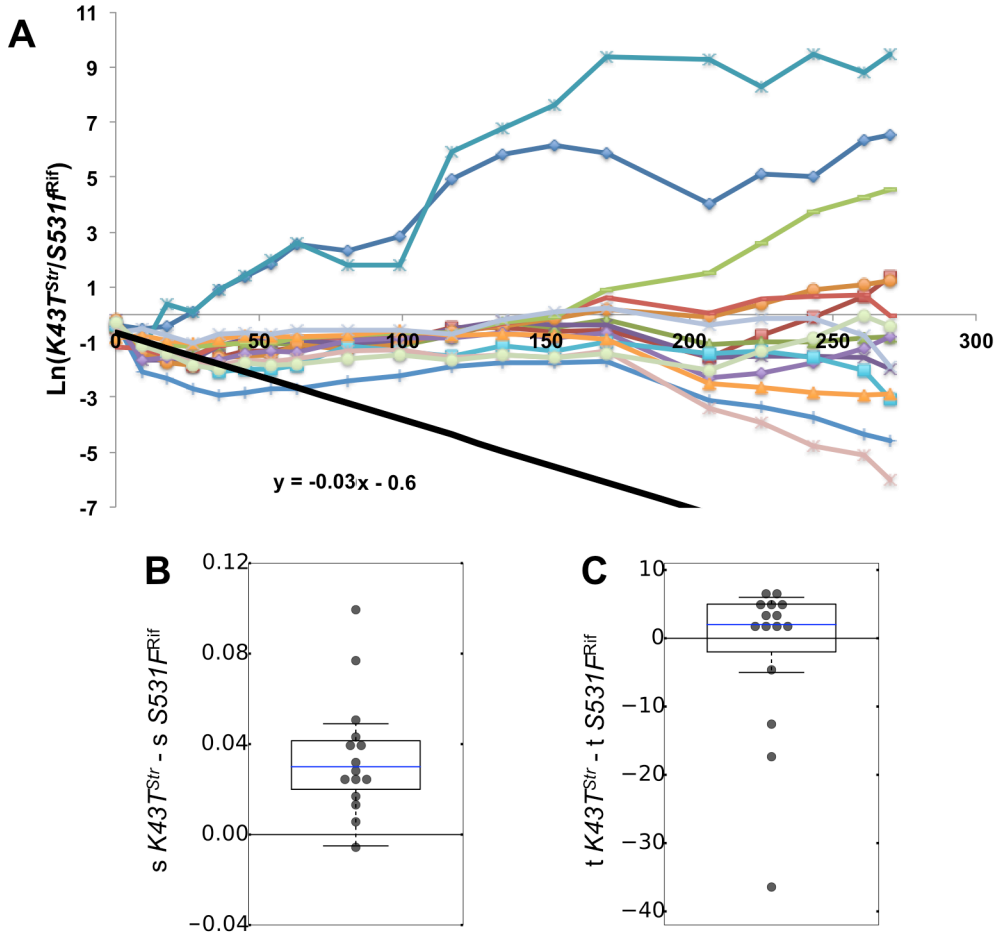


Figure 4. Differences in evolvability between Rifampicin and Streptomycin resistance alleles A) Long-term dynamics, for 30 days (280 generations) of evolution in 14 replicates of a population composed of resistance strain *K43T* and resistance strain *S531F*. Shown are the dynamics for the *K43T* background. The slope of the black line represents the initial difference in fitness between the resistances. **B)** Whisker-box shows the relative fitness differences inferred for new beneficial mutations between the two resistant backgrounds, with *K43T* background as a reference. **C)** Whisker-box shows the relative differences in time of appearance inferred for new beneficial mutations between the two resistance backgrounds.

		S531F (R1)		K43T (R2)	
	Initial K43T Frequency	W	T	W	T
Population 1	0.505	1.207	16	1.188	3
Population 2	0.319	1.349	20	1.359	26
Population 3	0.299	1.546	18	1.547	21
Population 4	0.349	1.26	16	1.253	18
Population 5	0.319	1.669	4	1.737	5
Population 6	0.359	1.457	4	1.474	10
Population 7	0.333	1.088	52	1.06	34
Population 8	0.256	1.325	7	1.334	12
Population 9	0.43	1.687	37	1.732	42
Population 10	0.335	1.097	3	1.09	4
Population 11	0.357	1.537	50	1.529	53
Population 12	0.388	1.16	11	1.144	6
Population 13	0.41	1.195	5	1.193	7
Population 14	0.379	1.095	43	1.058	6
Population 15	0.329	1.663	25	1.672	30

Table 3. Evolutionary Parameters estimated for the competition between strains *K43T* and *S531F*. The meaning of the parameters is as in Table 1.

Genetic characterization and fitness determination of evolved clones isolated from the long-term competition between H526Y^{Rif} *and S531F*^{Rif}

In the three cases of long-term competitions studied, the differences in evolvability inferred resulted from differences in the effects of the beneficial mutations acquired. We inferred that *H526Y*^{Rif} could acquire higher effect mutations, relative to those emerging in *S531F*^{Rif} clones, even though these two different alleles cause very similar fitness costs. To gain further insight into the

mutations acquired by the *H526Y*^{Rif} background in this competitive context, we performed whole genome sequencing (see Methods for a description of the procedure) of a sample of *H526Y*^{Rif} and of *S531F*^{Rif} clones, after these two lineages had evolved in competition (**Figure 3**). **Table 4** shows the mutations identified in each of the evolved backgrounds. We could identify 12 (in 10 distinct genes or intergenic regions) for possible beneficial (and compensatory) mutations in *H526Y*^{Rif} (all single point mutations), and 10 (all in different genes/intergenic regions) in *S531F*^{Rif} (seven of them were single point mutations , the other three involved transpositions of insertion sequence elements). Mutations in the *H526Y*^{Rif} background were observed at a higher frequency compared to the ones observed in *S531F*^{Rif}. This suggests that the mutations appearing in the *H526Y*^{Rif} background have stronger effects, since they are acquired by different replicate evolving populations, pointing to a rapid fixation due to their beneficial effect. To enquire if this is indeed the case we performed competition assays between each individual clone and the ancestral strains. Since the whole genome sequencing was performed with a mixture of clones (see Methods) we targeted sequencing each individual clone to gain access to the haplotypic composition of each clone. For the clones carrying the *H526Y*^{Rif} mutation, we targeted the most prevalent mutations detected (see **Table 4**). Fifteen out of the 16 *H526Y*^{Rif} clones carried at least one mutation in the target genes (**Supplementary Table 2**) and one clone was found to have two mutations, one single point mutation in *rpoA* and another in *waaZ*. The aminoacid change T196I in *rpoA* was present in 8 independently evolved clones. Since all of them have the same genetic background (*H526Y*^{Rif} YFP), this suggests that this mutation could already have been present prior to the long-term evolution experiment. All *S531F*^{Rif} clones, except one, carried one of the previously identified mutations (**Supplementary Table 2**) and *rpoC* was also a target for beneficial mutations in this Rif^{R} background. We then measured the competitive fitness of each of these evolved clones to directly assay their fitness advantage. In order to increase in frequency during the long-term propagation, these mutants had to outcompete their ancestral and also the competitor with a different resistance background. **Figure 5** shows that when competing either

against their respective ancestral or the other resistance clone, the selective effects of the evolved *H526Y*^{Rif} clones are stronger than the selective effects of the mutations acquired by the *S531F*^{Rif} background. For the latter background, the mutations are advantageous against their ancestral, supporting their increase to detectable frequencies, but most are neutral or even deleterious versus the opposite *H526Y*^{Rif} background. Remarkably, mutations in *rpoA* and *rpoC* provided the strongest competitive fitness advantages. Overall, these results therefore provide further support for the previously identified higher evolvability of the *H526Y*^{Rif} background and strongly suggest that the *H526Y*^{Rif} resistance may be easily maintained in populations.

Background	Genome Position	Gene(s)	Mutation	Annotation	Frequency	Function
H526Y	3,438,465	*rpoA*	C→T	T196I	51.5%	*RNA polymerase, alpha subunit*
	4,184,828	*rpoC*	T→C	S486P	25.5%	*RNA polymerase, beta prime subunit*
	4,187,042	*rpoC*	C→T	R1224C	6.1%	
	4,184,721	*rpoC*	A→C	H450P	2.9%	
	3,797,382	*waaZ*	T→C	S280P	19.7%	*Lipopolysaccharide core biosynthesis protein*
	3,946,224	*yifE*	G→A	G39D	12.0%	*Conserved protein, UPF0438 family, unknown function*
	4,292,389	*nrfG/gltP*	A→T	Intergenic	5.0%	***[nrfG]*** *Heme lyase (NrfEFG) for insertion of heme into c552, subunit NrfG/****[gltP]*** *glutamate/aspartate: proton symporte*
	3,637,091	*pitA*	C→T	A476V	2.6%	*Phosphate transporter, low-affinity; tellurite importer*
	374,196	*mhpE/mhpT*	T→C	Intergenic	2.4%	*4-hyroxy-2-oxovalerate/4-hydroxy-2-oxopentanoic acid aldolase, class I/putative 3-hydroxyphenylpropionic transporter*
	3,464,375	*bfr*	C→T	R125C	2.0%	*Bacterioferritin, iron storage and detoxification protein*
	1,743,180	*ydhQ*	T→C	S324S	3.1%	*Hypothetical protein*
	3,898,043	*yieL*	C→A	T186T	2.5%	*Putative xylanase*
S531F	4,183,521	*rpoC*	A→C	K50T	14.8%	*RNA polymerase, beta prime subunit*
	4,292,389	*nrfG/gltP*	A→T	Intergenic	19.6%	***[nrfG]*** *Heme lyase (NrfEFG) for insertion of heme into c552, subunit NrfG/* ***[gltP]*** *Glutamate/aspartate:proton symporte*
	2,749,808	*nadK/recN*	Δ1 bp	Intergenic	18.1%	***[nadK]*** *NAD kinase/* ***[recN]*** *Recombination and repair protein*
	1,940,499	*znuA*	T→C	F37L	17.7%	*Zinc transporter subunit: periplasmic-binding component of ABC superfamily*
	3,374,976	*sspA*	T→C	L156P	12.2%	*Stringent starvation protein A*
	300,420	*paoB*	T→A	V231E	11.8%	*PaoABC aldehyde oxidoreductase, FAD-containing subunit*
	2,036,728	*yedW*	IS1 Ins	Coding	9.9%	*Putative DNA-binding response regulator in two-component system with YedV*
	1,877,853	*yeaR*	IS186 Ins	Coding	8.4%	*Hypothetical protein*
	3,410,893	*envR*	IS5 Ins	Coding	6.8%	*DNA-binding transcriptional regulator*
	2,361,326	*yfaY*	A→G	V110V	9.4%	*Hypothetical protein*

Table 4. Potential compensatory mutations identified in the genomes of the clones evolved in the competition between resistances H526Y and S531F

Mutations were identified in the genomes of the evolved clones in comparison with a reference genome. Shown are the Single Nucleotide Polymorphisms (SNPs) or Insertion Sequence (IS) events identified in either of the backgrounds, and the frequencies at which they were detected. Mutations that occurred between genes (intergenic) are identified as such; otherwise all remaining mutations occurred within the gene indicated. The mutations previously identified to be compensatory are identified with a grey shaded cell.

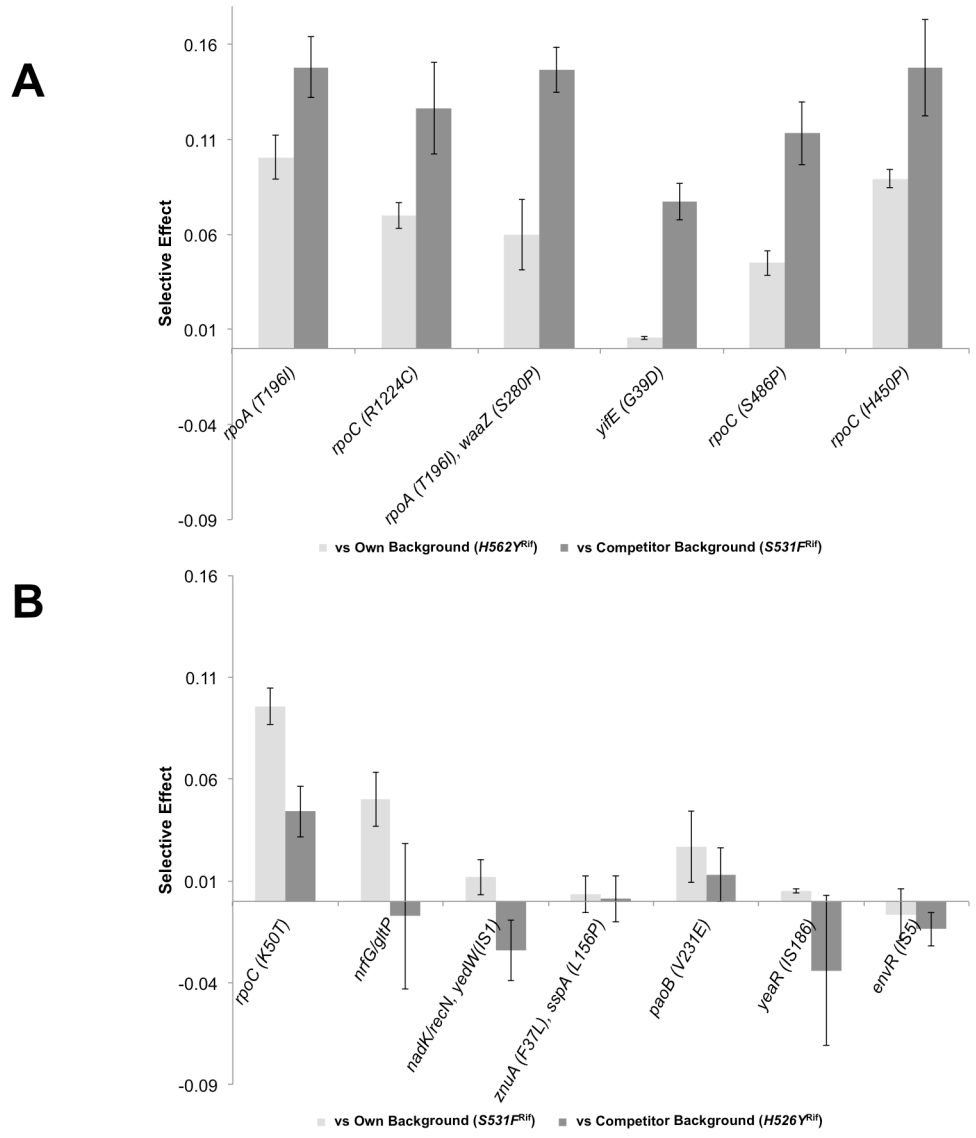


Figure 5. Differences in selective effects of evolved clones from the competition between *H526Y*^{Rif} and *S531F*^{Rif}. A) Fitness effects of selected clones evolved from the *H526Y*^{Rif} background, when in competition with their ancestor (light bars) or against their other resistant competitor (S531F^{Rif}, dark bars). **B)** Fitness effects of selected clones evolved from the *S531F*^{Rif} background, when in competition with their ancestor (light bars) or against their other resistant competitor (H526Y^{Rif}, dark bars).

DISCUSSION

In order to understand the relative role of cost versus evolvability in the maintenance of AR mutations we studied how subpopulations carrying different resistant alleles compete. This system mimics the composition of a population after drug exposure, where the different AR mutants co-exist at relatively high frequencies. This scenario has rarely been studied, though it possibly occurs in natural contexts, as has been observed in strains sampled from the same patient (Solomon H Mariam et al. 2011; Sun et al. 2012). Here we address this situation by following the fate of pairwise combinations of different alleles conferring resistance to the same or different drugs. We observe that the cost of resistance measured in competition against the sensitive bacterium is not always a good predictor of the difference in costs between two resistance mutations, and transitivity between these two fitness measures is not always observed. Therefore, to understand why some resistant alleles are rarely segregating while others are pervasive, it is important to also measure their selective coefficients when in coexistence in addition to measure their costs against the sensitive strain.

In this work we studied two sets of mutations whose fitness costs were barely distinguishable when compared with the sensitive strain. The first set comprised 3 alleles of the same gene (*H526Y*^{Rif}, *H526D*^{Rif} and *S531F*^{Rif}), thus conferring resistance to the same drug; the second group comprised alleles of different genes, (*S531F*^{Rif} and *K43T*^{Str}), thus conferring resistance to different antibiotics. In the first set we found that *H526Y*^{Rif} has a very significant cost (0.074) when competing with mutation *H526D*^{Rif} and it is less costly when competing with *S531F* (0.015), implying that, all else being equal, *H526Y*^{Rif} should rapidly go extinct when competing with *H526D*^{Rif}, but do so at a slower pace when in competition with *S531F*^{Rif}. The long-term outcome of these competitions indicated that this does not always occur and the *H526Y*^{Rif} allele can be maintained in both cases. The evolvability analysis undertaken suggests that the difference between *H526Y*^{Rif} and *H526D*^{Rif} could be attributed to a higher mean effect of beneficial mutations accessible to *H526Y*^{Rif}. This result was unexpected, because this pair of

mutations involves the same aminoacid replacement and the different single point mutations cause a fitness defect of similar magnitude. Given this, one could expect similar evolvabilities between the resistance backgrounds (Fisher 1930). The outcome of the competition between *H526Y*^{Rif} and *S531F*^{Rif} (mutations in different aminoacids but again in the same gene) shows a tendency for the former to increase in frequency, and this could also be due to access of the former background to higher effect mutations, when in comparison with the latter. This observation is further supported by the different competitions between *H526Y*^{Rif} and *S531F*^{Rif} clones, where competitive fitness assays showed stronger effects of the mutations acquired by *H526Y*^{Rif}, as predicted by our theoretical analysis. The two cases of competition between different AR mutations both show a long-term advantage of the *H526Y*^{Rif} allele, through its higher mean effect of mutations. Interestingly, *H526Y*^{Rif} has been reported to be the second most frequent mutation, in a variety of clinical settings, among the ones that confer Rif resistance in *M. tuberculosis*, even though it was also estimated that this mutation was the third mutation more costly in the same set (Gagneux et al. 2006; Trauner et al. 2014). This could be explained by the particularly high evolvability of this mutation. In the case of the pair of AR mutants composed of *K43T*^{Str} and *S531F*^{Rif}, differences in evolvability were expected *a priori* since these mutations alter genes responsible for different cellular traits. In fact, evolution was less reproducible in this situation, with different replicates following different dynamics. Our results indicate that *K43T*^{Str} acquired beneficial mutations of stronger effect than *S531F*^{Rif}, suggesting that their DEBM and hence their evolvability are different.

The ability to access specific subsets of beneficial mutations determines how evolvable an organism is. If there were no constraints, then all genotypes, regardless of their composition or competitive context, would be able to adapt in a predictable sequence of mutational events. However, pleiotropy and epistasis may limit the access to new beneficial mutations (Trindade et al. 2009; Khan et al. 2011; Woods et al. 2011), imposing different evolutionary outcomes in different genetic backgrounds, environmental conditions and/or competitive contexts. This is particularly relevant in the context of AR, as emergence of resistance mutants is

fairly common. The resistant clones would presumably be driven to extinction when competing against less costly strains, in the absence of the drugs. Here we show that it is very likely that such extinction events will not occur and, instead, mutations that buffer the effects of resistance alleles will accumulate (Maisnier-Patin et al. 2002; 2007; Hall & MacLean 2011; Borrell et al. 2013), allowing their maintenance. Our results indicate that different adaptive abilities and the access to strong effect beneficial mutations depends on the genetic background and the competitive context, determining the long-term fate of a given resistance allele.

Antibiotic based treatments focus on immediate clinical results, but the adaptive potential of resistant bacteria is subtle (Read et al. 2011; Stearns 2012) and we suggest it can affect their long-term fate within a host or between hosts. An interesting recent observation suggests that the effects of mutations conferring resistance to streptomycin tend to be smaller for genotypes that are well adapted to a given environment, relative to genotypes not adapted at all (Angst & Hall 2013). This observation, along with the one we have made here, indicates that it is relevant to determine the relative roles of cost versus evolvability in other environments of special clinical relevance (Miskinyte et al. 2013), in order to be better able to predict the evolution of pathogens carrying resistance alleles.

CONCLUSIONS

The frequency of antibiotic resistance constitutes an alarming concern for public health. A key factor determining the extinction or maintenance of resistance alleles is the fitness costs they may entail. High cost resistance alleles are expected to rapidly go extinct. However, this may not be an inescapable fate. If the availability of beneficial mutations is dependent on the genetic background, clones with less fit resistance alleles may also have higher evolvability, *i.e.* a higher potential for adaptation. If so, this will lead to the maintenance of resistances with a higher initial cost in populations. Here we perform competitions between strains of *Escherichia coli*, which carry resistance alleles of different costs, and estimate the relative differences in their adaptive potential. We demonstrate that costly resistance alleles can coexist with resistance alleles of lower cost for hundreds of

generations, suggesting that their adaptive potential can override the initial relative cost of resistance.

FUTURE PERSPECTIVE

Antibiotic resistance poses an ever-increasing danger to public health, and its maintenance is the result of a multitude of processes, which require increasing evaluation. How the costs of resistance depend both on the specific resistant alleles, the environment where the bacteria grow and the ecological context to which they are exposed should lead to a better understanding on how resistance can be reduced or avoided. The ability of resistant bacteria for acquiring compensatory mutations and revert to sensitive state across environments should also be evaluated with the help of increasing powerful genomic technics. This is especially important for bacteria carrying multiple resistances as these are becoming more and more common. Assaying the ability of resistance alleles to emerge and thrive in ecologically relevant contexts, which are very likely to include several resistances competing simultaneously and multiple biotic factors will become crucial for our proper understanding of their long-term pathogenicity.

EXECUTIVE SUMMARY

Long-term evolution of polymorphic antibiotic resistance populations

- In all three pairwise competitions studied, the more costly resistance avoided extinction with high probability.
- In the majority of populations the costly resistance did not sweep to fixation.
- The second most frequent rifampicin resistance mutation to segregate in natural pathogen populations (*H526Y*^{Rif}) shows a high cost but also higher resistance to extinction, across multiple competitive contexts.

Estimation of evolutionary parameters

- The distribution of effects for the beneficial mutations depends on the resistance background, with more costly resistance backgrounds acquiring mutations of stronger effects.
- Differences in rate of acquisition of beneficial mutations were not detected between the resistance backgrounds.

Sequencing and competitive fitness of evolved populations

- Whole genome sequencing of the evolving replicate populations with the *H526Y*^{Rif} and *S531F*^{Rif} competing clones revealed that the *H526Y*^{Rif} background acquired more mutations at higher frequencies pointing towards its increased adaptive potential.
- The number of targets identified for beneficial mutations was not significantly different between the resistance backgrounds, supporting the results from the theoretical analysis.
- Competitive fitness assays showed that the mean effect of beneficial mutations is different between these two resistant backgrounds (*H526Y*^{Rif} and *S531F*^{Rif}), as inferred from theoretical modeling.

Conclusions

- The initial relative difference in fitness costs between resistances is not predictive of their long tern evolution.
- The long-term outcome of competitions between pairs of distinct antibiotic resistance alleles is polymorphism for resistance.
- The results indicate that it is crucial to understand the ecological contexts and the adaptive potential of antibiotic resistance mutations in order to make informed clinical decisions regarding the treatment of bacterial infections.

ACKNOWLEDGEMENTS

We thank Guillaume Martin for the useful comments, from which we considered some, and Hajrabibi Ali for the experimental support. The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) /ERC grant agreement no 260421 – ECOADAPT, and PTDC/BIA-EVF/114622/2009, financed by Fundação para a Ciência e Tecnologia (FCT). I.G. acknowledges the salary support of LAO/ITQB & FCT. J.M.S. acknowledges the scholarship SFRH/BD/89151/2012, from FCT.

REFERENCES

Andersson DI, Hughes D. 2010. Antibiotic resistance and its cost: is it possible to reverse resistance? Nature Reviews Microbiology. 8:260–271. doi: 10.1038/nrmicro2319.

Andersson DI, Levin BR. 1999. The biological cost of antibiotic resistance. Current Opinion in Microbiology. 2:489–493.

Angst DC, Hall AR. 2013. The cost of antibiotic resistance depends on evolutionary history in Escherichia coli. BMC Evol Biol. 13:1–1. doi: 10.1186/1471-2148-13-163.

Baquero MR et al. 2004. Polymorphic Mutation Frequencies in Escherichia coli: Emergence of Weak Mutators in Clinical Isolates. Journal of Bacteriology. 186:5538–5542. doi: 10.1128/JB.186.16.5538-5542.2004.

Barrick JE, Kauth MR, Strelioff CC, Lenski RE. 2010. Escherichia coli rpoB Mutants Have Increased Evolvability in Proportion to Their Fitness Defects. Molecular Biology and Evolution. 27:1338–1347. doi: 10.1093/molbev/msq024.

Björkman J, Nagaev I, Berg OG, Hughes D, Andersson DI. 2000. Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. Science. 287:1479–1482.

Borrell S et al. 2013. Epistasis between antibiotic resistance mutations drives the evolution of extensively drug-resistant tuberculosis. Evolution, Medicine, and

Public Health. 2013:65–74. doi: 10.1093/emph/eot003.

Brandis G, Hughes D. 2013. Genetic characterization of compensatory evolution in strains carrying rpoB Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates. J. Antimicrob. Chemother. 68:2493–2497. doi: 10.1093/jac/dkt224.

Brandis G, Wrande M, Liljas L, Hughes D. 2012. Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. Molecular Microbiology. 85:142–151. doi: 10.1111/j.1365-2958.2012.08099.x.

Chait R, Craney A, Kishony R. 2007. Antibiotic interactions that select against resistance. Nature. 446:668–671. doi: 10.1038/nature05685.

Comas I et al. 2012. Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nature Genetics. 44:106–110. doi: 10.1038/ng.1038.

Couturier M, Desmet L, Thomas R. 1964. High pleiotropy of streptomycin mutations in Escherichia coli. Biochemical and Biophysical Research Communications. 16:244–248.

Cui L et al. 2010. An RpoB mutation confers dual heteroresistance to daptomycin and vancomycin in Staphylococcus aureus. Antimicrobial Agents and Chemotherapy. 54:5222–5233. doi: 10.1128/AAC.00437-10.

Davies J, Davies D. 2010. Origins and Evolution of Antibiotic Resistance. Microbiology and Molecular Biology Reviews. 74:417–433. doi: 10.1128/MMBR.00016-10.

Davis BH, Poon AFY, Whitlock MC. 2009. Compensatory mutations are repeatable and clustered within proteins. Proc. Biol. Sci. 276:1823–1827. doi: 10.1098/rspb.2008.1846.

De Vos M et al. 2013. Putative compensatory mutations in the rpoC gene of rifampin-resistant Mycobacterium tuberculosis are associated with ongoing transmission. Antimicrobial Agents and Chemotherapy. 57:827–832. doi: 10.1128/AAC.01541-12.

Farhat MR et al. 2013. Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. Nature Genetics. 45:1183–1189. doi: 10.1038/ng.2747.

Fisher RA. 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Forsberg KJ et al. 2012. The shared antibiotic resistome of soil bacteria and human pathogens. Science. 337:1107–1111. doi: 10.1126/science.1220761.

Gagneux S et al. 2006. The competitive cost of antibiotic resistance in Mycobacterium tuberculosis. Science. 312:1944–1946. doi: 10.1126/science.1124410.

Gifford DR, MacLean RC. 2013. Evolutionary reversals of antibiotic resistance in experimental populations of Pseudomonas aeruginosa. Evolution. 67:2973–2981. doi: 10.1111/evo.12158.

Hall AR, MacLean RC. 2011. Epistasis buffers the fitness effects of rifampicin-resistance mutations in Pseudomonas aeruginosa. Evolution. 65:2370–2379. doi: 10.1111/j.1558-5646.2011.01302.x.

Handel A, Regoes RR, Antia R. 2006. The role of compensatory mutations in the emergence of drug resistance. PLoS Comput Biol. 2:e137. doi: 10.1371/journal.

Hegreness M. 2006. An Equivalence Principle for the Incorporation of Favorable Mutations in Asexual Populations. Science. 311:1615–1617. doi: 10.1126/science.1122469.

Hermsen R, Deris JB, Hwa T. 2012. On the rapidity of antibiotic resistance evolution facilitated by a concentration gradient. Proc. Natl. Acad. Sci. U.S.A. 109:10775–10780. doi: 10.1073/pnas.1117716109.

Illingworth CJR, Mustonen V. 2012. A method to infer positive selection from marker dynamics in an asexual population. Bioinformatics. 28:831–837. doi: 10.1093/bioinformatics/btr722.

Jansen G, Barbosa C, Schulenburg H. 2014. Experimental evolution as an efficient tool to dissect adaptive paths to antibiotic resistance. Drug Resist. Updat. doi: 10.1016/j.drup.2014.02.002.

Jin DJ, Walter WA, Gross CA. 1988. Characterization of the termination phenotypes of rifampicin-resistant mutants. Journal of Molecular Biology.

Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. 2011. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. Science. 332:1193–1196. doi: 10.1126/science.1203801.

Koch A, Mizrahi V, Warner DF. 2014. The impact of drug resistance on Mycobacterium tuberculosis physiology: what can we learn from rifampicin? Emerg Microbes Infect. 3:e17. doi: 10.1038/emi.2014.17.

Laxminarayan R et al. 2013. Antibiotic resistance-the need for global solutions. Lancet Infect Dis. 13:1057–1098. doi: 10.1016/S1473-3099(13)70318-9.

Lenski RE. 1998. Bacterial evolution and the cost of antibiotic resistance. Int. Microbiol. 1:265–270.

Levin BR, Perrot V, Walker N. 2000. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. Genetics. 154:985–997.

Maisnier-Patin S, Andersson DI. 2004. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. Research in Microbiology. 155:360–369. doi: 10.1016/j.resmic.2004.01.019.

Maisnier-Patin S, Berg OG, Liljas L, Andersson DI. 2002. Compensatory adaptation to the deleterious effect of antibiotic resistance in Salmonella typhimurium. Molecular Microbiology. 46:355–366.

Maisnier-Patin S, Paulander W, Pennhag A, Andersson DI. 2007. Compensatory evolution reveals functional interactions between ribosomal proteins S12, L14 and L19. Journal of Molecular Biology. 366:207–215. doi: 10.1016/j.jmb.2006.11.047.

Mariam DH, Mengistu Y, Hoffner SE, Andersson DI. 2004. Effect of rpoB mutations conferring rifampin resistance on fitness of Mycobacterium tuberculosis. Antimicrobial Agents and Chemotherapy. 48:1289–1294.

Mariam SH, Werngren J, Aronsson J, Hoffner S, Andersson DI. 2011. Dynamics of Antibiotic Resistant Mycobacterium tuberculosis during Long-Term Infection and Antibiotic Treatment Gagneux, S, editor. PLoS ONE. 6:e21147. doi: 10.1371/journal.pone.0021147.s002.

Maughan H, Galeano B, Nicholson WL. 2004. Novel rpoB mutations conferring rifampin resistance on Bacillus subtilis: global effects on growth, competence, sporulation, and germination. Journal of Bacteriology. 186:2481–2486.

Merker M et al. 2013. Whole Genome Sequencing Reveals Complex Evolution Patterns of Multidrug-Resistant Mycobacterium tuberculosis Beijing Strains in Patients Metcalfe, JZ, editor. PLoS ONE. 8:e82551. doi: 10.1371/journal.pone.0082551.s002.

Miskinyte M et al. 2013. The Genetic Basis of Escherichia coli Pathoadaptation to Macrophages Monack, DM, editor. PLoS Pathog. 9:e1003802. doi: 10.1371/journal.ppat.1003802.s018.

Mwangi MM et al. 2007. Tracking the in vivo evolution of multidrug resistance in Staphylococcus aureus by whole-genome sequencing. Proceedings of the National Academy of Sciences. 104:9451–9456. doi: 10.1073/pnas.0609839104.

Mwangi MM et al. 2013. Whole-Genome Sequencing Reveals a Link Between β-Lactam Resistance and Synthetases of the Alarmone (p)ppGpp in Staphylococcus aureus. Microbial Drug Resistance. 19:153–159. doi: 10.1089/mdr.2013.0053.

Nolan CM et al. 1995. Evolution of rifampin resistance in human immunodeficiency virus-associated tuberculosis. Am. J. Respir. Crit. Care Med. 152:1067–1071. doi: 10.1164/ajrccm.152.3.7663785.

Poon A. 2005. The Coupon Collector and the Suppressor Mutation: Estimating the Number of Compensatory Mutations by Maximum Likelihood. Genetics. 170:1323–1332. doi: 10.1534/genetics.104.037259.

Read AF, Day T, Huijben S. 2011. The evolution of drug resistance and the curious orthodoxy of aggressive chemotherapy. Proc. Natl. Acad. Sci. U.S.A. 108 Suppl 2:10871–10877. doi: 10.1073/pnas.1100299108.

Reynolds MG. 2000. Compensatory evolution in rifampin-resistant Escherichia coli. Genetics. 156:1471–1481.

Romero E, Riva S, Berti M, Fietta AM, Silvestri LG. 1973. Pleiotropic effects of a rifampicin-resistant mutation in E. coli. Nature New Biol. 246:225–228.

Schrag SJ, Perrot V, Levin BR. 1997. Adaptation to the fitness costs of antibiotic resistance in Escherichia coli. Proceedings of the Royal Society B: Biological Sciences. 264:1287–1291. doi: 10.1098/rspb.1997.0178.

Silhavy TJ, Berman ML, Enquist LW. 1984. Experiments with gene fusions.

Sniegowski PD, Gerrish PJ. 2010. Beneficial mutations and the dynamics of adaptation in asexual populations. Philosophical Transactions of the Royal Society B: Biological Sciences. 365:1255–1263. doi: 10.1126/science.285.5426.422.

Sousa A, Magalhaes S, Gordo I. 2012. Cost of Antibiotic Resistance and the Geometry of Adaptation. Molecular Biology and Evolution. 29:1417–1428. doi: 10.1093/molbev/msr302.

Stearns SC. 2012. Evolutionary medicine: its scope, interest and potential. Proceedings of the Royal Society B: Biological Sciences. 279:4305–4321. doi: 10.1098/rspb.2012.1326.

Sun G et al. 2012. Dynamic Population Changes in Mycobacterium tuberculosis During Acquisition and Fixation of Drug Resistance in Patients. Journal of Infectious Diseases. 206:1724–1733. doi: 10.1093/infdis/jis601.

Trauner A, Borrell S, Reither K, Gagneux S. 2014. Evolution of Drug Resistance in Tuberculosis: Recent Progress and Implications for Diagnosis and Therapy. Drugs. 74:1063–1072. doi: 10.1007/s40265-014-0248-y.

Trindade S et al. 2009. Positive Epistasis Drives the Acquisition of Multidrug Resistance Zhang, J, editor. PLoS Genet. 5:e1000578. doi: 10.1371/journal.pgen.1000578.s005.

Velayati AA et al. 2009. Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran. Chest. 136:420–425. doi: 10.1378/chest.08-2427.

Wagner GP, Altenberg L. 1996. Perspective: Complex adaptations and the evolution of evolvability. Evolution. 967–976.

Walsh C. 2000. Molecular mechanisms that confer antibacterial drug resistance. Nature.

Watanabe Y, Cui L, Katayama Y, Kozue K, Hiramatsu K. 2011. Impact of rpoB mutations on reduced vancomycin susceptibility in Staphylococcus aureus. J. Clin. Microbiol. 49:2680–2684. doi: 10.1128/JCM.02144-10.

Woods RJ et al. 2011. Second-Order Selection for Evolvability in a Large Escherichia coli Population. Science. 331:1433–1436. doi: 10.1126/science.1198914.

Wrande M, Roth JR, Hughes D. 2008. Accumulation of mutants in ‘aging’ bacterial colonies is due to growth under selection, not stress-induced mutagenesis. Proc. Natl. Acad. Sci. U.S.A. 105:11863–11868. doi: 10.1073/pnas.0804739105.

Wright GD. 2010. The antibiotic resistome. Expert Opin. Drug Discov. 5:779–788. doi: 10.1517/17460441.2010.497535.

Zhang H et al. 2013. Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug

resistance. Nature Genetics. 45:1255–1260. doi: 10.1038/ng.2735.

Zhang Q et al. 2011. Acceleration of Emergence of Bacterial Antibiotic Resistance in Connected Microenvironments. Science. 333:1764–1767. doi: 10.1126/science.1208747.

SUPPLEMENTARY FIGURES AND TABLES

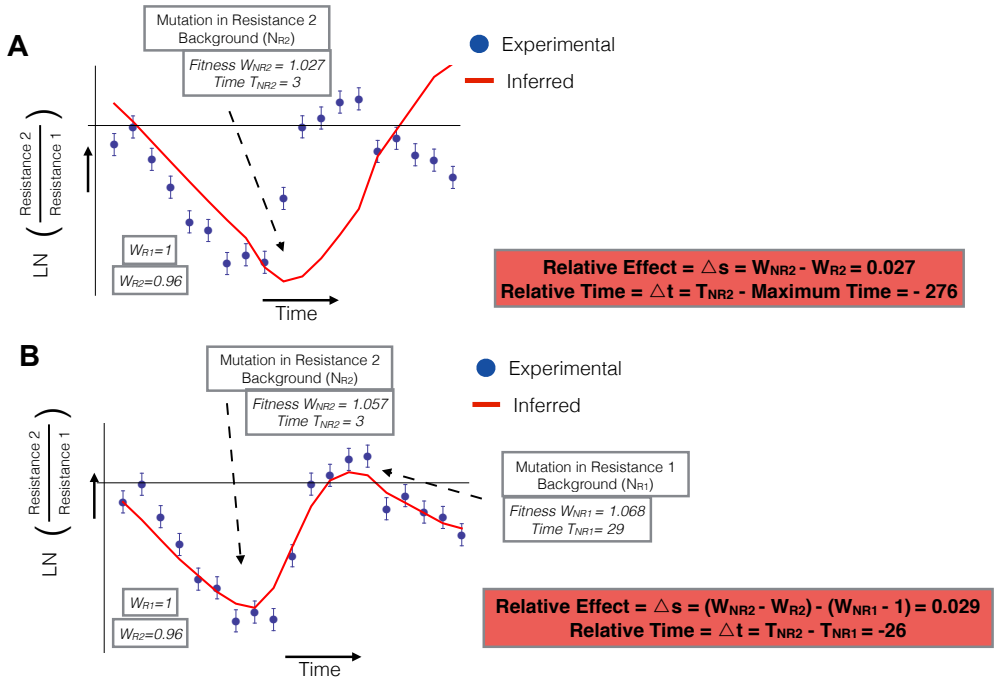


Figure S1. Example of the fitting process for a simulated experimental population. Experimental data, shown as blue circles, was simulated with a given set of parameters. Normally distributed noise (with mean 0 and standard deviation 0.2, shown also as error bars) was added to the experimental points. Red lines show the inferred trajectories given by simple model (one mutations) and the model assuming two mutations, with the fitness of the mutant (WN) and its time of appearance (TN) as parameters. Dotted arrows show the time at which sweeping beneficial mutants effectively change the dynamics of the resistance allele. Shown are also the calculations to obtain the relative parameters used in the distributions of relative effects and time. A) Model 1, with a single mutation. B) Model 2, with a mutation in each background, which has a lower AIC and is, therefore, chosen as the best model to explain the data.

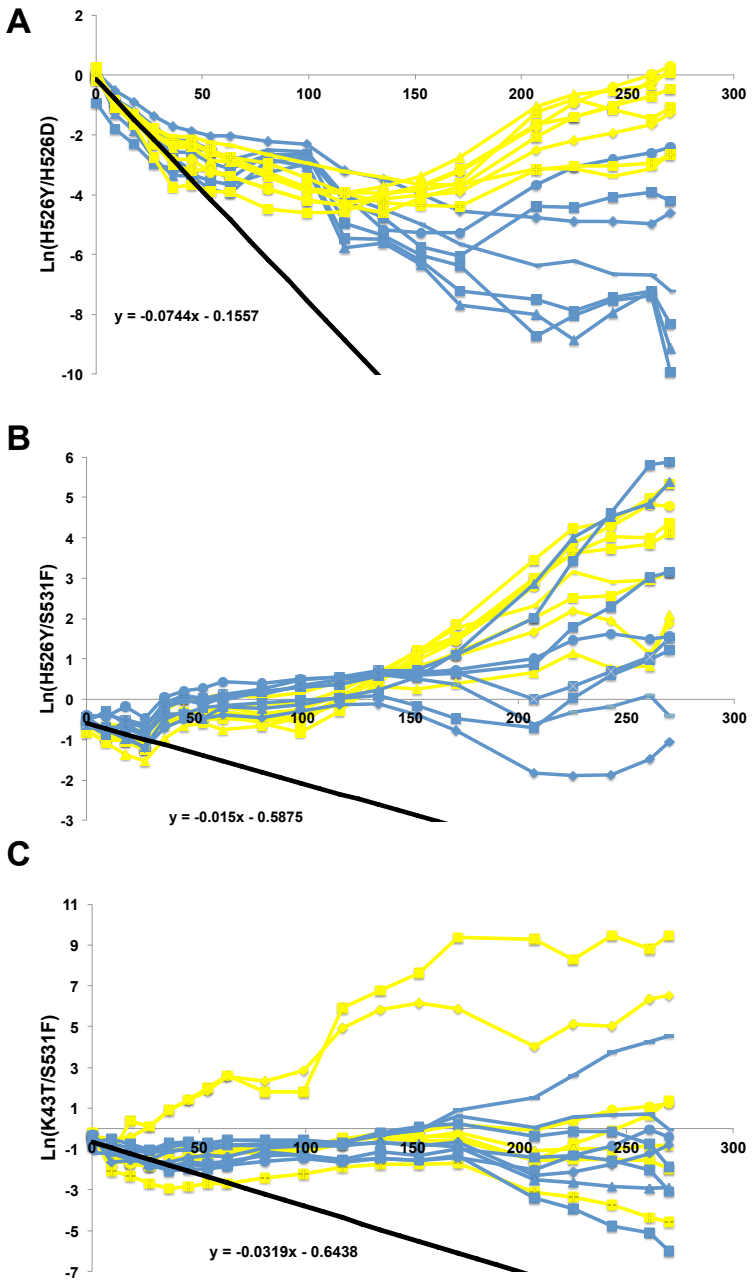


Figure S2. Long-term dynamics with the identification of the fluorescent backgrounds. Dynamics are shown as in Figure 2, but here the color of each line corresponds to the fluorescent marker of the mutation whose logarithm of the ratio is being plotted. **A)** Competitions between resistances *H526Y* and *H526D*. **B)** Competitions between resistances *H526Y* and *S531F*. **C)** Competitions between resistances *K43T* and *S531F*.

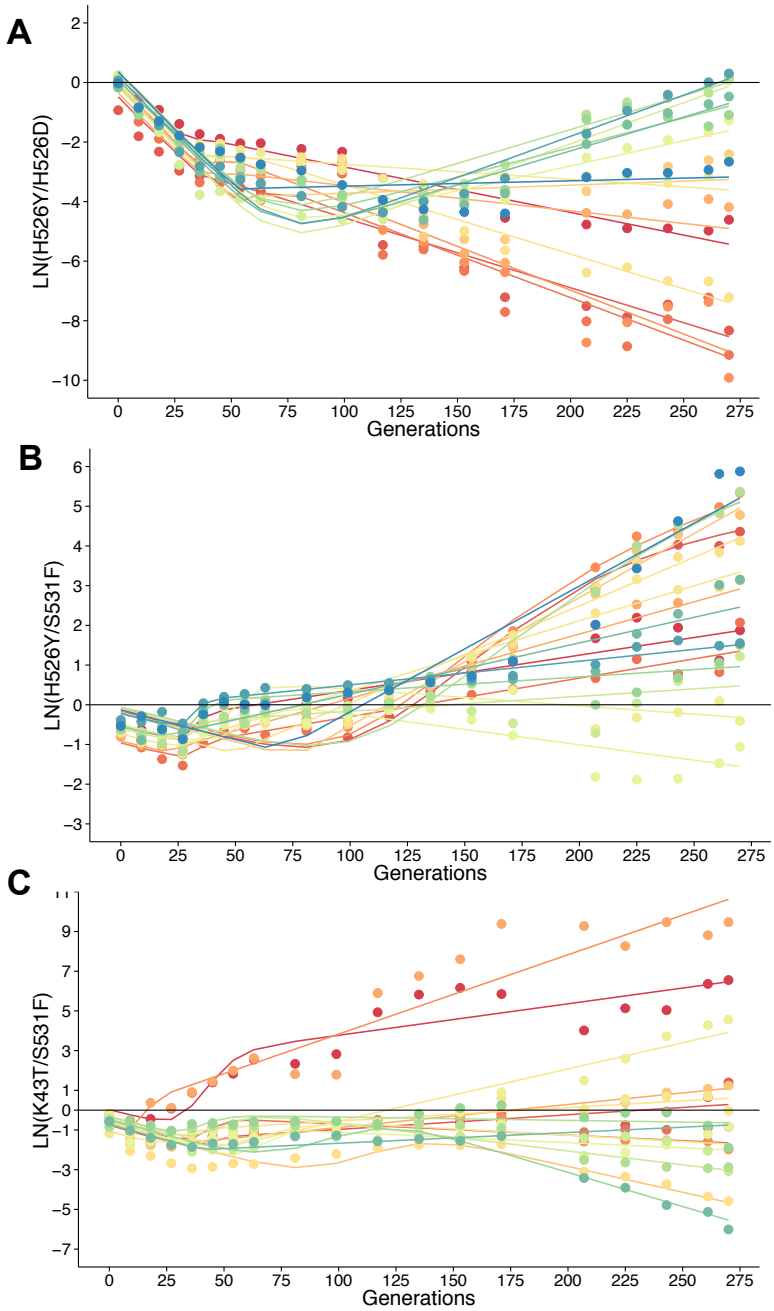


Figure S3. Long-term dynamics from the inferred parameters for each replicate population. Inferred dynamics are shown as solid lines and experimental data is shown as full circles. **A)** Competitions between resistances *H526Y* and *H526D*. **B)** Competitions between resistances *H526Y* and *S531F*. **C)** Competitions between resistances *K43T* and *S531F*.

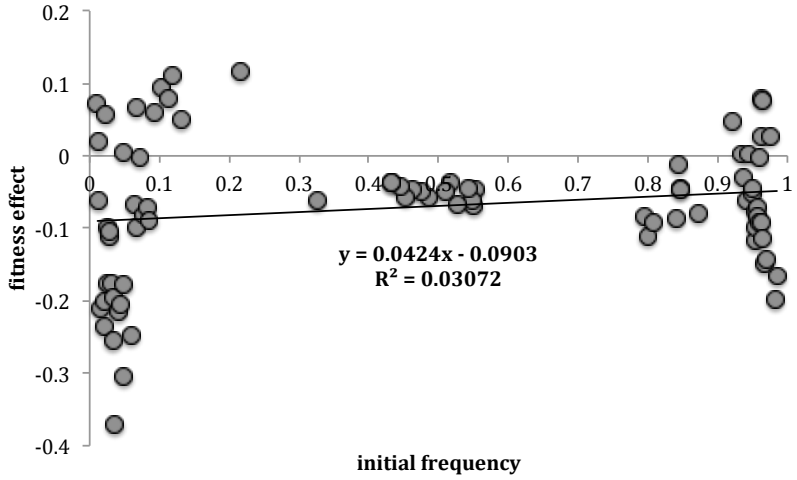


Figure S4. Test for negative frequency dependence selection in competitions between resistance strains *H526Y* and *H526D*. X-axis shows the initial frequency of the *H526Y* strain and Y-axis shows the fitness effect inferred from the slope of the dynamics for the first 24 hours of the competition.

	Antibiotic Resistance	Target Genes	Aminoacid change (location)	Fitness cost (relative to wild-type)	Standard Err
K43T	Streptomycin	*rpsL*	AAA - ACA (128)	**0.092**	0.034
S531F	Rifampicin	*rpoB*	TCC - TTC (1592)	**0.096**	0.011
H526Y	Rifampicin	*rpoB*	CAC - TAC (1376)	**0.073**	0.014
H526D	Rifampicin	*rpoB*	CAC - GAC (1576)	**0.064**	0.017

Table S1. Fitness costs imposed by the antibiotic resistance alleles (*K43T*, *S531F*, *H526Y*, *H526D*) measured in competition against the sensitive reference strain.

Genotype of H526Y Evolved Clones
galK:YFP, rpoB (H526Y), rpoA (T196I), waaZ (S280P)
galK:YFP, rpoB (H526Y), rpoA (T196I)
galK:YFP, rpoB (H526Y), rpoA (T196I)
galK:YFP, rpoB (H526Y), rpoA (T196I)
galK:YFP, rpoB (H526Y), rpoA (T196I)
galK:YFP, rpoB (H526Y), rpoA (T196I)
galK:YFP, rpoB (H526Y), rpoA (T196I)
galK:YFP, rpoB (H526Y), rpoA (T196I)
galK:CFP, rpoB (H526Y), rpoC (Unknown)
galK:CFP, rpoB (H526Y), yifE (G39D)
galK:CFP, rpoB (H526Y), rpoC (R1224C)
galK:CFP, rpoB (H526Y), rpoC (S486P)
galK:CFP, rpoB (H526Y), rpoC (S486P)
galK:CFP, rpoB (H526Y), rpoC (S486P)
galK:CFP, rpoB (H526Y), rpoC (H540P)

Genotype of S531F Evolved Clones
galK:YFP, rpoB (S531F), envR (IS5)
galK:YFP, rpoB (S531F), znuA (F37L), sspA (L156P)
galK:YFP, rpoB (S531F), znuA (F37L), sspA (L156P)
galK:YFP, rpoB (S531F), yeaR (IS186)
galK:YFP, rpoB (S531F), nadK/recN (Intergenic), yedW (IS1)
galK:YFP, rpoB (S531F), rpoC (K50T)
galK:YFP, rpoB (S531F), paoB (V231E)
galK:YFP, rpoB (S531F), nrfG/gltP (Intergenic)

Table S2. Genotypes of the evolved clones from the long–term competition between *H526Y*^{Rif} and *S531F*^{Rif}

CHAPTER VI

Trade-offs of Escherichia coli adaptation to an intracellular lifestyle in macrophages

Manuscript published in ***PLoS One***, in January 2016

The author of this thesis performed the theoretical inference of the adaptive dynamics, the analysis of the whole genome sequencing of the evolved populations and the bioinformatics analysis. Maria Azevedo performed the propagations, competitions, growth curves and metabolic assays. Ana Sousa and João Proença performed the *in vivo* experiments. Ana Sousa also performed the analysis of the whole genome sequencing of the evolved populations.

Trade-offs of *Escherichia coli* adaptation to an intracellular lifestyle in macrophages

Running Title: E. coli intracellular adaptation to macrophages

M. Azevedo[+1], A. Sousa[+1], J. Moura de Sousa[+1], J. A. Thompson[1], J. T. Proença[1] and I. Gordo*[1]

*igordo@igc.gulbenkian.pt

[+] These authors contributed equally to this work

[1]Instituto Gulbenkian de Ciência, Rua da Quinta Grande nº6, Oeiras, Portugal

Abstract

The bacterium *Escherichia coli* exhibits remarkable genomic and phenotypic variation, with some pathogenic strains having evolved to survive and even replicate in the harsh intra-macrophage environment. The rate and effects of mutations that can cause pathoadaptation are key determinants of the pace at which *E. coli* can colonize such niches and become pathogenic. We used experimental evolution to determine the speed and evolutionary paths undertaken by a commensal strain of *E. coli* when adapting to intracellular life. We estimated the acquisition of pathoadaptive mutations at a rate of 10^{-6} per genome per generation, resulting in the fixation of more virulent strains in less than a hundred generations. Whole genome sequencing of independently evolved clones showed that the main targets of intracellular adaptation involved loss of function mutations in genes implicated in the assembly of the lipopolysaccharide core, iron metabolism and di- and tri-peptide transport, namely *rfaI*, *fhuA* and *tppB*, respectively. We found a substantial amount of antagonistic pleiotropy in evolved populations, as well as metabolic trade-offs, commonly found in intracellular bacteria with reduced genome sizes. Overall, the low levels of clonal interference

detected indicate that the first steps of the transition of a commensal *E. coli* into intracellular pathogens are dominated by a few pathoadaptive mutations with very strong effects.

Introduction

Bacterial populations have an enormous potential to adapt to their environments. This is inferred from studies of molecular evolution and variation that find signatures of selection in many genes (Chattopadhyay et al. 2009; Guttman & Dykhuizen 1994). The remarkable pace of bacterial adaptation can also be directly demonstrated in the laboratory by following evolution in real time, over many generations, in controlled environments with specific selection pressures (Barrick et al. 2009; Kawecki et al. 2012; Gordo & Campos 2013). Many studies of microbial evolution in real time involve studying adaptation to simple abiotic environments consisting of single or multiple sugars (Maharjan 2006; Herron & Doebeli 2013), characterizing compensation to the costs of deleterious mutations, such as antibiotic resistance genes in drug free environments (Barrick et al. 2010; Sousa et al. 2012), or studying adaptation in spatially structured environments (Perfeito et al. 2008; van Ditmarsch et al. 2013; Schoustra et al. 2009). Complex environments, in which multiple, more natural, selective pressures are present, have received far less attention (Barroso-Batista et al. 2014). The vast majority of these experiments demonstrate the acquisition of adaptive mutations at high rates, with swift genetic and phenotypic changes. One way to quantify these evolutionary parameters is by following the dynamics of neutral markers in evolving clonal populations, where rapid and large allele frequency changes indicate the occurrence of a high rate adaptive mutations with strong selective effects (Moura de Sousa et al. 2013; Hegreness 2006; Koeppel et al. 2013).

Rapid adaptation is also detected in pathogen populations colonizing humans during infection (Sun et al. 2012). In these natural environments, where

bacteria are likely to encounter many different types of cells, key antagonistic interactions are imposed by the host innate immune system. Overcoming these interactions is often part of the transition from commensalism to pathogenesis (Leimbach et al. 2013; Pérez et al. 2013). Different strains of *E. coli* can be either commensals or versatile pathogens, and even switch between the two, and there is increasing evidence that some pathogenic strains evolved from commensal *E. coli* (Crossman et al. 2010; Tenaillon et al. 2010). Several natural *E. coli* pathovars have been studied, some of which use common mechanisms to increase their virulence. Many of such virulence traits are encoded in pathogenicity islands (blocks of genes found in a pathogen but not in related nonpathogenic strains (Hacker & Kaper 2000; Schmidt & Hensel 2004)), plasmids or prophages, highlighting the importance of successful horizontal gene transfer in pathogen adaptation to new niches. In addition to gene acquisition, gene loss can also contribute to the emergence and diversity of existing *E. coli* pathovars (Maurelli 2007), as well as other genome modifications which may lead to increased bacterial pathogenesis in the absence of horizontal transfer. These are usually called pathoadaptative mutations (Sokurenko et al. 1999). For instance, the knockout of *hemB*, an hemin biosynthetic gene, in *Staphylococcus aureus*, which leads to increased ability to persist intracellularly, constitutes a pathoadaptive mutation and mutations in *hemL* of *E. coli*, encoding glutamate-1-semialdehyde aminotransferase, can also confer pathogenic properties [Ramiro, Costa and Gordo, submitted]. Another common pathoadaptive mutation is the loss of the gene *mucA*, which in *Pseudomonas aeruginosa* increases its ability to evade phagocytosis and resist to pulmonary clearance (Limoli et al. 2014). In another remarkable example, Koli and colleagues (Koli et al. 2011) showed that two genetic changes in commensal *E. coli* K-12 were sufficient to reprogram its cellular transcription and render it invasive in eukaryotic cells, both *in vivo* and *ex vivo*. Macrophages (MFs), one of the major cell types of the innate immune system, are a typical intracellular niche for certain *E. coli* pathovars, including *Shigella*, enteroinvasive *E. coli* (EIEC) and adherent-invasive *E. coli* (AIEC). The former, for instance, is commonly found in patients of Crohn's disease, can adhere to

intestinal epithelial cells and invade and survive in epithelial cells and macrophages (Smith et al. 2013). Characterization of these pathoadaptive mutations is therefore important to understand the emergence of bacterial pathogenesis. We have previously studied the short-term adaptation of *E. coli* to recurrent encounters with macrophages and found that mucoid clones, which carry IS1 insertions into the regulatory region of *yrfF* and that overproduce colanic acid, repeatedly evolved (Miskinyte et al. 2013).

Here, we use experimental evolution to study *E. coli* adaptation to the intra-macrophage environment and to dissect the possible initial adaptive steps for a bacterium to adopt such a lifestyle. We used an established two-marker system to study bacterial adaptation *in vitro* and to determine the rate and fitness effects of pathoadaptive mutations. We then characterized phenotypically the bacteria that evolved and used whole genome sequencing to determine the most likely pathoadaptive evolutionary paths for the first steps in the transition into an intracellular environment.

Materials and Methods

Ethics statement

All experiments involving animals were approved by the Institutional Ethics Committee at the Instituto Gulbenkian de Ciência (project nr. A009/2010 with approval date 2010/10/15), following the Portuguese legislation (PORT 1005/92) which complies with the European Directive 86/609/EEC of the European Council. Endpoints to euthanize the animals were defined prior to the experiment. The specific signs used to make the decision of euthanizing the animals were: weight drop of 20% and/or body temperature decrease below 28ºC (for two consecutive days). Despite the frequent monitoring of the animals' health (at least two times a day), the aforementioned signals were not observed in any of the animals and, therefore, there was no need to perform euthanasia.

Strains and media

The murine macrophage cell line RAW 264.7 (Sigma-Aldrich) was maintained in RPMI 1640-GlutaMAX I (Gibco) supplemented with 1 mM Sodium Pyruvate (Invitrogen), 10 mM HEPES (Invitrogen), 100 U/ml penicillin/streptomycin (Gibco), 50 µM 2-mercaptoethanol solution (Gibco), 50 µg/ml Gentamicin solution (Sigma) and 10% heat-inactivated FBS (standard RPMI complete medium). Culture conditions were at 37ºC in a 5% CO_2 atmosphere.

All bacterial cultures were grown in the same conditions as the macrophage line but using only 100µg/mL of streptomycin (RPMI-Strep medium) instead of the three antibiotics present in RPMI complete medium. The same medium was used for the infection assays of MΦ with bacteria. The *Escherichia coli* strains used were MC4100-YFP and MC4100-CFP (MC4100, galK::CFP/YFP, Amp^R $Strep^R$), which express constitutively either the yellow (yfp) or the cyan (cfp) alleles of GFP integrated at the *galK* locus in MC4100 (*E. coli* Genetic Stock Center #6152) (Bateman & Seed 2012; Hegreness 2006)**.** Unlike certain pathogenic *E. coli* strains, our commensal strain is a derivative of K12 which is not able to replicate within macrophages [27].

Evolution Experiment

The evolution experiment was started from two single colonies of either MC4100-YFP or MC4100-CFP grown in RPMI-Strep in the same conditions as the cell line. The two bacterial cultures were mixed in equal proportion ($5x10^3$ colony forming units (cfu) each) and used to infect the activated MΦ, in 20 replicates.

Before the infection MΦs were centrifuged at 201 g for 5 min and re-suspended in RPMI-Strep. After this step ~ 10^5 cells per well were used to seed a 24-well microtiter plate and incubated over-night at 37ºC with 5% CO_2. Subsequently, activation was done by adding 2 µg/ml of CpG-ODN 1826 (5´TCCATGACGTTCCTGACGTT 3´ - Sigma) (Utaisincharoen et al. 2002) and incubating at 37ºC with 5% CO_2 for 24h. Following activation, cells were washed and infected with 10^4 bacteria mix (multiplicity of infection (MOI) = 1 cfu: 10 MΦs).

After infection we centrifuged the plates at 201 g for 5min (to increase the contact between MΦs and bacteria) and then incubated at 37ºC with 5% CO_2 for 24h (Bateman & Seed 2012). Next we discarded the extracellular bacteria, washed the MΦs with RPMI-Strep two times and added 100µg/mL of Gentamicin solution or 1h at 37ºC with 5% CO_2 (Mittal et al. 2010). Gentamicin penetrates poorly the macrophages and therefore whereas intracellular bacteria are protected from the bactericidal action of the antibiotic the extracellular are killed (Vaudaux & Waldvogel 1979). After washing out the gentamicin with PBS 1X, cells were lysed using a 0.1% Triton-X – PBS solution for 15 minutes (Bokil et al. 2011). Intracellular bacteria were collected, washed with PBS 1X and counted by flow cytometry using LSR Fortessa cytometer (BD Biosciences). From approximately 10^6 intracellular bacteria collected, we pooled 10^4 and infected a new batch of activated MΦs, in the same manner as described previously. This procedure was repeated for 26 days, a period after which fixation of one of the fluorescent markers could be observed for most of the replicate experiments, an indication of adaptation. This propagation protocol allows ~ 7 generations per day, calculated by Log_2 (Nf/Ni), where Nf is the number of intracellular bacteria 24h post-infection, and Ni is the bacterial inoculum used to infect the macrophages (Miskinyte et al. 2013).

Fitness measurements

Fitness increases of the evolved populations were estimated by competitive fitness assays in the presence or in the absence of MΦs. A sample of 30 clones carrying the fluorescence marker which achieved the highest frequency in a given population was competed against the ancestral strain labeled with a different marker. These samples of clones were assumed to be representative of the population. The competition assays for each evolved population were done in triplicate in the same conditions as the evolution experiment, for two passages - 48h. The neutrality of the fluorescent marker was tested by competition of the two ancestral strains (9 replicates). Relative fitness, expressed as a selection coefficient, was estimated by calculating the slope of the natural logarithm of the

ratio of evolved over ancestral bacteria per generation of ancestral bacteria (Miskinyte & Gordo 2014).

Whole genome re-sequencing and mutation prediction

Ancestral genome

The sequence reads were mapped to the reference strain *Escherichia coli* K12 MG1655 BW2952 (reference NC_012759.1). The two ancestors carry mutations in relation to the reference (Miskinyte et al. 2013). The sequenced evolved clones were then compared to the ancestral genome and the mutations identified are represented in **Fig. 3** and listed in **Table 2**.

Clone analysis

In the last time point of the evolution experiment, we isolated a clone from each evolved population carrying the fluorescent marker with higher frequency. In the populations where both markers reached similar frequencies at the last time point, one clone from each marker subpopulation was isolated. Each of these clones was then grown in 10mL of RPMI at 37ºC. DNA isolation from these cultures was subsequently obtained according to (Wilson 2001).

The DNA library construction and sequencing was carried out by the in-house genomics facility. Each sample was paired-end sequenced using an Illumina MiSeq Benchtop Sequencer. Standard procedures produced data sets of Illumina paired-end 250bp read pairs. Genome sequencing data have been deposited in the NCBI Read Archive http://www.ncbi.nlm.nih.gov/sra (accession no. SRP066892). The mean coverage per sample was ~35x. Mutations were identified using the BRESEQ pipeline (Barrick et al. 2009). To detect potential duplication events we used ssaha2 (Ning et al. 2001) with the paired-end information. This is a stringent analysis that maps reads only to their unique match (with less than 3 mismatches) on the reference genome. Sequence coverage along the genome was assessed with a 250 bp window and corrected for GC% composition by normalizing by the mean coverage of regions with the same GC%.

We then looked for regions with high differences (>1.4) in coverage. Large deletions were identified based on the absence of coverage. For additional verification of mutations predicted by BRESEQ, we also used the software IGV (version 2.1) (Robinson et al. 2011).

Phenotypic characterization of evolved clones

Growth in single carbon sources

The same samples of clones from the populations which were tested in the competition assays were used to estimate the growth curves in different carbon sources. Two media were used: M9 Minimal Media (MM) supplemented with maltose 0.4% or with glucose 0.4%. The growth curve assays were performed on a Bioscreen C microplate reader, using a volume of 150µL per sample and an inoculum of ~10^4 CFUs. Plates were incubated at 37ºC with shaking before each optical density measurement (OD at 600nm). All growth measurements were repeated at least twice.

Fitness of effect of *fhuA* mutant under oxidative stress

To test if the mutation on the *fhuA* gene conferred some advantage to the evolved bacteria in specific selective pressures characteristic of the macrophage intracellular environment, we grew ancestral and mutant clones under oxidative and iron limitation stresses. We combined different concentrations of Fe^{3+} (Iron (III) Chloride hexahydrate, Alfa Aesar #A16231) with the ferrichrome siderophore (Ferrichrome Iron-free, Santa Cruz Biotechnology # sc-255174) and added hydrogen peroxide (H_2O_2) (Hydrogen Peroxide solution 30% (w/w), Sigma # H1009). Ferrichrome captures iron III and the resulting complex is imported into the cell by the FhuA outer membrane transporter. Excess iron inside the cell may be detrimental in the presence of H_2O_2, due to the Fenton reaction. In the KO mutant of fhuA, ferrichrome-dependent uptake of iron does not occur, which could provide an advantage to the bacteria when exposed to oxidative stress. The mutant used for this experiment was the sequenced clone of population C (*fhuA* KO and *selC* IS), which was compared to an ancestral clone. The two clones were

first grown in M9 Minimal Media supplemented with 0.4% Glycerol in an orbital shaker at 37ºC with 230rpm, to an OD_{600nm} of 1 (stationary phase). The cultures were then diluted and grown again in the same conditions until they reached an OD_{600nm} of 0.4-0.6 (exponential phase). After normalization to the same O.D, samples were diluted 100x, divided in equal parts and centrifuged at 3220 g for 30 minutes, before being re-suspended in the same growth media either supplemented, or not, with Fe^{3+} and Ferrichrome at two different concentrations: 0μM Fe^{3+} + 100μM Ferrichrome and 100μM Fe^{3+} + 1000μM Ferrichrome. Samples were acclimatized at 37ºC with agitation for ~15 minutes before the addition of H_2O_2 to a final concentration of 2 mM. Samples were then left at 37ºC without agitation and collected after 1h, washed in PBS 1X and plated on LB agar. Plates were incubated for 16h at 37ºC, followed by CFU counting.

Analysis of *rfaI* conservation in other *E. coli* strains

A list of all sequenced strains of *E. coli* was retrieved from the European Bioinformatics Institute database (www.ebi.ac-uk/genomes, accessed on April 2014). The meta-information for all the strains (*i.e.*, laboratory origin, pathogen or commensal) was manually curated by accessing several different public microbial databases. The fasta sequences were retrieved for each of the genes comprising the *rfa* locus in *Escherichia coli* BW2952 (*rfaBCDFGIJLPQSYZ* and *waaAU*) and then BLASTed against the sequenced genomes of the genus *Escherichia* and *Shigella* (74 genomes in total), using Biopython. If, for a given strain, the query was returned as empty, we considered the gene to be absent. Otherwise, the gene was present but with varying degree of conservation, although not below 82% similarity.

***In vivo* test for increased pathogenesis**

C57/BL6 mice, aged 7-10 weeks (in house supplier, Instituto Gulbenkian de Ciência) were given food (RM3A(P); Special Diet Services, UK) and water *ad libitum*, and maintained with a 12 hour light cycle at 21ºC. The animals were infected intra-peritoneally with $2x10^7$ CFUs of either the ancestral clone or evolved

clone I (carrying two IS insertions, Fig. 3) diluted in 100µl of PBS. Furthermore, as a control, in each experimental block we injected a group of 2–3 mice with 100 ml of saline (these animals did not display any signs of disease). Mice were followed for 4 days post infection and their weights and temperatures were monitored daily. The infections were performed in two blocks, with n=3 mice per bacterial strain per block. A linear mixed effect model, with bacterial strain and day post-infection as factors and mouse as a random effect, was used to determine if significant increases in weight loss occurred in an infection with the evolved clone.

Results and Discussion

Dynamics of *E. coli* adaptation to intracellular life

We followed the evolutionary dynamics and adaptation of twenty independent populations of *E. coli* during repeated exposure to the intracellular environment of MFs. The bacterial populations were all founded from an equal mix of two ancestral clones, which were isogenic except for a distinct neutral fluorescent marker. Under the hypothesis that periodic selection will dominate the pathoadaptive process, the occurrence and spread of a strong beneficial mutation in one of the clones with a given fluorescent marker will cause the extinction of all other clones and hence the loss of diversity at the marker locus (Barroso-Batista et al. 2014; Atwood et al. 1951). A more complex pattern may emerge if adaptive mutations are very common and cause clonal interference (Hegreness 2006), which may slow the loss of neutral variation (Barroso-Batista et al. 2014), or if coexisting interdependent ecotypes emerge (Maharjan et al. 2012; Koeppel et al. 2013).

In our experimental evolution protocol, MFs (10^5/ml) were infected with *E. coli* for 24 hours, after which all extracellular bacteria were killed with gentamicin. 10^4 bacteria sampled from the intracellular environment of Mfs were then used to infect new uninfected MFs. The evolution experiment was followed for 26 days and the occurrence of adaptive mutations was detected through the observation of

rapid and consistent changes in the frequency of the neutral marker (**Fig. 1A** and **1B**). After 10 days of propagation, consistent changes in frequency started to be detected in some populations and by day 15 most of the populations showed significant deviations from the initial marker frequency (15 out of 20 populations showed deviations above 10%), suggesting that beneficial mutations had spread through the populations (**Fig. 1A** and **1B**). During the 26 days of evolution, in only one of the populations (O) the deviation from the initial marker frequency was less than 10%. A significant increase in the total number of bacteria after infection was also detected after 100 generations in all the lines evolved (**Fig. 1C** and **1D** and **Supplementary Table S1**). The increase in carrying capacity (K) of the evolving populations tends to be observed in synchronicity with the changes in the marker frequency, indicating that this fitness trait is being modified by occurring adaptive mutations.

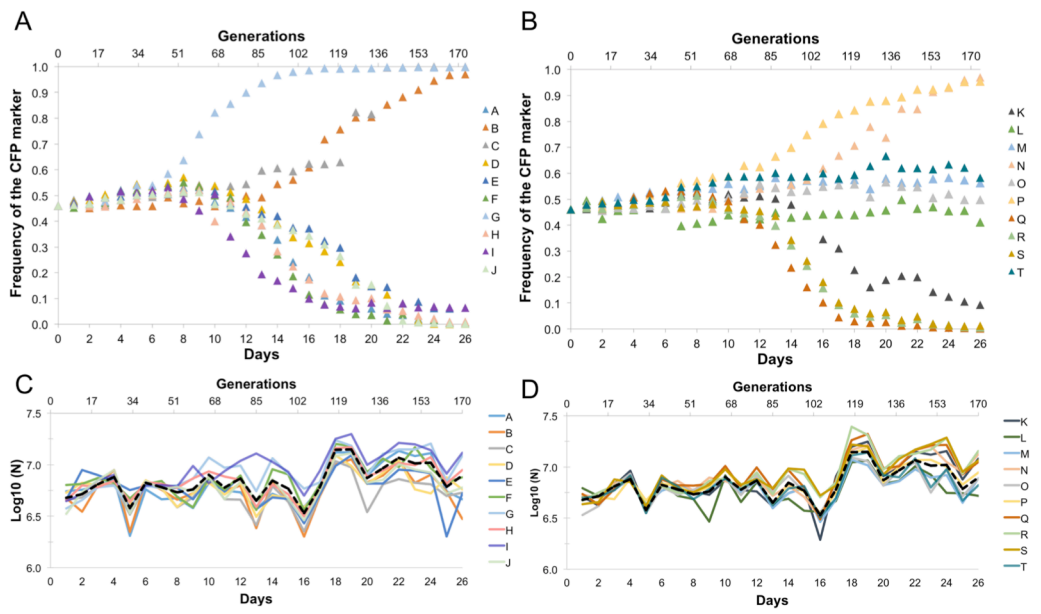


Fig. 1. Evolutionary dynamics of the populations evolved within macrophages. Dynamics of frequency of neutral marker in the 20 replicate evolving lines (in (A) lines A to J; in (B) lines K to T) and variation in population size along the evolution experiment (in (C) lines A to J; in (D) lines K to T).

Pathoadaptation occurs at a high rate and involves strong effect mutations

The rapid and consistent changes in the frequency of each of the fluorescent alleles imply the occurrence of strong beneficial mutations. Assuming a simple model of positive selection we can estimate their rate and effect through the deviations of the neutral markers (Hegreness 2006; Illingworth & Mustonen 2012). We have estimated these key evolutionary parameters using two different approaches: Marker Divergence Analysis (Hegreness 2006; Barrick et al. 2010), which summarizes the neutral marker dynamics using two parameters: the effective mutation rate (U_e) and the effective selection coefficient (S_e), by fitting simulations to the marker dynamics. This method, which assumes all mutations generated within a replicate population to have a given fixed effect, has been shown to perform acceptably for scenarios of low clonal interference (Moura de Sousa et al. 2013) and summarizes the adaptive dynamics of all the populations by a single value of U_e and S_e. The second method, Optimist (Illingworth & Mustonen 2012), determines the likelihood that the frequency of a neutral marker results from a given number of haplotypes, arising at a given time and segregating with a particular effect. For each particular replicate population, the number of haplotypes that best explains the marker frequency dynamics is chosen by the lowest Akaike Information Criteria (AIC), resulting in a distribution of the number of haplotypes, as well as their effects, for all the replicate populations.

From the dynamics in **Fig. 1A** and **1B**, the best estimates of U_e and S_e were 1.6×10^{-6} (mutations per genome per generation) and 0.26, respectively. Using the method implemented in Optimist, we find a mean increase in fitness of mutations of 0.09 (see **Table 1** for the estimated parameters and **Supplementary Fig. S1** for the corresponding simulated dynamics that best fit the experimental data). These estimates of the rate and strength of fitness effects of adaptive mutations can be compared with those obtained in bacterial adaptation to other environments, and using similar methods of inference. *E. coli* rates of adaptation to compensate for the costs of antibiotic resistance were found to lay in the range of 10^{-7}, and mean *s* in the range of 5 to 15%, dependent on the strain that evolved

(Barrick et al. 2010; Illingworth & Mustonen 2012). Using a different experimental system with more neutral markers, (Sousa et al. 2012) estimated higher rates of compensatory mutations to resistance $U \sim 10^{-5}$, with mean effects of 2.5 and 3.6% dependent on the resistance mutation. It is becoming well established that the distribution of effects of adaptive mutations markedly depends on the genetic background. For *E. coli* strains with the same genetic background as the ones used here, but adapting to a simpler environment (Luria-Bertani rich medium) Hegreness et al (Hegreness 2006) found $U_e = 2 \times 10^{-7}$ and $S_e = 0.05$. These estimates are considerably smaller than the estimates found here, when the strain faces harsher conditions. Since the same strain and the same method of estimation were used in our experiment, the comparison of the combined estimates demonstrates that the evolutionary parameters strongly depend on the environment. They furthermore support the idea that in more stressful environments, where strong biotic interactions prevail, higher rates and effects of adaptive mutations are to be expected (Dobzhansky 1950).

Population	# of	W	Time	W	Time
A	2	0.101	28	0.073	28
B	1	0.054	0		
C	1	0.109	63		
D	1	0.063	14		
E	1	0.056	7		
F	2	0.091	21	0.134	35
G	2	0.123	7	0.099	14
H	2	0.134	42	0.109	42
I	2	0.090	7	0.094	42
J	1	0.071	28		
K	2	0.141	63	0.121	63
L	0				
M	1	0.027	0		
N	1	0.053	0		
O	2	0.054	14	0.063	35
P	2	0.073	0	0.060	7
Q	2	0.112	35	0.111	77
R	2	0.101	35	0.103	77
S	2	0.151	56	0.124	63
T	2	0.043	7	0.089	98

Table 1. Inferred selective effects of beneficial haplotypes. The number of mutations inferred for a specific population is indicated in the 2nd column. W mut#1 and T mut#1 (3rd and 4th columns) indicate, respectively, the inferred fitness improvement and time of appearance (in generations) of the first mutant. W mut#2 and T mut#2 (5th and 6th columns) indicate the same inferred parameters for the second mutant. Shaded cells indicate a mutation inferred in the CFP background.

Competitive fitness assays reveal two distinct strategies of pathoadaptation

The changes in frequency of each fluorescent allele suggested a strong effect of the beneficial mutations that occurred. To support this inference and directly estimate the strength of these mutations, we performed competitive fitness assays, in the presence of macrophages, of evolved clones against the ancestral strain marked with a different fluorescence. **Fig. 2** shows that all populations

exhibit a significant fitness increase and are therefore better adapted to the environment with macrophages. The mean competitive fitness increase observed was 7%, with a minimum of 5% and a maximum of 12% (**Fig. 2**, blue bars). These values are in close agreement with those estimated from the changes in marker frequency alone and assuming the simplest model of positive selection (mean of 9% with a minimum of 4% and a maximum of 15% (see **Table 1** and **Supplementary Table 2**). Although there is a slight overestimation of the fitness effects inferred by the marker deviations, they can be explained by a number of reasons. Firstly, and contrary to what is assumed by the model, selection in such complex environments might not be constant, leading to non-linear effects of beneficial mutations. Secondly, theoretical approaches are known to overestimate the effects of mutations when there is more than one mutation (i.e., cases of higher clonal interference) (Illingworth & Mustonen 2012; Moura de Sousa et al. 2013). Finally, the AIC criteria (see methods) might be too stringent in selecting models that postulate an increased number of haplotypes, which will lead to stronger effect mutations. Nevertheless, both fitness measures are in agreement that the most likely form of selection taking place in this environment involves sweeps of beneficial mutations of strong effects.

One possible trait that could be expected to evolve as an adaptation to the selective pressure imposed in this experiment would be an increased ability to grow in the abiotic environment external to the macrophages (RPMI). If a variant with increased fitness in RPMI would emerge, then its frequency outside macrophages could increase and dominate the population; a likely scenario if such a mutant did not have any cost inside macrophages nor in the external environment as it becomes conditioned by those cells. To determine whether the evolved populations increased in fitness in RPMI, *i.e*. in the absence of macrophages, we performed competitive fitness assays against the ancestor in the medium alone (**Fig. 2**, orange bars). The results show that in 3 out of 14 populations there is, indeed, a significant fitness increase in the abiotic environment, suggesting that increasing growth in RPMI can be beneficial in the presence of macrophages. We note that during the evolution the abiotic

environment outside macrophages is likely to change, so a mutant which is beneficial in RPMI may change its advantage as this medium becomes conditioned by the presence of macrophages. The results also show a correlation between the changes in fitness in the absence of macrophages to the increase in fitness in their presence (Pearson r=0.688, P=0.0065, **Fig. 2**). In half of the populations (A,F,G,K,N, R and S) a clear trade-off was detected (**Fig. 2**). For these cases, accumulation of mutations with significant advantage in the presence of macrophages led also to a decreased competitive ability in their absence. This indicates a specialization in the transition to intracellular life. Together, the results suggest different adaptive strategies adopted by similar bacteria adapting independently to the same environment, but with distinct genetic mechanisms evolved to cope with the same antagonistic interaction.

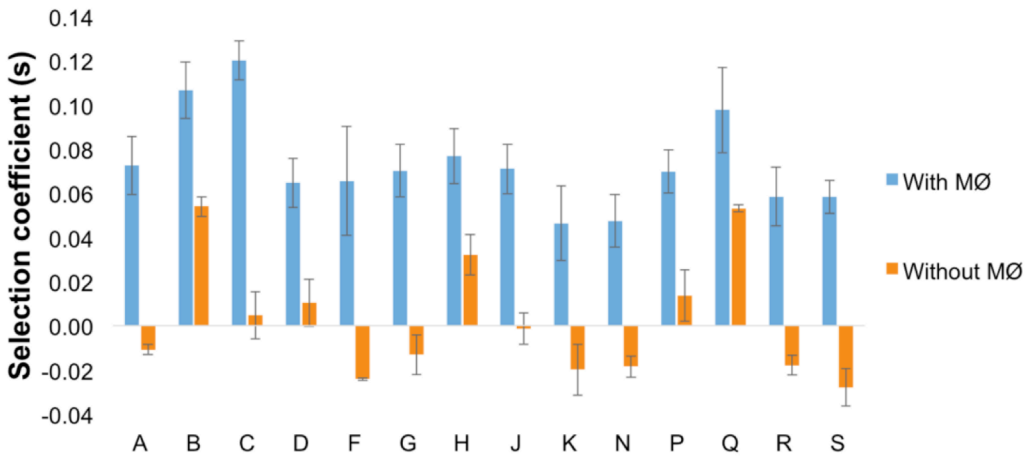


Fig. 2. Fitness measures of evolved clones. In blue, the competitive fitness of evolved populations in the presence of MΦ: evolved clones (a sample of 30 from each indicated population) were competed against the ancestral clone (1:1). In orange, competitive fitness assay in the absence of MΦ. Error bars correspond to 2SE.

Genetic basis of the intracellular adaptation reveals common evolutionary paths

Given the dynamics of neutral markers observed (see **Fig. 1A** and **1B**), the short duration of the experiment (~175 generations) and the estimates of only a few of beneficial mutations being responsible for the adaptive process (see **Table 2**), we predict that each population is dominated by a single clone with one or two mutations. In order to unravel the number of genetic changes that occurred and to reveal the underlying evolutionary paths taken by the populations, we performed whole genome sequencing of independently evolved clones. The evolved strains and their ancestor were sequenced to a minimum of 16x coverage on the Illumina Miseq platform. **Table 2** shows the genetic changes detected and **Fig. 3** their position along the chromosome. Overall, 25 different mutational targets were detected amongst the adapted clones. As expected, each clone carries an average of 2 mutations. Most of the mutations occurred in coding regions and 14 out of 34 in total involved insertions of transposable elements IS1, IS5 and IS186. The first two have been found to transpose at higher rates than other elements (Sousa et al. 2013) and are therefore more likely to contribute to adaptation. Among the gene targets for the mutations detected, two occurred in 4 and 8 clones (*fhuA* and *rfaI*, respectively) and one occurred in two independently evolved clones (*tppB*). Parallelism is a hallmark of adaptation since the probability that mutations in the same gene increase in frequency by random chance in at least two independent lines, over such a short period, is very low (Lang & Desai 2014; Adams & Rosenzweig 2014). Given the parallelism observed involving the gene targets *rfaI* and *fhuA*, we can safely assume that these changes are adaptive. Furthermore, the change hitting the coding region of *tppB* (either through an insertion or by a small deletion) in two independent clones, together with it being the sole detected mutation (B CFP and Q YFP) suggests that loss-of-function of *tppB*, coding for a proton-dependent transporter of di- and tri-peptides could be an important pathoadaptive mutation.

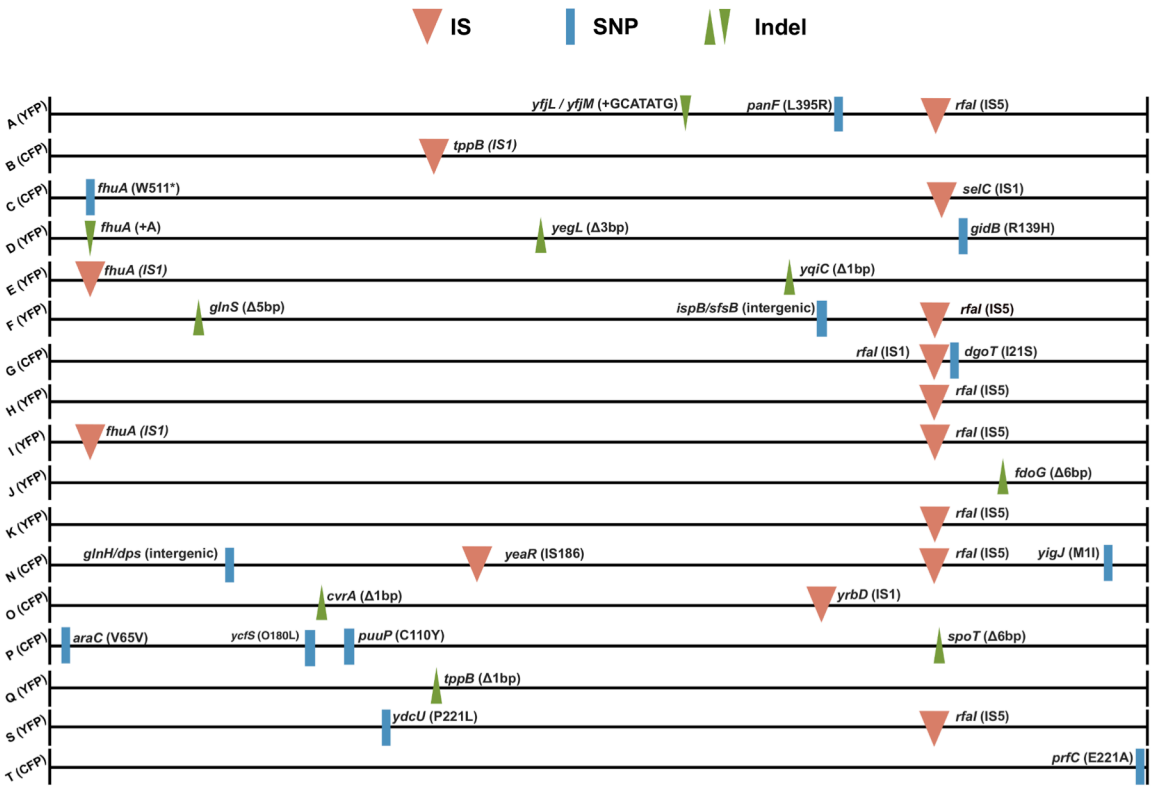


Fig. 3. Genomic maps of the mutations detected in the sequenced evolved clones. Red inverted triangles represent insertions of IS elements, blue lines mark single nucleotide polymorphisms and green triangles denote small insertions (pointing downwards) or deletions (pointing upwards). All events have either the aminoacid changes associated (for SNPs), the IS element inserted or the small number of base pairs deleted or inserted.

Clone (Coverage)	Genome Position	Gene	Mutation	Annotation	S	Tradeoff without Mfs
A YFP (36x)	2649245	*yfjL/yfjM*	+GCACTATG	intergenic (-258/+102 nt)	7%	Yes
	3293960	*panF*	T→G	L395R (CTG→CGG)		
	3689096	*rfaI*	IS5 + 4bp	coding (312/1020 nt)		
B CFP (35x)	1603229	*tppB*	IS1 + 5 bp	coding (378/1503 nt)	11%	No
C CFP (73x)	169014	*fhuA*	G→A	W511* (TGG→TAG)	12%	No
	3722624	*selC*	IS1 + 8bp	non coding (47/95 nt)		
D YFP (20x)	168462	*fhuA*	+A	coding (980/2244 nt)	6%	No
	2039547	*yegI*	Δ3 bp	coding (1944-1946/1947 nt)		
	3809621	*gidB*	C→T	R139H (CGC→CAC)		
E YFP (93x)	3070100	*yqiC*	Δ1 bp	coding (91/291 nt)		
	168929	*fhuA*	IS1 + 8bp	coding (1447/2244 nt)		
F YFP (16x)	3219882	*ispB/sfs*	G→A	intergenic (131/-197 nt)	7%	Yes
	608471	*glnS*	Δ5 bp	coding (396-400/1665 nt)		
	3689096	*rfaI*	IS5 + 4bp	coding (312/1020 nt)		
G CFP (40x)	3758025	*dgoT*	A→C	I21S (ATC→AGC)	7%	Yes
	3689065	*rfaI*	IS1 + 8bp	coding (343/1020 nt)		
H YFP (20x)	2541599	*hscA*	Δ3 bp	coding	8%	No
	3689096	*rfaI*	IS5 + 4bp	coding (312/1020 nt)		
I YFP (148x)	167493	*fhuA*	IS1 + 8bp	coding (11/2244 nt)		
	3689096	*rfaI*	IS5 + 4bp	coding (312/1020 nt)		
J YFP (346x)	3971199	*fdoG*	Δ6 bp	coding (2311-2316/3051 nt)	7%	No
K YFP (28x)	3689096	*rfaI*	IS5 + 4bp	coding (312/1020 nt)	5%	Yes
N CFP (34x)	75049	*glnH/dps*	A→T	intergenic (-295/+109 nt)	5%	Yes
	4410884	*yjgJ*	G→A	M1I (ATG→ATA)		
	1769910	*yeaR*	IS186 + 10bp	coding (122/360 nt)		
	3688872	*rfaI*	IS5 + 3bp	coding (536/1020 nt)		
O CFP (35x)	1129729	*cvrA*	Δ1 bp	coding (296/1737 nt)		
	3223632	*yrbD*	IS1 + 11bp	coding (9/522 nt) / intergenic		
P CFP (29x)	70580	*araC*	T→C	V65V (GTT→GTC)	7%	No
	1072984	*ycfS*	G→A	O180L (CCG→CTG)		
	1247732	*puuP*	C→T	C110Y (TGT→TAT)		
	3708996	*spoT*	Δ6 bp	coding (247-252/2115 nt)		
Q YFP (34x)	1602868	*tppB*	Δ1 bp	coding (17/1503 nt)	10%	No
S YFP (29x)	1404545	*ydcU*	C→T	P221L (CCG→CTG)	6%	Yes
	3689096	*rfaI*	IS5 + 4bp	coding (312/1020 nt)		
T CFP (30x)	4546582	*prfC*	A→C	E221A (GAA→GCA)		

Table 2. Mutations identified in the sequenced clones. Coverage for each clone is indicated in the first column. The 6th column (S) indicates the selective effect of the evolved clones in the presence of macrophages, compared to the ancestor, and the 7th column indicates whether there is a selective tradeoff (i.e., fitness in the absence of macrophages is lower than ancestor).

Loss of *rfaI* leads to a strong selective sweep during adaptation to macrophages

In 47% of the evolved populations, mutations in *rfaI* (all of them IS insertions presumably leading to gene inactivation) were detected, suggesting this to be a preferential target and, therefore, one with high beneficial effect. We followed the emergence of this adaptive mutation in one of the adapted populations (population I), by targeted PCR for the presence of IS5 element in *rfaI*, as this element had been identified in the evolved clone sequenced from this population. **Fig. 4** shows that the mutation could be detected by day 6, with a frequency 4.1% (SE 0.04) and rapidly swept to fixation, being detected in all tested clones (n=60) at day 26. We could directly estimate its selective effect, from its initial change in frequency, to be 0.09 (see inset of **Fig. 4**).

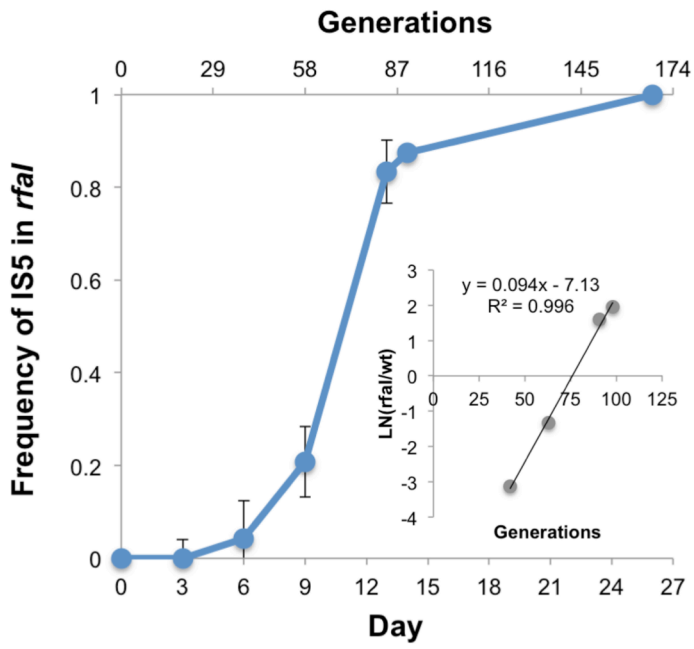


Fig. 4. Invasion of an adaptive mutation at the *rfaI* locus. The blue line shows the selective sweep leading to the fixation of an IS5 insertion likely causing gene loss of function. In the inset an estimate of the selection coefficient (*s*) of this mutation is given. *s* is the slope of the logarithm of the ratio between the number of clones carrying the mutation and those with the wild-type allele.

We also found a strong correlation between the presence in a clone of a mutation in *rfaI* and a competitive tradeoff (i.e., benefit in the presence of macrophages, but disadvantage in their absence) of the populations where that clone mutation emerged ($p<0.01$, Pearson correlation). *rfaI* is a glycosyltransferase and part of the lipo-polysaccharide (LPS) synthesis machinery present in bacteria. LPS are unique and complex glycolipids that provide characteristic components in the outer membranes of bacteria and as such are a critical component of their interaction with cells from the immune system (Beutler & Rietschel 2003). The *rfa* locus itself is composed of 15 different genes, which are responsible for generating different parts of the LPS structure (Schnaitman & Klena 1993). *rfaI* is involved in the outer part of the core oligosaccharide, connecting the lipid A (inner part) and the O-antigen (outer part) of LPS. Since the strain used in this study is devoid of O-antigen, the outer part of the core is the LPS terminal section, and it is likely acting as one of the main interfaces between the bacterial cell and the cells from the immune system. Modifications in the LPS structure, and the outer core in particular, are known to modify the behavior of bacterial cells regarding adhesion to epithelial cells and biofilm formation in enterohemorrhagic *E. coli* (Torres & Kaper 2003; Torres et al. 2005), and intracellular invasion of different serovars of *Salmonella enterica* (Hoare et al. 2006). Moreover, mutations in the outer core structure of *Brucella abortus* can induce pro-inflammatory responses and enhanced macrophage activation (Conde-Álvarez et al. 2012). Interestingly, many genes in the *rfa* locus itself are a target for bacterial persistence in *E. coli* (Girgis et al. 2012), and the operon seems to be poorly conserved in a vast group of *E. coli* pathovars, with several pathogenic (and non-pathogenic) strains missing many of its genes (including *rfaI*) (see Analysis of *rfaI* conservation in other *E. coli* strains (section) in Material and Methods). This suggests that, in these strains, the *rfa* genes could be common mutational targets. Together, both our results and these observations seem to indicate an important role of the LPS structure both in the interaction with the immune system and in the transition to a pathogenic lifestyle, implying the changes in *rfaI* detected in 8 independently evolving populations as a recurrent pathoadaptive target.

fhuA pathoadaptive mutant is beneficial under the combined pressure of iron limitation and oxidative stress

The beta barrel protein FhuA is involved in the active transport of ferric siderophores across the outer membrane of Gram-negative bacteria (Ferguson et al. 1998). Iron homeostasis is crucial to the lives of both bacteria and macrophages therefore both cells have exquisite mechanisms to achieve physiological levels of iron and to keep it in a safe intracellular non-toxic form. Although oxidative stress can be generated by aerobic respiration, it is also one of the microbicidal pressures generated by the macrophages in the harsh phagosomal environment. Superoxide and hydrogen peroxide (O_2^- and H_2O_2) are moderately reactive oxygen species, however, upon interaction with iron, the highly reactive hydroxyl radical (OH^-) can be created (Fenton reaction) (reviewed in (Andrews et al. 2003)). Since phagocytosed bacteria can face high levels of oxidative stress inside macrophages we tested the survival of a *fhuA* mutant in different conditions regarding the presence/absence of H_2O_2 (2mM H_2O_2) and different concentrations of Ferrichrome plus Fe^{3+}. Ferrichrome is a siderophore which binds iron III and enables it to be transported through the FhuA outer membrane transporter. We find that in the presence of Ferrichrome alone (100mM) or complexed with Fe^{3+} (100mM Fe^{3+}, 1000mM Ferrichrome), survival of the *fhuA* mutant is indistinguishable from that of ancestral bacteria (**Fig. 5A**). A similar result was obtained in the presence of H_2O_2 (**Fig. 5B**). A fitness advantage of the evolved clone was however detected in an environment comprising oxidative stress in conjunction with Ferrichrome, or with Ferrichrome and Fe^{3+}. Under these conditions the survival of *fhuA* mutant clones is significantly increased in relation to that of ancestral bacteria (P=0.001 and P=0.008 respectively) (**Fig. 5C**). The difference between the mutant and the ancestor is observed even in the absence of Fe^{3+} supplementation. This could be justified by the fact that bacteria are able to grow under limited amounts of this element, which is present under most biological conditions (Hartmann & Braun 1981). These results therefore indicate that this mutation may have evolved to decrease the amount of OH^- inside the bacterial cells.

An *in vitro* evolved double mutant of *rfaI* and *fhuA* shows increased pathogenic potential *in vivo*

We tested for increased virulence of one of the MΦ adapted clones. This clone carries two of the mutations that repeatedly emerged during the evolution: an insertion into *rfaI* and an insertion into *fhuA* (clone I **Fig. 3**). By infecting mice in the intra-peritoneal cavity with either the ancestral strain or the double mutant we find that, although mouse survival is similar for both strains, the weight loss caused by the infection of the evolved strain was significantly higher (P=0.046 for strain and P=0.003 for time, in a linear mixed effects model, with mouse as a random effect and strain and day of infection as factors, see **Fig. S2**). Given that the increased pathogenic potential of the double mutant was significant but not very strong we did not test each of the single mutants. Besides weight loss, a common phenotype to assay pathogenicity *in vivo*, we also measured temperature, but found no significant difference.

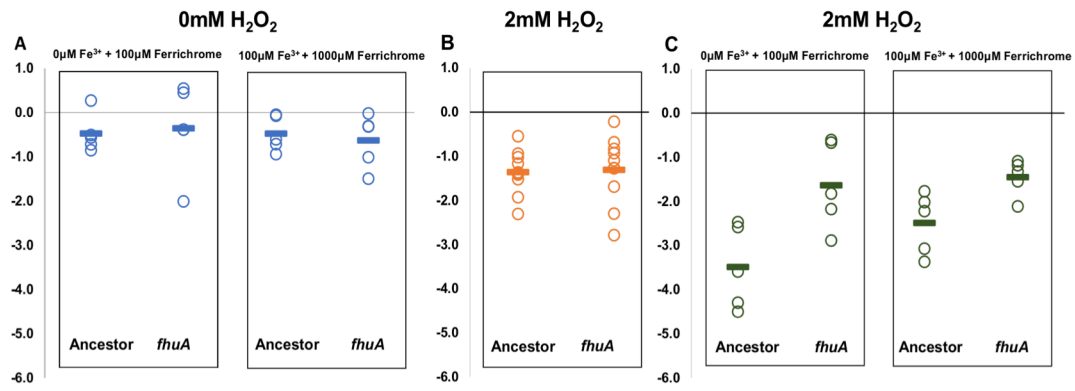


Fig. 5. Bacterial survival under iron and or oxidative stress. Log_{10}(Number of bacteria after 1h)- Log_{10} (Number bacteria at 0h) in the environments: (A) Effect of Fe^{3+} depletion tested in minimal media supplemented with the indicated concentrations of Fe^{3+} and Ferrichrome; no significant difference between evolved and ancestral clones could be detected (T-test, P=0.8 and P=0.6, n=5) (B) Effect of H_2O_2 stress tested in minimal media supplemented 2mM of H_2O_2; no significant difference between evolved and ancestral clones (T-test, P=0.8, n=10). (C) Fitness advantage of evolved clone is detected under both selective pressures, i.e. minimal media supplemented with the Fe^{3+}, Ferrichrome and 2mM of H_2O_2 (T-test, P=0.016, P=0.02, n=5).

Pathoadaptation to macrophages can lead to metabolic tradeoffs

Bacteria fully adapted to intracellular life tend to have small genomes (Moran 2002). Amongst the species of *E. coli*, *Shigella* strains have undergone a considerable amount of genome reduction (Hershberg et al. 2007). During its evolution from an extracellular inhabitant of the mammalian gut to an intracellular pathogen, *Shigella* accumulated a plethora of pseudogenes, with genes coding for carbon utilization, cell motility, transporter or membrane proteins more likely to become inactivated (Feng et al. 2011). While part of this gene loss may be the outcome of intensified genetic drift and inefficient selection, it can also be the result of positive selection for loss of anti-virulent functions, constituting adaptive losses in the intracellular niche (Andersson & Kurland 1998; Moran 2002). Such losses may entail antagonistic effects in extracellular environments. We tested the adapted clones for differences in their ability to grow on single carbon sources and found that some exhibited a strong metabolic trade-off when growing on either glucose or maltose (**Fig. 6**). We found that all the clones carrying the pathoadaptive loss of *rfaI* failed to reach high carrying capacity on minimal media with either of the sugars. In contrast, the mutants with pathoadaptive mutations in *tppB*, involved in the transport of peptides, showed increased growth in maltose (**Fig. 6**, red lines). Interestingly, subsequent mutations on the *rfaI* mutant background restore the ability to grow to similar levels as the ancestor, showing that the pleotropic effects of such pathoadaptation can be compensated to better grow on both poor and rich media.

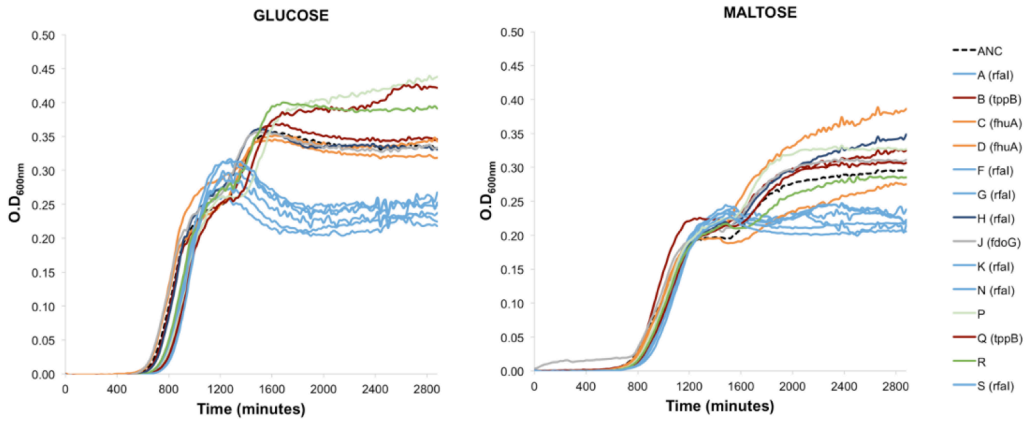


Fig. 6. Growth curves of evolved populations and the ancestral strain in minimal media with maltose or glucose (0.4%). Each color on the growth curves represents similar patterns amongst the populations.

Conclusions

Characterizing the evolutionary and genetic mechanisms underlying the transition from commensal to pathogenic lifestyle is paramount in understanding the particularities of what makes pathogens dangerous and often fatal to their hosts. Here, we followed the evolution of a commensal strain of *E. coli* under the selective pressure imposed by the intracellular niche of MΦs to identify the most probable paths of this adaptation. All evolved populations show an increased ability to survive in the presence of macrophages, as the result of acquisition of strong beneficial mutations, which we estimate and measure to be around 7 to 10%, on average. The characterization of their genetic basis unveiled mutation that were highly likely to be pathoadaptive mutations, namely those involving changes in LPS, crucial in the interaction with the immune system, and in iron metabolism, essential for both protecting against high levels of toxicity and to acquire the necessary resources to survive. Given the strong pressure imposed in our experimental system, our results show that commensal bacteria are able to acquire adaptations to increase their intracellular survival at a fast pace. Importantly, the adaptive mutations identified in this study suggest possible new

therapeutic targets to counteract pathogenic intracellular parasites.

Acknowledgments

We thank the members of Dr. Gordo's Lab for comments on the manuscript and Raffaella Gozzelino for help in the fitness assays of iron depletion.

References

Adams J, Rosenzweig F. 2014. Experimental microbial evolution: history and conceptual underpinnings. Genomics. 104:393–398. doi: 10.1016/j.ygeno.2014.10.004.

Andersson SG, Kurland CG. 1998. Reductive evolution of resident genomes. Trends in Microbiology. 6:263–268.

Andrews SC, Robinson AK, Rodríguez-Quiñones F. 2003. Bacterial iron homeostasis. FEMS Microbiology Reviews. 27:215–237.

Atwood KC, Schneider LK, Ryan FJ. 1951. Periodic selection in Escherichia coli. Proceedings of the National Academy of Sciences. 37:146–155.

Barrick JE et al. 2009. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature. 461:1243–1247. doi: 10.1038/nature08480.

Barrick JE, Kauth MR, Strelioff CC, Lenski RE. 2010. Escherichia coli rpoB Mutants Have Increased Evolvability in Proportion to Their Fitness Defects. Molecular Biology and Evolution. 27:1338–1347. doi: 10.1093/molbev/msq024.

Barroso-Batista J et al. 2014. The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps Coop, G, editor. PLoS Genet. 10:e1004182. doi: 10.1371/journal.pgen.1004182.s016.

Bateman SL, Seed PC. 2012. Epigenetic regulation of the nitrosative stress response and intracellular macrophage survival by extraintestinal pathogenic Escherichia coli. Molecular Microbiology. 83:908–925. doi: 10.1111/j.1365-2958.2012.07977.x.

Beutler B, Rietschel ET. 2003. Timeline: Innate immune sensing and its roots: the story of endotoxin. Nat Rev Immunol. 3:169–176. doi: 10.1038/nri1004.

Bokil NJ et al. 2011. Intramacrophage survival of uropathogenic Escherichia coli:

differences between diverse clinical isolates and between mouse and human macrophages. Immunobiology. 216:1164–1171. doi: 10.1016/j.imbio.2011.05.011.

Chattopadhyay S et al. 2009. High frequency of hotspot mutations in core genes of Escherichia coli due to short-term positive selection. Proc. Natl. Acad. Sci. U.S.A. 106:12412–12417. doi: 10.1073/pnas.0906217106.

Conde-Álvarez R et al. 2012. The lipopolysaccharide core of Brucella abortus acts as a shield against innate immunity recognition. PLoS Pathog. 8:e1002675. doi: 10.1371/journal.ppat.1002675.

Crossman LC et al. 2010. A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic Escherichia coli strain H10407. Journal of Bacteriology. 192:5822–5831. doi: 10.1128/JB.00710-10.

Dobzhansky T. 1950. Evolution in the tropics. American Scientist.

Feng Y, Chen Z, Liu S-L. 2011. Gene decay in Shigella as an incipient stage of host-adaptation. PLoS ONE. 6:e27754. doi: 10.1371/journal.pone.0027754.

Ferguson AD, Hofmann E, Coulton JW, Diederichs K. 1998. Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. Science.

Girgis HS, Harris K, Tavazoie S. 2012. Large mutational target size for rapid emergence of bacterial persistence. Proc. Natl. Acad. Sci. U.S.A. 109:12740–12745. doi: 10.1073/pnas.1205124109.

Gordo I, Campos PRA. 2013. Evolution of clonal populations approaching a fitness peak. Biology Letters. 9:20120239. doi: 10.1098/rsbl.2012.0239.

Guttman DS, Dykhuizen DE. 1994. Detecting selective sweeps in naturally occurring Escherichia coli. Genetics. 138:993–1003.

Hacker J, Kaper JB. 2000. Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. 54:641–679. doi: 10.1146/annurev.micro.54.1.641.

Hartmann A, Braun V. 1981. Iron uptake and iron limited growth of Escherichia coli K-12. Arch. Microbiol. 130:353–356.

Hegreness M. 2006. An Equivalence Principle for the Incorporation of Favorable Mutations in Asexual Populations. Science. 311:1615–1617. doi: 10.1126/science.1122469.

Herron MD, Doebeli M. 2013. Parallel evolutionary dynamics of adaptive diversification in Escherichia coli. PLoS Biol. 11:e1001490. doi: 10.1371/journal.pbio.1001490.

Hershberg R, Tang H, Petrov DA. 2007. Reduced selection leads to accelerated gene loss in Shigella. Genome Biology. 8:R164. doi: 10.1186/gb-2007-8-8-r164.

Hoare A et al. 2006. The outer core lipopolysaccharide of Salmonella enterica serovar Typhi is required for bacterial entry into epithelial cells. Infection and Immunity. 74:1555–1564. doi: 10.1128/IAI.74.3.1555-1564.2006.

Illingworth CJR, Mustonen V. 2012. A method to infer positive selection from marker dynamics in an asexual population. Bioinformatics. 28:831–837. doi: 10.1093/bioinformatics/btr722.

Kawecki TJ et al. 2012. Experimental evolution. Trends in Ecology & Evolution. 27:547–560. doi: 10.1016/j.tree.2012.06.001.

Koeppel AF et al. 2013. Speedy speciation in a bacterial microcosm: new species can arise as frequently as adaptations within a species. The ISME Journal. 7:1080–1091. doi: 10.1038/ismej.2013.3.

Koli P, Sudan S, Fitzgerald D, Adhya S, Kar S. 2011. Conversion of commensal Escherichia coli K-12 to an invasive form via expression of a mutant histone-like protein. mBio. 2. doi: 10.1128/mBio.00182-11.

Lang GI, Desai MM. 2014. The spectrum of adaptive mutations in experimental evolution. Genomics. 104:412–416. doi: 10.1016/j.ygeno.2014.09.011.

Leimbach A, Hacker J, Dobrindt U. 2013. E. coli as an all-rounder: the thin line between commensalism and pathogenicity. Curr. Top. Microbiol. Immunol. 358:3–32. doi: 10.1007/82_2012_303.

Limoli DH et al. 2014. Cationic Antimicrobial Peptides Promote Microbial Mutagenesis and Pathoadaptation in Chronic Infections Ausubel, FM, editor. PLoS Pathog. 10:e1004083. doi: 10.1371/journal.ppat.1004083.s011.

Maharjan R. 2006. Clonal Adaptive Radiation in a Constant Environment. Science. 313:514–517. doi: 10.1126/science.1129865.

Maharjan RP et al. 2012. The multiplicity of divergence mechanisms in a single evolving population. Genome Biology. 13:R41. doi: 10.1186/gb-2012-13-6-r41.

Maurelli AT. 2007. Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. FEMS Microbiology Letters. 267:1–8. doi: 10.1111/j.1574-6968.2006.00526.x.

Miskinyte M et al. 2013. The Genetic Basis of Escherichia coli Pathoadaptation to Macrophages Monack, DM, editor. PLoS Pathog. 9:e1003802. doi: 10.1371/journal.ppat.1003802.s018.

Miskinyte M, Gordo I. 2014. Fitness measurements of Evolved Esherichia coli.

Mittal R et al. 2010. Fcγ receptor I alpha chain (CD64) expression in macrophages is critical for the onset of meningitis by Escherichia coli K1. PLoS Pathog. 6:e1001203. doi: 10.1371/journal.ppat.1001203.

Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell. 108:583–586.

Moura de Sousa JA, Campos PRA, Gordo I. 2013. An ABC Method for Estimating the Rate and Distribution of Effects of Beneficial Mutations. Genome Biology and Evolution. 5:794–806. doi: 10.1093/gbe/evt045.

Ning Z, Cox AJ, Mullikin JC. 2001. SSAHA: a fast search method for large DNA databases. Genome Research. 11:1725–1729. doi: 10.1101/gr.194201.

Perfeito L, Pereira MI, Campos PRA, Gordo I. 2008. The effect of spatial structure on adaptation in Escherichia coli. Biology Letters. 4:57–59. doi: 10.1038/27900.

Pérez JC, Kumamoto CA, Johnson AD. 2013. Candida albicans Commensalism and Pathogenicity Are Intertwined Traits Directed by a Tightly Knit Transcriptional Regulatory Circuit Heitman, J, editor. PLoS Biol. 11:e1001510. doi: 10.1371/journal.pbio.1001510.s010.

Robinson JT et al. 2011. Integrative genomics viewer. Nature Biotechnology. 29:24–26. doi: 10.1038/nbt.1754.

Schmidt H, Hensel M. 2004. Pathogenicity islands in bacterial pathogenesis. Clinical Microbiology Reviews. 17:14–56.

Schnaitman CA, Klena JD. 1993. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. Microbiol. Rev. 57:655–682.

Schoustra SE, Bataillon T, Gifford DR, Kassen R. 2009. The Properties of Adaptive Walks in Evolving Populations of Fungus Barton, NH, editor. PLoS Biol. 7:e1000250. doi: 10.1371/journal.pbio.1000250.s009.

Smith EJ, Thompson AP, O'Driscoll A, Clarke DJ. 2013. Pathogenesis of adherent-invasive Escherichia coli. Future Microbiology. 8:1289–1300. doi: 10.2217/fmb.13.94.

Sokurenko EV, Hasty DL, Dykhuizen DE. 1999. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. Trends in Microbiology. 7:191–195.

Sousa A, Bourgard C, Wahl LM, Gordo I. 2013. Rates of transposition in Escherichia coli. Biology Letters. 9:20130838. doi: 10.1098/rsbl.2013.0838.

Sousa A, Magalhaes S, Gordo I. 2012. Cost of Antibiotic Resistance and the Geometry of Adaptation. Molecular Biology and Evolution. 29:1417–1428. doi: 10.1093/molbev/msr302.

Sun G et al. 2012. Dynamic Population Changes in Mycobacterium tuberculosis During Acquisition and Fixation of Drug Resistance in Patients. Journal of Infectious Diseases. 206:1724–1733. doi: 10.1093/infdis/jis601.

Tenaillon O, Skurnik D, Picard B, Denamur E. 2010. The population genetics of commensal Escherichia coli. Nature Reviews Microbiology. 8:207–217. doi: 10.1038/nrmicro2298.

Torres AG, Jeter C, Langley W, Matthysse AG. 2005. Differential binding of Escherichia coli O157:H7 to alfalfa, human epithelial cells, and plastic is mediated by a variety of surface structures. Applied and Environmental Microbiology. 71:8008–8015. doi: 10.1128/AEM.71.12.8008-8015.2005.

Torres AG, Kaper JB. 2003. Multiple elements controlling adherence of enterohemorrhagic Escherichia coli O157: H7 to HeLa cells. Infection and Immunity.

Utaisincharoen P, Anuntagool N, Chaisuriya P, Pichyangkul S, Sirisinha S. 2002. CpG ODN activates NO and iNOS production in mouse macrophage cell line (RAW 264.7). Clin. Exp. Immunol. 128:467–473.

van Ditmarsch D et al. 2013. Convergent Evolution of Hyperswarming Leads to Impaired Biofilm Formationin Pathogenic Bacteria. CellReports. 1–12. doi: 10.1016/j.celrep.2013.07.026.

Vaudaux P, Waldvogel FA. 1979. Gentamicin antibacterial activity in the presence of human polymorphonuclear leukocytes. Antimicrobial agents and chemoteraphy. 16:743–749.

Wilson K. 2001. Preparation of genomic DNA from bacteria. Curr Protoc Mol Biol. Chapter 2:Unit 2.4. doi: 10.1002/0471142727.mb0204s56.

Supplementary Figures and Tables

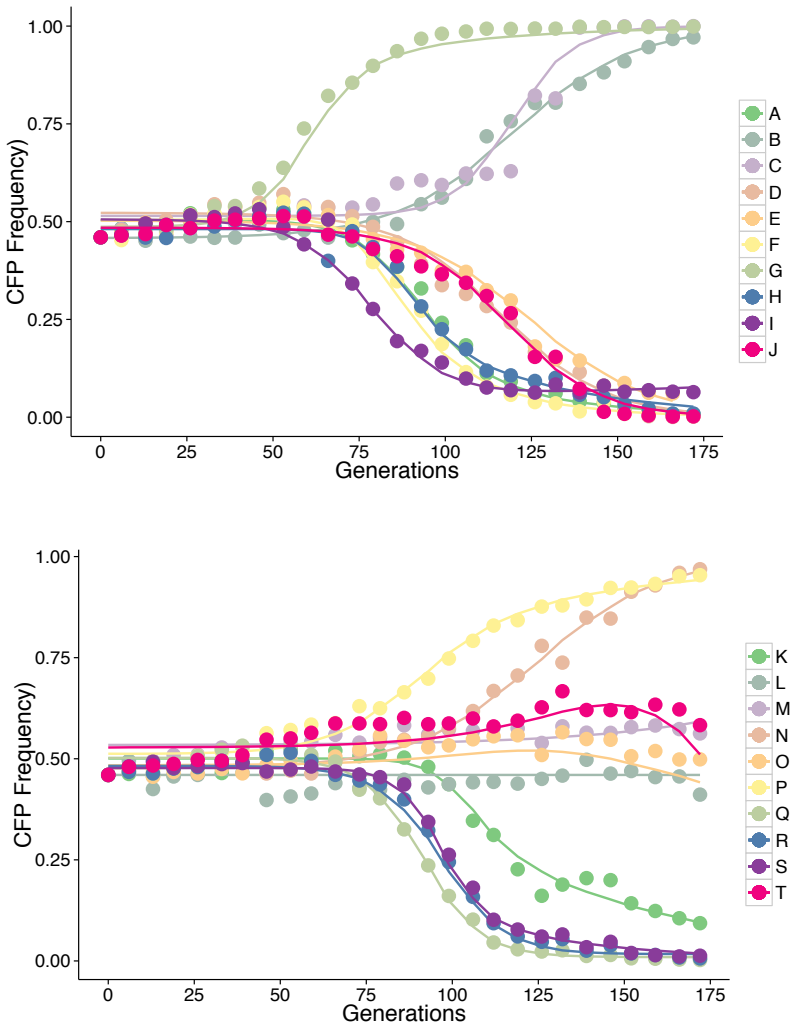


Fig. S1. **Inferred Evolutionary Dynamics.** Simulated dynamics of the model of positive selection (Illingworth & Mustonen 2012) with the parameters that provide the best fit to the data of changes in marker frequencies (displayed as points). Each color represents an independently evolved population.

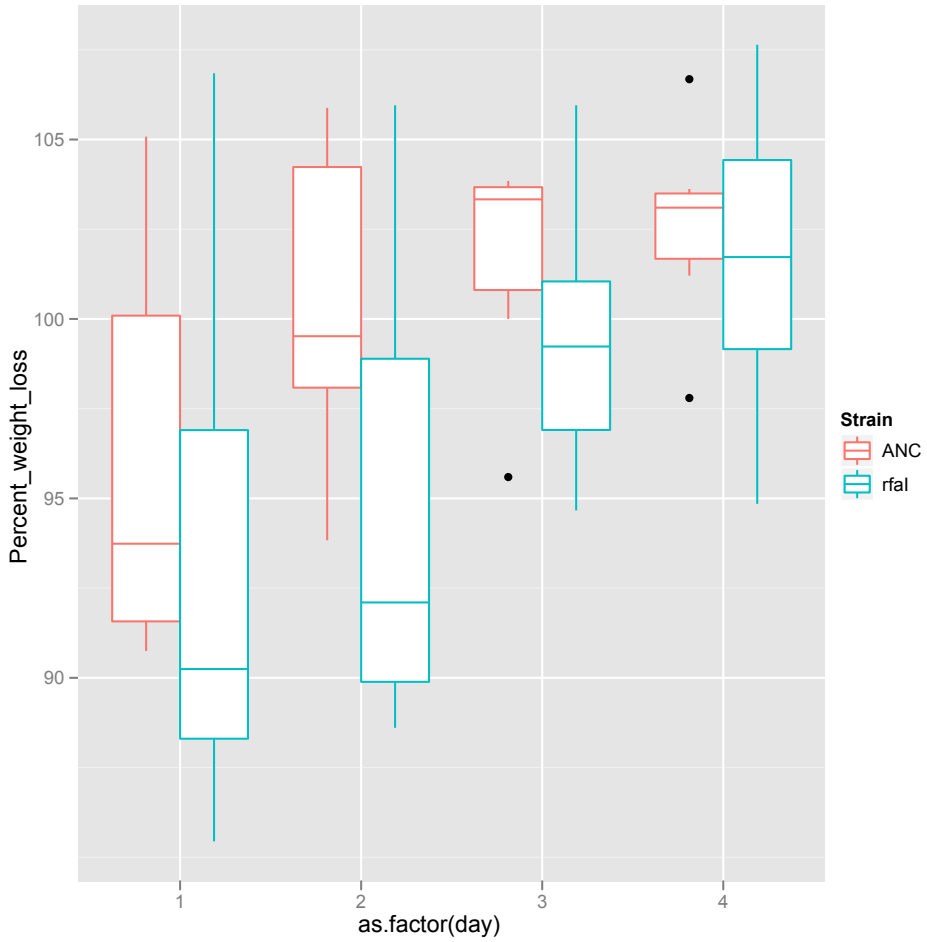


Fig. S2. ***rfaI* and *fhuA* double mutants increase weight loss in mice.** The change in weight of mice (as a percentage) after intra-peritoneal infection with ancestral or evolved (clone I) bacteria.

Population	*Slope*	*P-Value*
A	0.025	0.005
B	0.017	0.05
C	0.016	0.057
D	0.017	0.038
E	0.012	0.184
F	0.023	0.009
G	0.03	<0.001
H	0.022	0.009
I	0.03	<0.001
J	0.019	0.022
K	0.028	0.002
L	0.016	0.052
M	0.018	0.031
N	0.027	0.002
O	0.021	0.011
P	0.024	0.0056
Q	0.029	0.001
R	0.032	<0.001
S	0.03	<0.001
T	0.019	0.024

S1 Table. **Increase in bacterial loads along the experiment.** The majority of lines show a significant increase in bacterial loads. The slope of Log_{10}(CFUs) along the 26 days of evolution, from a linear regression, is indicated in the 1st column and P value of slope indicated in the 2nd column.

	Experimental	Inferred
A YFP	[0.06 - 0.08]	0.1
B CFP	[0.10 - 0.12]	0.05
C CFP	[0.11 - 0.13]	0.11
D YFP	[0.05 -0.07]	0.06
F YFP	[0.05 - 0.09]	0.13
G CFP	[0.06 - 0.08]	0.12
H YFP	[0.07 - 0.09]	0.13
J YFP	[0.06 - 0.08]	0.07
K YFP	[0.03 - 0.07]	0.14
N CFP	[0.04 - 0.06]	0.04
Q YFP	[0.8 - 0.12]	0.11
R YFP	[0.05 - 0.07]	0.1
S YFP	[0.05 - 0.07]	0.15

S2 Table. **Correspondence between experimental and inferred fitness.** The experimental values measured through competitive fitness assays are indicated with their errors (2SE), along with the fitness inferred through the marker dynamics of the respective population. In the majority of populations (with the exceptions of populations B, K and S), the two measures are either in agreement or the inferred fitness is slightly overestimated (see the main text for discussion).

CHAPTER VII

Clonal interference is sufficient to explain the pathoadaptive phenotype emerging during *Escherichia coli* adaptation to escape macrophage phagocytosis

Research included in the manuscript **"The Genetic Basis of Escherichia coli Pathoadaptation to Macrophages"**, published in ***PLoS Pathogens***, in December 2013

The author of this thesis performed all the theoretical simulations and their analysis described in this chapter. All figures were retrieved from the original publication.

Clonal interference is sufficient to explain the pathoadaptive phenotype emerging during Escherichia coli adaptation to escape macrophage phagocytosis

Abstract

Antagonistic interactions are common in the microbial world, driving the adaptation of microorganisms. Here we aim to understand the importance of clonal interference in an experiment involving the adaptation of *Escherichia coli* under antagonistic interaction with macrophages during 30 days. Morphological diversity was observed to rapidly emerge from this adaptive process, generating small colony variants and large translucid mucoid colonies. Moreover, a remarkable genetic parallelism was detected at the end of the experiment, across independently evolving populations. We developed a theoretical model that attempts to explain the dynamics of frequency change of the mucoid phenotype, taking into account the mutational events observed. We show that a simple model of clonal interference is able to generate the complex dynamics observed during adaptation to the presence of macrophages.

Introduction

Emergence of morphological diversity in Escherichia coli *adapting to the presence of macrophages*

Macrophages are a key component of the host defense mechanisms against pathogens (Wynn et al. 2013) and thus provide a strong antagonistic interaction to which bacteria must adapt. In **Chapter VI**, we discussed an ecological scenario where bacteria adapted to inhabit the intracellular milieu of macrophages, prompted by the supplementation of media with an antibiotic to

which they were susceptible. However, adaptation to the presence of macrophages is not limited to adopt an intracellular lifestyle, since escape and avoidance strategies can also evolve from this interaction (Baxt et al. 2013). In (Miskinyte et al. 2013), *E. coli* was allowed to evolve under continuous selective pressure of macrophages, but without the application of gentamycin, thus allowing bacteria to survive by more efficiently escaping the predatory behaviours of macrophages. In order to assert the speed and genetic basis of adaptation to this antagonistic interaction, six independently evolved populations were followed for 30 days, during which distinct colony morphologies emerged across the replicate populations and coexisted with the ancestral colony phenotype: small colony variants (SCVs) and large translucid mucoid (MUC) colonies. SCVs were observed in 5 out of the 6 populations, but only transiently and at low frequencies. In contrast, MUC clones, which were detected in all populations, rose in frequency and reached fixation in 5 out of the 6 populations, showing complex frequency dynamics (**Figure 1**). Both these phenotypes are known to emerge in clinical infections (Funada et al. 1978; Besier et al. 2008) and, therefore, it is important to understand their molecular and evolutionary dynamics.

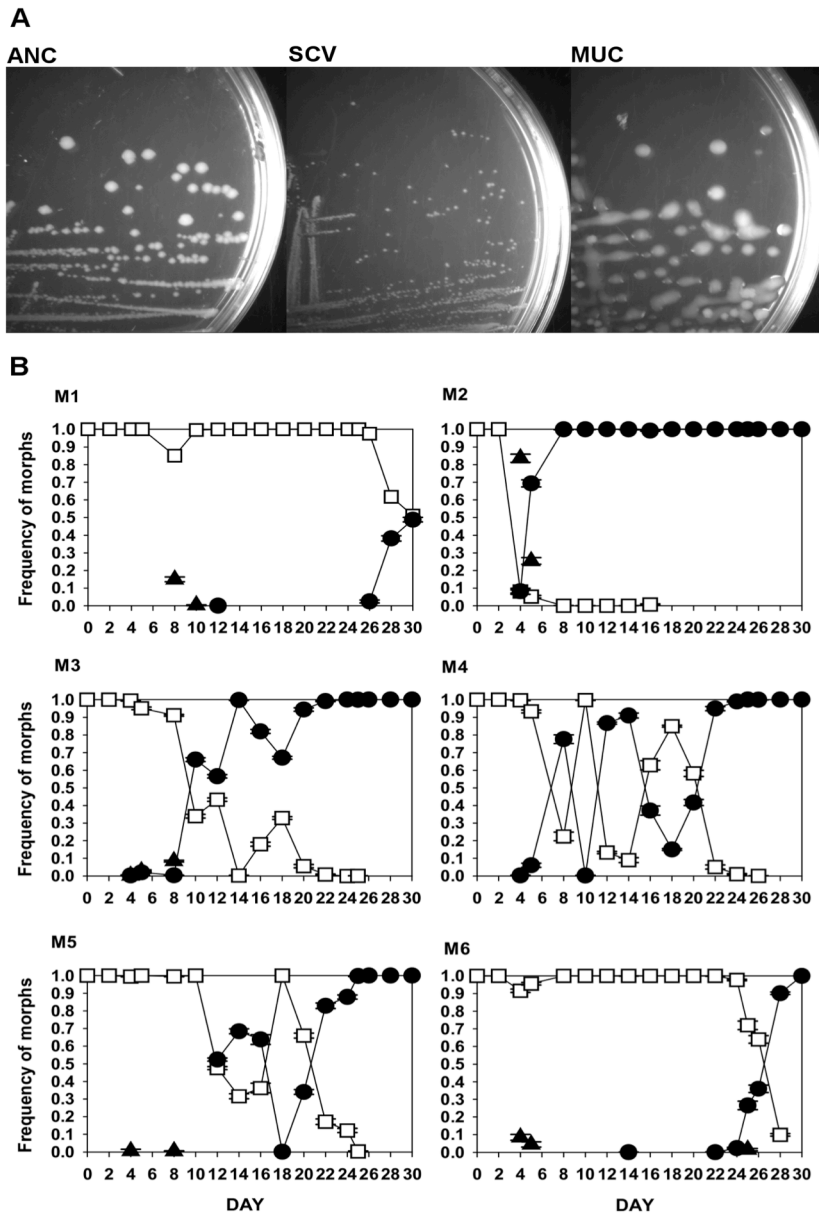


Figure 1. Emergency of morphological diversity in the bacterial populations adapting to MΦ.

(A) Examples of the variability for colony morphology that emerged in *E. coli* populations adapting to M_{Φ}, from left to right – ANC stands for morphology of ancestral, SCV for the small colony variants morphology and MUC for the mucoid colony morphology. (B) Dynamics of frequency change of the evolved phenotypes in each replicate evolving populations (M1 to M6): white squares indicate ANC, black triangles SCV, black circles MUC phenotypes.

Genetic basis of adaptation to macrophages

In order to uncover the molecular basis of the MUC phenotype dynamics, whole genome sequencing was performed not only for the evolved clones at day 30, for all populations, but also for one of the clones in population M3, at day 19. The mutations found are represented in **Figure 2** and **Table 1**, and the frequency of transposon insertions (IS) along the experiment was detected by PCR (**Figure 2b**). There is a high level of parallelism across the independently evolving populations, namely with the regulatory region of the *yrfF* gene being interrupted by an IS, in all six replicate populations. The homologue of *yrfF* in *Salmonella* (*igaA*) prevents the over-activation of the Rcs regulatory system, which controls the production of colanic acid capsule synthesis (Bernal & Pucciarelli 2004). This mutation is likely altering *E. coli*'s ability to produce colanic acid, in concordance to the MUC phenotype observed. To further understand the dynamics of adaptation in each independent evolved bacterial population, the frequency of the mutations found was assessed in mucoid clones sampled along the evolution experiment. Adaptation involved the competition between distinct haplotypes and the successive accumulation of beneficial mutations, mainly caused by IS insertions (**Figure 2**). Such haplotype dynamics is characteristic of clonal interference (Sniegowski & Gerrish 2010), where clones carrying distinct beneficial mutations compete for fixation.

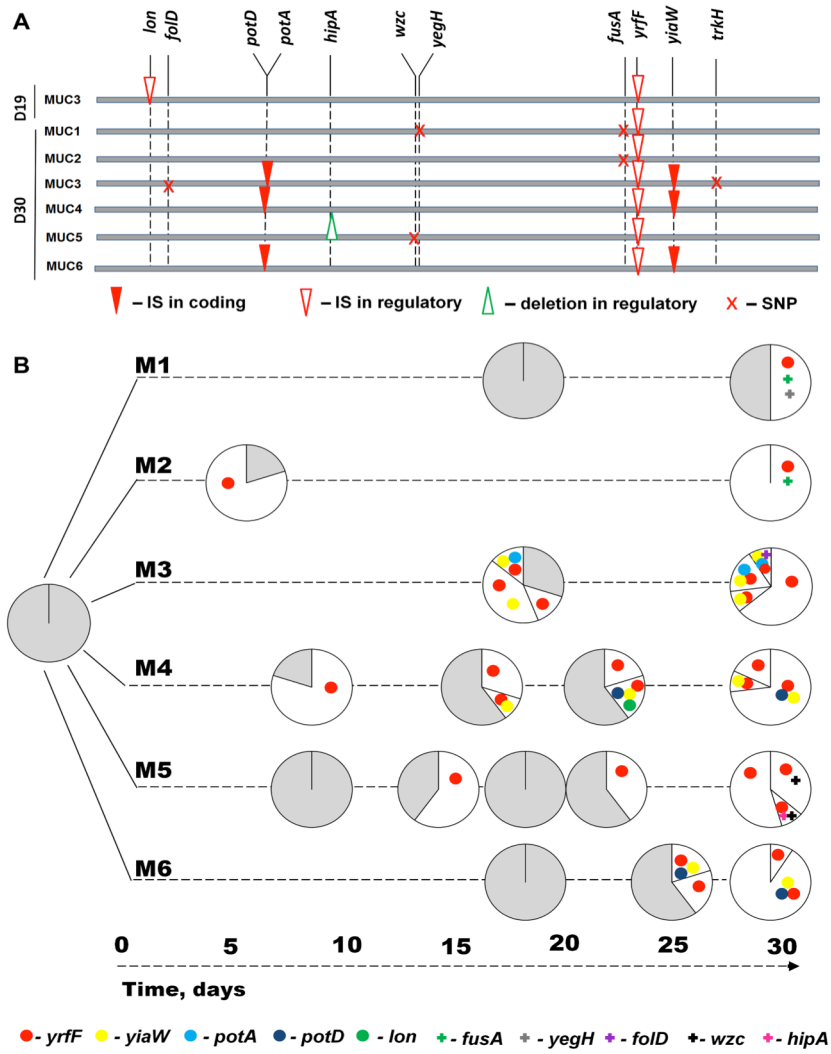


Figure 2. Genetic characterization of adaptive mutations and the dynamics of their appearance.

(A) Mutations identified in MUC1 to MUC6 clones isolated from M1 to M6 populations (evolved for 450 generations), represented along the *E. coli* chromosome. For simplicity, the genomes are represented linearly and are horizontally drawn. The types of mutations are represented in the following way: SNPs are shown as crosses, IS insertions as inverted triangles and deletions as triangles. Filled symbols represent mutation in the coding region of the gene and empty symbols in the regulatory region. (B) Emergence and spread of adaptive mutations in M1 to M6 populations. Dynamics of haplotype frequencies in evolving populations at different days of evolution experiment are represented by circles. The color and symbol (IS insertions are represented as circles and other mutations as crosses) of each sector represents different haplotypes and the area of the circle their frequency in the population. Grey area represents the frequency of clones in the population that were typed for existing mutations in the population and did not differ from ancestral haplotype.

Clone	Genome Position	Gene	Mutation	Annotation
MUC_M3_D19	360771	*clpX/lon*	intergenic (+88/−100)	IS186 +12 bp
	3411601	*nudE/yrfF*	intergenic (−273/−47)	IS1 +10
MUC1	2029672	*yegH*	G→T	A422S GCC→TCC
	3356932	*fusA*	A→C	S588A TCC→GCC
	3411601	*nudE/yrfF*	intergenic (−273/−47)	IS1 +10
MUC2	3356932	*fusA*	A→C	S588A TCC→GCC
	3411605	*nudE/yrfF*	intergenic (−277/−43)	IS1 +6
MUC3	459734	*folD/sfmA*	G→T	intergenic (−10/−461)
	1088154	*potA*	coding (589/1137 nt)	IS1 +10
	3411601	*nudE/yrfF*	intergenic (−273/−47)	IS1 +10
	3640515	*yiaW*	coding (263/324 nt)	IS1 +9
	3922002	*trkH*	T→A	L389Q CTG→CAG
MUC4	1084946	*potD*	coding (1032/1047 nt)	IS1 +9
	3411601	*nudE/yrfF*	intergenic (−273/−47)	IS1 +10
	3640515	*yiaW*	coding (263/324 nt)	IS1 +9
MUC5	1480525	*ydeS/hipA*	intergenic (−1603/+205)	Δ208 bp
	2024227	*wzc*	G→T	P645T CCG→ACG
	3411601	*nudE/yrfF*	intergenic (−273/−47)	IS1 +10
MUC6	1084946	*potD*	coding (1032/1047 nt)	IS1 +9
	3411601	*nudE/yrfF*	intergenic (−273/−47)	IS1 +10
	3640515	*yiaW*	coding (263/324 nt)	IS1 +9

Table 1. Mutations acquired by evolved clones identified through whole genome re-sequencing (WGS).
Mutations in intergenic regions have the two flanking genes listed (e.g., *clpX*/*lon*). SNPs are represented by an arrow between the ancestral and the evolved nucleotide. Whenever a SNP gives rise to a non-synonimous mutation the amino acid replacement is also indicated. The symbol Δ means a deletion. For intergenic mutations, the numbers in the "Mutation" row represent nucleotides relative to each of the neighboring genes, here + indicates the distance downstream of the stop codon of a gene – indicates the distance upstream of the gene, that is relative to the start codon. Insertions of IS elements are denoted by the specific IS element followed by the number of repeated bases caused by its insertion.

Model and Results

Model of clonal interference for the antagonistic interaction between bacteria and macrophages

In order to understand whether clonal interference could reproduce the mucoid and ancestral-like frequency dynamics observed, we modeled this process within the basic ecological scenario of our experiment. We assume a simple model for the interaction between bacteria and macrophages:

Eq. 1: $\frac{dB}{dt} = B(r - \frac{B}{K} + a_m M_\Phi e^{-\delta t})$

Here, ***B*** represents the number of bacteria along time, ***r*** their growth rate, ***K*** their carrying capacity and a_m is the factor of bacterial death by macrophages (a_m<0); M_Φ represents the number of macrophages and **δ** their death rate. For the initial conditions we used, in accordance with the experimental setup, $M_\Phi=10^6$ and $B(0)=10^6$. During a 24 hour period the infection dynamics expected under this model are similar to those obtained experimentally, with r=2.3 (per hour), $K=10^8$ and δ=0.1 (per hour) (**Figure 3**). As in the experiments, every 24 hours bacterial population numbers are reduced to $B(0)$ and macrophage numbers increased to M_Φ.

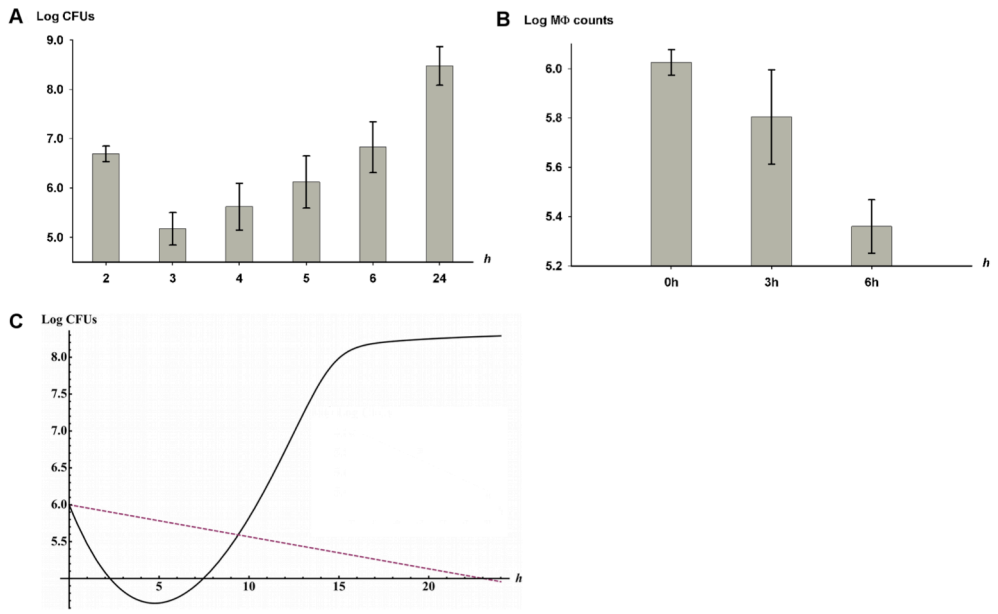


Figure 3: Infection dynamics of the ancestral strain.

Variation in numbers of bacteria (A) and M_Φ (B) during an infection with the ancestral clone (ANC) at MOI (1:1). (C) Simulated dynamics of a population of ancestral bacteria (black line) dividing in the presence of M_Φ (purple line) for 24 hours, following the deterministic model, with the following parameter values: $B(0)=10^6$; $M_\Phi=10^6$; $r=2.3$; $K=10^8$; $a_m=-3.7*10^{-6}$; $\delta=0.1$. We assume that M_Φ decay at linear rate of 0.1 following the data of (B).

Having these initial conditions, we then assume that new mutants arise and are not stochastically lost at a given rate. More specifically, the ancestral clone can mutate to two new types of adapted clones: one by transposition (upstream of *yrfF*) and another by point mutation. These mutations can have different rates (Sousa et al. 2013; Foster et al. 2015) and can also cause changes in both $\boldsymbol{r}$ and $\boldsymbol{a_m}$, i.e., the growth rate of mutant bacteria ($\boldsymbol{r_{muc}}$) can change and their ability to interact with macrophages ($\boldsymbol{a_{mmuc}}$) can also change. We assume that mucoid bacteria will exhibit an increased ability to escape macrophages $\boldsymbol{a_{mmuc}} < \boldsymbol{a_m}$ but also a decreased growth rate $\boldsymbol{r_{muc}} < \boldsymbol{r}$, due to the cost of producing

exopolysaccarides. While in some cases the mucoid haplotype displays an apparently simple phenotypic sweep (see, for instance, M2 or M6 in **Figure 1B**), in other instances the frequencies of the morphologies have a higher degree of complexity (M4 as an extreme cases in **Figure 1B**). This seems to indicate that the level of clonal interference could be different amongst the populations. Therefore, and in order to model different orders of clonal interference under an antagonistic interaction, we start with the following equations in order to describe the frequencies of the ancestral background and a derived mucoid haplotype:

Eq. 2a: $$\frac{dB}{dt} = B\left(r - \frac{(B + Muc)}{K} + a_m M_\Phi e^{-\delta t}\right) - U_{is} B$$

Eq. 3a: $$\frac{dMuc}{dt} = Muc\left(r_m - \frac{(B + Muc)}{K} + a_{mmuc} M_\Phi e^{-\delta t}\right) + U_{is} B$$

Here U_{is} is the rate of occurrence of successful transpositions leading to the mucoid haplotype. B represents the number of bacteria with the ancestral genotype and Muc of the mucoid genotype that emerges. From the experimental data, we know that the non-mucoid (or ancestral-like) phenotypes sometimes increase in frequency and are maintained in the populations (white squares in **Figure 1B** and grey areas in **Figure 2B**). Therefore, to simulate some of the dynamics in **Figure 1B**, we postulate that a phenotype derived from ancestral type (***B'***), which is non-mucoid, can emerge in the population at rate U, with growth r' and macrophage interaction a_{mb}. Note that both U and U_{is} are the spontaneous rate of mutation times the probability that such mutation is not lost by drift.

Eq. 2b: $$\frac{dB}{dt} = B\left(r - \frac{(B + Muc + B')}{K} + a_m M_\Phi e^{-\delta t}\right) - UB - U_{is} B$$

Eq. 4a: $$\frac{dB'}{dt} = B\left(r' - \frac{(B + Muc + B')}{K} + a_{mb} M_\Phi e^{-\delta t}\right) + UB$$

Moreover, because successional mutations are acquired in the background of the mucoid phenotype (**Figure 2B**), we also postulate that subsequent mucoid haplotypes can originate from the first one, at a rate $U_{is'}$ and with growth r'_m and macrophage interaction $a_{mmuc'}$. The complete set of equations for a model consisting of the segregation of an ancestral and 3 new haplotypes is the following:

Eq. 2c: $$\frac{dB}{dt} = B\left(r - \frac{(B + Muc + B' + Muc')}{K} + a_m M_\Phi e^{-\delta t}\right) - UB - U_{is} B$$

Eq. 3b:

$$\frac{dMuc}{dt} = Muc\left(r_m - \frac{(B + Muc + B' + Muc')}{K} + a_{mmuc} M_\Phi e^{-\delta t}\right) - U_{is'} Muc + U_{is} B$$

Eq. 4b: $$\frac{dB'}{dt} = B\left(r' - \frac{(B + Muc + B' + Muc')}{K} + a_{mb} M_\Phi e^{-\delta t}\right) + UB$$

Eq. 5: $$\frac{dMuc'}{dt} = Muc'\left(r'_m - \frac{(B + Muc + B' + Muc')}{K} + a_{mmuc'} M_\Phi e^{-\delta t}\right) + U_{is'} Muc$$

In this model of clonal interference, we make the simplifying assumption that two traits, growth rate (r) of bacteria and their ability (a_m) to escape M_Φ, are the most important for bacterial fitness in this environment. Both these traits can evolve, as it is evident from the emergence of new morphologies (in particular mucoid morphs) and phenotypic tests of the evolved clones. **Figure 4** shows simulations of adaptive dynamics over the period of the experiment (30 days), where the frequencies of mucoid (MUC) phenotypes are plotted and can be compared to those observed in the experiments (**Figure 1B**). The solutions for the equations were obtained in Mathematica v8.0 (the full script can be found as supplementary material in (Miskinyte et al. 2013)). These parameters were chosen because they depict the initial infection dynamics of the ancestral strain (**Figure 1**) and its relation with the derived evolved clones. We are able to find conditions that

can reproduce the observed dynamics of morphs under a scenario of intense clonal interference and accumulation of multiple beneficial mutations (see **Figure 5**). The parameters used to obtain these dynamics, which are indicative of the rates and strength of mutations that drive this clonal interference process, are shown in **Table 1**. For some of the dynamics, further haplotypes, both mucoid and non-mucoid, were required to explain the dynamics, and the parameters for these are shown in **Table 2**. The model is able to reproduce the observed changes in frequency of the mucoid and non-mucoid phenotypes, if we assume that distinct beneficial mutations occur and change each of the two fitness traits in both morphs. Specifically, we assume that a successful beneficial mutation occurs, which produces the first mucoid morph. Such mutation is assumed to increase the ability of bacteria to escape M_{Φ}, but also carry a cost in that it diminishes the growth rate (see **Figure 6** for the conditions under which such a mutation can invade). This mucoid morph can acquire further beneficial mutations that alter those traits values, such that a new derived mucoid haplotype can have a reduced cost of producing colanic acid and/or an increased ability to escape macrophages. Importantly, we also assume that clones with ancestral colony morphology can acquire beneficial mutations that increase their growth rate. Under these assumptions, and with the direct evidence that several distinct clones are segregating in the populations, complex dynamics are to be expected (**Figures 4** and **5**).

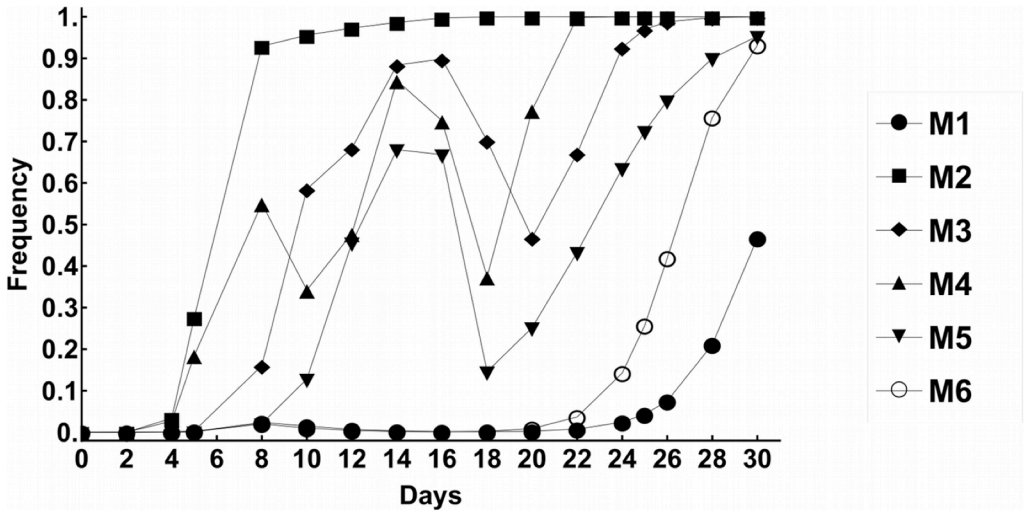


Figure 4. Predictions of model of clonal interference for changes in mucoid frequencies with time.

Simulations of the adaptive dynamics over the period of the experiment (30 days). The frequencies of mucoid phenotypes are plotted and can be compared to those observed in the experiments (**Figure 1B**). The values of parameters used are shown in **Tables 1** and **2** and the dynamics of haplotypes that compete for fixation are shown in **Figure 5**.

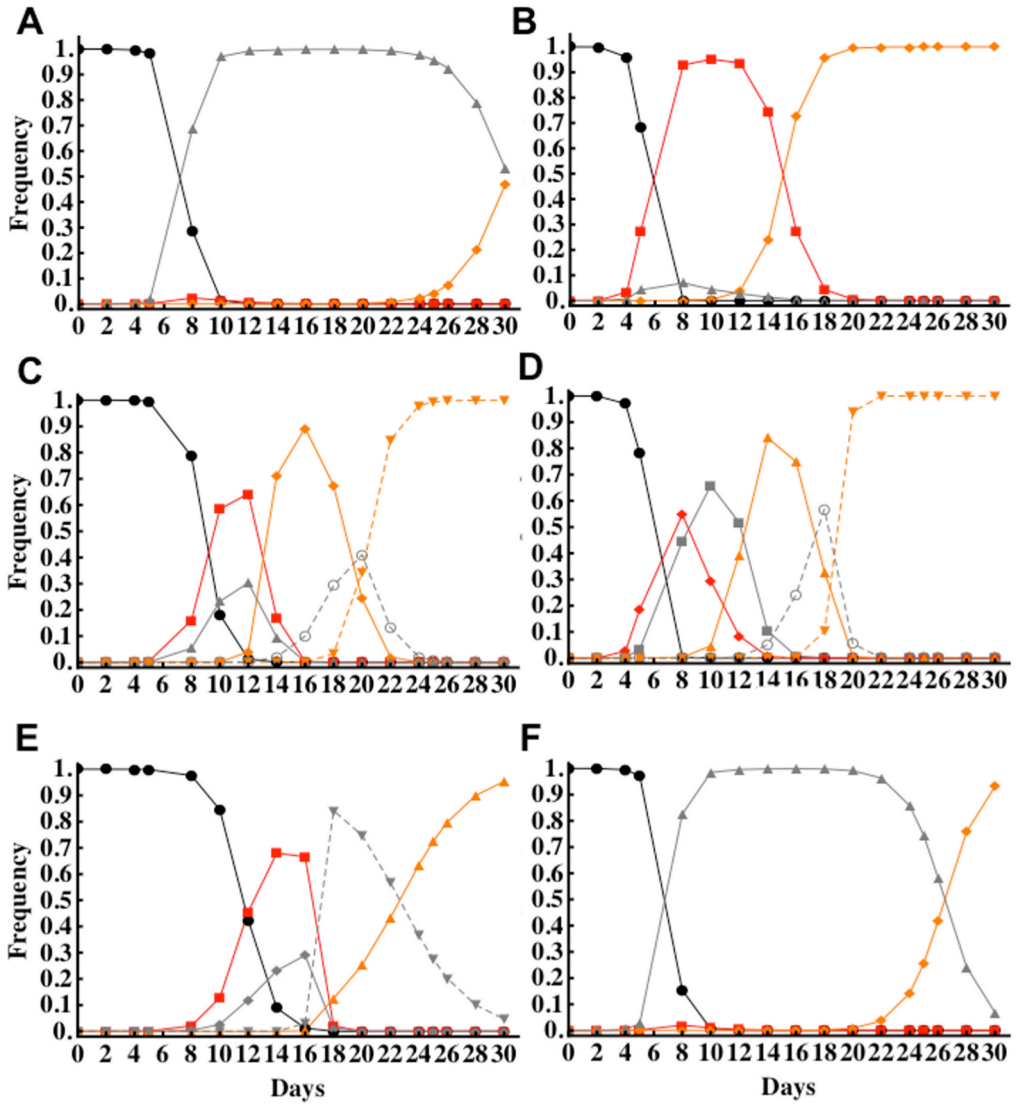


Figure 5. Dynamics for the different haplotypes under the model of clonal interference.

Simulated frequencies of the different haplotypes that result in the frequencies of the mucoid phenotypes of **Figure 4**. r = 2.3, am = −3.7×10−6 and the other parameters used are shown in **Table 1**. In this table, the cases where more haplotypes were assumed to reproduce the experimental dynamics are marked with *, and the additional parameters are in **Table 2**.

	U ($x10^{-7}$)	Uis ($x10^{-7}$)	r_m	a_{mmuc} ($x10^{-6}$)	Uis' ($x10^{-7}$)	r'	a_{mb} ($x10^{-6}$)	r_m'	$a_{mmuc'}$ ($x10^{-6}$)
M1	3	4	2.1758	-3.23	12	2.369	-3.7	2.2171402	-3.23
M2	0.96	4.45	2.21122	-3.2	3.84	2.392	-3.7	2.2554444	-3.2
M3*	0.43	4	2.185	-3.24	1.72	2.3609	-3.7	2.2724	-3.24
M4*	0.1	10	2.1988	-3.2	5	2.4104	-3.7	2.2691616	-3.2
M5*	0.43	6.94	2.180768	-3.28	0.107	2.346	-3.7	2.180768	-2.86
M6	3	4	2.1758	-3.23	12	2.3736	-3.7	2.2247555	-3.23

Table 1. Parameters used in modeling the dynamics of the different haplotypes. Parameters used for the dynamics in **Figure 5**. Cases where more haplotypes were assumed to reproduce the experimental dynamics are marked with *, and the additional parameters are in **Table 2**.

	U''(B'->B'')	r''	$a_{mb'}$ ($x10^{-6}$)	Uis''(Muc'->Muc'')	$r_{m''}$	$a_{mmuc''}$ ($x10^{-6}$)
M3*	4.3×10^{-8}	2.47309513	-3.7	1.72×10^{-7}	2.3428444	-3.24
M4*	1×10^{-8}	2.494764	-3.7	1×10^{-8}	2.3202177	-2.99
M5*	4.3×10^{-8}	2.482068	-3.7			

Table 2. Parameters for the additional haplotypes for the modelled dynamics. Parameters in the additional haplotypes required to obtain the dynamics in **Figure 5**.

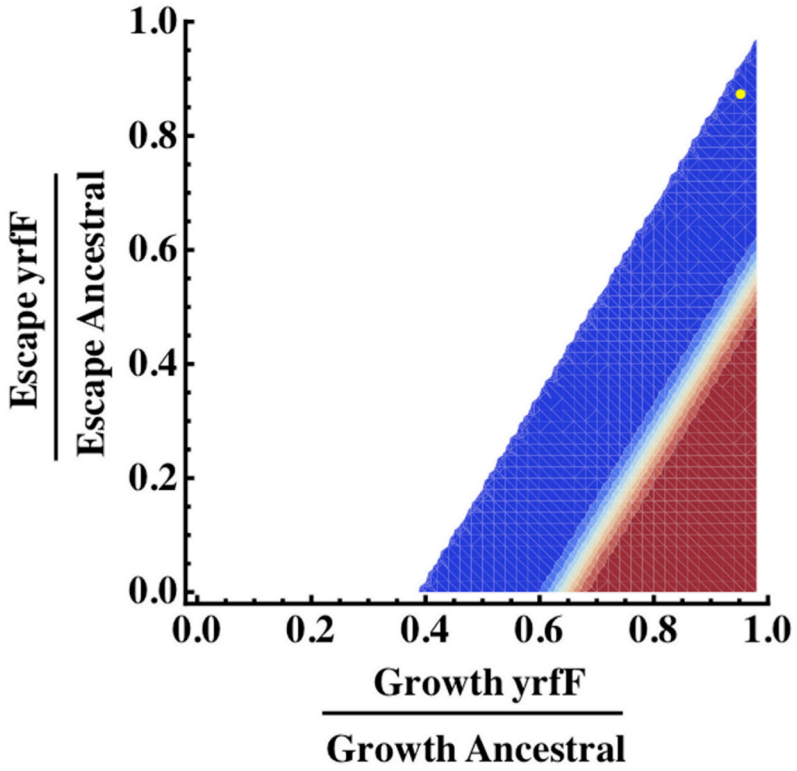


Figure 6. Region of parameter space theoretically expected for the invasion of first mucoid morph. Colored areas show the parameter region (r_m/r and a_{mmuc}/a_m) where a mucoid genotype (mimicking the IS insertion upstream of *yrfF* in the experiment) that has emerged is able to increase in frequency so that it can survive the bottleneck imposed every 24 hours in the experiment. The equations for these simulations are:

$$\frac{dB}{dt} = B\left(r - \frac{(B+Muc)}{K} + a_m M_\Phi e^{-\delta t}\right)$$

$$\frac{dMuc}{dt} = Muc\left(r_m - \frac{(B+Muc)}{K} + a_{mmuc} M_\Phi e^{-\delta t}\right)$$

with initial conditions Muc(0)=1, B(0)= 10^6 and the other parameter values as in **Figure 3**: M_0=10^6; r=2.3; K=10^8; a_m= 3.7*10^{-6}; d=0.1. Note that the escape parameter is negative (according to the mathematical model) and, therefore, a value lower than 1 indicates a higher ability to escape predation. Warmer colors show higher frequency of the mucoid genotype in the population after 24 hours of its emergence as a single copy. The yellow dot indicates the value of r_m and a_{mmuc}, of the first mucoid haplotype assumed to emerge in the 6 models that produced the dynamics in **Figures 4** and **5**.

Discussion

The complex adaptive dynamics driving the pathoadaptive process of *E. coli* in the presence of macrophages are difficult to disentangle in an environment that poses antagonistic biotic interactions and the generation of multiple phenotypes and haplotypes, as shown by the genetic analysis. One of the drivers of this process is clonal interference. Here we studied if clonal interference alone could explain the dynamics of the morphologies emerging during the evolution, under specific parameter values for the rate and effects of mutations causing the observed phenotypes. One of the strongest assumptions of the model, regarding the existence of only two main phenotypic traits (growth and escape from predation), might be too simplistic, even in this simple ecological setup. Other traits and environmental conditions that we did not consider might play a role. These include the structure of the environment, differential variation in the cell densities, or the possibility of frequency dependent selection emerging from the secreted expolysaccharide potentially becoming a public good. Therefore, it is important to note that we cannot exclude the possible occurrence of other forms of selection during the evolution experiment. Indeed it is known that even in simple abiotic environments adaptive diversification, involving frequency dependent selection can repeatedly evolve (Lenski et al. 1991; Maddamsetti et al. 2015). It is generally difficult to distinguish this form of selection from simple clonal interference. However, if strong negative frequency dependent selection would have occurred in our lines, it would have led to the maintenance of distinct lineages along the evolutionary process. Such expectation seems inconsistent at least with the fixation of the IS insertion upstream of *yrfF* and the observed fixation of the mucoid phenotype in the majority of the lines, but could explain further diversification across replicate populations and the complex dynamics observed in some of them.

Whilst our model of clonal interference, can reasonably describe the complex dynamics of the evolved phenotypes, we should stress the considerable amount of sensitivity in the assumed model. Indeed, a minor change in one of the

parameters results in very different haplotype dynamics. The sensitivity of the model to slight changes in parameter values is expected under intense clonal interference, but it is also suggests that the assumptions to generate the competition between the haplotypes might be an oversimplification of the real system. Nevertheless, the lack of theoretical approaches that are able to model selection under clonal interference and antagonistic interactions requires an initial analysis under the most basic scenarios of adaptation, before invoking other forms of selection. This is fundamental in order to distinguish cases where clonal interference cannot explain certain observations. Here we have shown that such a simple model, where distinct haplotypes carrying new transposon insertions and other mutations, increase in frequency and compete for fixation, can reasonably explain the adaptive dynamics, indicating that clonal interference might be the main driver of pathoadaptation in our basic ecological scenario. Nevertheless, further experiments in the lab involving competitions between mucoid and ancestral phenotypes suggest that frequency-dependent selection may have also played a role. Models that consider both forms of selection might therefore be necessary to better understand these complex adaptive dynamics.

References

Baxt LA, Garza-Mayers AC, Goldberg MB. 2013. Bacterial Subversion of Host Innate Immune Pathways. Science. 340:697–701. doi: 10.1126/science.1235771.

Bernal GD, Pucciarelli MG. 2004. Repression of the RcsC-YojN-RcsB phosphorelay by the IgaA protein is a requisite for Salmonella virulence. Molecular ….

Besier S, Zander J, Kahl BC, Kraiczy P. 2008. The thymidine-dependent small-colony-variant phenotype is associated with hypermutability and antibiotic resistance in clinical Staphylococcus aureus isolates. Antimicrobial agents ….

Foster PL, Lee H, Popodi E, Townes JP, Tang H. 2015. Determinants of spontaneous mutation in the bacterium Escherichia colias revealed by whole-genome sequencing. Proc. Natl. Acad. Sci. U.S.A. 201512136. doi:

10.1073/pnas.1512136112.

Funada H, Hattori KI, Kosakai N. 1978. Catalase-negative Escherichia coli isolated from blood. J. Clin. Microbiol.

Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. American Naturalist.

Maddamsetti R, Lenski RE, Barrick JE. 2015. Adaptation, Clonal Interference, and Frequency-Dependent Interactions in a Long-Term Evolution Experiment with Escherichia coli. Genetics. 200:619–631. doi: 10.1534/genetics.115.176677.

Miskinyte M et al. 2013. The Genetic Basis of Escherichia coli Pathoadaptation to Macrophages Monack, DM, editor. PLoS Pathog. 9:e1003802. doi: 10.1371/journal.ppat.1003802.s018.

Sniegowski PD, Gerrish PJ. 2010. Beneficial mutations and the dynamics of adaptation in asexual populations. Philosophical Transactions of the Royal Society B: Biological Sciences. 365:1255–1263. doi: 10.1126/science.285.5426.422.

Sousa A, Bourgard C, Wahl LM, Gordo I. 2013. Rates of transposition in Escherichia coli. Biology Letters. 9:20130838. doi: 10.1098/rsbl.2013.0838.

Wynn TA, Chawla A, Pollard JW. 2013. Macrophage biology in development, homeostasis and disease. Nature.

CHAPTER VIII

Discussion

Understanding the dynamics involved in the exploration of adaptive spaces, as well as the evolutionary parameters that drive this process, is essential not only for understanding adaptation, but also for predicting, preventing and reversing diverse pathologies inflicted by microbial infections. The studies presented in this thesis, as well as recently published research (e.g., (Strelkowa & Lässig 2012; Lang et al. 2013; Maharjan et al. 2015; Maddamsetti et al. 2015; Wilson et al. 2016; Barroso-Batista et al. 2014; Comas et al. 2012; Lieberman et al. 2014)), establish the influence of clonal interference in adaptation across a range of increasingly complex ecological scenarios and organisms. We started by developing tools and algorithms for estimating evolutionary parameters from experimental evolution studies that use neutral markers to track adaptive events. Both our method and ones developed by other groups (Hegreness 2006; Barrick et al. 2010; Illingworth & Mustonen 2012) were used to analyze the outcomes of evolving *Escherichia coli* populations. We studied how the genetic background affects adaptation, using compensation of single and multiple antibiotic resistant bacteria as a model. We also studied adaptation across different environmental contexts, including antagonistic interactions with cells of the immune system in an attempt to uncover likely paths of acquisition of pathogenic traits. Understanding how adaptation proceeds in these clinically relevant cases, using theoretical methods to infer important parameters, is paramount to quantify the strength of selection for new mutations that might lead to pathoadaptation.

Determining, from the entire spectrum of new variants, the rate, fraction and shape of the distribution of beneficial mutations is quite difficult to do experimentally, because many of these will be lost before reaching detectable frequencies. However, these evolutionary parameters underlie the adaptive paths that microorganisms can follow. It is thus important to understand whether strong effect mutations are rare (e.g., if new mutations are exponentially distributed) or if, on the other hand, they are the most likely, or the only expected mutations to occur. In **Chapter II**, we showed that the assumption that all mutations have a similar fitness effect, while useful in explaining certain adaptive dynamics

(Hegreness 2006; Barrick et al. 2010), might mislead the inference of the beneficial mutation rate (U) and the mean effect of mutations (E(S)), underestimating the former and overestimating the latter. Since there is experimental evidence that beneficial mutations have a considerable variation in their fitness effects (Kassen & Bataillon 2006), wrongly estimating the selective effects can mislead the interpretation of evolutionary outcomes. This is important, for instance, in studies addressing the evolvability of a genetic background (Barrick et al. 2010; Woods et al. 2011). We proposed a method to infer the underlying distribution of effects of arising beneficial mutations, as well as the rate at which they arise in an evolving population, by tracking two neutral markers and the population fitness over time. With the technological advances and the increasing widespread use of dozens or even hundreds of replicate populations in experimental evolution studies (Maisnier-Patin et al. 2002; Lenski et al. 2015; Lang et al. 2013; Barroso-Batista et al. 2014), methods that use direct measures from the adaptive dynamics are needed to extract important quantitative parameters from experimental data. The methodology we propose to estimate these evolutionary parameters is, like the work of other groups (Hegreness 2006; Barrick et al. 2010; Illingworth & Mustonen 2012), simple and relatively inexpensive to use, since it uses data routinely acquired during experimental evolution. Across all these methods, the difficulty in accurately assessing these parameters is strongly dependent on the magnitude of clonal interference (Sniegowski & Gerrish 2010), as well as in assumptions of frequency independent effects (Maddamsetti et al. 2015) and an additive effect of mutations (i.e., that no epistasis occurs, but see (Wiser et al. 2013) for an example of an underlying model that assumes epistasis to infer the dynamics of fitness increase in the LTEE). Moreover, there is a limitation imposed by the use of only two markers, since once one of these markers reaches fixation, new mutations cannot be detected. This problem can be alleviated by using additional information, such as the fitness of populations, as considered in our method, but the information given by the marker dynamics is lost. This is especially problematic for stronger effect mutations leading to rapid marker fixation. Alternative methods have been proposed in order to tackle the

question of determining the distribution of arising beneficial mutations, which involve single cell barcoding of the population (Levy et al. 2015). These methods allow for an exquisite level of resolution of the generation of diversity within a population but are, however, still technologically demanding. A possible tradeoff can be reached with experimental systems that use several neutral markers (such as microsatellites (Imhof & Schlotterer 2001; Perfeito et al. 2007), although an appropriate model for how these mutate is needed, as well as theoretical methods that allow the simulation of such data.

With the power and increasing accessibility of NGS data, the dynamics of adaptation are now being studied by following individual mutations and their fate in evolving populations. Recent experimental data disclosing these dynamics corroborate the idea that beneficial alleles generally do not segregate alone and are, instead, associated with further allelic modifications (Lang et al. 2013; Maddamsetti et al. 2015; Maharjan et al. 2015). More than simply competing against each other, these variants aggregate in cohorts, forming phenotypes that hide complex genetic compositions (Maharjan et al. 2015). Recently, these dynamics have also been tracked and comparatively studied between sexual and asexual *Saccharomyces cerevisiae* populations, showing the magnitude of linkage in segregating mutations across backgrounds (McDonald et al. 2016). In **Chapter III** we use a classical model of adaptation, Fisher's Geometric Model (FGM), to ask if these dynamics are commonly observed during the evolutionary ascent of a fitness peak in haploid populations. We observe that the patterns observed experimentally, the formation of cohorts and their "interference", can be reproduced under the simplest version of FGM, pointing to the ubiquity of these types of adaptive dynamics. Moreover, we were able to define a range of parameters where we expect these cohorts to have a higher prevalence. It would be interesting to test experimentally the dynamics of individual mutations in different environments, or in backgrounds with increased mutation rates, to test the predictions of the model. Since FGM can lead to epistasis (Weinreich & Knies 2013), it would also be compelling to disclose the fraction of cohorts that are

simply passenger mutations, hitchhiking with beneficial alleles, from those that are composed of synergistic beneficial mutations. In principle, this could be tested by separately testing the vector change caused by the different mutations in an ancestral genotype. It has recently been proposed that, under FGM, epistasis is expected to be more prevalent closer to the optimum (Blanquart et al. 2014). Therefore, it would be interesting to ask if the fraction of cohorts with synergistic mutations is dependent also on the distance to a fitness peak. Moreover, although the model we considered assumes that only competitive interactions occur, the dynamics and coexistence of different cohorts have been proposed to be also a consequence of frequency dependent interactions (Maddamsetti et al. 2015; Maharjan 2006). Since both processes can drive these dynamics, it is important to develop methods that can account for both processes, and that are able to distinguish their relative roles in the dynamics of cohorts of mutations.

The dynamics of beneficial mutations studied in **Chapters II** and **III** show that distinct evolutionary parameters lead to very different adaptive dynamics. Importantly, different genotypes (due to epistatic interactions) or environments (due to specific selective pressures) might entail distinct evolutionary trajectories caused, for instance, by different underlying distributions of arising beneficial effects. One very relevant scenario, in which it is crucial to disclose these fundamental aspects of adaptation, is the acquisition and maintenance of antibiotic resistant traits, a prevalent clinical problem (World Health Organization 2014; Hughes & Andersson 2015). While stabilization of single resistances through compensation has been extensively studied (see, for instance, (Levin et al. 2000; Reynolds 2000; Maisnier-Patin et al. 2007; Qi et al. 2016)), compensation of multidrug resistances, particularly when there are epistatic interactions between the resistance alleles (Hall & MacLean 2011; Trindade et al. 2009; Durão et al. 2015), has been far less explored. In **Chapter IV** we described the compensatory process in a multi-resistant *E. coli*, resistant to both rifampicin and streptomycin. Our initial hypothesis was that, since the two resistance alleles display strong negative synergistic epistasis, the pattern of compensation might be distinct from

the ones observed in the single resistance alleles. Indeed, our results show that the double resistant background followed an evolutionary path that seems to deviate from the ones pursued by the single resistances and, consequently, the observations acquired from studies of the latter might not be entirely transferrable to cases of multi-resistant strains. Evidence for the compensation of the epistatic interaction comes from a particular mutation ($rpoC^{Q1126K}$) that, while beneficial in a $Rif^{R}Str^{R}$ background (where it emerged), confers a deleterious effect upon allelic reconstruction in the single Rif^{R} ancestral background. An alternative explanation that is currently being tested is whether this particular allele is beneficial in the Str^{R} ancestral background, a result that would be surprising (since *rpoC* is a traditional target for compensation of rifampicin resistance (Brandis et al. 2012; De Vos et al. 2013; Comas et al. 2012)) but with important clinical implications. However, under the assumption that a subset of mutations is compensating specifically the genetic interactions between the resistance alleles, other questions arise. For instance, what is the fraction of this subset of compensatory mutations? Is it the majority? And how would these patterns of compensation look like in bacteria with resistance alleles that either a) are non-epistatic or b) display positive epistasis? Would an increased overlap of allelic targets be observed between these multi-resistant strains and their correspondent single resistant backgrounds? Or, on the other hand, are we really far from exhausting the compensatory targets for a given resistance allele? As in our study, the mutational targets identified in experiments involving compensatory adaptation allow further understanding of the mechanisms driving epistasis. Testing specific mutations that emerge during the compensatory process, and measuring relevant parameters, such as the speed of transcription and translation (for the particular resistance mutations we used in our work), could elucidate how these two fundamental cellular mechanisms interact. Another important question regards the types of compensatory targets that can be expected. Many studies so far have focused on genes that are known to be targets for compensatory adaptation, such as *rpoA*, *rpoB* or *rpoC* for the case of compensation to rifampicin resistance (Brandis & Hughes 2013) or genes coding for ribosomal proteins for streptomycin resistance (Maisnier-Patin et al. 2007).

However, the availability of NGS allows uncovering in further detail the genomic basis of compensation, both in laboratory (Qi et al. 2016) and clinical (Comas et al. 2012; Zhang et al. 2013; Lieberman et al. 2014) settings. In a recent study, the group of Craig MacLean has observed that compensation to rifampicin in *P. aeruginosa* occurs through mutations in the expected RNA polymerase genes when the resistance mutations are costly, but when the resistances are of low cost, adaptation occurs through mutations in the *lasR* gene, which is a highly pleiotropic regulator (Qi et al. 2016). Studying these alternative targets, and whether they have a direct relationship with the cellular mechanism of resistance, will further increase our knowledge of compensatory adaptation.

Whilst many compensatory mutations have been detected to segregate in clinical populations, especially for cases of infection by *M. tuberculosis* (e.g., (Comas et al. 2012; Zhang et al. 2013)), translating the studies of compensatory adaptation in the laboratory to both clinical and *in vivo* studies has not been straightforward (MacLean & Vogwill 2015). One of the reasons might be the one tackled in **Chapter IV**, but other factors might play an important role. For instance, the fitness cost of a resistance *in vitro* needs not to be the one expressed *in vivo* (Björkman et al. 2000; Martinez & Baquero 2000; Luangtongkum et al. 2012) or be similar across different environments (Trindade et al. 2012; Gifford et al. 2016); the usage of sub-lethal doses of antibiotics might lead to different resistance mechanisms (Andersson & Hughes 2014); or the compensation of resistance genes encoded in plasmids or other MGE, which typically have lower fitness costs (Vogwill & MacLean 2014), might occur through different processes. In **Chapter V**, we argue that ecology, through the coexistence of several resistance alleles within a population, might also play a role. The simple question we asked was whether the initial relative fitness cost of the resistances is predictive of its long-term fate. Should the more costly resistances be extinct or, on the other hand, can this disadvantage be overcome by an increased evolutionary potential to acquire compensatory mutations? To answer this question, we performed long-term competitions between *E. coli* subpopulations with different resistance alleles. This

is a scenario of standing genetic variation that is uncommonly studied in compensatory adaptation, even though it mimics ecological conditions that are much closer to a clinical infection (Wilson et al. 2016; Lieberman et al. 2014). The surprising outcome, for the 3 different combinations of resistant alleles we studied, is that the initial relative fitness cost is a poor predictor of the long-term maintenance of these alleles, as the more costly populations rarely got extinct, due to their evolvability. It would be interesting to study larger poly-resistant bacterial populations, in order to understand if evolvability plays a more outstanding role whenever there are more than 2 backgrounds in competition. Because compensation in different resistance alleles might occur within the same bacterial population, clonal interference will occur between the beneficial variants of either background. If their distributions of beneficial mutations are significantly distinct (for instance, due to epistasis with the resistance allele), then the outcome of clonal interference might bias the selection of beneficial variants from the background with a distribution with higher mean selective effect. Such a scenario would be unnoticed in the compensation of each of the resistant backgrounds alone. Whilst in the cases studied here, the evolutionary potential seems to play a major role in the maintenance of resistance, the coexistence of multiple resistant strains gives rise to plenty of opportunities for interactions to occur, and therefore shaping the composition of clinical infections. This is particularly likely for resistances that act through the production of public goods (Lee et al. 2010) and infections occurring in highly structured environments, such as cystic fibrosis (Lieberman et al. 2014; Winstanley et al. 2016). One interesting observation, that seems to point in this direction, is the apparent lack of transitivity in the fitness costs of the resistances we tested, where the cost against a sensitive ancestral is not transferrable to the competitions between the resistances. Why this lack of transitivity is observed (and what is its extent) is something that begs to be understood. Nevertheless, our results highlight the importance of measuring important evolutionary parameters in populations within contexts analogous to the one observed in clinically relevant scenarios (Eldholm et al. 2014), and thus closer to the ecological settings experienced by the microorganisms.

Ecology and environmental context seem then to be crucial in assessing adaptive dynamics in bacteria. An additional way to bridge the gap between real world microbial environments and the laboratory is through the study of adaptation in environments where biotic interactions are key to get closer to the conditions a given microbe is likely to encounter (Meyer & Kassen 2007; Gordo et al. 2014). The presence of cells from the innate immune system is one of the most relevant contexts for microbial evolution (Wynn et al. 2013; Schmid-Hempel 2008). The antagonistic interactions between these immune defenses and bacteria provide a strong environmental pressure and adaptation to withstand antimicrobial activity leads to severe pathogenic phenotypes (Sokurenko et al. 1999; Bokil et al. 2011; Brunke et al. 2014). Moreover, the ability of bacteria to survive in intracellular environments is also used as a strategy to avoid exposure to antibiotics, as in the case of *S. aureus* (Garzoni & Kelley 2011). What is, then, the genetic basis that leads to these pathogenic traits? In **Chapter VI** we studied the initial adaptive steps that might lead to the transition from commensalism to pathogenicity, by evolving *E. coli* in the intracellular environment of macrophages. Strong effect mutations, that increased the survival of evolved bacteria within macrophages relative to the ancestral, were acquired very rapidly, but this advantage was, in some cases, accompanied by a tradeoff in their absence. Adaptation to this environment was driven mostly by mutations in genes that code for two very relevant traits in the survival of *E. coli* in the presence of macrophages, the lipopolysaccharide structure and an iron transporter. Interestingly, clonal interference did not seem to be particularly prevalent here, and these few strong effect mutations dominated the adaptive process, possibly due to the lower effective population sizes resulting from the intracellular environment. However, in the conditions studied in **Chapter VII**, clonal interference was paramount in generating variation in the frequency of the emergent phenotypes. In the design of that study, we let both the bacteria and macrophages co-exist in the same environment and allow for mutations that could increase intracellular survival and/or extracellular fitness. The dominant mucoid phenotype only swept to fixation in 2 out of 6 replicate populations, while in the others its frequency varied

greatly along time. A model postulating clonal interference between haplotypes showed that the fast input of genetic mutations could, indeed, drive the dynamics observed in the frequencies of the morphologies. It is important to be able to distinguish whether clonal interference can explain the observed data, since several factors could account for a mode of selection that involves interactions between the evolved cells. Both the existence of spatial structure (Xavier & Foster 2007; Libberton et al. 2015) (through a static environment and the surface proportioned by macrophages) and the increased production of public goods (Lee et al. 2010) (colanic acid exopolysaccharides, in our study) could lead to frequency-dependent selection and the coexistence of several haplotypes and phenotypes (Stein et al. 2013; Maddamsetti et al. 2015; Cordero & Polz 2014). Moreover, in both our studies we detected single mutational events that increased the pathogenic traits in *E. coli*. Others studies that follow have observed similar patterns, namely in the fungi *Candida albicans*, where a single SNP can lead to a hypervirulent strain that is able to escape phagocytes (Brunke et al. 2014). In a clinical context, Lieberman and colleagues have studied the evolution of *Burkholderia dolosa* in cystic fibrosis, both across and within patients (Lieberman et al. 2011; 2014), identifying targets of pathoadaptive mutations recurrently acquired by this pathogen *in vivo*. Their results indicate the coexistence of different lineages within a single individual, but also genetic records of adaptation that include antibiotic resistance across individuals (Lieberman et al. 2014). However, the study of *Pseudomonas aeruginosa* in a similar cystic fibrosis context has indicated high levels of genetic convergence (Yang et al. 2011; Marvig et al. 2014), suggesting different evolutionary paths. These observations, both from our studies and others, show that the experimental evolution setup we used might be particularly suited to understand these important adaptive dynamics, and may contribute to anticipate the likely paths in the transition from commensal to pathogenic lifestyles.

Overall, the research explored in this thesis reinforces something fairly well known, but that sometimes is easily dismissed: microbes do not usually adapt in

simple environments, uncoupled from diversity and ecology, and, therefore, their adaption should also be studied under those more realistic scenarios. Experimental evolution has been, and still is, an invaluable tool to evolutionary biology, and has been key in recreating adaptive processes in increasingly complex ecological scenarios. The results we present and discuss here indicate that one of the fundamental characteristics of adaptation, the distribution of effects of beneficial mutations, depends on both the genetic background and the environmental context. Nonetheless, our studies are still very much abstractions of the real environments encountered by microbes, and further work to create conditions that are closer to the real world of microorganisms are necessary. In parallel, theoretical models used in the estimation of evolutionary parameters also need to be improved. Technological advances leading to the streamlining of data acquisition allows using further information from experimental studies, such as measures of population fitness, as we propose, or even data from whole genome sequencing of populations or clones. Simultaneously, using models that are considerably more complex, by taking into account epistasis between mutations (Hall & MacLean 2011; Wiser et al. 2013) or interactions between co-existing haplotypes that go beyond competitions (Maddamsetti et al. 2015; Stein et al. 2013), will be very useful to simulate data that can be compared with experimental observations.

Ultimately, an integrative approach to study adaptation – including experimental evolution, the assessment of genomic data at different time points and ecologically complex environments and theoretical approaches – is key to understand what is arguably the most essential process in biology. Evolutionary biology depends on the harmonious integration of these different approaches and the outcomes of these studies can also have a strong clinical and immediate application.

References

Andersson DI, Hughes D. 2014. Microbiological effects of sublethal levels of antibiotics. Nature Reviews Microbiology. doi: 10.1038/nrmicro3270.

Barrick JE, Kauth MR, Strelioff CC, Lenski RE. 2010. Escherichia coli rpoB Mutants Have Increased Evolvability in Proportion to Their Fitness Defects. Molecular Biology and Evolution. 27:1338–1347. doi: 10.1093/molbev/msq024.

Barroso-Batista J et al. 2014. The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps Coop, G, editor. PLoS Genet. 10:e1004182. doi: 10.1371/journal.pgen.1004182.s016.

Björkman J, Nagaev I, Berg OG, Hughes D, Andersson DI. 2000. Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. Science. 287:1479–1482.

Blanquart F, Achaz G, Bataillon T, Tenaillon O. 2014. Properties of selected mutations and genotypic landscapes under Fisher's geometric model. Evolution.

Bokil NJ et al. 2011. Intramacrophage survival of uropathogenic Escherichia coli: differences between diverse clinical isolates and between mouse and human macrophages. Immunobiology. 216:1164–1171. doi: 10.1016/j.imbio.2011.05.011.

Brandis G, Hughes D. 2013. Genetic characterization of compensatory evolution in strains carrying rpoB Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates. J. Antimicrob. Chemother. 68:2493–2497. doi: 10.1093/jac/dkt224.

Brandis G, Wrande M, Liljas L, Hughes D. 2012. Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. Molecular Microbiology. 85:142–151. doi: 10.1111/j.1365-2958.2012.08099.x.

Brunke S et al. 2014. One Small Step for a Yeast - Microevolution within Macrophages Renders Candida glabrata Hypervirulent Due to a Single Point Mutation Feldmesser, M, editor. PLoS Pathog. 10:e1004478. doi: 10.1371/journal.ppat.1004478.s014.

Comas I et al. 2012. Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nature Genetics. 44:106–110. doi: 10.1038/ng.1038.

Cordero OX, Polz MF. 2014. Explaining microbial genomic diversity in light of evolutionary ecology. Nature Reviews Microbiology. doi: 10.1038/nrmicro3218.

De Vos M et al. 2013. Putative compensatory mutations in the rpoC gene of rifampin-resistant Mycobacterium tuberculosis are associated with ongoing transmission. Antimicrobial Agents and Chemotherapy. 57:827–832. doi: 10.1128/AAC.01541-12.

Durão P, Trindade S, Sousa A, Gordo I. 2015. Multiple Resistance at No Cost: Rifampicin and Streptomycin a Dangerous Liaison in the Spread of Antibiotic Resistance. Molecular Biology and Evolution. 32:2675–2680. doi: 10.1093/molbev/msv143.

Eldholm V et al. 2014. Evolution of extensively drug-resistant Mycobacterium tuberculosis from a susceptibleancestor in a single patient. 1–11. doi: 10.1186/s13059-014-0490-3.

Garzoni C, Kelley WL. 2011. Return of the Trojan horse: intracellular phenotype switching and immune evasion by Staphylococcus aureus. EMBO Mol Med.

Gifford DR, Moss E, MacLean RC. 2016. Environmental variation alters the fitness effects of rifampicin resistance mutations in Pseudomonas aeruginosa. Evolution.

Gordo I, Demengeot J, Xavier K. 2014. Escherichia coliadaptation to the gut environment: a constant fight for survival. Future Microbiology. 9:1235–1238. doi: 10.2217/fmb.14.86.

Hall AR, MacLean RC. 2011. Epistasis buffers the fitness effects of rifampicin-resistance mutations in Pseudomonas aeruginosa. Evolution. 65:2370–2379. doi: 10.1111/j.1558-5646.2011.01302.x.

Hegreness M. 2006. An Equivalence Principle for the Incorporation of Favorable Mutations in Asexual Populations. Science. 311:1615–1617. doi: 10.1126/science.1122469.

Hughes D, Andersson DI. 2015. Evolutionary consequences of drug resistance: shared principles across diverse targets and organisms. Nature Publishing Group. 16:459–471. doi: 10.1038/nrg3922.

Illingworth CJR, Mustonen V. 2012. A method to infer positive selection from marker dynamics in an asexual population. Bioinformatics. 28:831–837. doi: 10.1093/bioinformatics/btr722.

Imhof M, Schlotterer C. 2001. Fitness effects of advantageous mutations in evolving Escherichia coli populations. Proceedings of the National Academy of Sciences. 98:1113–1117. doi: 10.1073/pnas.98.3.1113.

Kassen R, Bataillon T. 2006. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nature Genetics. 38:484–488. doi: 10.1038/ng1751.

Lang GI et al. 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature. 1–6. doi: 10.1038/nature12344.

Lee HH, Molla MN, Cantor CR, Collins JJ. 2010. Bacterial charity work leads to population-wide resistance. Nature. 467:82–85. doi: 10.1038/nature09354.

Lenski RE et al. 2015. Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with Escherichia coli. Proceedings of the Royal Society B: Biological Sciences. 282:20152292. doi: 10.1093/molbev/msq099.

Levin BR, Perrot V, Walker N. 2000. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. Genetics. 154:985–997.

Levy SF et al. 2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature. 519:181–186. doi: 10.1038/nature14279.

Libberton B, Horsburgh MJ, Brockhurst MA. 2015. The effects of spatial structure, frequency dependence and resistance evolution on the dynamics of toxin-mediated microbial invasions. Evol Appl. 8:738–750. doi: 10.1111/eva.12284.

Lieberman TD et al. 2014. Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. Nature Genetics. 46:82–87. doi: 10.1038/ng.2848.

Lieberman TD et al. 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nature Genetics. 43:1275–1280. doi: 10.1038/ng.997.

Luangtongkum T, Shen Z, Seng VW. 2012. Impaired fitness and transmission of macrolide-resistant Campylobacter jejuni in its natural host. Antimicrobial agents ….

MacLean RC, Vogwill T. 2015. Limits to compensatory adaptation and the persistence of antibiotic resistance in pathogenic bacteria. Evolution, Medicine, and Public Health. 2015:4–12. doi: 10.1093/emph/eou032.

Maddamsetti R, Lenski RE, Barrick JE. 2015. Adaptation, Clonal Interference, and Frequency-Dependent Interactions in a Long-Term Evolution Experiment with Escherichia coli. Genetics. 200:619–631. doi: 10.1534/genetics.115.176677.

Maharjan R. 2006. Clonal Adaptive Radiation in a Constant Environment. Science. 313:514–517. doi: 10.1126/science.1129865.

Maharjan RP, Liu B, Feng L, Ferenci T, Wang L. 2015. Simple phenotypic sweeps hide complex genetic changes in populations. Genome Biology and Evolution. 7:531–544. doi: 10.1093/gbe/evv004.

Maisnier-Patin S, Berg OG, Liljas L, Andersson DI. 2002. Compensatory adaptation to the deleterious effect of antibiotic resistance in Salmonella typhimurium. Molecular Microbiology. 46:355–366.

Maisnier-Patin S, Paulander W, Pennhag A, Andersson DI. 2007. Compensatory Evolution Reveals Functional Interactions between Ribosomal Proteins S12, L14 and L19. Journal of Molecular Biology. 366:207–215. doi: 10.1016/j.jmb.2006.11.047.

Martinez JL, Baquero F. 2000. Mutation frequencies and antibiotic resistance. Antimicrobial Agents and Chemotherapy.

Marvig RL, Sommer LM, Molin S, Johansen HK. 2014. Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. Nature Genetics. 47:57–64. doi: 10.1038/ng.3148.

McDonald MJ, Rice DP, Desai MM. 2016. Sex speeds adaptation by altering the dynamics of molecular evolution. Nature. 1–15. doi: 10.1038/nature17143.

Meyer JR, Kassen R. 2007. The effects of competition and predation on diversification in a model adaptive radiation. Nature.

Perfeito L, Fernandes L, Mota C, Gordo I. 2007. Adaptive Mutations in Bacteria: High Rate and Small Effects. Science. 317:813–815. doi: 10.1126/science.1142284.

Qi Q, Toll-Riera M, Heilbron K, Preston GM, MacLean RC. 2016. The genomic

basis of adaptation to the fitness cost of rifampicin resistance in Pseudomonas aeruginosa. Proceedings of the Royal Society B: Biological Sciences. 283:20152452. doi: 10.1038/ncomms6208.

Reynolds MG. 2000. Compensatory evolution in rifampin-resistant Escherichia coli. Genetics. 156:1471–1481.

Schmid-Hempel P. 2008. Parasite immune evasion: a momentous molecular war. Trends in Ecology & Evolution.

Sniegowski PD, Gerrish PJ. 2010. Beneficial mutations and the dynamics of adaptation in asexual populations. Philosophical Transactions of the Royal Society B: Biological Sciences. 365:1255–1263. doi: 10.1126/science.285.5426.422.

Sokurenko EV, Hasty DL, Dykhuizen DE. 1999. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. Trends in Microbiology. 7:191–195.

Stein RR et al. 2013. Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota Mering, Von, C, editor. PLoS Comput Biol. 9:e1003388. doi: 10.1371/journal.pcbi.1003388.s007.

Strelkowa N, Lässig M. 2012. Clonal interference in the evolution of influenza. Genetics.

Trindade S et al. 2009. Positive Epistasis Drives the Acquisition of Multidrug Resistance Zhang, J, editor. PLoS Genet. 5:e1000578. doi: 10.1371/journal.pgen.1000578.s005.

Trindade S, Sousa A, Gordo I. 2012. Antibiotic resistance and stress in the light of Fisher's model. Evolution.

Vogwill T, MacLean RC. 2014. The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. Evol Appl. 8:284–295. doi: 10.1111/eva.12202.

Weinreich DM, Knies JL. 2013. Fisher's geometric model of adaptation meets the functional synthesis: data on pairwise epistasis for fitness yields insights into the shape and size of phenotype space. Evolution. 67:2957–2972. doi: 10.1111/evo.12156.

Wilson BA, Garud NR, Feder AF, Assaf ZJ, Pennings PS. 2016. The population genetics of drug resistance evolution in natural populations of viral, bacterial and

eukaryotic pathogens. Mol Ecol. 25:42–66. doi: 10.1111/mec.13474.

Winstanley C, O'Brien S, Brockhurst MA. 2016. Pseudomonas aeruginosaEvolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections. Trends in Microbiology. 1–11. doi: 10.1016/j.tim.2016.01.008.

Wiser MJ, Ribeck N, Lenski RE. 2013. Long-term dynamics of adaptation in asexual populations. Science. 342:1364–1367. doi: 10.1126/science.1243357.

Woods RJ et al. 2011. Second-Order Selection for Evolvability in a Large Escherichia coli Population. Science. 331:1433–1436. doi: 10.1126/science.1198914.

World Health Organization. 2014. Antimicrobial Resistance - Global Report on Surveillance. http://apps.who.int/iris/bitstream/10665/112642/1/9789241564748_eng.pdf.

Wynn TA, Chawla A, Pollard JW. 2013. Macrophage biology in development, homeostasis and disease. Nature.

Xavier JB, Foster KR. 2007. Cooperation and conflict in microbial biofilms. Proceedings of the National Academy of Sciences. 104:876–881. doi: 10.1073/pnas.0607651104.

Yang L et al. 2011. Evolutionary dynamics of bacteria in a human host environment. Proceedings of the National Academy of Sciences. 108:7481–7486. doi: 10.1073/pnas.1018249108/-/DCSupplemental.

Zhang H et al. 2013. Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nature Genetics. 45:1255–1260. doi: 10.1038/ng.2735.