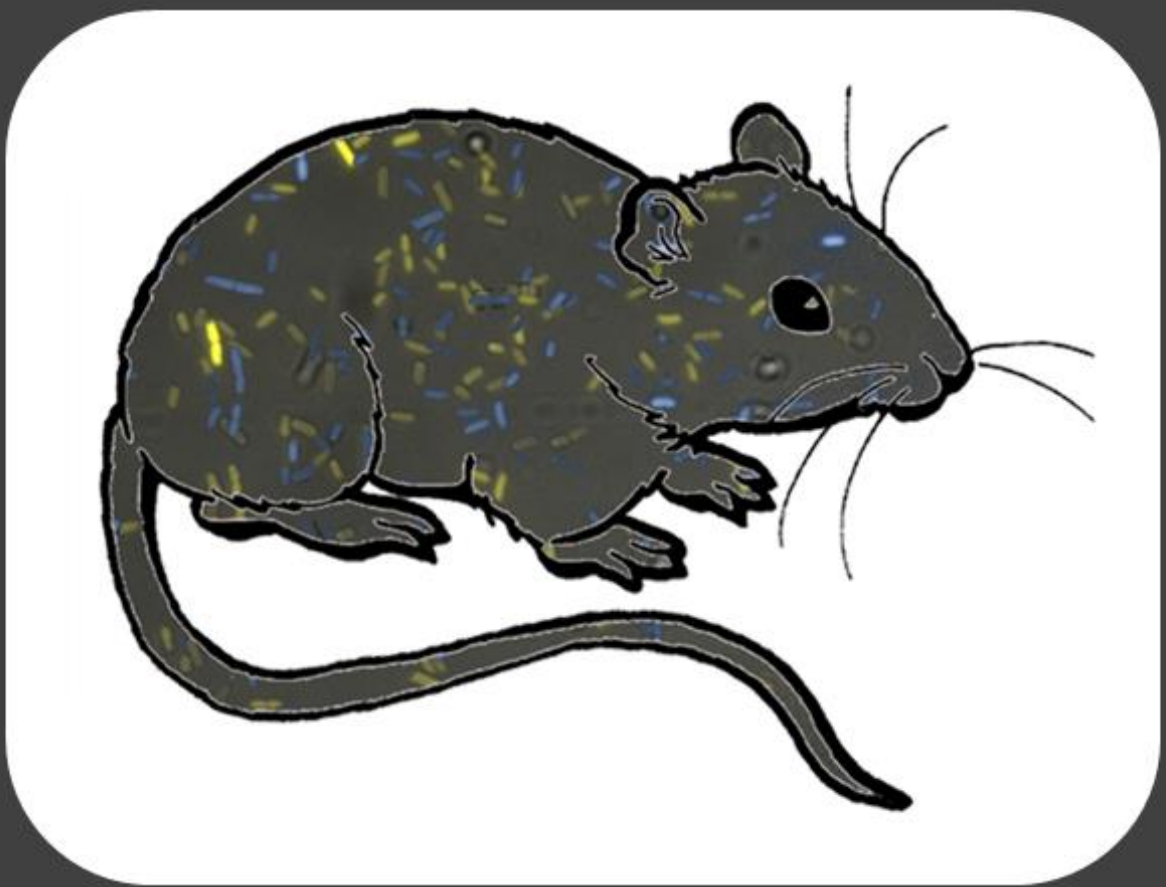
Adaptation of *Escherichia coli* to the mouse gut

João Miguel Barroso Batista



Dissertation presented to obtain the Ph.D degree in Evolutionary Biology

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

Oeiras,
October, 2016

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER /UNL

Knowledge Creation

Adaptation of *Escherichia coli* to the mouse gut

João Miguel Barroso Batista

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,

October, 2016

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER /UNL

Knowledge Creation

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/80257/2011 awarded to João Miguel Barroso Batista

Research work developed in coordination with

Instituto Gulbenkian de Ciência

Supervisor: Dr. Isabel Gordo

Co-supervisor: Dr. Jocelyne Demengeot

Acknowledgements

As in any other work, this thesis would not be possible without the contribution of several people. And for those, that one way or another helped me during my PhD, I would like to say thank you.

To begin, I want to thank my supervisor, Isabel Gordo. For giving me opportunity after opportunity, starting with my Master's project, then the internship and finally the PhD. It has been a long journey indeed, fruitful for both of us (I hope!). It was for sure a rewarding experience for me. Thank you for everything you taught me, for trusting and believing in me and for showing me how to be a better scientist. I like to think that your vision and your ability to see the big picture, your passion for your work and your permanent enthusiasm (despite my S. Tomé skepticism) are features that were somehow transmitted during these years, for they are invaluable skills for a good scientist.

I also would like to thank Jocelyne Demengeot, my co-supervisor. Thank you for bringing a different and interesting perspective to any scientific (and usually evolution-biased) discussion. I learned a lot with you and not only about immunology. Thank you for all the support, the good advice, and your general contribution during these years.

In addition, my thanks to Karina Xavier, third member of my thesis committee and collaborator. Thank you for considering me your "non-official" student and for always being available to discuss my project or my latest unexpected results.

To present and past members of my group, the Evolutionary Biology group. For everything that I learned from every single person, for all the shared knowledge and the interesting discussions. But also for those fun times spent eating a snack or drinking a beer. A particular thanks to my colleagues from the "caganita club", for all your ideas and feedback as well as the occasional help with mice experiments. To the members of the Lymphocyte Physiology group, always ready to teach me a new technique or to explain me some immunological stuff. Also, to present and past members and collaborators of the Bacterial Signaling group, for all the shared reagents or protocols and interesting discussions.

To my colleagues at IGC and ITQB, for all the shared meals and drinks, for the beer-hours and other enjoyable moments during these years.

For my friends, because decade-old friendships mean a lot. Thank you for all the diners, the parties and the get-togethers. Especially to those closer to me, thank you for the riverside snacks or the seaside lunches, the burger days or the viewpoints on Sunday. For all your support, in particular during this last part of my PhD. For all the times that I needed and you were there. For not giving up and for not letting me give in.

Finalmente, à minha família, em particular aos meus pais e irmão. Por todo o apoio e incentivos, pela paciência e orgulho. Por tudo isto e muito mais, obrigado.

Summary

The adaptation of organisms to the environment is a pervasive process in nature, which has been studied from the theoretical and experimental perspective over the years. Despite the plethora of *in vitro* studies addressing this subject, the adaptive process of individuals to a relevant natural environment is an area less explored. The mammalian gut, as a natural habitat inhabited by a complex cohort of microorganisms named the microbiota, represents a relevant host environment to study the adaption of commensal bacteria. However, the dynamics of adaptation of a commensal species in this environment are still poorly studied.

In this thesis we investigated the evolution of *Escherichia coli* in one of its natural habitats, the mouse gut. In **Chapter 2** we used a neutral marker system that allows tracking of adaptive events over time, to follow the adaptation of *E. coli* when colonizing the gut of streptomycin-treated mice. Over one month of colonization we detected fast adaptation to the gut environment, characterized by multiple adaptive mutations competing for fixation, a phenomenon called clonal interference. We described the genetic basis of *E. coli* adaptation to the mouse gut and revealed the striking level of parallelism present in independently evolved populations. We further evidenced the huge amount of intraspecific genetic diversity that can be maintained even during a phenotypic replacement.

In **Chapter 3** we addressed the contribution of a host factor on the evolution of *E. coli* colonizing the mouse gut. In line with the hypothesis of co-evolution between host immunity and their microbiota, we evaluated the influence of adaptive immunity on the evolution of a commensal species by comparing the adaptation of *E. coli* in immune-competent or immune-compromised animals. We found a slower rate of adaptation in immune-compromised mice, due to smaller and more variable effects of beneficial mutations, presumably modulated by the gut microbiota. Together with our findings of differentially selected genetic targets of adaption these results suggest that adaptive immunity alters the pace and predictability of *E. coli* adaptation to the gut.

Chapter 4 further characterizes the evolution of *E. coli* in the mouse gut by exploring the second steps of adaptation to this environment, consisting in the adaptive dynamics of an *E. coli* strain initially carrying a beneficial mutation (*gat*

mutant). We elucidated the different forms of selection that can coexist in this environment by estimating the selective effects of beneficial mutations through competitive fitness assays or analysis of their dynamics. We further demonstrated a case of reverse evolution, comprising a phenotypic reversion of the *gat*-negative phenotype.

The results presented in this thesis provide novel insights to our understanding of the microbe-microbe and host-microbe interactions that take place in the complex gut environment and are fundamental to the maintenance of the host homeostasis and health.

Resumo

A adaptação dos organismos ao ambiente é um processo ubíquo na natureza, que tem sido estudado ao longo dos anos segundo um ponto de vista teórico e experimental. Apesar dos vários estudos realizados *in vitro* que abordam este tema, o processo adaptativo dos indivíduos a ambientes naturais relevantes é uma área menos explorada. O intestino dos mamíferos, um ambiente natural habitado por uma complexa comunidade de microorganismos, denominado microbiota, representa um ambiente relevante no hospedeiro para estudar a adaptação de bactérias comensais. No entanto, as dinâmicas de adaptação de espécies comensais neste ambiente são ainda pouco estudadas.

Nesta tese estudámos a evolução da bactéria *Escherichia coli* num dos seus ambientes naturais, o intestino do ratinho. No **Capítulo 2** usámos um sistema de marcador neutro que permite identificar eventos adaptativos ao longo do tempo, para seguir a adaptação da *E. coli* quando coloniza o intestino de ratinhos tratados com estreptomicina. Durante um mês de colonização, detectámos adaptação rápida ao ambiente do intestino, caracterizada por múltiplas mutações competindo para se fixarem, um fenómeno chamado interferência clonal. Descrevemos a base genética da adaptação da *E. coli* ao intestino do ratinho e revelámos um impressionante nível de paralelismo presente em populações que evoluíram independentemente. Adicionalmente, também demonstrámos a enorme quantidade de diversidade intra-específica que pode ser mantida mesmo durante uma substituição fenotípica.

No **Capítulo 3** investigámos a contribuição de um factor do hospedeiro na evolução da *E. coli* quando coloniza o intestino do ratinho. De acordo com a hipótese de co-evolução entre o sistema imune do hospedeiro e o seu microbiota, avaliámos a influência da imunidade adaptativa na evolução duma espécie comensal através da comparação da adaptação da *E. coli* em ratinhos imunocompetentes ou imunocomprometidos. Encontrámos uma taxa de adaptação mais lenta em ratinhos imunocomprometidos, devido a menores e mais variáveis efeitos das mutações benéficas, presumivelmente modulados pela microbiota do intestino. Juntamente com a nossa descoberta de alvos de adaptação diferencialmente seleccionados, estes resultados sugerem que a

imunidade adaptativa altera o ritmo e a previsibilidade da adaptação da *E. coli* ao intestino.

O **Capítulo 4** caracteriza mais aprofundadamente a evolução da *E. coli* no intestino do ratinho, explorando os segundos passos de adaptação a este ambiente, que consistem nas dinâmicas adaptativas de uma estirpe de *E. coli* contendo inicialmente uma mutação benéfica (mutante *gat*). Elucidámos as diferentes formas de selecção que podem coexistir neste ambiente estimando os efeitos selectivos de mutações benéficas através de ensaios de competição de *fitness* ou análise das dinâmicas destas mutações. Adicionalmente revelámos um caso de evolução reversa, que consiste numa reversão fenotípica do fenótipo negativo *gat*.

Os resultados apresentados nesta tese fornecem novas perspectivas para a compreensão das interacções micróbio-micróbio e hospedeiro-micróbio que existem no complexo ambiente do intestino e são fundamentais para a manutenção da homeostasia e saúde do hospedeiro.

Table of contents

Thesis Outline

In this thesis we investigated the evolution of *Escherichia coli* in one of its natural environments, the mouse gut. In particular, we characterized the adaptive process of *E. coli* in this complex environment and addressed the contribution of one host factor, the adaptive immunity, to the evolution of this commensal species.

Chapter 1 broadly introduces the concept of adaptation of organisms to the environment, and provides examples of *in vitro* and *in vivo* studies of evolution. In addition, the mammalian gut is presented as a relevant natural habitat where multiple factors, including the microbiota and the host immune system interact. Finally, previous studies of *E. coli* in the mouse gut are reviewed.

Chapter 2 describes the adaptive process of *E. coli* when colonizing the gut of streptomycin-treated mice, as well as the genetic basis of adaptation, and demonstrates the importance of clonal interference in this environment.

Chapter 3 addresses the contribution of adaptive immunity to the evolution of *E. coli* in the mouse gut by comparing the adaptive process in immune-competent and immune-compromised mice.

Chapter 4 focuses on the second steps of adaptation to the mouse gut and details the different forms of selection acting in this environment, including a case of reverse evolution.

Chapter 5 compares the results presented in the previous chapters with other studies, and integrates this new data with the current knowledge on the complex host-microbe relationships taking place in the host environment.

CHAPTER 1

Introduction

1.1 – Adaptation of organisms to the environment

Adaptation of organisms is a continuous process, inherent to all living populations. Given its biological relevance, numerous efforts have been made, at both theoretical and experimental levels, to clarify the mechanisms by which evolution occurs. Nonetheless, several key questions remain open, regarding for example, the extent and limits of adaptation, the different regimes that it can follow or its genetic determinants. Importantly, although *in vitro* experimental evolution studies have provided valuable insights into this complex process, much less is known about adaptation of organisms in more complex and more natural habitats. Thus, the study of how adaptation depends on the environment remains one of the most interesting research subjects.

1.1.1 – The basic processes in evolution

The evolution of populations is shaped by different basic processes that change the frequency of alleles, including mutation, natural selection, genetic drift, migration and recombination (Olson-Manning et al., 2012).

Genetic mutations, defined as random alterations in the genome, can change the DNA sequence to different extents, depending on the type of mutation (Gordo et al., 2011). For example, single nucleotide polymorphisms (SNPs), are changes in individual nucleotides, while deletions and insertions consist in elimination or addition, respectively, of one or more nucleotides. A special case of insertions includes mobile genetic elements, which can move within the genome. For example insertion sequences (IS) elements, a class of transposable elements, are small DNA sequences that do not carry any accessory genes (Mahillon and Chandler, 1998). Other types of mutations comprise inversions or duplications of portions of the genome. While mutations are random, these may have fitness effects or may be neutral, depending on the type of mutation and the region of the genome they occur (Loewe and Hill, 2010). Importantly, multiple studies on the distribution of fitness effects of mutations have shown that the majority of mutations are either neutral or deleterious to the fitness of individuals, with few being beneficial (Orr, 2010). These beneficial or adaptive mutations are those that ultimately contribute to the adaptations of organisms to the environment.

However, when mutations first arise in the populations, they are subjected to the effects of genetic drift. Genetic drift, defined as random oscillations of allele frequencies in the population, is a stochastic process with increased relevance in smaller populations (Kimura and Ota, 1969). On the other hand, natural selection alters the frequency of mutations depending on their selective effects (Olson-Manning et al., 2012). Gene flow, characterized by migration of individuals between populations, and recombination, or genetic exchange between different individuals, are two others processes that can alter the genetic structure of the populations.

1.1.2 – Mechanisms of selection

The impact of natural selection on allele frequency depends on the types of mutations subjected to selection. Thus, deleterious mutations are eliminated from the populations, by negative or purifying selection, while beneficial mutations are subjected to positive selection, increasing in frequency in the populations (Loewe and Hill, 2010). The rate at which mutations increase or decrease in frequency depends on their fitness effect (Gordo et al., 2011). When the rate of emergence of beneficial mutations is low, an adaptive mutation tends to increase in frequency and reach fixation (selective sweep), replacing the ancestral genotype and purging genetic variability from the population (Barrick and Lenski, 2013). This form of selection, called periodic selection (Atwood et al., 1951), can occur when the population size and/or mutation rate are small. However, when the population size is large and genetic recombination is absent, such as populations of microorganisms, many adaptive mutations arise in the population and compete for fixation, a phenomenon named clonal interference (Gerrish and Lenski, 1998). In these conditions, adaptive mutations take longer to get fixed and genetic variability can be maintained in the population for longer periods. Another mechanism that can maintain variation in populations is negative frequency-dependent selection, where the selective effect of a mutation varies with its frequency (Levin, 1988). In this regime, beneficial mutations are advantageous when rare but less advantageous or, in extreme cases, even deleterious (negative frequency-dependent selection) when at high frequency. These different scenarios (Figure 1)

have been observed to different extents in populations of organisms, including microorganisms such as bacteria (Maddamsetti et al., 2015).

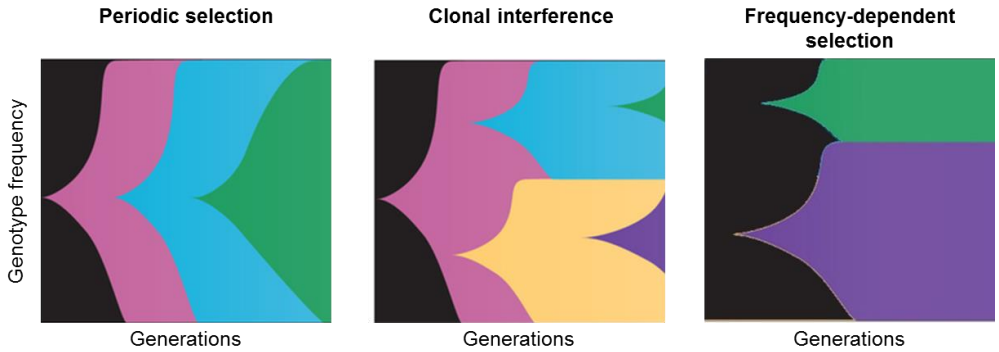


Figure 1. Regimes of selection. Frequencies of genotypes are showed in colored segments. In periodic selection, beneficial mutations get fixed sequentially, sweeping through the population. In clonal interference, many mutations arise, competing for fixation. In the third scenario, mutations are maintained at a given frequency by frequency-dependent selection. Figure adapted from Barrick and Lenski. 2013.

1.2 – Tools for the study of adaptation

In an effort to study the adaptation of organisms to the environment, different types of strategies have been adopted, including experimental evolution studies of populations in laboratory conditions or the observation of natural populations (Bailey and Bataillon, 2016). While studies of experimental evolution are performed using model organisms that are propagated in a multiplicity of controlled *in vitro* setups, the study of natural populations is based on the analysis of genetic data collected from different individuals and/or over time. A third complementary approach on evolutionary research focuses on mathematical modeling, including the evolution of digital organisms (Hindré et al., 2012). This approach aims not only to identify the determinants of evolutionary change but also to provide a theoretical framework whose assumptions can be tested or evaluated in real populations.

1.2.1 – Model organisms in experimental evolution

Model organisms are usually strains of relatively simple species that can be experimentally tractable and thus used to study in detail biological processes

(Dietrich et al., 2014). Importantly, the findings obtained with these model organisms can then be extrapolated to other organisms, including humans. While model organisms may not always be representative of their wild relatives or even the species they belong, they do have the particularity of possessing certain traits, absent in close non-model organisms, which facilitated their domestication and adaptation to laboratory environments (Alfred and Baldwin, 2015). Examples of organisms used in studies of experimental evolution include the fruit fly *Drosophila melanogaster*, the worm *Caenorhabditis elegans*, the yeast *Saccharomyces cerevisiae* and the bacterium *Escherichia coli*, among others.

1.2.2 – *E. coli*: from nature to the lab

As one of the most well-known and extensively used model organisms, *E. coli* has been the focus of many studies of experimental evolution. This Gram-negative bacillus was first described by Theodor Escherich in 1884, upon isolation from human feces (Blount, 2015). Easily isolated from human hosts and able to grow extremely well in different culture media, *E. coli* rapidly rose as a preferred model organism in biology. Although its main habitat is the mammalian gut, *E. coli* can also be found in other environments, such as the gut of other animals (for example, birds, reptiles and fish), soil and water (Blount, 2015). In the human gut, *E. coli* is one of the most common aerobes, although its frequency does not exceed 1% (Berg, 1996). In addition, the human-associated *E. coli* population includes both transient and long-term resident strains (Blount, 2015).

E. coli strains commonly used as model organisms in laboratory setups include strains B, C, W and the most widely used laboratory strain, K-12. This strain was first isolated from a patient recovering from diphtheria (Bachmann, 1996) but due to prolonged *in vitro* propagation has lost several features present in wild strains of *E. coli*. For example, this strain lacks not only the O antigen but also genetic elements such as the lambda phage and the F plasmid (Hobman et al., 2007). MG1655 is a derivative K-12 strain, which has been maintained in laboratory conditions with minimal genetic manipulation and thus can be considered a “wild-type” lab strain of *E. coli* (Blattner et al., 1997). Similarly to other described strains of *E. coli*, the genome of this strain has been sequenced,

which allowed the development of a multiplicity of genetic tools, including gene expression and genetic manipulation. All these features make *E. coli*, and in specific K-12 strain, an excellent model organism in studies of experimental evolution.

1.2.3 – Neutral marker system in the study of adaptation

As a valuable tool in experimental evolution, neutral markers have been used to track adaptive events as they occur. In bacteria, this system involves labelling an isogenic bacterial population with two or more genetic neutral markers and following the frequency of the markers over time (Lang and Desai, 2014). Deviation from the initial marker frequency is attributed to the appearance and expansion of a beneficial mutation in the population, since the markers hitchhike with beneficial mutations (Hegreness et al., 2006).

Markers used in evolution experiments are preferentially neutral in the environment and thus are not expected to be under selection. In addition, they should be easy to screen, leading to a phenotype that allows distinction of different lineages (Blundell and Levy, 2014). In particular, fluorescent markers have been successfully used as genetic markers due to their specific features (Day and Davidson, 2009). Genes coding for fluorescent proteins are completely foreign to the bacterial genome and thus their expression can be tightly controlled. Usually under the control of a strong promoter, fluorescent proteins are constitutively expressed by bacterial cells, effectively labelling the cell with a given fluorescence. This fluorescent signal can be detected using a fluorescent detector, such as a flow cytometer device, allowing the screening of a large number of cells in a time-effective way.

1.2.4 – Theoretical modeling of adaptation

Another advantage of neutral marker systems is that they can also be used, together with mathematical modeling, to infer parameters that characterize the evolutionary process, such as the distribution of fitness effects of beneficial mutations and the rate that they are produced in the evolving populations. Given that these are important parameters to understand the adaptation of populations to

a certain environment, several methods have been developed that try to estimate these parameters from adaptive dynamics of neutral markers. These methods usually rely on the simulations of digital populations with known parameters that are then compared with real populations (Figure 2).

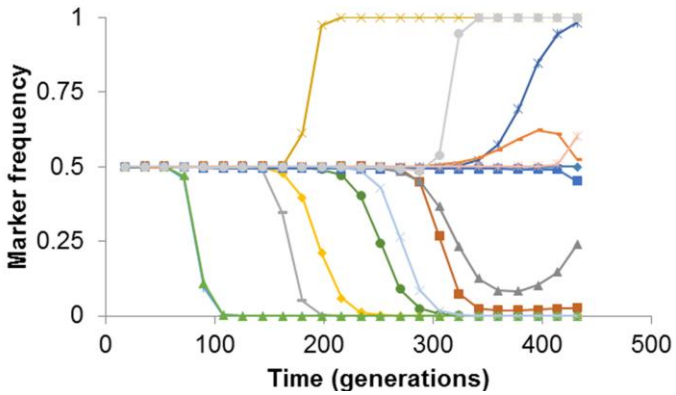


Figure 2. Neutral marker divergence dynamics in simulated populations. Each line represents a simulated evolving population based on a theoretical model of simple Darwinian selection that assumes an effective population size (N_e) of 10^7, a beneficial mutation rate (U_b) of $4.5x10^{-9}$ and a mean selective effect (s_b) of 0.10.

For example, one of such methods (Hegreness et al., 2006) compares evolutionary parameters (that account for the rate of initial divergence of the marker and the waiting time until divergence) extracted from simulated and experimentally-obtained marker divergence trajectories to infer the beneficial mutation rate and effect of the first beneficial mutation. This method has been used to estimate these parameters in different mutant strains of *E. coli* (Barrick et al., 2010). Another method relies on a maximum likelihood approach to infer the selective effects and establishment times of beneficial mutations based on changes in the frequency of the marker over time (Illingworth and Mustonen, 2012). As the previous method, it is based on neutral marker frequencies over time but has the advantage of taking into account the whole adaptive dynamic of a population, without discarding marker frequencies at later time-points. This method assumes the occurrence of beneficial mutations which establishment times and effects maximize the probability of observing the marker frequency data. Models assuming different numbers of beneficial mutations are then compared based on their likelihood score and the simplest possible model that can better explain the observed marker dynamics is chosen. With this methodology is thus possible to determine the fitness of multiple mutant haplotypes segregating in the population, or a distribution of haplotype fitnesses, with the minimal number of mutations that can explain the marker frequency dynamics.

1.3 – *In vitro* evolution of microorganisms

Important insights on the nature of adaptation have been acquired by performing studies of experimental evolution in laboratory conditions (Barrick and Lenski, 2013). Although some of these studies have been performed in eukaryotes (such as flies and fungi) (Kawecki et al., 2012), prokaryotic organisms, and in particular bacteria, have been the major focus of these works. Because bacteria have a short generation time, large population size and are easy to maintain in controlled conditions, these organisms are suitable candidates to follow adaptation in real time (Elena and Lenski, 2003). Moreover, the genome sequence of many bacterial strains has been described and a multiplicity of genetic tools is available, allowing their genetic manipulation. In fact, the recent advances in whole genome sequencing (WGS) led researchers to increasingly adopt an "evolve and resequencing" approach to uncover the genetic mechanisms of adaptation (Long et al., 2015)

Typically, replicate populations are propagated in a given *in vitro* environment for a variable number of generations and samples of the evolving populations are collected periodically and stored as a "fossil record". The increase in fitness is then detected by competitions between samples of the evolved populations and the ancestral (Elena and Lenski, 2003). Another type of analysis usually performed includes genome sequencing of evolved clones or population samples (Long et al., 2015). By comparing the genomes of ancestral and evolved bacteria it is possible to identify *de novo* mutations present in evolved genomes that are likely contributing to the adaptive process. Common to the majority of *in vitro* evolution experiments is the observation of fast adaptation to the environment, with the appearance of phenotypic innovations, high levels of parallelism and also considerable diversity within populations (Hindré et al., 2012).

In experimental evolution, bacterial populations can be propagated either in batch culture, with periodic cycles of growth, dilution and re-inoculation, or maintained in continuous culture, where resources are provided and a fraction of the population removed at a certain rate (chemostats) (Figure 3).

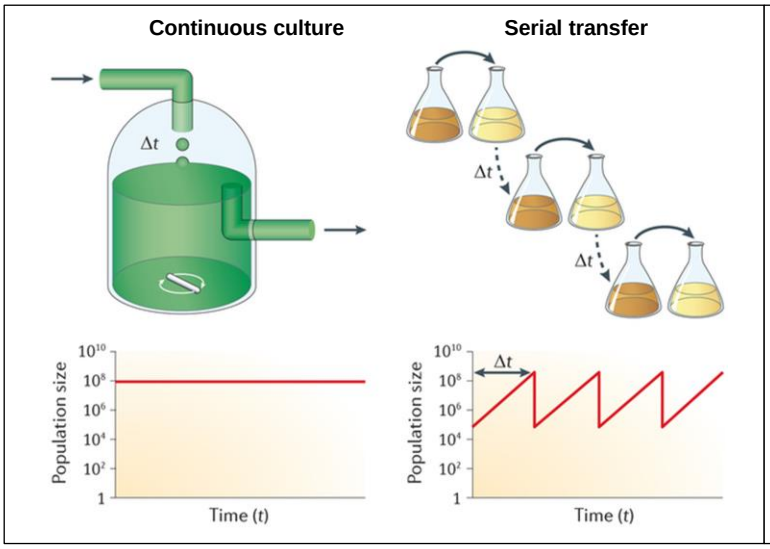


Figure 3. Propagation of cultures in experimental evolution. Bacterial populations are either maintained in a chemostat, with constant inflow of nutrients and an outflow of individuals and waste, or propagated in serial transfer, where a portion of the population is transferred to new medium that supports a new bout of growth (batch culture). Figure retrieved from Barrick and Lenski, 2013.

1.3.1 – Adaptation of *E. coli* to laboratory environments

The most well-known *in vitro* evolution experiment was initiated in the laboratory of Richard Lenski, where 12 populations of *E. coli* have been propagated in batch culture for more than 60 000 generations in a simple minimal medium supplemented with a limiting concentration of glucose (long term evolution experiment). As the longer *in vitro* evolution study to date, this experiment has provided invaluable data to better understand the dynamics and determinants of the evolutionary process. For example, a general pattern emerging from the analysis of competitive fitness assays between ancestral and evolved populations revealed a rapid increase in fitness, with deceleration of fitness increase over time. Despite this observation, WGS of evolved lines showed the accumulation of presumably adaptive mutations at a fairly constant rate (Barrick et al., 2009) (Figure 4).

Other features uncovered by sequencing analysis of the evolved populations include the evolution of an increased mutation rate (Sniegowski et al., 1997) and the evolution of a new trait, the ability to grow on citrate (Blount et al., 2008). Large-scale chromosomal rearrangements involving recombination between insertion sequences were also detected in clones evolved for 40 000 generations (Raeside et al., 2014). In addition, phenotypic changes such as cell size (Philippe et al., 2009), altered gene expression (Cooper et al., 2008) and changes in

regulatory networks (Philippe et al., 2007) were detected throughout the evolution experiment. Interestingly, reconstruction of the populations history allowed to detect different regimes of adaptation, including the signature of selective sweeps, clonal interference between different lineages or maintenance of variability by negative frequency-dependent interactions (Maddamsetti et al., 2015). Thus, the establishment of complex interactions between individuals and different adaptive strategies even in a seemingly simple *in vitro* environment highlights the remarkable adaptive potential of bacteria.

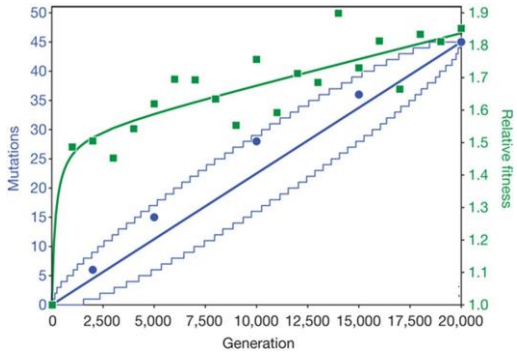


Figure 4. Rates of genomic evolution and fitness improvement in a long-term experiment with *E. coli*.
In blue are represented the number of mutations in evolved clones (circles) and a model where mutations accumulate linearly over time (line). In green are represented mean population fitness increments over time (squares) and a hyperbolic plus linear fit of this trajectory (line). Figure adapted from Barrick et al., 2009.

Studies of microbial adaptation have also been conducted in continuous culture, where *E. coli* populations were propagated in chemostats containing a growth-limiting nutrient. A common observation in this type of experiments is the improvement of growth rate in the nutrient-limited chemostat and altered metabolic strategies (Gresham and Hong, 2015). Although early studies performed in chemostats seemed to support periodic selection as the main mechanism through which adaptation occurred, later studies showed that clonal interference was pervasive also in this environment (Maharjan et al., 2015). Works in glucose-limited chemostats reported rapid diversification of the populations, with phenotypic convergence, but high heterogeneity at the genetic level (Maharjan et al. 2012). Furthermore, the mutations identified targeted preferentially nutrient-specific transporter genes. Other features of these adaptation studies comprise the appearance of mutator alleles (Maharjan et al., 2013) and the establishment of negative frequency-dependent selection interactions (Maharjan et al., 2012). In addition, transposable elements played an important role in driving adaptive

evolution to glucose-limited chemostats (Gaffé et al., 2011), although its importance in other nutrient-limited chemostats is still poorly known.

1.3.2 – Limitations of *in vitro* studies

Although *in vitro* studies of experimental evolution have provided important insights into the process of adaptation, evolution of natural populations rarely (if ever) occurs in strictly defined and simple conditions as those encountered in laboratory. In fact, in nature, organisms are subjected to multiple selective pressures, resulting from interactions not only with the environment but also with other organisms. All these factors that alter the habitat and thus are likely to play a role in the adaptive process, are however not considered when adapting populations to artificial *in vitro* environments. Given these reasons, biological processes that are elucidated and thoroughly described *in vitro* may occur differently in natural environments and thus have a different relevance in these conditions (Bailey and Bataillon, 2016).

1.4 – Dynamics of mammalian associated microorganisms

Although experimental evolution studies with bacteria have been performed in controlled *in vitro* conditions for quite some time, the study of the adaptive process in natural conditions has been addressed only more recently. The advances in the field of genome sequencing allowed the genetic characterization of isolates from hosts over time, thus tracking the adaptive evolution of bacteria in the context of the host. These host-associated bacteria can be classified depending on the type of interaction established with the host.

Classic pathogens can be defined as microorganisms that are able to cause disease in a healthy host, while opportunistic pathogens display a restricted pathogenic ability, requiring immune or microbial impairment of the host to cause disease (Hornef, 2015). Commensal bacteria are considered nonpathogenic organisms that colonize the host and may provide an advantage, though usually no identified benefit for the host is provided by the association (Ghosh, 2013). In the cases where a true mutualistic relationship is established, in which both host

and bacteria benefit from this interaction, the microorganism can be classified as symbiont. A more recently described group of microorganisms comprises the pathobionts, bacteria with pathogenic potential that cause disease indirectly, by stimulation of the host immune system. Thus, pathobionts are able to promote both immune maturation or inflammation, depending on the state of the host (Hornef, 2015). It is important to note that classification of bacteria as pathogens or commensals is not always clearly defined. For example, several bacterial pathogens, such as *Staphylococcus aureus*, *Helicobacter pylori* or *E. coli*, are usually considered commensals in humans (Didelot et al., 2016), since although carrying pathogenicity genes, these are not always expressed (asymptomatic carrier state) (Ghosh, 2013).

Generally, *in vivo* surveys of within-host evolution of host-associated bacteria have been mainly conducted for pathogens in the context of bacterial infections, given their relevance and impact in human health.

1.4.1 – Evolution of pathogens and opportunists

Through WGS of bacterial samples collected in the context of microbial infections it has been possible to obtain a detailed genetic structure of pathogen populations. This type of analysis, focusing on the intra-specific diversity of pathogens, has uncovered different evolutionary patterns, depending on the pathogen studied. For example, *Yersinia pestis* or *Salmonella typhi* are characterized by reduced genetic diversity, including mobile elements and nucleotide diversity (Bentley and Parkhill, 2015; Klemm and Dougan, 2016). Nonetheless, the genome of these organisms displays a high level of ISs and pseudogenes, i.e., genes inactivated by mutations. Interestingly, many of these inactivated genes were previously associated with virulence and interactions with the host, suggesting events of transition from broad-host-range to host-restricted pathogens (Bentley and Parkhill, 2015). Other pathogens display a considerable level of intra-specific diversity, with signatures of extensive horizontal gene transfer (HGT) of DNA and mobile elements. This is the case for *E. coli*, in which pathogenic strains harbor genes that are not present in other strains (Rasko et al.,

2008) and transition to pathogenesis is often associated with acquisition of virulence factors by HGT (Kaper et al., 2004).

Pseudomonas aeruginosa, a well-described opportunistic pathogen that colonizes the lungs of cystic fibrosis (CF) patients, has been the focus of several studies following the evolutionary dynamics of this bacterial species *in vivo*. In these studies, sequencing analysis of isolates collected from different patients and over a period of time has provided important information on the evolution of this pathogen, including estimates for the evolutionary rate, phylogeny of infectious bacterial strains and identification of accumulated mutations (Didelot et al., 2016) (Figure 5).

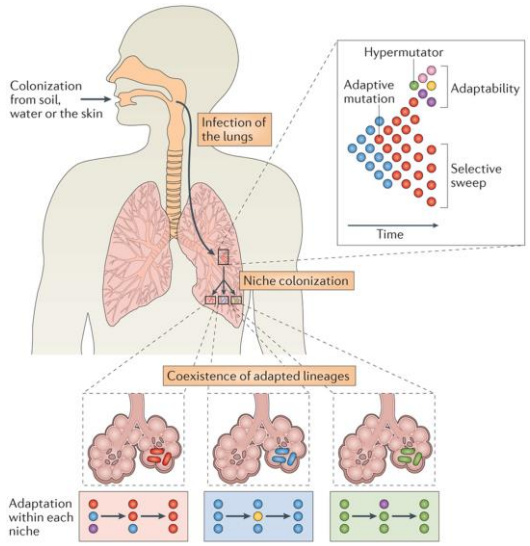


Figure 5. Within-host evolution of pathogens in the lungs of a patient with CF. Environmental acquired bacteria colonize the lungs and increase in population size, where adaptive mutations occur and sweep through the population. A common phenomenon is the appearance of mutators with increased evolutionary potential. As the infection progresses, pathogen niche specialization may occur, leading to the simultaneous existence of differentially adapted lineages. Figure retrieved from Didelot et al., 2016.

In one study, sequencing analysis of a collection of *P. aeruginosa* samples isolated over a period of 30 years from a group of Danish CF patients led to the identification of genetic mutations responsible for the transition from an opportunistic pathogen to a host-specific pathogen (Yang et al., 2011). In addition, this study revealed an initial period of rapid adaptation to the host followed by limited diversification of bacterial lineages. Another observation, common to other works (Feliziani et al., 2014), was the presence of hypermutator alleles in some sub-lineages, that infected almost 50% of the patients sampled (Marvig et al., 2013). An interesting result obtained from a longitudinal analysis of samples collected from young patients over a five-year period was the high level of parallelism in more than 50 genes across different lineages (Marvig et al., 2015).

This convergent molecular evolution targeted regulatory networks, central metabolism, antibiotic resistance and virulence factors, all features related with adaptation to the host environment.

Studies in *S. aureus* have also been performed to investigate bacterial evolution within the host both in the status of asymptomatic nasal carriage or blood infection. For example, WGS analysis of isolates from a single time point of 13 hosts (Golubchik et al., 2013) supported the idea that evolutionary dynamics of *S. aureus* in asymptomatic carriers are mainly characterized by purifying selection, with just a few adaptive events identified. Another work, that followed the progression from bacterial carriage to blood infection (Young et al., 2012) revealed that this transition was associated with only eight mutations, half of them leading to loss of function.

As a general pattern of these phylogenomic studies of bacterial pathogens, bacterial evolution *in vivo* involves not only adaptation to the host environment but also antimicrobial resistance. Unsurprisingly, antibiotic usage has driven a strong selection for antibiotic resistance on microbial pathogens, which determinants are frequently encoded in mobile DNA and thus susceptible to be horizontally transmitted (Klemm and Dougan, 2016).

1.4.2 – Evolution of commensals

Although characterization of pathogens dynamics *in vivo* provides interesting insights into the adaptation of bacteria in the context of infections, it is unlikely that these dynamics are the same for the majority of bacteria associated with the host. While a complete functional characterization of host-associated microbes has not yet been elucidated, it is clear that many of the species classified as commensals can establish interactions with the host or other microbes, contributing to the maintenance of host homeostasis and health. These interactions are expected to differ markedly from those described for pathogens and thus different selective pressures should play a role in these two scenarios. For example, the response of the host immune system against a pathogen or a commensal can be quite distinct, as are the numbers (population size) reached by invading microbes or stable inhabitants of the microbiota. All these factors presumably impact the adaptive

dynamics of the microbes, which thus should differ in a context of pathogenesis or commensalism. In fact, understanding how commensals evolve and adapt to the *in vivo* environment of the host is an important subject, especially in the light of the co-evolution hypothesis between host and microbiota (Ley et al., 2008a).

Though some of the works following the evolution of pathogens also encompass the asymptomatic stage before transition to infectious (Golubchik et al., 2013), within-host evolution of commensal bacteria, in particular a longitudinal characterization of adaptation in commensals is a subject that has been poorly explored. However, some attempts have been made to better understand the evolution of *Lactobacillus reuteri*, a mutualist inhabitant of the mammalian gut (Walter et al., 2011), through a comparative phylogenetic approach. Comparative sequencing analysis of strains isolated from different hosts revealed the existence of monophyletic, host-specific clades, suggesting that evolution of *L. reuteri* was characterized by diversification into host-adapted lineages (Oh et al., 2010) (Figure 6).

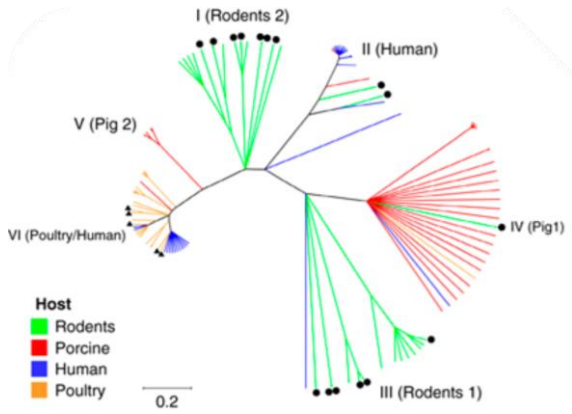


Figure 6. Host specialization in *L. reuteri*.
Genealogy of *L. reuteri* strains inferred by a coalescence Bayesian method (ClonalFrame), based on three loci. The scale bar displays time, in coalescent units.
Circles and triangles represent isolates from rats or turkey, respectively. Colors represent different host origin. Figure retrieved from Oh et al., 2010.

Supporting this hypothesis, rodent isolates displayed a strong ecological performance in these hosts, while isolates from other hosts did not. Further analysis of these strains identified genes specific to the intestinal environment of the different hosts, in particular those essential for ecological performance in the rodent gut (Frese et al., 2011).

1.4.3 – Dynamics of commensal and pathogenic *E. coli*

E. coli represents a particular case; primary a commensal inhabitant of the gut microbiota, can sometimes adopt the role of a pathogen, causing intra and extraintestinal diseases (Kaper et al., 2004). Strains of *E. coli* have been isolated from many vertebrate species, including mammals, reptiles and birds, as well as secondary habitats outside the host, such as the soil (Tenaillon et al., 2010). Traditionally considered a gut commensal, some evidence suggests that *E. coli* can establish a mutualistic relationship with the host, for instance by providing vitamins required by the host (Blount, 2015). In addition, by consuming oxygen, *E. coli* promotes the maintenance of an anoxic environment favored by the anaerobes and can also contribute to colonization resistance by competitive exclusion of pathogens (Blount, 2015). On the other side of the spectrum, certain strains of *E. coli* can become opportunistic pathogens when colonizing sick or impaired hosts, while others have the ability to cause disease in healthy hosts, and are thus considered pathogenic. These strains can be grouped into three pathotypes, depending on the associated pathology: enteric/diarrhoeal disease, urinary tract infection and sepsis/meningitis. Within these categories, pathogenic strains of *E. coli* can be further classified depending on the mechanisms of pathogenicity or body region they affect (e.g., enteropathogenic or uropathogenic *E. coli*) (Kaper et al., 2004).

Although the phylogeny of *E. coli* has suffered alterations over time, accompanying the development of genetic techniques, the bulk of data suggests the existence of four main phylogenetic groups (A, B1, B2 and D) and two accessory groups (C and E) (Chaudhuri and Henderson, 2012; Tenaillon et al., 2010). Although these phylogenetic groups (or phylogroups) may play distinct ecological roles, no clear association between phylogenetic groups and hosts has been identified. This almost absence of host-specific strains suggests that *E. coli* evolved to occupy niches within different hosts and secondary habitats (Tenaillon et al., 2010).

Reflecting the extensive strain diversity of *E. coli*, comparative genetic analysis of sequenced *E. coli* strains revealed a wide genomic plasticity and dynamism, with less than 20% of the genes shared among strains (Kaas et al.,

2012) (Figure 7). The remaining genes comprised a vast amount of genetic variation that could be acquired by HGT events, including prophages, transposable elements and accessory genes (Blount, 2015). Importantly, these genes can encode niche-specific fitness factors or pathogenic determinants (such as virulence factors). Interestingly, the presence of virulence factors has also been detected in commensal *E. coli* strains, suggesting that extraintestinal virulence may be a by-product of commensalism (Diard et al., 2010).

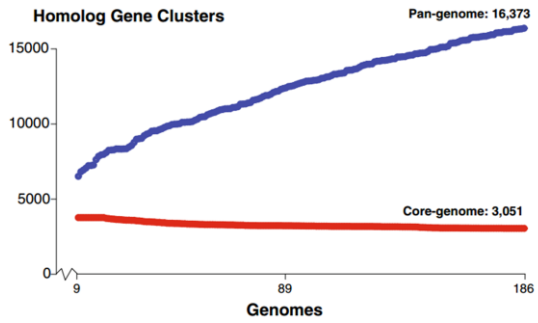


Figure 7. Genomic diversity of *E. coli*.
Gene clustering analysis of 186 *E. coli* strains revealed that the genes shared by all strains (core-genome, in red) correspond to less than 20% of the more than 16000 genes found in all strains (pan-genome, in blue). Figure retrieved Kaas et al., 2012.

A few studies have addressed the genetic structure of populations of *E. coli* colonizing its host, both in the context of commensalism and pathogenesis. Human hosts are commonly colonized over time by both transient and resident strains, in which resident strains are usually also dominant strains, found at high frequency in the host (Tenaillon et al., 2010). Works aiming to characterize the genetic diversity of *E. coli* isolated from human fecal samples found on average one to two genotypes per host (Alm et al., 2011), although an increased diversity of 3.5 strains was reported when analyzing biopsy samples from different regions of the lower intestinal tract (Gordon et al., 2015). While in this last study no evidence was found for region-specific phylogroups, intra-host analysis revealed that the distribution of phylogroups among intestinal samples varied depending on the phylogroup of the dominant strain. In a recent work, a collection of *E. coli* strains isolated from wild and domesticated animals, as well as humans, was analyzed in regard to the phylogroup composition (Smati et al., 2015). A variable prevalence of phylogroups, depending on the host species, was found, with a strong influence of domestication on the phylogroup diversity. The authors further defined three major clusters of *E. coli* animal commensal populations, or enterocolitypes, as well as two additional human-associated clusters. These

enterocolitypes were characterized by variable relative abundance of phylogroup strains and were associated with different host species, diets and habitats.

The genetic variation of pathogenic *E. coli* isolated from urinary tract infections (UTI) has also been addressed, through a multigenome microarray approach that allowed examining the genomic contents of unsequenced *E. coli* strains (Vejborg et al., 2011). This comparative genomics study revealed a strong correlation between genotype, specifically presence of virulence or fitness factors, and phylogenetic group of the strain. In addition, only small differences were reported for isolates associated with symptomatic and asymptomatic infections, although a significant correlation was reported between disease severity and presence of particular pathogenicity islands. The evolutionary dynamics of uropathogenic *E. coli* have also been explored by analyzing isolates of an *E. coli* clone shared among a family of five individuals and a pet dog for three years (Reeves et al., 2011). The strain studied was responsible for an acute UTI episode in the dog and could often be detected in fecal samples of the family members over that time period (Johnson et al., 2008). The results of this study included evidence for minimal adaptive change over the three-year period, with no recombination events or acquisition of genetic mobile elements, even in the isolate associated with the UTI episode (Reeves et al., 2011). In addition, the authors reported an important contribution of host transmission for the maintenance of the strain, with six host transfer events detected during the three-year period.

While these studies, and especially those focusing on the evolution of commensal bacteria, contribute to better understand bacterial diversity in the context of the host, this is still a poorly explored area. In particular, works performing a longitudinal and detailed characterization of the bacterial adaptive process within the host are lacking in the literature. Notably, this subject is even more relevant considering that within-host environments, such as the mammalian gut, represent habitats densely populated by a diverse cohort of microorganisms that are thought to have co-evolved with the host for millions of years (Ley et al., 2008a).

1.5 – The mammalian intestinal environment

The gut is the key organ for the digestive function, where the ingested food is digested and the nutrients necessary for the host survival are absorbed. The mammalian gut represents a complex and particular niche within the host, which differs from the rest of the body as it is colonized by a dense and diverse cohort of microorganisms, known as the microbiota. Displaying a certain level of structure, the intestine encompasses different anatomical regions where physiology, flow rate, substrates, host secretions, pH and oxygen levels, as well as microbial colonization vary (Donaldson et al., 2016) (Figure 8).

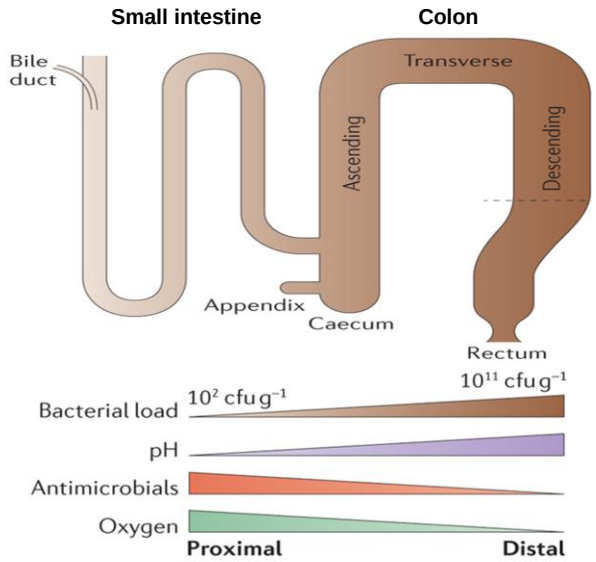


Figure 8. The mammalian gut as a complex environment with multiple microbial habitats. The gut displays gradients of oxygen, antimicrobial peptides (including bile acids) and pH that limit the bacterial colonization in the small intestinal, while supporting high bacterial loads in the colon. Figure adapted from Donaldson et al., 2016.

Moreover, it is also a place under immune surveillance, where immune mechanisms maintain and control the microbial communities. Thus, the gut is a complex environment where multiple factors interplay, establishing an equilibrium that is vital for host survival and well-being.

1.5.1 – The gut microbiota

The intestinal microbiota consists in a community of microorganisms that inhabits the mammalian gut. Whereas bacteria are undoubtedly the most studied and characterized component of this community, other organisms, such as archea, protozoa, fungi and viruses are also present in this environment. In fact, recent evidence suggests that these neglected communities, in particular the gut virome,

may affect the host by interacting with other members of the microbiota (Virgin, 2014).

Though the composition and density of microbial communities varies across the gastro-intestinal tract, it is in the large intestine that the majority of microorganisms inhabit. The numbers of bacteria composing the microbiota are still under debate, with the most conservative estimates pointing to an amount similar to host cells (Sender et al., 2016). In humans, this diverse community is thought to comprise more than 500 bacterial species, with considerable inter-individual diversity (Lozupone et al., 2012). While bacterial phyla such as Proteobacteria, Verrumicrobia, Actinobacteria, Fusobacteria and Cyanobacteria are represented in the gut microbiota, this community is mainly dominated by Firmicutes and Bacteroidetes. In fact, just these two phyla account for more than 90% of the bacteria in the human gut (Ley et al., 2008b).

The gut microbiota is assembled after birth and throughout the lifespan of an individual its composition changes markedly (Clemente et al., 2012). In humans, microbiota composition can be shaped by several factors, including diet, lifestyle, disease and antibiotic use (Lozupone et al., 2012). A longitudinal study following the development and maturation of the gut microbiota in infants over a period of one year has shown that different stages of host development are associated with different microbial compositions (Bäckhed et al., 2015). Thus, while initially dominated by facultative anaerobes (e.g., *Escherichia* and *Enterococcus*) or *Bacteroides*, the gut microbiota transited to a state characterized by high abundance of *Bifidobacterium* and *Lactobacillus*, associated with a more anaerobic environment and milk-based diet. In later stages, when children were no longer breast-fed, the gut microbiota was enriched in groups usually found in adults, such as *Clostridia* and *Bacteroides*, suggesting a shift to a more adult-like gut environment. Thus, the adult gut microbiota is generally dominated by anaerobes, with facultative species reaching only 0.1% of the community (Eckburg et al., 2005). A similar pattern of ecological succession of different bacterial communities was observed in germ-free (GF) mice colonized with a complex microbiota (Gillilland et al., 2012). In this study, one day after conventionalization of previously GF animals, the authors observed a transient bloom of

Proteobacteria in the cecum and jejunum, specifically the genus *Escherichia*. These results support the role of Enterobacteriaceae, including *Escherichia* spp., as pioneer organisms that change the oxygen, pH and nutritional conditions of the gut, allowing later colonization of the gut by strict anaerobes (Blount, 2015).

1.5.2 – Benefits conferred by the gut microbiota to the host

Evidence for the importance of the gut microbiota for host health and homeostasis comes mainly from works in GF animals, and the comparison with microbiota-harboring hosts. The observation that animals raised in GF conditions display altered intestinal functions, morphological defects at the gut level and impaired immune system (Smith et al., 2007), clearly indicates that the microbiota is important for a plethora of host functions, including digestion (Brestoff and Artis, 2013), development and immunity (Laukens et al., 2016) (Figure 9).

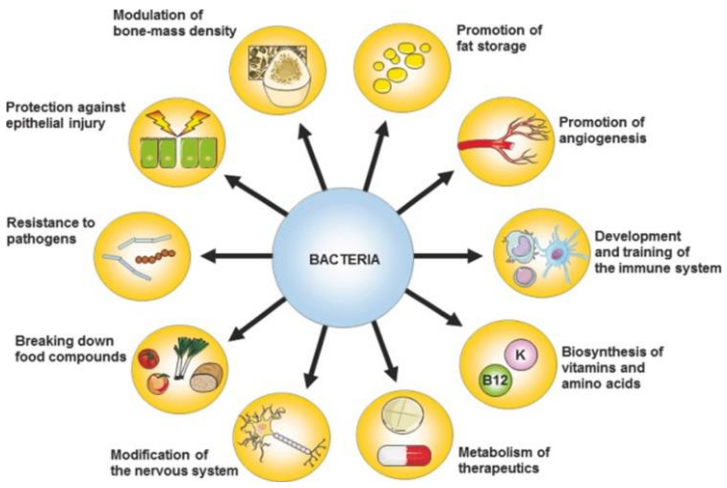


Figure 9. Benefits conferred to the host by the gut microbiota. Bacteria colonizing the mammalian gut provided important functions and benefits to the host. Notably, these include digestion, development and priming of the immune system and colonization resistance, among others. Figure retrieved from Laukens et al., 2016.

The gut microbiota plays a fundamental role in the digestion, by promoting the extraction, synthesis and absorption of nutrients and metabolites, such as bile salts, short chain fatty acids (SCFAs) and vitamins. Thus, gut bacteria provide the host the ability to process many compounds that otherwise would not be digested and eventually absorbed. For example, gut bacteria are involved in the synthesis, conversion, reabsorption and deconjugation of bile acids, molecules responsible by the digestion and absorption of fat. SCFAs such as acetate, propionate and butyrate, are another example of bacteria-derived metabolites. These compounds, resulting from bacterial fermentation of plant polysaccharides, play an important

role in the regulation of host metabolism and importantly, immunity, given their anti-inflammatory proprieties. Vitamins, in particular essential vitamins belonging to group B and K can also be produced by certain members of the microbiota (LeBlanc et al., 2013).

The microbiota is also crucial for the correct development of the host physiology and immunity. For instance, intestinal morphology is altered in the absence of microbes and GF animals display altered microvilli patterns and reduced turnover rate of the epithelial cells. The gut associated lymphoid tissues (GALT) of GF animals are also affected, with underdevelopment of immune structures such as Peyer's patches (PPs) and mesenteric lymph nodes (MLNs), and decreased antibody production. Consistently, GF animals have increased susceptibility to bacterial, viral and parasitic infections compared with animals harboring a complex microbiota (Round and Mazmanian, 2009). This phenomenon is to a certain extent due to the colonization resistance provided by the gut microbes that prevent colonization of an invasive bacterial species, which illustrates the co-dependence of host and microbes, thought to result from the extended co-evolution process between these two partners.

1.5.3 – Bacterial nutrition in the gut

Consistent with a primary function of absorption of nutrients, the small intestine (comprising ileum and jejunum) does not harbor a dense community of microorganisms, as it is found in the colon. Here, rates of intestinal transit, mucus thickness and antimicrobial secretion are different from those in the small intestine and thus more amenable to colonization by the microbiota (Louis et al., 2014). However, since most of the absorption of nutrients occurs in the small intestine, the main substrates reaching the large intestine are compounds previously undigested by the host enzymes. These include indigested carbohydrates, proteins and fats (Scott et al., 2013) (Figure 10). Carbohydrates can be diet-derived (resistant starches, non-starch polysaccharides, oligosaccharides and some mono- and di-saccharides) or host-derived (glycoproteins such as mucins). Proteins comprise not only residual dietary proteins but also enzymes produced by

the host. Fats resulting from the diet are mostly absorbed in the small intestine but a small percentage can be excreted in the feces.

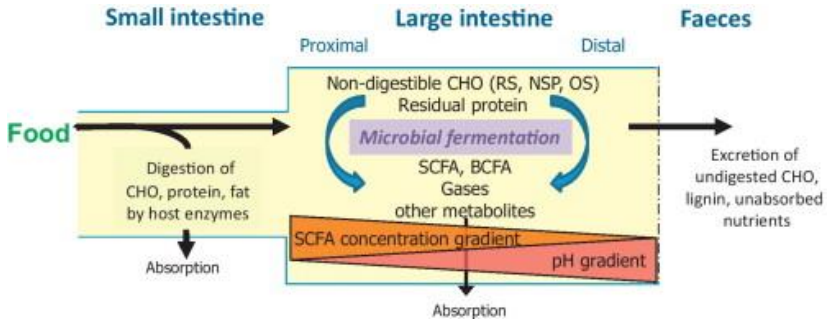


Figure 10. Digestion and microbial processing of ingested dietary components in the gut. Food-derived components not absorbed by the host, including non.digestible carbohydrates (CHO), resistant starch (RS), non-starch polysaccharides (NSP) and oligosaccharides (OS) can be processed by the microbiota into short and branched chain fatty acids (SCFA, BCFA) and other compounds. Figure retrieved from Scott et al., 2013.

Efficient degradation of these compounds involves many times specific enzymes, which are not present in all microbes. For example, starch degradation, requiring enzymes as amylases and pullulanases, can only be performed by bacterial groups with these catalytic activities, such as species of the genus *Bacteroides* (Flint et al., 2012). Emphasizing the metabolic flexibility of this group, members of the species *B. thetaiotaomicron* colonizing the gut of infant mice have been shown to alter their metabolism according to the diet of the hosts (Bjursell et al., 2006). Thus, in the initial lactation period, *B. thetaiotaomicron* can grow mainly in mucosal polysaccharides and other saccharides derived from the mother's milk, whereas in a later stage, after weaning, can metabolize plant-derived dietary polysaccharides. In addition, species of the genus *Bacteroides* are also able to degrade proteins (proteolytic activity) and ferment amino acids (Scott et al., 2013).

While most compounds are subjected to fermentation, some substrates can also be metabolized by anaerobic respiration, using as electron acceptors nitrate, sulphate and other organic compounds (Louis et al., 2014). Microbial metabolism of carbohydrates results mainly in the production of SCFAs and gases, but protein metabolism can also produce ammonia, branched chain fatty acids (BCFAs), phenols and indoles, amines and sulfides (Scott et al., 2013). Importantly, the SCFAs acetate, propionate and butyrate, are then used by the host in the processes of lipogenesis, gluconeogenesis, or as substrate for colonocytes,

respectively. In addition, products resulting from microbial metabolism can also be used by other microorganisms, promoting the establishment of cross-feeding interactions (Flint et al., 2012). As an example, *Bifidobacterium adolescentis* is able to metabolize starch or fructo-oligosaccharides, producing lactate and acetate as byproduct. These compounds can then be used by butyrate-producing bacteria, which are unable to use starch or fructo-oligosaccharides (Belenguer et al., 2006).

In agreement with different bacterial metabolic capabilities, the microbiota composition can change dramatically according to the host diet. For example, in humans, obese volunteers fed a low-carbohydrate diet displayed reduced numbers of certain butyrate-producing Clostridia species (Duncan et al., 2007), while in another work, shifts between plant-based and animal-based diets led to a change in the ratio of bile-tolerant Bacteroidetes and plant-degrading Firmicutes (David et al., 2014). Similarly, Bifidobacteria belonging to the phylum Actinobacteria have been shown to dominate the gut microbiota of breast-fed human babies (Favier et al., 2002), due to their efficient utilization of milk oligosaccharides (Flint et al., 2012). This impact of the diet on microbiota composition was further elucidated in a study in which gnotobiotic mice colonized with a consortium of representative members of the human gut microbiota (harboring major metabolic functions such as the ability to degrade complex polysaccharides, consume simple sugars, ferment amino acids and remove fermentation byproducts) were subjected to diet perturbations (Faith et al., 2011). Based on the responses to the perturbations, the authors developed a model that could predict up to 60% of the variation in species abundance induced by changes in the diet.

Globally, these studies demonstrate the importance of the microbiota in the nutrition of the host, in which complex microbe-microbe and host-microbe metabolic interactions are fundamental for the maintenance of host homeostasis and health.

1.5.4 – Immune mechanisms at the gut level

The gut microbiota is not in direct contact with the host, in fact it is separated by several layers of protection that aim to contain and monitor the microorganisms

living in the gut. In addition to physical barriers, such as the mucus layer or the epithelium, immune cells located at the gut mucosa actively mount innate and adaptive immune responses to the microbiota (Maynard et al., 2012). These cells are either organized in structures such as PP and MLN or dispersed throughout the epithelium and lamina propria, comprising the intestinal immune system (Mowat and Agace, 2014).

1.5.4.1 – The intestinal immune system

Similarly to the systemic immune system, the intestinal immune system is composed by both innate and adaptive immune cells. Innate immune cells express a limited set of pattern recognition receptors (PRRs) that recognize molecular patters expressed by microorganisms (MAMPs) (Molloy et al., 2012). On the other hand, B and T adaptive immune cells can express more specific antigen receptors, which activation leads to the clonal expansion of effector cells (McGhee and Fujihashi, 2012). These effector cells can not only exert immune responses but also shape innate immune responses. Importantly, although many of the cell types are shared between the systemic immune system and the intestinal immune system, the immune responses at the gut level have the particularity of being generated at the intestinal mucosa in response to the microbiota (Brandtzaeg and Pabst, 2004). Thus, the intestinal immune system, and in particular the adaptive immune branch of this system, is a flexible player that can not only monitor but also respond accordingly to microbial stimuli in the gut. This ability is evident, for instance in situations of microbial imbalance, called dysbiosis, when the immune system activates immune mechanisms aiming to control the microbial population and return to a status of equilibrium (Round and Mazmanian, 2009). The modulation of immune responses, accompanying changes in microbiota composition and function, is therefore a crucial feature that allows maintenance of homeostatic conditions in the gut.

1.5.4.2 – The defense layers in the gut

The microbes inhabiting the gut are separated from the host by several structural barriers (Figure 11). The first layer of defense in the gut comprises the

mucus. The mucus is formed by gel-forming mucins secreted by the goblet cells, in particular the Muc2 glycoprotein (Johansson et al., 2008). In addition to mucins, the mucus is also composed by small glycoproteins, proteins, glycolipids, and lipids (Conway and Cohen, 2015). The thickness of the mucus layer varies along the intestine: while it is thinner in the small intestine, it is thicker in the large. The loads of microorganisms in the gut also vary along the gut (Maynard et al., 2012). The mucus layer is subdivided in two layers: a loosely adherent outer layer and an inner adherent layer. In mice, the outer layer is thicker and can be easily removed, while the inner layer is thinner and firmly attached to the epithelium. These layers differ in the presence of microorganisms; as they densely populate the outer layer, they are absent in the inner layer (Johansson et al., 2008). This is due not only to structural differences in the physiochemical structure of the two mucus layers but also the presence of microbicidal products from the epithelium, including antimicrobial peptides (AMPs) and secreted IgA. In addition, mucin glycoproteins are toxic to some groups of bacteria (Mowat and Agace, 2014).

The epithelium is the second layer of defense and encompasses different types of cells with specialized functions: absorptive enterocytes (Snoeck et al., 2005), mucin-secreting goblet cells (Maynard et al., 2012), AMP–producing Panneth cells (Maynard et al., 2012), antigen transporter M cells (Mowat and Viney, 1997) and enteroendocrine cells. In addition to epithelial cells, intraepithelial lymphocytes (IELs) are also present in the epithelium (Sheridan and Lefrançois, 2010).

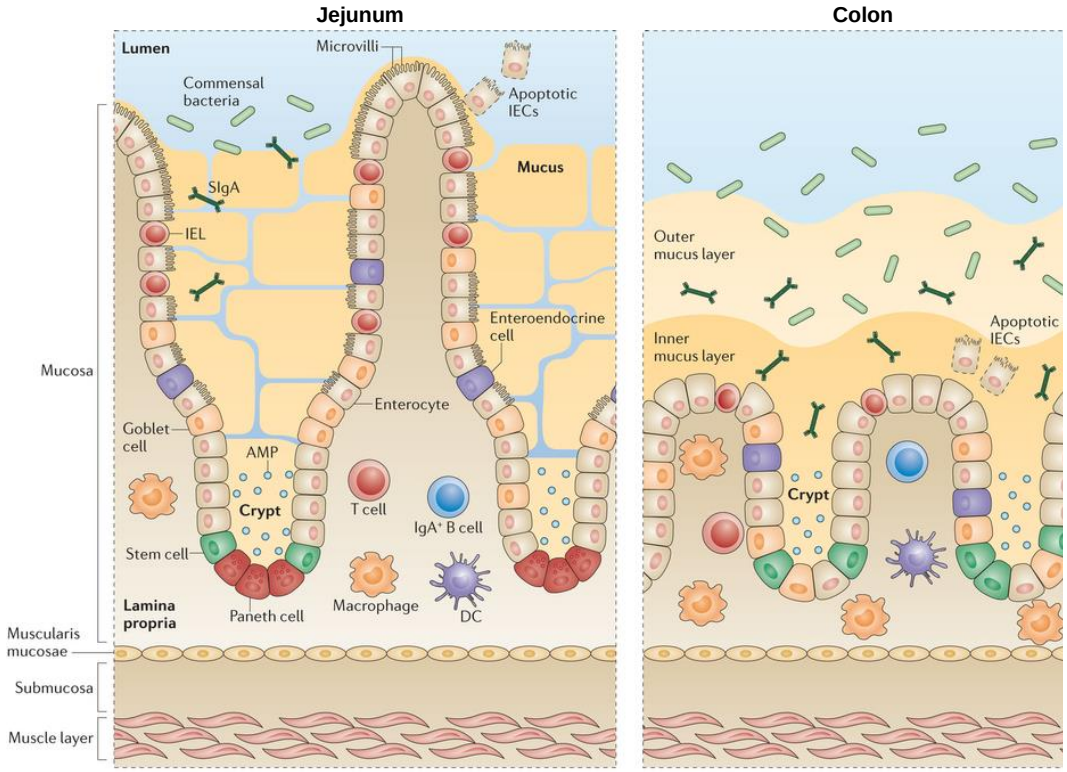


Figure 11. Intestinal barriers separating the microbiota from the host.
Schematic representation of the small intestine and colon, showing the mucus layer, epithelium and innate and adaptive cells in the gut mucosa. Figure retrieved from Mowat and Agace, 2014.

A third layer includes the lamina propria and a multiplicity of innate (dendritic cells, macrophages, eosinophils and mast cells) and adaptive (B and CD4 or CD8 T cells) immune cells that populate this region (Mowat and Agace, 2014)). The CD4 population is composed by several subsets, including T helper cells (Th1, Th2 and Th17) and T regulatory cells (Tregs). Notably, Th17 cells produce pro-inflammatory cytokines and are involved in inflammatory responses against microbes, while Tregs are involved in the regulation of inflammatory responses, thus acting as a counterweight to Th17 cells. In fact, Tregs can dampen aberrant inflammatory responses against members of the microbiota, promoting homeostasis in the gut mucosa (Kamada and Núñez, 2013). Expansion of both subsets in the intestine is dependent on cues from the microbiota, since their numbers are reduced in GF or antibiotic-treated animals (Kamada and Núñez, 2013). On the other hand, plasma B cells, present in high amounts in the lamina propria of the intestine, produce immunoglobulins, in particular IgA. This secreted

isotype is then translocated through the gut epithelium to the intestinal lumen, where it can bind to different bacterial groups, coating the cells. While the exact purpose of the coating is not completely understood, it has been proposed that the outcome of the binding may depend on the binding affinity: high-affinity binding to neutralize toxins and pathogens and low-affinity binding to block the adhesion of microbes to the epithelium (Cerutti, 2008). Other functions may be related with the enhancement of bacterial uptake at the gut mucosa to promote a positive feedback loop of B cell induction (Fransen et al., 2015) or in the maintenance, instead of exclusion of certain members of the microbiota in the gut (Kawamoto et al., 2014). Hence, IgA can potentially affect microbes by neutralizing or preventing the growth of pathogenic bacteria or supporting the maintenance of a diverse cohort of microorganisms in the gut (Palm et al., 2015).

Thus, in addition to structural barriers that limit the contact of microorganisms with the host, innate and adaptive mechanisms cooperate to selectively exclude or promote host-microbe interactions, constantly sensing and responding accordingly to stimulus from the microbiota and contributing to homeostasis in the gut.

1.6 – Reciprocal interactions between the gut microbiota and the host immune system

The microbiota composition is thought to result from a process of extended coevolution with the host that has been occurring for millions of years (Ley et al., 2008a). In particular, the immune mechanisms of the host that limit and control the microbiota are thought to be essential for the maintenance of this symbiotic relationship. Therefore, reciprocal interactions between the microbiota and the host, and in particular the host immune system, are expected to take place.

The interactions between host immune system and microbes starts after birth, when the host is first exposed to a significant amount of microbes at delivery. During the first stages of gut colonization, the immune system can change the type of immune responses, presumably according to the set of microbes colonizing the gut. Likewise, the presence of microbes during the post-natal development of the host immune system is crucial to promote the differentiation of certain lymphoid

tissues and cells during that period. In fact, some of the immune defects displayed by animals raised in GF conditions are never recovered even when exposed and colonized as adults by a complex cohort of microorganisms (Gensollen et al., 2016). This observation highlights the tight regulation necessary for a correct development of the host immune system and the establishment of the complex and important interactions between host and colonizing microbes.

To better understand the reciprocal interactions between the immune system and the microbiota, two complementary approaches are currently used. To investigate the impact of microbes on the development and expansion of immune cells, GF animals are colonized with single or multiple bacterial species (El Aidy et al., 2012). On the other hand, assessing the effect of the host immunity on the microbiota composition requires the analysis of the microbiota in hosts defective in different components of the immune system (Hasegawa and Inohara, 2014).

1.6.1 – Effects of the microbiota on the development of the immune system

Animals raised in GF conditions are invariably defective in immune mechanisms, which can be recovered to a certain extent upon conventionalization, i.e., colonization with microbiota from regular microbiota-harboring (conventional) animals. This effect has been elucidated in certain innate immune pathways, such as secretion of AMPs. For example, the gram-positive-killing C-type lectin Reg3γ is an AMP present at a minimal level in GF mice, which expression increases substantially upon colonization with a complex microbiota (Cash et al., 2006). Importantly, a similar effect has been shown to occur for more specific, adaptive immune mechanisms. In fact, the expansion and activity of cells from the adaptive immune system, such as effector T or B cells, have been shown to depend on signals from certain members of the microbiota, that thus influence the development of the immune system through more specific microbe-host interactions.

1.6.1.1 – T cells

Several works have tried to identify microbial groups that promote the differentiation and expansion of effector T cells, such as Th17 and Tregs (Figure

12), which play a fundamental role in the balance between inflammation and tolerance.

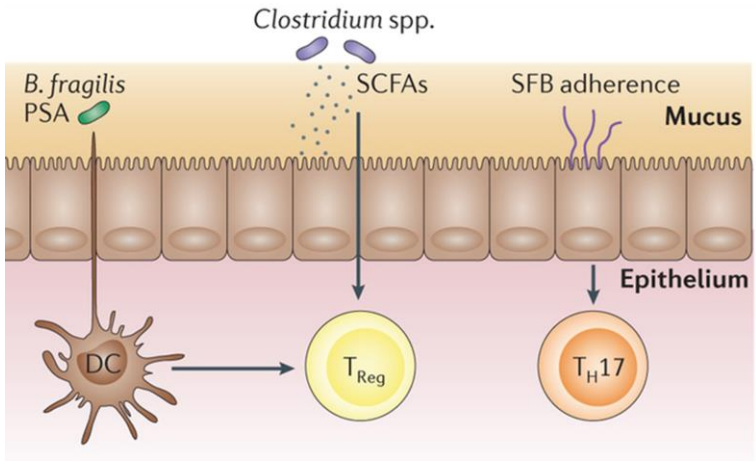


Figure 12. Immunomodulatory proprieties of members of the gut microbiota. Examples of bacterial groups that are able to drive the differentiation of anti (Treg) or pro- (TH17) inflammatory T cell subsets. Figure adapted from Donaldson et al., 2016.

For instance, expansion of inflammatory Th17 and also Th1 has been reported in mice monoassociated with segmented filamentous bacteria (SFB), a Clostridia-related bacterial species. This expansion was shown to confer protection against intestinal pathogens such as *Citrobacter rodentium* (Ivanov et al., 2009). In healthy hosts, this increase in pro-inflammatory T cells was also accompanied by an expansion in Treg cells (Gaboriau-Routhiau et al., 2009), that in this way maintain the homeostatic environment in the mouse gut, preventing aberrant inflammation. Due to their role in suppressing inflammatory immune responses, Tregs have also been the focus of several studies that identified microbial species that were able to drive *de novo* differentiation of Tregs in the intestinal mucosa. This process has been described in mice monoassociated with a cocktail of Clostridia species indigenous of the mouse gut (Atarashi et al., 2011), altered Schaedler flora (a consortia of eight microorganisms including *Clostridium* and *Lactobacillus* species (Brand et al., 2015)) and single or multiple bacteria isolated from human donors. Examples of these strains include *Bacteroides fragilis* (Round and Mazmanian, 2010), in specific the bacterial component polysaccharide A, a cocktail of clusters IV, XIVa and XVIII 17 Clostridia strains (Atarashi et al., 2013) and individual strains isolated from human hosts (Faith et al., 2014). Interestingly, in this last study several microbes could significantly induce the differentiation of Tregs in monocolonized mice, including species from

the order Bacteroidales or *E. coli*. This observation suggests that many bacterial species have the capacity to modulate the cell composition of the immune system globally promoting the maintenance of homeostasis.

1.6.1.2 – B cells

Multiple works in GF mice have showed that although the numbers of intestinal IgA-producing B cells are low in the absence of microbes, they increase substantially after microbial exposure (Macpherson et al., 2001). Early works described this induction in animals monocolonized with different species of bacteria, such as those belonging to the genus *Escherichia* and *Bacteroides* (Moreau et al., 1978), *Morganella* (Shroff et al., 1995) and, more recently, *B. thetaiotaomicron* (specifically to its capsular polysaccharide (Peterson et al., 2007)). Production of *B. thetaiotaomicron*-specific IgA was associated with a reduction in inflammatory responses against this bacterium, promoting the establishment of a non-inflammatory host-microbe relationship. SFB, that are located in close contact with the epithelial cells, have also been shown to promote the production of IgA, although in this case only a small percentage was specific for SFB (Talham et al., 1999). Induction of a slow but long-lasting specific IgA response has also been observed in GF mice colonized with an avirulent laboratory strain of *E. coli* (Hapfelmeier et al., 2010). Interestingly, the production of *E. coli*-specific IgA could be abolished upon colonization with a different microbiota, showing that the specificity of antibody responses can rapidly adapt to changes in microbiota composition.

Production of specific IgA against commensal species has also been detected in animals colonized with a complex microbiota (Palm et al., 2014). For example, IgA specific to *Enterobacter cloacae*, a predominant aerobe in the mouse microbiota (Macpherson et al., 2000) has been observed in mice harboring a complex microbiota. In fact, early studies showed that a significant percentage of colonic bacteria from human donors were coated with IgA (van der Waaij et al., 1996). Supporting these results, 16S rRNA sequencing analysis of the IgA-coated fraction of the fecal microbiota of six healthy humans concluded that many bacterial groups were coated by IgA, with rare bacteria accounting to 20% of the

total microbial diversity in this fraction (D'Auria et al., 2013). Interestingly, Gamaproteobacteria such as *Escherichia/Shigella*, were present in this fraction in almost all individuals, albeit at low frequency. In addition to coating commensal members of the microbiota, it has also been shown that IgA can bind commensals that potentially cause disease in susceptible animals (Palm et al., 2014). Thus, induction of B cells in the gut mucosa seems to represent an important immune mechanism strongly dependent on microbial exposure and contributing to a healthy cross-talk between microbiota and host.

1.6.2 – Effects of the immune system on the microbiota

Given the microbe-host interactions previously described, defects on host immunity are expected to have some impact at the level of the microbiota. An approach extensively applied to investigate the impact of the immune system on microbial composition is the analysis of microbial communities of genetically altered immune-deficient animals. This process has been facilitated by the rapid advances in genetic manipulation of mammalian model organisms (Rosenthal and Brown, 2007), which allowed the construction of hosts defective in genes related with a multiplicity of immune functions, both innate and adaptive. However, the results are not always clearly interpreted, since the gut microbiota can differ markedly between animal husbandries (Ericsson et al., 2015) and the phenotypes displayed by immune-compromised animals are often dependent on the microbiota composition (Yang et al., 2013). Thus, although some studies describe altered microbial composition in immune-deficient hosts relatively to wild-type, specific differences tend not to be consistent across different animal houses. This has been observed, for example, in studies analyzing the gut microbiota composition in mice with defective innate immune mechanisms involved in microbial patterns recognition, such as toll-like receptors (TLRs) or NOD-like receptors (NLRs). While some works report an altered gut microbiota and other phenotypes in these defective hosts (Petnicki-Ocwieja et al., 2009; Vijay-Kumar et al., 2010), the results could not be reproduced in other studies (Robertson et al., 2013; Ubeda et al., 2012), and thus the contribution of these pathways in shaping the microbiota composition remain unclear.

More relevant when considering the hypothesis of co-evolution between host immunity and microbiota, mice defective in adaptive immune components have generally shown to harbor an altered gut microbiota. Notably, this phenomenon has been observed in mice lacking Tregs (Josefowicz et al., 2012), B cells (Shulzhenko et al., 2011) and IgA (Suzuki et al., 2004). Interestingly, evidence suggests that T cells exert mainly an indirect effect in controlling the gut microbiota composition, by mediating the production of IgA by B cells through T-cell dependent mechanisms (Palm et al., 2015). Supporting this hypothesis, studies analyzing the microbiota composition of T or B cell knockout mice found a similar decrease in microbial diversity in relation to WT controls (Kawamoto et al., 2014). Recently, a study described a correlation between abundance of polyreactive IgAs and diversity of the gut microbiota, supporting the role of IgA as an active agent promoting the maintenance of microbial diversity (Fransen et al., 2015).

As a more drastic immune defect, *Rag* (*Rag1* or *Rag2*) knockout mice lack a crucial enzyme involved in the maturation of the T and B cells antigen receptors and thus lack both of these two cell types. These hosts are unable to mount adaptive immune responses, such as immunoglobulin production or control the levels of pro- (Th17) and anti-inflammatory (Tregs) T cell subsets in the gut mucosa. Although different studies reported alterations in the microbiota composition of these mice, the findings are not consistent across different animal houses. For example one work described similar results to those observed in IgA defective mice (Suzuki et al., 2004), comprising an expansion of SFB and other anaerobic bacteria in *Rag2*$^{-/-}$ mice. In another study the authors reported decreased diversity in the gut microbiota of *Rag1*$^{-/-}$ mice, mirroring the results obtained in B or T cell-deficient mice (Kawamoto et al., 2014). Consistent with these results, a recent study also described decreased diversity of microbial communities in hypomorphic *Rag2* mice compared to WT controls, as well as different abundances of several groups of Proteobacteria (Rigoni et al., 2016). Other authors found differences in the abundance of groups such as Bifidobacteria and *Clostridium leptum* (Thoene-Reineke et al., 2014) or Lachnospiraceae and Porphyromonadacea (Scholz et al., 2014) in *Rag1*$^{-/-}$ animals, or an increased

frequency of *Akkermansia* sp (Zhang et al., 2015) in *Rag2*$^{-/-}$ hosts compared to WT controls.

Globally, these results confirm the importance of the immune system, in particular the adaptive immune system, in shaping the composition of the gut microbiota. Together with the evidence that different microbial groups can promote the maturation and tuning of the immune system, these data highlights the reciprocal interactions between host and microbiota and provides support to the hypothesis of co-evolution between these two partners.

1.7 – *E. coli* and mouse as host-microbe model organisms

The area of host-microbe interactions and in particular how microorganisms colonize their hosts has been addressed using a variety of model systems, based on model organisms both for the host and the microbe. Available model organisms for the host display a range of complexity, including invertebrates (as the Hawaiian bobtailed squid and the fruit fly) or more complex vertebrate models (for example, zebrafish and mice) (Kostic et al., 2013). As the prime model for mammalian organisms, the mouse system has been extensively used in studies of microbial gut colonization. On the other hand, microbial model organisms include the most well-known bacterium, and a commensal member of the gut microbiota, *E. coli*.

1.7.1 – The laboratory mouse model

Initially selected for docility, tolerance to confinement and fecundity, lab mice differ from their wild counterparts as they display a faster growth rate, increased size and higher reproductive success (Austad, 2002). Other features altered during domestication include decreased physical fitness, shorter lifespans and smaller eyes and brains, as well as increased chromosome breakage and longer telomeres (Austad, 2002). Remarkably, inbreed lab mice lack the extensive genetic variability present in outbred wild populations of mice (Phifer-Rixey and Nachman, 2015). While phenotypic and genetic differences between wild and lab mice have been comprehensively addressed, the composition of their microbial communities is a subject still poorly explored. It has been shown, however, that

the gut microbiota of wild and wild-derived inbred mice, while displaying similar community diversity differ in the composition of the microbial communities (Kreisinger et al., 2014). Notably, pathogens and parasites (Maurice et al., 2015; Weldon et al., 2015) are commonly detected in wild mice, whereas in controlled inbred lab strains are generally absent. Another feature found in wild mice is a striking seasonal variation in the gut microbiota composition, possibility related with shifts in diet patterns of the animals (Maurice et al., 2015). In laboratory mice, shifts in microbiota composition have been shown to occur in immune-deficient mice (Maharshak et al., 2013) and/or in response to a dietary (Zhang et al., 2012), infection (Belzer et al., 2014) or drug (Antonopoulos et al., 2009; Schwab et al., 2014) perturbation, while in healthy mice kept in controlled conditions age-related changes have been described (Langille et al., 2014; Zhang et al., 2012). Despite these differences in relation to their wild relatives, the laboratory mouse is a good organism to investigate the gut microbiota and its importance and interaction with the host. In particular, the strict control over genetic and environmental variation accessible with this model represents a major asset when performing this type of experiments. Another important aspect is the availability of genetic tools, allowing the construction of mutant gene knock-out strains that can be used to investigate the importance of a given gene or trait (Beckers et al., 2009).

Furthermore, the lab mouse can also be used as a model for other mammalian organisms, in particular humans. Mice share with humans many of their overall physiology and cellular functions, including musculoskeletal, immune, endocrine and digestive systems (Vanhooren and Libert, 2013), the latter comprising roughly the same compartments in both organisms. On the other hand, mice differ from humans in important parameters such as size, metabolic rate, life history traits and, more relevant to the study of host-microbe interactions, intestine physiology, diet, pathogens and microbiomes (Perlman, 2016) (Figure 13). Despite these differences, the distal guts of mice and humans are strikingly dominated by the same bacterial phyla, which include the Firmicutes, Bacteroidetes and Actinobacteria (Spor et al., 2011). Similar microbiota composition between these two hosts has also been described at the family level, although at the genus level only 15% of the taxa were found to be shared (Ley et al., 2005).

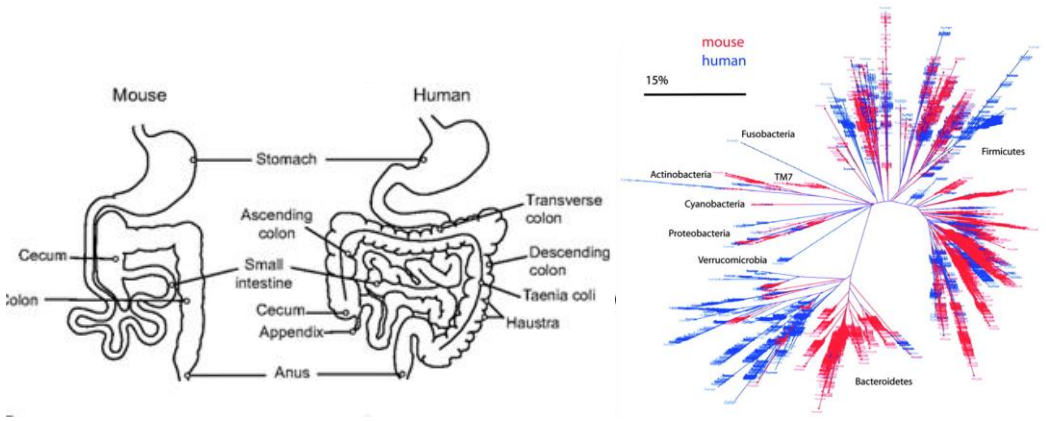


Figure 13. Anatomical and microbiota differences between mouse and human. Mice and humans differ in the compartmentalization of the colon and stomach. In addition, although the dominant bacterial phyla (Firmicutes and Bacteroides) are shared between the two hosts, 85% of the genera identified in mice were not fond in human. Figure adapted from Nguyen et al., 2015 and Ley et al., 2006.

In addition, mice can be produced in GF conditions (Al-Asmakh and Zadjali, 2015), allowing dissecting the specific contribution of different microbes to host phenotypes (Faith et al., 2014), as well as controlled manipulation of the microbiota composition, including colonization with microbial consortia from human donors (humanized mice (Laukens et al., 2016)). These features, together with proprieties intrinsic to the laboratory mouse, such as short generation times, easy breeding and maintenance, controlled biocontainment, and relatively low-cost, make this host a good experimental model (Austad, 2002). Thus, despite the limitations of the mouse model that not always allow a direct translation to humans (Nguyen et al., 2015), this model can still be considered a valuable and useful tool in the study of host-microbe interactions, demonstrated by the significant contribution it has provided to the field of microbiota research.

1.7.2 – *E. coli* as a microbial model organism

In addition to its status as the most famous microbial model organism, *E. coli* is also a relevant organism when addressing the topic of host-microbe interactions. As a pervasive inhabitant of the mammalian gut, commensal strains of *E. coli* have been commonly isolated from many animals, including humans (Tenaillon et al., 2010) and mice (Souza et al., 1999). Notably, the genus *Escherichia* (Gillilland et al., 2012) in mice, and in particular the species *E. coli* in

humans (Bäckhed et al., 2015), appear to be relevant pioneer microbes in the initial colonization of the gut. As a pathogen, it affects mainly humans and domestic animals (Kaper et al., 2004). While not a common mouse pathogen, several mouse models of *E. coli* infection have been developed trying to reproduce and study the disease process observed in humans (Mohawk and O'Brien, 2011; Savkovic et al., 2005). In addition, other strains isolated from sick (Diard et al., 2010; Fabich et al., 2008) or healthy (Adediran et al., 2014; Maltby et al., 2013; Wadolkowski et al., 1988) human donors have shown the ability to colonize the mouse gut without clinical symptoms, under certain models of colonization.

In fact, strains of *E. coli* have been used as microbial model organisms in several studies investigating bacterial colonization of the mouse gut and the determinants involved in the process. In particular, the topics of *E. coli* distribution (Foucault et al., 2010; Poulsen et al., 1994), growth rate (Li et al., 2015; Myhrvold et al., 2015; Poulsen et al., 1995; Rang et al., 1999), signal sensing (Lasaro et al., 2014), nutrition (Chang et al., 2004; Fabich et al., 2008), respiratory metabolism (Jones et al., 2007, 2011) and evolution (De Paepe et al., 2011; Leatham-Jensen et al., 2012; Lescat et al., 2016) in the mouse gut have been addressed. While a few works were performed with more natural but less described murine (Lasaro et al., 2014; Myhrvold et al., 2015; Poulsen et al., 1994, 1995; Rang et al., 1999) or human (Lescat et al., 2016) isolates of *E. coli*, the majority of these works used the K-12 MG1655 strain. Extensively used in many *in vitro* experiments and characterized by a fast growth, with few growth requirements and metabolic flexibility, this strain has been thoroughly described both at the genotypic and phenotypic levels. Importantly, it is also amenable to genetic manipulation as well as gene expression and sequencing (Hobman et al., 2007). Though considered a "domesticated" strain, that lost some of the features characteristic of natural isolates, K-12 MG1655 has been shown to share with commensal and environmental *E. coli* strains basic eco-physiological proprieties, such as carbon substrate uptake and breakdown capabilities and stress defense mechanisms (Ihssen et al., 2007). Thus, taking into account the limitations and innumerous advantages of the model, *E. coli*, particularly the well-known strain K-12 MG1655, can be used as a useful tool to address the topic of host-microbe interactions

taking place into the mammalian gut and gain valuable information on this still incompletely clarified subject.

1.7.3 – Models of colonization: streptomycin-treated and germ-free mouse models

In homeostatic conditions, the adult mouse gut microbiota is usually resistant to stable colonization with exogenous commensal *E. coli* strains, a process known as colonization resistance (van der Waaij et al., 1971). Although not completely elucidated, this phenomenon is thought to be related to direct or indirect competition with members of the microbiota. While the exact mechanisms that provide colonization resistance are still poorly understood, it has been shown that resistance can be abolished with antibiotic treatment, in particular streptomycin. Treatment with this antibiotic decreases the load of endogenous bacteria by one log and leads to profound changes in the composition of the gut microbiota (Thompson et al., 2015). In addition, it was also reported that antibiotic treatment leads to increased inflammation in the intestine, with production of reactive oxygen species (Spees et al., 2013). Thus, antibiotic treatment vacates a physical niche previously occupied by other microbes, eliminates competitors of *E. coli* and leads to an altered metabolic and respiratory environment where *E. coli* has an advantage. Alone or in combination, these factors ultimately lead to the breakdown of colonization resistance, allowing colonization of the mouse gut with *E. coli*. Although a less diverse habitat compared with untreated animals, the streptomycin-treated mouse gut is still a complex environment where *E. coli* is subjected to multiple selective pressures, including the presence of other microorganisms.

As an alternative to the streptomycin-treated mouse model, adult GF, or axenic animals can also be colonized with bacteria. As hosts devoid of microbiota, they lack colonization resistance and thus can be easily colonized with *E. coli* strains. Despite consisting in a simpler environment than microbiota-harboring animals, this model is useful to address specific questions related to the contribution of the microbiota to certain host phenotypes (Faith et al., 2014). In

addition, colonization of previously GF animals mimics the natural colonization that occurs after birth and thus allows the study of this process in controlled conditions.

1.7.4 – Nutrition of *E. coli* in the mouse gut

In the streptomycin-treated mouse gut, *E. coli* has been shown to form mixed biofilms in the mucus layer of the large intestine (Leatham-Jensen et al., 2012). However, while inhabiting the mucus layer, *E. coli* is unable to degrade the complex polysaccharides that result from the degradation of mucins, epithelial cell or dietary fibers (Conway and Cohen, 2015). Thus, it relies on the hydrolytic activity of other microorganisms that can use the complex glycoproteins as carbon source (Li et al., 2015). These microorganisms (e.g. *B. thetaiotaomicron*) are able to secrete hydrolases that degrade complex polysaccharides into simpler compounds. These compounds are then uptaken and metabolized, which results in the subsequent release of mono- and di-saccharides as subproducts. These available simpler carbon sources can then be used in metabolic reactions by *E. coli*.

Several studies addressing the nutrition of *E. coli* in the mouse gut have shown that different strains may have distinct metabolic requirements allowing coexistence in the intestine through the occupation of different niches (Leatham et al., 2009; Meador et al., 2014). Furthermore, some strains have the ability to switch their metabolic preferences depending on the competing strains, thus overcoming colonization resistance promoted by other *E. coli* strains. One example of this fine modulation of microbial metabolism is the switch performed by a pathogenic strain of *E. coli* from glycolytic to gluconeogenic nutrients when invading the mouse gut pre-colonized with a commensal *E. coli* strain (Schinner et al., 2015). In particular, the nutritional preferences of the commensal lab strain MG1655 when colonizing the streptomycin-treated mouse gut have been thoroughly described (Chang et al., 2004). Through the use of knockout mutants of *E. coli* in different metabolic pathways, the authors identified several carbon sources that are important for initiation or maintenance of gut colonization.

Globally, these works demonstrate the importance of bacterial nutrition in the mouse gut, suggesting that this may be an important factor shaping the evolution of *E. coli* in this habitat.

1.7.5 – Experimental evolution of *E. coli* in the mouse gut

Although the process of *E. coli* adaptation to the mouse gut is yet a poorly explored subject, some studies reported the selection of mutants with increased competitive advantage in the mouse gut upon colonization of streptomycin-treated mice with the *E. coli* strain MG1655. The described mutations targeted the *flhDC* operon (Gauger et al., 2007; Leatham et al., 2005) and the *envZ* gene (Leatham-Jensen et al., 2012). Mutants in the *flhDC* operon were rapidly selected in the mouse intestine (just three days after colonization), were non-motile and could grow better than the ancestral in cecal mucus and in media containing sugars present in the mucus. Conversely, *envZ* missense mutations conferred increased growth in specific sugars, such as galactose, suggesting that genotypes carrying these mutations may occupy a different nutritional niche than *flhDC* mutants in the mouse intestine. In addition, *envZ* mutants also displayed increased resistance to bile salts and colicin, a bacteriocin (Leatham-Jensen et al., 2012). Interestingly, the same mutants were reported to occur in the same strain of *E. coli* colonizing the GF mouse gut (De Paepe et al., 2011; Giraud et al., 2008), suggesting that mutations in these genes are adaptations to the nutritional environment of the mouse gut, independent of the microbiota.

Very recently, a long-term study followed the evolution of a human-derived *E. coli* strain in the gut of streptomycin-treated mice for over one year (Lescat et al., 2016). Sequencing analysis of individual clones revealed an average of seven mutations per genome and convergence among lineages at the gene (over 65%), operon or functional levels. Specifically, loci shown under selection comprised the *dgo* operon (involved in the galactonate pathway), the *rluD* and *gidB* complex (related with the maturation of ribosomal RNAs) and the intergenic region between *yijH* and *kptA*. While mutation in the *dgo* operon conferred increased growth in D-galactonate, those in *rluD* and *gidB* were advantageous only in the presence of streptomycin.

1.8 – Open questions at the start of this PhD and overall strategy to address these

Although studies of experimental evolution have addressed and described the adaptive process of bacteria in simple and/or artificial environments, the evolution of bacteria in natural environments, is a subject still largely unexplored. Moreover, the determinants and regimes of adaption, the contribution of the multiple selective pressures playing a role in the habitat, as well as the extent of adaptation in these natural environments are subjects currently unknown.

E. coli, the most well-known model organism, has been the focus of many *in vitro* studies of experimental evolution, that described in detail its adaptation to laboratorial environments (Barrick and Lenski, 2013). However, in its natural environment, the mammalian gut, the adaptive process of *E. coli* is far from elucidated. The mammalian gut is a complex environment, where bacteria are subjected to multiple pressures, including nutrients, the host immune system and the other microbes (Donaldson et al., 2016), that may shape microbial adaptation to this environment. While the importance of the nutritional environment has been the focus of some studies (De Paepe et al., 2011; Leatham-Jensen et al., 2012), the impact of host-microbe and microbe-microbe interactions to the adaptation of *E. coli* in the mouse gut, the prime mammalian model organism, is still poorly known.

Thus, it is within the scope of this thesis to characterize the adaptation of *E. coli* to the mouse gut and evaluate the mechanisms of selection taking place during adaptation to this habitat. In addition, one of the aims is also to assess the contribution of one important host factor, the adaptive immunity, to this process.

To address these questions we followed the evolution of fluorescently-labelled *E. coli* in the gut of streptomycin-treated mice, using a two-marker system to infer adaptive events occurring in the *E. coli* populations. Independently evolving populations were furthermore subjected to WGS to identify the genetic basis of *E. coli* adaptation. Additionally, the microbial ecosystem experienced by *E. coli* was characterized by 16s rRNA gene sequencing. Competitive fitness assays were also performed to evaluate the advantage of mutations identified by genome sequencing. To analyze the contribution of host immunity to *E. coli* evolution we

compared the dynamics, genetic basis and microbial environment of *E. coli* evolved in WT or mutant mice lacking the adaptive immune system (*Rag2*$^{-/-}$ mice).

References

Adediran, J., Leatham-Jensen, M.P., Mokszycki, M.E., Frimodt-Møller, J., Krogfelt, K.A., Kazmierczak, K., Kenney, L.J., Conway, T., and Cohen, P.S. (2014). An Escherichia coli Nissle 1917 missense mutant colonizes the streptomycin-treated mouse intestine better than the wild type but is not a better probiotic. Infect. Immun. *82*, 670–682.

Al-Asmakh, M., and Zadjali, F. (2015). Use of Germ-Free Animal Models in Microbiota-Related Research. J. Microbiol. Biotechnol. *25*, 1583–1588.

Alfred, J., and Baldwin, I.T. (2015). New opportunities at the wild frontier. eLife *4*, e06956.

Alm, E.W., Walk, S.T., and Gordon, D.M. (2011). The Niche of Escherichia coli. In Population Genetics of Bacteria, (American Society of Microbiology), p.

Antonopoulos, D.A., Huse, S.M., Morrison, H.G., Schmidt, T.M., Sogin, M.L., and Young, V.B. (2009). Reproducible community dynamics of the gastrointestinal microbiota following antibiotic perturbation. Infect. Immun. *77*, 2367–2375.

Atarashi, K., Tanoue, T., Shima, T., Imaoka, A., Kuwahara, T., Momose, Y., Cheng, G., Yamasaki, S., Saito, T., Ohba, Y., et al. (2011). Induction of Colonic Regulatory T Cells by Indigenous Clostridium Species. Science *331*, 337–341.

Atarashi, K., Tanoue, T., Oshima, K., Suda, W., Nagano, Y., Nishikawa, H., Fukuda, S., Saito, T., Narushima, S., Hase, K., et al. (2013). Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. Nature *500*, 232–236.

Atwood, K.C., Schneider, L.K., and Ryan, F.J. (1951). Periodic Selection in Escherichia Coli. Proc. Natl. Acad. Sci. U. S. A. *37*, 146–155.

Austad, S.N. (2002). A mouse's tale. Nat. Hist. *111*, 64–70.

Bachmann, B.J. (1996). Derivations and genotypes of some mutant derivatives of Escherichia coli K-12. In Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology, F.C. Neidhardt, ed. (Washington, DC: ASM Press), p.

Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., et al. (2015). Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe *17*, 690–703.

Bailey, S.F., and Bataillon, T. (2016). Can the experimental evolution programme help us elucidate the genetic basis of adaptation in nature? Mol. Ecol. *25*, 203–218.

Barrick, J.E., and Lenski, R.E. (2013). Genome dynamics during experimental evolution. Nat. Rev. Genet. *14*, 827–839.

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., and Kim, J.F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature *461*, 1243–1247.

Barrick, J.E., Kauth, M.R., Strelioff, C.C., and Lenski, R.E. (2010). Escherichia coli rpoB mutants have increased evolvability in proportion to their fitness defects. Mol. Biol. Evol. *27*, 1338–1347.

Beckers, J., Wurst, W., and de Angelis, M.H. (2009). Towards better mouse models: enhanced genotypes, systemic phenotyping and envirotype modelling. Nat. Rev. Genet. *10*, 371–380.

Belenguer, A., Duncan, S.H., Calder, A.G., Holtrop, G., Louis, P., Lobley, G.E., and Flint, H.J. (2006). Two Routes of Metabolic Cross-Feeding between Bifidobacterium adolescentis and Butyrate-Producing Anaerobes from the Human Gut. Appl. Environ. Microbiol. *72*, 3593–3599.

Belzer, C., Gerber, G.K., Roeselers, G., Delaney, M., DuBois, A., Liu, Q., Belavusava, V., Yeliseyev, V., Houseman, A., Onderdonk, A., et al. (2014). Dynamics of the Microbiota in Response to Host Infection. PLOS ONE *9*, e95534.

Bentley, S.D., and Parkhill, J. (2015). Genomic perspectives on the evolution and spread of bacterial pathogens. Proc. Biol. Sci. *282*, 20150488.

Berg, R.D. (1996). The indigenous gastrointestinal microflora. Trends Microbiol. *4*, 430–435.

Bjursell, M.K., Martens, E.C., and Gordon, J.I. (2006). Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period. J. Biol. Chem. *281*, 36269–36279.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. (1997). The complete genome sequence of Escherichia coli K-12. Science *277*, 1453–1462.

Blount, Z.D. (2015). The unexhausted potential of E. coli. eLife *4*, e05826.

Blount, Z.D., Borland, C.Z., and Lenski, R.E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. *105*, 7899–7906.

Blundell, J.R., and Levy, S.F. (2014). Beyond genome sequencing: lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer. Genomics *104*, 417–430.

Brand, M.W., Wannemuehler, M.J., Phillips, G.J., Proctor, A., Overstreet, A.-M., Jergens, A.E., Orcutt, R.P., and Fox, J.G. (2015). The Altered Schaedler Flora: Continued Applications of a Defined Murine Microbial Community. ILAR J. *56*, 169–178.

Brandtzaeg, P., and Pabst, R. (2004). Let's go mucosal: communication on slippery ground. Trends Immunol. *25*, 570–577.

Brestoff, J.R., and Artis, D. (2013). Commensal bacteria at the interface of host metabolism and the immune system. Nat. Immunol. *14*, 676–684.

Cash, H.L., Whitham, C.V., Behrendt, C.L., and Hooper, L.V. (2006). Symbiotic Bacteria Direct Expression of an Intestinal Bactericidal Lectin. Science *313*, 1126–1130.

Cerutti, A. (2008). Location, location, location: B-cell differentiation in the gut lamina propria. Mucosal Immunol. *1*, 8–10.

Chang, D.E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., Anderson, A.B., Grissom, J.E., Laux, D.C., Cohen, P.S., et al. (2004). Carbon nutrition of Escherichia coli in the mouse intestine. Proc Natl Acad Sci USA *101*, 7427–7432.

Chaudhuri, R.R., and Henderson, I.R. (2012). The evolution of the Escherichia coli phylogeny. Infect. Genet. Evol. *12*, 214–226.

Clemente, J.C., Ursell, L.K., Parfrey, L.W., and Knight, R. (2012). The impact of the gut microbiota on human health: an integrative view. Cell *148*, 1258–1270.

Conway, T., and Cohen, P.S. (2015). Commensal and Pathogenic Escherichia coli Metabolism in the Gut. Microbiol. Spectr. *3*.

Cooper, T.F., Remold, S.K., Lenski, R.E., and Schneider, D. (2008). Expression profiles reveal parallel evolution of epistatic interactions involving the CRP regulon in Escherichia coli. PLoS Genet. *4*, e35.

D'Auria, G., Peris-Bondia, F., Džunková, M., Mira, A., Collado, M.C., Latorre, A., and Moya, A. (2013). Active and secreted IgA-coated bacterial fractions from the human gut reveal an under-represented microbiota core. Sci. Rep. *3*, 3515.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature *505*, 559–563.

Day, R.N., and Davidson, M.W. (2009). The fluorescent protein palette: tools for cellular imaging. Chem. Soc. Rev. *38*, 2887–2921.

De Paepe, M., Gaboriau-Routhiau, V., Rainteau, D., Rakotobe, S., Taddei, F., and Cerf-Bensussan, N. (2011). Trade-off between bile resistance and nutritional competence drives Escherichia coli diversification in the mouse gut. PLoS Genet *7*, e1002107.

Diard, M., Garry, L., Selva, M., Mosser, T., Denamur, E., and Matic, I. (2010). Pathogenicity-associated islands in extraintestinal pathogenic Escherichia coli are fitness elements involved in intestinal colonization. J. Bacteriol. *192*, 4885–4893.

Didelot, X., Walker, A.S., Peto, T.E., Crook, D.W., and Wilson, D.J. (2016). Within-host evolution of bacterial pathogens. Nat. Rev. Microbiol. *advance online publication*.

Dietrich, M.R., Ankeny, R.A., and Chen, P.M. (2014). Publication Trends in Model Organism Research. Genetics *198*, 787–794.

Donaldson, G.P., Lee, S.M., and Mazmanian, S.K. (2016). Gut biogeography of the bacterial microbiota. Nat. Rev. Microbiol. *14*, 20–32.

Duncan, S.H., Belenguer, A., Holtrop, G., Johnstone, A.M., Flint, H.J., and Lobley, G.E. (2007). Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. Appl. Environ. Microbiol. *73*, 1073–1078.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., and Relman, D.A. (2005). Diversity of the Human Intestinal Microbial Flora. Science *308*, 1635–1638.

El Aidy, S., van Baarlen, P., Derrien, M., Lindenbergh-Kortleve, D.J., Hooiveld, G., Levenez, F., Doré, J., Dekker, J., Samsom, J.N., Nieuwenhuis, E.E.S., et al. (2012). Temporal and spatial interplay of microbiota and intestinal mucosa drive

establishment of immune homeostasis in conventionalized mice. Mucosal Immunol. *5*, 567–579.

Elena, S.F., and Lenski, R.E. (2003). Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet *4*, 457–469.

Ericsson, A.C., Davis, J.W., Spollen, W., Bivens, N., Givan, S., Hagan, C.E., McIntosh, M., and Franklin, C.L. (2015). Effects of vendor and genetic background on the composition of the fecal microbiota of inbred mice. PloS One *10*, e0116704.

Fabich, A.J., Jones, S. a, Chowdhury, F.Z., Cernosek, A., Anderson, A., Smalley, D., McHargue, J.W., Hightower, G.A., Smith, J.T., Autieri, S.M., et al. (2008). Comparison of carbon nutrition for pathogenic and commensal Escherichia coli strains in the mouse intestine. Infect Immun *76*, 1143–1152.

Faith, J.J., McNulty, N.P., Rey, F.E., and Gordon, J.I. (2011). Predicting a Human Gut Microbiota's Response to Diet in Gnotobiotic Mice. Science *333*, 101–104.

Faith, J.J., Ahern, P.P., Ridaura, V.K., Cheng, J., and Gordon, J.I. (2014). Identifying gut microbe-host phenotype relationships using combinatorial communities in gnotobiotic mice. Sci. Transl. Med. *6*, 220ra11.

Favier, C.F., Vaughan, E.E., De Vos, W.M., and Akkermans, A.D.L. (2002). Molecular monitoring of succession of bacterial communities in human neonates. Appl. Environ. Microbiol. *68*, 219–226.

Feliziani, S., Marvig, R.L., Luján, A.M., Moyano, A.J., Di Rienzo, J.A., Krogh Johansen, H., Molin, S., and Smania, A.M. (2014). Coexistence and Within-Host Evolution of Diversified Lineages of Hypermutable Pseudomonas aeruginosa in Long-term Cystic Fibrosis Infections. PLoS Genet *10*, e1004651.

Flint, H.J., Scott, K.P., Duncan, S.H., Louis, P., and Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. Gut Microbes *3*, 289–306.

Foucault, M.-L., Thomas, L., Goussard, S., Branchini, B.R., and Grillot-Courvalin, C. (2010). In vivo bioluminescence imaging for the study of intestinal colonization by Escherichia coli in mice. Appl Env. Microbiol *76*, 264–274.

Fransen, F., Zagato, E., Mazzini, E., Fosso, B., Manzari, C., El Aidy, S., Chiavelli, A., D'Erchia, A.M., Sethi, M.K., Pabst, O., et al. (2015). BALB/c and C57BL/6 Mice Differ in Polyreactive IgA Abundance, which Impacts the Generation of Antigen-Specific IgA and Microbiota Diversity. Immunity *43*, 527–540.

Frese, S.A., Benson, A.K., Tannock, G.W., Loach, D.M., Kim, J., Zhang, M., Oh, P.L., Heng, N.C.K., Patil, P.B., Juge, N., et al. (2011). The evolution of host specialization in the vertebrate gut symbiont Lactobacillus reuteri. PLoS Genet. *7*, e1001314.

Gaboriau-Routhiau, V., Rakotobe, S., Lécuyer, E., Mulder, I., Lan, A., Bridonneau, C., Rochet, V., Pisi, A., De Paepe, M., Brandi, G., et al. (2009). The Key Role of Segmented Filamentous Bacteria in the Coordinated Maturation of Gut Helper T Cell Responses. Immunity *31*, 677–689.

Gaffé, J., McKenzie, C., Maharjan, R.P., Coursange, E., Ferenci, T., and Schneider, D. (2011). Insertion sequence-driven evolution of Escherichia coli in chemostats. J. Mol. Evol. *72*, 398–412.

Gauger, E.J., Leatham, M.P., Mercado-Lubo, R., Laux, D.C., Conway, T., and Cohen, P.S. (2007). Role of motility and the flhDC Operon in Escherichia coli MG1655 colonization of the mouse intestine. Infect Immun *75*, 3315–3324.

Gensollen, T., Iyer, S.S., Kasper, D.L., and Blumberg, R.S. (2016). How colonization by microbiota in early life shapes the immune system. Science *352*, 539–544.

Gerrish, P.J., and Lenski, R.E. (1998). The fate of competing beneficial mutations in an asexual population. Genetica *102–103*, 127–144.

Ghosh, A.R. (2013). Appraisal of Microbial Evolution to Commensalism and Pathogenicity in Humans. Clin. Med. Insights Gastroenterol. *6*, 1–12.

Gillilland, M.G., Erb-Downward, J.R., Bassis, C.M., Shen, M.C., Toews, G.B., Young, V.B., and Huffnagle, G.B. (2012). Ecological succession of bacterial communities during conventionalization of germ-free mice. Appl. Environ. Microbiol. *78*, 2359–2366.

Giraud, A., Arous, S., De Paepe, M., Gaboriau-Routhiau, V., Bambou, J.-C., Rakotobe, S., Lindner, A.B., Taddei, F., and Cerf-Bensussan, N. (2008). Dissecting the genetic components of adaptation of Escherichia coli to the mouse gut. PLoS Genet *4*, e2.

Golubchik, T., Batty, E.M., Miller, R.R., Farr, H., Young, B.C., Larner-Svensson, H., Fung, R., Godwin, H., Knox, K., Votintseva, A., et al. (2013). Within-host evolution of Staphylococcus aureus during asymptomatic carriage. PloS One *8*, e61319.

Gordo, I., Perfeito, L., and Sousa, A. (2011). Fitness effects of mutations in bacteria. J Mol Microbiol Biotechnol *21*, 20–35.

Gordon, D.M., O'Brien, C.L., and Pavli, P. (2015). Escherichia coli diversity in the lower intestinal tract of humans. Environ. Microbiol. Rep. *7*, 642–648.

Gresham, D., and Hong, J. (2015). The functional basis of adaptive evolution in chemostats. FEMS Microbiol. Rev. *39*, 2–16.

Hapfelmeier, S., Lawson, M.A.E., Slack, E., Kirundi, J.K., Stoel, M., Heikenwalder, M., Cahenzli, J., Velykoredko, Y., Balmer, M.L., Endt, K., et al. (2010). Reversible Microbial Colonization of Germ-Free Mice Reveals the Dynamics of IgA Immune Responses. Science *328*, 1705–1709.

Hasegawa, M., and Inohara, N. (2014). Regulation of the gut microbiota by the mucosal immune system in mice. Int. Immunol. *26*, 481–487.

Hegreness, M., Shoresh, N., Hartl, D., and Kishony, R. (2006). An equivalence principle for the incorporation of favorable mutations in asexual populations. Science *311*, 1615–1617.

Hindré, T., Knibbe, C., Beslon, G., and Schneider, D. (2012). New insights into bacterial adaptation through in vivo and in silico experimental evolution. Nat. Rev. Microbiol. *10*, 352–365.

Hobman, J.L., Penn, C.W., and Pallen, M.J. (2007). Laboratory strains of Escherichia coli: model citizens or deceitful delinquents growing old disgracefully? Mol. Microbiol. *64*, 881–885.

Hornef, M. (2015). Pathogens, Commensal Symbionts, and Pathobionts: Discovery and Functional Effects on the Host. ILAR J. Natl. Res. Counc. Inst. Lab. Anim. Resour. *56*, 159–162.

Ihssen, J., Grasselli, E., Bassin, C., François, P., Piffaretti, J.-C., Köster, W., Schrenzel, J., and Egli, T. (2007). Comparative genomic hybridization and physiological characterization of environmental isolates indicate that significant (eco-)physiological properties are highly conserved in the species Escherichia coli. Microbiol. Read. Engl. *153*, 2052–2066.

Illingworth, C.J.R., and Mustonen, V. (2012). A method to infer positive selection from marker dynamics in an asexual population. Bioinformatics *28*, 831–837.

Ivanov, I.I., Atarashi, K., Manel, N., Brodie, E.L., Shima, T., Karaoz, U., Wei, D., Goldfarb, K.C., Santee, C.A., Lynch, S.V., et al. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell *139*, 485–498.

Johansson, M.E.V., Phillipson, M., Petersson, J., Velcich, A., Holm, L., and Hansson, G.C. (2008). The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. Proc. Natl. Acad. Sci. U. S. A. *105*, 15064–15069.

Johnson, J.R., Clabots, C., and Kuskowski, M.A. (2008). Multiple-Host Sharing, Long-Term Persistence, and Virulence of Escherichia coli Clones from Human and Animal Household Members. J. Clin. Microbiol. *46*, 4078–4082.

Jones, S.A., Chowdhury, F.Z., Fabich, A.J., Anderson, A., Schreiner, D.M., House, A.L., Autieri, S.M., Leatham, M.P., Lins, J.J., Jorgensen, M., et al. (2007). Respiration of Escherichia coli in the mouse intestine. Infect Immun *75*, 4891–4899.

Jones, S.A., Gibson, T., Maltby, R.C., Chowdhury, F.Z., Stewart, V., Cohen, P.S., and Conway, T. (2011). Anaerobic respiration of Escherichia coli in the mouse intestine. Infect. Immun. *79*, 4218–4226.

Josefowicz, S.Z., Niec, R.E., Kim, H.Y., Treuting, P., Chinen, T., Zheng, Y., Umetsu, D.T., and Rudensky, A.Y. (2012). Extrathymically generated regulatory T cells control mucosal TH2 inflammation. Nature *482*, 395–399.

Kaas, R.S., Friis, C., Ussery, D.W., and Aarestrup, F.M. (2012). Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes. BMC Genomics *13*, 577.

Kamada, N., and Núñez, G. (2013). Role of the gut microbiota in the development and function of lymphoid cells. J. Immunol. Baltim. Md 1950 *190*, 1389–1395.

Kaper, J.B., Nataro, J.P., and Mobley, H.L.T. (2004). Pathogenic Escherichia coli. Nat. Rev. Microbiol. *2*, 123–140.

Kawamoto, S., Maruya, M., Kato, L.M., Suda, W., Atarashi, K., Doi, Y., Tsutsui, Y., Qin, H., Honda, K., Okada, T., et al. (2014). Foxp3(+) T cells regulate immunoglobulin A selection and facilitate diversification of bacterial species responsible for immune homeostasis. Immunity *41*, 152–165.

Kawecki, T.J., Lenski, R.E., Ebert, D., Hollis, B., Olivieri, I., and Whitlock, M.C. (2012). Experimental evolution. Trends Ecol. Evol. *27*, 547–560.

Kimura, M., and Ota, T. (1969). The average number of generations until extinction of an individual mutant gene in a finite population. Genetics *63*, 701–709.

Klemm, E., and Dougan, G. (2016). Advances in Understanding Bacterial Pathogenesis Gained from Whole-Genome Sequencing and Phylogenetics. Cell Host Microbe *19*, 599–610.

Kostic, A.D., Howitt, M.R., and Garrett, W.S. (2013). Exploring host-microbiota interactions in animal models and humans. Genes Dev. *27*, 701–718.

Kreisinger, J., Cížková, D., Vohánka, J., and Piálek, J. (2014). Gastrointestinal microbiota of wild and inbred individuals of two house mouse subspecies assessed using high-throughput parallel pyrosequencing. Mol. Ecol. *23*, 5048–5060.

Lang, G.I., and Desai, M.M. (2014). The spectrum of adaptive mutations in experimental evolution. Genomics *104*, 412–416.

Langille, M.G., Meehan, C.J., Koenig, J.E., Dhanani, A.S., Rose, R.A., Howlett, S.E., and Beiko, R.G. (2014). Microbial shifts in the aging mouse gut. Microbiome *2*, 50.

Lasaro, M., Liu, Z., Bishar, R., Kelly, K., Chattopadhyay, S., Paul, S., Sokurenko, E., Zhu, J., and Goulian, M. (2014). Escherichia coli Isolate for Studying Colonization of the Mouse Intestine and Its Application to Two-Component Signaling Knockouts. J. Bacteriol. *196*, 1723–1732.

Laukens, D., Brinkman, B.M., Raes, J., De Vos, M., and Vandenabeele, P. (2016). Heterogeneity of the gut microbiome in mice: guidelines for optimizing experimental design. FEMS Microbiol. Rev. *40*, 117–132.

Leatham, M.P., Stevenson, S.J., Gauger, E.J., Krogfelt, K.A., Lins, J.J., Haddock, T.L., Autieri, S.M., Conway, T., and Cohen, P.S. (2005). Mouse Intestine Selects

Nonmotile flhDC Mutants of Escherichia coli MG1655 with Increased Colonizing Ability and Better Utilization of Carbon Sources. Infect Immun *73*, 8039–8049.

Leatham, M.P., Banerjee, S., Autieri, S.M., Mercado-Lubo, R., Conway, T., and Cohen, P.S. (2009). Precolonized human commensal Escherichia coli strains serve as a barrier to E. coli O157:H7 growth in the streptomycin-treated mouse intestine. Infect Immun *77*, 2876–2886.

Leatham-Jensen, M.P., Frimodt-Møller, J., Adediran, J., Mokszycki, M.E., Banner, M.E., Caughron, J.E., Krogfelt, K.A., Conway, T., and Cohen, P.S. (2012). The Streptomycin-Treated Mouse Intestine Selects Escherichia coli envZ Missense Mutants That Interact with Dense and Diverse Intestinal Microbiota. Infect Immun *80*, 1716–1727.

LeBlanc, J.G., Milani, C., de Giori, G.S., Sesma, F., van Sinderen, D., and Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. Curr. Opin. Biotechnol. *24*, 160–168.

Lescat, M., Launay, A., Ghalayini, M., Magnan, M., Glodt, J., Pintard, C., Dion, S., Denamur, E., and Tenaillon, O. (2016). Using long-term experimental evolution to uncover the patterns and determinants of molecular evolution of an Escherichia coli natural isolate in the streptomycin treated mouse gut. Mol. Ecol.

Levin, B.R. (1988). Frequency-dependent selection in bacterial populations. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *319*, 459–472.

Ley, R.E., Bäckhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005). Obesity alters gut microbial ecology. Proc. Natl. Acad. Sci. U. S. A. *102*, 11070–11075.

Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R., and Gordon, J.I. (2008a). Worlds within worlds: evolution of the vertebrate gut microbiota. Nat. Rev. Microbiol. *6*, 776–788.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., et al. (2008b). Evolution of mammals and their gut microbes. Science *320*, 1647–1651.

Li, H., Limenitakis, J.P., Fuhrer, T., Geuking, M.B., Lawson, M.A., Wyss, M., Brugiroux, S., Keller, I., Macpherson, J.A., Rupp, S., et al. (2015). The outer mucus layer hosts a distinct intestinal microbial niche. Nat. Commun. *6*, 8292.

Loewe, L., and Hill, W.G. (2010). The population genetics of mutations: good, bad and indifferent. Philos. Trans. R. Soc. B Biol. Sci. *365*, 1153–1167.

Long, A., Liti, G., Luptak, A., and Tenaillon, O. (2015). Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nat. Rev. Genet. *16*, 567–582.

Louis, P., Hold, G.L., and Flint, H.J. (2014). The gut microbiota, bacterial metabolites and colorectal cancer. Nat. Rev. Microbiol. *12*, 661–672.

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. Nature *489*, 220–230.

Macpherson, A.J., Gatto, D., Sainsbury, E., Harriman, G.R., Hengartner, H., and Zinkernagel, R.M. (2000). A Primitive T Cell-Independent Mechanism of Intestinal Mucosal IgA Responses to Commensal Bacteria. Science *288*, 2222–2226.

Macpherson, A.J., Hunziker, L., McCoy, K., and Lamarre, A. (2001). IgA responses in the intestinal mucosa against pathogenic and non-pathogenic microorganisms. Microbes Infect. Inst. Pasteur *3*, 1021–1035.

Maddamsetti, R., Lenski, R.E., and Barrick, J.E. (2015). Adaptation, Clonal Interference, and Frequency-Dependent Interactions in a Long-Term Evolution Experiment with Escherichia coli. Genetics genetics.115.176677.

Maharjan, R.P., Ferenci, T., Reeves, P.R., Li, Y., Liu, B., and Wang, L. (2012). The multiplicity of divergence mechanisms in a single evolving population. Genome Biol. *13*, R41.

Maharjan, R.P., Liu, B., Li, Y., Reeves, P.R., Wang, L., and Ferenci, T. (2013). Mutation accumulation and fitness in mutator subpopulations of Escherichia coli. Biol. Lett. *9*, 20120961.

Maharjan, R.P., Liu, B., Feng, L., Ferenci, T., and Wang, L. (2015). Simple Phenotypic Sweeps Hide Complex Genetic Changes in Populations. Genome Biol. Evol. *7*, 531–544.

Maharshak, N., Packey, C.D., Ellermann, M., Manick, S., Siddle, J.P., Huh, E.Y., Plevy, S., Sartor, R.B., and Carroll, I.M. (2013). Altered enteric microbiota ecology in interleukin 10-deficient mice during development and progression of intestinal inflammation. Gut Microbes *4*, 316–324.

Mahillon, J., and Chandler, M. (1998). Insertion Sequences. Microbiol. Mol. Biol. Rev. *62*, 725–774.

Maltby, R., Leatham-Jensen, M.P., Gibson, T., Cohen, P.S., and Conway, T. (2013). Nutritional basis for colonization resistance by human commensal Escherichia coli strains HS and Nissle 1917 against E. coli O157:H7 in the mouse intestine. PloS One *8*, e53957.

Marvig, R.L., Johansen, H.K., Molin, S., and Jelsbak, L. (2013). Genome Analysis of a Transmissible Lineage of Pseudomonas aeruginosa Reveals Pathoadaptive Mutations and Distinct Evolutionary Paths of Hypermutators. PLoS Genet *9*, e1003741.

Marvig, R.L., Sommer, L.M., Molin, S., and Johansen, H.K. (2015). Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. Nat. Genet. *47*, 57–64.

Maurice, C.F., CL Knowles, S., Ladau, J., Pollard, K.S., Fenton, A., Pedersen, A.B., and Turnbaugh, P.J. (2015). Marked seasonal variation in the wild mouse gut microbiota. ISME J. *9*, 2423–2434.

Maynard, C.L., Elson, C.O., Hatton, R.D., and Weaver, C.T. (2012). Reciprocal interactions of the intestinal microbiota and immune system. Nature *489*, 231–241.

McGhee, J.R., and Fujihashi, K. (2012). Inside the Mucosal Immune System. PLOS Biol *10*, e1001397.

Meador, J.P., Caldwell, M.E., Cohen, P.S., and Conway, T. (2014). Escherichia coli pathotypes occupy distinct niches in the mouse intestine. Infect. Immun. *82*, 1931–1938.

Mohawk, K.L., and O'Brien, A.D. (2011). Mouse models of Escherichia coli O157:H7 infection and shiga toxin injection. J. Biomed. Biotechnol. *2011*, 258185.

Molloy, M., Bouladoux, N., and Belkaid, Y. (2012). Intestinal Microbiota: Shaping local and systemic immune responses. Semin. Immunol. *24*, 58–66.

Moreau, M.C., Ducluzeau, R., Guy-Grand, D., and Muller, M.C. (1978). Increase in the population of duodenal immunoglobulin A plasmocytes in axenic mice associated with different living or dead bacterial strains of intestinal origin. Infect. Immun. *21*, 532–539.

Mowat, A.M., and Agace, W.W. (2014). Regional specialization within the intestinal immune system. Nat. Rev. Immunol. *14*, 667–685.

Mowat, A.M., and Viney, J.L. (1997). The anatomical basis of intestinal immunity. Immunol. Rev. *156*, 145–166.

Myhrvold, C., Kotula, J.W., Hicks, W.M., Conway, N.J., and Silver, P.A. (2015). A distributed cell division counter reveals growth dynamics in the gut microbiota. Nat. Commun. *6*, 10039.

Nguyen, T.L.A., Vieira-Silva, S., Liston, A., and Raes, J. (2015). How informative is the mouse for human gut microbiota research? Dis. Model. Mech. *8*, 1–16.

Oh, P.L., Benson, A.K., Peterson, D.A., Patil, P.B., Moriyama, E.N., Roos, S., and Walter, J. (2010). Diversification of the gut symbiont Lactobacillus reuteri as a result of host-driven evolution. ISME J. *4*, 377–387.

Olson-Manning, C.F., Wagner, M.R., and Mitchell-Olds, T. (2012). Adaptive evolution: evaluating empirical support for theoretical predictions. Nat. Rev. Genet. *13*, 867–877.

Orr, H.A. (2010). The population genetics of beneficial mutations. Philos. Trans. R. Soc. B Biol. Sci. *365*, 1195–1201.

Palm, N.W., de Zoete, M.R., Cullen, T.W., Barry, N.A., Stefanowski, J., Hao, L., Degnan, P.H., Hu, J., Peter, I., Zhang, W., et al. (2014). Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. Cell *158*, 1000–1010.

Palm, N.W., de Zoete, M.R., and Flavell, R.A. (2015). Immune-microbiota interactions in health and disease. Clin. Immunol. Orlando Fla *159*, 122–127.

Perlman, R.L. (2016). Mouse Models of Human Disease: An Evolutionary Perspective. Evol. Med. Public Health eow014.

Peterson, D.A., McNulty, N.P., Guruge, J.L., and Gordon, J.I. (2007). IgA response to symbiotic bacteria as a mediator of gut homeostasis. Cell Host Microbe *2*, 328–339.

Petnicki-Ocwieja, T., Hrncir, T., Liu, Y.-J., Biswas, A., Hudcovic, T., Tlaskalova-Hogenova, H., and Kobayashi, K.S. (2009). Nod2 is required for the regulation of commensal microbiota in the intestine. Proc. Natl. Acad. Sci. U. S. A. *106*, 15813–15818.

Phifer-Rixey, M., and Nachman, M.W. (2015). Insights into mammalian biology from the wild house mouse Mus musculus. eLife *4*, e05959.

Philippe, N., Crozat, E., Lenski, R.E., and Schneider, D. (2007). Evolution of global regulatory networks during a long-term experiment with Escherichia coli. BioEssays News Rev. Mol. Cell. Dev. Biol. *29*, 846–860.

Philippe, N., Pelosi, L., Lenski, R.E., and Schneider, D. (2009). Evolution of Penicillin-Binding Protein 2 Concentration and Cell Shape during a Long-Term Experiment with Escherichia coli. J. Bacteriol. *191*, 909–921.

Poulsen, L.K., Lan, F., Kristensen, C.S., Hobolth, P., Molin, S., and Krogfelt, K.A. (1994). Spatial distribution of Escherichia coli in the mouse large intestine inferred from rRNA in situ hybridization. Infect. Immun. *62*, 5191–5194.

Poulsen, L.K., Licht, T.R., Rang, C., Krogfelt, K.A., and Molin, S. (1995). Physiological state of Escherichia coli BJ4 growing in the large intestines of streptomycin-treated mice. J Bacteriol *177*, 5840–5845.

Raeside, C., Gaffé, J., Deatherage, D.E., Tenaillon, O., Briska, A.M., Ptashkin, R.N., Cruveiller, S., Médigue, C., Lenski, R.E., Barrick, J.E., et al. (2014). Large Chromosomal Rearrangements during a Long-Term Evolution Experiment with Escherichia coli. mBio *5*.

Rang, C.U., Licht, T.R., Midtvedt, T., Conway, P.L., Chao, L., Krogfelt, K.A., Cohen, P.S., and Molin, S. (1999). Estimation of growth rates of Escherichia coli BJ4 in streptomycin-treated and previously germfree mice by in situ rRNA hybridization. Clin. Diagn. Lab. Immunol. *6*, 434–436.

Rasko, D.A., Rosovitz, M.J., Myers, G.S.A., Mongodin, E.F., Fricke, W.F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N.R., Chaudhuri, R., et al. (2008). The Pangenome Structure of Escherichia coli: Comparative Genomic Analysis of E. coli Commensal and Pathogenic Isolates. J. Bacteriol. *190*, 6881–6893.

Reeves, P.R., Liu, B., Zhou, Z., Li, D., Guo, D., Ren, Y., Clabots, C., Lan, R., Johnson, J.R., and Wang, L. (2011). Rates of Mutation and Host Transmission for an Escherichia coli Clone over 3 Years. PLOS ONE *6*, e26907.

Rigoni, R., Fontana, E., Guglielmetti, S., Fosso, B., D'Erchia, A.M., Maina, V., Taverniti, V., Castiello, M.C., Mantero, S., Pacchiana, G., et al. (2016). Intestinal microbiota sustains inflammation and autoimmunity induced by hypomorphic RAG defects. J. Exp. Med. *213*, 355–375.

Robertson, S.J., Zhou, J.Y., Geddes, K., Rubino, S.J., Cho, J.H., Girardin, S.E., and Philpott, D.J. (2013). Nod1 and Nod2 signaling does not alter the composition of intestinal bacterial communities at homeostasis. Gut Microbes *4*, 222–231.

Rosenthal, N., and Brown, S. (2007). The mouse ascending: perspectives for human-disease models. Nat. Cell Biol. *9*, 993–999.

Round, J.L., and Mazmanian, S.K. (2009). The gut microbiome shapes intestinal immune responses during health and disease. Nat. Rev. Immunol. *9*, 313–323.

Round, J.L., and Mazmanian, S.K. (2010). Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. Proc. Natl. Acad. Sci. U. S. A. *107*, 12204–12209.

Savkovic, S.D., Villanueva, J., Turner, J.R., Matkowskyj, K.A., and Hecht, G. (2005). Mouse model of enteropathogenic Escherichia coli infection. Infect. Immun. *73*, 1161–1170.

Schinner, S.A.C., Mokszycki, M.E., Adediran, J., Leatham-Jensen, M., Conway, T., and Cohen, P.S. (2015). Escherichia coli EDL933 requires gluconeogenic nutrients to successfully colonize the intestines of streptomycin-treated mice precolonized with E. coli Nissle 1917. Infect. Immun. *83*, 1983–1991.

Scholz, F., Badgley, B.D., Sadowsky, M.J., and Kaplan, D.H. (2014). Immune mediated shaping of microflora community composition depends on barrier site. PloS One *9*, e84019.

Schwab, C., Berry, D., Rauch, I., Rennisch, I., Ramesmayer, J., Hainzl, E., Heider, S., Decker, T., Kenner, L., Müller, M., et al. (2014). Longitudinal study of murine

microbiota activity and interactions with the host during acute inflammation and recovery. ISME J. *8*, 1101–1114.

Scott, K.P., Gratz, S.W., Sheridan, P.O., Flint, H.J., and Duncan, S.H. (2013). The influence of diet on the gut microbiota. Pharmacol. Res. *69*, 52–60.

Sender, R., Fuchs, S., and Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol. *14*, e1002533.

Sheridan, B.S., and Lefrançois, L. (2010). Intraepithelial lymphocytes: to serve and protect. Curr. Gastroenterol. Rep. *12*, 513–521.

Shroff, K.E., Meslin, K., and Cebra, J.J. (1995). Commensal enteric bacteria engender a self-limiting humoral mucosal immune response while permanently colonizing the gut. Infect. Immun. *63*, 3904–3913.

Shulzhenko, N., Morgun, A., Hsiao, W., Battle, M., Yao, M., Gavrilova, O., Orandle, M., Mayer, L., Macpherson, A.J., McCoy, K.D., et al. (2011). Crosstalk between B lymphocytes, microbiota and the intestinal epithelium governs immunity versus metabolism in the gut. Nat. Med. *17*, 1585–1593.

Smati, M., Clermont, O., Bleibtreu, A., Fourreau, F., David, A., Daubié, A.-S., Hignard, C., Loison, O., Picard, B., and Denamur, E. (2015). Quantitative analysis of commensal Escherichia coli populations reveals host-specific enterotypes at the intra-species level. MicrobiologyOpen *4*, 604–615.

Smith, K., McCoy, K.D., and Macpherson, A.J. (2007). Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. Semin. Immunol. *19*, 59–69.

Sniegowski, P.D., Gerrish, P.J., and Lenski, R.E. (1997). Evolution of high mutation rates in experimental populations of E. coli. Nature *387*, 703–705.

Snoeck, V., Goddeeris, B., and Cox, E. (2005). The role of enterocytes in the intestinal barrier function and antigen uptake. Microbes Infect. *7*, 997–1004.

Souza, V., Rocha, M., Valera, A., and Eguiarte, L.E. (1999). Genetic Structure of Natural Populations of Escherichia coli in Wild Hosts on Different Continents. Appl. Environ. Microbiol. *65*, 3373–3385.

Spees, A.M., Wangdi, T., Lopez, C.A., Kingsbury, D.D., Xavier, M.N., Winter, S.E., Tsolis, R.M., and Bäumler, A.J. (2013). Streptomycin-induced inflammation enhances Escherichia coli gut colonization through nitrate respiration. mBio *4*.

Spor, A., Koren, O., and Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. Nat. Rev. Microbiol. *9*, 279–290.

Suzuki, K., Meek, B., Doi, Y., Muramatsu, M., Chiba, T., Honjo, T., and Fagarasan, S. (2004). Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. Proc. Natl. Acad. Sci. U. S. A. *101*, 1981–1986.

Talham, G.L., Jiang, H.Q., Bos, N.A., and Cebra, J.J. (1999). Segmented filamentous bacteria are potent stimuli of a physiologically normal state of the murine gut mucosal immune system. Infect. Immun. *67*, 1992–2000.

Tenaillon, O., Skurnik, D., Picard, B., and Denamur, E. (2010). The population genetics of commensal Escherichia coli. Nat. Rev. Microbiol. *8*, 207–217.

Thoene-Reineke, C., Fischer, A., Friese, C., Briesemeister, D., Göbel, U.B., Kammertoens, T., Bereswill, S., and Heimesaat, M.M. (2014). Composition of Intestinal Microbiota in Immune-Deficient Mice Kept in Three Different Housing Conditions. PLoS ONE *9*, e113406.

Thompson, J.A., Oliveira, R.A., Djukovic, A., Ubeda, C., and Xavier, K.B. (2015). Manipulation of the Quorum Sensing Signal AI-2 Affects the Antibiotic-Treated Gut Microbiota. Cell Rep. *10*, 1861–1871.

Ubeda, C., Lipuma, L., Gobourne, A., Viale, A., Leiner, I., Equinda, M., Khanin, R., and Pamer, E.G. (2012). Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. J. Exp. Med. *209*, 1445–1456.

Vanhooren, V., and Libert, C. (2013). The mouse as a model organism in aging research: Usefulness, pitfalls and possibilities. Ageing Res. Rev. *12*, 8–21.

Vejborg, R.M., Hancock, V., Schembri, M.A., and Klemm, P. (2011). Comparative genomics of Escherichia coli strains causing urinary tract infections. Appl. Environ. Microbiol. *77*, 3268–3278.

Vijay-Kumar, M., Aitken, J.D., Carvalho, F.A., Cullender, T.C., Mwangi, S., Srinivasan, S., Sitaraman, S.V., Knight, R., Ley, R.E., and Gewirtz, A.T. (2010). Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. Science *328*, 228–231.

Virgin, H.W. (2014). The virome in mammalian physiology and disease. Cell *157*, 142–150.

van der Waaij, D., Berghuis-de Vries, J.M., and Lekkerkerk Lekkerkerk-v (1971). Colonization resistance of the digestive tract in conventional and antibiotic-treated mice. J. Hyg. (Lond.) *69*, 405–411.

van der Waaij, L.A., Limburg, P.C., Mesander, G., and van der Waaij, D. (1996). In vivo IgA coating of anaerobic bacteria in human faeces. Gut *38*, 348–354.

Wadolkowski, E.A., Laux, D.C., and Cohen, P.S. (1988). Colonization of the streptomycin-treated mouse large intestine by a human fecal Escherichia coli strain: role of growth in mucus. Infect. Immun. *56*, 1030–1035.

Walter, J., Britton, R.A., and Roos, S. (2011). Host-microbial symbiosis in the vertebrate gastrointestinal tract and the Lactobacillus reuteri paradigm. Proc. Natl. Acad. Sci. U. S. A. *108 Suppl 1*, 4645–4652.

Weldon, L., Abolins, S., Lenzi, L., Bourne, C., Riley, E.M., and Viney, M. (2015). The Gut Microbiota of Wild Mice. PloS One *10*, e0134643.

Yang, I., Eibach, D., Kops, F., Brenneke, B., Woltemate, S., Schulze, J., Bleich, A., Gruber, A.D., Muthupalani, S., Fox, J.G., et al. (2013). Intestinal microbiota

composition of interleukin-10 deficient C57BL/6J mice and susceptibility to Helicobacter hepaticus-induced colitis. PloS One *8*, e70783.

Yang, L., Jelsbak, L., Marvig, R.L., Damkiær, S., Workman, C.T., Rau, M.H., Hansen, S.K., Folkesson, A., Johansen, H.K., Ciofu, O., et al. (2011). Evolutionary dynamics of bacteria in a human host environment. Proc. Natl. Acad. Sci. U. S. A. *108*, 7481–7486.

Young, B.C., Golubchik, T., Batty, E.M., Fung, R., Larner-Svensson, H., Votintseva, A.A., Miller, R.R., Godwin, H., Knox, K., Everitt, R.G., et al. (2012). Evolutionary dynamics of Staphylococcus aureus during progression from carriage to disease. Proc. Natl. Acad. Sci. *109*, 4550–4555.

Zhang, C., Zhang, M., Pang, X., Zhao, Y., Wang, L., and Zhao, L. (2012). Structural resilience of the gut microbiota in adult mice under high-fat dietary perturbations. ISME J. *6*, 1848–1857.

Zhang, H., Sparks, J.B., Karyala, S.V., Settlage, R., and Luo, X.M. (2015). Host adaptive immunity alters gut microbiota. ISME J. *9*, 770–781.

CHAPTER 2

The first steps of adaptation of *Escherichia coli* to the gut are dominated by soft sweeps

Manuscript published in *PLoS Genetics*, March 6th, 2014

The author of this thesis performed all the *in vivo* experiments, including the evolution experiment and the competitive fitness assays, the simulations, the growth curves, the phenotypic dynamics of the *gat* mutants and the fluctuation assays described in this chapter.

The first steps of adaptation of *Escherichia coli* to the gut are dominated by soft sweeps

João Barroso-Batista[1+], Ana Sousa[1+], Marta Lourenço, Marie-Louise Bergman[1], Daniel Sobral[1], Jocelyne Demengeot[1], Karina B. Xavier[1,2] and Isabel Gordo[1]*

[1]Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6
[2]Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa
+equal contribution
*author for correspondence: igordo@igc.gulbenkian.pt

Abstract

The accumulation of adaptive mutations is essential for survival in novel environments. However, in clonal populations with a high mutational supply, the power of natural selection is expected to be limited. This is due to clonal interference - the competition of clones carrying different beneficial mutations - which leads to the loss of many small effect mutations and fixation of large effect ones. If interference is abundant, then mechanisms for horizontal transfer of genes, which allow the immediate combination of beneficial alleles in a single background, are expected to evolve. However, the relevance of interference in natural complex environments, such as the gut, is poorly known. To address this issue, we have developed an experimental system which allows to uncover the nature of the adaptive process as *Escherichia coli* adapts to the mouse gut. This system shows the invasion of beneficial mutations in the bacterial populations and demonstrates the pervasiveness of clonal interference. The observed dynamics of change in frequency of beneficial mutations are consistent with soft sweeps, where different adaptive mutations with similar phenotypes, arise repeatedly on different haplotypes without reaching fixation. Despite the complexity of this ecosystem, the genetic basis of the adaptive mutations revealed a striking parallelism in independently evolving populations. This was mainly characterized

by the insertion of transposable elements in both coding and regulatory regions of a few genes. Interestingly, in most populations we observed a complete phenotypic sweep without loss of genetic variation. The intense clonal interference during adaptation to the gut environment, here demonstrated, may be important for our understanding of the levels of strain diversity of *E. coli* inhabiting the human gut microbiota and of its recombination rate.

Introduction

Mutation is the fuel of evolution and beneficial mutations the driver of organismal adaptation. When a small group of organisms founds a new niche it will rely on *de novo* mutations to adjust to such novel environment. If the rate of emergence of new beneficial mutations is low, adaptation will proceed through the asynchronous accumulation of such mutations, one at a time. This will result in short-term polymorphism, during which the frequency of the beneficial mutation will rapidly change. Such strong selective sweeps will purge linked neutral variation, a phenomenon long recognized as the signature of selective sweeps in bacterial populations (Atwood et al., 1951). Thus, for reasonably small populations and for low mutation rates, population mean fitness will increase in steps, where at each step neutral variation becomes completely depleted. The adaptive walk will therefore proceed by discrete movements along the fitness landscape.

Over the years studies of microbial adaptation in different environments have repeatedly detected deviations from this simple pattern (Hegreness et al., 2006; Kao and Sherlock, 2008; Maharjan et al., 2006; Perfeito et al., 2007; Woods et al., 2011). It is now much more commonly accepted that, in reasonably large microbial populations, many distinct adaptive mutations may arise and compete for fixation. The pattern of microbial adaptation has therefore been supportive of an evolutionary mechanism described in the sixties - the Hill-Roberston effect (Hill and Robertson, 1966). This has been coined clonal interference (CI) in the context of clonal bacterial populations (reviewed in (Sniegowski and Gerrish, 2010)). This effect is theoretically expected to limit the speed of adaptation in asexual versus sexual populations (Barton and Partridge, 2000). A great number of beneficial

mutations are lost and the distribution of mutations that fix can be greatly affected by interference (Gordo et al., 2011; Schiffels et al., 2011; Sniegowski and Gerrish, 2010).

But how strong is CI and what consequences does it entail? While it has been shown to occur in bacteria (Herron and Doebeli, 2013; Perfeito et al., 2007) and eukaryotes (Desai et al., 2007; Kao and Sherlock, 2008) in laboratory settings, its relevance in natural environments is poorly known. Interestingly, CI has recently been inferred to be an important determinant of the evolution of the influenza virus (Strelkowa and Lässig, 2012). Several examples from the study of rapid adaptation in natural populations have also shown the violation of the classical hard selective sweep model as well as the assumption of the mutation limited regime of adaptation. In contrast multiple adaptive alleles at the same locus can sweep through the populations at the same time, a phenomenon known as soft sweeps. These alleles can emerge from *de novo* mutation or from standing genetic variation (see (Messer and Petrov, 2013) for a revision). The phenomenon of CI is theoretically expected to impact the dynamics of adaptation if the effective population size (N_e) and/or the rate of occurrence of beneficial mutations (U_b) is large. More specifically, when the number of competing mutations is bigger than one, that is if $2N_e\ U_b\ \mathrm{Ln}(N_e\ s_b/2) > 1$ (where s_b is the mean effect of a beneficial mutation (Sniegowski and Gerrish, 2010)). Desai and Fisher (Desai and Fisher, 2007) made the case that in extremely large populations with very high mutational inputs, haplotypes with multiple beneficial mutations are expected to arise and increase in frequency. Furthermore, important advances of the theory of CI have also recently been made (Good et al., 2012; Lassig, 2012; Schiffels et al., 2011). Since the parameter values important to determine the importance and level of interference are expected to be dependent on the environment, the relevance of CI for bacterial evolution in natural conditions remains to be demonstrated.

E. coli K-12 has been an important model organism in many fields of biology, since its isolation from human feces in 1922 (Lederberg, 2004). Accumulation of mutations during its adaptation to the gut has been observed (De Paepe et al., 2011; Fabich et al., 2011; Leatham et al., 2005), but the fitness landscape

characteristic of this environment as well as the extent of interference *E. coli* experiences is not known.

We have studied the process of accumulation of beneficial mutations and the strength of their effects in a natural environment of *E. coli*, the mouse gut. We found that CI is pervasive *in vivo* and described the genetic basis of the initial steps of adaptation, which were observed to exhibit a striking parallelism. Remarkably we have found that in the same population, distinct mutations with equivalent functional effects (i.e., targeting the same gene or operon) reach detectable frequencies simultaneously. This leads to the occurrence of a phenotypic hard sweep without loss of variation at the genetic level. However, most of these mutations get extinct after a few generations, a signature of soft sweeps (Lee and Marx, 2013).

Material 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.

E. coli strains

All strains used were derived from *Escherichia coli* K-12, strain MG1655 (Blattner et al., 1997).

Strains DM08-YFP and DM09-CFP (MG1655, *galK::YFP/CFP* amp^R *(pZ12),* str^R *(rpsl150), ΔlacIZYA*) were used in the initial colonization experiment. These strains contain the yellow (*yfp*) or cyan (*cfp*) fluorescent genes linked to amp^R in the *galK* locus under the control of a *lac* promoter and were obtained by P1 transduction from previously constructed strains (Hegreness et al., 2006). To ensure constitutive expression of the fluorescent proteins the *lac* operon was deleted using the Datsenko and Wanner method (Datsenko and Wanner, 2000).

Fluorescent marker dynamics during mouse colonization

To study *E. coli* adaptation to the gut we used a streptomycin-treated mouse colonization model (T. Conway et al., 2004). Briefly, 6- to 8-week old C57BL/6 male mice raised in specific pathogen free (SPF) conditions were given autoclaved drinking water containing streptomycin (5g/L) for one day. After 4 hours of starvation for water and food the animals were gavaged with 100 µl of a suspension of 10^8 colony forming units (CFUs) of a mixture of YFP- and CFP-labeled bacteria (ratio 1:1) grown at 37°C in brain heart infusion medium to OD_{600} of 2. After gavage, the animals were housed separately and both food and water containing streptomycin were returned to them. Mice fecal pellets were collected for 24 days, diluted in PBS and plated in Luria Broth agar (LB agar). Plates were incubated overnight and the frequencies of CFP- or YFP-labeled bacteria were assessed by counting the fluorescent colonies with the help of a fluorescent stereoscope (SteREO Lumar, Carl Zeiss). Plating the fecal samples allowed checking for morphological phenotypic diversity of the colonies, a phenomenon reported previously in previous studies of adaptation of *E. coli* to the mouse gut (De Paepe et al., 2011; Fabich et al., 2011; Gauger et al., 2007; Giraud et al., 2008; Leatham et al., 2005). However, no consistent changes in colony morphology were observed during the evolution experiments.

A sample of each collected fecal pellet was daily stored in 15% glycerol at -80ºC for future experiments.

For the initial colonization a group of 5 mice was gavaged and followed for 24 days, and this procedure was repeated for two more groups of 5 mice. Thus a total of 15 mice were analyzed (Fig. 1a and Supplementary Fig. 1).

Competitive fitness assays *in vivo* to test for adaptation

We measured the relative fitness *in vivo* by bacterial clones isolated from mouse fecal samples after 24 days of colonization (approximately 400 generations). Individual colonies or mixtures of approximately 30 colonies with the same fluorescent marker (population of clones) were picked from mouse fecal platings, grown in 10 ml of Luria Broth medium (LB) supplemented with ampicillin (100 µg/ml) and streptomycin (100 µg/ml) and stored in 15% glycerol at -80ºC.

The isolated clones were competed against the ancestral strain labeled with the opposite fluorescent marker, at a ratio of 1 to 1, in 1-2 day co-colonization experiments following the same procedure described for the evolution experiments. The selective coefficient (fitness gain) of these clones *in vivo* (presented in Fig. 1b) was estimated as:

$$s_b = \ln\left(\frac{Rf_{ev/anc}}{Ri_{ev/anc}}\right)/t$$

where s_b is the selective coefficient of the evolved clone, $Rf_{ev/anc}$ *and* $Ri_{ev/anc}$ are the ratios of evolved to ancestral bacteria in the end (*f*) or in the beginning (*i*) of the competition and *t* is the number of generations per day. We assume *t*=18, in accordance with the 80 minute generation time estimated in previous studies on *E. coli* colonization of streptomycin-treated mouse (Poulsen et al., 1995; Rang et al., 1999; T. Conway et al., 2004).

Estimation of rate of beneficial mutations

To infer the effective evolutionary parameters (U_e and s_e) we used the method described in (Barrick et al., 2010), available online at http://barricklab.org/twiki/bin/view/Lab/ToolsMarkerDivergence. This procedure involves three main steps: simulate families of marker trajectories at many combinations of mutation rate (*U*) and selection coefficient (*s*), summarize the simulated and experimental marker trajectories in the statistics τ_e and α_e (that represent the time of divergence of marker frequency and the rate of divergence, respectively); compare the distributions of τ_e and α_e of the simulated dynamics and the experimental data and find the values of *U* and *s* that best explain the experimental marker dynamics.

Using the model described in (Barrick et al., 2010) we simulated sets of 100 replicate populations evolved under different parameter combinations of *U* and *s*. *U* ranged from 10^{-8} to 10^{-4} (with increments of 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5} and 10^{-4}) and *s* ranged from 0.07 to 0.1 (with increments of 0.005). We assumed a population size of 10^7, based on the weight of feces per day (1.5 ± 0.12 (2 s.e.m.) grams) that a mouse produces, on the observed number of CFUs (per gram of feces) along the experiments (which remained stable over the length of the experiment around 10^8,

see Supplementary Fig. 1), and on the number of generations per day, which is 18 (Poulsen et al., 1995). A population size of the same order of magnitude was measured by Leatham-Jensen *et al.* (Leatham-Jensen et al., 2012) in the intestinal mucus, where the majority of bacterial division occurs (T. Conway et al., 2004).

This simulated data consist of the logarithm of the ratio of the two marker frequencies (Rf = f_{YFP}(ti)/(f_{CFP}(ti)) at several time points (ti), (ln(Rf(ti))).

We then used both the simulated and experimental marker dynamics as input in the program marker_divergence_fit.pl, that summarizes the evolutionary dynamics in two statistics: τ_e and α_e. The first is the time, τ_e, where a significant deviation of ln(Rf(τ_e)) from ln(Rf(t = 0)) occurs. The second is the rate of change of ln(Rf(t)) with time, that is, 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(τ_e) and A(α_e). These distributions are then compared, using the program marker_divergence_significance.pl, to the distributions of τ_e and α_e that summarize the experimental data To(τ_e) and Ao(α_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.

Estimation of distribution of haplotypes fitness

To estimate the fitness effects of beneficial mutations that may contribute to adaptation we used a maximum likelihood approach to infer the number, fitness effect and time of establishment of beneficial mutations in an asexual population from individual marker frequencies trajectories (Illingworth and Mustonen, 2012). This method, which is available on web page: http://www.sanger.ac.uk/resources/software/optimist/, makes no *a priori* assumptions about the distribution of mutation effects. It has been tested against simulated data and has been shown to be able to: reproduce complex trajectories of neutral markers, make accurate estimates of the distribution of haplotype fitness effects, and provide good estimates of the number of mutations segregating at

high frequencies, at least for values of mutation rates that are not very large. We note that this method may miss many small effect beneficial mutations if these indeed occur at high rates.

It assumes an initial population of cells comprising two haplotypes with equal fitness, corresponding to two neutral markers. A beneficial mutation occurring in one of the backgrounds is represented as a new haplotype, which is assumed to exist at a given time (Tb_i) and with a frequency of 0.001 in the population. Since this new haplotype has increased fitness (W_i) relative to the initial wild-type population, its frequency will increase, leading to an increase in frequency of the fluorescent marker population in which the mutation has occurred. The simplest model that can be assumed is that where a single beneficial mutation occurs. That will imply a given time of establishment and effect of the mutation that maximize the probability of observing the marker frequency data. The second model that is considered is one where two mutations can occur, which will lead to a given likelihood of the data. Models assuming different numbers of beneficial mutations can thus be compared based on their likelihood score using Akaike information criteria (Illingworth and Mustonen, 2012). With this method we obtain a distribution of haplotype fitnesses with the minimal number of mutations that can explain the marker frequency dynamics (ω_h).

Whole genome re-sequencing and mutation prediction

Clone analysis: After 24 days of gut colonization one clone from populations 1.1 to 1.14 and the two ancestors (MG1655-YFP and MG1655-CFP) were isolated and used to seed 10 mL of LB (Line 1.15 was not analyzed since the mouse from this line died at that time point). These cultures were then grown at 37ºC with agitation. Subsequently DNA was isolated following a previously described protocol (Wilson, 2001). The DNA library construction and sequencing was carried out by BGI. Each sample was pair-end sequenced on an Illumina HiSeq 2000. Standard procedures produced data sets of Illumina paired-end 90bp read pairs with insert size (including read length) of ~470bp. Genome sequencing data have been deposited in the NCBI Read Archive, http://www.ncbi.nlm.nih.gov/sra (accession no. SRP017347). 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).

Ancestral genome: The sequence reads from MG1655 were mapped to the reference strain (Blattner et al., 1997). The two ancestors carried the mutations listed in Supplementary Table 1. The mutations underlined were present in the YFP ancestor but not in the CFP. The sequences of the 14 sequenced clones were then interrogated against this ancestral genome and the mutations identified are listed in Supplementary Table 2.

Population analysis: DNA isolation was obtained in the same way as described above for the clone analysis except that instead of one clone per population a mixture of >1000 clones per population was used. Three populations were sequenced: 1.10, 1.11 and 1.12.

The DNA library construction and sequencing was carried out by the IGC genomics facility. Each sample was pair-end sequenced on an Illumina MiSeq Benchtop Sequencer. Standard procedures produced data sets of Illumina paired-end 250bp read pairs. The mean coverage per sample was as ~260x for population 1.10, ~160x for population 1.11 and ~100x for population 1.12. Genome sequencing data have been deposited in the NCBI Read Archive http://www.ncbi.nlm.nih.gov/sra (accession no. SRP033025). Mutations were identified using the BRESEQ pipeline with the polymorphism option on (see Supplementary Table 7). The default settings were used except for: a) require a minimum coverage of 3 reads on each strand per polymorphism; b) eliminate polymorphism predictions occurring in homopolymers of length greater than 3; c) eliminate polymorphism predictions with significant (p-value<0.05) strand or base

quality score bias. All predicted polymorphisms were manually inspected using IGV.

Test for the advantage of the evolved clones in the presence of galactitol

All clones that were sequenced carried mutations in the *gat* operon. To test for any fitness advantage of these evolved clones, we measured their growth in M9 minimal medium (MM) supplemented with glycerol and with glycerol and galactitol, and compared growth in both conditions. All growth curves were conducted in 96-well plates incubated at 37ºC with aeration (Thermoshaker PHMP-4, Grant). After a first overnight growth in MM with glycerol (0.4%) the cultures were normalized to the same OD_{600} (approximately 0.1) and diluted 100-fold. We used 5 µl of the 10^{-2} dilution to inoculate in triplicate wells containing MM supplemented either with glycerol (0.4%) or glycerol (0.4%) and galactitol (0.4%). Optical density at 600 nm was monitored for 36 hours using a microplate reader (Victor3, PerkinElmer).

Emergence and dynamics of mutations in the *gat* operon

To investigate the dynamics of appearance and expansion of beneficial mutations in the *gat* operon we determined the frequency of bacteria unable to metabolize galactitol (*gat*-negative phenotype) within a given population of the evolution experiment.

For each population, a sample of the frozen fecal pellets was diluted in PBS and plated in Mac Conkey agar supplemented with 1% of galactitol and streptomycin (100 µg/ml); a differential medium used to monitor the ability of bacteria to use galactitol. Plates were incubated for 20 hours at 28ºC. The frequency of galactitol mutants for each time point was estimated by counting the number of white (galactitol mutants) and red colonies in Mac Conkey-galactitol plates.

Identification of adaptive mutations and estimate of haplotype frequencies in populations 1.1, 1.5, 1.11 and 1.12 during 24 days of adaptation to the mouse gut

In order to identify the adaptive mutations and estimate the haplotype frequencies, between 20 and 40 clones were analyzed per time point per population. These clones were obtained by plating directly the frozen fecal samples and thus correspond to samples representative of the evolving *E. coli* population. The presence of the adaptive mutations was queried by target PCR. The increase in size of the PCR product was indicative of the presence of an IS element. The detection of SNPs was done by target Sanger sequencing using the primers listed in Supplementary Table 8.

PCR reactions were performed in a total volume of 50 µl containing 1 µl bacterial culture, 10 µM of each primer (forward and reverse), 200 µM dNTPs, 1 U Taq polymerase and 1X Taq polymerase buffer. The PCR reaction conditions were as follows: 95ºC for 3 min followed by 34 cycles of 95 ºC for 30 s, 60 ºC for 30 s, 72 ºC for 2 min, followed by 72ºC for 5min. *srlR*, *gatC* and *gatZ* were sequenced directly from the PCR products using the primers listed in Supplementary Table 8.

To detect an insertion of 1 bp in *gatC*, the gene was PCR amplified and then submitted to an enzymatic restriction with MvaI enzyme. The mutant was identified based on the fragment restriction profile.

The 150 Kb duplication was detected by amplification of the new junction formed by the tandem duplication. The PCR was performed using the same conditions described before and the primers listed in Supplementary Table 8.

Direct estimation of the advantage of the gat alleles by competition

To determine the advantage of the gat alleles we performed *in vivo* fitness assays. The competition experiments lasted for three days and were performed according to the protocol described above for the evolution experiment.

In the first set of competitions (Supplementary Fig. 4) we competed a clone genetically similar to the ancestral but bearing an IS insertion in the *gatZ* gene (sequenced clone 4) against the ancestral. In the second set of competitions

(Supplementary Fig. 6) we competed the *gatZ* mutant against a *gatC* mutant (derived from sequenced clone 12). In addition to a 1bp insertion in the coding region of the *gatC* gene, this clone originally carried an unstable 150 kb duplication, lost during *in vitro* manipulations to create the mutant *gatC*.

Testing for negative frequency-dependent selection

To test for this form of selection we isolated around 30 YFP and 30 CFP clones from fecal samples after 24 days of colonization of population 1.13. We then performed competitive fitness assays *in vivo*. The mixture of the 30 YFP clones was competed against the mixture of the 30 CFP clones at three initial ratios (1:1, 10:1 and 1:100) for 2 days, using the procedure described for the evolution experiments.

Test for increased mutation rate

To test for the possible emergence of mutators during adaptation to the gut we determined the frequency of rifampicin-resistant mutants, in each of the evolved populations. We grew pre-cultures of the evolved populations by inoculating in 10 ml of LB a sample of each population from the last day of the experiments (approximately 400 generations of gut colonization). Pre-cultures were grown overnight at 37ºC with aeration in the presence of ampicillin (100 µg/ml) and streptomycin (100 µg/ml). The pre-cultures were then diluted and approximately 1000 cells (10 µl of a 10^{-4} dilution) were inoculated in triplicate in 1 ml of LB and incubated overnight. Aliquots of each tube were plated in LB agar and LB agar supplemented with rifampicin (100 µg/ml) and incubated overnight at 37ºC. The frequency of mutation to rifampicin resistance was calculated as the ratio between the number of rifampicin resistant mutants and the total number of individuals in each population.

Results and Discussion

Dynamics of adaptation of *E. coli* through changes in the frequency of a neutral marker

A genetically homogeneous population of *E. coli*, except for a chromosomally encoded fluorescence marker (see Material and Methods), was used to trace the occurrence of different adaptive mutations (Hegreness et al., 2006) and determine the strength of their fitness effects. This population, composed of equal amounts of two subpopulations expressing either a yellow (YFP) or cyan (CFP) fluorescent protein, was used to inoculate inbred mice orally. Subsequently, we followed fluorescent-markers frequency from daily collected fecal samples, for 24 days. The experiment was repeated three times to a total of 15 mice housed individually, where *E. coli* adaptation was followed.

The adaptive dynamics observed in 15 independently evolving populations are shown in Fig. 1a (see also Supplementary Fig. 1 for the *E. coli* loads over time). Some of the dynamics observed, exhibit a typical signature of CI. An initial increase in frequency of one neutral marker, followed by replacement by the subpopulation bearing another (e.g. Fig. 1a, population 1.9) was diagnostic of CI. Presumably, the markers hitchhike with successions of beneficial mutations (Hegreness et al., 2006). We also observed cases in which the frequencies of the markers remained stable, (Fig. 1a, populations 1.6 and 1.8), for up to 24 days. This time corresponds to approximately 400 generations, if we take the 80 minute generation time estimated in the gut (Poulsen et al., 1995). Stable marker dynamics are expected under two completely opposing scenarios: neutral evolution due to the lack of occurrence of adaptive mutations, or very intense CI caused by strong mutations of similar effect occurring in the different fluorescence backgrounds simultaneously. To distinguish between these scenarios and to further show that strong adaptive mutations are occurring, we performed *in vivo* competitive fitness assays of several clones with the ancestral strain. We tested clones from populations where a clear signal of adaptation was detected (populations 1.12 and 1.13 in Fig. 1a) or was not (1.6 and 1.8 in Fig. 1a).

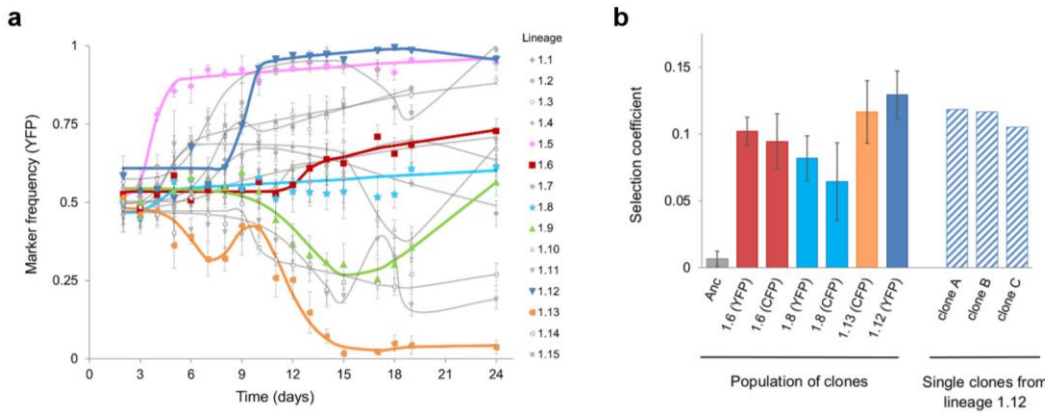


Figure 1. Evidence for rapid adaptation and CI *in vivo*.
a) Dynamics of marker frequency (with 95% confidence intervals) during the adaptation of *E. coli* to the gut upon colonization (populations 1.1 to 1.15). The predictions of the simplest model of Darwinian selection (Illingworth and Mustonen, 2012), for each set of data points are shown as lines. The lines correspond to the model that assumes multiple beneficial mutations (i=1,2,..,5) can occur in a given clone at a given time (Tb_i), and these clones have a given fitness (W_i). Tb_i and W_i are fitted by maximum likelihood and the best model, in terms of number of mutations, is chosen according to Akaike criteria. Representative examples of trajectories for the classical signature of a selective sweep (populations 1.5, 1.12 and 1.13) and for the maintenance of neutral diversity under with intense CI (populations 1.6, 1.8 and 1.9) are shown in colours. **b)** Direct evidence for adaptation and CI: the bars represent the mean selection coefficient (± 2 s.e.m, n=3) from an *in vivo* competitive fitness assay between evolved and ancestral clones. The first bar shows the neutrality of the fluorescent marker. The following six bars represent the results from competitions of a mixture of thirty clones (with a given fluorescent marker, indicated below the bar) isolated from the respective population (indicated below each bar). Both CFP and YFP mixtures of clones from lineages 1.6 and 1.8 show similar levels of adaptation, consistent with intense CI maintaining a high frequency of both neutral markers. The dashed bars show the results of competitions (n=2) of single clones isolated from lineage 1.12.

All clones tested were found to carry beneficial mutations (Fig. 1b). As expected, the clones isolated after a clear sweep (Fig. 1b, populations 1.12 and 1.13), showed a fitness increase. Importantly, clones sampled from the populations in which only slight changes in marker frequency were detected also showed a fitness increase (Fig. 1b, populations 1.6 and 1.8). This demonstrates that the evolution process involved multiple adaptive mutations competing for fixation. Consistent with the small changes in marker frequency observed, the competitive fitnesses were similar for clones isolated from the same evolving population, but bearing distinct neutral markers (Mann-Whitney test P=0.7 population 1.6, P=0.4 population 1.8).

The distribution of effects of mutations that contributed to adaptation

From the dynamics of neutral markers one can try to estimate the important evolutionary parameters of the adaptive process: mutation rate and selective effects (Barrick et al., 2010; Hegreness et al., 2006; Illingworth and Mustonen, 2012; Sousa et al., 2013). One method that has been used (Barrick et al., 2010) summarizes neutral marker dynamics in replicate experiments in two statistics**:** i) the time it takes for one of the markers to first change significantly in frequency and ii) the slope associated with that frequency change. The time for the first marker deviation corresponds to the time where the natural logarithm of the marker ratio significantly changes from the initial ratio. The slope associated with a frequency change is the slope of the linear regression of the natural logarithm of the ratio between the two markers after a significant frequency change is detected. These summary statistics lead to estimation of two effective evolutionary parameters (U_e and s_e) (Hegreness et al., 2006). Under intense CI U_e (the effective genomic mutation rate) will be an underestimate of U_b, and s_e (the effective selection coefficient of beneficial mutations) an overestimate of s_b (Sousa et al., 2013). For the dynamics in Fig. 1, the summary statistics imply U_e ~$7x10^{-7}$ and s_e ~0.075.

We also have used a recently developed method (Illingworth and Mustonen, 2012) to estimate the fitness of the haplotypes that segregate at sufficiently high frequency to lead to changes in marker dynamics. This method takes the frequencies of the neutral markers across all the time points of the experiment and allows for the occurrence of CI. We started by estimating the distribution of fitness effects of the minimum number of haplotypes, as proposed by Illingworth and Mustonen (Illingworth and Mustonen, 2012), to fit the marker frequency data. Assuming the simplest possible model of Darwinian selection, with constant selection across space and time (admittedly an oversimplified view of the gut), we fitted a series of models, which differ in the number of beneficial mutations. Starting from the simplest minimal model where only one beneficial mutation occurs, we sequentially increase the number of mutations until a maximum of five. For each model this approach fits, by maximum likelihood, the observed frequency changes at a marker locus with two alleles (our fluorescent alleles) and then

chooses the simplest possible model. The predictions described faithfully the empirical data in terms of marker frequencies (lines in Fig. 1a). With this method the distribution of haplotype fitnesses across all populations, can be determined (Fig. 2). The inferred fitness effects of segregating haplotypes are quite large with a mean of (15%) and some haplotypes leading to increases in fitness of more than 30%. We note that given the intense CI observed, this approach misses some of the mutations contributing to the dynamics (see below).

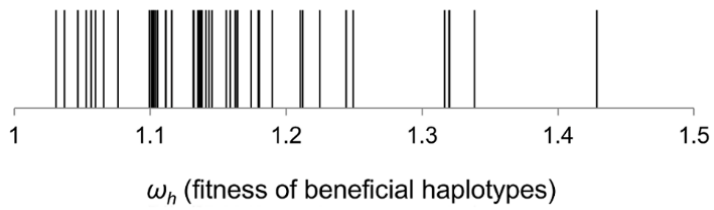


Figure 2. Distribution of fitness effects of beneficial haplotypes that contributed to adaptation. The fitness of beneficial haplotypes (ω_h) was estimated under a theoretical model which assumes the minimum number of beneficial mutations required to explain the marker dynamics.

Genetic basis of the initial adaptations to the mouse gut

To characterize genetically the first steps of the adaptive process *in vivo*, we performed whole genome sequencing (WGS) of the ancestral and independently evolved clones. Each evolved clone was isolated from a different mouse after colonization for 24 days (~400 generations). The mean number of mutations per evolved clone was 2.3, with a minimum of 1 and a maximum of 4 (Fig. 3 and Supplementary Table 2, see also Supplementary Table 1).

The genetic analysis showed that similar and parallel adaptive paths were taken during the initial adaptation to the gut (Fig. 3, Supplementary Table 2). The most striking recurrent event was detected at the level of mutations in the *gat* operon, which occurred in all the sequenced clones (79% IS insertions and 21% small indels). The *gat* operon consists of six genes that collectively allow for galactitol metabolism (Nobelmann and Lengeler, 1996). We found that galactitol had an inhibitory effect on the ancestral strain when grown in minimal medium with glycerol (Supplementary Fig. 2). However in all the evolved clones (carrying mutations either in the coding region of *gatA*, *gatC*, *gatZ* or in the regulatory region

of *gatY* (see Fig. 3)) this inhibition was surpassed (Supplementary Fig. 2). All these clones exhibited a *gat*-negative phenotype, inability to metabolize galactitol. Since galactitol is part of the host's metabolism of galactose, *E. coli* might be frequently exposed to this compound in the gut. This may have selected for the *gat*-negative phenotype.

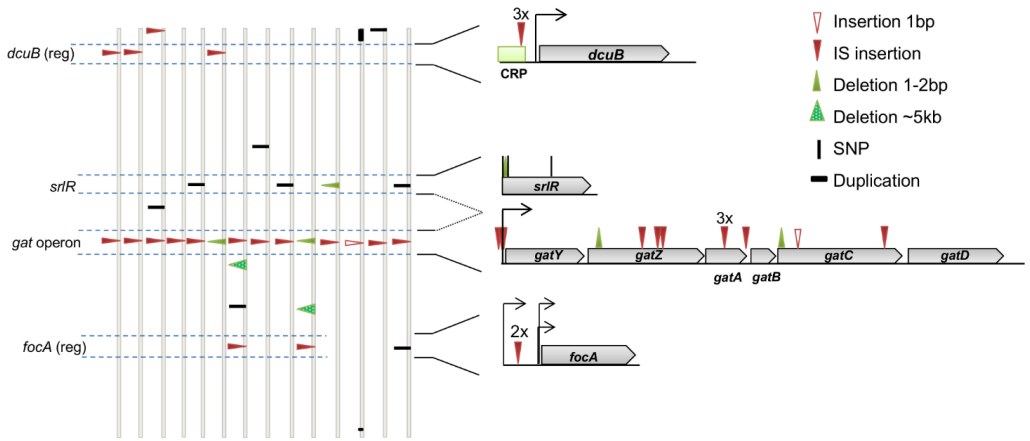


Figure 3. The genetic basis of adaptive mutations and the level of parallelism between populations.
Identified mutations in clones isolated from populations 1.1 to 1.14 (evolved *in vivo* for 24 days), represented along the *E. coli* chromosome. For simplicity, the genomes are represented linearly and vertically drawn. The type and position of mutations are shown by triangles for insertions and deletions, small vertical bars denote single nucleotide polymorphisms (SNPs), and one duplication in clone number 1.12 is depicted as a horizontal bar. See the symbol legend for other events. The genes *dcuB*, *srlR* and *focA* and one operon (*gat*) are highlighted. These represent regions of parallel mutation in at least two genomes. The genomic context of these mutations is represented on the right. (reg) after the gene name, means that the regulatory region, rather than the coding region, was affected. Numbers above marked mutations represent the number of times a particular mutation was detected at the same position.

Parallelism at the gene level was also observed: four clones carried mutations in *srlR*, which is a DNA-binding transcription factor that represses the *srl* operon involved in sorbitol metabolism (Yamada and Saier, 1987). At least two of the mutations inactivated the gene, since these caused a frameshift and a stop codon (Supplementary Table 2). The inactivation of *srlR* leads to the constitutive expression of the *srl* operon (Yamada and Saier, 1987), thus these mutants are expected to readily consume sorbitol, even at low concentrations. This might represent an advantage when facing limiting nutrients and strong competition, compatible with the hypothesis that the ability to use several limiting carbon

sources is an important advantage for *E. coli* inhabiting the mammalian gut (Chang et al., 2004; Gauger et al., 2007).

Two other parallelisms in the regulatory regions of *dcuB* (three clones) and *focA* (two clones) were observed. Both of these genes encode transmembrane transporters whose expression occurs almost exclusively in anaerobic conditions. The Dcu system mediates the uptake, exchange and efflux of C_4-dicarboxylic acids (Engel et al., 1994); fumarate is a C4-dicarboxylate acid and is the most important anaerobic electron acceptor for *E. coli* in the intestine (Jones et al., 2011). Additionally, *focA* is a putative transporter of formate (Suppmann and Sawers, 1994) and is the "signature compound" of *E. coli* anaerobic metabolism (Leonhartsberger et al., 2002). During anaerobic growth, pyruvate is cleaved into formate that is subsequently used as a major electron donor for the anaerobic reduction of fumarate (Macy et al., 1976) and nitrate or nitrite (Wimpenny and Cole, 1967). Hence, given their roles in anaerobic respiration it would be expected that *dcuB* and *focA* are under selective pressure in the gut and that the mutations observed may be important for regulating anaerobic respiration of *E. coli* in the gut.

Interestingly, one clone carried a non-synonymous mutation in *arcA*, a global regulator that governs respiratory flexibility of *E. coli*. *arcA* allows controlled switching between aerobic and anaerobic respiratory genes during fluctuating oxygen conditions, a trait crucial for *E. coli* to efficiently compete with the gut flora composed essentially of anaerobes (Jones et al., 2007, 2011).

Previous studies have found increased expression of *dcu*, *gat* and *srl* regulons during growth of *E. coli* on mucus (Fabich et al., 2011). Together with our findings of mutations on these genes, suggests that these regulons are under strong selection.

One common adaptation in *E. coli* reported to occur in the mouse gut is decreased motility (Gauger et al., 2007; Giraud et al., 2008; Leatham et al., 2005), which typically occurs in strains that are hypermotile as a result of an IS element upstream of the *flhDC* operon. Since our strain is not hypermotile, it is not surprising that mutations in this region were not observed in our study.

Taken together, our genetic analyses revealed that the initial adaptive steps of *E. coli* to the mouse gut involved gene inactivation or modulation through IS elements. In terms of the relative contribution of regulatory versus coding regions to adaptation (Carroll, 2008; Hoekstra and Coyne, 2007), we found that half of the IS insertions occurred in regulatory regions and one third of all mutations were located in these regions. We therefore observe that both changes in regulatory and coding regions contributed to *E. coli* adaptation to the gut. We also note that the first step of adaptation involved loss of function mutations (namely in the *gat* operon), which supports the recent view that null mutations may play an important role in early adaptation to new environments (Hottes et al., 2013).

Previous studies (Maharjan et al., 2006, 2012) focusing on evolution of *E. coli* populations in glucose-limited chemostats, a device used to simulate the human gut (Mason and Richardson, 1982), during a similar period of time (26 days), found a high degree of parallelism just as we observe here. Furthermore extensive phenotypic diversity was observed (Maharjan et al., 2012) and whole genome sequencing identified similar numbers of substitutions and a contribution of IS elements to the adaptive process (Gaffé et al., 2011). However the targets of selection in glucose-limited chemostats were very different from those found during its adaptation to the gut, reflecting the distinct specific selective pressures *E. coli* is subjected to in the different environments.

Phenotypic hard sweeps and genetic soft sweeps at the *gat* operon

Given the striking parallelism that occurred in the *gat* operon, through knockout mutations, we sought to determine the timing of its emergence in each of the populations. We followed the dynamics of the *gat*-negative phenotype during adaptation (Fig. 4). In eight populations the advantageous mutation could be detected just 2 days post-colonization. In one population an extraordinarily rapid phenotypic sweep could be seen: a population where initially the majority of clones were *gat*-positive changed completely its phenotype within 2 days (Fig. 4, population 1.5), showing the strong benefit associated with this phenotype.

In most populations the increase in frequency of the *gat*-negative phenotype was followed by a change in marker frequency (Supplementary Fig. 3a),

consistent with the expected dynamics of a strong beneficial mutation occurring in a given fluorescent background sweeping to fixation. However in some populations fixation of the *gat*-negative phenotype was observed without fixation of the neutral marker (Supplementary Fig. 3b). This suggests that independent mutations causing the same phenotype have occurred, *i.e.* clonal competition emerging from multiple alleles at the same locus with similar selective effects is driving these dynamics.

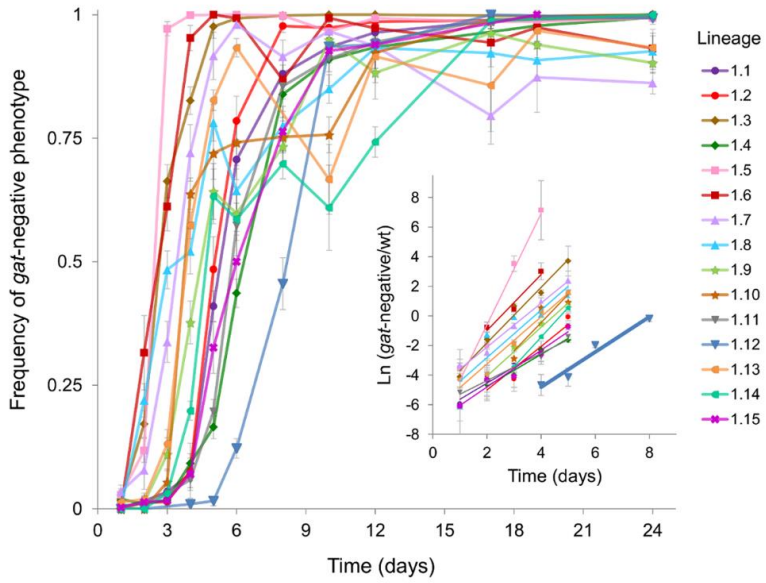


Figure 4. Emergence and spread of beneficial mutations in the *gat* operon. Dynamics of frequency change of the *gat*-negative phenotype over time are shown for all populations (1.1 to 1.15). Inset: The natural logarithm of the ratio of *gat*-negative individuals to wild type over the first 5 days of adaptation is shown as dots. Each group of points was fitted to a linear regression (represented as lines). Highlighted in bold is the population 1.12 for which the slope corresponds to an estimate of the selection coefficient of 0.075±0.01 (per generation).

We next sought to estimate the strength of selection associated with new beneficial alleles that emerge and escape stochastic loss from the initial change in frequency of the *gat*-negative phenotype. If we take population 1.12, where only one allele was detected in the *gat* operon, we can estimate the selective advantage conferred by this mutation (s_{gat}) from the slope of the Ln(x/(1-x)) along time, where x is the frequency of the *gat*-negative phenotype. When doing so we find that s_{gat} = 0.07 (±0.01) (Fig. 4 inset). In the other populations the strength of

selection is more difficult to estimate due to the emergence of many distinct alleles, which leads to an underestimate of their beneficial effect due to CI.

To determine directly the selective advantage of *gat* alleles we performed *in vivo* competitions against the ancestor. As shown in Supplementary Fig. 4, the mutant *gatZ* (clone 4YFP) had, on average, an advantage of 0.065 (±0.013) over the ancestral, which is compatible with that inferred from the dynamics of the *gat*-negative phenotype.

To gain further insight into the regime of interference detected by the change in frequency of the neutral markers, we studied polymorphism levels of four of the adapting populations. We followed the change in frequency of haplotypes at different points in time in two populations that showed a rapid change in marker frequency (populations 1.5 and 1.12) and in two other populations where the neutral marker was kept polymorphic throughout the period studied (1.1 and 1.11) (Tables S3 to S6). Each haplotype was the combined genotype at 7 selected loci: *gatY, gatZ, gatA, gatC, srlR, dcuB* and *focA*. These loci were previously observed to be mutational targets in two or more of the sequenced independently evolved clones (see Fig. 3), which we interpret as important players in adaptation to the gut, and thus likely exhibiting segregating polymorphisms in most populations. We note that we determined the haplotype structure of a sample of clones (between 20 and 40 clones per time point, per population) based on the mutations found in the sequenced clones, that is, we looked for IS insertions in *gatY, gatZ, gatA, gatC, dcuB* and *focA*, and SNPs or small indels in *gatZ, gatC and srlR* (see Methods). The presence of an extra mutation, a duplication of ~150Kb, was also tested in population 1.12.

Fig. 5 shows the dynamics of these haplotypes in the populations studied (see also Supplementary Fig. 5 for the sums of the frequencies of genotypes with possible equivalent phenotypes over time). Extensive haplotype diversity is found in all populations but to a lesser extent in populations 1.12 and 1.5. For example, in the first ~100 generations we observe 16 and 8 haplotypes segregating in populations 1.1 and 1.11, but only 3 and 5 in populations 1.12 and 1.5, respectively. This observation is concordant with the neutral marker dynamics, where populations 1.12 and 1.5 showed a rapid sweep of one of the markers while

the other two (1.1 and 1.11) maintained the ratio of these markers closer to 50% for a longer period.

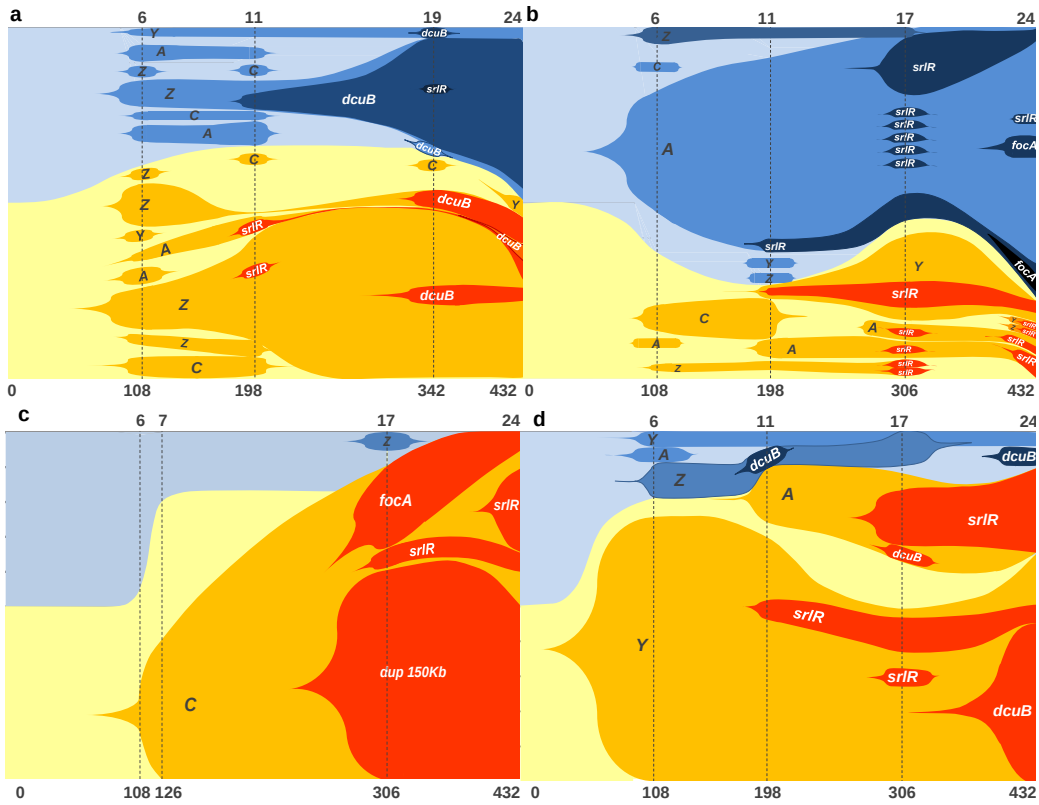


Figure 5. Graphic representation of the frequencies of newly generated haplotypes along 24 days (corresponding to 432 generations) of evolution inside the mouse gut (see Supplementary Tables 3 to 6 for numeric data). Shaded areas are proportional to the relative abundance of each haplotype. Yellow and blue shaded areas represent the two sub-populations of bacteria labeled either with *cfp* or *yfp* alleles. The ancestry relations between haplotypes can be inferred by the accumulation of new mutations in a previously existent genotype. Dash lines mark the time points in days (upper axis) or generations (lower axis) where the sampling took place. For the top two populations 1.1 (**a**) and 1.11 (**b**), 40 clones were sampled in each time point. For the bottom two populations 1.12 (**c**) and 1.5 (**c**) 20 clones were sampled in each time point.

The first mutation detected, without exception, impairs galactitol metabolism, thus surpassing the vital inhibitory effect of this compound on *E coli*'s growth. We observed multiple soft sweeps of mutations, with possibly similar fitness benefit, occurring independently in the *gat* operon and competing for fixation. We directly tested for this by performing competitions between clones carrying different *gat* alleles: mutations in *gatZ* against *gatC*. The results of these competitions show an equivalent selective coefficient of at least these two mutants (Supplementary Fig.

6), supporting the hypothesis that the different mutations in the *gat* operon present in the populations confer the same phenotype and fitness advantage. This provides an explanation for the observed coexistence between different alleles at the same locus and the phenotypic replacement detected.

After the mutations in the *gat* operon secondary adaptive mutations occurred and promoted the increase in the frequency of haplotypes carrying more than one beneficial mutation. Overall, the adaptive process appears non-mutation limiting, allowing polymorphism levels within populations to be kept high until the end of the period studied. A closer look to the haplotype dynamics points to the strong possibility of further mutations occurring beyond the ones we have targeted. For example, at generation 126 in population 1.12, the increase in frequency of the neutral allele *yfp* seems to be caused by some other mutation besides *gatC.* Furthermore the occurrence of a large duplication in the same population appears to cause a beneficial effect. To gain access to further beneficial mutations, we performed whole genome sequencing of this (1.12) and two other populations (1.11 and 1.10) at the end-point of the experiment (see Supplementary Table 7). This procedure aimed at querying the possibility of beneficial mutations at other loci significantly contributing to the adaptive process. In all populations we observed multiple mutations in some of the targets previously identified (*gat* operon, *srlR* and *focA*) as well as several other loci that were targets for mutation unique to a single population. The adaptive value of these last events is difficult to ascertain since these might have increased in frequency to detectable values due to linkage with other beneficial mutations or just by chance. This analysis allowed uncovering two new parallel mutational targets. One of these targets was the regulatory region of *yjjP* where an IS insertion occurred in population 1.11 and also on clone 3YFP (from population 1.3) and the other target was *asnA*, where either a deletion or a duplication of around 50bp was observed in two different populations (1.11 and 1.12). *yjjP* codes for an inner membrane structural protein of unknown function (Daley et al., 2005) and *asnA* codes for one of the two asparagine synthetases in *E. coli*, catalyzing the ammonia-dependent conversion of aspartate to asparagine (Cedar and Schwartz, 1969).

Potential role of epistasis in the adaptive process

Overall, the haplotype analysis found ~6% of triple mutants. These were all *gat*-negative and carried either a *srlR* and *dcuB* mutation or a *srlR* and *focA* mutation. *dcuB* and *focA* were never found in combination, suggesting that they may interact epistatically. This can be possibly due to their common involvement in anaerobic respiration modulation, thus fulfilling related functions. This last observation suggests a role for negative epistatic interactions whereby mutations are beneficial individually but deleterious when in combination. The fact that repeated selection of *dcuB* or *focA* mutations was observed, but these mutations were never found in the same genetic background, may indicate a disadvantage of the double mutant carrying mutations that individually are presumably beneficial. Another example of possible epistasis is the case of the mutations in the *gat* operon. These occur very rapidly and lead to the inactivation of the *gat* operon, conferring an estimated advantage of approximately 7%. It is therefore expected that additional mutations subsequently occurring in other *gat* genes would not confer a strong or any advantage, since the first mutation already resulted in a loss of function.

Test for negative frequency-dependent selection

Negative frequency dependent selection, *i.e.*, fitness advantage of clones when at low frequency and fitness disadvantage when at high frequency, is known to lead to maintenance of genetic diversity (Herron and Doebeli, 2013; Maharjan et al., 2012; Rainey, PB et al., 2000). Since the gut is a complex environment (T. Conway et al., 2004) where non-transitive ecological interactions may occur, we tested for variation in fitness effects with initial frequency of evolved clones. Negative frequency-dependent interactions could explain the long-term maintenance at low frequency of some of the fluorescent subpopulations observed in the adaptive dynamics (Fig. 1a, populations 1.5, 1.12 and 1.13). We tested population 1.13 for this type of selection, where CFP marker increased till 0.95 but never achieved fixation. In Supplementary Fig. 7 we show the results of the tested adapted clones in competitive fitness assays at different initial frequencies. While we observed that clones have an advantage when rare we also observe that such

selective advantage is present when at high frequency. No disadvantage was found at any of the initial frequencies tested, which suggests that negative frequency dependent selection is not a major force operating in this population.

No evidence for emergence of mutators

Our finding that mutations of large beneficial effects can accumulate rapidly *in vivo* suggested to us that mutators could have emerged. Mutators may play an important role in bacterial evolvability and their emergence typically involves mutations in genes of the DNA repair system (Miller, 1996). Depending on the gene mutated a mutator bacteria can acquire an increase in its genome-wide mutation rate of 10 up to 10 000 fold (Giraud et al., 2001a). Indeed co-colonization of wild-type and mutator bacteria in germ free mice (Giraud et al., 2001b) has shown that mutators can increase in frequency by hitchhiking with beneficial mutations to which they are genetically linked to. Furthermore mutators have been observed to sometimes emerge *de novo* in the early stages of adaptation to glucose-limited environments in laboratory experiments (Maharjan et al., 2013; Sniegowski et al., 1997). We therefore tested for an increased rate of mutation through fluctuation assays of several adapted clones. These tests however provided no evidence for significant increases in genomic mutation rate (Supplementary Fig. 8).

Conclusions

In conclusion, we demonstrate here that *E. coli* MG1655 adapts very rapidly to the intestine of streptomycin-treated mice. The first steps of adaptation of *E. coli* to the mouse gut are dominated by soft sweeps and adaptive mutations of large effect. We observed a regime of intense clonal interference where haplotypes carrying more than one beneficial mutation compete for fixation. This is the regime assumed under the Fisher-Muller theory that predicts that recombination speeds up adaptation by reducing competition between beneficial mutations. This data therefore provides support for the Fisher–Muller effect as an important mechanism for the maintenance of homologous gene recombination in bacteria (Wylie et al.,

2010). It shows that *E. coli* can adapt to this complex ecosystem very fast: within two days a phenotypic replacement could be detected. The first beneficial mutations targeted the same gene or operon and transposable elements made a substantial contribution to adaptation. These results demonstrate the remarkable adaptive potential of bacteria in one of the most complex environments of their natural ecology and may have important consequences for our understanding of the species and strain diversity in the gut microbiota.

Acknowledgements

We thank J. Xavier, L. Wahl, A. Athanasiadis, A. Coutinho, J. Thompson and Gordo's lab for critically reading the manuscript.

References

Atwood, K.C., Schneider, L.K., and Ryan, F.J. (1951). Periodic Selection in Escherichia Coli. Proc. Natl. Acad. Sci. U. S. A. *37*, 146–155.

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., and Kim, J.F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature *461*, 1243–1247.

Barrick, J.E., Kauth, M.R., Strelioff, C.C., and Lenski, R.E. (2010). Escherichia coli rpoB mutants have increased evolvability in proportion to their fitness defects. Mol. Biol. Evol. *27*, 1338–1347.

Barton, N., and Partridge, L. (2000). Limits to natural selection. Bioessays *22*, 1075–1084.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. (1997). The complete genome sequence of Escherichia coli K-12. Science *277*, 1453–1462.

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell *134*, 25–36.

Cedar, H., and Schwartz, J.H. (1969). The asparagine synthetase of Escherhic coli. I. Biosynthetic role of the enzyme, purification, and characterization of the reaction products. J. Biol. Chem. *244*, 4112–4121.

Chang, D.E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., Anderson, A.B., Grissom, J.E., Laux, D.C., Cohen, P.S., et al. (2004). Carbon nutrition of Escherichia coli in the mouse intestine. Proc Natl Acad Sci USA *101*, 7427–7432.

Daley, D.O., Rapp, M., Granseth, E., Melén, K., Drew, D., and von Heijne, G. (2005). Global topology analysis of the Escherichia coli inner membrane proteome. Science *308*, 1321–1323.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci USA *97*, 6640–6645.

De Paepe, M., Gaboriau-Routhiau, V., Rainteau, D., Rakotobe, S., Taddei, F., and Cerf-Bensussan, N. (2011). Trade-off between bile resistance and nutritional competence drives Escherichia coli diversification in the mouse gut. PLoS Genet *7*, e1002107.

Desai, M.M., and Fisher, D.S. (2007). Beneficial mutation selection balance and the effect of linkage on positive selection. Genetics *176*, 1759–1798.

Desai, M.M., Fisher, D.S., and Murray, A.W. (2007). The speed of evolution and maintenance of variation in asexual populations. Curr Biol *17*, 385–394.

Engel, P., Krämer, R., and Unden, G. (1994). Transport of C4-dicarboxylates by anaerobically grown Escherichia coli. Energetics and mechanism of exchange, uptake and efflux. Eur. J. Biochem. *222*, 605–614.

Fabich, A.J., Leatham, M.P., Grissom, J.E., Wiley, G., Lai, H., Najar, F., Roe, B. a, Cohen, P.S., and Conway, T. (2011). Genotype and phenotypes of an intestine-adapted Escherichia coli K-12 mutant selected by animal passage for superior colonization. Infect Immun *79*, 2430–2439.

Gaffé, J., McKenzie, C., Maharjan, R.P., Coursange, E., Ferenci, T., and Schneider, D. (2011). Insertion sequence-driven evolution of Escherichia coli in chemostats. J. Mol. Evol. *72*, 398–412.

Gauger, E.J., Leatham, M.P., Mercado-Lubo, R., Laux, D.C., Conway, T., and Cohen, P.S. (2007). Role of motility and the flhDC Operon in Escherichia coli MG1655 colonization of the mouse intestine. Infect Immun *75*, 3315–3324.

Giraud, A., Radman, M., Matic, I., and Taddei, F. (2001a). The rise and fall of mutator bacteria. Curr. Opin. Microbiol. *4*, 582–585.

Giraud, A., Matic, I., Tenaillon, O., Clara, A., Radman, M., Fons, M., and Taddei, F. (2001b). Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. Science *291*, 2606–2608.

Giraud, A., Arous, S., De Paepe, M., Gaboriau-Routhiau, V., Bambou, J.-C., Rakotobe, S., Lindner, A.B., Taddei, F., and Cerf-Bensussan, N. (2008). Dissecting the genetic components of adaptation of Escherichia coli to the mouse gut. PLoS Genet *4*, e2.

Good, B.H., Rouzine, I.M., Balick, D.J., Hallatschek, O., and Desai, M.M. (2012). Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. Proc. Natl. Acad. Sci. *109*, 4950–4955.

Gordo, I., Perfeito, L., and Sousa, A. (2011). Fitness effects of mutations in bacteria. J Mol Microbiol Biotechnol *21*, 20–35.

Hegreness, M., Shoresh, N., Hartl, D., and Kishony, R. (2006). An equivalence principle for the incorporation of favorable mutations in asexual populations. Science *311*, 1615–1617.

Herron, M.D., and Doebeli, M. (2013). Parallel Evolutionary Dynamics of Adaptive Diversification in Escherichia coli. PLoS Biol *11*, e1001490.

Hill, W.G., and Robertson, A. (1966). The effect of linkage on limits to artificial selection. Genet. Res. *8*, 269–294.

Hoekstra, H.E., and Coyne, J.A. (2007). The locus of evolution: evo devo and the genetics of adaptation. Evol. Int. J. Org. Evol. *61*, 995–1016.

Hottes, A.K., Freddolino, P.L., Khare, A., Donnell, Z.N., Liu, J.C., and Tavazoie, S. (2013). Bacterial Adaptation through Loss of Function. PLoS Genet. *9*.

Illingworth, C.J.R., and Mustonen, V. (2012). A method to infer positive selection from marker dynamics in an asexual population. Bioinformatics *28*, 831–837.

Jones, S.A., Chowdhury, F.Z., Fabich, A.J., Anderson, A., Schreiner, D.M., House, A.L., Autieri, S.M., Leatham, M.P., Lins, J.J., Jorgensen, M., et al. (2007). Respiration of Escherichia coli in the mouse intestine. Infect Immun *75*, 4891–4899.

Jones, S.A., Gibson, T., Maltby, R.C., Chowdhury, F.Z., Stewart, V., Cohen, P.S., and Conway, T. (2011). Anaerobic respiration of Escherichia coli in the mouse intestine. Infect. Immun. *79*, 4218–4226.

Kao, K.C., and Sherlock, G. (2008). Molecular characterization of clonal interference during adaptive evolution in asexual populations of Saccharomyces cerevisiae. Nat. Genet *40*, 1499–1504.

Lassig, M. (2012). Chance and risk in adaptive evolution. Proc Natl Acad Sci USA *109*, 4719–4720.

Leatham, M.P., Stevenson, S.J., Gauger, E.J., Krogfelt, K.A., Lins, J.J., Haddock, T.L., Autieri, S.M., Conway, T., and Cohen, P.S. (2005). Mouse Intestine Selects Nonmotile flhDC Mutants of Escherichia coli MG1655 with Increased Colonizing Ability and Better Utilization of Carbon Sources. Infect Immun *73*, 8039–8049.

Leatham-Jensen, M.P., Frimodt-Møller, J., Adediran, J., Mokszycki, M.E., Banner, M.E., Caughron, J.E., Krogfelt, K.A., Conway, T., and Cohen, P.S. (2012). The Streptomycin-Treated Mouse Intestine Selects Escherichia coli envZ Missense Mutants That Interact with Dense and Diverse Intestinal Microbiota. Infect Immun *80*, 1716–1727.

Lederberg, J. (2004). E. coli K-12. Microbiol Today *31*, 116.

Lee, M.-C., and Marx, C.J. (2013). Synchronous waves of failed soft sweeps in the laboratory: remarkably rampant clonal interference of alleles at a single locus. Genetics *193*, 943–952.

Leonhartsberger, S., Korsa, I., and Böck, A. (2002). The molecular biology of formate metabolism in enterobacteria. J. Mol. Microbiol. Biotechnol. *4*, 269–276.

Macy, J., Kulla, H., and Gottschalk, G. (1976). H2-dependent anaerobic growth of Escherichia coli on L-malate: succinate formation. J. Bacteriol. *125*, 423–428.

Maharjan, R., Seeto, S., Notley-McRobb, L., and Ferenci, T. (2006). Clonal adaptive radiation in a constant environment. Science *313*, 514–517.

Maharjan, R.P., Ferenci, T., Reeves, P.R., Li, Y., Liu, B., and Wang, L. (2012). The multiplicity of divergence mechanisms in a single evolving population. Genome Biol. *13*, R41.

Maharjan, R.P., Liu, B., Li, Y., Reeves, P.R., Wang, L., and Ferenci, T. (2013). Mutation accumulation and fitness in mutator subpopulations of Escherichia coli. Biol. Lett. *9*, 20120961.

Mason, T. g., and Richardson, G. (1982). Observations on the in vivo and in vitro competition between strains of Escherichia coli isolated from the human gut. J. Appl. Microbiol. *53*, 19–27.

Messer, P.W., and Petrov, D.A. (2013). Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol. Evol. *28*, 659–669.

Miller, J.H. (1996). Spontaneous mutators in bacteria: insights into pathways of mutagenesis and repair. Annu. Rev. Microbiol. *50*, 625–643.

Ning, Z., Cox, A.J., and Mullikin, J.C. (2001). SSAHA: a fast search method for large DNA databases. Genome Res. *11*, 1725–1729.

Nobelmann, B., and Lengeler, J.W. (1996). Molecular analysis of the gat genes from Escherichia coli and of their roles in galactitol transport and metabolism. J. Bacteriol. *178*, 6790–6795.

Perfeito, L., Fernandes, L., Mota, C., and Gordo, I. (2007). Adaptive mutations in bacteria: high rate and small effects. Science *317*, 813–815.

Poulsen, L.K., Licht, T.R., Rang, C., Krogfelt, K.A., and Molin, S. (1995). Physiological state of Escherichia coli BJ4 growing in the large intestines of streptomycin-treated mice. J Bacteriol *177*, 5840–5845.

Rainey, PB, Buckling, A, Kassen, R, and Travisano, M (2000). The emergence and maintenance of diversity: insights from experimental bacterial populations. Trends Ecol. Evol. *15*, 243–247.

Rang, C.U., Licht, T.R., Midtvedt, T., Conway, P.L., Chao, L., Krogfelt, K.A., Cohen, P.S., and Molin, S. (1999). Estimation of growth rates of Escherichia coli BJ4 in streptomycin-treated and previously germfree mice by in situ rRNA hybridization. Clin. Diagn. Lab. Immunol. *6*, 434–436.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Schiffels, S., Szollosi, G., Mustonen, V., and Lassig, M. (2011). Emergent Neutrality in Adaptive Asexual Evolution. Genetics *189*, 1361–1375.

Sniegowski, P.D., and Gerrish, P.J. (2010). Beneficial mutations and the dynamics of adaptation in asexual populations. Philos Trans R Soc Lond. B Biol Sci *365*, 1255–1263.

Sniegowski, P.D., Gerrish, P.J., and Lenski, R.E. (1997). Evolution of high mutation rates in experimental populations of E. coli. Nature *387*, 703–705.

Sousa, J.A.M. de, Campos, P.R.A., and Gordo, I. (2013). An ABC Method for Estimating the Rate and Distribution of Effects of Beneficial Mutations. Genome Biol. Evol. *5*, 794–806.

Strelkowa, N., and Lässig, M. (2012). Clonal interference in the evolution of influenza. Genetics *192*, 671–682.

Suppmann, B., and Sawers, G. (1994). Isolation and characterization of hypophosphite-resistant mutants of Escherichia coli: identification of the FocA protein, encoded by the pfl operon, as a putative formate transporter. Mol. Microbiol. *11*, 965–982.

T. Conway, K. A. Krogfelt, and P. S. Cohen (2004). The Life of Commensal Escherichia coli in the Mammalian Intestine. In EcoSal - Escherichia Coli and Salmonella: Cellular and Molecular Biology, R. Curtiss III, ed. (Washington, DC: ASM Press), p.

Wilson, K. (2001). Preparation of genomic DNA from bacteria. Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al *Chapter 2*, Unit 2.4.

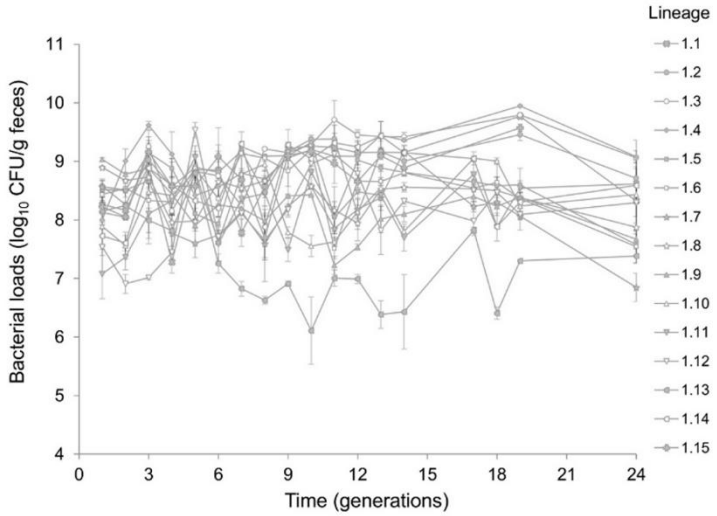
Wimpenny, J.W., and Cole, J.A. (1967). The regulation of metabolism in facultative bacteria. 3. The effect of nitrate. Biochim. Biophys. Acta *148*, 233–242.

Woods, R.J., Barrick, J.E., Cooper, T.F., Shrestha, U., Kauth, M.R., and Lenski, R.E. (2011). Second-order selection for evolvability in a large Escherichia coli population. Science *331*, 1433–1436.

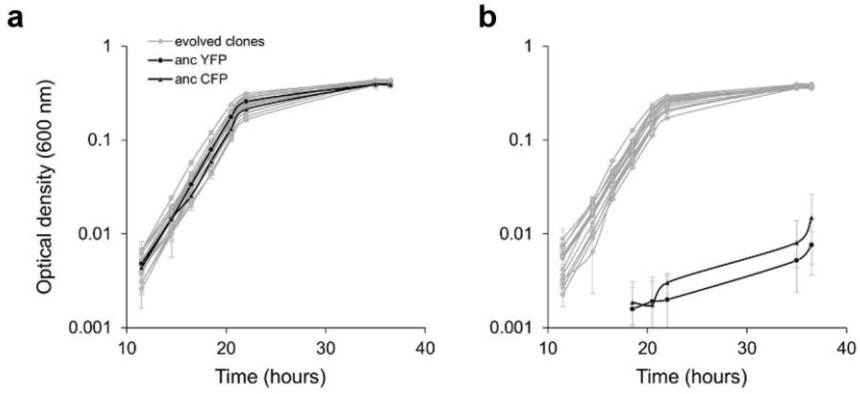
Wylie, C.S., Trout, A.D., Kessler, D.A., and Levine, H. (2010). Optimal strategy for competence differentiation in bacteria. PLoS Genet. *6*.

Yamada, M., and Saier, M.H., Jr (1987). Physical and genetic characterization of the glucitol operon in Escherichia coli. J. Bacteriol. *169*, 2990–2994.

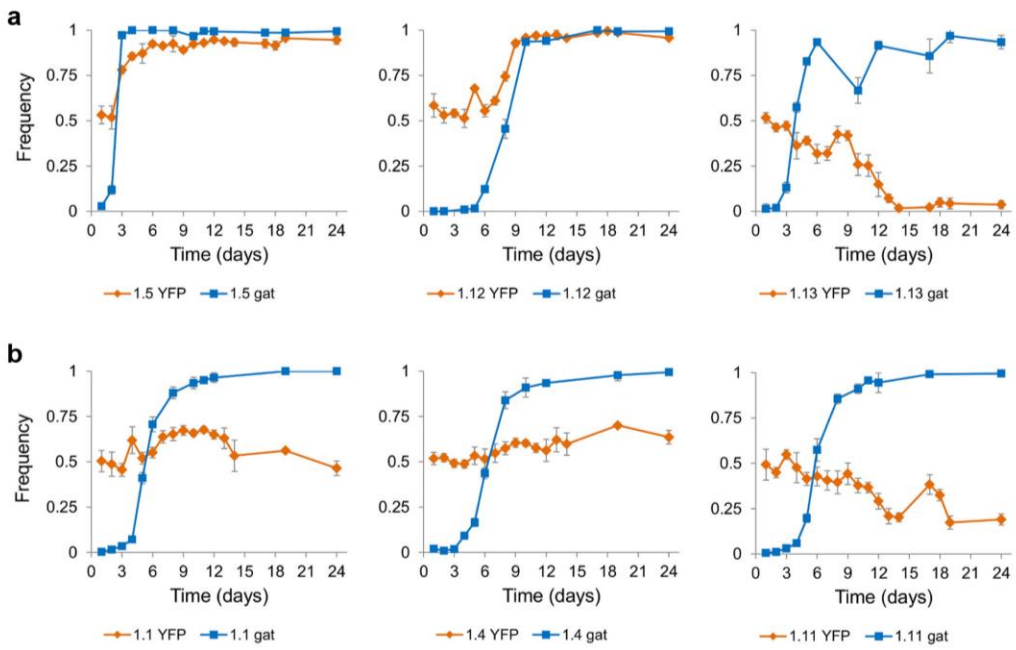
Supplementary Figures and Tables



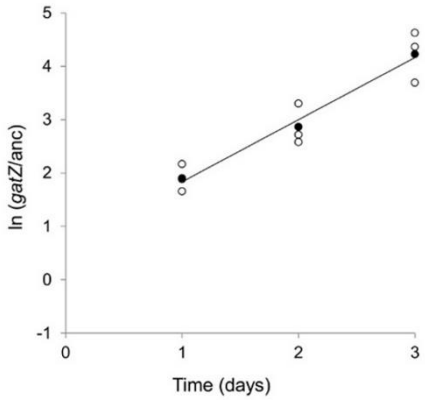
Supplementary Figure 1. Colonization of the mouse gut by *Escherichia coli*.
Bacterial loads per gram of feces (with 95% confidence intervals) during 24 days of adaptation of *E. coli* to the mouse gut (populations 1.1 to 1.15).



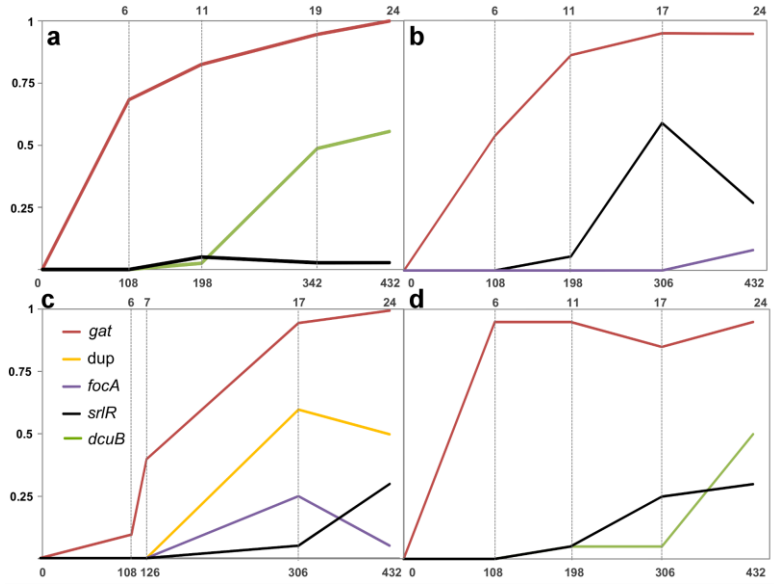
Supplementary Figure 2. Comparison of growth curves of ancestral strain and evolved clones.
Growth curves of ancestral (black) and evolved clones (grey) in MM with glycerol (**a**) and MM with glycerol and galactitol (**b**). Error bars represent the standard error of the mean of three independent measurements.



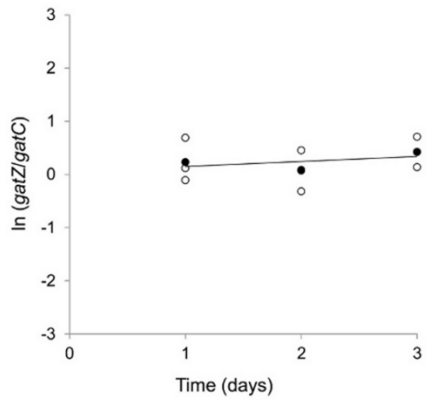
Supplementary Figure 3. Emergence and spread of beneficial mutations in the *gat* operon.
Dynamics of frequency change of the *gat*-negative phenotype (blue squares) and of the neutral fluorescent marker (orange diamonds) are shown for representative examples of populations where increase in frequency and eventual fixation of the *gat*-negative phenotype was accompanied by strong divergence of the fluorescent marker ((**A**) populations 1.5, 1.12 and 1.13) or not ((**B**) 1.1, 1.4 and 1.11).



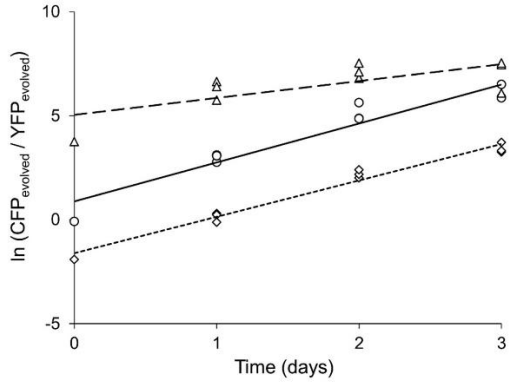
Supplementary Figure 4. Direct estimate of the selective advantage of a *gat* allele over the ancestral in the mouse gut.
Results of three independent competition experiments between *gatZ* mutant and the ancestral. We estimated a selective advantage of *gatZ* (corresponding to the slope (±2 s.e.m) of the linear regression of ln(*gatZ*/anc) along the three days of competition) of 0.065 (±0.013), R^2=0.99 over the ancestral.



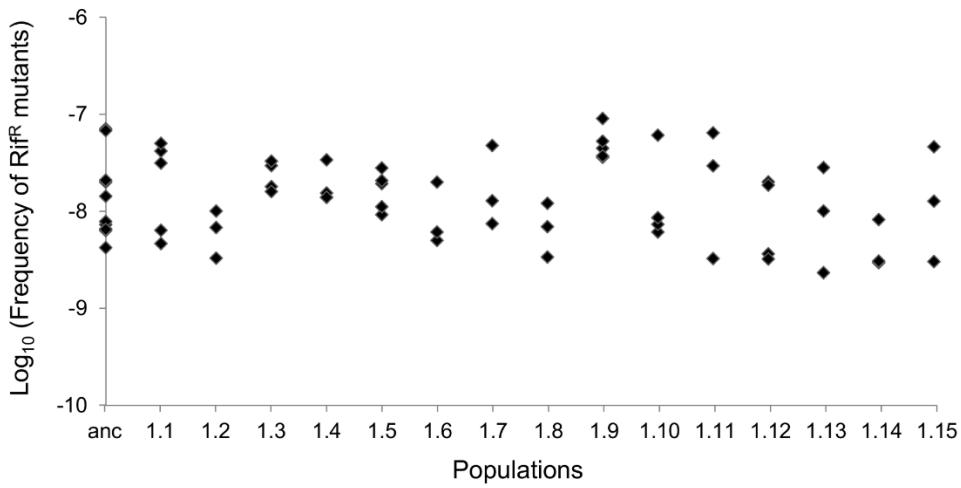
Supplementary Figure 5. Sums of frequencies of genotypes at different loci along time for the populations represented in Fig. 5.
As in Fig. 5 these frequencies are represented along 24 days (corresponding to 432 generations) of evolution inside the mouse gut (see Supplementary Tables 3 to 6 for numeric data). **a** corresponds to population 1.1, **b** to 1.11, **c** to 1.12 and **d** to 1.5.



Supplementary Figure 6. Direct estimate of the selective advantage of different *gat* alleles in the mouse gut.
Results of three independent competition experiments between *gatZ* and *gatC* mutants. We estimated a selective advantage of *gatZ* (corresponding to the slope (±2 s.e.m) of the linear regression of ln(*gatZ*/*gatC*) along the three days of competition) of 0.005 (±0.016), R^2=0.30 over *gatC*.



Supplementary Figure 7. Test for negative frequency-dependent selection of evolved clones.
30 CFP and 30 YFP clones isolated from population 1.13 after 24 days of colonization were competed *in vivo* at initial ratios CFP:YFP of 1:1 (open circles), 1:10 (open diamonds) and 100:1 (open triangles) for 3 days (corresponding to approximately 54 generations). Three independent competition experiments were performed for each ratio. The selective advantages per generation of the CFP in relation to the YFP population were calculated for each ratio of CFP over YFP. These are based on the slopes of the linear regression of ln(CFP/YFP) along time. The slopes (±2 s.e.m) are: 0.11 (±0.04), R^2=0.95 for 1:1 (dotted line); 0.10 (±0.02), R^2=0.99 for 1:10 (solid line) and 0.06 (±0.05), R^2=0.77 for 100:1(dashed line). A selective advantage was found for all the cases tested irrespective of their initial frequency, indicating that negative frequency-dependent selection is not the major process underling adaptation in the mouse gut.



Supplementary Figure 8. Fluctuation tests of evolved clones to test for the emergence of mutator alleles during adaptation to the gut.
Each symbol represents an independent measurement of the frequency of spontaneous mutants resistant to rifampicin. Measurements of the mutation frequency were performed for each of the 15 mice evolved populations (1.1 to 1.15) as well as for the respective ancestors. No significant difference was found between the estimations of the mutation frequencies for the ancestors and evolved populations (Mann-Whitney test with Bonferroni correction, P>0.05).

Supplementary Table 1. Mutations identified in the genomes of the ancestral strain.
Mutations were identified in the genome of the ancestor 0YFP by comparison with the reference genome (Blattner et al., 1997). Mutations in intergenic regions have the two flanking genes listed (e.g., fdrA/ylbF). Genes within brackets mean that the mutation happened within the gene. SNPs are represented by an arrow between the ancestral and the evolved nucleotide. Whenever a SNP gives rise to a non-synonymous mutation the amino acid replacement is also indicated. The symbol Δ means a deletion event and a + symbol represents an insertion of the nucleotide that follows the symbol. For intergenic mutations, the numbers in the annotation row represent nucleotides relative to each of the neighboring genes, where + indicates the distance downstream of the stop codon of a gene and - indicates the distance upstream of the gene, that is relative to the start codon. The mutations present in the 0YFP but not in the 0CFP, are underlined.

Clone	Genome Position	Gene	Mutation	Annotation
0YFP	547694	*fdrA/ylbF*	A→G	intergenic (+123/-1156)
	547835	*fdrA/ylbF*	+G	intergenic (+264/-1015)
	1395405	*[ynaJ]–[ttcA]*	Δ13,756 bp	Multigenic
	1976527	*insB–insA*	Δ776 bp	
	2369558	*arnT*	4 bp x 2	Duplication
	3422257	*rrlD*	A→C	Noncoding
	3422258	*rrlD*	T→A	Noncoding
	3422259	*rrlD*	C→T	Noncoding
	3434719	*trkA*	G→A	E60E GAG→GAA
	3472447	*rpsL*	T→C	K43R AAA→AGA
	3844290	*uhpT*	A→C	F301V TTT→GTT
	3957957	*ppiC/rep*	C→T	intergenic (-121/-743)
	4095684	*rhaB/rhaS*	T→C	intergenic (-213/-75)
	360473	*lacA–lacI*	Δ6264 bp	Multigenic
	788169	*[galK]*	Δ1034 bp	
	4294082	*RIP321*	Δ338 bp	

Supplementary Table 2. The number and nature of adaptive events across independently evolved clones.

In the list of mutations, the initials IS denote the abbreviation of insertion sequence element at the indicated position. The asterisk means that the corresponding SNP originated a STOP codon. For further details see Supplementary Table 1 legend. The last column shows the number of mutations segregating in the lineage from which the sequenced clone was isolated, as inferred from the theoretical model.

Clone	Genome Position	Gene	Mutation	Annotation	Inferred mutations
1CFP	2173589	*gatZ*	IS Ins	coding (755/1263)	2
	4346888	*dcuB/dcuR*	IS Ins	intergenic (-121/+450)	
2YFP	2173759	*gatZ*	IS Ins	coding (585/1263)	2
	4346888	*dcuB/dcuR*	IS Ins	intergenic (-121/+450)	
3YFP	2560011	*yffN*	G→A	C122Y TGC→TAC	2
	2175242	*gatY/fbaB*	IS Ins	intergenic (-16/+292)	
	4601260	*yjjP/yjjQ*	IS Ins	intergenic (-379/-244)	
4YFP	2173531	*gatZ*	IS1 Ins	coding (813/1263)	3
5YFP	2827493	*srlR*	G→C	G142A GGC→GCC	2
	2172869	*gatA*	IS Ins	coding (203/453)	
6YFP	2172262	*gatC*	del 1 bp	coding (39/1356)	2
	4346888	*dcuB/dcuR*	IS Ins	intergenic (-121/+450)	
7YFP	1420379	*ydaV*	C→A	L125M CTG→ATG	4
	1902231	*[manZ]–[kdgR]*	Δ5,451 bp		
	2175298	*gatY/fbaB*	IS Ins	intergenic (-72/+236)	
	953904	*focA/ycaO*	IS Ins	intergenic (-212/+194)	
8YFP	3268729	*garK*	A→G	F355S TTC→TCC	2
	2172869	*gatA*	IS Ins	coding (203/453)	
9YFP	2827073	*srlR*	A→T	K2I AAA→ATA	2
	2172636	*gatA*	IS Ins	coding (436/453)	
10CFP	2174223	*gatZ*	Δ2 bp	coding (120-121/1263)	4
	1388754	*[ycjY-ynaI]*	Δ5315 bp	large deletion	
	953904	*focA/ycaO*	IS5 Ins	intergenic (-212/+194)	
11CFP	2827095	*srlR*	Δ1 bp	coding (27/774)	4
	2172869	*gatA*	IS Ins	coding (203/453)	
12YFP	2172079	*gatC*	+C	coding (222/1356)	3
	4500113	*[insG-yaaI]*	2x151716 bp	large duplication	
13CFP	4637714	*arcA*	G→T	R206S CGC→AGC	2
	2171153	*gatC*	IS Ins	coding (1148/1356)	
14CFP	943941	*dmsC*	G→A	W229* TGG→TAG	2
	2175263	*gatY/fbaB*	IS Ins	intergenic (-37/+271)	
	2827117	*srlR*	C→T	Q17* CAG→TAG	

Supplementary Table 3. Frequencies of newly generated haplotypes along 24 days of evolution of population 1.1 inside the mouse gut.

Yellow and blue shading identify haplotypes belonging to the two sub-populations of bacteria labeled either with *cfp* or *yfp* alleles. The ancestral haplotype is, by definition, devoid of mutations. The genome position for a given mutation is only indicated for mutations that are not IS insertions. Mutations in intergenic regions have the two flanking genes listed (e.g., *dcuB/dcuR*). Whenever a SNP gives rise to a non-synonymous mutation the amino acid replacement is indicated. Synonymous mutations are represented as Syn. The symbol Δ means a deletion event and a + symbol represents an insertion of the nucleotide that follows the symbol. The asterisk means that the corresponding SNP originated a STOP codon. IS insertions at given position are indicated as IS Ins. The exact location of IS insertions is not indicated since it was determined (see Material and Methods), however the detection strategy allowed us to distinguish between different insertions in the same gene. Therefore every haplotype indicated in Supplementary Table 3 was confirmed to be unique. See Fig. 5 for a graphical representation of the data in this table.

Genome Position	Gene	Mutation	Haplotype frequencies				
			0 gen	108 gen	198 gen	342 gen	432 gen
			0.5	0.17	0.10		
	dcuB/dcuR	IS Ins				0.03	
	gatZ	IS Ins		0.07	0.05		0.03
	gatZ *dcuB/dcuR*	IS Ins IS Ins			0.03	0.30	0.42
2827852	*gatZ* *dcuB/dcuR* *srlR*	IS Ins IS Ins Δ 1bp				0.03	
2173887	*gatZ*	Δ 5bp		0.02			
2171894	*gatC*	Δ 1bp			0.03		
	gatY	IS Ins		0.02	0.03		0.06
	gatY *dcuB/dcuR*	IS Ins IS Ins				0.03	
	gatC	IS Ins		0.02	0.03		
	gatA	IS Ins		0.05	0.08		
	gatA	IS Ins		0.05	0.03		
			0.5	0.15	0.08	0.03	
	gatZ	IS Ins		0.12	0.38	0.46	0.28
2827549	*gatZ* *srlR*	IS Ins FE161QK			0.03		
	gatZ *dcuB/dcuR*	IS Ins IS Ins				0.05	0.03
	gatY	IS Ins		0.02			0.06
	gatC	IS Ins		0.05	0.05		

2171454	*gatC*	Q155*			0.03		
2172118	*gatC*	+C				0.03	
2174029	*gatZ*	∆1bp		0.02			
2173899	*gatZ*	Δ 5bp		0.02	0.05		
2173900	*gatZ*	Δ 5bp		0.12			
2173900	*gatZ* *dcuB/dcuR*	Δ 5bp IS Ins				0.05	0.08
	gatA	IS Ins		0.05			
	gatA	IS Ins		0.02	0.03		
	gatA *dcuB/dcuR*	IS Ins IS Ins					0.03
	gatA	IS Ins			0.03		0.03
2827529	*srlR*	G154E					

Supplementary Table 4. Frequencies of newly generated haplotypes along 24 days of evolution of population 1.5 inside the mouse gut.
See Supplementary Table 3 for further details.

Genome Position	Gene	Mutation	Haplotype frequencies				
			0 gen	108 gen	198 gen	306 gen	432 gen
			0.5			0.05	
	dcuB/dcuR	IS Ins					0.05
	gatA	IS Ins		0.05			
	gatA	IS Ins		0.10		0.10	
	gatA *dcuB/dcuR*	IS Ins IS Ins			0.05		
	gatY/fbaB	IS Ins		0.05	0.05		0.05
			0.5	0.05	0.05	0.10	
	gatY	IS Ins		0.75	0.65	0.40	0.15
2827489	*gatY* *srlR*	IS Ins P141S			0.05	0.05	
2827490	*gatY* *srlR*	IS Ins P141L				0.05	
2827493	*gatY* *srlR*	IS Ins G142E					0.05
	gatY	IS Ins					0.45

2827493	*dcuB/dcuR*	IS Ins					
	gatA	IS Ins			0.15	0.05	
	gatA *srlR*	IS Ins G142A				0.15	0.25
	gatA	IS Ins				0.05	
	dcuB/dcuR	IS Ins					

Supplementary Table 5 – Frequencies of newly generated haplotypes along 24 days of evolution of population 1.11 inside the mouse gut.
See Supplementary Table 3 for further details.

Genome Position	Gene	Mutation	Haplotype frequencies				
			0 gen	108 gen	198 gen	306 gen	432 gen
			0.5	0.23	0.08		
	gatA	IS Ins		0.33	0.50	0.15	0.57
2827492	*gatA* *srlR*	IS Ins G142S			0.03	0.08	0.03
2827493	*gatA* *srlR*	IS Ins G142D				0.03	
2827728	*gatA* *srlR*	IS Ins del1bp				0.03	
2827177	*gatA* *srlR*	IS Ins del7bp				0.18	0.03
2827627	*gatA* *srlR*	IS Ins E187K				0.03	
2827076	*gatA* *srlR*	IS Ins P3L				0.03	
2827726	*gatA* *srlR*	IS Ins ins7bp				0.03	
2827764	*gatA* *srlR*	IS Ins Syn					0.03
	gatA *focA*	IS Ins IS Ins					0.05
2827492	*gatA* *srlR* *focA*	IS Ins G142S IS Ins					0.03

2173900	*gatZ*	Δ 5bp		0.05	0.03	0.03	
	gatC	IS Ins		0.03			
	gatY	IS Ins			0.03		
	gatZ	IS Ins			0.03		
			0.5	0.23	0.06	0.05	0.05
2173878	*gatZ*	Δ 5bp		0.03	0.03		
2173878 2827278	*gatZ* *srlR*	Δ 5bp del2bp				0.03	
2173878 2827492	*gatZ* *srlR*	Δ 5bp G142S				0.03	
	gatC	IS Ins		0.08	0.11		
	gatA	IS Ins			0.06	0.03	
2827394	*gatA* *srlR*	IS Ins L109P				0.03	
2827496	*gatA* *srlR*	IS Ins G143V					0.05
	gatA	IS Ins		0.03		0.03	0.03
2827664	*gatA* *srlR*	IS Ins A99E				0.03	
2827529	*gatA* *srlR*	IS Ins G154E					0.03
2174150	*gatY*	del total			0.03	0.13	0.03
2174150 2827172	*gatY* *srlR*	del total T35N			0.03	0.10	0.03
2827172	*gatY* *srlR*	IS Ins T35N					0.03
2827276	*gatZ* *srlR*	IS Ins Δ 2bp					0.03

Supplementary Table 6. Frequencies of newly generated haplotypes along 24 days of evolution of population 1.12 inside the mouse gut.
See Supplementary Table 3 for further details.

Genome Position	Gene	Mutation	Haplotype frequencies				
			0 gen	108 gen	126 gen	306 gen	432 gen
			0.5	0.48	0.20	0.05	
2172078	*gatC*	+C				0.05	
			0.5	0.42	0.40		
	gatC	+C		0.09	0.40		0.15
	gatC *dup*	+C				0.60	0.50
	gatC *focA*	+C IS Ins				0.25	0.05
2827117	*gatC* *srlR*	+C Q17*				0.05	0.05
	gatC	+C					0.25
2827234	*srlR*	L97V					

Supplementary Table 7. The number and nature of adaptive events across independently evolved populations.

In the list of mutations del/dup denotes that either a deletion or a duplication of the indicated size occurred but it is not possible to distinguish between the two. For further details see Supplementary Table 1 legend. Underlined mutations were also found in the sequenced clones isolated from the corresponding populations (see Supplementary Table 2).

Population	Genome Position	Gene	Mutation	Annotation
1.1	735886	*ybfQ*	C→T	pseudogene (219/240 nt)
	1464225	*ydbA*	T→C	pseudogene (1704/2513 nt)
	2174223	*gatZ*	Δ2 bp	coding (120-121/1263 nt)
	2827492	*srlR*	G→A	G142S (GGC→AGC)
	1706868	*rsxC*	C→G	A642A (GCC→GCG)
	953904	*focA/ycaO*	IS Ins	
	953895	*focA/ycaO*	IS Ins	
	2171093	*gatC*	IS Ins	
	2172273	*gatC*	IS Ins	
1.11	2172869	*gatA*	IS Ins	coding (203/453 nt)
	2827172	*srlR*	C→A	T35N (ACC→AAC)
	4601110	*yjjP/yjjQ*	IS Ins	intergenic (-229/-387)
	2174150	*[gat Z-gatY]*	Δ1111bp	
	4601260	*yjjP/yjjQ*	IS Ins	
	3925441	*asnA*	del/dup 40bp	
	761046	*sucB*	del/dup 67bp	
	3996061	*uvrD*	del/dup 39bp	
	2176626	*fbaB/yegT*	IS Ins	
	2545378	*murP*	del/dup 15bp	
1.12	2827234	*srlR*	G→A	M56M (GTG→ATG)
	953904	*focA/ycaO*	IS Ins	
	3925441	*asnA*	del/dup 39bp	
	281295	*yagA/yagE*	del/dup 35bp	
	1284572	*narJ*	del/dup 47bp	
	2172079	*gatC*	+C	coding (222/1356 nt)
	2440543	*mnmC*	del/dup 106bp	
	4570195	*yjiS/yjiT*	IS Ins	
	4500113	*[insG-yaaI]*	2x 151716 bp	large duplication

Supplementary Table 8. Oligonucleotide primers used in this work.

Gene	Forward (5' - 3')	Reverse (5' - 3')	Usage
dcuB	GGCTGAAGGT GGAAGACGAA	ACATTTCGCGT GTTTCCTGC	Amplification of *dcuB*
focA	AGCGGATGTTT CGTTGCTTT	TGCTGCACATC AGTCGTTGT	Amplification of *focA*
gatABCD	TCCCACCGCAT CAATATAGCC	CAGTCCGGGG AATTATCAGCA	Amplification of *gatABCD*
gatZY	CACCTTTGGCG AGCATCTCA	AAAACACGCGC ACTTTGCTA	Amplification of *gatZ and gatY*
gatB	GATCCACTTTG GCAGTGGTG	CAGTCCGGGG AATTATCAGCA	Amplification of *gatB*
gatY	GCCACAATCGG CAATCACTT	AAAACACGCGC ACTTTGCTA	Amplification of *gatY*
dupl 150kb	GTTCGTTGCGC ATCAGTACG	GCTCTACCCGC AGGTCAAAA	Amplification of the new junction
srlR	GCATGCGGGT GATTTACAGC	TTCCGGTAAAC GGCTTGCTT	Amplification / sequencing of *srlR*
gatC	ATTAGCCGCCA GTTGGGTG		sequencing of gatC
	TGCCGATAATC AGCCCCATC		sequencing of gatC
	CCAGCCAACAT CGACCACAT		sequencing of gatC
gatZ	ATATCGCCTCG CGTAAAGCA		sequencing of gatZ
	ACCGTTTCTGG TGCTAACGG		sequencing of gatZ

CHAPTER 3

Adaptive immunity increases the pace and predictability of evolutionary change in commensal gut

Manuscript published in *Nature Communications*, November 30th, 2015

The author of this thesis performed all the experimental work and analysis described in this chapter.

Adaptive immunity increases the pace and predictability of evolutionary change in commensal gut

João Barroso-Batista[1], Jocelyne Demengeot[1] and Isabel Gordo[1]*

[1]Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6
*author for correspondence: igordo@igc.gulbenkian.pt

Abstract

Co-evolution between the mammalian immune system and the gut microbiota is believed to have shaped the microbiota's astonishing diversity. Here we test the corollary hypothesis that the adaptive immune system, directly or indirectly, influences the evolution of commensal species. We compare the evolution of *Escherichia coli* upon colonization of the gut of wild-type and *Rag2*$^{-/-}$ mice, which lack lymphocytes. We show that bacterial adaptation is slower in immune-compromised animals, a phenomenon explained by differences in the action of natural selection within each host. Emerging mutations exhibit strong beneficial effects in healthy hosts but substantial antagonistic pleiotropy in immune-deficient mice. This feature is due to changes in the composition of the gut microbiota, which differs according to the immune status of the host. Our results indicate that the adaptive immune system influences the tempo and predictability of *E. coli* adaptation to the mouse gut.

Introduction

The maintenance of a healthy status in mammals depends on the interaction between their microbiota and the immune system. The partnership between gut commensals and the host is hypothesized to result from millions of years of co-evolution, where the host immune response is tightly regulated to tolerate commensals, which in turn shape its development. Indeed, it recently became clear that specific bacterial species, such as segmented filamentous bacteria

(SFB) (Ivanov et al., 2009), *Clostridia* spp (Atarashi et al., 2011) and *Bacteroides fragilis* (Round and Mazmanian, 2010), can promote the expansion of regulatory and pro-inflammatory T cells, promoting immune homeostasis and contributing to restoration of health. Further support for a role of microbes in controlling host physiology comes from fecal transplants, which can affect their weight (Ridaura et al., 2013) and even protect hosts from invasion of pathogens (Cammarota et al., 2014). The patterns of microbiota species composition also differ between healthy and immune-compromised mice lacking an adaptive immune system (Kawamoto et al., 2014; Zhang et al., 2015), suggestive of the hypothesis of co-evolution between host and gut commensals (Ley et al., 2008). However, a direct support for co-evolution requires the study of how host genetics affects evolutionary change within the gut microbial species it carries and vice versa. These issues have been poorly studied, probably partly due to the difficulty in conducting studies that demonstrate the direct influence of microbial evolution on mammalian host evolution. It is however, easier to address the influence of host genetics on microbial ecology and evolution.

E. coli, the most abundant aerobe and one of the first species to colonize the human gut, offers a powerful system to investigate how and to what extent the genetic composition of a commensal changes with time when facing a healthy or a compromised host immune system. For example, in patients with inflammatory bowel diseases, such as Crohn's disease and ulcerative colitis, which are increasing in incidence in the human population and are associated with different mutations in genes with important immune functions, often show an increased abundance of Proteobacteria, namely *E. coli*, which tends to be sampled, in comparison with healthy hosts (Mukherjee et al., 2015; Sartor and Mazmanian, 2012).

Using a classical *E. coli* colonization system - the streptomycin-treated mouse - we have recently shown that *E. coli* undergoes rapid adaptation to the intestinal tract (Barroso-Batista et al., 2014). This adaptation was detected through observation of phenotypic sweeps, which occurred across populations of *E. coli* evolving independently in genetically identical immune-competent hosts. Here, we have tested the hypothesis that the process of adaptation of *E. coli* can be altered

in immune-compromised mice lacking an adaptive immune system (*Rag2*$^{-/-}$). We show that evolution of *E. coli* is slower in these hosts than in immune-competent animals, due to smaller selective effects of the first beneficial mutations that emerge in these bacteria. We also demonstrate that the strength of natural selection is dependent on the composition of the microbiota, which differs between immune-competent and immune-compromised animals. Furthermore, we describe the genetic targets of *E. coli* adaptation and find adaptive mutations specific to the host genetic background. Finally, the findings that emerging mutations exhibit strong beneficial effects in healthy hosts but substantial antagonistic pleiotropy in immune-deficient mice, support the notion that the adaptive immune system enhances the predictability of adaptive evolution of bacteria comprising the microbiota. Together**,** these results indicate that not only the microbiotic environment but also the pace and the path of adaptation of a commensal species can be altered by the host immune system.

Material and Methods

Ethics statement

All experiments involving animals were approved by the Institutional Ethics Committee at the Instituto Gulbenkian de Ciência (project no. 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.

E. coli and mouse strains

All bacterial strains used in this study were derived from *E. coli* K-12, strain MG1655 (Blattner et al., 1997). Strains DM08-YFP and DM09-CFP (MG1655, *galK::YFP/CFP* ampicillin resistant, amp^{R} *(pZ12),* streptomycin resistant, str^{R} *(rpsl150), ΔlacIZYA*), described elsewhere (Barroso-Batista et al., 2014) were used in the colonization experiments. For *in vivo* fitness assays (see below) two previously described (Barroso-Batista et al., 2014) *gat* mutants of *E. coli* were used: one harboring an insertion sequence (IS) in the coding region of *gatZ* (clone 4YFP) and another carrying an insertion of 1 bp in *gatC* (derived from clone

12YFP). These clones were isolated after 24 days of evolution in WT mouse gut and differed from the ancestral by a single mutation (as confirmed by WGS). Both clones display the *gat*-negative phenotype.

C57BL/6 (WT) and C57BL/6 *Rag2*$^{-/-}$ mice (*Rag2*$^{-/-}$), were used as a model of immune-competent and immune-compromised hosts, respectively. *Rag2*$^{-/-}$ animals carry a null mutation in the *rag* gene essential for the immunoglobulin and T-cell receptor gene recombination and thus lack mature T and B lymphocytes. Six- to eight-week-old WT and *Rag2*$^{-/-}$ male mice bred at the IGC rodent facility and maintained under strict specific pathogen-free (SPF) conditions (barrier facility with autoclaved caging, food and water) were used for all the experiments, unless otherwise stated. For co-housing experiments, 1-month-old WT and *Rag2*$^{-/-}$ mice (7 animals each) were placed in the same cage for 4 weeks, before *E. coli* colonization. GF C57BL/6 (WT) and C57BL/6 *Rag2*$^{-/-}$ (*Rag2*$^{-/-}$) animals were bred and raised at the IGC gnotobiology facility in dedicated axenic isolators (La Calhene/ORM), an activity partially sponsored by the EU FP7 InfrafrontierI3-EMMA consortium. Young adults were transferred into sterile ISOcages (Tecniplast) before the competition experiments. All animals were kept in individual cages during both the evolution and the competition experiments.

Dynamics of adaptation of *E. coli* in the mouse gut

To study *E. coli* adaptation to the gut we used a streptomycin-treated mouse colonization model (T. Conway et al., 2004). Briefly, SPF-raised mice (see *E. coli* and mouse strains above) were given autoclaved drinking water containing streptomycin (5g/l) for one day. After 4 hours of starvation for water and food, the animals were gavaged with 100 µl of a suspension of 10^8 colony forming units (CFUs) of a mixture of YFP- and CFP-labeled bacteria (ratio 1:1) grown at 37°C in brain heart infusion medium to optical density $(OD)_{600}$ of 2. After gavage, the animals were housed in individual cages, and both food and water containing streptomycin were returned to them. Fecal pellets were collected for 24 days, diluted in PBS and plated in Luria Broth (LB) agar. Plates were incubated overnight and the frequencies of CFP- or YFP-labeled bacteria were assessed by counting the fluorescent colonies with the help of a fluorescent stereoscope

(SteREO Lumar, Carl Zeiss). A sample of each collected fecal pellet was daily stored in 15% glycerol at -80ºC for further experiments.

For the evolution experiment, groups of 5 WT and 5 *Rag2*$^{-/-}$ mice were gavaged and followed for 24 days. This procedure was repeated three times, such that a total of 15 mice of each genotype were analyzed. While the results of *E. coli* evolution in WT mice were published earlier (Barroso-Batista et al., 2014), experiments in both genotypes were conducted at the same time.

Dynamics of *gat*-negative phenotype

To investigate the dynamics of appearance and expansion of beneficial mutations in the *gat* operon, we determined the frequency of bacteria unable to metabolize galactitol (*gat*-negative phenotype) within a given bacterial population of the evolution experiment (Barroso-Batista et al., 2014). We plated frozen fecal pellets in Mac Conkey agar supplemented with 1% galactitol and streptomycin (100 µg/ml). Plates were incubated for 20 hours at 28°C. The frequency of galactitol-metabolizing mutants for each time point was estimated by counting the number of white (*gat* mutants) and red colonies in Mac Conkey-galactitol plates.

Division rate of *E. coli* in the mouse gut

To determine the growth rate of *E. coli* in the mouse gut we used a modified version of a previously described method (Poulsen et al., 1995) based on the correlation between growth rate and ribosomal content of bacterial cells. First, we established a calibration curve (see below) that relates the growth rate of *E. coli* populations growing in a given medium and the fluorescent intensity of a probe specific to *E. coli* 23S rRNA (as a measure of ribosomal content). This calibration curve was used to determine the ribosomal content of *E. coli* cells recovered from fecal pellets of WT and *Rag2*$^{-/-}$ mice and infer their growth rates to establish if any differences in the division time of *E. coli* exist in the different hosts.

To obtain a range of growth rates for *E. coli*, we grew *E. coli* strain DM09 (ancestral) in M9 minimal medium supplemented with different carbon sources, which support different growth rates. For the initial inoculum, DM09 was grown for 24 hours in 5 ml of LB supplemented with streptomycin at 37ºC with aeration. The

culture was washed 3 times in M9 with no carbon source and diluted to OD 2. 2 µl of a 10^{-1} dilution (~10^5 cells) were used to inoculate 198 µl of M9 minimal medium supplemented with one of the following carbon sources: ribose 0.4%, sorbitol 0.4%, xylose 0.4%, fructose 0.4%, maltose 0.4%, arabinose 0.4%, gluconate 0.5% or gluconate 1%. The growth curve assays were performed in a Bioscreen plate incubated in Bioscreen C Microbiology Reader equipment at 37ºC with continuous shaking. ODs at 600nm were monitored for 48 hours. All growth measurements were repeated four times. Specific growth rates were calculated based on the slope of the linear regression of the OD increase over time, during exponential growth. This parameter was then used to calculate the growth rates of *E. coli* in the different media. We then grew new cultures in the same conditions as described above, harvesting the cells in mid-log phase. The cells were fixed in 4% paraformaldehyde overnight, washed two times in PBS 1x and stored at -80ºC until use. For the whole-cell hybridization we used the protocol described by Poulsen *et al.* (Poulsen et al., 1995) with some modifications (Skinner et al., 2013). Briefly, fixed *E. coli* cells from the *in vitro* cultures were washed once with PBS, once with solution I (35% formamide, 100mM Tris (pH 7.5), 0.1% SDS, 0.9 M NaCl), and hybridized with a probe specific to *E. coli* 23S rRNA (EC 1531; 5'-CACCGTAGTGCCTCGTCATCA-3') labeled with the fluorochrome Cy3 (final concentration 2.5 ng/µl). The hybridizations were carried for 16 hours at 37ºC. The cells were then washed once with solution I and twice with solution II (35% formamide, 100mM Tris (pH 7.5), 0.9 M NaCl). Next, the cells were incubated for 15 min at 37ºC, centrifuged and ressuspended in 1x PBS. Cy3 fluorescence was measured by flow cytometry (LSR Fortessa, BD) and median fluorescence analyzed with FlowJo v10. For each sample, two hybridizations were performed. The average of median fluorescence intensity of Cy3 was correlated with growth rate for each growth condition, thus enabling a measure of division time of *E. coli*.

WT and *Rag2*$^{-/-}$ mice (n=6) were colonized with *E. coli* strain DM09 (ancestral) following the protocol described above for the evolution experiments. Fecal pellets were collected at day 1 and 3 after gavage, homogenized in PBS 1x and fixed in 4% paraformaldehyde overnight. Following fixation, cells were washed twice in 1x PBS and stored at -80ºC until use. Whole-cell hybridization with the probe specific

for *E. coli* 23S rRNA was performed according to the protocol described before for the *in vitro* cultures. On the basis of the fluorescence intensity of the probe obtained with the hybridization and the previously obtained calibration curve, we inferred a given duplication time for *E. coli* colonizing WT and *Rag2*$^{-/-}$ mice (see Fig. 2b and Supplementary Fig. 4).

Comparison of *E. coli* mutation rate between hosts

To estimate the equilibrium frequency for antibiotic resistance clones we determined the fraction of *E. coli* clones carrying spontaneous resistance to different antibiotics. Slightly deleterious mutations conferring resistance to nalidix acid and rifampicin (point mutations in *gyrA* and *rpoB*, respectively) or furazolidone (mutations in *nfsA*, including IS insertions) are expected to spontaneously occur and be continuously eliminated by purifying selection in *E. coli* populations colonizing the mouse gut, hence resistance alleles are expected to reach a mutation-selection balance and their frequency to reach a stable equilibrium which is proportional to the mutation rate.

Rag2$^{-/-}$ and WT mice were colonized with *E. coli* strain DM08 (ancestral) following the protocol described above for the evolution experiments, and the frequency of spontaneous resistance was determined in fecal pellets collected over 15 days. We tested 6 animals of each genotype for rifampicin and nalidixic acid resistance and 10 *Rag2*$^{-/-}$ and 11 WT mice for furazolidone resistance. Fecal pellets were collected, homogenized in PBS 1x and dilutions plated either in LB agar supplemented with streptomycin (100 µg/ml) or streptomycin (100 µg/ml) plus nalidixic acid (40 µg/ml), rifampicin (100 µg/ml) or furazolidone (1.25 µg/ml). The frequency of mutants resistant to each antibiotic was calculated as the ratio between the number of antibiotic-resistant clones and the total number of cells in each *E. coli* population. We also estimated the transposition frequency based on the fraction of furazolidone resistance mutants harboring insertions in *nfsA*.

***In vivo* competition fitness assays**

We measured the relative fitness *in vivo* of an *E. coli* clone harboring a IS insertion in the *gatZ* gene (clone 4YFP (Barroso-Batista et al., 2014)) isolated from

fecal samples after 24 days of adaptation to the gut of WT mice. This clone was competed against the ancestral strain labeled with the opposite fluorescent marker (DM09) (Fig. 2 and 3) or against a *gatC* mutant (derived from clone 12YFP) (Barroso-Batista et al., 2014) carrying an insertion of 1bp in the *gatC* gene (Fig. 4).

The competitions were performed at a ratio of 1:1, over 3 days of co-colonization, following the same procedure described for the evolution experiments in SPF mice; with no streptomycin treatment for competitions carried out in GF mice. We estimated the selective advantage of *gatZ* over the ancestral from the slope of the linear regression of ln(*gatZ*/anc) along time.

Microbiota analysis

To assess the gut microbiota composition of unmanipulated mice we analyzed fecal samples collected from 5 independently housed WT and *Rag2*$^{-/-}$ mice belonging to different litters and in two experimental blocks (3+2 mice). These animals were bred and maintained in similar SPF conditions in the same barrier facility (see *E. coli* and mouse strains). With this experimental design, we aimed both to characterize a representative sample of the microbiota composition associated with each host genotype and minimize the cage effects that could bias our analysis.

To characterize the microbiotal context in which *E. coli* was evolving, we analyzed fecal samples collected from WT and *Rag2*$^{-/-}$ mice at day 3 of the evolution experiment, that is, streptomycin-treated mice that had been colonized with *E. coli* for 3 days and housed individually. Similarly to unmanipulated animals, these mice were also bred and raised in SPF conditions but treated with streptomycin before and during colonization with *E. coli* (see Dynamics of adaptation of *E. coli* in the mouse gut above). We analyzed a total of 10 mice of each genotype, with a minimum of 3 animals from each experimental block (WT: 3+4+3; *Rag2*$^{-/-}$: 3+3+4).

The extraction was performed with a QIAamp DNA Stool Mini Kit (QIAGEN), according to the manufacturer's instructions and an additional step of mechanical disruption (Thompson et al., 2015).

16S rRNA gene amplification and sequencing was carried out at the IGC Genomics Unit. Samples were amplified using primers specific to the V3-V4 region of the 16s rRNA gene and pair-end sequenced on an Illumina MiSeq Benchtop Sequencer, following Illumina recommendations. Samples were sequenced to a depth of at least 20000 high-quality sequences. Reads were processed and analyzed using mother (Schloss et al., 2009) software, following the MiSeq SOP (Schloss et al., 2011) on the mothur wiki (http://www.mothur.org/wiki/MiSeq_SOP) with some modifications (Thompson et al., 2015). To minimize possible biases related with variable sequencing depth across samples, a subsampling of 20000 high-quality reads was performed. Sequences were aligned in accordance with SILVA alignment (Pruesse et al., 2007) and unique sequences were clustered into operational taxonomic units (OTUs) using a 97% identity cutoff. A phylogenetic tree was inferred using clearcut (Sheneman et al., 2006) on the 16S rRNA sequence alignment generated by mothur. Alpha and beta measurements of diversity, both OTU- and phylogeny-based, were obtained using mothur. The Chao (Chao, 1984) and Shannon (Shannon, 1948) indices were used to estimate community richness and diversity, respectively. The phylogenetic diversity was calculated as the total of the unique branch length in the phylogenetic tree. As a measure of beta diversity we evaluated the structure of microbial communities, a parameter that takes into account not only the presence and absence of organisms but also their abundance. To assess similarities in community structure OTU- (Bray-Curtis (Bray and Curtis, 1957)) or phylogeny-based (weighted UniFrac (Lozupone et al., 2007)) measures of beta diversity were calculated. Differences in phylogenetic community membership were evaluated using unweighted UniFrac (Lozupone et al., 2011). PCoA on the distance matrices of the beta diversity metrics were performed using mothur and R software package (http://www.R-project.org).

Haplotypic structure of *Rag*$^{-/-}$ populations

To estimate the *E. coli* haplotypic frequencies evolving in the *Rag2*$^{-/-}$ mice we analyzed ~20 clones isolated from fecal pellets from the last day of the evolution experiment for each of the evolved lines. These clones were screened for the

presence of mutations shown to be adaptive in WT animals (Barroso-Batista et al., 2014): mutations in the *gat* operon, insertions of IS elements in the regulatory regions of *focA* and *dcuB* and point mutations in the coding region of *srlR*. We typed *gat* mutants based on the inability to metabolize galactitol using the procedure described above for the dynamics of *gat*-negative phenotype. The remaining mutations were identified by target PCR; insertions of IS elements were identified based on the increase in size of the PCR products and Sanger sequencing was used to detect SNPs. PCR reactions were performed in the same conditions and with primers as previously described (Barroso-Batista et al., 2014), but in a total volume of 12.5 µl.

WGS of *E. coli* populations

To identify the pool of mutations segregating in *E. coli* populations evolved in *Rag2*$^{-/-}$ and WT mice we performed whole genome sequencing of a large sample of each population. We isolated more than 1000 clones from mice fecal pellets from day 24 of evolution, and we extracted the DNA from this mixture of clones following a protocol previously described (Barroso-Batista et al., 2014). The DNA library construction and sequencing was carried out by the IGC genomics facility. Each sample was pair-end sequenced on an Illumina MiSeq Benchtop Sequencer. Standard procedures produced data sets of Illumina paired-end 250-bp read pairs. The mean coverage per sample was between 100x and 180x for both WT and *Rag2*$^{-/-}$. Subsampling of 3 million reads was performed to obtain an average of 100x coverage. Mutations were identified using the BRESEQ pipeline v0.23 with the polymorphism option on. The default settings were used except for: a) requiring a minimum coverage of three reads on each strand per polymorphism; b) elimination of polymorphism predictions occurring in homopolymers of length greater than 3; c) elimination of polymorphism predictions with significant (p-value<0.05) strand or base quality score bias. In addition, unassigned new junctions supported by a minimum of two reads in each direction and insertions of IS elements where only one new junction was identified were also accepted. A cutoff of 5% of frequency was used to identify mutations segregating at high frequency in the populations. We defined parallel mutations as mutational events

that occurred in a minimum of two animals (either WT or *Rag2*$^{-/-}$) and that reached a minimum frequency of 10% in at least one population.

Statistical analysis

All statistical analyses were performed in R software: http://www.r-project.org/.

To test for differences in the initial deviation of the fluorescent marker, we used a binomial test that compared the number of WT and *Rag2*$^{-/-}$ lines in which the slope of the linear regression of ln(YFP/CFP) deviated significantly from 0 on the first 6 days of colonization.

We used a linear mixed model on the logarithm of bacterial counts per gram of feces, with host genotype and time as fixed effects and individuals as random effects to compare the temporal dynamics of bacterial loads between *Rag2*$^{-/-}$ and WT mice. To test whether the bacterial loads over time were affected by host genotype we contrasted a model with the interaction between these two effects with a model lacking this interaction. Additionally, we also tested if the bacterial loads differed globally between the two genotypes, by comparing models including or excluding the host genotype as a fixed effect.

To assess the influence of host genotype on the temporal dynamics of *gat*-negative phenotype, we employed a generalized linear mixed model, with host genotype and time as fixed effects and individuals as random effects. The *gat* frequencies were weighted by the number of counted CFUs and the initial frequency was constrained to the same value in WT and *Rag2*$^{-/-}$ dynamics.

Mann-Whitney *U*-tests were used to evaluate differences on the estimated generation time or mutation frequency between WT and *Rag2*$^{-/-}$ mice.

Analysis of variance (ANOVA) was used to investigate the impact of the microbiota and the immune state on the selective advantage of the *gatZ* mutation, followed by Tukey's *post hoc* tests to assess differences between groups. Two-tailed F-tests were used to determine whether the variance for the selective advantage of *gatZ* was increased in *Rag2*$^{-/-}$ compared with WT animals.

Mann-Whitney *U*-tests were used to identify genera differentially represented in WT or *Rag2*$^{-/-}$ mice. The *P* values were adjusted for multiple comparisons using the Benjamini and Hochberg correction. Two-tailed F-tests were used to determine

differences in the variance of the genera. Analysis of molecular variance (AMOVA) were used to test for differences in the structure or membership of the gut microbiota of WT and *Rag2*$^{-/-}$ animals.

To determine whether the parallel mutations identified by WGS were more represented in one host that in the other we performed binomial tests.

Results

E. coli adaptation is slower in immune-compromised mice

To study *E. coli* adaptation in the gut of *Rag2*$^{-/-}$ mice we colonized 15 animals (see Methods) with two *E. coli* strains, isogenic except for the presence of a neutral fluorescent marker (cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP)). We first measured the frequency of the neutral marker and its dynamics by daily monitoring of *E. coli* numbers and fluorescence in the fecal content. Much smaller changes in marker frequency were detected in *Rag2*$^{-/-}$ compared with wild-type (WT) mice (Barroso-Batista et al., 2014) during the first 6 days of colonization (Fig. 1a, binomial test P=0.03), suggesting a slower rate of adaptation in immune-compromised hosts. This delayed evolutionary change was not due to smaller population size, as similar loads of *E. coli* were recovered from both animals (Supplementary Fig. 1). After this initial period, the marker frequency started to diverge (Fig. 1b) in some *Rag2*$^{-/-}$ animals, as had been observed in WT mice (Barroso-Batista et al., 2014). Some populations showed a signature of a selective sweep, where a single marker steadily increased towards fixation (lines R1.1 and R1.12). Other populations showed the classical signature of clonal interference (R1.11 and R1.15), where clones carrying a given fluorescence first increased in frequency and then were replaced by clones with a different fluorescence, likely due to beneficial mutations with stronger effect arising in the latter. We next asked whether the slower pace of adaptation in *Rag2*$^{-/-}$ mice could be due to a longer time for sweeps to occur. In WT mice, the first adaptive mutations to appear involve changes in genes of the *gat* operon, conferring *E. coli* with a *gat*-negative phenotype (Barroso-Batista et al., 2014). In *Rag2*$^{-/-}$ mice the same phenotype emerged and swept (Supplementary Fig. 2; Fig. 1c), but at a

slower pace (z=-25.5, $P<10^{-3}$). Though mutations in this operon ultimately reached high frequency in the majority of lineages evolving in *Rag2*$^{-/-}$ mice, they could only be detected after 3 days of adaptation, contrasting with what was seen in WT mice. This occurred within the same experimental groups (Supplementary Fig. 2, upper panels versus lower panels). Although we detected populations where the increase in frequency of the phenotype was as fast as in WT (Supplementary Fig. 2, lines R1.3 and R1.10), this increase was slower in the majority of the lines. Thus, the *gat*-negative phenotype arose faster (Fig. 1c), and variance in this phenotype was eliminated quicker (Fig. 1d), in WT compared with *Rag*$^{-/-}$ animals. Both patterns are consistent with faster adaptation under the pressures of an adaptive immune system.

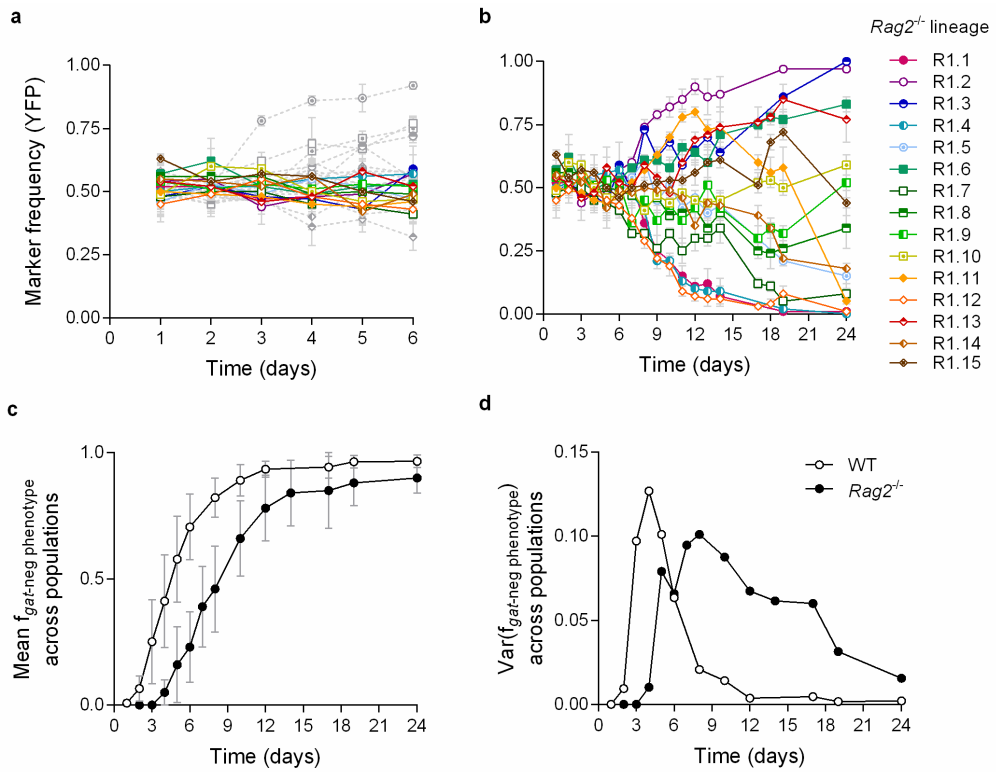


Figure 1. *E. coli* adapts at a slower pace in immune-compromised mice.
Dynamics of marker frequency (±2 s.e.m., representing the error associated with the sampling when calculating the frequency of the marker) during the first 6 (**a**) or 24 (**b**) days of adaptation of *E. coli* upon colonization of *Rag2*$^{-/-}$ (colored) or WT (grey) mice (reproduced from Barroso-Batista et al., 2014 for clarity). (**c**) Average frequency of the *gat*-negative phenotype (±2 s.e.m) over time across WT (white) and *Rag2*$^{-/-}$ (black) animals (n=15, for each host genotype) and (**d**) the variance for the frequency.

Duplication time of *E. coli* in the mouse gut

We then sought to identify the mechanism responsible for the observed slower adaptive pace in *Rag2*$^{-/-}$ mice. Adaptation rate depends on generation time, population size, mutation rate and the strength of selection on newly adaptive alleles (Charlesworth, 2010). Population size estimates based on bacterial loads recovered from both groups of animals (Supplementary Fig. 1) suggested that this factor would not play a leading role in the observed differences. To determine the division rate of *E. coli* in the mouse gut, we used *in situ* hybridization with a probe specific for *E. coli* 23S ribosomal RNA (rRNA) to estimate cellular rRNA content, which strongly correlates with bacterial division rate (Supplementary Fig. 3), using an adapted version of a previously described method (Poulsen et al., 1995). We colonized WT and *Rag2*$^{-/-}$ mice with *E. coli* and collected fecal samples at day 1 and 3 after inoculation, when *E. coli* had already reached the same load as observed during the course of the evolution experiment (Supplementary Fig. 1). On the basis of the fluorescence intensity of hybridized *E. coli* cells, we inferred an average duplication time of 66 (±3, 2 s.e.m.) and 76 (±3, 2 s.e.m.) min in *Rag2*$^{-/-}$ and WT mice, respectively (Fig. 2a). As the generation time for *E. coli* was significantly smaller in *Rag2*$^{-/-}$ than in WT mice (Mann-Whitney *U*-test, W=111.5, $P<10^{-3}$), this parameter also could not explain the different rates of adaptation observed between hosts.

Having observed a shorter *E. coli* duplication time in *Rag2*$^{-/-}$ compared with WT animals, we tested the hypothesis that the observed delay in the emergence of the adaptive phenotype in *Rag2*$^{-/-}$ mice was due to a lower spontaneous mutation rate in these hosts. We determined the spontaneous mutation frequency towards resistance to different antibiotics, by measuring the number of clones resistant to rifampicin, nalidixic acid or furazolidone in *E. coli* populations colonizing *Rag2*$^{-/-}$ and WT mice over a period of 15 days (Fig. 2b). Resistance against these antibiotics can be conferred through mutations in specific *E. coli* genes (*rpoB*, *gyrA* and *nfsA*, respectively). Resistance to rifampicin (Rif^{R}) and nalidixic acid (Nal^{R}) involves mainly point mutations, while resistance to furazolidone (Fzd^{R}) can also be acquired by transposition of IS elements (Whiteway et al., 1998). Assuming that such resistant alleles are slightly

deleterious, and thus kept at mutation-selection balance in a large population, the fraction of resistant mutants is proportional to the mutation rate (Charlesworth, 2010). We estimated an average $\log_{10}$ mutation frequency for Rif^R of -7.53 in *Rag2*$^{-/-}$ animals, not significantly different from that in WT mice (Mann-Whitney *U*-test, *W*=284.5, *P*=0.95). A similar result was found for Nal^R mutants (-7.76 in WT and -7.76 in *Rag2*$^{-/-}$ mice, Mann-Whitney *U*-test, *W*=202, *P*=0.65) and Fzd^R (-5.65 in WT and -5.49 in *Rag2*$^{-/-}$ mice, Mann-Whitney *U*-test, *W*=1613, *P*=0.10). We also measured the *in vivo* frequency of spontaneous resistant mutants to furazolidone, where resistance was achieved through transpositions, therefore providing the first estimate of the *in vivo* spontaneous transposition frequency. This is an important parameter in *E. coli* adaptation to the gut given that approximately half of the adaptive mutations identified in WT mice were caused by insertion of transposable elements (Barroso-Batista et al., 2014). We estimated an average $\log_{10}$ transposition frequency of -5.99 in WT and a similar frequency of -5.75 in *Rag2*$^{-/-}$ mice (Mann-Whitney *U*-test, *W*=83, *P*=0.09). Taken together, these results indicate that both the point mutation frequency and transposition frequency are similar in immune-compromised and immune-competent animals, suggesting that the slower rate of adaptation observed in the former was not due to an overall decreased mutation rate of *E. coli*.

Altered selective pressures in immune-compromised mice

We next asked whether the fitness effects of adaptive mutations were different in the gut of *Rag2*$^{-/-}$ and WT mice. We determined the fitness effect (*s*) of a mutant (*gatZ*) harboring a single beneficial mutation through *in vivo* competition assays against the ancestral *E. coli* (see Methods). In WT mice (Fig. 2c, left panel; Supplementary Table 1) we estimated a mean advantage, per hour (s_{gatZ}), of 0.068 (±0.008, 2 s.e.m.). In contrast, the selective effect of the *gatZ* mutant was smaller in *Rag2*$^{-/-}$ mice (s_{gatZ} = 0.03 (±0.01, 2 s.e.m.)) compared with WT (ANOVA with Tukey's *post hoc* test, $P<10^{-3}$). However, higher variation in the selective effects of the *gatZ* mutation was found in *Rag2*$^{-/-}$ mice (Fig. 2c, right panel; Supplementary Table 1), with a two-tailed test for variance in s_{gatZ} between *Rag2*$^{-/-}$ and WT being marginally significant (*F*=0.30, *P*=0.09). The mutation shows antagonistic

pleiotropy, that is, in some *Rag2*$^{-/-}$ mice it was advantageous (*Rag2*$^{-/-}$ 1 and 5) while in others it was neutral or slightly deleterious (*Rag2*$^{-/-}$ 4 and 10).

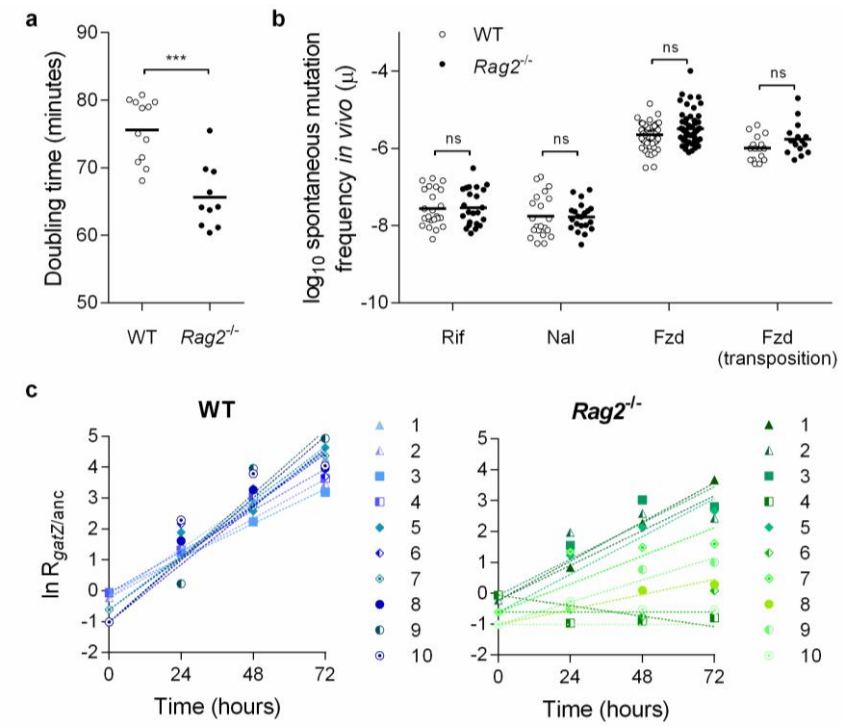


Figure 2. Selection strength, not mutation rate, is responsible for lower rate of adaptation in immune-compromised mice.

(**a**) Higher growth rate of *E. coli* in immune-compromised mice. The doubling time of *E. coli* colonizing the gut of WT or *Rag2*$^{-/-}$ is estimated by the amount of rRNA (see Supplementary Fig. 3). Each symbol corresponds to an average of two independent quantifications made for each mouse (n=6) during days 1 and 3 of colonization. The doubling time is significantly lower in *Rag2*$^{-/-}$ than in WT mice, indicating that *E. coli* divides faster in the gut of *Rag2*$^{-/-}$ than in WT animals (Mann-Whitney *U*-test, W=111.5, $P<10^{-3}$). ***$P < 0.001$. (**b**) Similar frequency of spontaneous mutation in immune-competent and immune-compromised mice: the $\log_{10}$ frequency of spontaneous resistance (μ) in WT (open circles) and in *Rag2*$^{-/-}$ (full circles) animals is proportional to the mutation rate at mutation-selection balance. Symbols correspond to measurements of mutation frequency over a period of 15 days for rifampicin (n=6), nalidixic acid (n=6) and furazolidone resistance (n=11 and 10). We estimated an average $\log_{10}$ μ_{WT} ~-7.56 and $\mu_{Rag2-/-}$ ~-7.53 for rifampicin (Mann-Whitney *U*-test, W=284.5, P=0.95), μ_{WT} ~-7.76 and $\mu_{Rag2-/-}$ ~-7.77 for nalidixic acid (Mann-Whitney *U*-test, W=202, P=0.65) and μ_{WT} ~-5.65 and $\mu_{Rag2-/-}$ ~-5.49 for furazolidone resistances (Mann-Whitney *U*-test, W=1613, P=0.10). The frequency of spontaneous transpositions corresponds to the fraction of mutations conferring furazolidone resistance, which are caused by insertion of transposable elements: $\mu_{TE\ WT}$ ~-5.99 and $\mu_{TE\ Rag2-/-}$ ~-5.75 (Mann–Whitney *U*-test, W=83, P=0.09). ns (not significant) $P > 0.05$. (**c**) Mean selective advantage of an emerging *gat* allele is lower in *Rag2*$^{-/-}$ than in WT mice. Results of short-term competitive fitness experiments (n=10) between a *gatZ* mutant and the ancestral *E. coli*. *Rag2*$^{-/-}$ (right panel) or WT (left panel) mice from the same litter are represented by symbols of the same shape (triangles, squares, diamonds or circles). The selective advantage of *gatZ* along the three days of competition is inferred from the slope of the linear regression of ln(*gatZ*/anc) (dashed lines, showed for each mouse), corresponding to the selection coefficient.

We note that variability for the fitness effect of the mutant was observed in mice from the same litter (for example, *Rag2*$^{-/-}$ littermates 3 and 4 or 5 and 6 in Fig. 2c). The strength of natural selection is therefore the principal source of variation in the evolutionary dynamics between the hosts. These results suggest that the selective pressures in the gut of immune-compromised and WT mice are different.

Differences in host immune status are associated with differences in microbiota species composition: the microbiota of mice lacking T and B cells has been shown to differ from that in WT controls (Zhang et al., 2015). To investigate the potential role of the microbiota in driving the changes in adaptation rate observed in this study, we estimated s_{gatZ} by direct *in vivo* competition against the ancestral in (i) WT and *Rag2*$^{-/-}$ mice previously co-housed for one month, a procedure that leads to homogenization of the microbiota between different individuals (Elinav et al., 2013) (Fig. 3a; Supplementary Table 2), and (ii) germ free (GF) WT and *Rag2*$^{-/-}$ mice devoid of microbiota (Fig. 3b; Supplementary Table 3). Analyzing the full dataset (Fig 3c) including competitions in independently housed WT and *Rag2*$^{-/-}$ animals, revealed a significant overall interaction between host immune status and the microbiota (ANOVA; $F_{(2,50)}=9.64$; $P<10^{-3}$), upon the fitness effect of the emerging mutation. In WT animals, co-housing slightly (s_{gatZ} = 0.05 ± 0.01) but not significantly (ANOVA with Tukey's *post hoc* test, $P=0.3$, Fig. 3c) decreased the mutant fitness (Fig. 3a, left panel). This advantage was similar to that measured in co-housed *Rag2*$^{-/-}$ (s_{gatZ} = 0.050±0.009, ANOVA with Tukey's *post hoc* test, $P>0.99$) which was higher than in independently housed *Rag2*$^{-/-}$ (ANOVA with Tukey's *post hoc* test, $P=0.02$). Another marked effect of the co-housing protocol was a reduction in the variance for the fitness effect of *gatZ* in *Rag2*$^{-/-}$ mice, which became more similar to that observed in WT animals (F-test, $F= 1.32$, $P=0.67$), indicating loss of antagonistic pleiotropy.

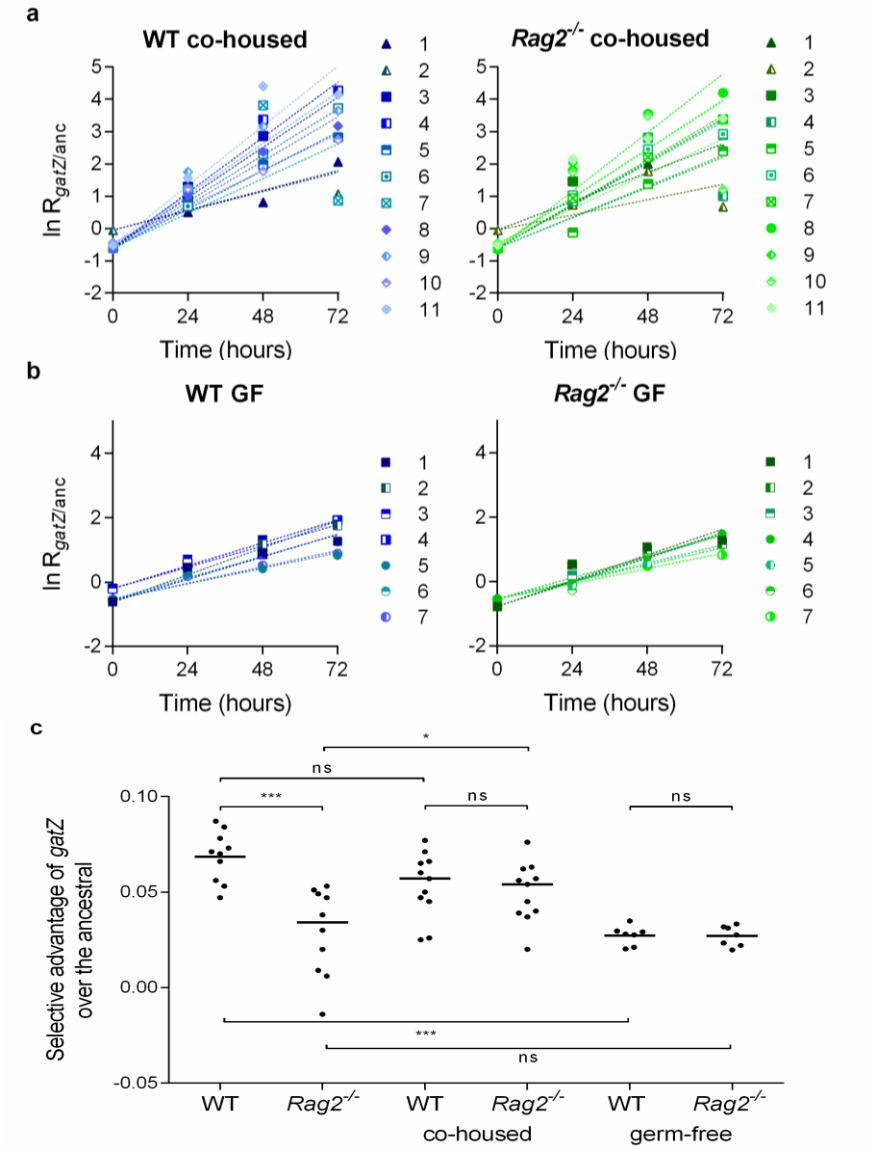


Figure 3. The microbiota influences the selective advantage of adaptive mutations.
Competitive fitness experiments of the emerging *gatZ* allele against the ancestral strain in (**a**) WT co-housed with *Rag2*$^{-/-}$ mice (left panel, n=11) and *Rag2*$^{-/-}$ co-housed with WT mice (right panel, n=11) and in (**b**) GF WT (left panel, n=7) and GF *Rag2*$^{-/-}$ (right panel, n=7) mice. In **a**, mice co-housed in the same group are represented with the same shape (triangles, squares or diamonds). Advantage of *gatZ* over the ancestral, per hour was calculated as in Fig.2c. (**c**) Selective advantage of the *gatZ* mutant, per hour, inferred from the slope of the linear regression of ln(*gatZ*/anc), over 3 days of *in vivo* competition against the ancestral in mice either WT, *Rag2*$^{-/-}$, WT co-housed, *Rag2*$^{-/-}$ co-housed, GF WT and GF *Rag2*$^{-/-}$. The competitive advantages differed between *Rag2*$^{-/-}$ and WT (ANOVA with Tukey's *post hoc* test, $P<10^{-3}$), but upon co-housing the advantage in *Rag2*$^{-/-}$ mice (co-housed) increased (ANOVA with Tukey's *post hoc* test, $P=0.02$) and became similar to WT animals (co-housed) (ANOVA with Tukey's *post hoc* test, $P>0.99$). The advantage of the WT did not significantly change upon co-housing (ANOVA with Tukey's *post hoc* test, $P=0.3$). In the absence of microbiota (GF) the selective effect of *gatZ* is similar in WT (circles)

and *Rag2*$^{-/-}$ (triangles) mice (ANOVA with Tukey's *post hoc* test, P>0.99) and smaller than in WT mice harboring microbiota (ANOVA with Tukey's *post hoc* test, P<10^{-3}). ns (not significant), $P > 0.05$, *$P < 0.05$, ***$P < 0.001$.

Analysis of GF animals provided further evidence of a major role for the microbiota in shaping the selective pressure on the *gatZ* mutant. First, s_{gatZ} was smaller in GF animals than in microbiota-bearing WT animals (ANOVA with Tukey's *post hoc* test, P<10^{-3}, Fig. 3c), even though a shorter duplication time was found for *E. coli* in the former (see Supplementary Fig. 4). Second, both the mean and the variance for s_{gatZ} were similar between GF WT and *Rag2*$^{-/-}$ mice (ANOVA with Tukey's *post hoc* test, $P>0.99$, Fig. 3c; and F-test, $F=0.97$, $P=0.97$). Finally, irrespective of the functional state of the immune system, the variance for s_{gatZ} decreased dramatically in GF compared with microbiota-harboring animals (F-test, $P<10^{-3}$). Globally, these results confirm the microbiota as a major player modulating the selective effect of beneficial mutations.

When competing two single *gat* mutants, one carrying an IS insertion in *gatZ* (previously used for the competitions against the ancestral) and the other a SNP in *gatC* (see Methods), in independently housed (Fig. 4a; Supplementary Table 4) or co-housed (Fig. 4b; Supplementary Table 5) WT and *Rag2*$^{-/-}$ mice, similar results were obtained. Again, we found a significant interaction between the microbiota and the presence of an adaptive immune system (ANOVA; $F_{(2,31)}=12.9$; $P=0.001$). As expected, in WT mice (Fig. 4a, left panel) the mutants were almost neutral ($s_{gatZvsgatC}$ = 0.01±0.004). However in *Rag2*$^{-/-}$ mice (Fig. 4a, right panel) $s_{gatZvsgatC}$ was on average smaller (-0.03±0.02, ANOVA with Tukey's *post hoc* test, $P<10^{-3}$, Fig. 4c). Extensive variability for the fitness effects was observed, with an extreme case in which *gatZ* was found to have a strongly deleterious effect (Fig. 4a, *Rag2*$^{-/-}$ mouse 3). On the other hand, when performing the same competitions in co-housed mice (Fig. 4b), the average $s_{gatZvsgatC}$ in *Rag2*$^{-/-}$ mice increased (ANOVA with Tukey's *post hoc* test, $P<10^{-3}$, Fig. 4c), approximating the value observed in co-housed WT ($s_{gatZvsgatC}$ = 0.02±0.01, ANOVA with Tukey's *post hoc* test $P>0.99$, Fig. 4c). Finally, we also detected a significant increase in the variance for s_{gatZ} when competed against *gatC* in independently housed *Rag2*$^{-/-}$ compared with WT mice (F-test, $F=0.04$, $P<10^{-3}$).

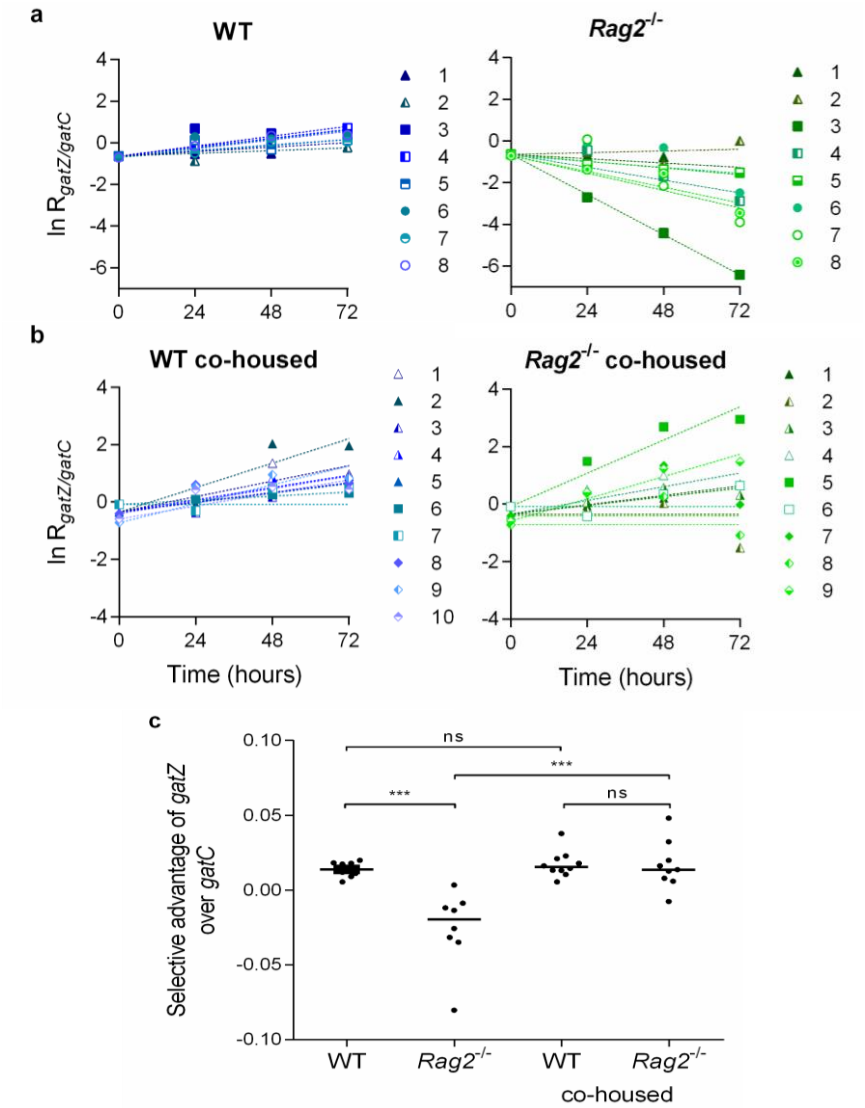


Figure 4. Evidence for antagonistic pleiotropy in immune-compromised mice.
Selective effect of *gatZ* allele relative to *gatC* allele in (**a**) WT (left panel) and *Rag2*$^{-/-}$ (right panel), (**b**) WT co-housed with *Rag2*$^{-/-}$ mice (left panel) and *Rag2*$^{-/-}$ co-housed with WT mice (right panel). Competitions were performed in independently housed WT and *Rag2*$^{-/-}$ mice (n=8) or in mice co-housed for one month (n=9). In **a**, mice from the same litter are represented by the same shape symbols (triangles, squares or circles) and in **b**, mice co-housed in the same group are represented with the same shape (triangles, squares or diamonds). (**c**) Comparison of the selective effects of *gatZ* over *gatC* inferred from the slope of the linear regression of ln(*gatZ*/*gatC*), over 3 days of *in vivo* competition in WT, *Rag2*$^{-/-}$, WT co-housed and *Rag2*$^{-/-}$ co-housed. The mean relative fitness effect differs between *Rag2*$^{-/-}$ and WT (ANOVA with Tukey's *post hoc* test, $P<10^{-3}$), but upon co-housing no significant difference was detected between animals (ANOVA with Tukey's *post hoc* test P>0.99). Co-housing with WT results in an increase in the mean selective effect of *gatZ* vs *gatC* in *Rag2*$^{-/-}$ mice (ANOVA with Tukey's *post hoc* test, $P<10^{-3}$) but not in WT mice (ANOVA with Tukey's *post hoc* test P>0.99). ns (not significant), $P > 0.05$, ***$P < 0.001$.

Our combined fitness data provide strong evidence for antagonistic pleiotropy specific to immune-compromised animals, a feature that reduces the predictability of *E. coli* evolution in these hosts. Moreover, our results indicate that the host, the microbiota and their interactions influence the pace of adaptive evolution of *E. coli* to the mouse gut. Together, these findings suggest that a microbiota shaped by the adaptive immune system provides stronger selective pressures upon an individual bacterial species residing within this environment than that which arises in the absence of this host component.

Microbiota characterization of WT and *Rag2*$^{-/-}$ mice

Recent studies have reported differences in microbiota composition between immune-competent and immune-compromised animals (Dimitriu et al., 2013; Kawamoto et al., 2014; Scholz et al., 2014; Shulzhenko et al., 2011; Suzuki et al., 2004; Thoene-Reineke et al., 2014; Zhang et al., 2015). To determine whether changes in the microbiota within WT and *Rag2*$^{-/-}$ raised in our facility were associated with the slower rate and increased variation observed in *E. coli* adaptation in immune-compromised animals, we assessed the composition of this community following high-throughput 16S rRNA gene sequencing of DNA extracted from fecal samples of unmanipulated animals (Supplementary Fig. 5) and from day 3 of the evolution experiment (Fig. 5).

In unmanipulated hosts (n=5 for each, see Methods) the phylogenetic diversity was significantly higher in WT when compared with *Rag2*$^{-/-}$ animals (Mann-Whitney *U*-test, W=23.5, P=0.03, Supplementary Fig. 5a). Consistent with this result, PCoA on the phylogeny-based distance of community membership showed a significant difference between genotypes (AMOVA on unweighted UniFrac distance, P=0.003, Supplementary Fig. 5b). However, at the genus level, difference in abundance was only observed for some low-frequency genera, albeit not statistically significant (Supplementary Fig. 5c). Similarly, OTU-based measures of community richness and diversity revealed only a trend, not significant, for decreased community diversity in *Rag2*$^{-/-}$ (Mann-Whitney *U*-test, W=17, P=0.4 and W=21, P=0.1, respectively, Supplementary Fig. 5d, e). PCoA analysis of OTU- and phylogeny-based distances of community structure revealed

no significant differences (AMOVA on Bray-Curtis distance, *P*=0.1 and weighted UniFrac distance, *P*=0.4, Supplementary Fig. 5e, f). Importantly, reads identified as *Escherichia/Shigella* could be detected at very small and similar frequencies (maximum 0.03%) in both *Rag2*$^{-/-}$ and WT animals (Mann-Whitney *U*-test, *W*=8.5, *P*=0.46) suggesting that direct competition between our experimental *E. coli* and native *Escherichia*, if any, would be similar in both hosts.

We then analyzed 10 WT and 10 *Rag2*$^{-/-}$ animals from day 3 of the evolution experiment, that is, treated with streptomycin and colonized with *E. coli*. The gut microbiota was dominated at the genus level by a few bacterial groups (Fig. 5a), compatible with the strong effect of this treatment, as shown before (Thompson et al., 2015). In this setting, we identified genera differentially represented in WT and *Rag2*$^{-/-}$ mice (Fig. 5b). Most notably *Barnesiella, Allobaculum* and an unclassified genus of Clostridiales were more abundant in *Rag2*$^{-/-}$ mice, while *Bacteroides*, and *Paenibacillus* were more prevalent in WT animals. In line with the variance observed when monitoring the *gat* mutant fitness, we detected increased variance in *Rag2*$^{-/-}$ mice for these genera, particularly that of *Barnesiella* (F-test, *F*=16.3, $P<10^{-3}$), for which the relative frequencies ranged between 2 and 74%, while in WT hosts the range was from 0 to 17%. In mice where the rate of increase of the *gat*-negative phenotype (as a proxy for rate of adaptation) was the highest (R1.10) or the lowest (R1.11) (see Supplementary Fig. 2), the microbiota was composed of a small (2%) or large (51%) frequency of *Barnesiella*, respectively.

Although the richness of microbial communities was similar between streptomycin-treated WT and *Rag2*$^{-/-}$ animals (Mann-Whitney *U*-test, *W*=39, *P*=0.4, Fig. 5c), both OTU-based and phylogenetic diversity were increased in the latter (Mann-Whitney *U*-test, *W*=20, *P*=0.03 and *W*=17, *P*=0.01, Fig. 5d and e). Strikingly, when comparing the beta diversity metrics of treated WT and *Rag2*$^{-/-}$ animals, PCoA analysis of Bray-Curtis (OTU-based) and weighted UniFrac (phylogeny-based) distances revealed that the structure of microbial communities significantly differed between the two genotypes (AMOVA on Bray-Curtis distance, *P*=0.01 and weighted UniFrac distance, *P*=0.03, Fig. 5f, g). However, only marginally significant differences were found for the phylogenetic membership (AMOVA on the unweighted UniFrac distance, *P*=0.08, Supplementary Fig. 6).

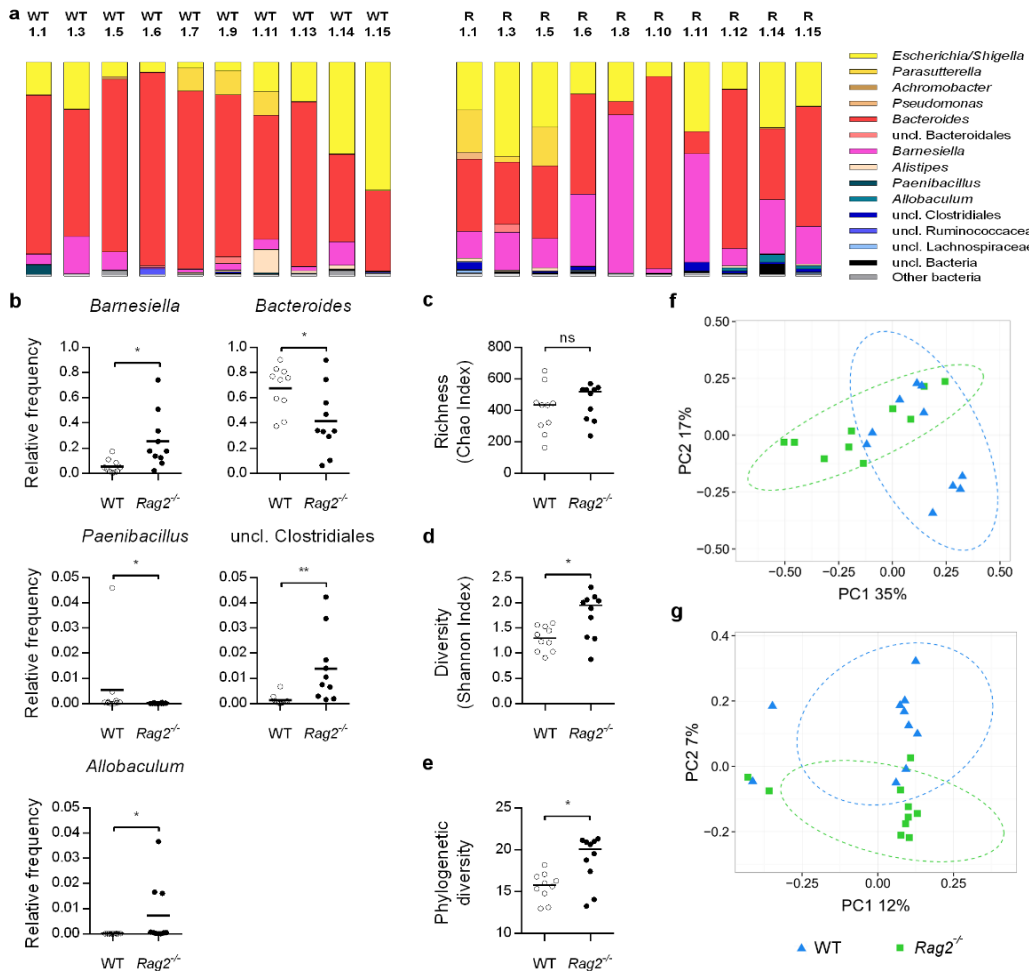


Figure 5. Microbiota composition differs between immune-competent and immune-compromised hosts

(**a**) Microbiota composition at the genus level of 10 WT and 10 *Rag2*$^{-/-}$ mice from day 3 of the evolution experiment (streptomycin-treated and colonized with *E. coli*). Genera with a relative abundance larger than 1% are displayed. The colored segments represent the relative frequency of each genus. (**b**) Relative frequencies of genera differentially represented in WT and *Rag2*$^{-/-}$ animals: *Barnesiella* (Mann-Whitney *U*-test with the Benjamini and Hochberg correction, W=11, P=0.021), *Bacteroides* (Mann-Whitney *U*-test with the Benjamini and Hochberg correction, W=18, P=0.049), *Paenibacillus* (Mann-Whitney *U*-test with the Benjamini and Hochberg correction, W=14.5, P=0.029), unclassified Clostridiales (Mann-Whitney *U*-test with the Benjamini and Hochberg correction, W=6, P=0.0065) and *Allobaculum* (Mann-Whitney *U*-test with the Benjamini and Hochberg correction, W=13, P=0.025). *P < 0.05, **P < 0.01. Alpha diversity estimates of microbiota community richness (**c**), diversity (**d**) and phylogenetic diversity (**e**) in WT and *Rag2*$^{-/-}$ animals. Both OTU-based diversity and phylogenetic diversity are increased in *Rag2*$^{-/-}$ animals compared with WT (Mann-Whitney *U*-test, W=20, P=0.03 and W=17, P=0.01, respectively). NS (not significant) P > 0.05, *P < 0.05. Principal coordinates analysis (PCoA) of the Bray-Curtis (**f**) and weighted UniFrac (**g**) distance matrices of fecal microbiota of WT and *Rag2*$^{-/-}$ mice. The first two coordinates are shown. Ellipses centered on the averages of the metric distances with a 90% confidence interval for the first two coordinates were drawn on the associated PCoA.

These results indicate that the microbiota ecosystem experienced by *E. coli* in our experimental setting differs between WT and *Rag2*$^{-/-}$ hosts and between *Rag2*$^{-/-}$ individuals, a feature strikingly reminiscent of the *gat* mutant fitness.

Genetic targets of adaptation in WT and *Rag2*$^{-/-}$ mice

We and others have previously shown that over a 3-week period of evolution, more than one genetic target can be involved in the genetics of *E. coli* adaptation to the mouse gut (Barroso-Batista et al., 2014; Leatham et al., 2005; Leatham-Jensen et al., 2012). We therefore thought to determine whether the haplotypic structure of adaptation in *Rag2*$^{-/-}$ mice would be similar to that of WT animals (Barroso-Batista et al., 2014). We sampled clones (~20) from independent *E. coli* populations recovered from each *Rag2*$^{-/-}$ mouse after 24 days of evolution. Clones were typed for the presence of mutations previously found to be under selection in the gut of WT mice (Barroso-Batista et al., 2014), namely: IS insertions in the intergenic regions of *focA/ycaO* (formate transporter), *dcuB/dcuR* (fumarate transporter) and SNPs in *srlR* (regulator of the sorbitol metabolism).

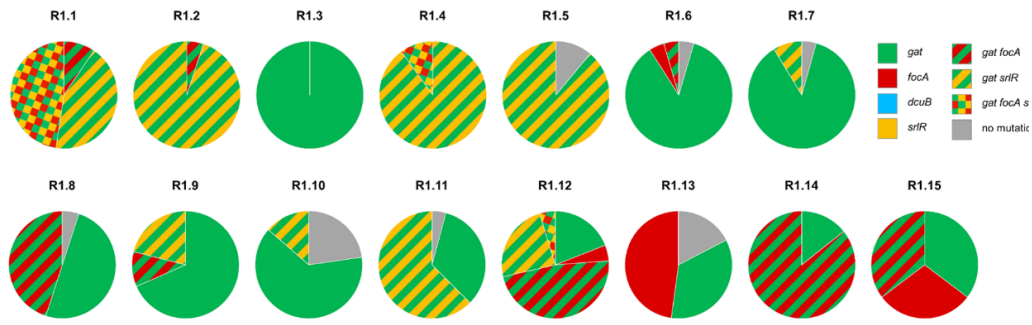


Figure 6. Host genetics influences the genetic basis of *E. coli* adaptation in the gut. Frequencies of haplotypes in *E. coli* populations evolved in the gut of *Rag2*$^{-/-}$ mice for 24 days. Clones (n~20) isolated from the last day of evolution of lineages R1.1 to R1.15 (Fig. 1b) were screened for the presence of mutations previously identified in WT animals: *gat* phenotype (green), insertions in the regulatory region of *focA* (red) and *dcuB* (blue) and SNPs in the coding region of *srlR* (orange). Clones in which none of these mutations were found are shown in grey. Clonal interference is detected by the competition between clones carrying different beneficial mutations (stripes of different colors represent double mutants and clones carrying three mutations are shown as a checkered pattern).

In *Rag2*$^{-/-}$ animals (Fig. 6), IS insertions near *focA/ycaO* as well as SNPs in the *srlR* gene were common (found in 10 and 9 out of 15 mice, respectively). These mutations were segregating at different frequencies within each host: 10 to

90% for *focA/ycaO* and 10 to 100% for *srlR* mutations. Strikingly insertions near *dcuB/dcuR* were not found in the *E. coli* that had evolved in immune-deficient animals (Fig. 6), although this target was repeatedly identified in clones sampled from WT mice (Barroso-Batista et al., 2014).

The clear occurrence of clonal interference - where clones carrying distinct haplotypes competing for fixation within a population - among *E. coli* evolving in both *Rag2*$^{-/-}$ (Fig. 6) and WT mice (Barroso-Batista et al., 2014) together with the similar rate of spontaneous transposition estimated in these hosts (Fig. 2b), suggests that mutation is non-limiting in the gut of both animals. In turn, this property argues for insertions near *dcuB* to be beneficial in WT, but neutral or even slightly deleterious in *Rag2*$^{-/-}$. To test this hypothesis and identify other targets of adaptation we performed whole genome sequencing (WGS) of large population samples (> 1000 clones) that had evolved in each animal. This approach allows the identification of mutations segregating at high frequency, and thus likely to be beneficial, but not those present at low frequency. As expected, WGS (Table 1; Supplementary Table 6 and 7) revealed the *gat* operon, the *srlR* locus and the intergenic region of *focA/ycaO* as the main targets of adaptation in both hosts. Remarkably, *dcuB* was hit in 7/14 WT and in none of the 15 *Rag2*$^{-/-}$ population tested (binomial test, P=10^{-3}), confirming the beneficial effects of insertions near this gene in immune-sufficient and not in immune-deficient mice.

The WGS approach also revealed new *bona fide* targets for *E. coli* adaptation to the mouse gut (Table 1). These are targets that occurred in at least two *E. coli* populations evolving in mice living in allopatry. We detected IS insertions, deletions and nonsense mutations in the coding region of *kdgR*, a transcriptional repressor (Murray and Conway, 2005) regulating the metabolism of sugar acids (Peekhaus and Conway, 1998), and insertions of IS elements in the intergenic region of *yjjP* and *yjjQ*, that code, respectively for an predicted inner membrane protein and a putative transcription factor (Stratmann et al., 2008). We also found an additional target specific to WT mice: *yeaR*, encoding a protein induced in the presence of nitric compounds (Lin et al., 2007).

Table 1. Targets of adaptation differ in *E. coli* evolving in WT and *Rag2*$^{-/-}$ animals

Gene	Function	Frequency of mice	
		WT	*Rag2*$^{-/-}$
gat operon	Metabolism of galactitol	100% (±7%)	100% (±7%)
srlR	Metabolism of Sorbitol	50% (±13%)	60% (±13%)
focA/ycaO	Anaerobic respiration (formate transporter)	29% (±12%)	40% (±13%)
yjjP/yjjQ	Inner membrane protein	57% (±13%)	13% (±9%) ***
kdgR	Metabolism of sugar acids	14% (±9%)	13% (±9%)
dcuB/dcuR	Anaeobic respiration (fumarate, succinate transporter)	50% (±13%)	0% ***
yeaR	Metabolism of nitric compounds	14% (±9%)	0%
arcB	Regulation of respiration	0%	13% (±9%)
frlR	Metabolism of fructosamines	0%	13% (±9%)
rimJ	Modification of ribosomal proteins	0%	13% (±9%)

Frequency (±2 s.e.m) of WT (n=14) or *Rag2*$^{-/-}$ (n=15) populations in which parallel mutations at a given locus were found segregating at high frequency (>10%). These loci, identified through WGS of population samples, were observed in at least two populations and thus correspond to *bona fide* targets of beneficial mutations. Mutations whose emergence differs significantly between the two host genetic backgrounds are highlighted in bold, with the level of significance displayed (binomial test ***P<0.001).

Interestingly, we identified parallel mutational targets detected only in immune-compromised mice (*arcB*, *frlR* and *rimJ*). *arcB* encodes a transmembrane sensor kinase and together with *arcA* regulates the expression of the Arc modulon in response to changes in oxygen levels (Georgellis et al., 2001). *frlR* codes for a predicted regulator of the fructoselysine operon (Wiame et al., 2002) that is responsible for the metabolism of fructosamines. *rimJ* encodes an alanine acetyltransferase involved in modification of ribosomal proteins (Yoshikawa et al., 1987).

Discussion

We investigated the process of *E. coli* evolution in the gut of immune-compromised mice and compared it with the evolutionary patterns observed for the same ancestral strain in WT mice. We observed a difference in the initial

divergence of the neutral marker in immune-compromised compared with immune-competent animals, compatible with a slower rate of evolution in the former. This finding is further supported by a delayed emergence of the adaptive *gat* phenotype, whose dynamics displayed a slower pace and increased variability in its fitness effects amongst these hosts. In a system of intense clonal interference, as is the case here, the power of the two-marker system in detecting adaptive mutations becomes rapidly reduced. Consequently, the lack of differences in marker dynamics between hosts over longer periods is not interpretable.

Our findings of a higher growth rate of *E. coli* in *Rag2*$^{-/-}$ animals compared with WT is compatible with a weakened control of bacterial growth in immune-compromised hosts, suggesting a more benign environment for *E. coli*. The same explanation may also be valid for the germ-free mice, where even lower duplications times, regardless of the immune state, were observed. In agreement with this conjecture, the selective effects of the first sweeping beneficial mutations were smaller in immune-compromised hosts, indicating stronger selective pressures in WT mice.

The bulk of our work indicates that the adaptive immune system indirectly increases this selective pressure through the shaping of the microbiota, which in turn interacts with *E. coli*. First, alteration (co-housing) or removal (GF) of the microbiota modulates the selective effects, clearly revealing that this is a major factor controlling the selective pressures in the mouse gut. Second, we could indeed detect differences in the relative abundance of some genera, notably *Barnesiella* spp, between the two host genotypes. Bacteria belonging to this genus have been shown to confer protection against vancomycin-resistant *Enterococcus* (VRE), through some undefined direct or indirect mechanism (Ubeda et al., 2013). Our results obtained upon antibiotic treatment in *Rag2*$^{-/-}$ animals, not only show an increase in the mean frequency of *Barnesiella* but also a substantial degree of inter-individual variation, supporting a role for the host adaptive immune system in maintaining coherence in the ecology and evolution of the commensal microbes.

The influence of the host immune system, in particular the adaptive immune system, on the shaping of the gut microbiota has been the focus of recent studies. Among these, the microbiota composition of WT and *Rag1*$^{-/-}$ or *Rag2*$^{-/-}$ mice (both

Rag1 and *Rag2* null mutations result in total ablation of the adaptive immune system) has been compared using 16S analysis (Dimitriu et al., 2013; Scholz et al., 2014; Thoene-Reineke et al., 2014; Zhang et al., 2015). While each study reports differences in the representation of several microbial groups, these are not the same across animal facilities. For example, one work revealed an expansion of anaerobes, especially SFB, in the gut of *Rag2*$^{-/-}$ mice (Suzuki et al., 2004), while others described differences in the abundance of Bifidobacteria and *Clostridium leptum* (Thoene-Reineke et al., 2014), or in the phylogenetic community membership such as increased Lachnospiraceae and decreased Porphyromonadacea in *Rag1*$^{-/-}$ mice (Scholz et al., 2014). None of these specific differences were noticeable in our colonies. In addition a previous study reported an increased frequency of *Akkermansia* sp in *Rag2*$^{-/-}$ hosts when compared with WT controls (Zhang et al., 2015). While this genus was represented at notable frequency (5%) in one unmanipulated *Rag2*$^{-/-}$ animal in our study, it was undetectable in the streptomycin-treated mice analyzed here, irrespectively of their genotype. However, we detected differences in phylogenetic memberships (unweighted UniFrac) between unmanipulated WT and *Rag2*$^{-/-}$ animals, confirming that the two genotypes do carry different microbial gut compositions. Moreover, our observation of reduced phylogenetic diversity in unmanipulated *Rag2*$^{-/-}$ corroborates other works, where immune-compromised mice, including mutants lacking either or both T and B cells display decreased microbial diversity (Kawamoto et al., 2014). A similar finding was reported recently, where the abundance and diversity of B cell-secreted immunoglobulin-A was correlated with the diversity of the gut microbiota (Fransen et al., 2015).

More directly relevant to the scope of this work, we found differences in both phylogeny- and OTU-based measures of alpha diversity between streptomycin-treated WT and *Rag2*$^{-/-}$ mice. Together with our findings of distinct community structures (weighted UniFrac and Bray-Curtis distances) between the two genotypes, these results suggest that the pre-existing differences in microbial composition between WT and *Rag2*$^{-/-}$ are exacerbated upon antibiotic treatment. Overall, and in agreement with the majority of previous works, our results indicate that microbial composition differs between immune-competent and -compromised

animals, supporting the hypothesis that adaptive immunity plays an important role in the shaping of the microbiota.

The genetic basis of *E. coli* adaptation determined by WGS of evolved populations revealed that most of the mutations were related with bacterial metabolism and respiration, indicating that the main selective pressure was adaptation to the metabolic environment of the intestine. As the microbiota provides key metabolites (Faith et al., 2014) and given our findings of different microbial profiles in WT and *Rag2*$^{-/-}$ mice, it is possible that the gut environment of these hosts differs in the availability of different metabolites. In turn, mutations found exclusively or mainly in one host may reflect the specific metabolic composition of that environment. One example, and the most extreme case, is the mutational target *dcuB* found in half of the WT populations but absent in *Rag2*$^{-/-}$ mice. This gene codes for a transporter of C4-dicarboxylic acids such as fumarate and succinate (Engel et al., 1994). In the mouse gut, succinate can be produced by *Bacteroides* (Ferreyra et al., 2014), a genus that our microbiota analysis revealed to be present at higher relative abundance in WT than in *Rag2*$^{-/-}$ mice. It is possible that differences in the availability of succinate, produced by the microbiota, may have driven the differential selection for mutations in genes related to the uptake and metabolism of this compound in the two groups. The same may be true to *yeaR*, which codes for a protein induced in the presence of nitric compounds (Lin et al., 2007), that can be used by *E. coli* to perform anaerobic respiration in the intestine of the host (Jones et al., 2011). Similarly, mutations in *frlR*, found in *Rag2*$^{-/-}$ mice, may be related to the levels of fructosamines, compounds that are present in the gut and likely metabolized by bacteria (Wiame et al., 2002). The mutations identified may lead to inactivation of *frlR* and consequently, to the constitutive expression of the *frl* operon, conferring an advantage for *E. coli* in the gut when nutrients are available at low-concentration. Another example is *kdgR*, a transcriptional repressor of the Entner-Doudoroff aldolase-coding *eda* gene (Murray and Conway, 2005). Eda is a key enzyme in the Entner-Doudoroff pathway (Peekhaus and Conway, 1998) responsible for the metabolism of sugar acids, and mutants of this pathway are known to display impaired growth in the mouse gut (Chang et al., 2004; Leatham

et al., 2005; Sweeney et al., 1996), which is consistent with our finding of selection for mutations inactivating *kdgR*.

Overall, our data reveal that the path and rate of adaptation of commensal bacteria is influenced by the presence of a healthy adaptive immune system, providing experimental evidence to the first corollary of the co-evolution hypothesis between commensal microbiota and the host immune system. Our findings further suggest that the shaping of the gut microbiota composition by the adaptive immune system results in a coherent environment under which *E. coli* evolves along a stronger selective pressure and a more predictable path.

Acknowledgements

We thank A. Sousa and J. Sousa for help in the sequence analysis and continuous discussions along the work, R. Ramiro and N. Martins for help in the statistical analysis, R. Oliveira and R. Areal for help with the microbiota analysis and V. Barreto, J. Thompson and Gordo's lab members for critically reading the manuscript. The research leading to these results has received funding from the European Research Council under the European Community' Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement n^{o} 260421–ECOADAPT and from the Portuguese Science Foundation FCT (PTDC/BIA-EVF/118075/2010). We thank the IGC animals house Unit and the gnotobiology service partially funded through FCT grant RECI/IMI-IMU/0038/2012 and EU-FP7 infrastructure grants EMMA and InfrafrontierI3. I.G. acknowledges salary support of LAO/ITQB & FCT. J.B.B. was supported by grant SFRH/BD/80257/2011 from FCT. Genome sequencing and 16S rRNA gene sequencing data have been deposited in the NCBI Read Archive database with accession code PRJNA297801.

References

Atarashi, K., Tanoue, T., Shima, T., Imaoka, A., Kuwahara, T., Momose, Y., Cheng, G., Yamasaki, S., Saito, T., Ohba, Y., et al. (2011). Induction of Colonic Regulatory T Cells by Indigenous Clostridium Species. Science *331*, 337–341.

Barroso-Batista, J., Sousa, A., Lourenço, M., Bergman, M.-L., Sobral, D., Demengeot, J., Xavier, K.B., and Gordo, I. (2014). The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps. PLoS Genet. *10*, e1004182.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. (1997). The complete genome sequence of Escherichia coli K-12. Science *277*, 1453–1462.

Bray, J.R., and Curtis, J.T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecol. Monogr. *27*, 326–349.

Cammarota, G., Ianiro, G., and Gasbarrini, A. (2014). Fecal microbiota transplantation for the treatment of Clostridium difficile infection: a systematic review. J. Clin. Gastroenterol. *48*, 693–702.

Chang, D.E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., Anderson, A.B., Grissom, J.E., Laux, D.C., Cohen, P.S., et al. (2004). Carbon nutrition of Escherichia coli in the mouse intestine. Proc Natl Acad Sci USA *101*, 7427–7432.

Chao, A. (1984). Nonparametric Estimation of the Number of Classes in a Population. Scand. J. Stat. *11*, 265–270.

Charlesworth, B. (2010). Elements of Evolutionary Genetics (Roberts and Company Publishers).

Dimitriu, P.A., Boyce, G., Samarakoon, A., Hartmann, M., Johnson, P., and Mohn, W.W. (2013). Temporal stability of the mouse gut microbiota in relation to innate and adaptive immunity. Environ. Microbiol. Rep. *5*, 200–210.

Elinav, E., Thaiss, C.A., and Flavell, R.A. (2013). Analysis of microbiota alterations in inflammasome-deficient mice. Methods Mol. Biol. Clifton NJ *1040*, 185–194.

Engel, P., Krämer, R., and Unden, G. (1994). Transport of C4-dicarboxylates by anaerobically grown Escherichia coli. Energetics and mechanism of exchange, uptake and efflux. Eur. J. Biochem. *222*, 605–614.

Faith, J.J., Ahern, P.P., Ridaura, V.K., Cheng, J., and Gordon, J.I. (2014). Identifying gut microbe-host phenotype relationships using combinatorial communities in gnotobiotic mice. Sci. Transl. Med. *6*, 220ra11.

Ferreyra, J.A., Wu, K.J., Hryckowian, A.J., Bouley, D.M., Weimer, B.C., and Sonnenburg, J.L. (2014). Gut microbiota-produced succinate promotes C. difficile infection after antibiotic treatment or motility disturbance. Cell Host Microbe *16*, 770–777.

Fransen, F., Zagato, E., Mazzini, E., Fosso, B., Manzari, C., El Aidy, S., Chiavelli, A., D'Erchia, A.M., Sethi, M.K., Pabst, O., et al. (2015). BALB/c and C57BL/6 Mice Differ in Polyreactive IgA Abundance, which Impacts the Generation of Antigen-Specific IgA and Microbiota Diversity. Immunity *43*, 527–540.

Georgellis, D., Kwon, O., and Lin, E.C.C. (2001). Quinones as the Redox Signal for the Arc Two-Component System of Bacteria. Science *292*, 2314–2316.

Ivanov, I.I., Atarashi, K., Manel, N., Brodie, E.L., Shima, T., Karaoz, U., Wei, D., Goldfarb, K.C., Santee, C.A., Lynch, S.V., et al. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell *139*, 485–498.

Jones, S.A., Gibson, T., Maltby, R.C., Chowdhury, F.Z., Stewart, V., Cohen, P.S., and Conway, T. (2011). Anaerobic respiration of Escherichia coli in the mouse intestine. Infect. Immun. *79*, 4218–4226.

Kawamoto, S., Maruya, M., Kato, L.M., Suda, W., Atarashi, K., Doi, Y., Tsutsui, Y., Qin, H., Honda, K., Okada, T., et al. (2014). Foxp3(+) T cells regulate

immunoglobulin A selection and facilitate diversification of bacterial species responsible for immune homeostasis. Immunity *41*, 152–165.

Leatham, M.P., Stevenson, S.J., Gauger, E.J., Krogfelt, K.A., Lins, J.J., Haddock, T.L., Autieri, S.M., Conway, T., and Cohen, P.S. (2005). Mouse Intestine Selects Nonmotile flhDC Mutants of Escherichia coli MG1655 with Increased Colonizing Ability and Better Utilization of Carbon Sources. Infect Immun *73*, 8039–8049.

Leatham-Jensen, M.P., Frimodt-Møller, J., Adediran, J., Mokszycki, M.E., Banner, M.E., Caughron, J.E., Krogfelt, K.A., Conway, T., and Cohen, P.S. (2012). The Streptomycin-Treated Mouse Intestine Selects Escherichia coli envZ Missense Mutants That Interact with Dense and Diverse Intestinal Microbiota. Infect Immun *80*, 1716–1727.

Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R., and Gordon, J.I. (2008). Worlds within worlds: evolution of the vertebrate gut microbiota. Nat. Rev. Microbiol. *6*, 776–788.

Lin, H.-Y., Bledsoe, P.J., and Stewart, V. (2007). Activation of yeaR-yoaG Operon Transcription by the Nitrate-Responsive Regulator NarL Is Independent of Oxygen- Responsive Regulator Fnr in Escherichia coli K-12. J. Bacteriol. *189*, 7539–7548.

Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., and Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. ISME J. *5*, 169–172.

Lozupone, C.A., Hamady, M., Kelley, S.T., and Knight, R. (2007). Quantitative and Qualitative β Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. Appl. Environ. Microbiol. *73*, 1576–1585.

Mukherjee, P.K., Sendid, B., Hoarau, G., Colombel, J.-F., Poulain, D., and Ghannoum, M.A. (2015). Mycobiota in gastrointestinal diseases. Nat. Rev. Gastroenterol. Hepatol. *12*, 77–87.

Murray, E.L., and Conway, T. (2005). Multiple Regulators Control Expression of the Entner-Doudoroff Aldolase (Eda) of Escherichia coli. J. Bacteriol. *187*, 991–1000.

Peekhaus, N., and Conway, T. (1998). What's for dinner?: Entner-Doudoroff metabolism in Escherichia coli. J. Bacteriol. *180*, 3495–3502.

Poulsen, L.K., Licht, T.R., Rang, C., Krogfelt, K.A., and Molin, S. (1995). Physiological state of Escherichia coli BJ4 growing in the large intestines of streptomycin-treated mice. J Bacteriol *177*, 5840–5845.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. *35*, 7188–7196.

Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., Griffin, N.W., Lombard, V., Henrissat, B., Bain, J.R., et al. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science *341*, 1241214.

Round, J.L., and Mazmanian, S.K. (2010). Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. Proc. Natl. Acad. Sci. U. S. A. *107*, 12204–12209.

Sartor, R.B., and Mazmanian, S.K. (2012). Intestinal Microbes in Inflammatory Bowel Diseases. Am. J. Gastroenterol. Suppl. *1*, 15–21.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., et al. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Appl. Environ. Microbiol. *75*, 7537–7541.

Schloss, P.D., Gevers, D., and Westcott, S.L. (2011). Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. PLoS ONE *6*, e27310.

Scholz, F., Badgley, B.D., Sadowsky, M.J., and Kaplan, D.H. (2014). Immune mediated shaping of microflora community composition depends on barrier site. PloS One *9*, e84019.

Shannon, C.E. (1948). A Mathematical Theory of Communication. Bell Syst. Tech. J. *27*, 379–423.

Sheneman, L., Evans, J., and Foster, J.A. (2006). Clearcut: a fast implementation of relaxed neighbor joining. Bioinforma. Oxf. Engl. *22*, 2823–2824.

Shulzhenko, N., Morgun, A., Hsiao, W., Battle, M., Yao, M., Gavrilova, O., Orandle, M., Mayer, L., Macpherson, A.J., McCoy, K.D., et al. (2011). Crosstalk between B lymphocytes, microbiota and the intestinal epithelium governs immunity versus metabolism in the gut. Nat. Med. *17*, 1585–1593.

Skinner, S.O., Sepúlveda, L.A., Xu, H., and Golding, I. (2013). Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization. Nat. Protoc. *8*, 1100–1113.

Stratmann, T., Madhusudan, S., and Schnetz, K. (2008). Regulation of the yjjQ-bglJ Operon, Encoding LuxR-Type Transcription Factors, and the Divergent yjjP Gene by H-NS and LeuO. J. Bacteriol. *190*, 926–935.

Suzuki, K., Meek, B., Doi, Y., Muramatsu, M., Chiba, T., Honjo, T., and Fagarasan, S. (2004). Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. Proc. Natl. Acad. Sci. U. S. A. *101*, 1981–1986.

Sweeney, N.J., Laux, D.C., and Cohen, P.S. (1996). Escherichia coli F-18 and E. coli K-12 eda mutants do not colonize the streptomycin-treated mouse large intestine. Infect. Immun. *64*, 3504–3511.

T. Conway, K. A. Krogfelt, and P. S. Cohen (2004). The Life of Commensal Escherichia coli in the Mammalian Intestine. In EcoSal - Escherichia Coli and Salmonella: Cellular and Molecular Biology, R. Curtiss III, ed. (Washington, DC: ASM Press), p.

Thoene-Reineke, C., Fischer, A., Friese, C., Briesemeister, D., Göbel, U.B., Kammertoens, T., Bereswill, S., and Heimesaat, M.M. (2014). Composition of Intestinal Microbiota in Immune-Deficient Mice Kept in Three Different Housing Conditions. PLoS ONE *9*, e113406.

Thompson, J.A., Oliveira, R.A., Djukovic, A., Ubeda, C., and Xavier, K.B. (2015). Manipulation of the Quorum Sensing Signal AI-2 Affects the Antibiotic-Treated Gut Microbiota. Cell Rep. *10*, 1861–1871.

Ubeda, C., Bucci, V., Caballero, S., Djukovic, A., Toussaint, N.C., Equinda, M., Lipuma, L., Ling, L., Gobourne, A., No, D., et al. (2013). Intestinal Microbiota Containing Barnesiella Species Cures Vancomycin-Resistant Enterococcus faecium Colonization. Infect. Immun. *81*, 965–973.

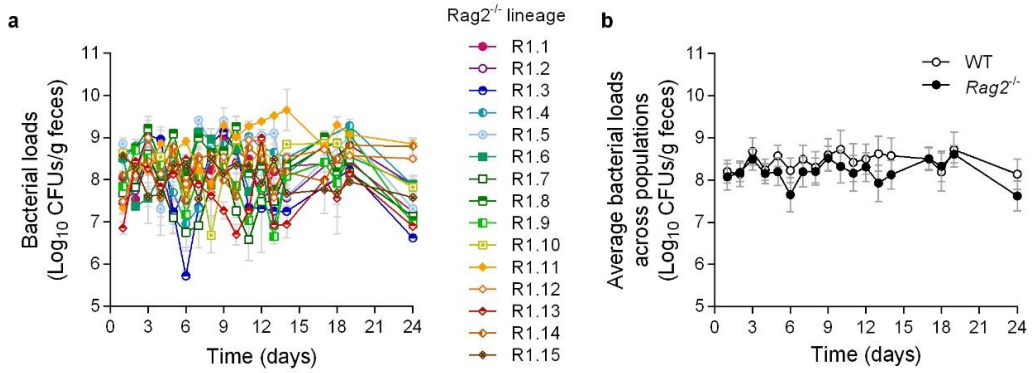
Whiteway, J., Koziarz, P., Veall, J., Sandhu, N., Kumar, P., Hoecher, B., and Lambert, I.B. (1998). Oxygen-Insensitive Nitroreductases: Analysis of the Roles of nfsA and nfsB in Development of Resistance to 5-Nitrofuran Derivatives in Escherichia coli. J. Bacteriol. *180*, 5529–5539.

Wiame, E., Delpierre, G., Collard, F., and Schaftingen, E.V. (2002). Identification of a Pathway for the Utilization of the Amadori Product Fructoselysine in Escherichia coli. J. Biol. Chem. *277*, 42523–42529.

Yoshikawa, A., Isono, S., Sheback, A., and Isono, K. (1987). Cloning and nucleotide sequencing of the genes rimI and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of Escherichia coli K12. Mol. Gen. Genet. MGG *209*, 481–488.

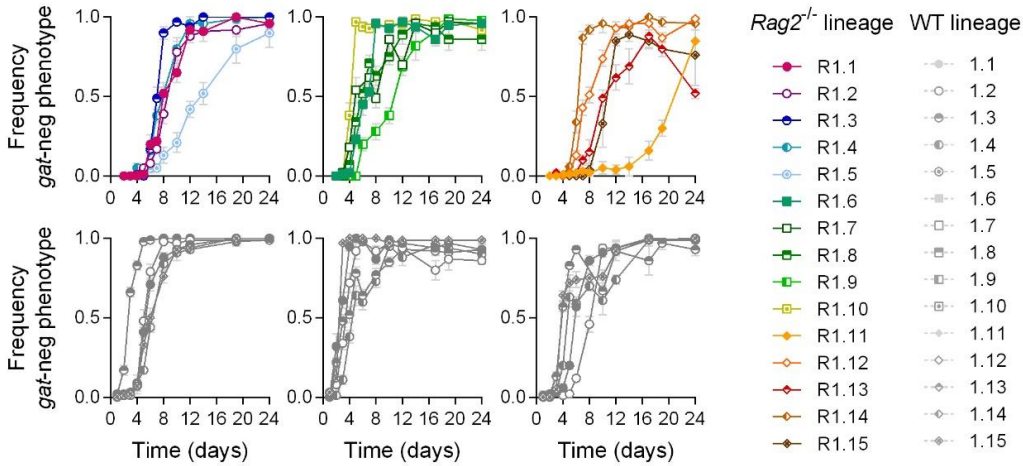
Zhang, H., Sparks, J.B., Karyala, S.V., Settlage, R., and Luo, X.M. (2015). Host adaptive immunity alters gut microbiota. ISME J. *9*, 770–781.

Supplementary Figures and Tables



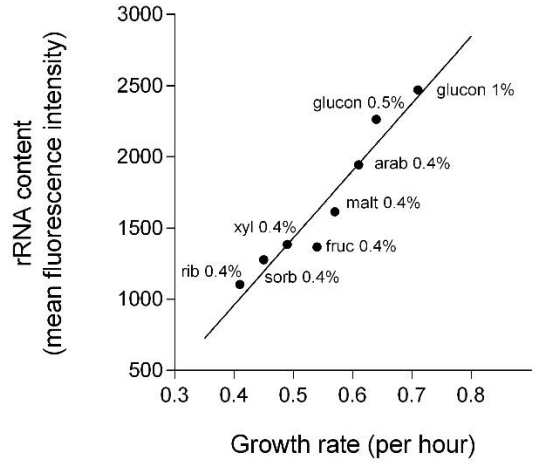
Supplementary Figure 1. Similar *E. coli* load in WT and *Rag2*$^{-/-}$ mice during the evolution experiments

a) *E. coli* colonization of the gut of *Rag2*$^{-/-}$ mice. Bacterial loads per gram of feces (±2 s.e.m) during 24 days of adaptation of *E. coli* to the mouse gut (populations R1.1 to R1.15). **b)** Average bacterial loads (±2 s.e.m) over time in WT (white circles) or *Rag2*$^{-/-}$ (black circles) populations. The host genotype does not influence the temporal dynamics of bacterial loads over time (interaction host genotype and time, ANOVA χ^2_{17}=18.64, p = 0.35) nor the bacterial loads (host genotype, ANOVA χ^2_1=3.35, p = 0.07).



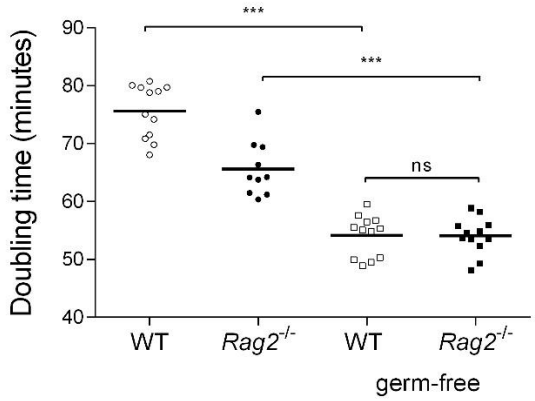
Supplementary Figure 2. Dynamics of the *gat*-negative phenotype

Emergence and expansion of beneficial mutations in the *gat* operon in *Rag2*$^{-/-}$ (upper panels) and WT (lower panels) mice, which cause the increase in frequency of a *gat*-negative phenotype. Each plot displays a group of 5 *Rag2*$^{-/-}$ (upper panels) or WT (lower panels) mice, corresponding to 3 independent experimental blocks.



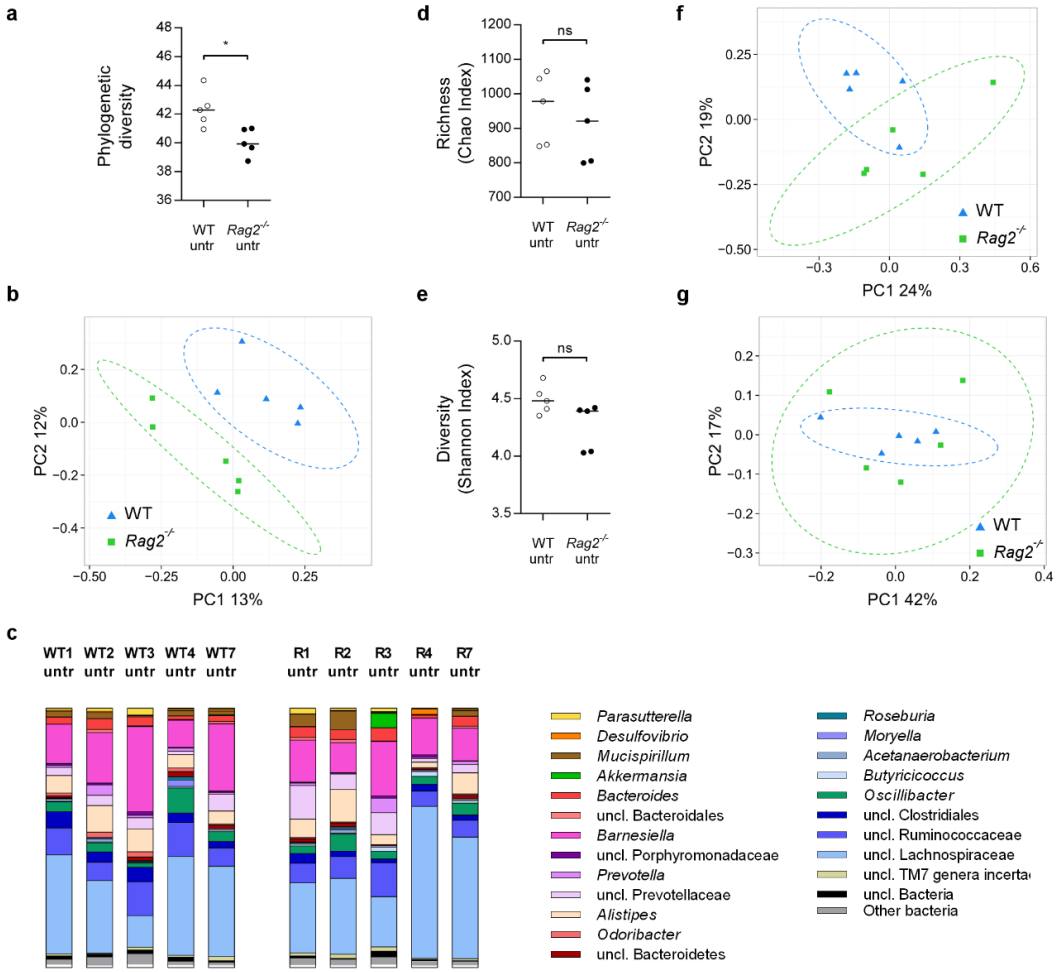
Supplementary Figure 3. Cellular content of ribosomes (cy3 intensity per cell) inferred by whole-cell hybridization

E. coli MG1655 strain DM09-CFP (ancestral) was grown in M9 minimal medium supplemented with (from left to right): ribose 0.4%, sorbitol 0.4%, xylose 0.4%, fructose 0.4%, maltose 0.4%, arabinose 0.4%, gluconate 0.5% and gluconate 1%. Circles represent the average of two hybridization measurements and four growth rates estimates (±2 s.e.m.)



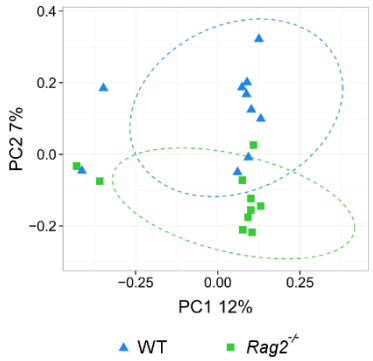
Supplementary Figure 4. Higher growth rate of *E. coli* in GF mice

The doubling time of *E. coli* colonizing the gut of WT or *Rag2*$^{-/-}$ is estimated by the amount of rRNA (see Supplementary Fig. 3). Each symbol corresponds to an average of 2 independent quantifications made for each mouse (n=6) during day 1 and 3 of colonization. The growth rate is significantly higher in GF compared to microbiota-harboring animals (WT vs WT GF, Mann-Whitney U test with Benjamini and Hochberg correction, W= 144, P<10^{-3} and *Rag2*$^{-/-}$ vs *Rag2*$^{-/-}$ GF, Mann-Whitney U test with Benjamini and Hochberg correction, W=120, P<10^{-3}).



Supplementary Figure 5. Microbiota composition of naïve (unmanipulated) WT and *Rag2*$^{-/-}$ mice.

a) Alpha diversity estimates of microbiota community phylogenetic diversity in WT and *Rag2*$^{-/-}$ animals. Phylogenetic diversity was increased in the WT mice (Mann–Whitney U test, W=23.5, P=0.03). *P < 0.05. **b)** Principal coordinates analysis (PCoA) of the unweighted UniFrac distance matrices of faecal microbiota of WT and *Rag2*$^{-/-}$ mice. The first two coordinates are shown. Ellipses centered on the averages of the metric distances with a 90% confidence interval for the first two coordinates were drawn on the associated PCoA. **c)** Microbiota composition at the genus level of 5 WT and 5 *Rag2*$^{-/-}$ mice before streptomycin treatment and *E. coli* colonization. Genera with a relative abundance larger than 1% are displayed. The colored segments represent the relative abundance of each genus. Alpha diversity estimates of microbiota community richness **(d)** and OTU-based diversity **(e)**. Both parameters were similar in *Rag2*$^{-/-}$ and WT animals (Mann–Whitney U test, W=17, P=0.4 and W=21, P=0.1, respectively). Ns (not significant,) P > 0.05. Principal coordinates analysis (PCoA) of the weighted UniFrac **(f)** or Bray Curtis **(g)** distance matrices of faecal microbiota of WT and *Rag2*$^{-/-}$ mice. The first two coordinates are shown. Ellipses centered on the averages of the metric distances with a 90% confidence interval for the first two coordinates were drawn on the associated PCoA.



Supplementary Figure 6. Analysis of the phylogenetic membership of WT and *Rag2*$^{-/-}$ mice from the day 3 of the evolution experiment (streptomycin-treated and colonized with *E. coli*).
Principal coordinates analysis (PCoA) of the unweighted UniFrac distance matrices of faecal microbiota of WT and *Rag2*$^{-/-}$ mice. The first two coordinates are shown. Ellipses centered on the averages of the metric distances with a 90% confidence interval for the first two coordinates were drawn on the associated PCoA.

Supplementary Table 1. *In vivo* competitions between a *gatZ* mutant and the ancestral strain in WT and *Rag2*$^{-/-}$ mice.
Table displaying the change in ln(*gatZ*/anc), i.e., the natural logarithm of the ratio between the frequencies of the *gatZ* mutant and the ancestral strain, over 3 days (72 hours) of competition in WT and *Rag2*$^{-/-}$ mice (see also Fig. 2c). ***s*** represents the selective advantage of *gatZ* along the three days of competition, inferred from the slope of the linear regression of ln(*gatZ*/anc).

Mice		**Time (hours)**				***s***
		0	**24**	**48**	**72**	
WT	**1**	-0.23	2.32	2.87	4.28	**0.066**
	2	-0.23	1.49	2.32	3.47	**0.053**
	3	-0.07	1.23	2.23	3.19	**0.047**
	4	-0.07	1.31	3.04	3.63	**0.056**
	5	-0.62	1.89	2.58	4.63	**0.073**
	6	-0.62	2.17	3.30	3.70	**0.070**
	7	-0.62	1.66	2.72	4.37	**0.071**
	8	-1.02	1.61	3.26	3.97	**0.078**
	9	-1.02	0.22	3.96	4.93	**0.087**
	10	-1.02	2.28	3.79	4.05	**0.084**
Rag2$^{-/-}$	**1**	-0.23	0.83	2.26	3.67	**0.053**
	2	-0.23	1.96	2.58	2.43	**0.047**
	3	-0.07	1.55	3.02	2.80	**0.049**
	4	-0.07	-0.97	-0.90	-0.80	**-0.014**
	5	-0.62	1.23	2.13	2.66	**0.051**
	6	-0.62	-0.53	-0.76	0.08	**0.006**

	7	-0.62	1.34	1.48	1.59	**0.038**
	8	-1.02	-0.31	0.09	0.28	**0.020**
	9	-1.02	-0.49	0.76	0.99	**0.030**
	10	-1.02	-0.26	-0.52	-0.55	**0.009**

Supplementary Table 2. *In vivo* competitions between a *gatZ* mutant and the ancestral strain in WT and *Rag2*$^{-/-}$ co-housed mice.

Table displaying the change in ln(*gatZ*/anc), i.e., the natural logarithm of the ratio between the frequencies of the *gatZ* mutant and the ancestral strain, over 3 days (72 hours) of competition in WT and *Rag2*$^{-/-}$ co-housed mice (see also Fig. 3a). ***s*** represents the selective advantage of *gatZ* along the three days of competition, inferred from the slope of the linear regression of ln(*gatZ*/anc).

Mice		**Time (hours)**				***s***
		0	**24**	**48**	**72**	
WT co-housed	**1**	-0.05	0.50	0.81	2.06	**0.026**
	2	-0.05	0.79	2.04	1.06	**0.025**
	3	-0.60	1.29	2.87	3.73	**0.065**
	4	-0.60	0.87	3.37	4.26	**0.071**
	5	-0.60	0.77	1.97	2.82	**0.050**
	6	-0.60	0.69	2.31	3.73	**0.060**
	7	-0.60	1.20	3.82	0.86	**0.045**
	8	-0.65	1.02	2.37	3.17	**0.057**
	9	-0.54	1.75	3.16	3.65	**0.066**
	10	-0.45	1.24	1.78	2.75	**0.047**
	11	-0.52	1.57	4.40	4.14	**0.077**
Rag2$^{-/-}$ co-housed	**1**	-0.05	0.84	2.00	2.45	**0.037**
	2	-0.05	0.73	1.78	0.67	**0.020**
	3	-0.60	1.46	2.45	2.90	**0.056**
	4	-0.60	0.84	2.81	1.01	**0.039**
	5	-0.60	-0.13	1.38	2.39	**0.040**
	6	-0.60	1.02	2.45	2.93	**0.054**
	7	-0.60	0.94	2.21	3.38	**0.057**
	8	-0.65	1.79	3.55	4.20	**0.076**
	9	-0.54	2.03	2.83	3.35	**0.062**
	10	-0.45	1.74	3.49	1.21	**0.045**
	11	-0.52	2.14	2.79	3.41	**0.063**

Supplementary Table 3. *In vivo* competitions between a *gatZ* mutant and the ancestral strain in WT and *Rag2*$^{-/-}$ germ-free mice.

Table displaying the change in ln(*gatZ*/anc), i.e., the natural logarithm of the ratio between the frequencies of the *gatZ* mutant and the ancestral strain, over 3 days (72 hours) of competition in WT and *Rag2*$^{-/-}$ germ-free mice (see also Fig. 3b). ***s*** represents the selective advantage of *gatZ* along the three days of competition, inferred from the slope of the linear regression of ln(*gatZ*/anc).

Mice		**Time (hours)**				***s***
		0	**24**	**48**	**72**	
WT germ-free	**1**	-0.61	0.46	0.92	1.27	**0.029**
	2	-0.61	0.43	1.17	1.76	**0.035**
	3	-0.20	0.71	1.32	1.79	**0.030**
	4	-0.20	0.56	0.87	1.92	**0.028**
	5	-0.54	0.24	0.41	0.83	**0.020**
	6	-0.54	0.43	0.99	1.25	**0.028**
	7	-0.54	0.19	0.51	0.88	**0.021**
Rag2$^{-/-}$ germ-free	**1**	-0.77	0.55	1.08	1.28	**0.033**
	2	-0.77	0.52	1.00	1.18	**0.032**
	3	-0.77	0.20	0.95	1.25	**0.031**
	4	-0.54	0.06	0.76	1.48	**0.028**
	5	-0.54	-0.11	0.56	1.19	**0.023**
	6	-0.54	-0.27	0.53	1.13	**0.022**
	7	-0.54	-0.09	0.49	0.83	**0.020**

Supplementary Table 4. *In vivo* competitions between two *gat* mutants (*gatZ* and *gatC*) in WT and *Rag2*$^{-/-}$ mice.

Table displaying the change in ln(*gatZ*/*gatC*), i.e., the natural logarithm of the ratio between the frequencies of the *gatZ* and *gatC* mutants, over 3 days (72 hours) of competition in WT and *Rag2*$^{-/-}$ mice (see also Fig. 4a). ***s*** represents the selective advantage of *gatZ* along the three days of competition, inferred from the slope of the linear regression of ln(*gatZ*/anc).

Mice		**Time (hours)**				***s***
		0	**24**	**48**	**72**	
WT	**1**	-0.64	-0.54	-0.51	0.25	**0.009**
	2	-0.64	-0.87	-0.24	-0.21	**0.006**
	3	-0.64	0.69	0.45	0.42	**0.020**
	4	-0.64	-0.11	0.08	0.71	**0.018**
	5	-0.64	0.12	-0.32	0.14	**0.011**
	6	-0.70	0.30	0.20	0.41	**0.018**
	7	-0.70	-0.48	0.05	0.06	**0.012**
	8	-0.70	-0.15	0.34	0.37	**0.017**
Rag2$^{-/-}$	**1**	-0.64	-0.71	-0.77	-1.50	**-0.009**
	2	-0.64	-0.54	-1.09	0.01	**0.003**

	3	-0.64	-2.71	-4.42	-6.42	**-0.080**
	4	-0.64	-0.45	-1.69	-2.88	**-0.026**
	5	-0.64	-1.13	-1.36	-1.52	**-0.014**
	6	-0.70	-0.07	-0.32	-2.48	**-0.012**
	7	-0.70	0.08	-2.15	-3.90	**-0.035**
	8	-0.70	-1.37	-1.55	-3.45	**-0.032**

Supplementary Table 5. *In vivo* competitions between two *gat* mutants (*gatZ* and *gatC*) in WT and *Rag2*$^{-/-}$ co-housed mice.

Table displaying the change in ln(*gatZ*/*gatC*), i.e., the natural logarithm of the ratio between the frequencies of the *gatZ* and *gatC* mutants, over 3 days (72 hours) of competition in WT and *Rag2*$^{-/-}$ co-housed mice (see also Fig. 4b). ***s*** represents the selective advantage of *gatZ* along the three days of competition, inferred from the slope of the linear regression of ln(*gatZ*/anc).

Mice		**Time (hours)**				***s***
		0	**24**	**48**	**72**	
WT co-housed	**1**	-0.35	-0.17	1.36	0.97	**0.023**
	2	-0.35	-0.10	2.04	1.96	**0.038**
	3	-0.35	-0.37	0.18	0.90	**0.018**
	4	-0.35	0.11	0.84	0.71	**0.016**
	5	-0.35	-0.21	0.51	0.58	**0.015**
	6	-0.08	0.10	0.25	0.31	**0.006**
	7	-0.08	-0.30	0.48	0.50	**0.011**
	8	-0.40	0.61	0.54	0.67	**0.013**
	9	-0.72	0.55	0.95	0.83	**0.021**
	10	-0.59	0.43	0.53	0.44	**0.013**
Rag2$^{-/-}$ co-housed	**1**	-0.35	-0.06	0.21	0.71	**0.014**
	2	-0.35	-0.20	0.04	-1.51	**-0.008**
	3	-0.35	0.06	0.57	0.32	**0.013**
	4	-0.35	0.50	1.00	0.70	**0.020**
	5	-0.08	1.49	2.69	2.95	**0.048**
	6	-0.08	-0.42	0.34	0.64	**0.008**
	7	-0.40	0.37	1.35	-0.01	**0.016**
	8	-0.72	0.39	0.26	-1.08	**0.006**
	9	-0.59	0.42	1.23	1.48	**0.032**

Supplementary Table 6. The number and nature of adaptive events across independently evolved populations of *E. coli* after 24 days of adaptation in WT mice

Mutations found in *E. coli* populations after adaptation to the gut of WT mice. For intergenic mutations the two flanking genes are listed, otherwise the mutation occurred in the gene coding region. SNPs are represented by an arrow between the ancestral and the evolved nucleotide. The symbol Δ means a deletion event and a + symbol represents an insertion of the nucleotide that follows the symbol. The initials IS denote the abbreviation of insertion sequence element at the indicated position. del/dup indicates that either a deletion or a duplication of the indicated size occurred but it is not possible to distinguish between the two. * indicates that the mutation corresponds to a supported unassigned new junction whereas ¤ denotes a IS insertion where only one new junction was identified.

Population	Genome position	Gene	Mutation	Frequency	
1.1	4346885	*dcuB / dcuR*	IS5 +4 bp	77%	
	2173581	*gatZ*	IS1 +9 bp	77%	
	2173490	*gatZ*	IS1 +8 bp	36%	
	3527427	*hslR*	del/dupl 56 bp	6%	*
	4035239	*ileT / ileT/alaT*	del/dupl 11 bp	5%	*
	1269768	*ldrC / chaA*	G→T	7%	
	3873556	*yidX*	T→G	8%	
	4601137	*yjjP / yjjQ*	IS5	5%	¤
	3214187	*yqjH*	del/dupl 82 bp	6%	*
1.2	921810	*cspD*	C→T	6%	
	4346885	*dcuB / dcuR*	IS5 +4 bp	48%	
	2172866	*gatA*	IS5 +4 bp	5%	
	2175295	*gatY / fbaB*	IS1 +9 bp	60%	
	2173751	*gatZ*	IS1 +9 bp	30%	
	2173522	*gatZ*	IS1 +9 bp	7%	
	968583	*lpxK/ycaQ / ycaQ*	del/dupl 42 bp	6%	*
	3996061	*uvrD*	del/dupl 39 bp	5%	*
	2180108	*yegW*	IS3	24%	¤
	4601256	*yjjP / yjjQ*	IS5 +4 bp	28%	
1.3	4346885	*dcuB / dcuR*	IS5 +4 bp	28%	
	2172866	*gatA*	IS5 +4 bp	7%	
	2175234	*gatY / fbaB*	IS1 +9 bp	13%	
	2175282	*gatY / fbaB*	IS5 +4 bp	5%	
	2173544	*gatZ*	IS1	27%	
	2173511	*gatZ*	IS1	6%	¤
	2173225	*gatZ*	IS1 +9 bp	7%	
	1847931	*sppA*	C→T	8%	
	2711851	*srmB*	del/dupl 79 bp	7%	*
	4601140	*yjjP / yjjQ*	IS5 +12 bp	25%	
	4601120	*yjjP / yjjQ*	IS2 +5 bp	10%	
1.4	4346885	*dcuB / dcuR*	IS5 +4 bp	12%	
	2172869	*gatA*	IS5	10%	¤

	2175284	*gatY / fbaB*	IS1 +9 bp	24%	
	2175250	*gatY / fbaB*	IS1 +9 bp	22%	
	2175266	*gatY / fbaB*	IS1	8%	¤
	2173522	*gatZ*	IS1 +9 bp	16%	
	2174014	*gatZ*	IS1 +9 bp	7%	
	2827379	*srlR*	IS1	7%	¤
	285098	*yagG*	del/dupl 22 bp	5%	*
	2143874	*yegE / yehD/yehE*	del/dupl 46 kb	5%	*
	3214187	*yqjH*	del/dupl 82 bp	5%	*
1.5	446451	*cyoE*	del/dupl 39 bp	6%	*
	4346885	*dcuB / dcuR*	IS5 +4 bp	31%	
	2172866	*gatA*	IS5 +4 bp	44%	
	2174422	*gatY*	IS1 +9 bp	45%	
	695730	*glnV*	Δ112 bp	6%	
	1907559	*kdgR*	G→A	8%	
	2827493	*srlR*	G→C	39%	
	4601137	*yjjP / yjjQ*	IS5	5%	¤
1.6	4346885	*dcuB / dcuR*	IS5 +4 bp	60%	
	3379861	*degQ*	del/dupl 30 bp	6%	*
	2566222	*eutH / eutG*	Δ100 bp	6%	
	2172784	*gatA*	IS1 +9 bp	14%	
	2172866	*gatA*	IS5 +4 bp	7%	
	2172489	*gatB*	IS1 +9 bp	8%	
	2172260	*gatC*	Δ1 bp	27%	
	2175282	*gatY/fbaB*	IS1	5%	¤
	2173533	*gatZ*	IS5 +4 bp	6%	
	2174345	*gatZ / gatY*	IS1 +8 bp	27%	
	1510987	*ydcT*	del/dupl 65 bp	5%	*
1.7	953901	*focA / ycaO*	IS5 +4 bp	68%	
	3269213	*garK*	C→T	12%	
	2172535	*gatB*	IS1	10%	¤
	2172265	*gatC*	IS1 +9 bp	7%	
	2175290	*gatY / fbaB*	IS1 +9 bp	66%	
	1907949	*kdgR*	IS5 +4 bp	8%	
	1908116	*kdgR*	IS5 +4 bp	6%	
	1902230	*manZ / kdgR*	Δ 5452 bp	72%	*
	1573764	*yddB*	IS5	7%	¤
1.8	2172866	*gatA*	IS5 +4 bp	6%	
	2172221	*gatC*	IS5 +4 bp	49%	
	2172265	*gatC*	IS1 +9 bp	15%	
	2173970	*gatZ*	IS1	10%	¤
	2174027	*gatZ*	IS1 +8 bp	7%	
	2440543	*mnmC*	del/dupl 103 bp	7%	*
	2827492	*srlR*	G→A	28%	

	3634862	*yhiN*	del/dupl 14 bp	7%	*
1.9	2880256	*casB*	del/dupl 18 bp	6%	*
	3379861	*degQ*	del/dupl 30 bp	8%	*
	2172633	*gatA*	IS5 +4 bp	29%	
	2172265	*gatC*	IS1 +9 bp	9%	
	2175296	*gatY / fbaB*	IS1 +9 bp	11%	
	2173104	*gatZ*	IS1 +9 bp	25%	
	1434252	*ompN*	del/dupl 16 bp	5%	*
	2827073	*srlR*	A→T	24%	
	1877853	*yeaR*	Δ1 :: IS186 +6 bp :: Δ1	10%	
	2763434	*yfjL / yfjM*	Δ8 bp	8%	
	4601117	*yjjP / yjjQ*	IS5 +4 bp	27%	
	4601148	*yjjP / yjjQ*	IS5 +4 bp	16%	
	4601266	*yjjP / yjjQ*	IS5 +4 bp	6%	
	4601006	*yjjP / yjjQ*	IS2	9%	¤
1.10	3379861	*degQ*	del/dupl 30 bp	6%	*
	953892	*focA / ycaO*	IS5 +13 bp	66%	
	2171086	*gatC*	IS1 +8 bp	47%	
	2172265	*gatC*	IS1 +9 bp	15%	
	2174223	*gatZ*	Δ2 bp	33%	
	4601134	*yjjP / yjjQ*	IS5 +4 bp	5%	
1.11	2172866	*gatA*	IS5 +4 bp	83%	
	2174150	*gatZ / gatY/fbaB*	Δ1111 bp	14%	*
	2827172	*srlR*	C→A	16%	
	1691840	*uidB*	del/dupl 20 bp	5%	*
	4601110	*yjjP / yjjQ*	IS5 +4 bp	72%	
	4601256	*yjjP / yjjQ*	IS5 +4 bp	7%	
1.12	953901	*focA / ycaO*	IS5 +4 bp	20%	
	2172079	*gatC*	.→C	97%	
	2827234	*srlR*	G→A	36%	
	2827117	*srlR*	C→T	6%	
	12154	*yaaI / dnaK*	IS4	47%	¤
1.13	4637714	*arcA*	G→T	87%	
	2171150	*gatC*	IS5 +4 bp	93%	
	3265169	*tdcA / tdcR*	IS5 +4 bp	11%	
	572125	*ybcN*	IS2 +5 bp	6%	
1.14	4346885	*dcuB / dcuR*	IS5 +4 bp	70%	
	953892	*focA / ycaO*	IS5 +4 bp	58%	
	2172912	*gatA*	IS5 +4 bp	15%	
	2175256	*gatY / fbaB*	IS1 +9 bp	81%	
	2827117	*srlR*	C→T	13%	
	1877853	*yeaR*	IS186 +6 bp	66%	
	4601151	*yjjP / yjjQ*	IS5	7%	¤

Supplementary Table 7. The number and nature of adaptive events across independently evolved populations of *E. coli* after 24 days of adaptation in *Rag2*$^{-/-}$ mice

Mutations found in *E. coli* populations after adaptation to the gut of *Rag2*$^{-/-}$ mice. See Supplementary Table 1 legend for further details.

Population	Genome position	Gene	Mutation	Frequency	
R1.1	3925453	*asnA*	inv 15 bp	6%	
	953892	*focA / ycaO*	IS5 +4 bp	56%	
	2172791	*gatA*	IS1 +9 bp	8%	
	2172073	*gatC*	IS1 +9 bp	80%	
	2827355	*srlR*	T→C	59%	
	2827528	*srlR*	G→A	13%	
	2827712	*srlR*	+CCCCAA	9%	
	2827388	*srlR*	Δ1 bp	5%	
	3634862	*yhiN*	del/dupl 14 bp	5%	*
	4601143	*yjjP / yjjQ*	IS5	5%	¤
R1.2	3925441	*asnA*	del/dupl 39 b	7%	*
	2172008	*gatC*	Δ 7508 bp	90%	
	2827081	*srlR*	C→T	83%	
	3214187	*yqjH*	del/dupl 82 bp	8%	*
R1.3	3350864	*arcB*	+AAA	37%	
	3925441	*asnA*	del/dupl 39 b	5%	*
	3502248	*frlR*	IS5 +4 bp	29%	
	3502642	*frlR*	Δ1 bp	7%	
	2172835	*gatA*	IS1 +9 bp	29%	
	2175289	*gatY / fbaB*	IS1	62%	
	2173225	*gatZ*	IS1 +9 bp	8%	
	156651	*yadN*	IS5 +4 bp	10%	
	4592986	*yjjM*	IS1 +9 bp	7%	
	4601157	*yjjP / yjjQ*	IS1 +8 bp	5%	
	2039432	*zinT*	IS5 +4 bp	14%	
R1.4	2173491	*gatZ*	IS5 +4 bp	94%	
	2827492	*srlR*	G→A	87%	
R1.5	2172972	*gatA*	IS1 +9 bp	84%	
	2752245	*ratB*	del/dupl 23 bp	5%	*
	2827793	*srlR*	T→G	85%	
	3634862	*yhiN*	del/dupl 14 bp	6%	*
R1.6	2172774	*gatA*	IS1	6%	¤
	2173014	*gatA*	IS5	5%	¤
	2175284	*gatY / fbaB*	IS1 +9 bp	81%	
	4035239	*ileT / ileT/alaT*	del/dupl 21 bp	7%	*
	932176	*lrp*	G→T	8%	
R1.7	3925441	*asnA*	del/dupl 39 b	5%	*
	3496103	*cysG*	del/dupl 5 bp	5%	*

	251816	*dinB*	del/dupl 48 bp	5%	*
	2172211	*gatC*	IS1 Δ4 bp	22%	
	2175244	*gatY / fbaB*	IS1 +9 bp	71%	
	2827492	*srlR*	G→A	43%	
	2891617	*ygcN / gcvP*	del/dupl 153 kb	63%	*
R1.8	3925441	*asnA*	del/dupl 39 b	5%	*
	953892	*focA / ycaO*	IS5 +4 bp	63%	
	3502032	*frlD / frlR*	IS4 +12 bp	7%	
	2172866	*gatA*	IS5 +4 bp	51%	
	2175266	*gatY / fbaB*	IS1 +9 bp	16%	
	2173571	*gatZ*	IS1 +9 bp	9%	
	1125212	*rimJ*	IS5 +4 bp	15%	
	761046	*sucB*	del/dupl 67 bp	7%	*
	931615	*trxB / lrp*	IS5 +4 bp	5%	
	3240783	*uxaA*	del/dupl 26 bp	5%	*
	3856228	*yidJ*	del/dupl 23 bp	5%	*
	3152375	*yqhC*	del/dupl 13 bp	6%	*
R1.9	4359175	*cadC*	A→T	8%	
	2173093	*gatA*	IS1 +8 bp	5%	
	2173093	*gatZ*	IS1 +9 bp	38%	
	2174199	*gatZ*	G→A	37%	
	1907809	*kdgR*	Δ2 bp	26%	
	2827627	*srlR*	G→A	26%	
	1221585	*ymgD*	C→T	25%	
R1.10	3379861	*degQ*	del/dupl 30 bp	5%	*
	628178	*entA*	del/dupl 97 bp	6%	*
	2172912	*gatA*	IS5 +4 bp	45%	
	2172866	*gatA*	IS5 +4 bp	13%	
	2175244	*gatY / fbaB*	IS1 +9 bp	35%	
	1907971	*kdgR*	IS1	12%	¤
	4197319	*nfi*	del/dupl 40 bp	5%	*
	298917	*paoC*	del/dupl 53 bp	7%	*
	3878678	*recF*	del/dupl 62 bp	7%	*
	1125212	*rimJ*	IS5 +4 bp	12%	
	2827490	*srlR*	C→T	21%	
	2030708	*yedJ*	del/dupl 50 bp	6%	*
R1.11	2173028	*gatA*	IS1 +9 bp	76%	
	2172073	*gatC*	IS1 +9 bp	6%	
	2173584	gatZ	IS1	*6%*	¤
	2173160	*gatZ*	IS1 +9 bp	6%	
	3325215	*rlmE*	del/dupl 37 bp	5%	*
	2827712	*srlR*	G→A	50%	
	2827492	*srlR*	G→A	13%	
R1.12	953901	*focA / ycaO*	IS5 (+) +4 bp	85%	

	1985465	*ftnB / yecJ*	del/dupl 9 bp	5%	*
	2172866	*gatA*	IS5 (+) +4 bp	97%	
	2827184	*srlR*	A→G	36%	
R1.13	3350023	*arcB*	T→C	33%	
	953919	*focA / ycaO*	IS1	7%	¤
	2172153	*gatC*	IS1 +9 bp	5%	
	2173559	*gatZ*	IS1 +9 bp	47%	
	3353621	*gltB*	del/dupl 52 bp	5%	*
	1482153	*hrpA*	del/dupl 35 bp	6%	*
	4035239	*ileT / ileT/alaT*	del/dupl 21 bp	6%	*
	2373853	*menC*	IS186	7%	¤
	2317269	*rcsC*	del/dupl 17 bp	5%	*
	2066430	*yoeH*	del/dupl 5 bp	5%	*
R1.14	953901	*focA / ycaO*	IS5 +4 bp	87%	
	2172650	*gatA*	IS1 +8 bp	83%	
	2171899	*gatC*	IS1	8%	¤
	1632600	*tfaQ*	del/dupl 72 bp	6%	*
	4368228	*yjeH*	del/dupl 111 bp	7%	*
R1.15	3925441	*asnA*	del/dupl 39 b	9%	*
	953901	*focA / ycaO*	IS5 +4 bp	75%	
	2172889	*gatA*	IS1 +8 bp	30%	
	2173012	*gatA*	IS1 +9 bp	10%	
	2171613	*gatC*	IS1 +9 bp	26%	
	3527427	*hslR*	del/dupl 56 bp	8%	*
	2620189	*purM*	G→T	30%	
	4470157	*pyrB*	del/dupl 49 bp	5%	*
	4383948	*yjeO*	Δ1 bp	5%	

CHAPTER 4

Natural selection in bacteria colonizing the intestinal tract: a serendipitous case of reverse evolution

Manuscript in preparation.

The author of this thesis performed the *in vivo* evolution experiment, the estimation of the evolutionary parameters and the phenotypic dynamics of the *gat* reversion described in this chapter.

Natural selection in bacteria colonizing the intestinal tract: a serendipitous case of reverse evolution

Ana Sousa[1*], Marta Lourenço[1], João Barroso-Batista[1], Karina B. Xavier[1,2], Isabel Gordo[1*]

[1]Instituto Gulbenkian de Ciência, Oeiras, Portugal.

[2]Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Lisbon, Portugal.

Abstract

The relative role of drift versus selection underlying the evolution of bacteria in the gut microbiota remains poorly understood. We followed the emergence of intra-species diversity in a commensal *Escherichia coli* strain previously adapted to the mouse gut. Remarkably rampant reverse evolution, at both the genetic and phenotypic levels within and among populations to reacquisition of metabolic function, occurred. This shows that recurrent violations of Dollo's law of evolutionary irreversibility can be common in the microbiota. Natural selection, in its different forms, dominated the evolutionary change observed. Periodic selection reducing diversity was rare while frequency-dependent selection actively maintaining polymorphism was common. Furthermore the continuous emergence of similar phenotypes due to distinct mutations, clonal interference, was found to be pervasive. Evolutionary change within the gut is therefore repeatable with adaptive mutations accumulating without strong constrains on genetic background but coerced by fluctuating selection modulating diversity.

Introduction

Gut microbiota composition is presently recognized as an important indicator of health status. Time series data has shown that the gut microbiota typically

comprises a diverse community of species and that reduction in its diversity is frequently associated with illness (Lozupone et al., 2012). Seldom investigated, but potentially as important, is intra-species evolution (Ahern et al., 2014; Greenblum et al., 2015). In fact, studies providing an understanding on how intraspecific microbiota variation emerges and changes over time are lacking (Waldor et al., 2015). Therefore, important questions such as whether the extant intra-species diversity is mainly due to migration and genetic drift or the result of natural selection on new mutations are still unanswered. While some mutations segregating in natural populations may be neutral, the large size of bacterial communities inhabiting the mammalian gut suggests that here the polymorphism is mostly the result of deterministic forces (Falush, 2009). Natural selection is the "principle by which each slight variation, if useful, is preserved" (Darwin, 1859). If a variation on a trait created by mutation is always useful, positive Darwinian selection will drive the replacement of that trait by the new one. Additionally, if multiple useful variants emerge, the one coding for the best trait will end up winning the competition. Finally, if a variant is useful in some contexts but not in others, it may not fully dominate but still be preserved in the population. The importance of natural selection and its strength versus other evolutionary processes in shaping intra-species variation in the guts of hosts living in their natural environments is hard to dissect. However in a natural yet controlled system, the influence of natural selection can be tracked thanks to the genetic signatures it leaves in the populations where it acts.

Mouse colonization models offer a great opportunity to understand microbiota evolution and emergence of diversity, test its repeatability and measure the strength of natural selection. Here, using a common model of gut colonization, we show that all classical forms of natural selection, *i.e*. periodic selection, balancing selection and clonal interference are ubiquitous in determining strain variation within the gut. Moreover we unravel a recurrent case of violation of Dollo’s law, which proposes that evolution is unidirectional and irreversible (Marshall et al., 1994). The results demonstrate *de novo* acquisition of a primordial lost phenotype via a genetic route that can include compensatory mutation but also genetic reversion leaving no trace of the past.

Material 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.

Bacterial strains and culture conditions

All strains used were derived from MG1655, a K12 commensal strain of *E. coli*. (Blattner et al., 1997) Strains JB19-YFP and JB18-CFP (MG1655, *galK::YFP/CFP* cm^R , str^R *(rpsl150), ΔlacIZYA,* Ins (1bp) *gatC*) were used for the evolution experiment here reported. These strains differ from the ancestral MG1655 fluorescent strains DM08-YFP and DM09-CFP used in a previous evolution experiment (MG1655, *galK::YFP/CFP* amp^R , str^R *(rpsl150), ΔlacIZYA*)(Barroso-Batista et al., 2014)) by a mutation in the *gatC* gene (1bp insertion in the coding region). To construct these strains the ampicillin resistance cassette in the ancestral strains DM08-YFP and DM09-CFP was replaced with a chloramphenicol resistant cassette using the Datsenko and Wanner method (Datsenko and Wanner, 2000). The yellow (yfp) and cyan (cfp) fluorescent genes linked to cm^R were then transferred by P1 transduction to a derivative of clone 12YFP (Barroso-Batista et al., 2014), an evolved clone of DM08-YFP, isolated after 24 days of adaptation in the gut of WT mice, that carried an insertion of 1bp in *gatC*.

To measure the effect of the identified parallel mutations in gene expression we tested the following clones isolated from the evolution experiment here reported (see Supplementary Table 1): 18YFP (*focA srlR*), 22YFP (*dcuB*), 25YFP (*yjjP/yjjQ radA insX-insA*), 29CFP (*arcA*) and the ancestral strains DM08-YFP and DM09-CFP. Five biological replicates from each clone were performed for both aerobic and anaerobic conditions.

To directly measure the effects of the mutations involved in the 2^{nd} step of adaptation by *in vivo* fitness assays we used several single mutants derived from clones isolated from a previous evolution experiment (Barroso-Batista et al., 2014)

(5YFP (*srlR*) and 6YFP (*dcuB*)) or the experiment here reported (17YFP (*arcA*), 24YFP (*radA*) and 2 clones isolated from population 2.14 screened by PCR for mutation at *focA* and *yjjP* locus, respectively). The *oppB* single mutant was obtained by transducing the knockout of this gene (*oppB*::*Kan*) from the KEIO collection (Baba et al., 2006) to the ancestral DM08-YFP background. These clones, initially carrying an additional mutation in one of *gat* operon genes (*gatA* or *gatC*), were made single mutants by P1 transduction from JW2074 (*gatR*::Kan), a *gatR* knockout mutant from the KEIO collection. *gatR* is already interrupted by a transposable element in the ancestral MG1655 strain, and therefore transduction with P1 from a mutant strain *gatR*, leads to an effective replacement of the neighbor genes of the *gat* operon to their wild type status, while maintaining a knockout mutation in *gatR*. As reference strains, we used derivatives of the ancestral DM08-YFP and DM09-CFP strains in which the *gatR* gene was replaced with *gatR*::Kan.

To test for frequency dependent selection by competitive fitness assays we used the single *srlR* mutant previously mentioned (derived from clone 5YFP but with a wild type *gat* operon), as well as the ancestral strain DM09-CFP.

To test if clones evolved *in vivo* had a different growth ability we performed *in vitro* competitions against the ancestral strains DM08-YFP and DM09-CFP of clones isolated from 14 independent populations from a previous evolution experiment (sequenced clones 1 to 14 from populations 1.1 to 1.14, (Barroso-Batista et al., 2014)) and this evolution experiment (sequenced clones 16 to 30, from populations 2.1 to 2.15, Supplementary Table 1).

To distinguish between *gat*-negative and *gat*-positive bacteria we used the differential medium Mac Conkey agar supplemented with 1% galactitol and streptomycin (100µg/ml). Plates were incubated at 30ºC. The frequency of galactitol mutants was estimated by counting the number of white (auxotrophic for galactitol) and red colonies.

To perform the fluctuation test for the *gat*-negative phenotype we used the selective M9 Minimal Medium (MM) agar, supplemented with D-arabitol (10mM) and glycerol (0.4%) or Luria Broth (LB) agar supplemented with furazolidone (1.25 µg/ml).

For the *in vitro* competition assays we used MM supplemented with 3mM of $MgSO_4$ and either sorbitol, ribose, mannose, gluconate or glucuronate at a concentration of 0.02%. Additionally a mixture with the different carbon sources (composed of 0.1% from each of the five carbon sources) was also tested.

Evolution experiment

In order to study *E. coli* adaptation to the gut we used the classical streptomycin-treated mouse colonization model (T. Conway et al., 2004). Briefly, 6- to 8-week-old C57BL/6 male mice raised in specific pathogen free (SPF) conditions were given autoclaved drinking water containing streptomycin (5g/L) for one day. After 4 hours of starvation for water and food the animals were gavaged with 100µl of a suspension of 10^8 colony forming units (CFUs) of a mixture of MG1655-YFP-*gatC* and MG1655-YFP-*gatC* bacteria (ratio 1:1) grown at 37°C in brain heart infusion medium to optical density $(OD)_{600}$ of 2. After the gavage, all the animals were housed separately and both the water with streptomycin and the food were returned to them. Mice fecal samples were collected for 24 days and diluted in PBS, from which a sample was stored in 15% glycerol at -80ºC and the remaining was plated in LB agar supplemented with streptomycin. Plates were incubated overnight at 37ºC and then with the help of a fluorescent stereoscope (SteREO Lumar, Carl Ziess) the fluorescent colonies were counted to assess the frequencies of CFP- and YFP-labelled bacteria. These fluorescent proteins are used as neutral markers with which we can follow the appearance of beneficial mutations, since these markers hitchhike with the beneficial mutations that spread in the populations (Hegreness et al., 2006).

***In vivo* competitive fitness assays**

To test the *in vivo* advantage of clones sampled from the evolving populations at the last time point of the evolution experiment (Fig. 1B) samples of either YFP or CFP clones isolated from day 24 (sub-populations) were competed against the respective ancestor labelled with the opposite fluorescent marker. These sub-populations were composed of mixtures of approximately 30 colonies with the same fluorescent marker isolated after plating the appropriate dilution of mice fecal

pellets. These mixtures of clones were then grown in 10 ml of LB supplemented with chloramphenicol (100 µg/ml) and streptomycin (100 µg/ml) and stored in 15% glycerol at -80ºC. *In vivo* competitions of evolved sub-populations against the ancestral were performed at a ratio of 1 to 1, following the same procedure described above for the evolution experiment. Mice fecal pellets were collected for 3 days, diluted in PBS and frozen in 15% glycerol at -80ºC. Total numbers and relative proportions of YFP- and CFP-labeled *E. coli* were subsequently determined by flow cytometry, using a BD LSRFortessa cytometer.

The selective coefficient (fitness gain) of these clones *in vivo* (presented in Figure 1) was estimated as:

$$s_b = \ln\left(\frac{Rf_{ev/anc}}{Ri_{ev/anc}}\right)/t$$

where s_b is the selective advantage of the evolved clone, $Rf_{ev/anc}$ and $Ri_{ev/anc}$ are the ratios of evolved to ancestral bacteria in the end (*f*) or in the beginning (*i*) of the competition and *t* is the number of generation per day. We assumed *t*=18, in accordance with the 80 minute generation time estimated in previous studies on *E.coli* colonization of streptomycin-treated mouse (Barroso-Batista et al., 2015; Poulsen et al., 1995; Rang et al., 1999).

To assess the selective advantage *in vivo* of the mutations involved in the 2nd step of adaptation (Fig. 2) single mutants harboring mutations in the *arcA*, *dcuB*, *focA*, *oppB*, *yjjP* and *srlR* loci (see “Bacterial strains and culture conditions” above) were competed against the ancestral in streptomycin-treated mice, following the protocol previously described. Mice fecal samples were collected for 24 days, diluted in PBS and plated in Mac Conkey agar supplemented with galactitol (1%) and streptomycin (100µg/ml). After overnight incubation at 30ºC, the colonies were screened for the *gat* phenotype, based on their white or red color. In addition, CFP- and YFP-labelled bacteria were counted with a fluorescent stereoscope (SteREO Lumar, Carl Ziess). The selective advantage of the mutants was calculated as for the mixtures of clones described above.

Whole genome re-sequencing and mutation prediction

Clone analysis: After 24 days of colonization one clone from each independently evolving populations (2.1 to 2.15) was isolated. It was grown in 10 ml of LB at 37ºC with agitation for DNA extraction. Subsequently DNA was isolated following a previously described protocol (Wilson, 2001). The DNA library construction and sequencing was carried out by BGI. Each sample was pair-end sequenced on an Illumina HiSeq 2000. Standard procedures produced data sets of Illumina paired-end 90 bp read pairs with insert size (including read length) of 470 bp. Genome sequencing data have been deposited in the NCBI Read Archive database with the accession no. SRP063701. 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).

Population analysis: DNA isolation was obtained in the same way as described above for the clone analysis except that now it derived from a mixture of >1000 clones per population grown in LB agar. Four populations, from the evolution experiment, were sequenced: 2.7, 2.10, 2.14 and 2.15. Those were sequenced for three time points during the adaptive period (generation 198 (day11), generation 306 (day17) and generation 432 (day24)). The DNA library construction and sequencing was carried out by the IGC genomics facility. Each sample was pair-end sequenced on an Illumina MiSeq Benchtop Sequencer. Standard procedures produced data sets of Illumina paired-end 250 bp read pairs. The mean coverage per sample was between ~90x and ~150x for population 2.7, between ~100x and ~120x for population 2.10, between ~111x and ~93x for population 2.14 and between ~70x and ~115x for population 2.15. Mutations were identified using the BRESEQ pipeline (version 0.26) with the polymorphism option

on. The default settings were used except for: a) requirement of a minimum coverage of 3 reads on each strand per polymorphism; b) eliminating polymorphism predictions occurring in homopolymers of length greater than 3; c) polymorphism predictions with significant (*P*=0.05) strand or base quality score bias were discarded.

Identification of adaptive mutations and estimate of haplotype frequencies in selected populations of the evolution experiment

In order to estimate the haplotype frequencies depicted in Fig. 2, 3 and 4, two complementary strategies were employed. In addition to the WGS of the populations, targeted PCR of the identified parallel mutations was performed. For the targeted PCR, 20 to 80 clones from different time points were isolated (from populations 2.7, 2.10, 2.14 and 2.15). The isolation procedure consisted in diluting the frozen fecal samples in PBS and plating the appropriate dilution in LB agar plates supplemented with streptomycin (100 µg/ml), and incubating overnight at 37˚C. The frequencies of CFP or YFP bacteria were measured by counting the CFUs in a stereoscope (SteREO Lumar, Carl Zeiss).

Mutations were screened by PCR followed by electrophoresis in 1% agarose gel, at 50V for 1h30min. IS insertions were scored by increases (around 700-1500bp) in the size of the PCR fragment. Single nucleotide polymorphisms (SNPs) in *srlR* were identified by sequencing this locus using an ABI 3130XL and ABI 377 Automatic Sequencer. The PCR reactions were performed in the same conditions as previously described (Barroso-Batista et al., 2014).

Analysis of gene expression changes caused by IS insertions

To determine the effects of the IS insertions identified during the 2nd steps of adaptation we measured the expression of *focA*, *dcuB*, *arcA*, *yjjY* and *yjjP* by qPCR in two environments with different levels of oxygen.

Aerobic Conditions: The clones were initially grown for 24h at 37ºC with aeration in MM with glycerol (0.02%). The cultures were diluted 10-fold and 100 µl of the dilution were inoculated in in 10ml MM supplemented with a mixture of the following carbon sources: sorbitol, ribose, mannose, gluconate and glucoronate, at

individual concentration of 0.01%. The cultures were grown at 37ºC, with aeration, until an OD_{600} of 0.5. Five milliliters of the bacterial culture were then harvested by centrifugation at 4ºC for 5 minutes at the maximum speed. The resulting pellet was ressuspended in lysozyme solution (5 mg lysozyme /ml DEPC treated water, Sigma protocol) and incubated at 37ºC for 30 minutes, promoting disruption of the bacterial cell wall and allowing for RNA extraction (see below).

Anaerobic conditions: The protocol used was the same as in the aerobic conditions with the following alterations: the second overnight growth was performed at 37ºC in an anaerobic chamber with the atmosphere of 5% H_2, 15% CO_2, 80% N_2 (Plas Labs, Lansing, MI, USA), and at approximately OD_{600} of 0.2 the cultures were placed in dry ice to prevent their growth and the cells were harvested by centrifugation from 10 ml of bacterial culture.

RNA extraction, DNAse treatment, RT-PCR and qPCR: The RNA extraction was performed with the Qiagen RNeasy Mini Kit. RNA concentration and quality were evaluated with Nanodrop 2000. DNase treatment was performed with the RQ1 DNase (Promega), 0.5µl of DNase and 1µl buffer were added to 1µg of RNA and incubated for 30 minutes at 37ºC. After this, 1 µl stop solution was added and then incubated for 15 minutes at 65ºC to inactivate the DNase. The resulting RNA was used for the reverse transcription which consisted in mixing with 1 µg of RNA, with 0.5 µl random primers and DEPC-water (final volume of 15 µl) and then incubated at 70ºC for 5min. Afterwards the M-MLV Reverse Transcriptase Protocol (Promega) were performed, to the first mix was added 5 µl of RT buffer, 0.5 µl RT enzyme and 2 µl dNTP mix, and then incubated 10 min at 25ºC, 50min at 50ºC and 10 min at 70ºC.

We used a relative quantification method of analysis with normalization against a reference gene. qPCR was executed in BioRad CFX 384 with itaq universal sybr green supermix (BioRad). cDNA was diluted 100-fold before used in the qPCR. The qPCR reaction conditions were as follows: one cycle of 2 min at 50ºC and then 39 cycles of 10 min at 95ºC, 30 sec at 95ºC, 1 min at 57ºC and finally 30 s at 72ºC. Primers used are listed in Supplementary Table 7. Melting curve analysis was performed to verify product homogeneity. All reactions

included three replicates for each sample. Data were normalized by the Pfaffl method (Pfaffl, 2001) using the *hfq* housekeeping gene of *E. coli* as a reference.

Fluctuation test for the *gat*-negative phenotype

To test for the possibility of a higher mutation rate at the galactitol operon, we determined the frequency of spontaneous *gat*-negative phenotype mutants when plated on D-arabitol. D-arabitol is known to be toxic for bacteria that are able to metabolize galactitol (*gat*-positive phenotype) (Reiner, 1977) and so the growth of *gat*-positive bacteria is much slower, allowing to differentiate between *gat*-positive and *gat*-negative clones. The ancestral strains DM08-YFP and DM09-CFP were grown overnight in 10 ml of LB at 37ºC with aeration. After growth, the total number of cells in the cultures was measured using BD LSR Fortessa (BD Biosciences) and approximately 1000 cells were used to inoculate 1 ml of LB (10 replicates of each strain) and incubated overnight. Aliquots of each replicate tube were plated in LB agar and MM agar supplemented with D-arabitol (10 mM) and glycerol (0.4%) and incubated overnight at 37ºC. The number of spontaneous *gat*-negative mutants and total number of cells grown on LB were used to estimate the mutation rate using the maximum likelihood approach as implemented in FALCOR (Hall et al., 2009).

Similarly, a fluctuation assay for measuring the spontaneous rate of emergence of furazolidone resistant mutants was used as proxy for the spontaneous rate of random gene inactivation. We then used this number to compare with the rate for *gat*-negative phenotype. The experiment was performed in the same conditions as described above except that the cultures were plated in LB supplemented with furazolidone (1.25 µg/ml).

Test for frequency dependent selection of *srlR* mutation in the mouse gut

To test the hypothesis that *srlR* mutations have disadvantage from high frequency but advantage from low frequency *in vivo*, a representative clone from a single *srlR* mutant (see “Bacterial strains and culture conditions” above) was competed against the ancestor at different initial frequencies. We followed the protocol described before for the evolution experiment but this time mice were

gavaged with mixture of *srlR*-YFP and anc-CFP at a ratio of 1:9 or 9:1 (3 replicates per condition). Mice fecal samples were collected for 7 days, diluted in PBS, and plated in Mac Conkey agar supplemented with galactitol (1%) and streptomycin (100µg/ml). *gat* and fluorescent marker frequencies were assessed as previously described (see "*In vivo* competitive assays" above).

In vitro tests for changes in the growth ability of evolved clones

To test whether *E. coli* clones evolved *in vivo* had a different nutritional profile when growing *in vitro*, we performed competitions between each of the sequenced clones (obtained both from (Barroso-Batista et al., 2014) and from the present work) and the ancestor of the first colonization of the opposite fluorescence. Competitions were performed in triplicate, in MM supplemented with different carbon sources and in two different oxygen conditions. All competitions were conducted in 96-well plates incubated at 37ºC (Thermoshaker PHMP-4, Grant) under (aerobiosis) or in an anaerobic chamber (anaerobiosis). The competitor and reference strains were initially acclimated to the growth media for two overnights in MM supplemented with glycerol (0.02%) and then 10^5-10^6 cells of both strains inoculated in MM containing 0.02% of either sorbitol, ribose, mannose, gluconate or glucuronate or a mixture of all five carbon sources at individual concentrations of 0.01%. The strains were allowed to compete for 24 hours and the initial and final ratios of both strains were determined by flow cytometry, using a BD LSRFortessa (BD Biosciences) cytometer. The relative fitness of the evolved clones (Supplementary Table 2) was estimated as previously described (see "*In vivo* competitive assays" above).

Competitions in anaerobic conditions were performed for each of the evolved clones following the protocol above described but with some alterations. After an initial overnight aerobic growth in MM with glycerol (0.02%), the cultures were diluted 10-fold, inoculated in MM with glycerol and acclimated overnight by incubation in an anaerobic chamber (5% H_2, 15% CO_2, 80% N_2) (Plas Labs, Lansing, MI, USA), at 37ºC. After acclimatization, the competitor and reference strain were inoculated in MM supplemented with individual or a mixture of carbon sources and allowed to compete for 48 hours. To determine the initial and final

ratios of the competing strains, serial dilutions of the mixtures were plated in LB supplemented with streptomycin (100µg/ml) and the resulting CFUs counted in a stereoscope (SteREO Lumar, Carl Zeiss).

Statistical analysis

To determine significant differences in gene expression between mutant and ancestral strain, the unpaired t test was used, with a significance defined as P value of <0.05.

Differences in the selective advantage of clones competed *in vitro* against the ancestral were evaluated with paired one-tailed distribution t test, with a significance defined as *P* value of <0.05.

Results and Discussion

Dynamics of adaptation of *E. coli* through changes in the frequency of a neutral marker

By following the evolution of a commensal *E. coli* strain in the mouse gut we have previously observed the emergence and spread of adaptive mutations within only 3 days of colonization (Barroso-Batista et al., 2014). The 1st step of adaptation consisted in the selective inactivation of the operon that allows *E. coli* to metabolize galactitol (Barroso-Batista et al., 2014). Distinct alleles conferring this phenotype and bearing a similar selective effect (7.5% benefit, per generation) recurrently emerged in all populations. Thus the distribution of fitness effects of mutations that constitute the 1st step of adaptation to the gut is a simple Dirac-Delta distribution. Using this experimental system, where neutral fluorescence markers can unravel the spread of adaptive mutations, we now study the nature of the 2nd step of adaptation. We followed the evolution of a clonal population that carries the first beneficial phenotype (inability to metabolize galactitol). In this population the phenotype is conferred by a single base pair insertion in the coding region of the *gatC* gene (which codes for a subunit of the galactitol transporter), thus preventing galactitol uptake. This clonal population was also made dimorphic for a neutral marker to allow determination of the emergence of further adaptive

changes (Hegreness et al., 2006). The dynamics of neutral marker frequency resulting from the 2nd step of adaptation are shown in Fig. 1a.

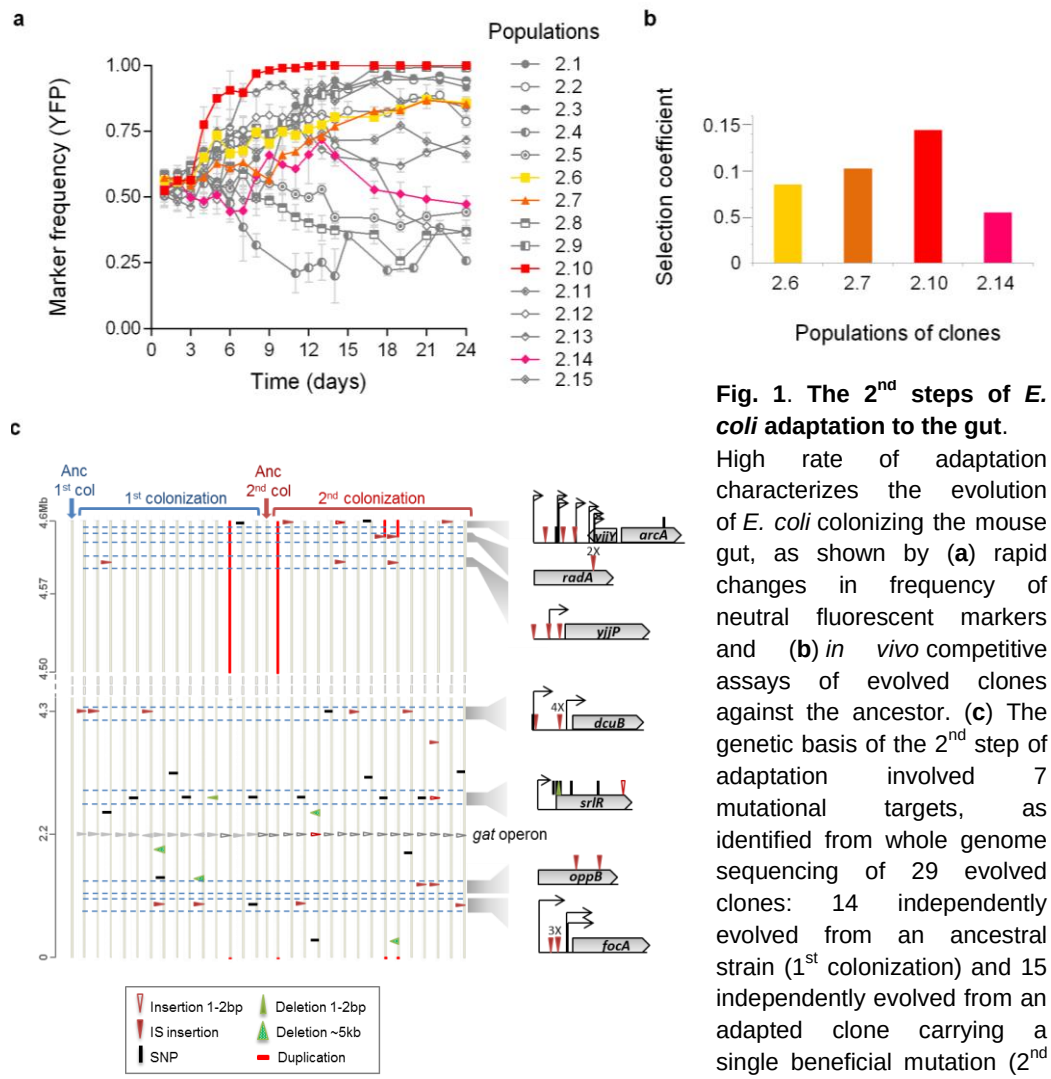


Fig. 1. The 2nd steps of *E. coli* adaptation to the gut. High rate of adaptation characterizes the evolution of *E. coli* colonizing the mouse gut, as shown by (**a**) rapid changes in frequency of neutral fluorescent markers and (**b**) *in vivo* competitive assays of evolved clones against the ancestor. (**c**) The genetic basis of the 2nd step of adaptation involved 7 mutational targets, as identified from whole genome sequencing of 29 evolved clones: 14 independently evolved from an ancestral strain (1st colonization) and 15 independently evolved from an adapted clone carrying a single beneficial mutation (2nd colonization).

These dynamics can be summarized in two effective evolutionary parameters (Barrick et al., 2010; Barroso-Batista et al., 2014; Hegreness et al., 2006), the mutation rate, U_e and the fitness effects of adaptive mutations, S_e. These quantitatively describe a simple adaptive process where new beneficial mutations emerge and spread in the populations (Hegreness et al., 2006). The estimates of the evolutionary parameters (U_e= 7.1 x 10^{-7}, S_e= 0.09) for the 2nd step of adaptation are remarkably similar to those previously obtained for the first adaptive

mutations (U_e= 7 x 10^{-7}, s_e= 0.075) (Barroso-Batista et al., 2014). This indicates that the subsequent steps of adaptation involve mutations occurring at similar rate and with similar selective strength. Direct *in vivo* competitions between populations of evolved clones and their ancestor further corroborate that the fitness benefit of the accumulated mutations is large, ranging from 4 to 14% fitness increase (Fig. 1b).

To query if a single or multiple genetic targets underlie the 2nd step of adaptation and to determine how repeatable evolution is, we performed whole genome sequencing (WGS) of 15 independently evolved clones (from day 24 of the colonization, see Fig. 1c and Supplementary Table 1). Seven new targets of mutation are recurrently detected, revealing their adaptive nature (Long et al., 2015). Taking the frequency of independent parallel emergence as a proxy for the probability of a mutational target being the second adaptive event, we find that the most frequent is *srlR*, a locus that codes for the repressor of the sorbitol operon (probability 24%). This is followed by *dcuB* and *focA* (probability 21% and 14%), which code for membrane transporters of C4-dicarboxylates (e.g. fumarate) and formate, respectively (Engel et al., 1994; Suppmann and Sawers, 1994), and then *arcA* (probability 14%), a dual transcriptional regulator mainly involved in governing the respiratory flexibility of *E. coli* (Jones et al., 2007, 2011). Another frequent mutation (detected in 14% of the clones) involved a large duplication (ranging from 34Kb to 157Kb). Two additional targets were observed less frequently: *yjjP* (in 10% of the clones) which codes for a membrane protein of unknown function (Stratmann et al., 2008) and *oppB* (in 7% of the clones) which codes for a component of the oligopeptide ABC transporter (Andrews et al., 1986). These targets indicate that *E. coli* is adapting by tuning the expression of a global respiratory regulator and changing its carbohydrate metabolism and transport. Consistent with this metabolic adaptation the competitive fitness of several gut adapted clones is increased when they are grown *in vitro* in media containing carbon sources known to be present in the intestine (Supplementary Table 2) in the presence or absence of oxygen. Half (53%) of the mutations comprising the second-step of adaptation were caused by IS insertions (Supplementary Table 1). Among these insertions, 78% are located in regulatory regions, suggesting that

mutations altering gene regulation are an important driver in the adaptation of these populations (Wittkopp and Kalay, 2012). The effects of IS insertions targeting the regulatory regions of *focA*, *dcuB*, *arcA*, *yjjY* and *yjjP* were determined by measuring gene expression of representative mutants evolved during the gut colonization. While IS insertions associated with the first adaptive step caused gene inactivation the transposition events comprising the 2nd step of adaptation modulate gene expression. This modulation was mostly positive but conditional on the amount of oxygen experienced during bacterial growth (Supplementary Fig. 1), suggesting that IS insertions contribute to adapt to environmental fluctuations likely to occur in the gut environment (Leone et al., 2015).

The order of adaptive steps reflects both their effects and mutation rate heterogeneity

The order of the mutational steps along an adaptive walk is influenced by the mutation rate, the fitness effects of mutations and/or possible epistatic interactions (Rokyta et al., 2005; de Visser and Krug, 2014). We investigated if these processes played a role in *E. coli*'s adaptation to the mouse gut. Using a fluctuation assay, we found evidence that the *gat* operon is a mutational hotspot with its functional inactivation occurring at a spontaneous rate of 10^{-5} (95% CI, [$4x10^{-4}$, $6.7x10^{-6}$]) (Supplementary Fig. 2) per genome per generation. This rate is two orders of magnitude higher than that of a random locus. This finding supports mutation rate differences as an important contributor for the *gat*-negative phenotype being the first adaptive event. However it does not exclude a role for direct or epistatic selection. Therefore we estimated the fitness effects of the 2nd step mutations in the wildtype (*gat*-positive) background, through competitive fitness assays against the ancestor (*gat*-positive) (Fig. 2).

We found substantial variation for the selective effects of these mutations, when compared with the values estimate for the first adaptive mutations (all with a similar selective effect (Barroso-Batista et al., 2014)). Most mutations showed a competitive advantage: 0.05, 0.06, 0.02, 0.02 and 0.1 for *focA*, *yjjP*, *dcuB* and *oppB* and *arcA*, respectively, though the selection coefficients (with one exception) were weaker than that for the *gat*-negative mutations. Remarkably the mutation

inactivating the *srlR* locus, which renders the bacteria constitutively expressing the sorbitol operon (Supplementary Fig. 1), steadily decreases in frequency (s_{srlR}=-0.03) during the first days of competition (Fig. 2) but it is able to be maintained at low frequency in the populations. This maintenance is suggestive of this locus being a target for balancing selection. To directly test for this form of selection occurring in the gut we performed competitive fitness assays at different starting frequencies. Indeed *srlR* mutants show a strong disadvantage when competing with the ancestral at high frequency but are able to be maintained at lower frequencies (Fig. 3a and b).

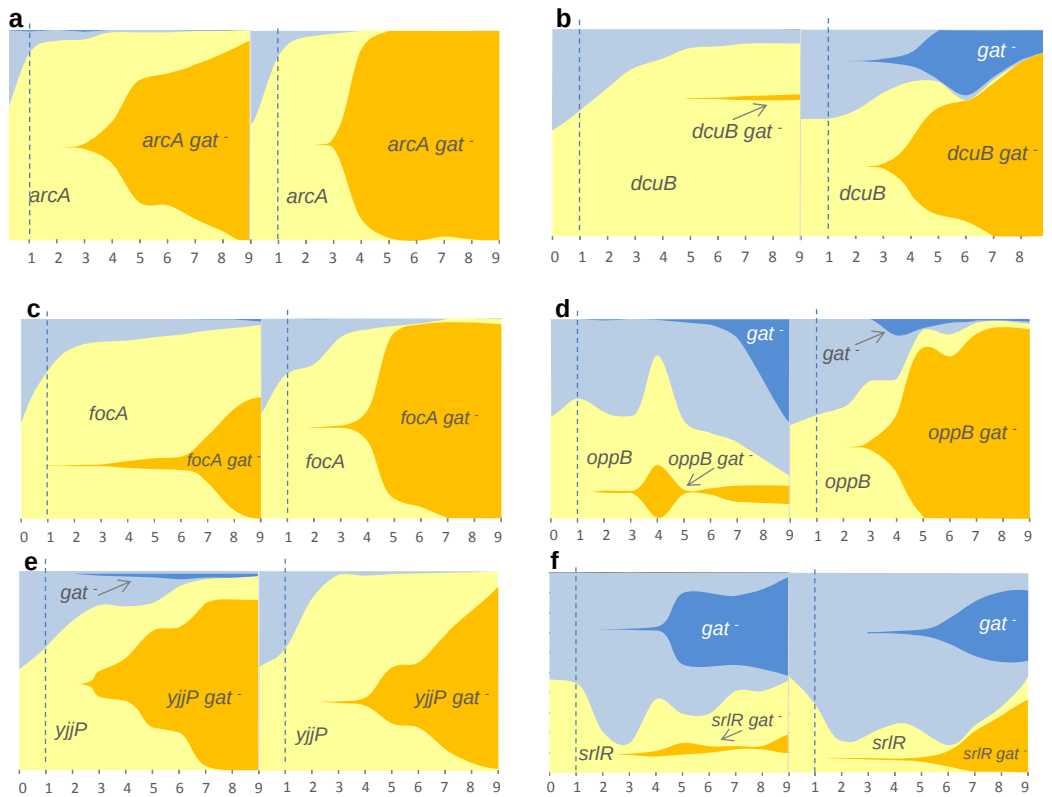


Fig. 2. The mutations occurring during the 2nd step of adaptation show strong beneficial effects in the wild type background (*gat*-positive).
Variable selection coefficients (considering the time interval in days marked by the dashed line), measured by competition of single mutants against the ancestral (n=2), are detected for the mutations in: (**a**) *arcA*, (**b**) *dcuB*, (**c**) *focA*, (**d**) *oppB*, (**e**) *yjjP* and (**f**) *srlR*). *gat*-negative mutations occur in both genetic backgrounds, indicating the absence of historical contingency along the adaptive walk.

Taken together, these results show that substantial variation in selection coefficients underlies the steps following the first adaptive event. In all

competitions performed we observed the appearance of the *gat*-negative phenotype on both ancestral and mutant backgrounds, including the *srlR* competitions where balancing selection and clonal interference could be simultaneously detected. This strongly suggests that little contingency and high repeatability accompanies the evolutionary path taken by *E. coli* during colonization.

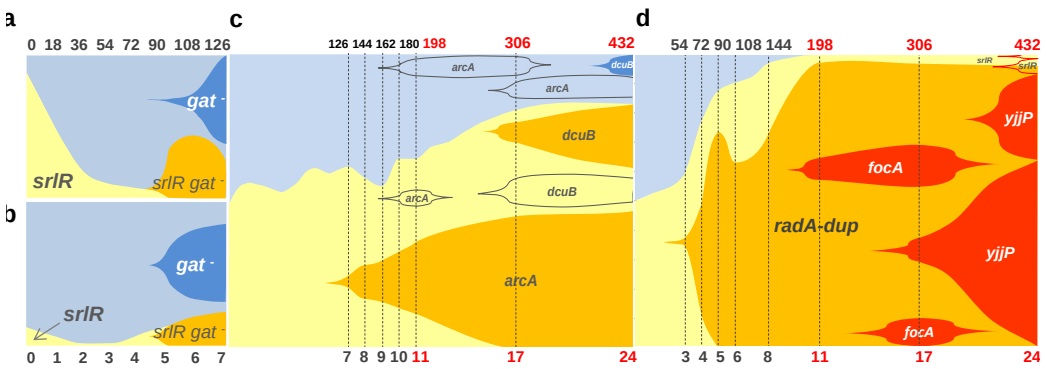


Fig. 3. Natural selection in three different forms explains *E. coli* variation in the mouse gut. Evidence for balancing selection, expressed as negative-frequency dependent selection, maintaining variation at the *srlR* locus: *srlR* alleles show (**a**) disadvantage when at high frequency but (**b**) are maintained at low frequency. Evidence for positive Darwinian selection in the gut: (**c**) clonal interference, where clones carrying different adaptive mutations at the same and in different loci compete for fixation in the gut; (**d**) periodic selection resulting in the rapid replacement of a clone, *i.e.* a hard selective sweep in which the haplotype *radA*-dup reaches close to 100% frequency in the population. Time is depicted in generations (top axis) or days (bottom axis).

Evidence for periodic selection and clonal interference driving intra-species variation

To monitor adaptation at the genome wide scale through time we analyzed levels of polymorphism in the evolving populations. Both WGS and typing of selected mutations were used to unravel the pervasiveness of clonal interference (Fig. 3c and d and Supplementary Table 3). Distinct alleles at each locus are segregating in these sampled populations: two alleles for *focA*, *yjjP*, *srlR*, three for *dcuB* and four for *arcA*, further supporting the adaptive nature of the identified targets and showing that the evolutionary process is not limited by mutation. An example of periodic selection, where a rapid and strong mutation (*radA* dup) sweeps close to fixation, is also revealed (Fig. 3d). The selective coefficient

estimated for this haplotype, while it is sweeping but before other mutations can be detected at high frequency, is $s_{rad\ dup}$ = 14% (Supplementary Fig. 3). This estimate is similar to the competitive advantage of the population where it emerged (see Fig. 1b, population 2.10). A large selective benefit associated with the *arcA* mutation is also estimated s_{arcA} = 9% (Fig. 3c and Supplementary Fig. 3), which is very similar to that measured in direct competition between this mutant and the ancestor of the first colonization (Fig. 2a). This suggests lack of epistasis between *arcA* adaptive mutations and the *gat* operon.

Together with our findings on frequency-dependent selection maintaining the *srlR* polymorphism (Fig. 3a and b), our evidence for period selection and clonal interference in evolving populations (Fig. 3c and d) show that these three forms of selection underlie the second steps of adaptation of *E. coli* to the mouse gut.

Re-acquisition of the *gat*–positive phenotype

The genetic basis of the 2nd steps of adaptation (Fig. 1c) revealed that a second mutation in *gatC*, able to reconstruct the opening frame of this previously pseudogenized gene, had occurred. This suggests that the unlikely re-gain of function could have re-evolved. Surprisingly this clone was able to metabolize galactitol. This indicates that fluctuating selection for and against this function characterizes the evolutionary process in the gut. If so, this phenotypic reversion should be both pervasive and fluctuate in frequency along the colonization time. Indeed the analysis of polymorphism of two populations revealed the recurrent emergence of the *gat*-positive phenotype (Fig. 4a and b).

At day 11 in population 2.14 (Fig. 4a) the *gat*-positive phenotype was observed in both fluorescent backgrounds but only one new *gat* allele could be detected by whole population sequencing (involving a 1bp deletion, as indicated in Supplementary Table 3). This demonstrates that not only population sequencing underestimates the amount of polymorphism but also that the *gat*-positive phenotype is able to rapidly invade a *gat*-negative population. The back-mutation observed in population 2.15 (Fig. 4b) further corroborates this conclusion demonstrating that the ancestral gene can invade a population where loss of

function at this locus had previously evolved. This genetic signature is compatible with strong fluctuating selection occurring in the gut of streptomycin-treated mice.

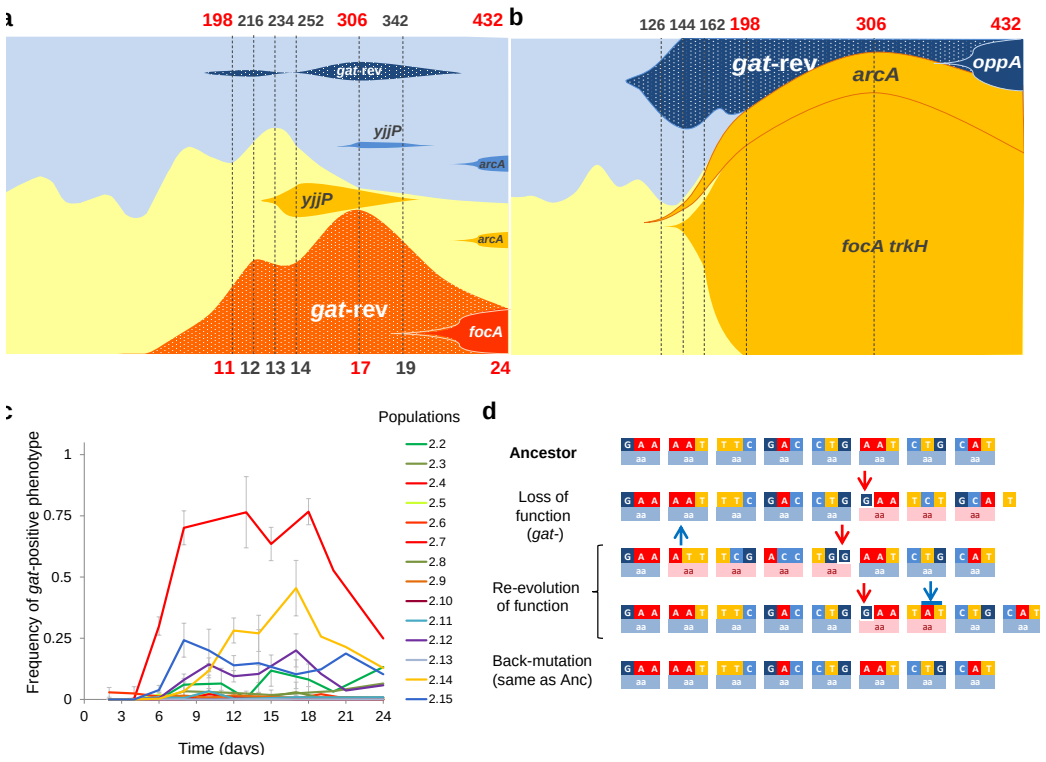


Fig. 4. Regain of a previously lost phenotype (*gat*-negative phenotype) with recurrent violation of Dollo's law both at genetic and phenotypic levels.
(**a**) and (**b**) Example of emergence of polymorphism in two evolving populations where regain of function is observed - *gat*-rev. (**c**) Reacquisition of the ancestral metabolic ability is pervasive across independently evolving populations and suggests that fluctuating selection is driving adaptation. (**d**) The genetic basis of the *gat*-reversion phenotype (*gat*-rev) involves both compensatory and back-mutation.

To enquire about the pervasiveness of reverse evolution and determine the strength of natural selection at restoring the ancestral phenotype, we tested all the remaining populations for the spread of a *gat*-positive phenotype. Consistent with a strong selective pressure to revert, 80% of the populations presented some level of re-evolution of the ancestral phenotype (Fig. 4c). From these, some showed evidence for fluctuating selection where the frequency of *gat*-positive bacteria can be as high as 75% at the intermediate time of the colonization process but then decreases at later time points (Fig. 4c, e.g. population 2.4 and 2.14). In other populations the increase in frequency is more modest but still strong fluctuations in the frequency of the phenotype are seen. Since the experiment started with a

clonal population fixed for a non-functional *gatC* gene, the process of reverse evolution occurred through *de novo* mutation and not from standing genetic variation. Indeed, genetic analysis of clones displaying the *gat*-positive phenotype showed that re-acquisition of this phenotype could occur due to different genetic mutations, including both compensatory or back-mutations (Fig. 4d).

This data points to a fitness landscape possibly characterized by fluctuating selection where opposite phenotypes can either alternate or coexist, which is coherent with the observed polymorphism for presence/absence of this operon in wild strains of *E. coli* (Woodward and Charles, 1983).

Overall we present quantitative estimates of the fitness effects of mutations within a natural ecosystem where both ecological and evolutionary processes occur and defy Dollo's law of evolutionary irreversibility. The data obtained establish that strong effect beneficial mutations exhibiting all classical forms of natural selection shape the genetic diversity of a commensal species of the mouse microbiota. They suggest that, even in mammalian hosts with identical genetic backgrounds and diets, reproducible adaptations emerge albeit with a significant level of host individuality.

Acknowledgements

The research leading to these results was funded by the European Research Council under the European Community' Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no 260421–ECOADAPT. Genome sequencing data have been deposited in the NCBI Read Archive, accession no. SRP063701.

References

Ahern, P.P., Faith, J.J., and Gordon, J.I. (2014). Mining the human gut microbiota for effector strains that shape the immune system. Immunity *40*, 815–823.

Andrews, J.C., Blevins, T.C., and Short, S.A. (1986). Regulation of peptide transport in Escherichia coli: induction of the trp-linked operon encoding the oligopeptide permease. J. Bacteriol. *165*, 428–433.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, 2006.0008.

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., and Kim, J.F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature *461*, 1243–1247.

Barrick, J.E., Kauth, M.R., Strelioff, C.C., and Lenski, R.E. (2010). Escherichia coli rpoB mutants have increased evolvability in proportion to their fitness defects. Mol. Biol. Evol. *27*, 1338–1347.

Barroso-Batista, J., Sousa, A., Lourenço, M., Bergman, M.-L., Sobral, D., Demengeot, J., Xavier, K.B., and Gordo, I. (2014). The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps. PLoS Genet. *10*, e1004182.

Barroso-Batista, J., Demengeot, J., and Gordo, I. (2015). Adaptive immunity increases the pace and predictability of evolutionary change in commensal gut bacteria. Nat. Commun. *6*, 8945.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. (1997). The complete genome sequence of Escherichia coli K-12. Science *277*, 1453–1462.

Darwin, C. (1859). On the Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life (J. Murray).

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci USA *97*, 6640–6645.

Engel, P., Krämer, R., and Unden, G. (1994). Transport of C4-dicarboxylates by anaerobically grown Escherichia coli. Energetics and mechanism of exchange, uptake and efflux. Eur. J. Biochem. *222*, 605–614.

Falush, D. (2009). Toward the Use of Genomics to Study Microevolutionary Change in Bacteria. PLOS Genet *5*, e1000627.

Greenblum, S., Carr, R., and Borenstein, E. (2015). Extensive strain-level copy-number variation across human gut microbiome species. Cell *160*, 583–594.

Hall, B.M., Ma, C.-X., Liang, P., and Singh, K.K. (2009). Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinforma. Oxf. Engl. *25*, 1564–1565.

Hegreness, M., Shoresh, N., Hartl, D., and Kishony, R. (2006). An equivalence principle for the incorporation of favorable mutations in asexual populations. Science *311*, 1615–1617.

Jones, S.A., Chowdhury, F.Z., Fabich, A.J., Anderson, A., Schreiner, D.M., House, A.L., Autieri, S.M., Leatham, M.P., Lins, J.J., Jorgensen, M., et al. (2007). Respiration of Escherichia coli in the mouse intestine. Infect Immun *75*, 4891–4899.

Jones, S.A., Gibson, T., Maltby, R.C., Chowdhury, F.Z., Stewart, V., Cohen, P.S., and Conway, T. (2011). Anaerobic respiration of Escherichia coli in the mouse intestine. Infect. Immun. *79*, 4218–4226.

Leone, V., Gibbons, S.M., Martinez, K., Hutchison, A.L., Huang, E.Y., Cham, C.M., Pierre, J.F., Heneghan, A.F., Nadimpalli, A., Hubert, N., et al. (2015). Effects of diurnal variation of gut microbes and high-fat feeding on host circadian clock function and metabolism. Cell Host Microbe *17*, 681–689.

Long, A., Liti, G., Luptak, A., and Tenaillon, O. (2015). Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nat. Rev. Genet. *16*, 567–582.

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. Nature *489*, 220–230.

Marshall, C.R., Raff, E.C., and Raff, R.A. (1994). Dollo's law and the death and resurrection of genes. Proc. Natl. Acad. Sci. U. S. A. *91*, 12283–12287.

Ning, Z., Cox, A.J., and Mullikin, J.C. (2001). SSAHA: a fast search method for large DNA databases. Genome Res. *11*, 1725–1729.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res. *29*, e45–e45.

Poulsen, L.K., Licht, T.R., Rang, C., Krogfelt, K.A., and Molin, S. (1995). Physiological state of Escherichia coli BJ4 growing in the large intestines of streptomycin-treated mice. J Bacteriol *177*, 5840–5845.

Rang, C.U., Licht, T.R., Midtvedt, T., Conway, P.L., Chao, L., Krogfelt, K.A., Cohen, P.S., and Molin, S. (1999). Estimation of growth rates of Escherichia coli BJ4 in streptomycin-treated and previously germfree mice by in situ rRNA hybridization. Clin. Diagn. Lab. Immunol. *6*, 434–436.

Reiner, A.M. (1977). Xylitol and D-arabitol toxicities due to derepressed fructose, galactitol, and sorbitol phosphotransferases of Escherichia coli. J. Bacteriol. *132*, 166–173.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Rokyta, D.R., Joyce, P., Caudle, S.B., and Wichman, H.A. (2005). An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nat. Genet. *37*, 441–444.

Stratmann, T., Madhusudan, S., and Schnetz, K. (2008). Regulation of the yjjQ-bglJ Operon, Encoding LuxR-Type Transcription Factors, and the Divergent yjjP Gene by H-NS and LeuO. J. Bacteriol. *190*, 926–935.

Suppmann, B., and Sawers, G. (1994). Isolation and characterization of hypophosphite-resistant mutants of Escherichia coli: identification of the FocA protein, encoded by the pfl operon, as a putative formate transporter. Mol. Microbiol. *11*, 965–982.

T. Conway, K. A. Krogfelt, and P. S. Cohen (2004). The Life of Commensal Escherichia coli in the Mammalian Intestine. In EcoSal - Escherichia Coli and Salmonella: Cellular and Molecular Biology, R. Curtiss III, ed. (Washington, DC: ASM Press), p.

de Visser, J.A.G.M., and Krug, J. (2014). Empirical fitness landscapes and the predictability of evolution. Nat. Rev. Genet. *15*, 480–490.

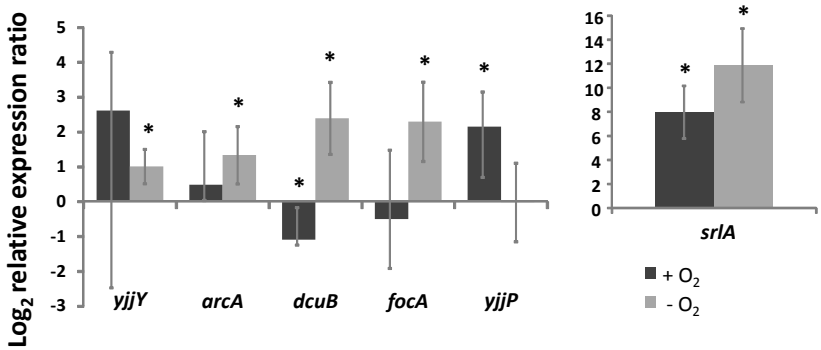
Waldor, M.K., Tyson, G., Borenstein, E., Ochman, H., Moeller, A., Finlay, B.B., Kong, H.H., Gordon, J.I., Nelson, K.E., Dabbagh, K., et al. (2015). Where Next for Microbiome Research? PLOS Biol *13*, e1002050.

Wilson, K. (2001). Preparation of genomic DNA from bacteria. Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al *Chapter 2*, Unit 2.4.

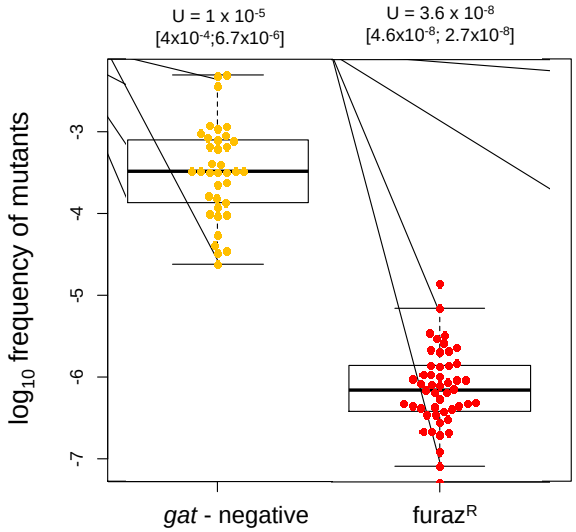
Wittkopp, P.J., and Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. *13*, 59–69.

Woodward, M.J., and Charles, H.P. (1983). Polymorphism in Escherichia coli: rtl atl and gat regions behave as chromosomal alternatives. J. Gen. Microbiol. *129*, 75–84.

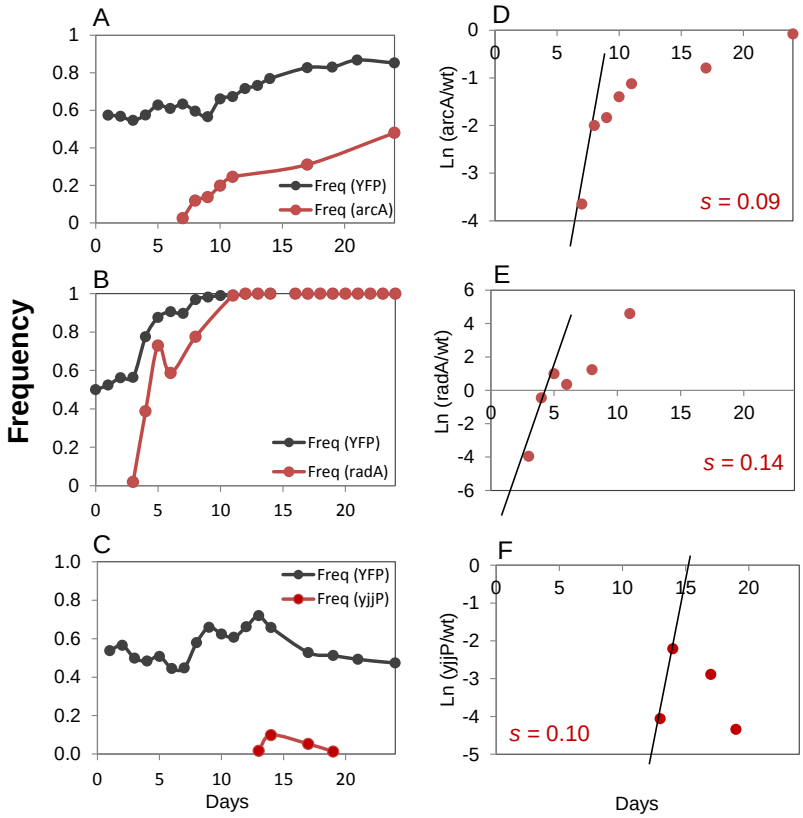
Supplementary Figures and Tables



Supplementary Fig. 1 - Transposable elements modulate gene expression.
The graph represents the log_2 relative expression rate of each gene in relation to the ancestral clone, obtained by RT-qPCR. The expression rate is above 0 for clones where the IS insertion caused an increase in gene expression (in relation to the ancestor); if this value is below 0 than the IS insertion decreased gene expression (in comparison to the ancestor). Error bars represent 2 times the standard error. Asterisks indicate significant differences at the level $P < 0.05$. Effects of IS insertions on gene expression were conditional on oxygen availability.



Supplementary Fig. 2 - The *gat* locus is a mutational hotspot.
The mean frequency of spontaneous *gat*-negative mutants is 6.9 x 10^{-4}, which is ~500 times higher than the frequency of spontaneous resistant to furazolidone (1.3 x 10^{-6}). After correcting for the difference in the mutational target size of both phenotypes (resistance to furazolidone is conferred by the inactivation of the *nfsA* gene, which is ~7 times smaller than the *gat*-operon) we still observed that the frequency of *gat*-negative phenotype is ~76 times higher than the frequency for furazolidone resistance. The estimated mutations rates (*U*) (per genome, per generation) and their respective confidence intervals, for each phenotype, are shown on the top.



Supplementary Fig. 3 - Selective coefficients (*s*) of three 2nd step mutations measured directly in the evolving populations.

These estimates correspond to a timeframe where no further mutations were detected (or were at very low frequency), that is before clonal interference affected the estimates (D-E). (A-C) correspond to populations 2.7, 2.10 and 2.14. All three mutations were detected in the YFP background.

Supplementary Table 1 - Identity of the mutations detected in the sequenced clones.
Mutations were identified by comparison with the ancestor. The ancestor differs from the reference MG1655 strain (NCBI Reference Sequence: NC_000913.2) in the positions reported in (Barroso-Batista et al., 2014) for the 0YFP clone plus 1pb insertion (+C) in the coordinate 2172079 (*gatC*, coding (222/1356 nt)). Mutations in intergenic regions, as for example *yjjY/yjtD,* list the two flanking genes. Numbers in the annotation row represent nucleotides relative to each of the neighboring genes, where + indicates the distance downstream of the stop codon of a gene and - indicates the distance upstream of the gene, that is relative to the start codon. Genes inside brackets means that the mutation is localized within the gene. Single nucleotide substitutions (SNP) are represented by an arrow between the ancestral and the evolved nucleotide. Whenever a SNP gives rise to a non-synonymous mutation the amino acid replacement is also indicated. One asterisk means that the corresponding SNP originated a STOP codon. The symbol Δ means a deletion event and a + symbol represents an insertion of the nucleotide that follows the symbol. IS denotes the abbreviation of insertion sequence element at the indicated position.

Clone	Genome Position	Gene	Mutation	Annotation
16YFP	4500113	*[nhaA-IS4]*	2x157093 bp	large duplication
17YFP	4638771	*yjjY/yjtD*	IS5 Ins	intergenic (+206/ -191)
18YFP	953901	*focA/ycaO*	IS5 Ins	intergenic (-212/+191)
	2827229	*srlR*	G→A	G54D (GGC→GAC)
19CFP	317172	*ykgB*	IS2 Ins	coding (368- 372/594 nt)
	2556721	*intZ–[eutA]*	Δ6790 bp	
	2172079	*gatC*	+TC	coding (222/1356 nt)
20CFP	4347121	*dcuB/dcuR*	A→T	intergenic (- 354/+217)
21YFP	4601066	*yjjP/yjjQ*	IS2 Ins	intergenic (- 185/ -430)
	4638684	*yjjY/yjtD*	+TTAT	intergenic (+119/ -281)
22YFP	4347105	*dcuB/dcuR*	IS2 Ins	intergenic (-338/+229)
23CFP	3174801	*cpdA*	G→A	Q19* (CAA→TAA)
	4638719	*yjjY/yjtD*	C → A	intergenic (+154/ -246)
24YFP	4625061	*radA*	IS1 Ins	coding (1127- 1135/1383 nt)
	2827492	*srlR*	G → A	G142S (GGC→AGC)
	4625061	*[IS1-radA]*	2x 34425 bp	large duplication
25YFP	279155	*insX–insA*	Δ11,471 bp	between IS1
	4600992	*yjjP/yjjQ*	IS1 Ins	intergenic (- 111/- 500)
	4625061	*radA*	IS1 Ins	coding (1127- 1135/1383 nt)
	4625061	*[IS1-radA]*	2x 34425 bp	large duplication
26YFP	1859452	*ydjL/yeaC*	G → T	intergenic (- 96/+274)
	4346885	*dcuB/dcuR*	IS5 Ins	intergenic (- 118/+450)
27CFP	1301748	*oppB*	IS5 Ins	coding (826-829/921 nt)
	2827045	*gutM/srlR*	A → T	intergenic (+43/- 24)
28YFP	1301338	*oppB*	IS186 Ins	coding (416-421/921 nt)
	2827745	*srlR*	+T	coding (677/774 nt)
	3805903	*waaQ*	IS1 Ins	coding (211-219/1035 nt)
29CFP	4638628	*yjjY/yjtD*	IS2 Ins	intergenic (+63/- 333)
30YFP	9531240	*focA/ycaO*	IS1Ins	intergenic (- 174/+224)
	4031240	*trkH*	G → A	G275R GGG→AGG

Supplementary Table 2 - Phenotypic characterization of gut adapted clones.
In vitro competitive fitness assays of evolved clones against the ancestral shows increased fitness when growing under presence or absence of oxygen in several carbon sources present in the mouse intestine. Numbers represent the selective coefficient (*s*) of each clone in relation to the reference strain (see Material and Methods for details on how we performed the competitions and estimated *s*). Yellow, green and red shading means that *s* is: not significantly different, higher or lower than the ancestral clone, respectively. The column "Genotype" depicts the genes affected by mutation in each clone. Clones are ordered by genotype similarity.

		Anaerobiosis	Aerobiosis					
Clone	Genotype	Mix 0.5%	Mix 0.5%	Sorbitol	Gluconate	Mannose	Glucoronate	Ribose
5YFP	*gatZ*	0.18	-0.02	0.13	-0.04	0.11	-0.07	0.13
13YFP	*Duplication*	-0.01	0.03	0.00	0.01	0.02	0.01	-0.05
16YFP	*Duplication*	0.07	0.05	0.04	0.12	-0.06	0.14	0.56
24YFP	*radA*	0.05	0.00	0.09	-0.04	-0.26	-0.13	-0.17
25YFP	*del11kb yjjP radA*	0.07	0.02	0.06	0.05	-0.06	0.01	0.52
4YFP	*yjjP yffN*	0.12	0.02	0.00	-0.04	0.02	0.02	0.36
21YFP	*yjjP arcA*	0.24	0.03	0.02	-0.07	-0.08	-0.11	0.08
29YFP	*arcA*	-0.02	0.04	0.07	-0.06	-0.07	0.07	0.02
14CFP	*arcA*	0.10	0.05	0.01	0.17	0.17	0.14	-0.11
17YFP	*arcA*	0.20	-0.01	-0.08	0.08	0.12	0.01	-0.07
23CFP	*cpdA arcA*	-0.19	0.01	-0.01	0.06	0.16	0.07	0.25
3YFP	*dcuB*	0.06	0.04	0.03	-0.09	0.01	0.00	0.29
7YFP	*dcuB*	0.06	0.01	-0.03	-0.03	-0.01	0.04	0.25
2CFP	*dcuB*	-0.02	0.02	0.09	0.14	0.18	0.17	0.20
20CFP	*dcuB*	-0.05	0.01	0.05	0.19	0.11	0.25	0.52
22YFP	*dcuB*	0.15	0.02	0.07	-0.08	-0.08	-0.08	0.14
26YFP	*ydjL/yeaC dcuB*	0.11	0.03	0.04	0.06	-0.11	0.02	0.60
6YFP	*srlR*	0.06	0.03	0.11	-0.09	-0.18	-0.07	0.22
10YFP	*srlR*	0.07	0.04	0.10	-0.08	-0.08	-0.04	0.16
12CFP	*srlR*	0.09	0.05	0.17	0.14	0.02	0.18	0.08
15CFP	*dmsC srlR*	0.19	0.06	0.17	0.15	-0.03	0.18	0.08
27CFP	*oppB srlR*	-0.07	0.02	0.06	0.07	0.11	0.10	-0.16
28YFP	*oppB srlR rfaQ*	-0.02	0.02	0.12	0.11	-0.21	0.14	0.16
18YFP	*focA/ycaO + srlR*	0.51	0.02	0.22	0.00	-0.29	0.08	0.01
30YFP	*focA trkH*	0.03	0.00	0.25	0.26	-0.04	0.24	0.64
11CFP	*del5kb focA*	-0.02	0.05	0.13	0.15	0.18	0.20	0.15
8YFP	*ydaV del5kb focA*	0.06	0.04	-0.04	-0.01	-0.77	0.10	0.16
9YFP	*garK*	0.08	0.05	0.00	0.00	0.03	0.06	0.02
19CFP	*ykgB del7kb*	-0.01	0.05	0.09	0.19	0.12	0.21	0.45

Supplementary Table 3 - Genetic and phenotypic polymorphism in four evolved populations. Polymorphism detected by a phenotypic assay is highlighted in bold. Mutations were detected by whole genome sequencing of samples of each population at different time points (underlined) or by target sequencing (shown in black). In addition, the frequency of the *gat*-rev phenotype in each of the fluorescence backgrounds was estimated by plating in selective media (shown in bold). The full data is graphically represented in Fig. 3c and d and 4a and b.

Pop.	Genome position	Gene	Mut.	Annotation	Frequency of mutation along time (days)						
					7	8	9	10	11	17	24
2.7	2858252	*pphB*	+G	coding (471/657 nt)					0.04	0	0
	4346885	*dcuB/ dcuR*	IS5 Ins	intergenic (-123/+453)					0	0.1	0.09
	4347105	*dcuB/ dcuR*	IS2 Ins	intergenic (-338/+233)					0	0.06	0.22
	4347126	*dcuB/ dcuR*	IS2 Ins	intergenic (-359/+212)					0	0	0.08
	4638684	*yjjY/ yjtD*	+TTA T	intergenic (+119/-281)					0.07	0	0
	4638673	*yjjY/ yjtD*	IS5 Ins	intergenic (+108/-292)					0	0.06	0.07
	4638771	*yjjY/ yjtD*	IS5 Ins	intergenic (+206/-194)	0.03	0.12	0.14	0.2	0.26	0.49	0.48
	4638805	*yjjY/ yjtD*	IS1 Ins	intergenic (+240/-160)					0.07	0.11	0

Pop.	Genome position	Gene	Mut.	Annotation	Frequency of mutation along time (days)					
					1	2	3	4	5	6
2.10	770545	*mngB/ cydA*	IS5 ins	intergenic (+711/-136)						
	953890	*focA/ ycaO*	IS1 ins	intergenic (-201/+205)						
	953892	*focA/ ycaO*	IS5 ins	intergenic (-203/+203)						
	1314238	*yciG/ trpA*	IS5 ins	intergenic (-17919/+202)						
	2827162	*srlR*	A→G	T32A (ACA→GCA)						
	2827271	*srlR*	+T							
	4600992	*yjjP/ yjjQ*	IS1 ins	intergenic (-111/-508)						
	4601134	*yjjP/ yjjQ*	IS5 ins	intergenic (-256/-363)						
	4625061	*radA*	IS1 ins	coding (1127/1383 nt)	0	0	0.02	0.39	0.73	0.59

					8	11	17	24
2.10	770545	*mngB/ cydA*	IS5 ins	intergenic (+711/-136)		0	0	0.09
	953890	*focA/ ycaO*	IS1 ins	intergenic (-201/+205)		0.04	0.08	0
	953892	*focA/ ycaO*	IS5 ins	intergenic (-203/+203)		0	0.05	0
	1314238	*yciG/ trpA*	IS5 ins	intergenic (-17919/+202)		0.13	0	0
	2827162	*srlR*	A→G	T32A (ACA→GCA)		0	0	0.06
	2827271	*srlR*	+T			0	0	0.02
	4600992	*yjjP/ yjjQ*	IS1 ins	intergenic (-111/-508)		0	0	0.34
	4601134	*yjjP/ yjjQ*	IS5 ins	intergenic (-256/-363)		0	0.13	0.7
	4625061	*radA*	IS1 ins	coding (1127/1383 nt)	0.77	0.97	0.98	0.95

Pop.	Genome position	Gene	Mut.	Annotation	Frequency of mutation along time (days)					
					2	4	6	8	10	11
2.14	953892	*focA/ ycaO*	IS5 Ins	intergenic (-203/+203)						0
	1740247	*cfa*	Δ1 bp	coding (811/1149 nt)						0
		***gat*-rev phen.**			**0**	**0**	**0**	**0.3**	**0.12**	
		***gat*-rev phen.**			**0**	**0**	**0**	**0**	**0.01**	
	2172089	*gatC*	Δ1 bp	coding (212/1356 nt)						0.05
	3718679	hok/ insJ	C→T	intergenic (-56/-24)						0
	4601151	*yjjP/ yjjQ*	IS5 Ins	intergenic (-270/-349)						0
		yjjP/ yjjQ	IS Ins							0
		yjjP/ yjjQ	IS Ins							0
	4638771	*yjjP/ yjjQ*	IS5 Ins	intergenic (+206/-194)						0

					12	13	14	17	24
2.14	953892	*focA/ ycaO*	IS5 Ins	intergenic (-203/+203)				0	0.17
	1740247	*cfa*	Δ1 bp	coding (811/1149 nt)				0	0.06
		***gat*-rev phen.**			**0.28**		**0.27**	**0.45**	**0.13**
		***gat*-rev phen.**			**0.02**		**0**	**0.05**	**0**

	2172089	*gatC*	Δ1 bp	coding (212/1356 nt)				0.13	0.02
	3718679	hok/ insJ	C→T	intergenic (-56/-24)				0	0.07
	4601151	*yjjP/ yjjQ*	IS5 Ins	intergenic (-270/-349)	0	0.02	0.1	0.05	0
		yjjP/ yjjQ	IS Ins					0.02	0
		yjjP/ yjjQ	IS Ins					0	0.06
	4638771	*yjjP/ yjjQ*	IS5 Ins	intergenic (+206/-194)				0	0.06

Pop.	**Genome position**	**Gene**	**Mut.**	**Annotation**	**Frequency of mutation along time (days)**					
					6	**7**	**8**	**9**	**10**	**11**
2.15	953863	*focA/ ycaO*	IS1 Ins	intergenic (-174/+224)		0	0.04	0.2		0.61
	1268674	ldrA/ ldrB	G→T	intergenic (-176/+252)						0.05
	1296579	*adhE*	Δ3 bp	coding (764-766/2676 nt)						0
	1324623	*yciQ*	Δ36 bp	coding (1854-1889/1896 nt)						0
	1299449	*oppA*	IS5 Ins	coding (-222-247/1632 nt)						0
		***gat*-rev phen.**			**0**		**0**		**0**	
		***gat*-rev phen.**			**0.04**		**0.24**		**0.2**	
	2172078	*gatC*	Δ1 bp	coding (223/1356 nt)						0.18
	3506347	*php*	Δ1 bp	coding (266/879 nt)						0
	3925453	asnA	Inv 15bp	coding (276-291/993 nt)						0.05
	4031240	*trkH*	G→A	G25R (GGG→AGG)						0.6
	4638637	*yjjY/ yjtD*	IS5 Ins	intergenic (+72/-325)		0.01	0.01	0.04		0.11

					12	**14**	**17**	**19**	**21**	**24**
2.15	953863	*focA/ ycaO*	IS1 Ins	intergenic (-174/+224)			0.84			0.77
	1268674	ldrA/ ldrB	G→T	intergenic (-176/+252)			0			0
	1296579	*adhE*	Δ3 bp	coding (764-766/2676 nt)			0			0.5

	1324623	*yciQ*	Δ36 bp	coding (1854-1889/1896 nt)			0			0.16
	1299449	*oppA*	IS5 Ins	coding (-222-247/1632 nt)			0			0.21
		***gat*-rev phen.**			**0**	**0**	**0**	**0**	**0**	**0**
		***gat*-rev phen.**			**0.1**	**0.14**	**0.06**	**0.14**	**0.12**	**0.16**
	2172078	*gatC*	Δ1 bp	coding (223/1356 nt)			0.03			0.18
	3506347	*php*	Δ1 bp	coding (266/879 nt)			0.04			0
	3925453	asnA	Inv 15bp	coding (276-291/993 nt)			0			0
	4031240	*trkH*	G→A	G25R (GGG→AGG)			0.84			0.73
	4638637	*yjjY/ yjtD*	IS5 Ins	intergenic (+72/-325)			0.13			0.19

CHAPTER 5

Discussion

As a transversal subject to all fields of biology, the adaptation of organisms to the environment remains one of the most interesting and important subjects whose mechanisms are not yet completely understood. Despite the multiplicity of *in vitro* studies that contributed considerably to elucidate the details of the adaptive process in those conditions, much less is known regarding the evolution of populations inhabiting natural environments. In particular, the way diversity is generated and maintained, or how the different environmental factors impact the adaptation of organisms, are points which relevance in natural habitats have been poorly addressed. With the rapid and recent advances in genetic tools, sequencing analysis and model organisms, it is now possible to conduct experiments that attempt to answer these questions in controlled yet close to natural conditions.

The present work focuses on the adaptation of a commensal species to one of the most complex and relevant natural environments, the mammalian gut. As a within-host environment, the gut represents a complex habitat influenced by multiple factors, but the dynamics of commensal microbial adaptation to this environment are still poorly known. In **Chapter 2** we followed the adaptation of *E. coli* to the gut of streptomycin-treated mice and characterized the evolutionary process in this habitat. Similarly to what has been observed in many in *vitro studies*, clonal interference between beneficial mutations was shown to play a major role in adaptation of *E. coli* to the mouse gut. The abundance of presumably adaptive mutations identified by WGS suggests that in this environment, mutation is a non-limiting process, contributing to the maintenance of polymorphisms in the evolving populations. In addition to clonal interference, other forms of selection, further elucidated in **Chapter 4**, were shown to coexist in a natural habitat. These include the classic selective sweeps and frequency-dependent interactions between different genotypes. Another observation common to *in vitro* studies (Hindré et al., 2012) was the high level of parallelism at the genetic level detected in independently-evolving lineages of *E. coli*, a remarkable observation in such a complex environment.

Generally, the majority of mutations identified by WGS were common to the several evolution experiments (the two bouts of adaptation in healthy mice and the evolution in immune-compromised animals) and targeted mainly genes related

with metabolism and respiration. These observations emphasize the hypothesis that nutrition in the intestine is one of the major selective pressures that *E. coli* adapts to when colonizing the streptomycin-treated mouse gut. Our findings thus generally agree with others (Leatham et al., 2005; Leatham-Jensen et al., 2012; Lescat et al., 2016) that described the selection of mutations conferring the ability to grow better in different carbon sources present in the gut. Conversely, no overlap for the genetic targets of adaptation at the gene level was observed between our studies and others. These probably reflect the specific conditions of each experiment, which include the genetic background of the bacterial strain used, the mouse diet or the microbiota composition. Nevertheless, we did find mutations in genes that others studies showed to be important for the colonization of the mouse gut by *E. coli* (e.g., *kdgR*, metabolism of gluconate (Chang et al., 2004; Leatham et al., 2005)). Interestingly, many mutations were found in intergenic regions or in genes involved in the regulation of gene expression (such as repressors) emphasizing the importance of the regulation of bacterial metabolic processes. This phenomenon was particularly evident in mutations characterizing the second steps of adaptation to the gut (**Chapter 4**), where most led to an increase in gene expression in an oxygen-dependent manner.

Chapter 3 addresses the importance of a component of the environment to the evolution of a microbial species, specifically the impact of the host adaptive immune system on the adaptation of *E. coli* to the mouse gut. While the role of immunity in the case of bacterial infections is a subject fairly explored, the same is not true in commensal species that inhabit the gut. The importance of the interactions between immune system and microbes is in fact even more relevant considering the co-evolution process that is thought to occur between microbiota and the host immune system (Ley et al., 2008). Although an assessment of the co-evolution hypothesis between host and microbes requires the reciprocal analysis of both partners, determining the impact of the host on the evolution of the microbe, as done here, still provides valuable information to elucidate the complex co-evolution process.

Up to this day, our understanding of the interaction between the adaptive immune system of the host and commensal *E. coli* remains limited. Differentiation

of Tregs has been shown to occur upon gut colonization of GF mice with a human-isolated *E. coli* strain (Faith et al., 2014) and production of *E. coli*-specific IgA has been detected in mice colonized with a lab strain of *E. coli* (Hapfelmeier et al., 2010). However, both these findings were reported exclusively in monocolonized mice, where stimulation of the host immunity may be facilitated possibly due to increased microbial contact with the host epithelium. Thus, to my knowledge, the ability of the adaptive immune system to recognize and respond directly to a colonizing *E. coli* strain (either by IgA production or expansion of T cell subsets) in the presence of a complex microbiota, as is the case of streptomycin-treated mice, has not been established. We could not find evidence for a direct effect of the adaptive immune system on *E. coli* evolution in our experiments, at least during the time period studied (24 days). While mutations were found differentially represented depending on host background, none of the genetic targets identified has been described as related directly with immune functions of the host. Furthermore, the selective strength of a beneficial mutation shown to be more variable and on average smaller in immune-compromised animals displayed a similar effect in immune-compromised or immune-competent GF mice.

We did however find a major indirect effect of adaptive immunity on *E. coli* evolution and physiology by modulation of the gut microbiota that interacts with *E. coli*. First, our results of an altered microbiota composition in immune-compromised animals are in agreement with several other studies that attributed to the adaptive immune system the function of maintaining a diverse gut microbiota (Kawamoto et al., 2014). Second, in streptomycin-treated animals colonized with *E. coli*, our analysis revealed differences in the relative abundances of presumably metabolically relevant genera between immune-compromised and immune-competent hosts. Considering that these bacterial groups (*Bacteroides*, for example) likely interact with *E. coli*, a change in the abundance of these species is expected to impact the metabolic environment of the gut that *E. coli* adapts to. Supporting this conjecture, previous studies have shown the importance of microbe-microbe interactions in the mouse gut, through the establishment of complex cross-feeding interactions (Conway and Cohen, 2015; Ferreyra et al., 2014). For example, *E. coli*, which lacks mucolytic activity, is able to colonize the

outer mucus layer in the mouse gut by metabolizing simpler carbon sources that are the product of metabolism of other bacterial organisms, such as *B. thetaiotaomicron* (Li et al., 2015). This metabolic impact is further supported by the differences in abundance of mutations related with bacterial metabolism (e.g., *dcuB*), when comparing selection in immune-compromised or immune-competent mice.

Overall, the slower rate of adaptation detected in immune-compromised compared to immune-competent animals suggests that the selective pressures are on average stronger in the latter than in the former. Likewise, our estimates for the growth rate support the idea that the gut of immune-compromised animals represents a more benign environment, where adaptation of *E. coli* occurs at a slower pace. An interesting observation is that despite differences in growth rate between the two hosts, the total loads of fecal *E. coli*, as a proxy for population size in the gut, were not significantly different. An increased growth rate may reflect a weakened control of bacterial growth or inability to mount an appropriated direct or indirect immune response. In fact, it has been shown that in immune-compromised *Rag* mice innate immune mechanisms, such as secretion of AMPs, are upregulated, presumably to compensate for the lack of the adaptive immune responses in the gut mucosa (Korn et al., 2014; Vaishnava et al., 2011). Thus, it is possible that while growing faster in the gut of immune-compromised animals, *E. coli* is also more exposed to other stresses, preventing overt expansion of the population. An alternative explanation is a faster gastrointestinal transit in *Rag2*$^{-/-}$ animals, which would require a faster growth rate of *E. coli* to avoid elimination from the gut. A similar phenomenon has been observed in bacterial populations evolving in chemostats, a device used to simulate the mammalian gut (Mason and Richardson, 1982). When growing in continuous cultures, populations adapt their division rate according to the flow of the chemostat (Gresham and Hong, 2015). Interestingly, increased rates of epithelial turnover were described in the gut of mice lacking B and T cells (Nishiyama et al., 2002), which may support this hypothesis.

The growth rate of *E. coli* in the mouse gut is a topic that has been addressed by several authors (Poulsen et al. 1995; Rang et al. 1999; Li et al. 2015; Myhrvold

et al. 2015). Whereas there is a general agreement that the maximum growth takes place in the mucus layer overlaying the gut epithelium, estimates for this parameter vary depending on the study. These differences may reflect the genetic background of the bacterial strain, the microbiota composition of the host or the methodologies used. These techniques are either based on the dilution of a cell marker over time (e.g., fluorescence, radioactivity) or on the correlation between growth rate and another parameter (such as amount of rRNA). While the second method is an indirect inference, the first method only allows to determine the growth rate over a short number of bacterial divisions. For example, a division rate of 3 hours in the mucus layer was reported for a lab strain *E. coli* in the first 24 hours of colonization of GF mice (Li et al., 2015). These estimates were based on dilution of fluorescence or loss of radioactivity techniques. A similar measurement was obtained for the fecal populations of a mouse-isolated strain in the first 12 hours of colonization of mice pre-treated with streptomycin (Myhrvold et al., 2015). On the other hand, studies based on the correlation between growth rate and cellular ribosomal contents reported a growth rate ranging from 70 to 90 minutes for a rat-isolated strain colonizing the gut of GF or streptomycin-treated mice over a period of 6 days (Rang et al., 1999). Although these measurements were performed in bacteria isolated from fecal samples, a comparable rate of 30 to 80 minutes was reported in *E. coli* isolated from the mucus layer over 20 days of colonization in streptomycin-treated mice (Poulsen et al., 1995). Our estimates for the generation time of *E. coli* in the gut of WT mice (76 minutes) are thus in agreement with these previous studies that were performed with the same methodology.

In **Chapter 4** we explored the second steps of adaptation of *E. coli* to the mouse gut, by following the evolution of a strain harboring a mutation in a gene of the *gat* operon. Mutations in this operon invariably lead to the *gat*-negative phenotype, which was selected in all the lines of the first evolution experiment. Supporting the strong beneficial effect associated with this phenotype, competitive fitness assays of gat mutants against the ancestral strain showed an advantage of the mutants. However, complete fixation of the phenotype (at least to our limit of detection) was only observed in some lines of the first evolution experiment.

Moreover, several lines displayed fluctuations in the frequency of this phenotype. While these observations could be explained by the interference of other beneficial mutations, an alternative scenario may include fluctuations in the environmental conditions promoting the maintenance of the *gat* polymorphism. In fact, the re-evolution of the ancestral phenotype observed in the second evolution experiment strongly supports this hypothesis. Despite starting with a clonal phenotype (and genotype), shown to be adaptive in that environment, most of the populations displayed phenotypic reversion, with variable genetic basis. Genetic reversions have rarely been observed in previous studies of experimental evolution, and when observed the initial mutation typically carried a fitness cost, such as mutations conferring antibiotic resistance (Gifford and MacLean, 2013). In addition, reversions usually occur in an environment where the initial selective pressure (for example antibiotic) is no longer relevant. Cases like the one we report here, where a previously selected phenotype reverts to the ancestral state in a similar environment have not been previously described. Interestingly, the dynamics of reversion we report here also display large oscillations, compatible with fluctuating selective pressures associated with this phenotype throughout the second evolution experiment. However, the advantage conferred by a functional or inactivated *gat* operon is not yet completely clear. As an operon involved in the metabolism of galactitol it may confer a metabolic advantage in the metabolism of other carbon sources. Alternatively, pleiotropic effects affecting other cellular functions cannot be discarded, such as resistance to stress factors. Trade-off mechanisms between nutritional ability and stress resistance have been shown to occur in populations of *E. coli* adapting to the gut of streptomycin-treated and GF mice, thus maintaining in the populations different types of mutations (De Paepe et al., 2011; Leatham-Jensen et al., 2012). A similar scenario may occur, although as an extreme, in the case of the *gat* operon. Thus, depending on the environmental pressures, an intact or mutated gat operon could provide an advantage, resulting in its strong selection or counter-selection. In addition, the evidence that the *gat* locus is a mutational hotspot, at least to inactivate the operon, may facilitate this switch between the two states.

The global observation from the multiplicity of studies in experimental evolution performed over the years consists in a deceleration in the rate of adaptation over time, with fixation of mutations with increasingly smaller effects (Dettman et al., 2012). The same phenomenon has been observed over successive steps of adaptation of the fungus Aspergillus to the lab (Schoustra et al., 2009). Interestingly, the fitness dynamics of *E. coli* adaptation in Lenski's long term evolution experiment were best explained by a model predicting a deceleration in fitness improvement but without an upper limit, i.e., no fitness limit (Lenski et al., 2015; Wiser et al., 2013). In our second evolution experiment, while the majority of selective effects estimated for the parallel mutations were smaller than that observed for mutations targeting genes from the *gat* operon, some were shown to confer an identical strong beneficial effect. Globally, these observations may suggest that the potential for adaptation of microbial populations, especially in complex environments, may be greater than previously thought.

The present thesis provides novel insights into the evolution of a commensal species when colonizing the mammalian gut and the influence of a major host factor, host immunity, in this process. We show that over a period of one month, an extraordinary amount of variation can be created *de novo* and maintained in evolving populations of *E. coli*, through multiple selection regimes. We further demonstrate the pervasiveness of clonal interference and parallelism in this natural environment. This extensive level of intra-species diversity here demonstrated reveals an additional layer of complexity when addressing microbe-microbe and host-microbe interactions in the gut environment. Importantly, this data should be integrated with previous and future studies focusing on the gut microbiota, to better understand the complex network of species, genomes and metabolites that characterize this complex environment. The results here reported on the indirect contribution of host adaptive immunity to bacterial evolution provide important information to elucidate the complex interactions between immune system of the host and commensal microbes. In a medical context, this data emphasizes the importance of personalized medical therapies for the treatment of

immune-compromised patients, taking into account microbiota composition and diversity.

References

Chang, D.E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., Anderson, A.B., Grissom, J.E., Laux, D.C., Cohen, P.S., et al. (2004). Carbon nutrition of Escherichia coli in the mouse intestine. Proc Natl Acad Sci USA *101*, 7427–7432.

Conway, T., and Cohen, P.S. (2015). Commensal and Pathogenic Escherichia coli Metabolism in the Gut. Microbiol. Spectr. *3*.

De Paepe, M., Gaboriau-Routhiau, V., Rainteau, D., Rakotobe, S., Taddei, F., and Cerf-Bensussan, N. (2011). Trade-off between bile resistance and nutritional competence drives Escherichia coli diversification in the mouse gut. PLoS Genet *7*, e1002107.

Dettman, J.R., Rodrigue, N., Melnyk, A.H., Wong, A., Bailey, S.F., and Kassen, R. (2012). Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. Mol. Ecol. *21*, 2058–2077.

Faith, J.J., Ahern, P.P., Ridaura, V.K., Cheng, J., and Gordon, J.I. (2014). Identifying gut microbe-host phenotype relationships using combinatorial communities in gnotobiotic mice. Sci. Transl. Med. *6*, 220ra11.

Ferreyra, J.A., Wu, K.J., Hryckowian, A.J., Bouley, D.M., Weimer, B.C., and Sonnenburg, J.L. (2014). Gut microbiota-produced succinate promotes C. difficile infection after antibiotic treatment or motility disturbance. Cell Host Microbe *16*, 770–777.

Gifford, D.R., and MacLean, R.C. (2013). Evolutionary reversals of antibiotic resistance in experimental populations of Pseudomonas aeruginosa. Evol. Int. J. Org. Evol. *67*, 2973–2981.

Gresham, D., and Hong, J. (2015). The functional basis of adaptive evolution in chemostats. FEMS Microbiol. Rev. *39*, 2–16.

Hapfelmeier, S., Lawson, M.A.E., Slack, E., Kirundi, J.K., Stoel, M., Heikenwalder, M., Cahenzli, J., Velykoredko, Y., Balmer, M.L., Endt, K., et al. (2010). Reversible Microbial Colonization of Germ-Free Mice Reveals the Dynamics of IgA Immune Responses. Science *328*, 1705–1709.

Hindré, T., Knibbe, C., Beslon, G., and Schneider, D. (2012). New insights into bacterial adaptation through in vivo and in silico experimental evolution. Nat. Rev. Microbiol. *10*, 352–365.

Kawamoto, S., Maruya, M., Kato, L.M., Suda, W., Atarashi, K., Doi, Y., Tsutsui, Y., Qin, H., Honda, K., Okada, T., et al. (2014). Foxp3(+) T cells regulate immunoglobulin A selection and facilitate diversification of bacterial species responsible for immune homeostasis. Immunity *41*, 152–165.

Korn, L.L., Thomas, H.L., Hubbeling, H.G., Spencer, S.P., Sinha, R., Simkins, H.M., Salzman, N.H., Bushman, F.D., and Laufer, T.M. (2014). Conventional CD4+ T cells regulate IL-22-producing intestinal innate lymphoid cells. Mucosal Immunol. *7*, 1045–1057.

Leatham, M.P., Stevenson, S.J., Gauger, E.J., Krogfelt, K.A., Lins, J.J., Haddock, T.L., Autieri, S.M., Conway, T., and Cohen, P.S. (2005). Mouse Intestine Selects Nonmotile flhDC Mutants of Escherichia coli MG1655 with Increased Colonizing Ability and Better Utilization of Carbon Sources. Infect Immun *73*, 8039–8049.

Leatham-Jensen, M.P., Frimodt-Møller, J., Adediran, J., Mokszycki, M.E., Banner, M.E., Caughron, J.E., Krogfelt, K.A., Conway, T., and Cohen, P.S. (2012). The Streptomycin-Treated Mouse Intestine Selects Escherichia coli envZ Missense Mutants That Interact with Dense and Diverse Intestinal Microbiota. Infect Immun *80*, 1716–1727.

Lenski, R.E., Wiser, M.J., Ribeck, N., Blount, Z.D., Nahum, J.R., Morris, J.J., Zaman, L., Turner, C.B., Wade, B.D., Maddamsetti, R., et al. (2015). Sustained

fitness gains and variability in fitness trajectories in the long-term evolution experiment with Escherichia coli. Proc. Biol. Sci. *282*.

Lescat, M., Launay, A., Ghalayini, M., Magnan, M., Glodt, J., Pintard, C., Dion, S., Denamur, E., and Tenaillon, O. (2016). Using long-term experimental evolution to uncover the patterns and determinants of molecular evolution of an Escherichia coli natural isolate in the streptomycin treated mouse gut. Mol. Ecol.

Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R., and Gordon, J.I. (2008). Worlds within worlds: evolution of the vertebrate gut microbiota. Nat. Rev. Microbiol. *6*, 776–788.

Li, H., Limenitakis, J.P., Fuhrer, T., Geuking, M.B., Lawson, M.A., Wyss, M., Brugiroux, S., Keller, I., Macpherson, J.A., Rupp, S., et al. (2015). The outer mucus layer hosts a distinct intestinal microbial niche. Nat. Commun. *6*, 8292.

Mason, T. g., and Richardson, G. (1982). Observations on the in vivo and in vitro competition between strains of Escherichia coli isolated from the human gut. J. Appl. Microbiol. *53*, 19–27.

Myhrvold, C., Kotula, J.W., Hicks, W.M., Conway, N.J., and Silver, P.A. (2015). A distributed cell division counter reveals growth dynamics in the gut microbiota. Nat. Commun. *6*, 10039.

Nishiyama, Y., Hamada, H., Nonaka, S., Yamamoto, H., Nanno, M., Katayama, Y., Takahashi, H., and Ishikawa, H. (2002). Homeostatic regulation of intestinal villous epithelia by B lymphocytes. J. Immunol. Baltim. Md 1950 *168*, 2626–2633.

Poulsen, L.K., Licht, T.R., Rang, C., Krogfelt, K.A., and Molin, S. (1995). Physiological state of Escherichia coli BJ4 growing in the large intestines of streptomycin-treated mice. J Bacteriol *177*, 5840–5845.

Rang, C.U., Licht, T.R., Midtvedt, T., Conway, P.L., Chao, L., Krogfelt, K.A., Cohen, P.S., and Molin, S. (1999). Estimation of growth rates of Escherichia coli

BJ4 in streptomycin-treated and previously germfree mice by in situ rRNA hybridization. Clin. Diagn. Lab. Immunol. *6*, 434–436.

Schoustra, S.E., Bataillon, T., Gifford, D.R., and Kassen, R. (2009). The Properties of Adaptive Walks in Evolving Populations of Fungus. PLoS Biol *7*, e1000250.

Vaishnava, S., Yamamoto, M., Severson, K.M., Ruhn, K.A., Yu, X., Koren, O., Ley, R., Wakeland, E.K., and Hooper, L.V. (2011). The Antibacterial Lectin RegIIIγ Promotes the Spatial Segregation of Microbiota and Host in the Intestine. Science *334*, 255–258.

Wiser, M.J., Ribeck, N., and Lenski, R.E. (2013). Long-term dynamics of adaptation in asexual populations. Science *342*, 1364–1367.

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt