

RESEARCH ARTICLE

# A Mutational Hotspot and Strong Selection Contribute to the Order of Mutations Selected for during *Escherichia coli* Adaptation to the Gut

Marta Lourenço, Ricardo S. Ramiro, Daniela Güleresi, João Barroso-Batista, Karina B. Xavier, Isabel Gordo\*, Ana Sousa\*

Instituto Gulbenkian de Ciência, Oeiras, Portugal

\* [igordo@igc.gulbenkian.pt](mailto:igordo@igc.gulbenkian.pt) (IG); [amsousa@igc.gulbenkian.pt](mailto:amsousa@igc.gulbenkian.pt) (AS)



 OPEN ACCESS

**Citation:** Lourenço M, Ramiro RS, Güleresi D, Barroso-Batista J, Xavier KB, Gordo I, et al. (2016) A Mutational Hotspot and Strong Selection Contribute to the Order of Mutations Selected for during *Escherichia coli* Adaptation to the Gut. *PLoS Genet* 12(11): e1006420. doi:10.1371/journal.pgen.1006420

**Editor:** Tim F Cooper, University of Houston, UNITED STATES

**Received:** May 13, 2016

**Accepted:** October 11, 2016

**Published:** November 3, 2016

**Copyright:** © 2016 Lourenço et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** DFG Genome sequencing data have been deposited in the NCBI Read Archive, accession no. SRP063701.

**Funding:** The research leading to these results was funded by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no 260421-ECOADAPT, the Deutsche Forschungsgemeinschaft Grant SFB 680 and the Portuguese Science Foundation FCT (PTDC/BIA-EVF/118075/2010 and RECI/IMI-IMU/0038/2012).

## Abstract

The relative role of drift versus selection underlying the evolution of bacterial species within the gut microbiota remains poorly understood. The large sizes of bacterial populations in this environment suggest that even adaptive mutations with weak effects, thought to be the most frequently occurring, could substantially contribute to a rapid pace of evolutionary change in the gut. We followed the emergence of intra-species diversity in a commensal *Escherichia coli* strain that previously acquired an adaptive mutation with strong effect during one week of colonization of the mouse gut. Following this first step, which consisted of inactivating a metabolic operon, one third of the subsequent adaptive mutations were found to have a selective effect as high as the first. Nevertheless, the order of the adaptive steps was strongly affected by a mutational hotspot with an exceptionally high mutation rate of  $10^{-5}$ . The pattern of polymorphism emerging in the populations evolving within different hosts was characterized by periodic selection, which reduced diversity, but also frequency-dependent selection, actively maintaining genetic diversity. Furthermore, the continuous emergence of similar phenotypes due to distinct mutations, known as clonal interference, was pervasive. Evolutionary change within the gut is therefore highly repeatable within and across hosts, with adaptive mutations of selection coefficients as strong as 12% accumulating without strong constraints on genetic background. *In vivo* competitive assays showed that one of the second steps (*focA*) exhibited positive epistasis with the first, while another (*dcuB*) exhibited negative epistasis. The data shows that strong effect adaptive mutations continuously recur in gut commensal bacterial species.

## Author Summary

The relative contribution of random loss and migration versus *de novo* mutation to the overall diversity of the gut microbiota is far from understood. Population sizes of bacterial communities inhabiting the gut can be very large and therefore, both weak and strong

JBB and AS were supported by grants SFRH/BD/80257/2011 and SFRH/BPD/111725/2015. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

effect beneficial mutations theoretically have the opportunity to contribute to adaptation. Here, by discretizing the adaptive steps that occur during colonization by a gut commensal, we uncover a mutational hotspot, large effect mutations and different forms of natural selection as the core ingredients to the emergence of diversity in this environment. We show that, on occasion, strong periodic selection can create and reduce diversity but then recurrent mutation generates new adaptive variants that compete for fixation. Selection for keeping particular adaptive variants at low frequency, balancing selection, was also shown to be pervasive. Unexpectedly, given the complexity of the gut ecosystem, we find a highly repeatable evolutionary process, motivated by a mutational hotspot and strong effect adaptive mutations continuously recurring in commensal bacterial species.

## Introduction

The composition of the gut microbiota can exert a strong influence on host physiology, behavior and health. Time series data have shown that the gut microbiota typically comprises a diverse community of species and that reduction of such diversity is frequently associated with illness [1]. Less studied, but potentially as important, is the diversity at the level of each species [2,3]. In fact, studies providing an understanding on how intraspecific variation in the microbiota emerges and changes over time are lacking [4]. 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 remain unanswered [5]. While some mutations segregating in natural populations may be neutral, the large size of bacterial communities inhabiting the mammalian gut suggests that here, polymorphism is more likely to result from deterministic forces [6].

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. Direct measurements of selective effects of spontaneously emerging mutations in this environment are extremely rare due to its complex nature. Nevertheless, even mutations with very weak adaptive effects, thought occur most frequently [7], could substantially contribute to a rapid pace of bacterial evolutionary change in the gut. Here, using experimental evolution of *E. coli* colonizing a natural environment, coupled with whole genome sequencing and *in vivo* competitive assays, we unravel some of the targets of adaptive evolution and the strength of the effects of beneficial mutations *in vivo*. Mouse colonization models offer a great opportunity to observe the emergence of diversity, test its repeatability amongst different hosts and measure the strength of natural selection in bacterial populations comprising the gut microbiota. 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, and contribute to strain variation within the gut. Furthermore, we were able to i) quantify important evolutionary parameters in this system, such as the effect of second step mutations, ii) evaluate the repeatability of evolution and iii) assess possible constraints on the order of the observed adaptive events.

## Results

### Multiple targets and a high degree of parallelism characterize the second steps of *E. coli* gut adaptation

By following the evolution of a commensal *E. coli* strain inhabiting the guts of streptomycin treated mice, we previously observed the emergence and spread of adaptive mutations within

only 3 days of colonization [8]. The first step of adaptation targeted a single locus and consisted of the selective inactivation of the *gat* operon, which enables *E. coli* to metabolize galactitol [8]. Distinct alleles conferring this phenotype, with similar selective effects ( $8 \pm 0.01\%$  (2SE) benefit), recurrently emerged in all independently evolving *E. coli* populations recovered from different hosts. Using this experimental system, we have now studied the subsequent steps of adaptation. We colonized 15 mice with a clonal population of *E. coli* carrying the first beneficial phenotype (inability to metabolize galactitol). This phenotype is conferred by a single base pair insertion into the coding region of the *gatC* gene, which codes for a subunit of the galactitol transporter, to thus prevent galactitol uptake. The colonizing population was also made dimorphic for a fluorescent marker to enable the emergence of further adaptive changes to be determined and their effects measured by competitive fitness assays [9].

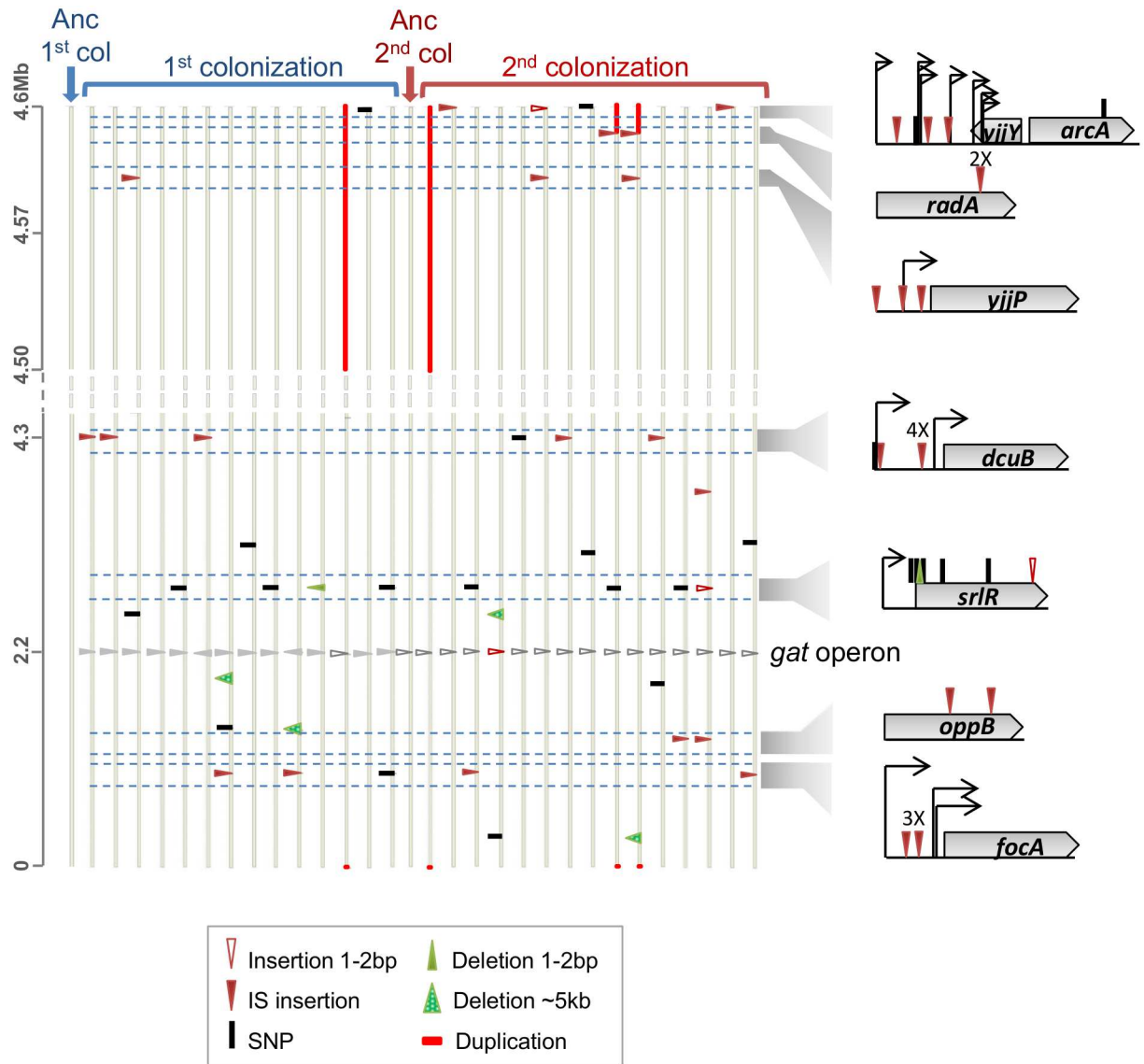
To ask whether single or multiple genetic targets underlie the second step of adaptation, and to determine how repeatable evolution is, we performed whole genome sequencing (WGS) of 15 independently evolved clones (sampled at day 24 post colonization, see Fig 1 and S1 Table).

A total of 30 mutations were detected, including 12 mutations in coding regions, 13 mutations in intergenic regions, as well as 3 large duplications and 2 large deletions. From the mutations observed in coding regions, 3 were non-synonymous, 1 was a nonsense mutation, and 6 mutations involved insertion sequence (IS) insertions. Finally, 8 IS insertions and 5 small insertion deletion mutations occurred in intergenic regions. The average number of mutations per sequenced clone was 2, a number similar to that observed in clones sampled from the first colonization (2.3 [8],  $n = 14$ , Fig 1).

To determine the likely targets of adaptation we compared the mutations occurring in the clones sampled from both colonizations. As shown in Fig 1, some mutational targets were similar between the first and second colonizations. Pooling all the clones, 7 new targets of mutation (*srlR*, *arcA*, *yjjP*, *oppB*, *radA*-dup, *dcuB* and *focA*) were recurrently detected in different mice, revealing their adaptive nature [10]. The most frequent was *srlR*, a locus that codes for the repressor of the sorbitol operon; followed by insertions in the intergenic region upstream of *dcuB* and *focA*, which code for membrane transporters of C4-dicarboxylates (e.g. fumarate) and formate, respectively [11,12]. The regulatory region of *arcA*, a dual transcriptional regulator predominantly involved in controlling the respiratory flexibility of *E. coli* [13,14], was also highly targeted. Another frequent mutation involved a large duplication (ranging from 34Kb to 157Kb). Two additional targets were observed less frequently: the promoter region of *yjjP*, which codes for a membrane protein of unknown function [15], and the coding region of *oppB*, which codes for a component of the oligopeptide ABC transporter [16]. These two loci were targeted by IS element insertions.

## Transposon insertions drive adaptation to the mouse gut by fine-tuning gene expression

Half of the mutations identified in the sequenced clones ( $n = 29$ ) were caused by IS insertions (Fig 1 and S1 Table), specifically IS5, IS1, IS2 and IS186, which have a high rate of transposition, between  $10^{-6}$  and  $10^{-5}$  per element per generation, in *E. coli* [17]. In fact, IS insertions occurred in 6 out of the 7 targets of adaptation identified through parallelism. Among these insertions, 78% were located in regulatory regions (S1 Fig), suggesting that mutations altering gene regulation are an important driver of adaptation in these populations [18]. Only one of the second step mutations involving an IS insertion likely caused a loss of function (insertion of an IS element in the coding region of *oppB*). The phenotypic effects of IS adaptive insertions in the regulatory regions of *focA*, *dcuB*, *arcA*, *yjjY* and *yjjP* were determined by comparing gene expression of representative mutants evolved during gut colonization with that of the ancestral

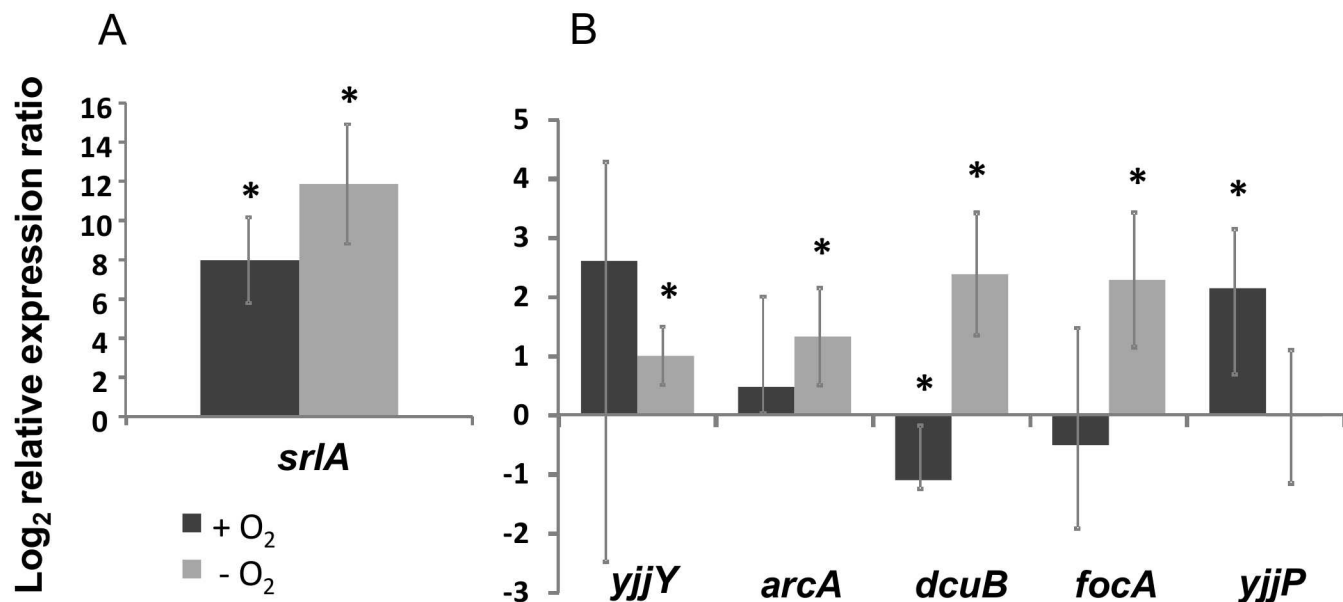


**Fig 1. The genetic basis of the 2<sup>nd</sup> step of adaptation involved 7 parallel mutational targets.** Mutations were identified by whole genome sequencing of 29 evolved clones: 14 independently evolved clones from an ancestral strain (first colonization) [8] and 15 independently evolved clones from an adapted clone carrying a single beneficial mutation (second colonization). For simplicity, the genomes are represented linearly (vertical bars). The genomic context of the parallel mutations is represented on the right. Type and position of mutations are shown as triangles for insertions and deletions and small vertical bars denote single nucleotide polymorphisms (SNPs). Four duplications are depicted as red horizontal bars. See the symbol legend for other events. Regions of parallel mutation are highlighted. Numbers above marked mutations represent the number of clones where a particular mutation was detected at the same position.

doi:10.1371/journal.pgen.1006420.g001

clone (see [Material and Methods](#)). Since *E. coli* in the gut experiences variation in the amount of oxygen conditions [13,14], gene expression was assayed both in the presence and absence of oxygen.

In anaerobic conditions the transcription level of *arcA*, *focA* and *dcuB* was higher in the evolved clones than in the ancestor. In aerobic conditions the effect of IS insertions on expression was more variable, ranging from a 1-fold reduction of *dcuB* expression to a 2-fold increase in the expression of *yjiP* (Fig 2A).



**Fig 2. Transposable elements modulate gene expression.** The graph represents the log<sub>2</sub> relative expression rate of each gene in the evolved clone in relation to that of the ancestral clone. This was obtained by RT-qPCR and analyzed according to the method described in [19], using *hfq* as an housekeeping gene. The expression rate is above 0 for evolved clones where a SNP in *srlR* (A) or an IS insertion in the depicted genes (B) caused an increase in gene expression (in relation to the ancestor); if this value is below 0 gene expression was decreased (in comparison to the ancestor). Asterisks indicate significant differences at the level  $P < 0.05$  (unpaired T-test). Effects of IS insertions on gene expression were conditional on oxygen availability. Error bars represent 2 times the standard error. Representative clones used for RT-qPCR measurements were the following (see S1 Table): 18YFP (*focA srlR*), 22YFP (*dcuB*), 25YFP (*yjjP/yjjQ radA insX-insA*), 29CFP (*arcA*) and the ancestral strains DM08-YFP and DM09-CFP.

doi:10.1371/journal.pgen.1006420.g002

Mutational targets provide clues on the traits under selection in a given environment. Three of the targets identified (*arcA*, *focA* and *dcuB*) are related with microaerobic/anaerobic metabolism. For example, fumarate (transported by DcuB) has been described as the most important anaerobic electron acceptor in the intestine of streptomycin-treated mice model [14], whereas formate (transported by FocA) is the signature compound of anaerobic metabolism for *E. coli* [20]. The gut is described as a microaerobic environment [13,14] where *E. coli* can take profit of its respiratory flexibility. ArcA is known to be an important regulator of this trait and to play an important role in the ability of *E. coli* to colonize [13,14]. Specifically, ArcA is a global repressor of carbon oxidation pathways [21], repressing the expression of 74 operons and inducing the expression of 11 operons (including *focA*) under anaerobic conditions. Interestingly, the 4 large duplications, observed in different hosts, involved a genomic region that includes *arcA*. Since duplications can raise gene dosage we hypothesize that increased expression of this gene constitutes a major beneficial phenotype.

In sum, the mutational targets indicate that *E. coli* is adapting to the intestinal tract by tuning the expression of a global respiratory regulator (*arcA*) and changing its carbohydrate metabolism and transport (*srlR*, *dcuB*, *focA* and *oppB*). Consistent with this metabolic adaptation, several gut-adapted clones have an increased fitness when competing with the ancestral in media containing carbon sources known to be present in the intestine [22]. This increased competitive ability occurs both in the presence and absence of oxygen (S2 Table).

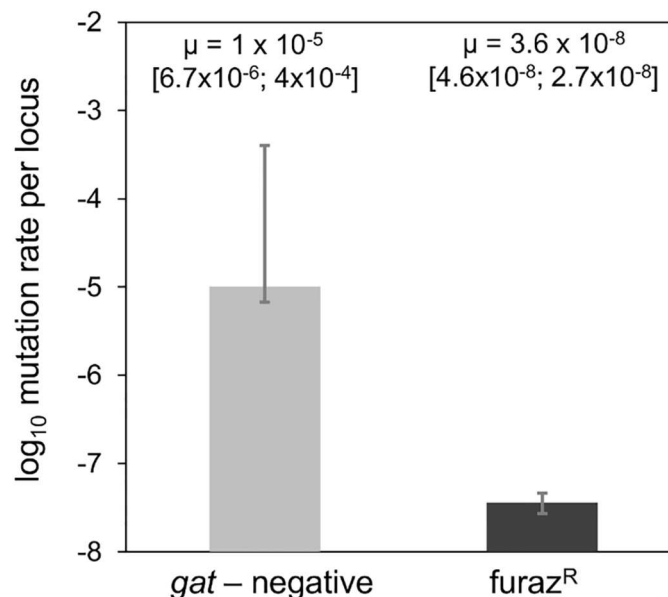
Another interesting example of gain-of-function mutations, not involving transpositions, occurred at the *srlR* locus (which codes for the transcriptional repressor of the sorbitol operon). Around half of the mutations observed in this gene (considering the 29 clones from the two colonizations) were either single base pair insertions or nonsense mutations, which thus

presumably inactivated this gene. Indeed the expression of *srlA* (the first gene of the sorbitol operon) was increased 12-fold and 8-fold in the absence and in the presence of oxygen, respectively (Fig 2B), in the gut-adapted clones, supporting the hypothesis that mutations in *srlR* abrogated expression of this repressor and lead to increased expression of the sorbitol operon. Consistently, the mutants carrying the *srlR* mutation have an advantage when competing for sorbitol (S2 Table).

### 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 as well as the stability of the environment [23,24]. We investigated whether these processes play a role in *E. coli* adaptation to the mouse gut. We measured the spontaneous mutation rate to the functional inactivation of the *gat* operon ( $\mu$ ), using a fluctuation assay in rich medium (LB), and found a  $\mu$  of  $10^{-5}$  (95% CI, [ $6.7 \times 10^{-6}$ ,  $4 \times 10^{-4}$ ]) per locus per generation (Fig 3).

We compared this rate with the rate of inactivating mutations occurring at another locus of the genome (*nfsA*), which results in resistance to furazolidone (*furaz*<sup>R</sup>). For this locus we found a rate of  $4 \times 10^{-8}$  (95% CI, [ $4.6 \times 10^{-8}$ ,  $2.7 \times 10^{-8}$ ]), far closer to the expected rate for loss of function mutations ( $10^{-8}$  per gene per cell division [26]). After correcting for locus size differences, the inactivation rate of the *gat* operon is one order of magnitude higher than that of *nfsA*, which is taken as a random locus. The estimation of the mutation rate in the fluctuation test performed above assumes that the mutations causing the phenotype for which  $\mu$  is being determined are neutral. To account for a potential bias in  $\mu$  estimate, which might have resulted from a lack of



**Fig 3. The *gat* locus is a mutational hotspot.** The graph shows the  $\log_{10}$  mutation rate ( $\mu$ ) (per locus, per generation) for each phenotype: *gat*-negative (grey bar) and furazolidone resistance (black bar) and their respective confidence intervals. Resistance to furazolidone is conferred by the inactivation of the *nfsA* gene. The values of  $\mu$  were estimated by performing fluctuation assays and analyzing the results using the maximum likelihood approach implemented in FALCOR [25]. The mean frequency of spontaneous *gat*-negative mutants is ~300 times higher than the frequency of spontaneous resistants to furazolidone. After correcting for the difference between loci sizes the frequency of *gat*-negative phenotype is still ~38 times higher than the frequency of furazolidone resistance.

doi:10.1371/journal.pgen.1006420.g003

neutrality of the *gat*-negative phenotype in LB, we determined its fitness effect when in direct competition with the *gat*-positive ancestor in this environment. The *gat*-negative phenotype was found to be deleterious *in vitro*, with an estimated selection coefficient of  $-0.06 \pm 0.01$  (2SE). A similar competition between a *nfsA* knock mutant and the wild type bacterium resulted in a selection coefficient of  $-0.007 \pm 0.005$  for the mutant, thus the knock out of this gene is very close to neutrality. Our estimate of  $\mu$  is therefore conservative, since the strong deleterious effect of the phenotype being scored can only cause an underestimation of  $\mu$ . These results support the conclusion that the high rate of spontaneous inactivation of the *gat*-operon could have been an important contributory factor for the emergence of the *gat*-negative phenotype as the first adaptive event to occur.

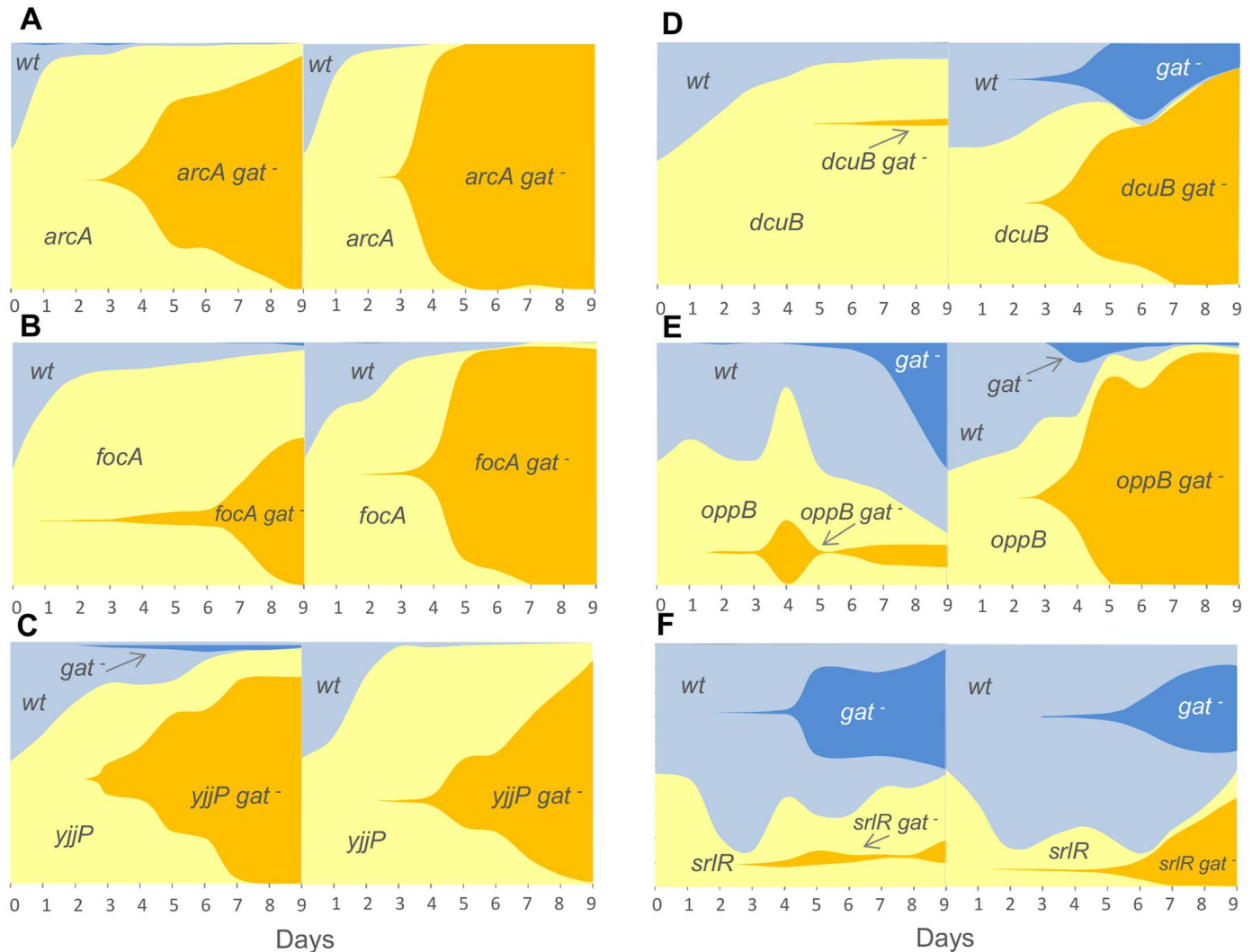
Heterogeneity in the mutation rate does not exclude a possible contribution of direct or epistatic selection to the order of the adaptive events. To test for this, and to determine the selective effects of beneficial mutations in the gut, we estimated the selection coefficients of 6 out of the 7 second step mutations when each occurs either in a wild-type (*gat*-positive) or in an evolved (*gat*-negative) background. This was done through *in vivo* competitive fitness assays against the ancestors of the first (*gat*-positive) (Figs 4 and 5) and second colonizations (*gat*-negative) (Fig 5). The *radA*-dup mutation was not tested, as this duplication was highly unstable during *in vitro* manipulation. We note that despite the allelic variation within each target of adaptation, only one allele representative of each locus was tested. While for the *gat* locus we have previously shown that different alleles have equivalent selective effects [8,27], we acknowledge that for the second targets of adaptation, different alleles may have different phenotypes.

We found substantial variation for the selective effects of these mutations in both backgrounds ( $P = 6 \times 10^{-6}$  for the *gat*-positive background and  $P = 6 \times 10^{-6}$  for *gat*-negative background, ANOVA). This contrasts with the effect of the mutations responsible for the first step of adaptation (for which a selective effect,  $s_{gat-neg} = 0.08 \pm 0.01$ , was estimated [8], Fig 5). Most mutations showed a weaker competitive advantage on the *gat*-positive background than those causing the first adaptive step (ANOVA with Tukey *post-hoc* correction for multiple testing, Fig 5 and S3 Table). Interestingly the effect of the *arcA* mutation was significantly larger than the first step mutations ( $P = 0.01$ , ANOVA Contrasts). These results indicate that alongside differences in the mutation rate, selection might also have contributed to the order of the adaptive steps.

Remarkably, the mutation inactivating the *srIR* locus, leading to constitutive expression of the sorbitol operon (Fig 2B), steadily decreased in frequency ( $s_{srIR} = -0.03$ ) during the first days of competition (Fig 4F). It was subsequently maintained at low frequency, an observation that would not be expected for a strictly deleterious allele. This mutation also emerged in multiple, independent populations (Fig 1), all of which suggests that *srIR* is a likely target for balancing selection, a hypothesis further investigated below.

Importantly, we observed the appearance of the *gat*-negative phenotype on the ancestral background, and in all backgrounds carrying a single second step mutation (Fig 4). This strongly suggests that high repeatability and little historical contingency on genetic background, accompany the evolutionary path taken by *E. coli* during gut colonization. Here we refer to historical contingency as the situation where the beneficial effect of a focal mutation is contingent on the ancestral background since it becomes deleterious in an evolved background, thus limiting the number of evolutionary paths. [28].

To determine whether epistatic interactions occur between the mutations, and how the first and second steps differ in terms of magnitude, we further measured the effects of these second step mutations in the *gat*-negative background. We observed that mutations in *arcA*, *yjjP* and *oppB* had a similar effect in *gat*-negative and *gat*-positive backgrounds, whereas the fitness effect of mutations in *dcuB* and *focA* differed ( $P = 0.029$  for both of the mutations. Mann-Whitney-Wilcoxon). *dcuB* showed negative epistasis, since its effect is undistinguishable from

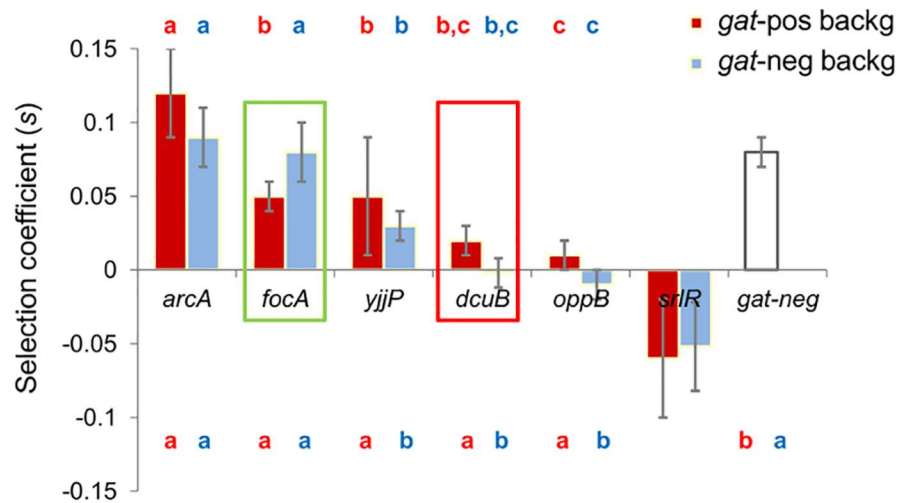


**Fig 4. The mutations occurring during the second step of adaptation show beneficial effects in the wild type background (*gat*-positive).** Graphical representation of the frequencies of each mutant (labelled with a *yfp* allele) carrying one of the second step mutations: (A) *arcA*, (B) *focA*, (C) *yjjP*, (D) *dcuB*, (E) *oppB* and (F) *srlR* and the ancestor (labelled with a *cfp* allele). Shaded areas are proportional to the frequencies of each phenotype. Increasingly darker tones of yellow or blue represent newly generated *gat*-negative phenotypic mutants in the respective fluorescent background. *gat*-negative mutants (likely originating from multiple mutations and typed for the phenotype) invade both genetic backgrounds, indicating that its beneficial effect could be observed both in the background of the first and second step mutations. In this sense no historical contingency along the adaptive walk is expected since the fixation of any of the second step mutations could not prevent the emergence and spread of the *gat*-negative phenotype.

doi:10.1371/journal.pgen.1006420.g004

neutrality on the more adapted background (the *gat*-negative) while it is beneficial ( $0.02 \pm 0.01$ ) in the less adapted background (*gat*-positive). The high frequency of emergence of mutations in *dcuB* during the evolution experiment, however, strongly support its having a beneficial effect in the *gat*-negative background, suggesting that *dcuB* is, in fact, a likely target for frequency-dependent selection. Conversely, the mutation in *focA* had a significantly higher effect in the more adapted, *gat*-negative, background. This provides an interesting example of positive epistasis, a phenomenon which, though documented, occurs much less frequently than that of negative epistasis [29–31].





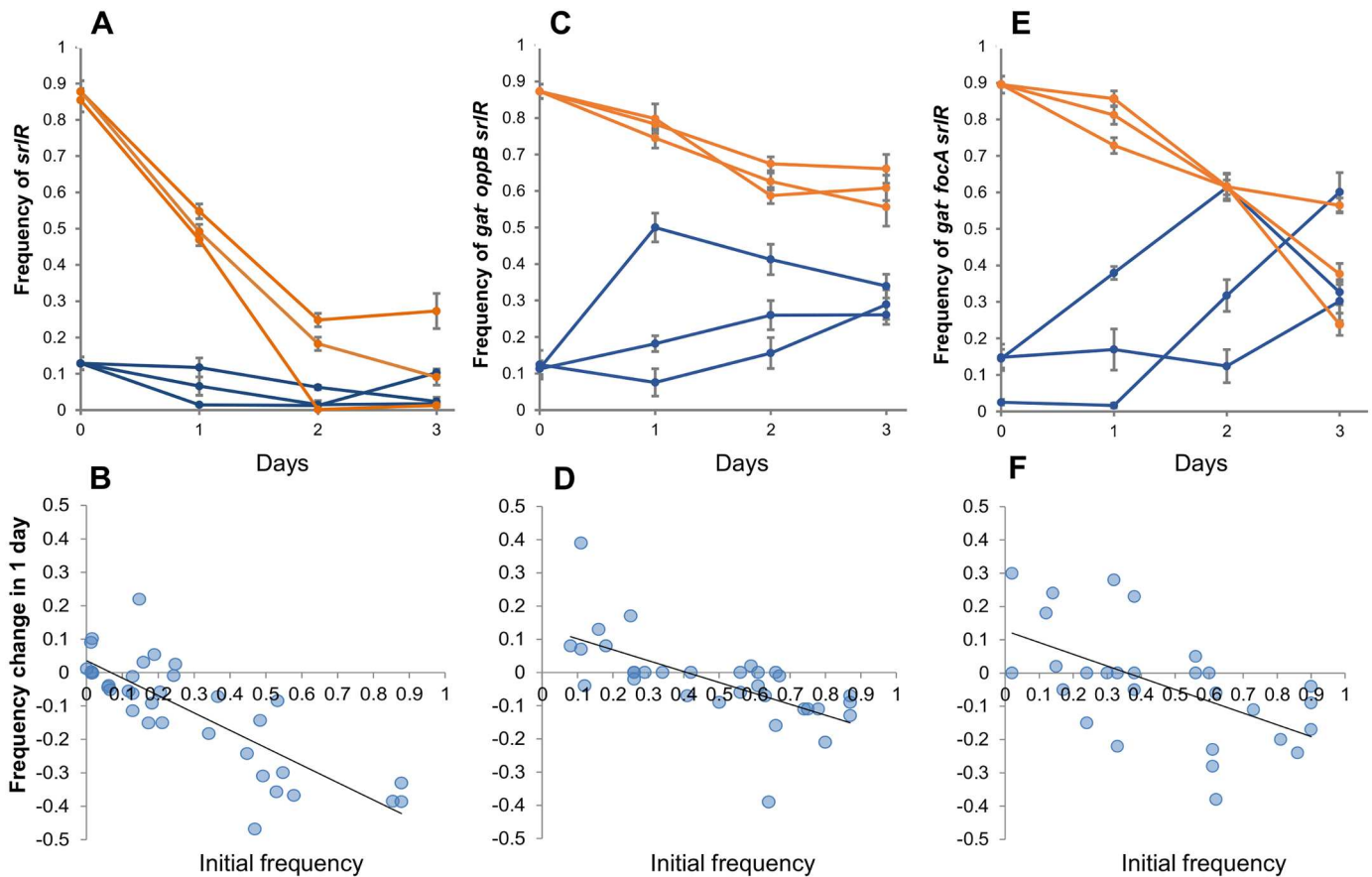
**Fig 5. The second step mutations show variable effects both within and between backgrounds (*gat*-positive and *gat*-negative).** All second step mutations except for the *srlR*, are either beneficial or neutral in the two backgrounds (red for the *gat*-positive background and blue for the *gat*-negative background). The selection coefficient of each mutation is an average of 4 independent competitions (except for *yjiP* on the *gat*-positive background where  $n = 3$ . See also S2 Fig) and was estimated as described in the Methods. The time interval used to estimate the selection coefficient was 3 days in all cases, except for the *arcA* mutation in the *gat*-positive background (since it reached a very high frequency after 2 days), and for the *srlR* mutation in the *gat*-negative background where 1 day was used (since it shows FDS- see Fig 6). Letters above the bar chart indicate which selective effects are significantly different from each other in the respective background (ANOVA with Tukey *post-hoc* correction for multiple testing. Please see S3 Table for the significance associated with each pairwise comparison). Letters below the bar chart indicate which selective effects are significantly different from the *gat*-negative mutation (ANOVA Contrasts, please see S3 Table for the *P* value associated with each pair wise comparison). Error bars (2SE). Rectangles (green and red) denote the mutations with a significantly different effect in the *gat*-positive versus *gat*-negative backgrounds (*i.e.* mutations showing an epistatic interaction with the background, green rectangle denotes positive epistasis and red negative epistasis.  $P < 0.05$ , Mann-Whitney-Wilcoxon Test).

doi:10.1371/journal.pgen.1006420.g005

Consistent with theoretical models of adaptation towards a fitness optimum, several empirical studies have found that the effects of beneficial mutations tend to decrease as adaptation proceeds [32–34]. Interestingly, we have found that two of the second step mutations ( $s_{arcA} = 0.09 \pm 0.02$  and  $s_{focA} = 0.08 \pm 0.02$ ) were as large as the first step ( $s_{gat-neg} = 0.08 \pm 0.01$ ) ( $P = 0.01$  and  $P = 0.05$ , respectively. ANOVA Contrasts). This observation suggests that the gut maybe more complex than a more strictly constant *in vitro* environment or that there are multiple discrete traits to optimize even in a constant environment. A key distinction between a fixed and a moving landscape is that in the first the adaptive walk is short and strong diminishing returns are expected. In a moving landscape [35–37], adaptation is continuous and the step sizes may not necessarily decrease with time.

### Evidence for strong negative frequency-dependent selection on the inactivation of the sorbitol repressor

Multispecies communities of microbes are a likely scenario for the emergence of important ecological interactions. One possible outcome of these interactions is frequency-dependent selection (FDS), a situation where the fitness of a phenotype changes according to its frequency [5]. Negative FDS occurs if a phenotype is deleterious when common but beneficial when rare: this form of selection is capable of maintaining genetic diversity. The results of the competition assays with the *srlR* mutation (Fig 4F) indicate that negative FDS may have played a role in the evolution of the populations. To directly test this hypothesis, we performed new competitive



**Fig 6. Evidence for balancing selection.** *In vivo* competitive fitness assays where the mutant starts at a frequency of ~10% (blue lines) or ~90% (orange lines) in relation to the reference clone. Panel (A) shows the competition between *srIR* and wild type ancestor (DM09); panel (C) shows the competition between 18YFP (*gatC oppB srIR*) and JB19-YFP (*gatC*) and panel (E) shows the competition between 27CFP (*gatC focA srIR*) and JB18-CFP (*gatC*). The temporal dynamics of the frequency of the mutants show negative FDS, maintaining variation at the *srIR* locus, as *srIR* alleles are disadvantageous when at high frequency but are maintained at low frequency (B, D and F, where the change in frequency from day *i* to day *i*+1 as a function of frequency at day *i*, along the temporal dynamics is shown). This form of selection dominates irrespective of the presence of additional mutations, such as those in *focA*, previously determined to be under positive Darwinian selection (Fig 4B).

doi:10.1371/journal.pgen.1006420.g006

fitness assays starting with different frequencies of the wild type and *srIR* mutant strains, (same genotypes as in Fig 4F). Indeed, *srIR* mutants show a strong disadvantage when competing with the ancestor at high frequency but were maintained at lower frequencies (Fig 6A). When plotting the change in frequency of this mutant as a function of its initial frequency, a significant linear correlation is observed (Fig 6B—Linear regression Adjusted  $R^2 = 0.66$ ,  $P < 0.0001$ ; slope =  $-0.52$  (0.06 SE), intercept =  $0.04$  (0.02 SE), which supports FDS.

Interestingly, the effect of the *srIR* mutation was so strong that frequency dependence was maintained, irrespective of the background in which it occurred. As shown in Fig 6C and 6E, two different genotypes (*gatC oppB srIR* and *gatC focA srIR*), carrying a *srIR* mutation, had a competitive advantage when rare, but a disadvantage when at high frequency (Fig 6D—Linear regression, Adjusted  $R^2 = 0.41$   $P < 0.0001$ ; slope =  $-0.32$  (0.07 SE), intercept =  $0.13$  (0.04 SE) and Fig 6F—Linear regression Adjusted  $R^2 = 0.30$   $P = 0.0015$ ; slope =  $-0.35$  (0.1 SE), intercept =  $0.13$  (0.05 SE)). This shows that positive Darwinian selection on *focA* (Figs 4B and 5), previously shown to drive this mutation rapidly to fixation, can be highly influenced by negative FDS if the *srIR* mutation co-occurs with this mutation. The strength of selection on *oppB*

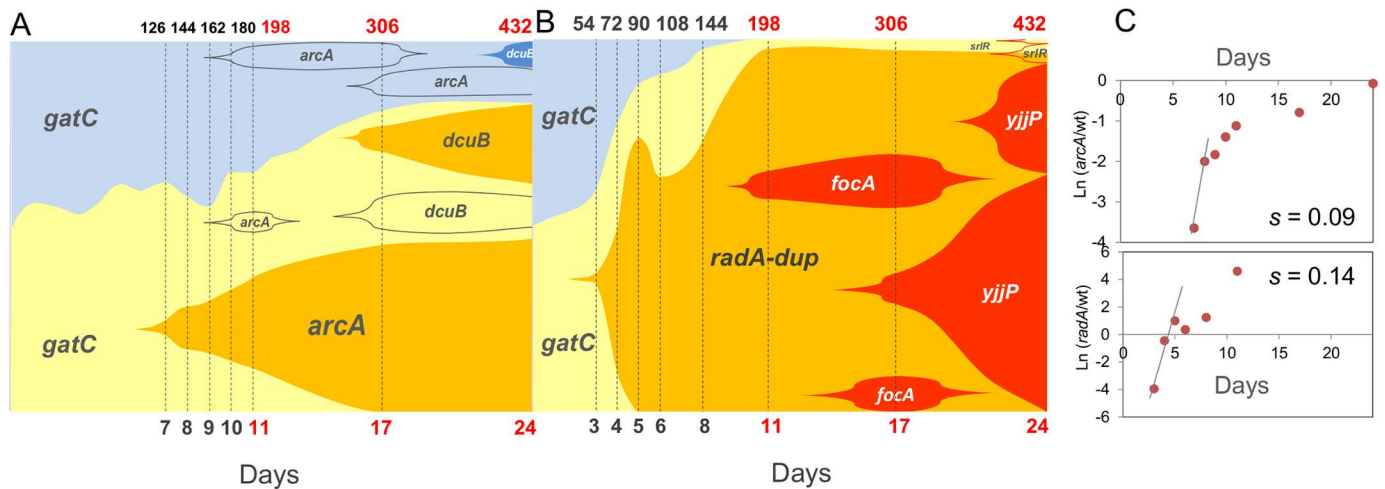
(Fig 5) is much weaker than that of *focA*, yet when combined with *srlR* it shows a similar dynamics (Fig 6C). This suggests either an epistatic interaction between *srlR* and *oppB* or that *oppB* is also under FDS. Overall, these observations support the hypothesis that the *srlR* mutation *per se* is able to maintain diversity for long periods of time.

### Evidence for periodic selection and clonal interference driving intra-species variation

To monitor adaptation at the genome-wide scale, we analyzed levels of polymorphism across time in two of the evolving populations colonizing the mouse gut. We used both WGS of samples of the populations and typing of selected mutations in clones to unravel the pervasiveness of clonal interference.

Distinct alleles at each locus were found segregating in the two populations assayed (Fig 7A and 7B and S3 Table): 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.

Importantly, population 2.10 is remarkable in that it shows evidence of co-occurrence of all classical forms of natural selection (periodic selection, clonal interference and FDS), within less than a month. First, there is an example of a molecular pattern close to that expected under periodic selection, where a strong haplotype involving an IS insertion in *radA* and a large duplication (*radA*-dup) rapidly sweeps close to fixation (Fig 6B). Second, after day 11 a



**Fig 7. Clonal interference and periodic selection co-occur during adaptation of *E. coli* to the mouse gut.** Muller plot of the frequencies of newly generated haplotypes during 24 days (corresponding to 432 generations) of evolution inside the mouse gut (see S4 Table for numeric data). Shaded areas are proportional to the frequencies of each haplotype. Yellow and blue shaded areas represent the two sub-populations of bacteria labeled either with *cfp* or *yfp* alleles. Increasingly darker tones of yellow or blue represent the accumulation of target beneficial mutations in clones with a given fluorescence background (either *cfp* or *yfp*). In cases where the fluorescent background is not known, the new mutation (as assayed by whole genome population sequencing) is indicated without change in the shading tone. Evidence for positive Darwinian selection in the gut: clonal interference, where clones carrying different adaptive mutations at the same and in different loci compete for fixation in the gut in population 2.7 (A); periodic selection results 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 2.10 (B). Dashed lines mark the time points in days (lower axis) or generations (lower axis) where the sampling of clones (in black) or WGS (in red) of the population took place. For population 2.7, between 40 and 50 clones were analyzed at each time point, whereas in population 2.10, between 20 and 60 clones were analyzed each day (please refer to Methods for the further details of this analysis). The effects of two second step mutations (*radA*-dup and *arcA*) were estimated through their frequency increase in the populations where they emerged (C). This quantification was only possible because additional mutations are very unlikely to be segregating at the same time. Only mutations observed more than once in any of the sequenced clones are represented here. Population 2.7 has one additional mutation not represented in the figure and population 2.10 has two other mutations (see S4 Table).

doi:10.1371/journal.pgen.1006420.g007

signature of clonal interference between different alleles of *focA* and *yjjP* is observed. Finally, towards the end of the experiment, clones carrying mutations in *srlR* emerge, which should result in negative FDS (as shown in Fig 6).

The selection coefficient associated with the *radA-dup* mutant can be estimated from its change in frequency while sweeping through the population, but before other mutations can be detected. The estimate for the effect of this mutation is  $s_{rad\ dup} = 0.14$  (Fig 7C). This is the highest selection coefficient for all of the mutations of the second steps of adaptation (Fig 5), though it is based on a single observation.

In the same manner, a selective benefit for the *arcA* mutation can also be estimated (Fig 7C). A selection coefficient of 0.09 is estimated between day 7 and 8 when the mutation was first detected and before day 11 when another mutation (*pphB*) was also observed (S4 Table). Even though this is a crude estimate, it is remarkably consistent to that measured in direct competition between a mutant carrying the same IS5 insertion in the *arcA* regulatory region and the ancestor of the second colonization (Fig 6 and S3 Table  $s_{arcA} = 0.09 \pm 0.02$ ).

## Discussion

The rules governing the adaptive process of microbial populations in natural communities are far from understood. Numerous aspects of host physiology [38] and even behavior are influenced by the microbiota [39], making this a priority environment to appreciate how much microbial evolution occurs within the lifetime of a host. Nevertheless, the amount of evolution taking place in this environment, as well as its contribution to the overall diversity resulting from *de novo* mutation is currently underappreciated. We followed evolution in a natural environment under controlled conditions and studied two consecutive bouts of adaptation. The first bout was caused by the emergence of multiple mutations (both within and among hosts) causing a similar phenotype, loss of function of the operon encoding galactitol metabolism (*gat*-negative phenotype), and an associated selective effect of  $0.08 \pm 0.01$  per generation [8]. The ability to use galactitol is a polymorphic trait in wild strains of *E. coli* [40]. The second bout of adaptation targeted at least seven different loci. No strong signals of deceleration in adaptation rate were detected, judging from the strong effect of the second step mutations estimated either from their frequency increase in the evolving populations (Fig 6C,  $s_{rad-dup} = 0.14$  and  $s_{arcA} = 0.09$ ) or from direct *in vivo* competition assays against the ancestral strain of the second colonization, which had a *gat*-negative phenotype (Fig 5). This observation contrasts with the pattern of diminishing returns epistasis previously found in *E. coli* adapting to a glucose limiting environment [29] and in *Methylobacterium extorquens* AM1 evolved in batch culture with methanol as the sole carbon source [41]. The extent to which the fitness effects of further mutations that accumulate *in vivo* may become smaller as adaptation proceeds remains to be determined in future work.

By discretizing the adaptive steps we aimed to understand not only the rhythm but also the repeatability of adaptation in the gut environment. WGS of clones isolated from different mice revealed 7 parallel targets comprising: three membrane transporters, one repressor of a metabolic operon, one major regulator involved in the aerobiosis/anaerobiosis transition, one large duplication, and a protein of unknown function. Interestingly all mutations tested, which occurred in regulatory regions, either up-regulated the targeted gene directly, or downstream genes in the regulatory cascade. The fact that *E. coli* adapted to the gut environment by up-regulating (directly or indirectly) nutrient transport functions bears an interesting resemblance to the observations of microbes adapting to limiting nutrient conditions in chemostats [42]. For instance, *E. coli* adapted to low lactulose chemostats by duplicating the *lac* operon or abolishing its regulation [43] and *Sacharomyces cerevisiae* adapted to glucose-limited conditions by amplifying a region

including high affinity transporters [44]. Similarities between adaptation in the gut and in chemostats maybe expected to some extent, given that the later were built for continuous culture, where the flow is controlled by a peristaltic pump mimicking digestive transit. Furthermore, chemostats enable conditions to be set to ensure a slow growth rate by manipulating the concentration of the limiting nutrient, perhaps simulating the conditions experienced in the gut by *E. coli* [22]. Remarkably, in chemostats with limiting glucose concentrations, a great amount of clonal interference and mutations exhibiting frequency dependence was shown to occur [45,46].

Epistasis is an important factor that can impose strong constraints on the adaptive process of bacterial populations [47,48]. This has been extensively studied *in vitro* [29,30,49,50] but only rarely *in vivo* [51]. In the context of bacterial evolution in the gut we evaluated whether epistasis contributed to the order of the adaptive steps that characterized *E. coli* evolution during colonization of the mouse intestine. Our setup provides an ideal situation to address this question, since the order of adaptive events was absolutely conserved with a single universal first adaptive phenotype. One possible explanation for the order of events could result from the second mutations being weaker or even deleterious in the ancestral genotype. In fact we estimated that at least one third of the second step mutations were as strong as those responsible for the first step, showing that strong effect mutations were still available for adaptation. However, many showed a smaller average effect (mean  $s_{(oppB, yjjP, dcuB, focA)} = 0.03$ ). Therefore, it is possible that the large effect size of the first-step may have contributed for the order observed. A much more likely explanation can, however, be provided by differences in mutation rate. Indeed, we found that the *gat* operon is inactivated at a much higher rate than a random locus ( $10^{-5}$  versus  $3.6 \times 10^{-8}$ ). This fact very likely contributed to the 100% parallelism observed at the phenotypic level, inactivation of the galactitol metabolism, despite the much reduced parallelism at the genetic level [8]. Besides the *gat*, shown here, other metabolic operons have been previously found to be mutational hotspots. One example is the *ribose* operon which was the most common target in one experiment of *E. coli* adaptation to glucose minimal medium [52]. Inactivation of this operon occurred at a high rate ( $\sim 5 \times 10^{-5}$  per genome, per duplication), similar to that of *gat*. However it conferred a 1–2% benefit to *E. coli* growing on glucose, considerable smaller than the 8% benefit conferred by *gat* inactivation in *E. coli* colonizing the mouse gut.

Overall we present some of the first quantitative estimates of the fitness effects of beneficial mutations occurring in bacteria colonizing a natural ecosystem, where both ecological and evolutionary processes occur at fast time scales. The data obtained establish that strong effect beneficial mutations exhibiting all classical forms of natural selection shape the genetic diversity of a commensal species inhabiting the mouse gut microbiota. They show that, even in mammalian hosts with identical genetic backgrounds and diets, reproducible adaptations emerge albeit with a significant level of host individuality.

## Materials and Methods

### Ethics Statement

All experiments involving animals were approved by the Institutional Ethics Committee at the Instituto Gulbenkian de Ciencia (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 *Escherichia coli* [53]. Strains JB19-YFP and JB18-CFP (MG1655, *galk::YFP/CFP*  $cm^R$ ,  $str^R$  (*rpsL150*),  $\Delta$ *lacIZYA*, 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<sup>R</sup>, str<sup>R</sup> (*rpsL150*),  $\Delta$ *lacIZYA*) [9]) 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. The yellow (*yfp*) and cyan (*cfp*) fluorescent genes linked to cm<sup>R</sup> were then transferred by P1 transduction to a derivative of clone 12YFP [8], 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* and a large duplication. During the genetic manipulations the large duplication was lost, confirmed by whole genome sequencing, leaving this clone with a single mutation, 1bp insertion in *gatC*.

To measure the effect of the identified parallel mutations in gene expression we tested the following clones isolated from the evolution experiment reported here (see S1 Table): 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<sup>nd</sup> step of adaptation (in the *gat*-positive background) through *in vivo* fitness assays we used several single mutant clones isolated from a previous evolution experiment [8] (5YFP (*srlR*) and 6YFP (*dcuB*)) or the experiment here reported (17YFP (*arcA*), 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 [54] to the ancestral 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*, results in the 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*.

All mutants used to test the effect of the 2<sup>nd</sup> step mutations in the *gat*-negative background were derived from the collection of mutants constructed to test the effect of the same mutations in the *gat*-positive background. These mutants, except for *srlR*, were made *gat*-negative by P1 transduction of the *gatC* allele present in the ancestor of the 2<sup>nd</sup> colonization. Selection of *gat*-negative mutants was achieved by incorporating 10mM of D-Arabitol in the selection plates. All clones were confirmed to have the intended *gatC* allele by target PCR and restriction analysis as described in [8]. The *srlR* mutant was obtained by transduction of the *gatC::Kan* from the KEIO collection and was competed against a reference strain with the same *gatC::Kan*.

To test for FDS we competed the single *srlR* mutant previously mentioned (derived from clone 5YFP but with a wild type *gat* operon), against the ancestral strain DM09 (MG1655-CFP). We also competed the clones 18YFP (*gatC focA srlR*) and 27CFP (*gatC oppB srlR*) against the reference strains JB18-CFP (*gatC*) and JB19-YFP (*gatC*), respectively.

In order 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) and this evolution experiment (sequenced clones 16 to 30, from populations 2.1 to 2.15, S1 Table).

*In vitro* competitions to measure the effect of the *gat* operon and *nfsA* gene inactivation were performed using the clones 4YFP [8] and a clone obtained by P1 transduction of *nfsA::Kan* (from the KEIO collection), DM08-YFP and DM09-CFP. Mutant clones were constructed

in the two fluorescent backgrounds and competed against the ancestors DM08-YFP or DM09-CFP, depending on the fluorescent background.

To distinguish between *gat*-negative and *gat*-positive bacteria we used the differential medium MacConkey agar supplemented with galactitol 1% and streptomycin (100 µg/ml). Plates were incubated at 30°C. 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<sub>4</sub> 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 used.

## Evolution experiment

In order to study *E. coli*'s adaptation to the gut we used the classical streptomycin-treated mouse colonization model [55] and performed the evolution experiment using the same conditions as before [8]. 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<sup>8</sup> 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 OD<sub>600</sub> 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 -80°C and the remaining was plated in Luria Broth agar supplemented with streptomycin (LB plates). Plates were incubated overnight at 37°C and then with the help of a fluorescent stereoscope (SteREO Lumar, Carl Zeiss) 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 [9].

## *In vivo* competitive assays

To test the *in vivo* advantage of 12 clones carrying the 2<sup>nd</sup> step mutations in (*n* = 4 per clone) (Figs 4 and 5) and 3 clones carrying the *srlR* mutation (*n* = 3 per clone) we performed competitive assays against the respective ancestor labelled with the opposite fluorescent marker. *In vivo* competitions were performed at a ratio of 1 to 1 for all clones except the ones where we tested for FDS. To test for FDS we performed ratios of 1:9 and 9:1, following the same procedure described above for the evolution experiment. Mice fecal pellets were collected daily, diluted in PBS and frozen in 15% glycerol at -80°C. Total numbers and relative proportions of YFP- and CFP-labeled *E. coli* were subsequently determined by plating appropriate dilutions in either LB agar supplemented with streptomycin (100 µg/ml) or MacConkey supplemented with streptomycin (100 µg/ml) and galactitol 1%. After overnight incubation at 30°C, 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 Zeiss).

The selection coefficient (fitness gain) of the clones *in vivo* (presented in Fig 1) was estimated as:  $s_b = \ln\left(\frac{R_{f_{ev/anc}}}{R_{i_{ev/anc}}}\right) / t$ , where  $s_b$  is the selective advantage of the evolved clone,  $R_{f_{ev/anc}}$  and

$R_{ev/anc}$  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 [56,57].

### *In vitro* tests for changes in 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 [8] 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 37°C (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 (S2 Table) 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 the following alterations: after an initial aerobic growth overnight 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<sub>2</sub>, 15% CO<sub>2</sub>, 80% N<sub>2</sub>) (Plas Labs, Lansing, MI, USA), at 37°C. 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).

### 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 and grown in 10 mL of LB at 37°C with agitation for DNA extraction (following a previously described protocol [58]). 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. Mutations were identified using the BRESEQ pipeline [59]. To detect potential duplication events we used ssaha2 [60] 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) [61]. Data presented in S1 Table.

**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. Two populations, from the evolution experiment, were sequenced: 2.7 and 2.10. 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 and between ~100x and ~120x for population 2.10. 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. Data presented in [S3 Table](#).

### Analysis of gene expression changes caused by IS insertions

To determine the effects of the IS insertions identified during the 2<sup>nd</sup> steps of adaptation we measured the expression of *focA*, *dcuB*, *arcA*, *yjyY* and *yjyP* by RT-qPCR in two environments with different levels of oxygen. Five biological replicates and three technical replicates per clone were performed.

**Aerobic Conditions:** The clones were initially grown for 24h at 37°C with aeration in MM with glycerol (0.02%). The cultures were diluted 10-fold and 100  $\mu$ l of the dilution were inoculated in 10ml of M9 minimal medium (MM) supplemented with a mixture of the following carbon sources: sorbitol, ribose, mannose, gluconate and glucuronate, at individual concentration of 0.01%. The cultures were grown at 37°C, with aeration, until an OD<sub>600</sub> of 0.5. Five milliliters of the bacterial culture were then harvested by centrifugation at 4°C for 5 minutes at the maximum speed. The resulting pellet was resuspended in lysozyme solution (5 mg lysozyme/ml DEPC treated water, Sigma protocol) and incubated at 37°C 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 37°C in an anaerobic chamber with the atmosphere of 5% H<sub>2</sub>, 15% CO<sub>2</sub>, 80% N<sub>2</sub> (Plas Labs, Lansing, MI, USA), and at approximately OD<sub>600</sub> 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), 0.5  $\mu$ l of DNase and 1  $\mu$ l buffer were added to 1  $\mu$ g of RNA and incubated for 30 minutes at 37°C. After this, 1  $\mu$ l stop solution was added and then incubated for 15 minutes at 65°C to inactivate the DNase. The resulting RNA was used for the reverse transcription which consisted in mixing with 1  $\mu$ g of RNA, with 0.5  $\mu$ l random primers and DEPC-water (final volume of 15  $\mu$ l) and then incubated at 70°C for 5min. Afterwards the M-MLV Reverse Transcriptase Protocol (Promega) were performed, to the first mix was added 5  $\mu$ l of RT buffer, 0.5  $\mu$ l RT enzyme and 2  $\mu$ l dNTP mix, and then incubated 10 min at 25°C, 50min at 50°C and 10 min at 70°C.

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 50°C and then 39 cycles of 10 min at 95°C, 30 sec at 95°C, 1 min at 57°C and finally 30 s at 72°C. Primers used are listed in [S5 Table](#). Melting curve

analysis was performed to verify product homogeneity. All reactions included three replicates for each sample. Data were normalized by the Pfaffl method [19] 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 difference in the mutation rate of 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) [62] 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 37°C 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 37°C. 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 [25].

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).

### 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 6A and 6B 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 screened (from populations 2.7 and 2.10) using the same primers and PCR conditions as in [8]. Because all target mutations correspond to IS insertions an increase in the PCR band is indicative of the presence of an IS. Frequencies are depicted in S3 Table.

### 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. All statistical analysis were conducted with the statistical software R [63].

### Accession Numbers

Genome sequencing data have been deposited in the NCBI Read Archive, <http://www.ncbi.nlm.nih.gov/sra> (accession no. SRP063701).

### Supporting Information

**S1 Fig. Position of the adaptive IS insertions.** Exact position (except for the small sequence duplicated upon IS insertion, shaded in green) of the IS insertions in the regulatory regions of *dcuB*, *focA*, *arcA* and *yjiP*. The genomic coordinates of the sequences represented are indicated

between brackets. The start codon of each gene is shaded in yellow and marked with the letter M (methionine). Arrows represent positions of transcription start site. Boxed sequences represent binding sites for transcriptional activators (in green) or repressors (in red). D, N, F, C, A, I, H correspond to the following DNA-binding transcriptional dual regulators: DcuR-Phosphorylated, NarL Phosphorylated, FNR, CRP-cAMP, ArcA-phosphorylated, IHF and H-NS. The sequence annotation is according to EcoCyc [64].

(TIF)

**S2 Fig. Dynamics of the 2<sup>nd</sup> step mutations competing in the two backgrounds (*gat*-positive and *gat*-negative) against the respective ancestors.** Shown are the frequencies of the 2<sup>nd</sup> step mutations, labelled with the *yfp* allele, along 3 days of competition. (A) Competitions in the *gat*-positive background. (B) Competitions in the *gat*-negative background. The natural logarithm of the ratio of each mutant to the ancestor over the first 3 days of competition was used to estimate the selection coefficients depicted in S3 Table and Fig 5. Four independent competitions were performed to test each mutation. Error bars represent 2SE.

(TIF)

**S1 Table. Identity of the mutations detected in the sequenced clones.** Mutations were identified by comparison with the ancestor. The ancestor differs from (NCBI Reference Sequence: NC\_000913.2) in the positions reported in (1) for the 0YFP clone plus 1pb insertion (+C) in the coordinate 2172079 (*gatC*, coding (222/1356 nt)). Mutations in intergenic regions, as for example *yjyY/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.

(XLSX)

**S2 Table. 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 selection coefficient (*s*) of each clone in relation to the reference strain (see Material and Methods for details on how we performed the competitions and estimate *s*). Yellow, green and red shading means that *s* is not significantly different, is 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.

(XLSX)

**S3 Table. Selection coefficients of the 2<sup>nd</sup> step mutations.** Rows named “*s* (*gat*-pos)” and “*s* (*gat*-neg)” show the selection coefficients of the 2<sup>nd</sup> step mutations (listed in the first row) measured in direct competition (*in vivo*) with the respective ancestors (see also the S2 Fig and Fig 5). The error estimates associated with each measurement represents 2SE. Shaded in red are the estimates in the *gat*-positive background and in blue in the *gat*-negative. The asterisks denote the significance of the pairwise comparisons of the effects of mutations in the each of the backgrounds (ANOVA with Tuckey correction for multiple comparisons. \*  $P \leq 0.05$ , \*\*  $P \leq 0.01$  and \*\*\*  $P \leq 0.001$ ). The diagonal shows the comparison of the effect of the same

mutation in the two backgrounds (Mann-Whitney-Wilcoxon). The green rectangle highlights a mutation with positive epistasis with the background whereas the red triangle highlights negative epistasis. N.A. means not applicable. N.S. stands for non-significant.

(XLSX)

**S4 Table. Genetic polymorphism in two evolved populations.** Mutations were detected by whole genome sequencing of samples of each population at different time points (shown in red) or by target sequencing (shown in black). The full data is graphically represented in Fig 5.

(XLSX)

**S5 Table. Primers used for the RT-qPCR experiment.** The first row indicates the gene targeted by the qPCR primers. The resulting measurements, relative expression levels of the target genes, are depicted in Fig 2.

(XLSX)

## Acknowledgments

We thank Thomas Lenormand for critical discussion on testing the causes of the order of the adaptive steps of *E. coli* colonizing the gut. We thank Jessica Thompson for comments on the manuscript.

## Author Contributions

**Conceptualization:** AS ML IG.

**Formal analysis:** AS ML RSR IG JBB.

**Funding acquisition:** IG.

**Investigation:** ML AS RSR DG JBB.

**Methodology:** AS ML IG.

**Resources:** IG KBX.

**Supervision:** AS IG.

**Writing – original draft:** AS IG ML.

## References

1. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012; 489: 220–230. doi: [10.1038/nature11550](https://doi.org/10.1038/nature11550) PMID: [22972295](https://pubmed.ncbi.nlm.nih.gov/22972295/)
2. Ahern PP, Faith JJ, Gordon JI. Mining the Human Gut Microbiota for Effector Strains that Shape the Immune System. *Immunity*. 2014; 40: 815–823. doi: [10.1016/j.immuni.2014.05.012](https://doi.org/10.1016/j.immuni.2014.05.012) PMID: [24950201](https://pubmed.ncbi.nlm.nih.gov/24950201/)
3. Greenblum S, Carr R, Borenstein E. Extensive Strain-Level Copy-Number Variation across Human Gut Microbiome Species. *Cell*. 2015; 160: 583–594. doi: [10.1016/j.cell.2014.12.038](https://doi.org/10.1016/j.cell.2014.12.038) PMID: [25640238](https://pubmed.ncbi.nlm.nih.gov/25640238/)
4. Waldor MK, Tyson G, Borenstein E, Ochman H, Moeller A, Finlay BB, et al. Where Next for Microbiome Research? *PLOS Biol*. 2015; 13: e1002050. doi: [10.1371/journal.pbio.1002050](https://doi.org/10.1371/journal.pbio.1002050) PMID: [25602283](https://pubmed.ncbi.nlm.nih.gov/25602283/)
5. Levin BR. Frequency-dependent selection in bacterial populations. *Philos Trans R Soc Lond B Biol Sci*. 1988; 319: 459–472. PMID: [2905487](https://pubmed.ncbi.nlm.nih.gov/2905487/)
6. Falush D. Toward the Use of Genomics to Study Microevolutionary Change in Bacteria. Guttman DS, editor. *PLoS Genet*. 2009; 5: e1000627. doi: [10.1371/journal.pgen.1000627](https://doi.org/10.1371/journal.pgen.1000627) PMID: [19855823](https://pubmed.ncbi.nlm.nih.gov/19855823/)
7. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. *Nat Rev Genet*. 2007; 8: 610–618. doi: [10.1038/nrg2146](https://doi.org/10.1038/nrg2146) PMID: [17637733](https://pubmed.ncbi.nlm.nih.gov/17637733/)

8. Barroso-Batista J, Sousa A, Lourenço M, Bergman M-L, Sobral D, Demengeot J, et al. The First Steps of Adaptation of *Escherichia coli* to the Gut Are Dominated by Soft Sweeps. *Coop G*, editor. *PLoS Genet*. 2014; 10: e1004182. doi: [10.1371/journal.pgen.1004182](https://doi.org/10.1371/journal.pgen.1004182) PMID: [24603313](https://pubmed.ncbi.nlm.nih.gov/24603313/)
9. Hegreness M. An Equivalence Principle for the Incorporation of Favorable Mutations in Asexual Populations. *Science*. 2006; 311: 1615–1617. doi: [10.1126/science.1122469](https://doi.org/10.1126/science.1122469) PMID: [16543462](https://pubmed.ncbi.nlm.nih.gov/16543462/)
10. Long A, Liti G, Luptak A, Tenailon O. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nat Rev Genet*. 2015; 16: 567–582. doi: [10.1038/nrg3937](https://doi.org/10.1038/nrg3937) PMID: [26347030](https://pubmed.ncbi.nlm.nih.gov/26347030/)
11. Engel P, Krämer R, Uden G. Transport of C4-dicarboxylates by anaerobically grown *Escherichia coli*. Energetics and mechanism of exchange, uptake and efflux. *Eur J Biochem FEBS*. 1994; 222: 605–614.
12. Suppmann B, Sawers G. 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*. 1994; 11: 965–982. PMID: [8022272](https://pubmed.ncbi.nlm.nih.gov/8022272/)
13. Jones SA, Chowdhury FZ, Fabich AJ, Anderson A, Schreiner DM, House AL, et al. Respiration of *Escherichia coli* in the Mouse Intestine. *Infect Immun*. 2007; 75: 4891–4899. doi: [10.1128/IAI.00484-07](https://doi.org/10.1128/IAI.00484-07) PMID: [17698572](https://pubmed.ncbi.nlm.nih.gov/17698572/)
14. Jones SA, Gibson T, Maltby RC, Chowdhury FZ, Stewart V, Cohen PS, et al. Anaerobic Respiration of *Escherichia coli* in the Mouse Intestine. *Infect Immun*. 2011; 79: 4218–4226. doi: [10.1128/IAI.05395-11](https://doi.org/10.1128/IAI.05395-11) PMID: [21825069](https://pubmed.ncbi.nlm.nih.gov/21825069/)
15. Stratmann T, Madhusudan S, Schnetz K. Regulation of the yjjQ-bglJ Operon, Encoding LuxR-Type Transcription Factors, and the Divergent yjjP Gene by H-NS and LeuO. *J Bacteriol*. 2008; 190: 926–935. doi: [10.1128/JB.01447-07](https://doi.org/10.1128/JB.01447-07) PMID: [18055596](https://pubmed.ncbi.nlm.nih.gov/18055596/)
16. Andrews JC, Short SA. opp-lac Operon fusions and transcriptional regulation of the *Escherichia coli* trp-linked oligopeptide permease. *J Bacteriol*. 1986; 165: 434–442. PMID: [3080404](https://pubmed.ncbi.nlm.nih.gov/3080404/)
17. Sousa A, Bourgard C, Wahl LM, Gordo I. Rates of transposition in *Escherichia coli*. *Biol Lett*. 2013; 9: 20130838–20130838. doi: [10.1098/rsbl.2013.0838](https://doi.org/10.1098/rsbl.2013.0838) PMID: [24307531](https://pubmed.ncbi.nlm.nih.gov/24307531/)
18. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet*. 2012; 13: 59–69. doi: [10.1038/nrg3095](https://doi.org/10.1038/nrg3095) PMID: [22143240](https://pubmed.ncbi.nlm.nih.gov/22143240/)
19. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001; 29: e45. PMID: [11328886](https://pubmed.ncbi.nlm.nih.gov/11328886/)
20. Leonhartsberger S, Korsa I, Böck A. The molecular biology of formate metabolism in enterobacteria. *J Mol Microbiol Biotechnol*. 2002; 4: 269–276. PMID: [11931558](https://pubmed.ncbi.nlm.nih.gov/11931558/)
21. Park DM, Akhtar MS, Ansari AZ, Landick R, Kiley PJ. The Bacterial Response Regulator ArcA Uses a Diverse Binding Site Architecture to Regulate Carbon Oxidation Globally. *Burkholder WF*, editor. *PLoS Genet*. 2013; 9: e1003839. doi: [10.1371/journal.pgen.1003839](https://doi.org/10.1371/journal.pgen.1003839) PMID: [24146625](https://pubmed.ncbi.nlm.nih.gov/24146625/)
22. Chang D-E, Smalley DJ, Tucker DL, Leatham MP, Norris WE, Stevenson SJ, et al. Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc Natl Acad Sci U S A*. 2004; 101: 7427–7432. doi: [10.1073/pnas.0307888101](https://doi.org/10.1073/pnas.0307888101) PMID: [15123798](https://pubmed.ncbi.nlm.nih.gov/15123798/)
23. Rokyta DR, Joyce P, Caudle SB, Wichman HA. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nat Genet*. 2005; 37: 441–444. doi: [10.1038/ng1535](https://doi.org/10.1038/ng1535) PMID: [15778707](https://pubmed.ncbi.nlm.nih.gov/15778707/)
24. de Visser JAGM, Krug J. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet*. 2014; 15: 480–490. doi: [10.1038/nrg3744](https://doi.org/10.1038/nrg3744) PMID: [24913663](https://pubmed.ncbi.nlm.nih.gov/24913663/)
25. Hall BM, Ma C-X, Liang P, Singh KK. Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics*. 2009; 25: 1564–1565. doi: [10.1093/bioinformatics/btp253](https://doi.org/10.1093/bioinformatics/btp253) PMID: [19369502](https://pubmed.ncbi.nlm.nih.gov/19369502/)
26. Hottes AK, Freddolino PL, Khare A, Donnell ZN, Liu JC, Tavazoie S. Bacterial Adaptation through Loss of Function. *Matic I*, editor. *PLoS Genet*. 2013; 9: e1003617. doi: [10.1371/journal.pgen.1003617](https://doi.org/10.1371/journal.pgen.1003617) PMID: [23874220](https://pubmed.ncbi.nlm.nih.gov/23874220/)
27. Barroso-Batista J, Demengeot J, Gordo I. Adaptive immunity increases the pace and predictability of evolutionary change in commensal gut bacteria. *Nat Commun*. 2015; 6: 8945. doi: [10.1038/ncomms9945](https://doi.org/10.1038/ncomms9945) PMID: [26615893](https://pubmed.ncbi.nlm.nih.gov/26615893/)
28. Weinreich DM, Watson RA, Chao L. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evol Int J Org Evol*. 2005; 59: 1165–1174.
29. Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*. 2011; 332: 1193–1196. doi: [10.1126/science.1203801](https://doi.org/10.1126/science.1203801) PMID: [21636772](https://pubmed.ncbi.nlm.nih.gov/21636772/)

30. Chou H-H, Chiu H-C, Delaney NF, Segrè D, Marx CJ. Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science*. 2011; 332: 1190–1192. doi: [10.1126/science.1203799](https://doi.org/10.1126/science.1203799) PMID: [21636771](https://pubmed.ncbi.nlm.nih.gov/21636771/)
31. Rokyta DR, Joyce P, Caudle SB, Miller C, Beisel CJ, Wichman HA. Epistasis between beneficial mutations and the phenotype-to-fitness Map for a ssDNA virus. *PLoS Genet*. 2011; 7: e1002075. doi: [10.1371/journal.pgen.1002075](https://doi.org/10.1371/journal.pgen.1002075) PMID: [21655079](https://pubmed.ncbi.nlm.nih.gov/21655079/)
32. Miller CR, Joyce P, Wichman HA. Mutational Effects and Population Dynamics During Viral Adaptation Challenge Current Models. *Genetics*. 2011; 187: 185–202. doi: [10.1534/genetics.110.121400](https://doi.org/10.1534/genetics.110.121400) PMID: [21041559](https://pubmed.ncbi.nlm.nih.gov/21041559/)
33. Schoustra SE, Bataillon T, Gifford DR, Kassen R. The properties of adaptive walks in evolving populations of fungus. *PLoS Biol*. 2009; 7: e1000250. doi: [10.1371/journal.pbio.1000250](https://doi.org/10.1371/journal.pbio.1000250) PMID: [19956798](https://pubmed.ncbi.nlm.nih.gov/19956798/)
34. Couce A, Tenaillon OA. The rule of declining adaptability in microbial evolution experiments. *Front Genet*. 2015; 6: 99. doi: [10.3389/fgene.2015.00099](https://doi.org/10.3389/fgene.2015.00099) PMID: [25815007](https://pubmed.ncbi.nlm.nih.gov/25815007/)
35. Kopp M, Hermisson J. The genetic basis of phenotypic adaptation II: the distribution of adaptive substitutions in the moving optimum model. *Genetics*. 2009; 183: 1453–1476. doi: [10.1534/genetics.109.106195](https://doi.org/10.1534/genetics.109.106195) PMID: [19805820](https://pubmed.ncbi.nlm.nih.gov/19805820/)
36. Gordo I, Campos PRA. Evolution of clonal populations approaching a fitness peak. *Biol Lett*. 2012; 9: 20120239–20120239. doi: [10.1098/rsbl.2012.0239](https://doi.org/10.1098/rsbl.2012.0239) PMID: [22764110](https://pubmed.ncbi.nlm.nih.gov/22764110/)
37. Mustonen V, Lässig M. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends Genet TIG*. 2009; 25: 111–119. doi: [10.1016/j.tig.2009.01.002](https://doi.org/10.1016/j.tig.2009.01.002) PMID: [19232770](https://pubmed.ncbi.nlm.nih.gov/19232770/)
38. Sommer F, Bäckhed F. Know your neighbor: Microbiota and host epithelial cells interact locally to control intestinal function and physiology. *BioEssays*. 2016; 38: 455–464. doi: [10.1002/bies.201500151](https://doi.org/10.1002/bies.201500151) PMID: [26990415](https://pubmed.ncbi.nlm.nih.gov/26990415/)
39. Cryan JF, Dinan TG. More than a Gut Feeling: the Microbiota Regulates Neurodevelopment and Behavior. *Neuropsychopharmacology*. 2015; 40: 241–242. doi: [10.1038/npp.2014.224](https://doi.org/10.1038/npp.2014.224) PMID: [25482171](https://pubmed.ncbi.nlm.nih.gov/25482171/)
40. Woodward MJ, Charles HP. Polymorphism in *Escherichia coli*: rtl atl and gat regions behave as chromosomal alternatives. *J Gen Microbiol*. 1983; 129: 75–84. doi: [10.1099/00221287-129-1-75](https://doi.org/10.1099/00221287-129-1-75) PMID: [6339679](https://pubmed.ncbi.nlm.nih.gov/6339679/)
41. Carroll SM, Lee M-C, Marx CJ. Sign epistasis limits evolutionary trade-offs at the confluence of single- and multi-carbon metabolism in *Methylobacterium extorquens* AM1. *Evol Int J Org Evol*. 2014; 68: 760–771. doi: [10.1111/evo.12301](https://doi.org/10.1111/evo.12301) PMID: [24164359](https://pubmed.ncbi.nlm.nih.gov/24164359/)
42. Gresham D, Hong J. The functional basis of adaptive evolution in chemostats. *FEMS Microbiol Rev*. 2014; n/a–n/a. doi: [10.1111/1574-6976.12082](https://doi.org/10.1111/1574-6976.12082) PMID: [25098268](https://pubmed.ncbi.nlm.nih.gov/25098268/)
43. Horiuchi T, Tomizawa JI, Novick A. Isolation and properties of bacteria capable of high rates of beta-galactosidase synthesis. *Biochim Biophys Acta*. 1962; 55: 152–163. PMID: [14448925](https://pubmed.ncbi.nlm.nih.gov/14448925/)
44. Brown CJ, Todd KM, Rosenzweig RF. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Mol Biol Evol*. 1998; 15: 931–942. PMID: [9718721](https://pubmed.ncbi.nlm.nih.gov/9718721/)
45. Maharjan RP, Liu B, Feng L, Ferenci T, Wang L. Simple phenotypic sweeps hide complex genetic changes in populations. *Genome Biol Evol*. 2015; 7: 531–544. doi: [10.1093/gbe/evv004](https://doi.org/10.1093/gbe/evv004) PMID: [25589261](https://pubmed.ncbi.nlm.nih.gov/25589261/)
46. Maharjan RP, Ferenci T, Reeves PR, Li Y, Liu B, Wang L. The multiplicity of divergence mechanisms in a single evolving population. *Genome Biol*. 2012; 13: R41. doi: [10.1186/gb-2012-13-6-r41](https://doi.org/10.1186/gb-2012-13-6-r41) PMID: [22682524](https://pubmed.ncbi.nlm.nih.gov/22682524/)
47. Jerison ER, Desai MM. Genomic investigations of evolutionary dynamics and epistasis in microbial evolution experiments. *Curr Opin Genet Dev*. 2015; 35: 33–39. doi: [10.1016/j.gde.2015.08.008](https://doi.org/10.1016/j.gde.2015.08.008) PMID: [26370471](https://pubmed.ncbi.nlm.nih.gov/26370471/)
48. de Visser JAGM, Cooper TF, Elena SF. The causes of epistasis. *Proc R Soc B Biol Sci*. 2011; 278: 3617–3624. doi: [10.1098/rspb.2011.1537](https://doi.org/10.1098/rspb.2011.1537) PMID: [21976687](https://pubmed.ncbi.nlm.nih.gov/21976687/)
49. Plucaïn J, Hindre T, Le Gac M, Tenaillon O, Cruveiller S, Medigue C, et al. Epistasis and Allele Specificity in the Emergence of a Stable Polymorphism in *Escherichia coli*. *Science*. 2014; 343: 1366–1369. doi: [10.1126/science.1248688](https://doi.org/10.1126/science.1248688) PMID: [24603152](https://pubmed.ncbi.nlm.nih.gov/24603152/)
50. Chou H-H, Delaney NF, Draghi JA, Marx CJ. Mapping the Fitness Landscape of Gene Expression Uncovers the Cause of Antagonism and Sign Epistasis between Adaptive Mutations. Achtman M, editor. *PLoS Genet*. 2014; 10: e1004149. doi: [10.1371/journal.pgen.1004149](https://doi.org/10.1371/journal.pgen.1004149) PMID: [24586190](https://pubmed.ncbi.nlm.nih.gov/24586190/)

51. Neverov AD, Kryazhimskiy S, Plotkin JB, Bazykin GA. Coordinated Evolution of Influenza A Surface Proteins. Malik HS, editor. PLOS Genet. 2015; 11: e1005404. doi: [10.1371/journal.pgen.1005404](https://doi.org/10.1371/journal.pgen.1005404) PMID: [26247472](https://pubmed.ncbi.nlm.nih.gov/26247472/)
52. Cooper VS, Schneider D, Blot M, Lenski RE. Mechanisms Causing Rapid and Parallel Losses of Ribose Catabolism in Evolving Populations of *Escherichia coli* B. J Bacteriol. 2001; 183: 2834–2841. doi: [10.1128/JB.183.9.2834-2841.2001](https://doi.org/10.1128/JB.183.9.2834-2841.2001) PMID: [11292803](https://pubmed.ncbi.nlm.nih.gov/11292803/)
53. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of *Escherichia coli* K-12. Science. 1997; 277: 1453–62. PMID: [9278503](https://pubmed.ncbi.nlm.nih.gov/9278503/)
54. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006; 2: 2006.0008. doi: [10.1038/msb4100050](https://doi.org/10.1038/msb4100050) PMID: [16738554](https://pubmed.ncbi.nlm.nih.gov/16738554/)
55. Krogfelt KA, Cohen PS, Conway T. The Life of Commensal *Escherichia coli* in the Mammalian Intestine. EcoSal Plus. 2013; 1. doi: [10.1128/ecosalplus.8.3.1.2](https://doi.org/10.1128/ecosalplus.8.3.1.2) PMID: [26443354](https://pubmed.ncbi.nlm.nih.gov/26443354/)
56. Poulsen LK, Licht TR, Rang C, Krogfelt KA, Molin S. Physiological state of *Escherichia coli* BJ4 growing in the large intestines of streptomycin-treated mice. J Bacteriol. 1995; 177: 5840–5845. PMID: [7592332](https://pubmed.ncbi.nlm.nih.gov/7592332/)
57. Rang CU, Licht TR, Midtvedt T, Conway PL, Chao L, Krogfelt KA, et al. Estimation of growth rates of *Escherichia coli* BJ4 in streptomycin-treated and previously germfree mice by in situ rRNA hybridization. Clin Diagn Lab Immunol. 1999; 6: 434–436. PMID: [10225851](https://pubmed.ncbi.nlm.nih.gov/10225851/)
58. Wilson K. Preparation of genomic DNA from bacteria. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, et al., editors. Current Protocols in Molecular Biology. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2001. p. 00:2.4.1–2.4.5. <http://doi.wiley.com/10.1002/0471142727.mb0204s56>
59. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, et al. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature. 2009; 461: 1243–1247. doi: [10.1038/nature08480](https://doi.org/10.1038/nature08480) PMID: [19838166](https://pubmed.ncbi.nlm.nih.gov/19838166/)
60. Ning Z, Cox AJ, Mullikin JC. SSAHA: a fast search method for large DNA databases. Genome Res. 2001; 11: 1725–1729. doi: [10.1101/gr.194201](https://doi.org/10.1101/gr.194201) PMID: [11591649](https://pubmed.ncbi.nlm.nih.gov/11591649/)
61. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29: 24–26. doi: [10.1038/nbt.1754](https://doi.org/10.1038/nbt.1754) PMID: [21221095](https://pubmed.ncbi.nlm.nih.gov/21221095/)
62. Reiner AM. Xylitol and D-arabitol toxicities due to derepressed fructose, galactitol, and sorbitol phosphotransferases of *Escherichia coli*. J Bacteriol. 1977; 132: 166–173. PMID: [334721](https://pubmed.ncbi.nlm.nih.gov/334721/)
63. R Core Team. R: A language and environment for statistical computing [Internet]. R Foundation for Statistical Computing, Vienna, Austria; 2016. <https://www.R-project.org/>
64. Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muniz-Rascado L, et al. EcoCyc: a comprehensive database of *Escherichia coli* biology. Nucleic Acids Res. 2011; 39: D583–D590. doi: [10.1093/nar/gkq1143](https://doi.org/10.1093/nar/gkq1143) PMID: [21097882](https://pubmed.ncbi.nlm.nih.gov/21097882/)