Published in final edited form as:

Mol Ecol. 2017 April ; 26(7): 1802–1817. doi:10.1111/mec.13851.

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

Mathilde Lescat^{#1,2,3}, Adrien Launay^{#1,4}, Mohamed Ghalayini^{1,2}, Mélanie Magnan^{1,4}, Jérémy Glodt^{1,4}, Coralie Pintard^{1,4}, Sara Dion^{1,4}, Erick Denamur^{1,4,5}, and Olivier Tenaillon^{1,4,#}

¹INSERM, IAME, UMR 1137, Paris, France

²Univ Paris Nord, Sorbonne Paris Cité, Bobigny, France

³APHP, Hôpitaux Universitaires Paris Seine Saint-Denis, Bondy, France

⁴Univ Paris Diderot, Sorbonne Paris Cité, Paris, France

⁵APHP, Hôpitaux Universitaires Paris Nord Val de Seine, Paris, France

These authors contributed equally to this work.

Abstract

Though microbial ecology of the gut is now a major focus of interest, little is known about the molecular determinants of microbial adaptation in the gut. Experimental evolution coupled with whole genome sequencing can provide insights of the adaptive process. *In vitro* experiments have revealed some conserved patterns: intermediate convergence, epistatic interactions between beneficial mutations and mutations in global regulators. To test the relevance of these patterns and to identify the selective pressures acting *in vivo*, we have performed a long-term adaptation of an *E. coli* natural isolate, the streptomycin resistant strain 536, in the digestive tract of streptomycin treated mice. After a year of evolution, a clone from 15 replicates was sequenced. Consistently with *in vitro* observations, the identified mutations revealed a strong pattern of convergence at the mutation, gene, operon and functional levels. Yet, the rate of molecular evolution was lower than in *in vitro* and no mutations in global regulators were recovered. More specific targets were observed: the *dgo* operon, involved in the galactonate pathway that improved growth on D-galactonate, and *rluD* and *gidB*, implicated in the maturation of the ribosomes, which mutations improved growth only in the presence of streptomycin. As *in vitro*, the non-random associations of mutations within the same pathways suggested a role of epistasis in shaping the adaptive landscape. Overall, we show that “evolve and sequence” approach coupled to an analysis of

#Corresponding author: Olivier Tenaillon, IAME, INSERM UMR 1137, Faculté de Médecine Paris Diderot, Site Xavier Bichat, 16 rue Henri Huchard, 75018 Paris, France. olivier.tenaillon@inserm.fr.

Data accessibility

The data used to produce the figure 1, 4, 5 and 6 are accessible on the Dryad website : (<http://datadryad.org>) with the number doi: 10.5061/dryad.4g503 and in the table S1

The data used to produce the figures 2 and 3 are presented in the table S2 and in the MicroScope website (<http://www.genoscope.cns.fr/agc/microscope/home>).

convergence, when applied to a natural isolate, can be used to study adaptation *in vivo* and uncover the specific selective pressures of that environment.

Keywords

Escherichia coli; experimental evolution; gut colonization mouse model; 536; streptomycin resistance; D-galactonate

Background

Bacteria have evolved for billions of years and colonized all available ecological niches. Many studies have been performed to understand how evolution has shaped the genome of these complex organisms. Comparative genomics has unravelled the dynamics of gene acquisition and losses in bacterial genomes (Touchon *et al.* 2009), and the general constraints acting on genome composition (Hershberg & Petrov 2010). Yet, much less is known about bacterial genomes evolution in the short term in their natural habitat. While we know through the emergence of resistance to antibiotics that bacterial evolution can be caught in the act, most of our evidence of short-term evolution in the wild is linked to drastic selective pressures such as antibiotics (Wong *et al.* 2012) or escape of the immune system (Flint *et al.* 2016). Can other selective pressures be strong enough to drive significant changes in bacterial genomes in the short term? To investigate that possibility, we have decided to study the evolution of a natural isolate of *Escherichia coli* in the gut of mice.

E. coli is one of the most characterised bacterial species; it has been used as a model organism in the laboratory to unravel the multiple facets of bacterial biology from biochemistry to genetics. Yet, the natural environment of *E. coli* is not the laboratory. In the wild, *E. coli* is a versatile species, known as both a widespread gut commensal of the vertebrates and a dangerous pathogen that can also be retrieved in the environment (Tenaillon *et al.* 2010). As a pathogen *E. coli* is responsible for about 1 million human deaths yearly due to intra and now mostly extra intestinal diseases (Russo & Johnson 2003). The species is a structured species subdivided in seven phylogenetic groups A, B1, B2, C, D, E and F (Jauregui *et al.* 2008). These groups are not randomly distributed. Previous studies have shown an association between the phylogenetic groups and virulence, with most extra intestinal *E. coli* pathogens (including urinary tract infection and meningitis strains) belonging to phylogenetic group B2 (Bingen *et al.* 1998; Picard *et al.* 1999). The prevalence of the different phylogenetic groups as commensals varies largely across host species and even within host populations (Tenaillon *et al.* 2010). For instance, the B2 phylogenetic group can be frequently retrieved in wild animal's guts (Gordon & Cowling 2003; Lescat *et al.* 2013), rarely found (<10%) as a commensal among human tropical populations while it is dominant (>30%) among the populations of industrialized countries (Tenaillon *et al.* 2010). Indeed, over the last decades, the prevalence of the B2 phylogroup has significantly increased in frequency in some western countries such as France. B2 phylogroup strains being both highly efficient extra intestinal pathogens, and more and more prevalent as commensals in industrialized countries, they are becoming an increasing public health

concern. Yet, the molecular determinants explaining the differences in prevalence of the different subgroups of *E. coli* among different hosts or populations remain largely unknown.

To identify the molecular bases that may lead *E. coli* to specialize to certain environmental conditions, two main approaches can be used: comparative genomics of natural isolates and experimental evolution. Thanks to the sequencing revolution, more than two hundred complete genomes of *E. coli* are now accessible and thousands more are underway. We can subsequently compare these genomes in order to identify some genetic characteristics likely to contribute to the phenotypic differentiation of the subgroups. For instance, we compared 130 sequenced genomes to identify the core region that was the most different between B2 to non-B2 strains: the *nhaAR* operon involved in pH control. Though we could find a contribution of the operon in extra intestinal virulence, we were not able to observe its implication in colonization of the mouse gut (Lescat *et al.* 2014) nor in any pH linked phenotypes we could measure in the laboratory. This example illustrates the limitation of comparative genomics: though some specific genomic signatures may be found, we have not access to the environmental conditions that shaped them and it is therefore difficult to assess their evolutionary relevance.

An alternative approach relies in evolving bacteria in controlled conditions. *In vitro* experimental evolution has been used for decades to study bacterial adaptation. Over the last years, the coupling of experimental evolution with whole genome sequencing allowed progresses in the understanding of the genetic bases of adaptation. The most charismatic experiment in the field is the one initiated more than 25 years ago by Richard Lenski. His 12 lineages long-term evolution experiment (LTEE) has now reached more than 60,000 generations. The genomic data as well as targeted sequencing approaches have revealed convergence among the first targets of adaptation involving several global regulators (Maddamsetti *et al.* 2015). Another large scale study confirmed these results using 115 lineages adapting for 2,000 generations to high temperature: the first target of adaptation was the RNA polymerase one of the most pleiotropic gene in the genome, and a large convergence was observed between replicates (Tenaillon *et al.* 2012). These studies and others (Herring *et al.* 2006; Hindré *et al.* 2012; Le Gac *et al.* 2013) suggest that the first steps of adaptation may involve global modifications of gene expression rather than a specific small-scale optimization of some specific pathways. Interestingly, global regulators such as *rpoS*, have also been found to be involved in the diversification that seems to occur during infections (Levert *et al.* 2010). However, the relevance of these *in vitro* or *in infectio* observations in the primary habitat of *E. coli*, the gut of vertebrates, remains to be tested.

In vivo assays have also been performed to study adaptation to a more relevant environment for *E. coli*. In the early 2000's, Giraud *et al.* tested the implication of mutator alleles of *E. coli* in the adaptation to the gut of germ free mice and showed that mutator bacteria possessed a short-term advantage and long-term disadvantage in colonizing new environments (Giraud *et al.* 2001). Later, Giraud *et al.* also used germ-free mice inoculated with *E. coli* MG1655 (Giraud *et al.* 2008) and observed a rapid adaptation by a single point mutation in EnvZ/OmpR, a two component signal transduction system, which controls more than 100 genes. These observations suggest that global regulators could also be the targets of adaptation in the mouse gut. However, these works are entailed by an important limitation:

the gut of adult germ free mice is an artificial environment. Mice that have developed without flora have abnormal immune system and gut physiology (Khouri *et al.* 1969; Welling *et al.* 1980; Umekasi 2014). Furthermore, in these experiments, *E. coli* was alone in the gut. In the wild, *E. coli* is a minority species of the gut microbiota, which is largely dominated by numerous anaerobes (Arumugam *et al.* 2011). As a consequence, *E. coli* is facing intense competition for nutrients, and may also have access to other nutrients processed by the rest of the microbiota. More recently, Barroso-Batista *et al.* performed an evolution experiment using the laboratory strain *E. coli* K-12 in streptomycin treated mice. Streptomycin treatment allows the colonization of the digestive tract by a streptomycin resistant *E. coli*, without altering theoretically the anaerobic microbiota (Barroso-Batista *et al.* 2014). They observed the rapid apparition of mutations in the different lineages leading to convergence in inactivation of the *gat* operon conferring advantage to the strains bearing the mutations and unable to utilize the galactitol. Here the first target of adaptation was not a global regulator. However, one limitation of these studies is the use of a laboratory strain that has been subject to many genetic alterations and that has not seen a gut for close to a century. *E. coli* K-12 is a strain evolved in *in vitro* conditions since 1922, date of its isolation in Palo Alto, from a diphtheria convalescent patient. Genetic modifications, adaptation to oxygen rich media from the laboratory as well as accumulation of mutations favoured by UV irradiations or strong bottlenecks may modify its adaptation to the mouse gut.

In this context, the aims of this study are (i) to identify the basis of molecular adaptation in *in vivo* commensal conditions using long-term evolution during one year of a natural isolate of *E. coli* belonging to the phylogenetic group B2 in the digestive tract of streptomycin treated mice and (ii) to test the phenotypic significances of the molecular modifications observed and their possible advantages in this natural environment.

Materials and Methods

Bacterial strains

We used *E. coli* K-12 BW25113 (phylogenetic group A) from the KEIO collection (Baba *et al.* 2006) and 536 (phylogenetic group B2, serotype O6:K5:H31) reference strains. The 536 strain has been isolated from a patient suffering from urinary tract infection in the early 1980's (Berger *et al.* 1982) and subsequently shown to be a good gut colonizer in streptomycin treated mouse (Diard *et al.* 2010). This strain was acquired by our laboratory in 1998 from Jörg Hacker who isolated it and then stored at -80°C with glycerol.

Mutant strains of K-12, obtained from the KEIO collection (K-12 *dgoR*:kan, K-12 *rluD*:kan, K-12 *gidB*:kan) and of 536, obtained during the course of this study using the lambda-red disruption method (Datsenko & Wanner 2000) (536 *dgoR*:cm, 536 *rluD*:cm, 536 *gidB*:cm, 536 *gidB rluD*:cm), were also used. For the flux cytometry assays, we used K-12 *pdgoR* issued from the Zaslaver collection (Zaslaver *et al.* 2006) which bore the plasmid *pdgoR* which contained the *gfp* gene placed under the control of the promoter of the *dgoRKADT* operon. After extraction of this plasmid, we electroporated it in the K-12 *dgoR*:kan mutant strain. We thus obtained K-12 *dgoR pdgoR* strain.

Mouse intestinal colonization assay and sampling strategy

All animal experiments were performed in compliance with the recommendations of the French Ministry of Agriculture and approved by the French Veterinary Services (accreditation A 75-18-05). The mouse gut colonization model was conducted after approval by the Debré-Bichat Ethics Committee for Animal Experimentation (Protocol Number 2012-17/722-0076) in accordance to the European Decret and French law on the protection of animals. Intestinal colonization by 536 *E. coli* strain was performed using a mouse model as described elsewhere (Diard *et al.* 2010). Thirty 6-week-old female outbred mice (Charles River CD-1) were housed in 15 cages (numbered from C1 to C15) during one year (two mice per cage named Ma and Mb). Five days before inoculation of the 536 strain, and throughout the experiment, streptomycin was added to the sterile drinking water at a final concentration of 5 g/L. Likewise the mice were fed with sterile food all along the experiment. Mice free of coliform flora (controlled by plating the faeces on Drigalski plates) were inoculated *per os* with 10^6 bacteria in 200 µl of physiological water. The day after the inoculation and during the experiment (Days 1, 5, 33, 47, 155, 188, 217, 294, 336 and 363, the intestinal population of *E. coli* was estimated by plating dilutions of weighted fresh faeces for each mouse on lysogeny broth (LB) agar with streptomycin (5 g/L) and supplementary samples were performed for sequencing assays (Days 61, 124, 155 and 250). Faeces were harvested for each mouse by putting the mouse in a sterile plastic cage until it defecated. The faeces were stored at -80°C with glycerol for further analysis. For sequencing, the frozen faeces were diluted and plated on LB agar with streptomycin. One colony per LB plate was then randomly picked, grown in LB, stored at -80°C with glycerol and thereafter called “clone”. In addition, to study the reproducibility between the sample of the two mice from each cage, 50 colonies were picked from the faeces on the LB plate for each mouse (100 colonies in total) in cage 1 at Day 363. For the polymorphism study, we picked a pool of about 50 colonies from 4 mice (C2Ma, C3Ma, C8Ma and C11Ma) at both the last time point (Days 250, 336, 363 and 363, respectively) and an intermediate time point (Days 124, 155, 155 and 155, respectively).

High throughput sequencing

From the glycerol stocks, the clones were plated on LB plates, and one colony was randomly picked, and DNA was extracted using Wizard® genomic DNA purification kit (Promega). The genomes of the evolved clones, as well as that of the ancestor 536 strain used for inoculation were sequenced on the Illumina HiSeq 2000 (IntegraGen), with one line of single-end 51-bp reads per genome. We also sequenced after DNA extraction the pool of colonies for the reproducibility and polymorphism assays.

Sequence analyses

Point mutations were identified by comparing sequence reads to the genome of the ancestral strain 536, considered as the reference, using breseq 0.23, a pipeline for analyzing resequenced microbial genomes (Deatherage & Barrick 2014).

Reference reconstruction—In a first step, we sequenced the 536 reference genome and compared the reads to the original strain genome (Brzuszkiewicz *et al.* 2006) from the

MicroScope website (<https://www.genoscope.cns.fr/agc/microscope/mage/viewer.php>) to identify differences using breseq 0.23 (Deatherage & Barrick 2014). This pipeline maps the reads on an annotated reference genome in gbk format to precisely identify the mutations that occurred: the type of the mutation, the genes it fell into, or the closest ones, and in the case of point mutation the changes in amino acid associated. We then used gdttools utility program from breseq 0.27.1 (Deatherage & Barrick 2014) to integrate the identified mutations inside the reference genome. We repeated this operation of comparison and modification of the reference five times until there was no more mutation identified. The iterations are required to correct mutations in close neighbourhood of one another. Using this method, we identified a total of 351 mutations (230 being point mutations, and 121 indels) between our ancestral 536 strain and the one from the database we used, a number compatible with errors due to Sanger sequencing method at the genome level.

Detection of mutations—Once the reference genome was obtained, we compared this reference to the sequences of the different evolved clones using the same approach. The identified mutations were then manually checked. We did the same for the pooled colonies and also obtained the frequencies for each mutation from breseq 0.27.1 (Deatherage & Barrick 2014).

Quantification of convergence—To quantify how significant convergence at the gene level is, for each gene of the genome, we computed the expected number of mutations E_i , that would alter this gene if mutations were randomly spread in the chromosome. We compared that expectation to the observed results O_i through the computation of a G statistics for nonsynonymous mutations: in which the sum is made on all genes with some observed mutations. The number of degrees of freedom of G statistics being difficult to compute when many classes are null (genes with no mutations) we compared the observed G value to that of simulated dataset with mutations randomly spread (Tenaillon *et al.* 2016).

Phenotypic assays

Flux cytometry—The strain K-12 *pdgoR* and the mutant strain K-12 *ΔgoR pdgoR* were compared using flux cytometry as described elsewhere (Bleibtreu *et al.* 2013).

Growth curves—For the comparative growth assays, strains were grown at 37°C in four media, all supplemented with streptomycin at 5 g/L: LB and minimum Davies medium (DM) with glucose, gluconate or galactonate (20 mmol/L for each sugar). Each medium was adjusted at pH 7 and pH 5.7 with MOPS® and TAPS® products, respectively (Sigma). LB is a complex media, whereas DM is a minimal media with only one source of carbon. All the strains studied were grown overnight (O/N) in LB media in deep-well plates at 37°C with constant stirring at 280 rpm. O/N cultures were prediluted at 1/100 in physiological serum and strains were inoculated each at 1/100 in a Costar® 96 flat-bottomed well plate and recovered by 50µl of mineral oil. Growth was recorded by an Infinite 200 Tecan®, which measured the OD600 in each well every 5 minutes at 37°C, while stirring for 24 hours. Growth assays were repeated 3 times. The maximum growth rate (MGR) was calculated from growth curves obtained by Tecan®. OD600 in nm were collected and Log-transformed. Curves were calculated from a smoothed spline function. The MGR was defined as the time

point at which the maximum value of the derivative of the smoothed function was observed. The doubling times (DTs) in mn have then been computed as followed: $DT = \text{Log}2 / (\text{MGR} * 60)$.

Statistical analysis

The DTs were compared by strain and by medium using a Welch test. All statistics were computed using R (R Development Core Team, 2009, Vienna, Austria) and statistical significance was determined at a p-value of less than 0.05.

Results

Mouse colonization assay and clone sampling

Thirty mice in 15 independent cages (2 mice per cage) were colonized by the *E. coli* 536 strain. As it is known that mice are coprophages, we postulated that sampling one clone of one randomly selected mouse per cage should be representative of the population of *E. coli* colonizing both mice within a single cage. To validate this approach, we sampled in cage 1 at Days 61, 155, 250 and 363, one clone from each mouse as well as a mix of 50 clones at Day 363 for each mouse and performed sequencing of the individual clones and the mix. We observed a single mutation difference between the two individual clones per cage and this site was the only polymorphic site recovered in the pooled sequencing. These results validate our strategy, sampling a single mouse per cage being relevant.

The 536 strain was maintained in all alive mice throughout the experiment that lasted 363 days. However, ten mice died in the course of the experiment (figure 1), most of these were issued from different cages except for 2 cages (cages 7 and 14) in which the two mice died before day 363 (supplementary table 1). The average population density was $1.4 \cdot 10^9$ unity forming colony (UFC)/g of faeces along the experiment. More precisely, density first decreased from $3 \cdot 10^9$ UFC/g of faeces to $2 \cdot 10^4$ UFC/g of faeces at day 188 and stabilized at $6 \cdot 10^7$ until the end of experiment at day 363 (figure 1). One clone per cage has been used for sequencing. The sampling points corresponds (i) to the end of the experiment when at least one of the two mice of the cage was still alive (Day 363 for 10 cages: C1, C4, C5, C6, C8, C9, C10, C11, C13, C15), (ii) the time at which the last mouse of the cage died (Day 250 for cage C2, C7 and C14), and for three cages we could not recover *E. coli* from the frozen faeces and we sampled the last point that produced an *E. coli* clone (Day 124 for cage C12 and Day 336 for C3) (supplementary table 1).

Description of mutations in the 536 evolved clones

Using breseq pipeline (Deatherage & Barrick 2014), we identified 95 independent mutations that occurred during the experiment of mice gut evolution in the 15 genomes as shown in the supplementary table 2. Amongst those mutations, 70 were point mutations, 17 were small indels (less than 50bp), 6 were large indel events (>25bp) and two were IS insertion. For the point mutations, 44 were non-synonymous including 10 stop mutations; 13 were synonymous and 13 intergenic. The synonymous mutations recovered were only transitions and 11 out of 13 were of C-G to T-A type, a bias reported before (Wielgoss *et al.* 2011). Assuming 18 generations a day in the mice gut (Rang *et al.* 1999), we could compute a

mutation rate of $1.85 \cdot 10^{-10}$ per base-pair per generation using synonymous mutation. This is consistent with previous estimates of mutation rate of $8.9 \cdot 10^{-11}$ found in the LTEE (Wielgoss *et al.* 2011). The ratio of rate of non-synonymous to the rate of synonymous, Ka/Ks, was 1.90 (Confidence Interval (CI) estimated by bootstrapping mutations: 1.09-4.01). Such a value suggests that about 50% of non-synonymous mutations (including nonsense mutations) are the product of natural selection. Stop codons were even more overrepresented in the dataset compared to the expectation based on synonymous (Kstop/Ks=6.72, CI: 2.78-15.86) or on all non-stop mutations (Kstop/Ka_nonstop=4.51, CI: 1.96-8.39). This suggests a strong selection for gene inactivation.

Among the 17 small indels (<50bp), 13 lead to frameshifts, and none were intergenic. For the 10 lineages that survived over the 363 days, an average of 7 mutations were found per sequenced genome (4 to 11).

Convergence among lineages

Convergence was observed at several levels. First, while 95 mutations were recovered, they corresponded indeed to 76 different mutations. Six mutations were recovered in 2 to 8 lineages. For instance 8 lineages shared the same C to T mutation in an intergenic region between *yjiH* and *kptA*, and 5 shared a G to A non-sense mutation in *dgoR*. While these shared mutations could be a signal of cross contamination, the star like phylogeny of the samples and the mapping of mutations on that phylogeny suggest that these common mutations must have occurred multiple times (figure 2). In other words, using the previous two examples, the lineages having the *yjiH/kptA* mutations were not fully included in, fully encompassing or mutually exclusive with the lineages having the afore mentioned *dgoR* mutations: some lineages have both mutations, one, the other or none. This suggests that if the population evolved asexually, in at least one locus the same mutations occurred several times independently. These convergent mutations account for 26% (25/95) of the mutations found.

Convergence could be integrated further at the gene level. Nine genes were found to be mutated repeatedly in different cages. Adding a large deletion recovered 3 times, the mutations showing some sign of convergence among lineages at the gene level accounted for 56% (53/95) of the mutations found. With the limited number of mutations found in each genome, this is a clear sign of convergence (see below). This signal is further reinforced by the fact that all genes affected multiple times presented at least 2 different mutations at the origin of the convergence (table 1). The fact that different alleles are found in the different lineages, excludes the possibility that convergence is only due to a single mutation that was by chance at high frequency in the initial stock used to inoculate the different mice. For instance, ECP_4610, a putative transporter is affected in 5 lineages with 5 different mutations.

To quantify how significant this convergence at the gene level is, for each gene of the genome, we computed a statistical test for convergence (see method) in the form of a G statistics (Tenaillon *et al.* 2016). We found $G=482.3$ that has to be compared to a mean value of 399.2 for simulations, with a Z-score of 8.3. This suggests that the observed value is more than 8 standard deviations away from the simulations mean and results in a p-value lower

than 2.10^{-16} . This overwhelming signal of convergence suggests the filtering action of natural selection.

Among this list of genes showing signs of convergence, two clusters of linked genes can be further identified. *dgoR*, *dgoD* and *dgoT* are implicated in the same metabolic pathways. This observation suggests further that convergence may occur at a higher level of integration than the gene. In the same manner, 6 mutations in the *rluD* gene were found, and 7 in *gidB*. Both genes are implicated in the maturation of the ribosomes, and both are linked to streptomycin resistance.

Polymorphism within lineages

To have a better insight on the dynamics of adaptation, we sequenced a pool of about 50 colonies from 4 mice (C2Ma, C3Ma, C8Ma and C11Ma) at both the last time point (Days 250, 336, 363 and 363, respectively) and an intermediate time point (Days 124, 155, 155 and 155, respectively). Using an analysis of polymorphism in these data, we can detect mutations with a frequency higher than 5% in most samples. The analysis of the mutations found at the population level confirmed the presence of the mutations of interest we detected in the clone sequenced. Polymorphism data revealed however if mutations were fully fixed (frequency of 100%) in the population or if they had reached intermediate frequencies. Moreover, several mutations affecting the same gene were found simultaneously in some populations. This suggests some competition between these various alleles. Interestingly, all cases of multiple allele affecting the same gene were restricted to the genes we identified as target of selection with the clone sequencing approach (supplementary table 2). This confirms that we have identified the biggest targets of selection.

The loci under selection showed traces of polymorphism in two ways (supplementary table 3). Either a single allele was present in the population but not fixed or several alleles were found simultaneously in the population. *gidB* showed the strongest sign of polymorphism. In mouse C11Ma, at day 155, the whole population seems to have mutated *gidB*, but 4 alleles are involved: a 25bp deletion, a frameshift, P79L and Q23* at respective frequencies of 8%, 60.2%, 8.7% and 31.3% (note that the frequencies are based on local coverage statistics so the sum is not supposed to be precisely 1). At day 363, the 25bp deletion overtook the other alleles and reached a frequency of 100%. In mice C2Ma, two alleles of *gidB*, Q167* and the 25bp deletion, are found polymorphic at both sampling time. *dgoR* also had traces of polymorphism, with C8Ma having two competing alleles (360bp at 75.3% and Q212* at 20.9%) at one time point (Day 155). Out of the 3 mice studied for polymorphism that had mutations in *dgoR* at the last sampling time point, only in one case fixation was reached, while in the others the frequency of the alleles was of 55.8% and 74.1%. The study of polymorphism revealed also new alleles in some of the locus of interest. A mutation in the promoter of *dgoR* reached a frequency of 7.4% in C2Ma at day 124 and another allele between *yjiH* and *kptA* reached a frequency of 24% in C8Ma at Day 155 before being out competed by another allele. Finally, in mouse C11Ma, while we did not find any mutation in *dgoR* in the clone sequenced at the last time point, a mutation was recovered at high frequency in that gene at an earlier time point, but disappeared later on.

The comparison of the alleles found polymorphic in the different populations studied revealed another gene with evidence of convergence. Clone sequencing revealed a frameshift in *uxuR* in mouse C11Ma. The polymorphism analysis revealed that this mutation was fixed in that population. In 2 other mice we found mutations affecting *uxuR*. These mutations were at low frequencies (Q116* at 16.7% in mouse C2Ma, A213V at 6.6% in mouse C3Ma) which explains why we have not sampled them with the clone approach. UxuR is a transcriptional factor that regulates a cluster of operons involved in transport and degradation of the sugar acids β -D-glucuronides, glucuronate, gluconate and L-galactonate.

We further focused on both *dgoRKADT* operon and *rluD* and *gidB* complex as the other important convergence targets, the intergenic region between *yijH* and *kptA* (hit 11 times but with only 3 alleles) or ECP_4610 (hit 5 times) affect proteins of unknown functions. To limit the use of animals, functional tests were made only *in vitro*.

Galactonate operon

Role of *dgoR* and *dgoRKADT* operon—Among the most notable mutated genes, the genes related to the galactonate operon, *dgoRKADT*, seemed particularly interesting. Seventeen mutations comprising 6 stop codons, 2 frameshifts, and 3 missenses were recovered in *dgoR*, 4 missenses in *dgoD* and 2 missenses in *dgoT* such that in 11 out of the 15 cages the operon was mutated (figure 3). The operon is composed of *dgoR* (coding for the repressor), *dgoK* (coding for a kinase), *dgoA* (coding for an aldolase), *dgoD* (coding for a dehydratase) and *dgoT* (coding for the D-galactonate transporter). This operon is responsible for the utilization of the D-galactonate sugar in a pathway that is somehow parallel to the Entner Doudouff pathway (EDp): contrary to this pathway the input is D-galactonate rather than gluconate but likewise to the EDp the outputs are pyruvate and D-glyceraldehyde 3P (Lamble *et al.* 2004). We confirmed that *dgoR* is the repressor of the pathway, which is expressed at low level in the absence of D-galactonate and allows the inhibition of the expression of the operon. We observed this regulator activity using reporter strains carrying the *gfp* gene under control of the *dgoRKADT* operon promoter in K-12 strain. In LB, which does not contain galactonate, no fluorescence was observed. The presence of D-galactonate, induced fluorescence that was revealed using flow cytometry. Moreover the strain *E. coli* K-12 *dgoR* (Baba *et al.* 2006) using the same construct presented a constitutive induction of fluorescence in both media (figure 4). These results confirmed the repressor activity of DgoR, which had not yet been observed. To test the effects of the *dgoR* mutations selected for in our experiments, one evolved lineage of 536 obtained at the end of the experiment from the mouse C8Ma and the ancestor 536 were transformed with the reporter construct plasmid *pdgoR*. It showed a constitutive expression of fluorescence similar to the one found in *E. coli* K-12 *dgoR*, confirming that inactivation of the repressor has been selected for (data not shown).

Interestingly, mutations in *dgoD* always occurred in lineages having mutation in *dgoR* and *dgoT* mutations always occurred in lineages having also *dgoR* and *dgoD* mutations ($p=0.009$). For instance lineages 5 and 13 both had the 3 genes modified, using 6 different mutations.

To test for a potential benefit linked to mutations in the *dgoRKADT* operon, we cultured the 15 evolved lineages in different conditions. Growth curves were assayed in different media (LB and minimum medium with glucose or gluconate as the sole carbon source) at different pH (5.7 and 7) and in minimum medium with D-galactonate at pH 5.7 because the sugar is not soluble at neutral pH (figure 5). We then compared the DTs of the seven evolved strains with mutations in *dgoR*, the two evolved strains with mutations in *dgoR* and *dgoD* and the two evolved strains with mutations in *dgoR*, *dgoD* and *dgoT* and the four evolved lineages unaffected in *dgoR* as well that of the ancestor strains. A significant difference in growth was found in minimum medium with D-galactonate between *dgoR* mutant strains and the others using a Welch test ($p=0.000114$). In more details, the data suggest that *dgoR* and *dgoR-dgoD* ($p=0.001$) mutants have a higher growth rate than strains non mutated in *dgo* operon, and that strains with the three mutations *dgoR*, *dgoD* and *dgoT* grow even better than the ones with only two ($p=0.0014$). Of note, truncating (non sense and frame shift) mutations were observed only in the regulator coding gene *dgoR*. This suggest that despite an induction of the operon of *dgoRKADT* operon by D-galactonate, inactivation of the repressor of the operon promotes growth in D-galactonate, and that growth is further improved when genes implicated in the pathway of degradation of the D-galactonate are mutated. No difference was observed in LB, minimum medium with glucose and minimum medium with gluconate between the wild type and the mutant strains at both pH 5.7 and 7 (data not shown).

Focus on *rluD* and *gidB*

rluD and *gidB*, are other genes that attracted our attention. Specifically, 6 mutations, including frame shifts were found in *rluD*, and 7 mutations, including 25bp deletions were recovered in *gidB*. *rluD* encodes for a pseudo uridin synthetase that catalyses the formation of pseudo uridine on the 23S rRNA, in other words *rluD* is involved in 23S rRNA maturation. Similarly *gidB* is implicated in the maturation of the 16S subunit of the ribosomes as it encodes for a S-adenoside methionine dependant methyltransferase. In other bacterial species, both genes have been shown to be involved in streptomycin resistance, as streptomycin antibiotic targets the ribosome (Okamoto *et al.* 2007; Sato & Iino 2010). It is therefore very likely that these mutations have been selected for the adaptation to the antibiotic selective pressure that was maintained during our experiment. Moreover the selected functions of both genes seem to be associated as no evolved lineage presented mutations in both genes. Given the frequency of these mutations (7 and 6 mutations), we could find a significant repulsion between the mutations (test of non random association using a correlation: $p<0.001$), suggesting that once one is found in one of the genes the acquisition of a mutation in the other is no more selected for.

We first wanted to test if these mutations increased resistance to streptomycin in our strain. First, we tested if *E. coli* K-12 mutants of those two genes, K-12 *rluD* and K-12 *gidB* had a higher resistance level by determining the minimum inhibition concentrations (MIC) of streptomycin. K-12 *rluD* and K-12 *gidB* had an MIC of 20 and 80 mg/L, compared to 16 mg/L for K-12, confirming that the genes could play a role in resistance. We then measured the MIC of streptomycin of strain 536 and the 15 lineages. The mutation K87R responsible for the resistance to streptomycin has been identified previously by others and is responsible

for high resistance to streptomycin (Pelchovich *et al.* 2013). Indeed, MIC was over 150 g/L (more than 30 times higher than the concentration used in the experiment in the water of mice) and fairly similar across evolved lineages and the ancestor. We decided not to go further with MIC as such an incredibly high level of resistance prevents any accurate and reproducible measurement.

To go further, we then inactivated both genes independently or in combination in strain 536 and performed growth curves in LB supplemented or not with streptomycin at 5 g/L. The underlying idea was that the mutations could compensate the cost of resistance to streptomycin that has been shown to be costly. Wild type strain 536 and 536 *gidB* grew faster than 536 *rluD* and 536 *rluD gidB* ($p = 0.001$) in LB rejecting the idea that the mutations were compensating a general cost of resistance. Yet, the mutant strains grew faster than the wild type strain 536 in LB with streptomycin (figure 6), revealing that these mutations could readily be selected for under the antibiotic selective pressure used. *rluD* mutants were dominant over *gidB* mutations in the sense that *rluD* was sufficient to promote the reducing of the DT; adding *gidB* to it did not bring any further improvement. Hence, once *rluD* is inactivated, we do not expect to find any further mutations in *gidB*, however, under the conditions we tested, if *gidB* is mutated first, *rluD* could still be selected for. Hence our experimental test in LB could only partially explain the observed full repulsion between the two mutated genes.

Discussion

Experimental evolution coupled to whole genome sequencing is a powerful tool to study bacterial adaptation (Long *et al.* 2015). Over the last decades, many *E. coli* populations have been adapted to different media using that approach. Genomics was then used to uncover the molecular determinants of these adaptations and to follow how reproducible adaptation is. Most studies have been done *in vitro* and have shown some similarities. First, the response to selection is proportional to the initial maladaptation: clones with initial low fitness tend to improve faster (Couce & Tenaillon 2015). Second, intermediate convergence is observed among replicate lineages. Moreover this convergence is involving different mutations that affect the same genes or functional units. Though integration at the gene level seems the most relevant to study convergence, mutations in different genes involved in the same function can also be identified as a source of convergence (Tenaillon *et al.* 2012). Third, some epistatic interactions may be found between mutations within genes or among genes (Tenaillon *et al.* 2012). Fourth, global regulators tend to be frequently recruited during the first steps of adaptations (Hindré *et al.* 2012). As most studies, especially long-term ones have been done *in vitro* or *in vivo* with laboratory strains, the relevance of these observations remained to be questioned. We therefore wanted to know what fraction of these results would still hold if we were to adapt a natural isolate of *E. coli* to a more natural environment, the mouse gut. We therefore performed a 363-day assay of colonization of the digestive tract of 30 streptomycin treated mice in 15 independent cages with the B2 strain *E. coli* 536, and sequenced one evolved clone from each cage. Assuming 18 generations per day for *E. coli* in streptomycin treated mice (Poulsen *et al.* 1995), this represents about 6,534 generations.

Similarities and differences with *in vitro* lessons of experimental evolution

We found some similarities and differences with the previous experiments done.

- We recovered an important signal of convergence between lineages at the mutation, gene/operon and function levels. Convergence at the genome level was highly significant; it was variable among lineages. For instance, 3 out of 10 affected genes of in lineage 9 were targeted in other lineages, while the 6 genes affected in lineage 6 were at least found in another lineage. On average, 56% of the mutations affected genes that were also found mutated in at least another lineage. For instance, lineages 5 and 13 shared mutations in *dgoR*, *dgoD*, *dgoT*, *rluD* and in between *yjiH-kptA*. They have therefore 5 common targets (hit by different mutations for 3 out of 5) out of 9 or 7 affected genes respectively. They have hence a more than 50% convergence at the gene level.

-As found *in vitro*, this convergence was based on different mutations in the affected genes, and could be integrated into higher functional levels: *rluD-gidB* modifications of the ribosomal RNA for instance.

-Also some interactions among mutations within genes or among genes could be found. For instance, though 11 mutations were found in *dgoR*, no lineage was found with a double mutation in *dgoR* ($p < 0.004$), which suggest that once the gene is inactivated, no further mutations in the gene are selected for. Among genes, mutations in *gidB* and *rluD* were mutually exclusive, while mutations in *dgoT* only occurred in lineages having a mutation in both *dgoR* and *dgoT*. These nonrandom associations of mutations suggest an important role of epistasis in shaping the adaptive landscape.

One important difference with previous work was that no global regulator was recovered among the mutations found. *In vitro*, for instance *rpoB* is a very frequent target of adaptation (Conrad *et al.* 2010; Tenaillon *et al.* 2012). Despite a extremely high level of conservation within the species, with just 4 non-synonymous mutations in *rpoB* among 128 *E. coli* isolated genomes for a 4 Kb gene (Lescat *et al.* 2014), the mutations is one of the most frequent target of adaptation in many systems (Long *et al.* 2015)). The lack of mutations in global regulators in our system, suggests that the conditions encountered may be less artificial than *in vitro* conditions. Yet, if no global regulators are hit, we may still question the targeted functions to the mouse gut to uncover how specific they are from our experimental conditions or from our strain?

Functional targets of adaptation

One of the dominant targets of adaptation is the inactivation of *dgoR* that leads to an enhanced ability to grow on D-galactonate. Other studies performed in chemostats with specific sugar as nutrients allowed the observation of mutation in genes enhancing the increase of utilization of these sugars (Ferenci 2007; Lee & Palsson 2010). However, the specific convergence in the *dgoR* gene has never been reported so far. Particularly, it was not recovered by Barroso-Batista *et al* (Barroso-Batista *et al.* 2014) who used also streptomycin treated mice to study the evolution of the K-12 strain. Many hypotheses could be proposed. These observations could be linked to the strain studied. We used the strain *E. coli* 536 which was isolated after a urinary tract infection in a human patient (Brzuszkiewicz *et al.*

2006), whereas in other studies the strain used for the experiment of colonization was *E. coli* K-12 which is the typical laboratory strain (Blattner *et al.* 1997). One possibility is that strain 536 could harbor an abnormal genetic background relative to other strains. For instance, Barroso-Batista *et al.* found that the first response of *E. coli* K-12 to the gut was mutations inactivating *gat* operon, but this operon is abnormal in *E. coli* K-12: it is constitutively expressed due to an insertion sequence integration in its repressor (Barroso-Batista *et al.* 2014). Hence the observed response could be more a specificity of the strain than a signature of the environment. When we looked at *dgoR*, in *E. coli* 536, we could not find any clear specificity. Our functional assays showed that *dgoR* does repress efficiently the operon in the presence of glucose and release the repression in the presence of D-galactonate. We performed further analysis of *dgoR* gene history using fully sequenced genome of *E. coli* isolates. We observed that the entire operon *dgoRKADT* was missing in the strains *E. coli* from the serotype O157:H7 belonging to the E group. This group is usually responsible for enterohemorrhagic diarrhea in patients but it is also retrieved as commensal in digestive tracts of cattle (Cray & Moon 1995). This particular observation could be linked to the specific digestive tract of cattle. We further used more than 110 *E. coli* genome sequences (Lescat *et al.* 2014) to retrieve the mean pairwise rate of synonymous (Ks) and nonsynonymous substitutions (Ka) of *dgoR*, *dgoD* and *dgoT*. For each of these genes, the ratio Ka/Ks was much less than 1 (0.0055, 0.013 and 0.019 respectively) which comforts the idea that the operon is under strong purifying selection.

We then tried to identify the source of D-galactonate in the digestive tract of animals or human individuals, but nothing has been described about this sugar except the big amount produced in specific tissues of patients or mice with a specific genetic disease (Yager *et al.* 2004; Ficicioglu *et al.* 2005). However, this sugar can be produced from galactose using the De Ley Doudoroff pathway present in specific aerobic bacteria such as *Stenotrophomonas maltophilia* (Brechtel *et al.* 2002). Saffarian *et al.* recently identified that these bacteria especially *S. maltophilia* can be retrieved in the crypts of mice (Saffarian *et al.* 2015). Besides, we know that *S. maltophilia* can be retrieved in the mucous membranes of patients under antibiotic selective pressure especially with imipenem treatment, this bacteria being naturally resistant to this antibiotic as well as to all aminoglycosides, including streptomycin. We therefore retrospectively searched for the presence of this species in mouse feces stored at -80°C in glycerol, but failed to find it. Surprisingly, in other experiments with CD1 mice used for colonization assay, we identified the presence of this species in the feces but were not able to find if the species could produce D-galactonate (data not shown). We can also hypothesize that *dgoRKADT* operon may be used to catabolize some alternative sugars found in the digestive tract of mice, and that this sugar is unable to relieve the repression by *dgoR* and favors therefore its inactivation. Overall, from the data we have, we can say that selection for *dgoR* inactivation seem not to be a specificity of the strain we used, and must be due to the environmental conditions we evolved the strains in. However, it is hard to know if the selection for *dgoR* loss is generally selected in the mouse gut or if it is selected here as an indirect consequence of the use of streptomycin that alters the composition of the microbiota as Leatham-Jensen *et al.* observed missenses in other genes of an *E. coli* K-12 strain evolved model of colonization (Leatham-Jensen *et al.* 2012). Nevertheless, these convergences illustrate the ability of *E. coli* to adapt to the

specificity of its gut environment, whether they are linked to the host, its diet or *E. coli* interactions with the gut microbiota (Maltby *et al.* 2013).

Our second target of adaptation was directly linked to the use of antibiotic in our experimental setting. Thirteen out of the 15 lineages had mutations either in *gidB* or in *rluD*, two genes involved in maturation of ribosomal RNAs and linked to the *rpsL* gene in which mutations may be responsible of resistance to streptomycin. Experiments in which bacteria are evolved in presence of antibiotics represents one of the typical cases of selection that favors the fast emergence of mutations. It is also a case of medical importance due to the drastic increase of resistance among pathogenic bacteria (Andersson & Hughes 2010; Toprak *et al.* 2011). Two types of mutations are generally selected for in such a regime: mutations that increase resistance to the antibiotic on the one hand and compensatory mutations on the other. The latter are mutations that are selected as they compensate the cost of some previously acquired costly resistance mutations (Andersson & Hughes 2010). In our case, as the concentration of antibiotic used was much lower than the MIC of the strain (5 g/L compared to an MIC of more than 150 g/L), we were therefore not expecting resistance mutations to be selected for. Indeed, in the case of streptomycin, subinhibitory concentrations higher than 1/4 of the MIC can select for resistance mutations (Gullberg *et al.* 2011), while in the present setting, strains were exposed to concentrations lower than 1/30 of the MIC. We found however that *gidB* and *rluD* inactivations were not compensatory mutations, but mutations that improved growth only in the presence of the antibiotic. Interestingly *gidB* was found to be inactivated in several streptomycin resistant laboratory strains (Jeong *et al.* 2009) suggesting that the selection of its inactivation may occur quite broadly under streptomycin selective pressure.

Finally, the pattern of convergence suggests that two other sets of genes of unknown functions are involved in adaptation to the streptomycin treated mice gut. ECP4610, a putative transcriptional regulator is affected 8 times and the intergenic region between *yjiH* and *kptA* is modified 11 times. It would be interesting to look in more details at the phenotypic effects of these mutations to uncover the specificity of their contribution to adaptation.

Rates of adaptation

Besides the targets of adaptation, our data gave also the rate at which mutations are selected for in the mice gut. In experimental evolution, fast rate of substitutions can be found, with about 11 mutations in 2,000 generations in a thermal adaptation experiment using a very maladapted strain (Tenaillon *et al.* 2012), 15 mutations after 5,000 generations in a less challenging environment (Barrick *et al.* 2009) *In vivo*, Barroso-Batista *et al.* found a mean of 2.3 mutations per lineage over 400 generations (Barroso-Batista *et al.* 2014). Here we found 95 mutations over 4,838 days of evolution, a rate 4.4, 2.4 and 4.7 times lower than the one found in the afore mentioned experiments if we assume 18 generations per day (Rang *et al.* 1999) and no mutation in the global regulators. Hence, the selective pressures may not be as strong as in other setting, reflecting that despite the antibiotic selective pressure, the strain was more adapted to its environment than the ones of the previous experiments. The polymorphism data were suggesting low selective pressure as in mice C3Ma and C2Ma

some polymorphic alleles shifted their frequency by less than 40% over more than 125 days of evolution (table S3) and over 4 populations sequenced after more than 125 days of evolution, a single mutation was found fixed (data not shown). The value of K_a/K_s was substantially higher than one ($K_a/K_s=1.90$, 95% CI 1.09-4.02) but less than the one found in other studies. For instance, a K_a/K_s close to 5 was found in thermal adaptation suggesting that 80% of non-synonymous mutations were the product of selection in that setting. Such high values of K_a/K_s are not rare, for instance in the long-term experimental evolution, the population sequenced (Barrick *et al.* 2009) had no synonymous mutations fixed after 20,000 generations. The lower value found here is coherent with a much weaker selective pressure in the mouse gut compared to *in vitro* evolution.

Overall, our data show that an “evolve and sequence” approach (Long et al 2015) coupled to an analysis of convergence can be used to uncover the molecular bases of short-term adaptation in an *in vivo* study. With that approach, we identified several genes as targets of selection in the gut of streptomycin treated mice and showed that adaptation could be caught in the act in a somehow natural setting with a natural isolate. This suggests that short term adaptation may contribute to some of the patterns of prevalence we observed in the wild (Tenaillon et al 2010). While some of the dominant targets, *gidB/rluD* and *dgo* operon do not seem to be specific of the strain we used, others such as ECP4610 and mutations between *yjiH* and *kptA* are more enigmatic and could be specific of the strain, especially ECP4610 that is only found in few strains. Further studies with various strains will be required to investigate the fraction of that adaptation that is dependent on the genetic background of the host. Finally, the presence of genetic responses linked to the antibiotic used to maintain the strain, call for the development of some colonization assays without such a selective pressure.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding information

This work was supported by the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013)/ERC Grant 310944. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. MG was supported by the grant FDM20150633803 obtained from the « Fondation pour la Recherche Médicale ».

Bibliographic references

- Andersson DI, Hughes D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nature Reviews Microbiology*. 2010; 8:260–271. [PubMed: 20208551]
- Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut microbiome. *Nature*. 2011; 473:174–180. [PubMed: 21508958]
- Baba T, Ara T, Hasegawa M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology*. 2006; 2 2006.0008.
- Barrick JE, Yu DS, Yoon SH, et al. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*. 2009; 461:1243–1247. [PubMed: 19838166]
- Barroso-Batista J, Sousa A, Lourenço M, et al. The First Steps of Adaptation of *Escherichia coli* to the Gut Are Dominated by Soft Sweeps. *PLoS Genetics*. 2014; 10(3):4182. 1371/journal.pgen.1004182.

- Berger H, Hacker J, Juarez A, Hughes C, Goebel W. Cloning of the chromosomal determinants encoding hemolysin production and mannose-resistant hemagglutination in *Escherichia coli*. *Journal of Bacteriology*. 1982; 152:1241–1247. [PubMed: 6754701]
- Bingen E, Picard B, Brahimi N, et al. Phylogenetic Analysis of *Escherichia coli* Strains Causing Neonatal Meningitis Suggests Horizontal Gene Transfer from a Predominant Pool of Highly Virulent B2 Group Strains. *Journal of Infectious Diseases*. 1998; 177:642–650. [PubMed: 9498443]
- Blattner FR, Plunkett G 3rd, Bloch CA, et al. The complete genome sequence of *Escherichia coli* K-12. *Science (New York, N.Y.)*. 1997; 277:1453–1462.
- Bleibtreu A, Gros P-A, Laouénan C, et al. Fitness, Stress Resistance, and Extraintestinal Virulence in *Escherichia coli*. *Infection and immunity*. 2013; 81:2733–2742. [PubMed: 23690401]
- Brechtl E, Huwig A, Giffhorn F. l-Glucitol Catabolism in *Stenotrophomonas maltophilia* Ac. *Applied and Environmental Microbiology*. 2002; 68:582–587. [PubMed: 11823194]
- Bruszkiewicz E, Brüggemann H, Liesegang H, et al. How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:12879–12884. [PubMed: 16912116]
- Conrad TM, Frazier M, Joyce AR, et al. RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:20500–20505. [PubMed: 21057108]
- Couce A, Tenaillon OA. The rule of declining adaptability in microbial evolution experiments. *Frontiers in Genetics*. 2015; 6(99)doi: 10.3389/fgene.2015.00099
- Cray WC, Moon HW. Experimental infection of calves and adult cattle with *Escherichia coli* O157:H7. *Applied and Environmental Microbiology*. 1995; 61:1586–1590. [PubMed: 7747972]
- Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97:6640–6645. [PubMed: 10829079]
- Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods in Molecular Biology (Clifton, N.J.)*. 2014; 1151:165–188.
- Diard M, Garry L, Selva M, et al. Pathogenicity-associated islands in extraintestinal pathogenic *Escherichia coli* are fitness elements involved in intestinal colonization. *Journal of Bacteriology*. 2010; 192:4885–4893. [PubMed: 20656906]
- Ferenci, T. Bacterial Physiology, Regulation and Mutational Adaptation in a Chemostat Environment. *Advances in Microbial Physiology*. Poole, RK., editor. Academic Press; 2007. p. 169-315.
- Ficicioglu C, Yager C, Segal S. Galactitol and galactonate in red blood cells of children with the Duarte/galactosemia genotype. *Molecular Genetics and Metabolism*. 2005; 84:152–159. [PubMed: 15670721]
- Flint A, Butcher J, Stintzi A. Stress Responses, Adaptation, and Virulence of Bacterial Pathogens During Host Gastrointestinal Colonization. *Microbiology Spectrum*. 2016; 4(2)doi: 10.1128/microbiolspec.VMBF-0007-2015
- Giraud A, Arous S, Paepé MD, et al. Dissecting the Genetic Components of Adaptation of *Escherichia coli* to the Mouse Gut. *PLoS Genet*. 2008; 4:e2. [PubMed: 18193944]
- Giraud A, Matic I, Tenaillon O, et al. Costs and Benefits of High Mutation Rates: Adaptive Evolution of Bacteria in the Mouse Gut. *Science*. 2001; 291:2606–2608. [PubMed: 11283373]
- Gordon DM, Cowling A. The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology (Reading, England)*. 2003; 149:3575–3586.
- Gullberg E, Cao S, Berg OG, et al. Selection of Resistant Bacteria at Very Low Antibiotic Concentrations. *PLoS Pathogens*. 2011; 7(7):e1002158. Epub 2011 Jul 21. doi: 10.1371/journal.ppat.1002158 [PubMed: 21811410]
- Herring CD, Raghunathan A, Honisch C, et al. Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nature Genetics*. 2006; 38:1406–1412. [PubMed: 17086184]

- Hershberg R, Petrov DA. Evidence that mutation is universally biased towards AT in bacteria. *PLoS genetics*. 2010; 6:e1001115. [PubMed: 20838599]
- Hindré T, Knibbe C, Beslon G, Schneider D. New insights into bacterial adaptation through *in vivo* and *in silico* experimental evolution. *Nature Reviews Microbiology*. 2012; 10:352–365. [PubMed: 22450379]
- Jaureguy F, Landraud L, Passet V, et al. Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics*. 2008; 9:560. [PubMed: 19036134]
- Jeong H, Barbe V, Lee CH, et al. Genome sequences of *Escherichia coli* strains REL606 and BL21(DE3). *Journal of Molecular Biology*. 2009; 394:644–652. [PubMed: 19786035]
- Khoury KA, Floch MH, Hersh T. Small intestinal mucosal cell proliferation and bacterial flora in the conventionalization of the germfree mouse. *The Journal of Experimental Medicine*. 1969; 130:659–670. [PubMed: 4896909]
- Lamble HJ, Milburn CC, Taylor GL, Hough DW, Danson MJ. Gluconate dehydratase from the promiscuous Entner–Doudoroff pathway in *Sulfolobus solfataricus*. *FEBS Letters*. 2004; 576:133–136. [PubMed: 15474024]
- Leatham-Jensen MP, Frimodt-Møller J, Adediran J, et al. The Streptomycin-Treated Mouse Intestine Selects *Escherichia coli* *envZ* Missense Mutants That Interact with Dense and Diverse Intestinal Microbiota. *Infection and Immunity*. 2012; 80:1716–1727. [PubMed: 22392928]
- Lee D-H, Palsson BØ. Adaptive Evolution of *Escherichia coli* K-12 MG1655 during Growth on a Nonnative Carbon Source, 1-1,2-Propanediol. *Applied and Environmental Microbiology*. 2010; 76:4158–4168. [PubMed: 20435762]
- Le Gac M, Cooper TF, Cruveiller S, Médigue C, Schneider D. Evolutionary history and genetic parallelism affect correlated responses to evolution. *Molecular Ecology*. 2013; 22:3292–3303. [PubMed: 24624420]
- Lescat M, Clermont O, Woerther PL, et al. Commensal *Escherichia coli* strains in Guiana reveal a high genetic diversity with host-dependant population structure. *Environmental microbiology reports*. 2013; 5:49–57. [PubMed: 23757130]
- Lescat M, Reibel F, Pintard C, et al. The Conserved *nhaAR* Operon Is Drastically Divergent between B2 and Non-B2 *Escherichia coli* and Is Involved in Extra-Intestinal Virulence. *PLoS ONE*. 2014; 9(9):e108738. eCollection 2014. doi: 10.1371/journal.pone.0108738 [PubMed: 25268639]
- Leveret M, Zamfir O, Clermont O, et al. Molecular and Evolutionary Bases of Within-Patient Genotypic and Phenotypic Diversity in *Escherichia coli* Extraintestinal Infections. *PLoS Pathogens*. 2010; 6(9):e1001125. doi: 10.1371/journal.ppat.1001125 [PubMed: 20941353]
- Long A, Liti G, Luptak A, Tenaillon O. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nature Reviews Genetics*. 2015; 16:567–582.
- Maddamsetti R, Lenski RE, Barrick JE. Adaptation, Clonal Interference, and Frequency-Dependent Interactions in a Long-Term Evolution Experiment with *Escherichia coli*. *Genetics*. 2015; 200:619–631. [PubMed: 25911659]
- Maltby R, Leatham-Jensen MP, Gibson T, Cohen PS, Conway T. 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*. 2013; 8:e53957. [PubMed: 23349773]
- Okamoto S, Tamaru A, Nakajima C, et al. Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. *Molecular Microbiology*. 2007; 63:1096–1106. [PubMed: 17238915]
- Pelchovich G, Schreiber R, Zhuravlev A, Gophna U. The contribution of common *rpsL* mutations in *Escherichia coli* to sensitivity to ribosome targeting antibiotics. *International journal of medical microbiology: IJMM*. 2013; 303:558–562. [PubMed: 23972615]
- Picard B, Garcia JS, Gouriou S, et al. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infection and Immunity*. 1999; 67:546–553. [PubMed: 9916057]
- 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. *Journal of Bacteriology*. 1995; 177:5840–5845. [PubMed: 7592332]

- Rang CU, Licht TR, Midtvedt T, et al. Estimation of Growth Rates of *Escherichia coli* BJ4 in Streptomycin-Treated and Previously Germfree Mice by In Situ rRNA Hybridization. *Clinical and Diagnostic Laboratory Immunology*. 1999; 6:434–436. [PubMed: 10225851]
- Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. *Microbes and infection / Institut Pasteur*. 2003; 5:449–456.
- Saffarian A, Mulet C, Naito T, et al. Draft Genome Sequences of *Acinetobacter parvus* CM11, *Acinetobacter radioresistens* CM38, and *Stenotrophomonas maltophilia* BR12, Isolated from Murine Proximal Colonic Tissue. *Genome Announcements*. 2015; 3(5) pii: e0108915. doi: 10.1128/genomeA.01089-15
- Sato T, Iino T. Genetic analyses of the antibiotic resistance of *Bifidobacterium bifidum* strain Yakult YIT 4007. *International Journal of Food Microbiology*. 2010; 137:254–258. [PubMed: 20051305]
- Tenaillon O, Barrick JE, Ribick N, et al. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature*. 2016; 536(7615):165–70. [PubMed: 27479321]
- Tenaillon O, Rodríguez-Verdugo A, Gaut RL, et al. The molecular diversity of adaptive convergence. *Science (New York, N.Y.)*. 2012; 335:457–461.
- Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. *Nature Reviews Microbiology*. 2010; 8:207–217. [PubMed: 20157339]
- Toprak E, Veres A, Michel J-B, et al. Evolutionary paths to antibiotic resistance under dynamically sustained drug stress. *Nature genetics*. 2011; 44:101–105. [PubMed: 22179135]
- Touchon M, Hoede C, Tenaillon O, et al. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genetics*. 2009; 5:e1000344. [PubMed: 19165319]
- Umekasi Y. Use of gnotobiotic mice to identify and characterize key microbes responsible for the development of the intestinal immune system. *Proceedings of the Japan Academy Series B, Physical and Biological Sciences*. 2014; 90:313–332.
- Welling GW, Groen G, Tuinte JH, Koopman JP, Kennis HM. Biochemical effects on germ-free mice of association with several strains of anaerobic bacteria. *Journal of General Microbiology*. 1980; 117:57–63. [PubMed: 7391821]
- Wielgoss S, Barrick JE, Tenaillon O, et al. Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution Experiment With *Escherichia coli*. G3 (Bethesda, Md.). 2011; 1:183–186.
- Wong A, Rodrigue N, Kassen R. Genomics of adaptation during experimental evolution of the opportunistic pathogen *Pseudomonas aeruginosa*. *PLoS genetics*. 2012; 8:e1002928. [PubMed: 23028345]
- Yager C, Ning C, Reynolds R, Leslie N, Segal S. Galactitol and galactonate accumulation in heart and skeletal muscle of mice with deficiency of galactose-1-phosphate uridylyltransferase. *Molecular Genetics and Metabolism*. 2004; 81:105–111. [PubMed: 14741191]
- Zaslaver A, Bren A, Ronen M, et al. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nature methods*. 2006; 3:623–628. [PubMed: 16862137]

Figure 1. Population density of *Escherichia coli* 536 per gram of feces of the streptomycin treated mice in 15 independent cages after oral colonization of the strain during 363 days.

The population density is represented in logarithmic scale. All density populations from the same mice at the different points in the time are relied by a grey curve. The percentage of alive mice is also represented with the black curve.

Figure 2. Mutation pattern among the evolved lineages.

The dendrogram of the evolved lineages (on the left) was computed using parsimony method and the presence or absence of mutations among the lineages. The convergence at the mutation level is showed in red (light red absence of mutation, bright red: presence of mutation) and illustrates similar mutations that appeared in different lineages with the name of the gene or regions it falls into, and the position on the *Escherichia coli* 536 genome. The specificity at the gene level is shown in blue (light blue: absence of mutations, bright blue: presence of mutation) for the *dgoR*, *dgoT*, and *dgoD* genes, and in green (light green: absence of mutations, bright green: presence of mutation) for the *gidB* and *rluD* genes illustrating mutations falling in the same genes, but with possible differences.

Figure 3. Diagrammatic representation of the *dgoRKADT* operon of the 15 evolved lineages compared to the *Escherichia coli* 536 ancestor strain.

The mutations are shown at scale all along the operon. Each evolved lineage mutation is represented with a symbol defined in the legend.

Figure 4. Comparison of total fluorescence of a reporter plasmid of the promoter activity of the *dgoRKADT* operon in *Escherichia coli* K-12 derived strains.

Boxplots of the total fluorescence of the strain K-12 pdgoR and the mutated strain K-12 *dgoR* pdgoR have been compared in LB and in minimum medium D-galactonate.

Figure 5. Mutated strains of *Escherichia coli* 536 in *dgoR* grew faster than wild type strains in minimum medium D-galactonate.

Boxplots of the doubling times in minutes of 11 mutated strains of *E. coli* 536 in *dgoR* and 5 non mutated strains in *dgoR* with the ancestor, in minimum medium D-galactonate at pH 5.7. We found a significant difference between the *dgoR* mutated strains and the wild type strains using a Welch test ($p = 0.001$).

Figure 6. Inactivation of *rluD* and *gidB* genes confers an advantage to the *Escherichia coli* strains in LB with streptomycin.

Boxplots of the doubling times in minutes of 536, 536 *gidB*, 536 *rluD* and 536 *rluD gidB* in LB, and in LB with streptomycin at 5 g/L. We observed in one side that wild type strain 536 and 536 *gidB* grew faster than 536 *rluD* and 536 *rluD gidB* in LB and that the mutant strains grew faster than the wild type strains 536 in LB with streptomycin ($p < 0.001$).

Table 1

Summary of the mutations that were found in the evolved lineages.

The name of the genes as well as the orientation (defined by the arrow), the position along the genome, and the flanking genes in case of an intergenic mutation are furthermore displayed, with the mutations, its type, the number of occurrence, and a description of the gene product. Stop codons are indicated with the symbol *.

Gene	Position	Mutation	Mutation type	Number of occurrence of the mutation	Mouse	Gene info
<i>rluD</i> ←	2742937	IS100	IS insertion	1	C8Ma	23S rRNA pseudouridine synthase
	2743335	151 bp	deletion	1	C10Ma	
	2743427	7 bp	deletion	1	C1Ma	
	2743564	6 bp	deletion	3	C5Ma ; C9Mb ; C13Mb	
<i>dgoT</i> ←	4055032	C→T	A380T (GCA→ACA)	1	C13Mb	D-galactonate transporter
	4055282	C→T	M296I (ATG→ATA)	1	C5Ma	
<i>dgoD</i> ←	4056705	G→A	P230S (CCA→TCA)	1	C5Ma	Galactonate dehydratase
	4056714	T→C	T227A (ACG→GCG)	1	C6Ma	
	4056917	G→A	A159V (GCG→GTG)	1	C13Mb	
	4057008	T→C	I129V (ATC→GTC)	1	C1Ma	
<i>dgoR</i> ←	4058921	G→A	Q212* (CAG→TAG)	5	C2Ma; C5Ma; C12Mb; C14Ma; C15Ma	DNA binding transcriptional regulator
	4059047	360 bp	deletion	1	C8Ma	
	4059269	8 bp x 2	duplication	1	C9Mb	
	4059403	G→A	S51L (TCG→TTG)	1	C1Ma	
	4059431	G→A	R42C (CGC→TGC)	1	C3Ma	
	4059436	G→A	T40I (ACC→ATC)	1	C13Mb	
	4059478	G→T	S26* (TCG→TAG)	1	C6Ma	
	4106867	G→A	Q167* (CAA→TAA)	1	C2Ma	
<i>gidB</i> ←	4106902	25 bp	deletion	4	C3Ma; C6Ma; C7Mb; C11Ma	S-adenosine methionine dependent methyltransferase-
	4107005	G→A	Q121* (CAG→TAG)	1	C15Ma	
	4107347	80 bp	deletion	1	C14Ma	

Gene	Position	Mutation	Mutation type	Number of occurrence of the mutation	Mouse	Gene info
ECP_4610 ←	4816907	A→T	L185Q (CTG→CAG)	1	C3Ma	Putative transcriptional regulator
	4816998	10 bp→10 bp	substitution	1	C8Ma	
	4817034	C→T	G143S (GGC→AGC)	1	C5Ma	
	4817174	G→A	T96I (ACT→ATT)	1	C14Ma	
	4817222	G→A	A80V (GCC→GTC)	1	C1Ma	
<i>yjH</i> ← / → <i>kptA</i>	4863869	C→T	intergenic (-68/-182)	8	C1Ma; C2Ma; C3Ma; C5Ma; C7Mb; C9Mb; C10Ma; C13Mb	Putative inner membrane protein/2-phosphotransferase
	4863869	C→G	intergenic (-68/-182)	1	C6Ma	
	4863894	A→G	intergenic (-93/-157)	2	C8Ma; C15Ma	
Group of 77 genes	294397	68069 bp	deletion	3	C6Ma; C11Ma; C14Ma	Prophagic region