High Recombinant Frequency in Extraintestinal Pathogenic *Escherichia coli* Strains

Jerónimo Rodríguez-Beltrán,[1,2] Jérôme Tourret,[3,4] Olivier Tenaillon,[3] Elena López,[5] Emmanuelle Bourdelier,[3] Coloma Costas,[1,2] Ivan Matic,[6] Erick Denamur,[3] and Jesús Blázquez*[,1,2]

[1]Instituto de Biomedicina de Sevilla (IBIS)-CSIC, Sevilla, Spain

[2]Centro Nacional de Biotecnología-CSIC, Madrid, Spain

[3]INSERM, IAME, UMR 1137, Paris, France; Université Paris Diderot, Sorbonne Paris Cité, Paris, France

[4]Assistance Publique—Hôpitaux de Paris, Groupe Hospitalier Pitié-Salpêtrière Charles Foix; Université Pierre et Marie Curie, Paris, France

[5]Servicio de Microbiología, Hospital Universitario La Paz, IdiPaz, Madrid, Spain

[6]INSERM U1001, Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine Paris Descartes, Paris, France

***Corresponding author:** E-mail: jblazquez-ibis@us.es.

Associate editor: Csaba Pal

Abstract

Homologous recombination promotes genetic diversity by facilitating the integration of foreign DNA and intrachromosomal gene shuffling. It has been hypothesized that if recombination is variable among strains, selection should favor higher recombination rates among pathogens, as they face additional selection pressures from host defenses. To test this hypothesis we have developed a plasmid-based method for estimating the rate of recombination independently of other factors such as DNA transfer, selective processes, and mutational interference. Our results with 160 human commensal and extraintestinal pathogenic *Escherichia coli* (ExPEC) isolates show that the recombinant frequencies are extremely diverse (ranging 9 orders of magnitude) and plastic (they are profoundly affected by growth in urine, a condition commonly encountered by ExPEC). We find that the frequency of recombination is biased by strain lifestyle, as ExPEC isolates display strikingly higher recombination rates than their commensal counterparts. Furthermore, the presence of virulence factors is positively associated with higher recombination frequencies. These results suggest selection for high homologous recombination capacity, which may result in a higher evolvability for pathogens compared with commensals.

Key words: **homologous recombination, evolution, ExPEC, commensal, RecA.**

Article

Introduction

Two types of genetic mechanisms allow bacterial genomes to diversify and therefore adapt to new or changing environments. On the one hand, inherent errors in the replication processes through point mutations, insertion sequence mobility, or genomic rearrangements lead to the transmission of modified genomes from one generation to the next. On the other hand, import of exogenous DNA, a process termed horizontal gene transfer (HGT), may lead to the immediate acquisition of either new functions or new alleles of existing genes through homologous recombination (HR). When the DNA sequences acquired by HGT are unable to replicate autonomously, HR can contribute both to its integration in the bacterial chromosome and to its propagation in the species using the conserved regions flanking the integration site (Schubert et al. 2009).

Therefore, recombination, which is a major DNA repair pathway, is also a key factor in bacterial genomic evolution. Through HGT, HR allows bacteria to expand their energy sources, evade the immune response of their hosts, acquire antibiotic resistance, and increase virulence (Guttman and Dykhuizen 1994; Lawrence and Roth 1996; Lawrence and Ochman 1998). Moreover, in *Escherichia coli* species it has been estimated that recombination is much more frequent than mutation (Touchon et al. 2009). For instance, a nucleotide has a 50–100 times higher probability of being involved in genetic recombination than being mutated (Guttman and Dykhuizen 1994). Computer simulations and experimental results suggest that high recombination rates may combine beneficial mutations appearing in different backgrounds faster than the incremental accumulation of them (Tenaillon et al. 2000; Cooper 2007). Recombination, therefore, can accelerate the rate at which a bacterial population adapts to environmental conditions (Levin and Cornejo 2009).

Obviously, bacteria that share ecological niches have more opportunities for genetic exchange than those living in distinct environments. The acquisition and use of premade and pretested new traits, such as antibiotic-inactivating enzymes or virulence mechanisms, provide a broad spectrum of possibilities to bacteria, which may use new functions previously developed by others. In environments with a high possibility for genetic exchange (where many related and/or genetically compatible species are present in the niche), the genetic exchange, through plasmids, transposons, phages

and/or HR, may be favored over mutation (Tenaillon et al. 2000). On the contrary, genetically isolated bacterial populations (i.e., with low possibility of exchanging DNA) have to find their own solutions through their own resources (i.e., mutation and/or gene shuffling by HR) and, consequently, will benefit from increased mutation and intragenomic recombination rates. A paradigmatic example is *Mycobacterium tuberculosis*, in which, for instance, acquisition of antibiotic resistance occurs exclusively by mutation (Ramaswamy and Musser 1998). Thus, the rates and effects of mutation and recombination on evolution are drastically conditioned by ecology (Didelot and Maiden 2010).

Both mutation and recombination processes are genetically controlled. As such, they are themselves under the action of natural selection. Both processes can therefore evolve within species through a process called second order selection (Tenaillon et al. 1999) in which variants are selected for based on their impact on genomic diversity. For instance, a high mutation rate clone, or mutator, can be selected for (Sniegowski et al. 1997; Oliver et al. 2000) or counterselected (Trobner and Piechocki 1984; Wielgoss et al. 2013) based on the selective impact of the mutation it generates (Tenaillon et al. 1999). Mutators are favored by selection when the advantage of beneficial mutations is greater than the cost of being a mutator due to the overproduction of lethal and deleterious mutations (Mao et al. 1997; Taddei et al. 1997). Mutator alleles have been extensively studied in *E. coli* where a large diversity of mutation rates can be found among natural isolates (LeClerc et al. 1996; Matic et al. 1997; Denamur et al. 2002; Denamur and Matic 2006). In most cases, bacterial mutator strains had a mismatch repair deficient genotype (LeClerc et al. 1996; Matic et al. 1997; Oliver et al. 2000, 2002). However, much less is known about the diversity of recombination rates within a bacterial species. Yet, for a species like *E. coli* living in the promiscuous gut of vertebrates (containing large numbers of cells, species and niches that facilitate genetic exchange) (Stecher et al. 2013) and having a highly plastic genome, it can be expected that recombination should be a major player of adaptation.

Escherichia coli is a versatile and ubiquitous bacterial species that colonizes the gut of vertebrates and is also a leading human pathogen, causing chronic and acute infections, the management of which can be severely complicated by antibiotic resistance. This apparent dichotomy in bacterial lifestyle has been linked to the acquisition of particular gene clusters, called pathogenicity islands (PAIs), which in great part confer *E. coli* its pathogenic capacity (Groisman and Ochman 1996; Hacker and Kaper 2000). However, according to the coincidental-evolution hypothesis (Le Gall et al. 2007; Diard et al. 2010), virulence determinants, at least for extraintestinal pathogenic *E. coli* (ExPEC), can be intestinal colonization and survival factors, that is, acquisition of virulence could be a byproduct of commensalism. Independently of the forces selecting for virulence, *E. coli* is a perfect candidate for studying the transition between commensalism and pathogenicity.

The extant of recombinational diversity and its potential link with virulence have been studied before with genomic approaches. Data from multilocus sequence typing (MLST) indicate that rates of evolution were accelerated in *E. coli* pathogenic strains by increased HR events, suggesting that recombination and virulence are causally related (Wirth et al. 2006). Although genomic techniques, such as MLST and whole-genome sequencing, are successfully being used to understand how mutation and recombination impact bacterial evolution (Tenaillon et al. 2000; Spratt et al. 2001; Wirth et al. 2006; Touchon et al. 2009; Didelot et al. 2012), some important caveats have not been addressed. First, genetic events cannot be disentangled from the selection process and the DNA transfer process itself; consequently, only the signature of their combined action can be analyzed (Denamur et al. 2010). Second, MLST relies on the sequences of a relatively small number of house-keeping genes (usually seven) and therefore it is prone to introduce misleading results because it overestimates the weight of single gene fragments, which provide little information on the overall genomic composition (Falush et al. 2006). Furthermore, it has been pointed out that it lacks the resolution necessary for ascertaining the underlying diversity in bacterial populations determined by other approaches (Noller et al. 2003). Third, it is plausible that during a period of stress, enhanced recombination or mutation rates could be transitorily selected (Bjedov et al. 2003). These episodic increased rates of adaptability are difficult to distinguish with these techniques, as they are based on final nucleotide differences and hence lack temporal resolution (Paul et al. 2013).

We therefore decided to analyze the intrinsic capacity of recombining homologous DNA sequences, independently of the DNA transfer, selective processes and mutational interferences, in a collection of well-characterized human *E. coli* natural isolates from different origins, including both commensals and extraintestinal pathogens (Picard et al. 1999; Duriez et al. 2001; Hommais et al. 2005). Correlations between recombinant and mutant frequencies, isolate origin, phylogenetic group, number of virulence genes, and pathogenic capacity have been studied.

Results

Spontaneous Mutagenesis

Spontaneous mutagenesis of the strains was estimated by monitoring their capacity to generate mutations conferring resistance to rifampicin (RifR). Figure 1*A* shows the distribution of RifR mutant frequencies of both commensals and pathogens (raw data are shown in supplementary table S1, Supplementary Material online). As previously described (LeClerc et al. 1996; Matic et al. 1997; Denamur et al. 2002), mutant frequencies of most strains fall into a narrow peak (around 10^{-8}), except for a small group of strains, defined as mutators. Median values of both populations were almost identical (9.3×10^{-9} and 1.5×10^{-8}, respectively) (fig. 1*B*). The proportion of mutators is about 2.3-fold higher in the group of pathogen strains (10% vs. 4.4%), although this difference was not significant at the 95% level (Fisher's Exact Test $P = 0.21$). In our study, strains were considered mutators when they exhibited mutant frequencies of at

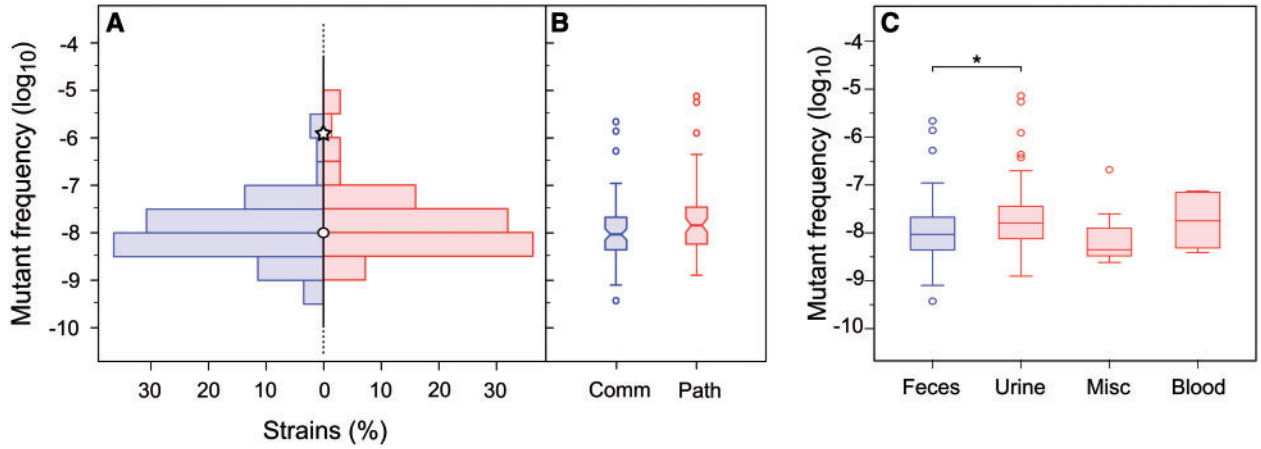

FIG. 1. RifR mutant frequencies. (*A*) Histogram showing the distributions of mutant frequencies of 90 commensals (left) and 70 pathogens (right). Frequencies of the *Escherichia coli* K-12 strains ME12 (circle) and ME12 Δ*mutS* (star) are shown as a reference. (*B*) Boxplot representing median values (horizontal line in the box) of mutant frequencies of commensals and pathogens. (*C*) Distributions of mutant frequencies according to the origin of the isolates. The differences between mutant frequencies of fecal and urinary strains are statistically significant (pairwise Wilcoxon test $P = 0.005$). For all boxplots, the depth of the box represents the interquartile range (50% of the data) and the whiskers extend to 1.5 times the interquartile range. Notches around medians represent 95% confidence intervals.

least 10-fold higher than the median value of all studied strains ($n = 160$; median = 1.06×10^{-8}).

As previously reported (Denamur et al. 2002), differences in the distribution of mutant frequencies can be observed between commensal ($n = 90$) and uropathogenic ($n = 48$) strains (pairwise Wilcoxon test $P = 0.005$) (fig. 1*C*).

Homologous Recombination

To study HR, we developed a genetic assay (a detailed description of the process can be found in the Supplementary Material online), which scores the recombination between two truncated *tetA* alleles separated by an antibiotic resistance cassette. Recombination restores the functional *tetA* gene, thereby conferring tetracycline resistance, which can be selected for. Therefore, this assay allows quantification of the frequency of recombinants.

We tested that our system relied on HR by introducing the plasmid pRhomo, conferring resistance to ampicillin and kanamycin but not to tetracycline (AmpRKanRTetS) (supplementary fig. S1, Supplementary Material online), carrying the recombination testing system, into the *E. coli* K-12 MG1655 derivatives ME12 and ME12 *recA* (Elez et al. 2007). When a recombinant event occurs between the two *tetA* alleles, plasmid pRhomo-Tet (TetR) is generated (supplementary fig. S1, Supplementary Material online). The frequency of AmpRTetR recombinants was in the order of 10^{-3} for the wild-type strain ME12, a value similar to that obtained with other tests (Elez et al. 2007; Lopez et al. 2007). As expected, HR was about 100-fold lower in the strain lacking *recA* (see values for K-12 in supplementary fig. S2, Supplementary Material online), indicating that reconstruction of the functional *tetA* is mostly dependent on RecA, and that it therefore relies on HR. The *recA* requirement was also studied in 13 randomly chosen strains from our panel (five commensals and eight pathogens). These strains, made *rec*A-deficient (*recA*::*kan*) as explained in Supplementary Material online, showed a decreased rate of HR with respect to their *recA*-proficient counterparts in all cases (supplementary fig. S2 and table S1, Supplementary Material online). For all tested strains, the RecA-dependent recombination is higher than 95% (with many strains in which the dependence is over 99.9%). Exceptions are strains P53 and P70 (with a RecA-dependence of 72% and 83%, respectively). Thus, in our conditions, the RecA-independent recombination seems to be rare in most cases. DNA-polymerase slippage and/or RecA-independent DNA crossing over events (Lovett et al. 1993) may account for the observed RecA-independent recombination. Furthermore, a strong dependence on the RecA function has been described previously in assays where DNA is acquired from other cells through F′ transfer (see e.g., Matic et al. 1995). Therefore, the use of plasmid pRhomo to study the frequency of *recA*-dependent HR was validated.

In addition, to ascertain that plasmidic recombination measured with pRhomo is comparable to chromosomal recombination, we introduced a single copy of a DNA fragment containing two pieces of the *tetA* gene separated by a kanamycin resistance gene (*tetA1-kan-tetA2*) into the chromosome of MG1655 K-12 and its *recA*-isogenic mutant, generating the strains MG1655 *att*Tn*7*::*tetA1-kan-tetA2* and MG1655 *att*Tn*7*::*tetA1-kan-tetA2 recA* (for a detailed description, see supplementary material and fig. S3, Supplementary Material online). The chromosomal recombination events in the *recA*-deficient strain were almost ten times less frequent (3.2×10^{-6}) than in wild-type MG1655 (1.63×10^{-5}) (supplementary fig. S4, Supplementary Material online). This confirms that, whether on a plasmid or inserted as a single copy in the chromosome, the *tetA* fragments mostly undergo *recA*-dependent recombination. Note that when *tetA* sequences are in the chromosome, the frequency of TetR recombinants is lower than that of the plasmidic construction (10^{-5} vs. 10^{-3}), indicating that, as expected, the availability of substrates for recombination influences the frequency of recombinants. To further explore this in our strains, we measured the recombinant frequencies of five

strains of our collection whose genomes have been sequenced (*E. coli* B, F11, HS, 536, and CFT073) (supplementary table S1, Supplementary Material online) and the *E. coli* MG1655 derivative ME12 strain harboring either the plasmidic or the chromosomal system. A strong linear correlation (Pearson's $r = 0.82$, $P = 0.04$) between recombinant frequencies obtained with both systems was observed (supplementary fig. S5, Supplementary Material online), demonstrating that pRhomo is an accurate tool to measure recombinant frequency. It should be noted that our assay with pRhomo cannot distinguish between intra- and interplasmidic recombination. However, our data indicate that if both possibilities occur, the recombinant frequencies still correlate with those from the chromosomal assay and are mainly mediated by RecA.

Recombinant frequencies of three independent colonies of each of the 160 strains (90 commensal and 70 pathogens), containing plasmid pRhomo, were calculated as the median number of recombinants (AmpRTetR) divided by the median number of viable cells on Lysogeny Broth (LB)-agar plates containing only ampicillin. Figure 2 shows the histogram (*A*) and the boxplot (*B*) with the distribution of recombinant frequencies of commensal and pathogen isolates. Interestingly, a high variability of recombinant frequencies, ranging from 10^{-1} to 10^{-9}, is observed for both groups. Our results clearly demonstrate that altogether pathogens have higher recombinant frequencies than commensals (Mann–Whitney *U* test $P < 10^{-8}$), with medians of 6.46×10^{-4} and 9.94×10^{-6}, respectively. An analysis of frequencies, classifying the strains by the isolate origin, showed significant differences (pairwise Wilcoxon test $P < 0.003$, in all cases) when comparing commensal strains with pathogenic strains from urine and also with the miscellaneous group, which includes isolates from diverse extraintestinal infections (mainly from pus) but excluding urine and blood (fig. 2*C*). Interestingly, recombination of pathogens isolated from blood showed no statistical differences with that of commensal strains (fig. 2*C*).

Estimation of recombinant frequencies by using pRhomo could depend on the different copy number of the plasmid in each strain (in principle, the higher plasmid copy number the higher probability of producing a recombination event). To rule out this possibility, the number of copies of pRhomo was determined for a group of strains randomly chosen ($n = 28$, 14 commensals and 14 ExPEC), with different frequencies of recombinants. The K-12 ME12 strain, which showed approximately 500 copies of pRhomo per cell (see supplementary methods, Supplementary Material online), was used as a reference. No correlation between recombinant frequency and plasmid copy number was observed ($n = 29$, Spearman's rho = 0.13, $P = 0.45$) (supplementary fig. S6, Supplementary Material online), indicating that, although the availability of substrates for recombination can influence the frequency of recombinants (see above), the observed differences in recombinant frequencies cannot be explained by the differences in the number of plasmid molecules present in each strain.

Escherichia coli constantly alternates between various habitats. In particular, ExPEC are characterized by the presence of virulence factors, which allow surviving and growing in host niches with specialized defenses, where they are responsible for severe infections. The most common site of infection for ExPEC is the urinary tract. For this reason and because the highest difference in recombinant frequency was observed between commensals and urinary tract infection strains (fig. 2*C*), we wanted to compare frequencies measured in LB and human urine. Figure 3 shows the recombinant frequencies of 22 strains (13 ExpEC and 9 commensal)

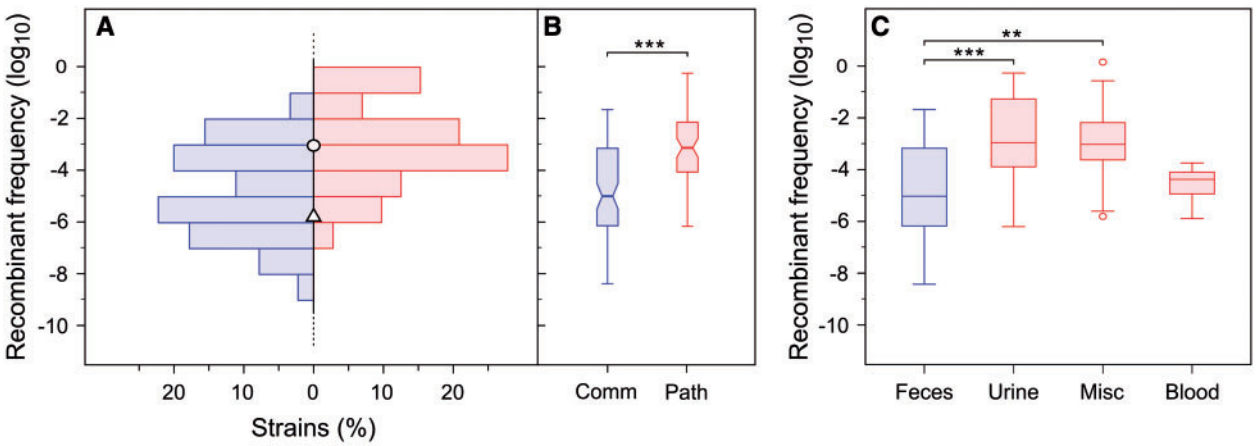

FIG. 2. TetR recombinant frequencies. (*A*) Histogram with the distributions of TetR recombinant frequencies of 90 commensal (left) and 70 pathogen (right) isolates. Frequencies of the *Escherichia coli* K-12 strains ME12 (circle) and ME12 Δ*recA* (triangle) are shown as a reference. (*B*) Boxplots of the distributions of recombinant frequencies of commensals and pathogens. Stars indicate a highly significant difference between distributions according to the Mann–Whitney *U* test (*P* value $< 10^{-8}$). (*C*) Boxplots representing the median values for strains grouped by origin of the isolates. Significant differences were found between strains isolated from the feces of healthy individuals and those isolated from urine and miscellaneous infections (Wilcoxon pairwise test; $^{**}P < 0.003$, $^{***}P < 10^{-5}$). For all boxplots, the horizontal line represents the median value, the depth of the box represents the interquartile range (50% of the data), and the whiskers extend to 1.5 times the interquartile range. Notches around medians represent 95% confidence intervals.

harboring pRhomo measured in LB and urine. Our results indicate that growth in urine dramatically modifies recombinant frequency. Although most strains (18 of 22; 14 with differences statistically significant) increased their recombinant frequencies (from 4- to 10,000-fold) upon growing in urine, the increase is not a general phenomenon (three strains showed decreased frequency and one strain showed no difference between LB and urine). This, however, confirms that environment may condition the rate of recombination, adding a level of complexity to the generation of variability. Increased recombination in urine was also observed in the strains MG1655 *att*Tn*7*::*tetA1-kan-tetA2* and MG1655 *att*Tn*7*::*tetA1-kan-tetA2 recA* (supplementary fig. S4, Supplementary Material online). The difference in the recombinant frequency in urine between these two strains indicates that recombination is mainly *recA*-dependent in this environment. To explore the generality of this *recA*-dependence in urine, the recombinant frequencies of the 13 *recA*-deficient strains (five commensals and eight pathogens) from our panel were studied in this medium. In all cases except one (12 of 13), the *recA*-deficient strains showed lower recombination frequencies in urine than their *recA*-proficient counterparts (supplementary fig. S7, Supplementary Material online).

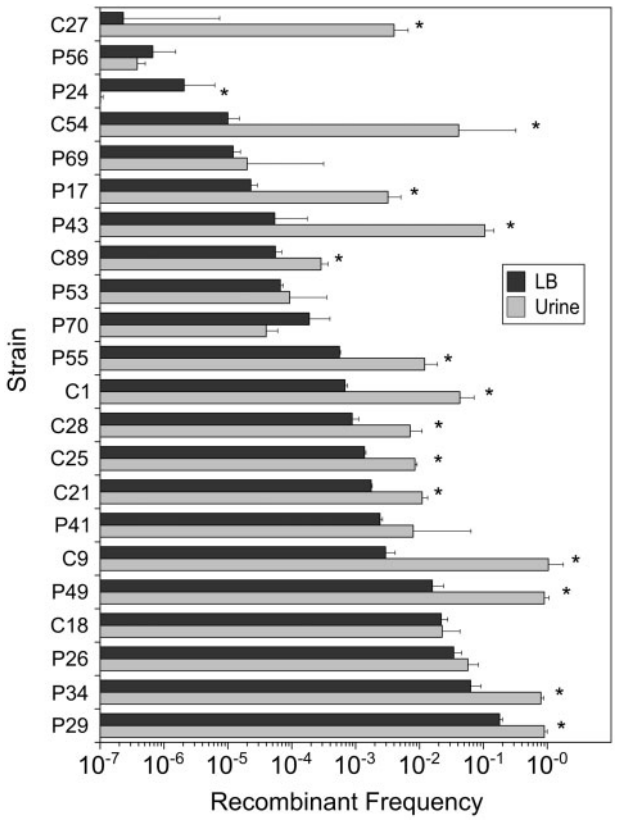

Fig. 3. TetR recombinant frequencies in LB and urine. Frequencies (median ± SEM) measured in a subset of 22 pRhomo transformed *Escherichia coli* strains in LB (black bars) and urine (gray bars) are shown. Strain names begin with either "P" or "C" indicating that the strain is a pathogen or a commensal, respectively. The asterisk denotes significant differences (*t*-test for log10-transformed values; $P < 0.05$) between the frequency in urine and in LB. To facilitate understanding, strains appear ordered by recombinant frequency in LB.

Recombinant Frequency and Number of Virulence Genes

The presence of different extraintestinal virulence genes was previously studied in 116 of our strains (Picard et al. 1999; Duriez et al. 2001). Therefore, we analyzed the correlation of recombinant frequencies and the number of virulence genes of 116 strains (71 commensals and 45 pathogens). Taking into account that pathogens have higher recombinant frequencies than commensals and also a higher number of virulence genes, a correlation between recombinant frequency and the number of virulence genes was suspected. Our results demonstrate that there is a modest, yet significant, positive correlation ($n = 116$, Spearman's rho = 0.23, $P = 0.0127$) between recombinant frequency and the number of virulence genes (fig. 4). However, no correlation was observed between recombination and pathogenicity, measured as the intrinsic virulence of the strains estimated by a mouse sepsis assay (Johnson et al. 2006) although, in this case, there are not enough commensal strains with data to perform a statistically significant study.

In addition, no correlation could be established between frequencies of recombinants and mutation, that is, strains with higher recombinant frequency did not have higher mutant frequency, or between mutation and number of virulence genes. Furthermore, no statistically significant differences were found in mutant frequency or in HR among phylogenetic groups (Kruskal–Wallis test; $P > 0.15$).

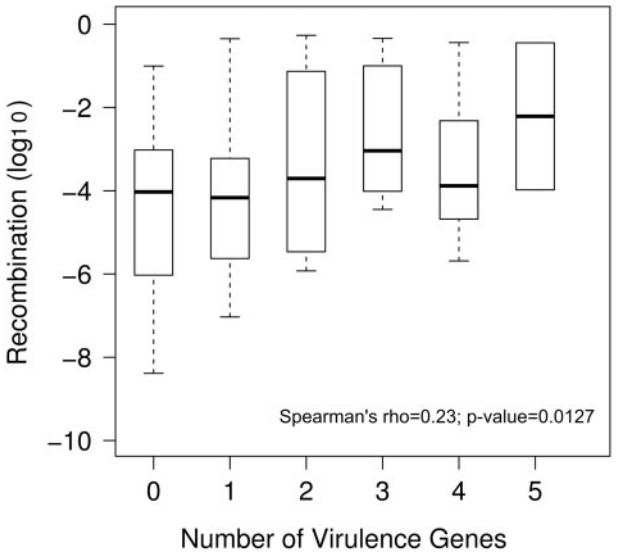

Fig. 4. Recombinant frequency of strains grouped by number of virulence genes. Boxplots represent the median values of recombination frequency of strains grouped by number of virulence genes. A positive correlation (Spearman's rho = 0.23, $P = 0.0127$) can be identified. Strains with more virulence determinants tend to have higher median values of recombination (horizontal line in the boxplot). The depth of the box represents the interquartile range (50% of the data) and the whiskers extend to 1.5 times the interquartile range.

Discussion

Because the *E. coli* species includes both commensal and pathogenic strains and they are easy to isolate and grow, it has been used for years as a model to study population genetics and the transition between commensalism and pathogenicity. Different techniques, such as serotyping, biotyping, random amplified polymorphic DNA, restriction fragment length polymorphism and multilocus enzyme electrophoresis, were used in the first population genetics studies. These studies suggested a clonal structure of the *E. coli* species, whose evolution was largely dominated by mutation (see e.g., Tenaillon et al. 2010, and references therein). Advances in molecular technology have permitted the refinement of results on mutation and recombination at the molecular level, demonstrating that recombination is a key parameter in the evolution of *E. coli* (Schubert et al. 2009; Tenaillon et al. 2010). However, although these new genetic techniques are very useful for studying the effect of mutation and recombination on bacterial evolution, they present a series of drawbacks, including low resolution and incapacity to separate genetic events from the process of selection.

It is generally admitted that PAIs are not solely vertically transmitted, but spread within the *E. coli* species by recombination. For instance, PAIs can be transferred between different *E. coli* strains by F-plasmid mediated mobilization, demonstrating that HGT and HR play major roles in horizontal transfer of PAIs within the *E. coli* species (Schubert et al. 2009). Wirth et al. (2006) suggested that *E. coli* pathogenic strains have undergone increased rates of recombination in genes distributed throughout the genome, likely accelerating the evolution to pathogenicity. As a consequence, they proposed that recombination and virulence are causally related. However, these results relied on the hypothesis that *E. coli* had four phylogenetic subgroups and clustered all strains not belonging to these groups as recombinants (Falush et al. 2006; Denamur et al. 2010). We now know that there are close to seven solid phylogenetic groups in the species (Clermont et al. 2013). Therefore, the question of the link between recombination and virulence has not been solved yet.

We decided to test more directly whether or not *E. coli* pathogenic strains facing, in principle, higher selection pressures from host defenses than commensals, are more prone to adaptive evolution by HR than their commensal counterparts. Guided by this important evolutionary question, we have analyzed the actual potential for genetic change (mutation and HR) of a panel of well-characterized naturally occurring *E. coli* strains, including both commensals and extraintestinal pathogens, from different origins. We have experimentally characterized both mutant frequency and the rate of recombination between homologous DNA sequences, and studied the relationship of these parameters with the isolate origin, phylogenetic group, number of virulence genes, and pathogenicity. Note here that our experimental approach exclusively permits the study of the process that allows recombination between homologous DNA sequences, which is, obviously, only a part of the whole process of HGT.

Concerning the variability of mutation rate, we corroborate the results found in previous studies (LeClerc et al. 1996; Matic et al. 1997). We show the existence of some variability but no significant differences between mutant frequencies of commensal and ExPEC strains. However, differences in the distribution of mutant frequencies can be observed between commensal and uropathogenic strains, confirming previous results from Denamur et al. (Denamur et al. 2002).

The results for HR are surprisingly much more marked. First, the diversity of recombinant frequency observed extends over many orders of magnitude. Hence the ability to recombine homologous DNA, as measured by our system, is extremely diverse within the species: Although some clones show 10% of recombinants other show one in a billion. Second, these frequencies are significantly enhanced in the environment most commonly encountered by ExPEC, that is, urine, as demonstrated for a group of strains. This means that recombination is variable not only among strains but also among environments. Third and most importantly, our results indicate that ExPEC strains have a significantly higher frequency of recombinants (about 2 orders of magnitude) than commensal isolates. This is also associated with a significant positive correlation between recombinant frequencies with the number of virulence genes of each strain.

These results suggest that, as for mutation rate, which shows both diversity among strains and among environments (Bjedov et al. 2003), selection can act on the rate of recombination within *E. coli*. The large diversity we observed is a prerequisite for selection to be able to operate. Upon arrival in a new environment, the benefits linked to the acquisitions of some foreign genetic material present in other strains, such as PAIs, or the acquisition of different alleles (or its generation by intrachromosomic gene shuffling) promoting adaptation to local conditions, may drive the selection for higher recombination rates. According to that hypothesis, the difference between commensals and pathogens suggests that pathogens may have been facing more challenging environments than their commensal counterparts. Moreover, it suggests that thanks to this enhanced recombination, ExPEC may have an increased potential for further adaptive evolution through recombination. It is important to highlight that, although HR is a key step in the overall HGT process, selection could also act on other steps, such as the rate of DNA transfer.

On the other hand, the intensity of intrinsic oxidative stress appears to be quite variable among natural isolates of *E. coli* grown in urine (Aubron et al. 2012). Because oxidative stress is responsible for DNA damage (Imlay 2013) and recombination is a major DNA repair pathway, increased recombination rates may also be selected by their higher capacity to cope with the stressful environment represented here by human urine. These hypotheses may not be mutually exclusive, as both roles of recombination (DNA-repair and the acquisition of new alleles) can be subjects of selection and act synergistically. However, regardless of the nature of the selective pressure acting on recombination, the resulting increased

variability may play an important role in bacterial evolution in a manner similar to what has been suggested for stress-induced mutagenesis (Tenaillon et al. 2004).

Overall, our results demonstrate that a large diversity in the rate of HR exists within *E. coli* species and that this diversity may correlate with the life style (commensal or pathogen). Further work will have to be done to unravel the mechanisms and the selective forces driving this recombination diversity and to uncover its dynamics and long-term impact on bacterial species genetic diversity.

Materials and Methods

Bacteria, Plasmids, and Primers

Escherichia coli K-12 strains, plasmids, and primers used in this study are shown in table 1 and supplementary table S2, Supplementary Material online.

Commensal and Pathogenic *E. coli* Strains

In total, 160 human *E. coli* strains (90 commensal strains, isolated from feces of healthy persons, and 70 extraintestinal pathogens involved in various pathologies) from well-characterized collections gathered in the 1980s from different countries (France, Croatia, Germany, Mali, Canada, the Unites States, Australia) (Picard et al. 1999; Duriez et al. 2001; Hommais et al. 2005), were used for this study. Main characteristics and all experimental data utilized in this work are shown in the supplementary table S1, Supplementary Material online.

Virulence Genes

The presence of seven different extraintestinal virulence determinants (*sfa/foc, pap, afa, hly, cnf, aer,* and *ibeA*) was previously determined by polymerase chain reaction (PCR) (Picard et al. 1999).

Pathogenicity

The intrinsic extraintestinal virulence of the strains was studied by scoring the lethality in a mouse septicemia model as described previously (Picard et al. 1999; Johnson et al. 2006). Briefly, ten mice were inoculated by 10^8 colony forming units of bacteria subcutaneously in the neck and were observed for 7 days. In this model, the K-12 MG1655 strain does not kill mice whereas the ExPEC strain CFT073 kills all the inoculated mice (Johnson et al. 2006).

Media, Antibiotics, and Growth Conditions

Strains were grown in LB. Antibiotics and concentrations (in μg/ml) were ampicillin (Amp, 100), kanamycin (Kan, 30), gentamicin (Gm, 10), tetracycline (Tet, 10), and rifampicin (Rif, 100). Human urine was collected from several healthy male donors who were not taking any medication, pooled, filtered, and stored at −20 °C before use.

Construction of Plasmids and Strains

Construction of plasmids pRhomo (supplementary fig. S1, Supplementary Material online) and strains MG1655 *att*Tn*7*::*tetA1-kan-tetA2* and MG1655 *att*Tn*7*::*tetA1-kan-tetA2 recA* (supplementary fig. S2, Supplementary Material online) is described in detail in Supplementary Material online.

Estimation of RifR Mutant Frequencies

The mutant frequencies of 160 strains were estimated by monitoring their capacity to generate mutations conferring RifR in at least three independent clones for each strain. Between 10^2 and 10^3 cells from an overnight culture were inoculated into LB tubes and grown with shaking 24 h at 37 °C. Appropriate dilutions of the culture were plated onto LB petri dishes containing rifampicin (100 μg/ml) and incubated for 24 h at 37 °C. Mutant frequencies were

Table 1. *Escherichia coli* K-12 Strains and Plasmids Used in This Work.

K-12 Strain	Genotype/Phenotype	Origin/Reference
DH5α	Δ*(lacZ)M15, recA1, endA1, hsdR17, supE44*	Laboratory stock
MG1655	K-12 wild-type strain	Laboratory stock
MG1655 *recA*	MG1655 *recA*938::Tn*9–200*	Wertman et al. 1986
ME12	MG1655 *lacZ* Δ*3′-lacZ* Δ*5′-yfp*	Elez et al. 2007
ME12 *recA*	ME12 Δ *recA::Kan* (KmR)	Elez et al. 2007
Plasmid	Main Characteristics/Resistance	Origin/Reference
pGEM-T	Cloning PCR products; AmpR	Promega, USA
pGEM-T Easy	pGEM-T variant; AmpR	Promega, USA
pUC19	Cloning vector; AmpR	Invitrogen, USA
pBBR1MCS3	Cloning vector; TetR	Kovach et al. 1995
pBBR1MCS5	Cloning vector; GmR	Kovach et al. 1995
pUCGm*lox*	Vector with *aacC1*; GmR	Quenee et al. 2005
pRhomo	pGEM-T easy with *tetA* Δ*5′-aacC1-tetA*Δ*3′*; AmpR, GmR	This study
pGRG36	Contains Tn7 transposition machinery, a thermosensitive origin of replication and an origin of conjugation; AmpR	McKenzie and Craig 2006
pTetKanTet	A pGRG36 derivative which contains two incomplete fragments of the *tetA* resistance gene interspaced by a kanamycin resistance gene (*tetA1-kan-tetA2*); KmR, AmpR	This study

calculated as the median number of rifampicin resistant colonies divided by the median number of viable colonies on LB-agar plates without antibiotic.

Estimation of TetR Recombinant Frequencies

In total, 160 strains (90 commensal and 70 pathogens) were transformed with plasmid pRhomo. Three independent transformants from each strain were purified onto LB-agar plates containing ampicillin and gentamicin and allowed to grow overnight at 37 °C. To diminish the probability of inoculating cells that contain plasmids already recombined, between 10^1 and 10^2 cells from each purified transformant were inoculated into LB with ampicillin but not gentamicin. Cultures were grown overnight with shaking until saturation. Appropriate dilutions were plated onto LB agar containing ampicillin and tetracycline and incubated for 24 h. TetR recombinant frequencies were calculated as the median number of recombinants (AmpR and TetR) divided by the median number of viable on LB-agar plates containing only ampicillin.

MG1655 *att*Tn*7*::*tetA1-kan-tetA2*, MG1655 *recA att*Tn*7*::*tetA1-kan-tetA2*, *E. coli* B, HS, F11, 536 and CFT073 either pRhomo-transformed or with the *att*Tn*7* inserted construction, and a subset of 22 pRhomo-transformed strains were processed similarly: 10^2 cells from three independent overnight cultures were inoculated in parallel in LB-Amp or urine-Amp.

Statistical Analysis

Shapiro–Wilk test for normality was conducted for all distributions. When possible, parametric Student's *t*-test was applied to log-transformed data. When normality was not assessed, Mann–Whitney *U* test for nonparametric data was chosen for comparisons between distributions. After a significant Kruskall–Wallis test, Wilcoxon signed rank test with Bonferroni correction was used for multiple comparisons. To assess correlation between two variables, Pearson's product–moment or Spearman's rank correlation tests were used for lineal and monotonic relationships, respectively. Fisher's exact test was conducted to test for differences in the number of mutators on each group. Statistical test with *P* values under 0.05 was considered to be significant. All statistical analyses were performed using R software (http://www.R-project.org, last accessed March 23, 2015).

Supplementary Material

Supplementary methods, references, figures S1–S7, and tables S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Ministerio de Economía y Competitividad, Instituto de Salud Carlos III, the Spanish Network for Research in Infectious Diseases RD12/0015/0029, and Fondo de Investigación Sanitaria Grant PI13/00063. J.T. was supported by a grant from the Fondation pour la Recherche Médicale. The authors are grateful to Nicolas Plault for technical assistance and Michael McConnell for useful scientific comments and English revision of the manuscript.

References

Aubron C, Glodt J, Matar C, Huet O, Borderie D, Dobrindt U, Duranteau J, Denamur E, Conti M, Bouvet O. 2012. Variation in endogenous oxidative stress in *Escherichia coli* natural isolates during growth in urine. *BMC Microbiol.* 12:120.

Bjedov I, Tenaillon O, Gerard B, Souza V, Denamur E, Radman M, Taddei F, Matic I. 2003. Stress-induced mutagenesis in bacteria. *Science* 300:1404–1409.

Clermont O, Christenson JK, Denamur E, Gordon DM. 2013. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep.* 5:58–65.

Cooper TF. 2007. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biol.* 5:e225.

Denamur E, Bonacorsi S, Giraud A, Duriez P, Hilali F, Amorin C, Bingen E, Andremont A, Picard B, Taddei F, et al. 2002. High frequency of mutator strains among human uropathogenic *Escherichia coli* isolates. *J Bacteriol.* 184:605–609.

Denamur E, Matic I. 2006. Evolution of mutation rates in bacteria. *Mol Microbiol.* 60:820–827.

Denamur E, Picard B, Tenaillon O. 2010. Population genetics of pathogenic *Escherichia coli*. In: Robinson DA, Falush D, Feil EJ, editors. Bacterial population genetics in infectious disease. Hoboken (NJ): John Wiley & Sons, Inc. p. 269–286.

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

Didelot X, Maiden MC. 2010. Impact of recombination on bacterial evolution. *Trends Microbiol.* 18:315–322.

Didelot X, Meric G, Falush D, Darling AE. 2012. Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. *BMC Genomics* 13:256.

Duriez P, Clermont O, Bonacorsi S, Bingen E, Chaventre A, Elion J, Picard B, Denamur E. 2001. Commensal *Escherichia coli* isolates are phylogenetically distributed among geographically distinct human populations. *Microbiology* 147:1671–1676.

Elez M, Radman M, Matic I. 2007. The frequency and structure of recombinant products is determined by the cellular level of MutL. *Proc Natl Acad Sci U S A.* 104:8935–8940.

Falush D, Torpdahl M, Didelot X, Conrad DF, Wilson DJ, Achtman M. 2006. Mismatch induced speciation in *Salmonella*: model and data. *Philos Trans R Soc Lond B Biol Sci.* 361:2045–2053.

Groisman EA, Ochman H. 1996. Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* 87:791–794.

Guttman DS, Dykhuizen DE. 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* 266:1380–1383.

Hacker J, Kaper JB. 2000. Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol.* 54:641–679.

Hommais F, Pereira S, Acquaviva C, Escobar-Paramo P, Denamur E. 2005. Single-nucleotide polymorphism phylotyping of *Escherichia coli*. *Appl Environ Microbiol.* 71:4784–4792.

Imlay JA. 2013. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat Rev Microbiol.* 11:443–454.

Johnson JR, Clermont O, Menard M, Kuskowski MA, Picard B, Denamur E. 2006. Experimental mouse lethality of *Escherichia coli* isolates, in relation to accessory traits, phylogenetic group, and ecological source. *J Infect Dis.* 194:1141–1150.

Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop RM 2nd, Peterson KM. 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166:175–176.

Lawrence JG, Ochman H. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci U S A.* 95:9413–9417.

Lawrence JG, Roth JR. 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143:1843–1860.

Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, Tenaillon O. 2007. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. *Mol Biol Evol.* 24:2373–2384.

LeClerc JE, Li B, Payne WL, Cebula TA. 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* 274: 1208–1211.

Levin BR, Cornejo OE. 2009. The population and evolutionary dynamics of homologous gene recombination in bacterial populations. *PLoS Genet.* 5:e1000601.

Lopez E, Elez M, Matic I, Blazquez J. 2007. Antibiotic-mediated recombination: ciprofloxacin stimulates SOS-independent recombination of divergent sequences in *Escherichia coli*. *Mol Microbiol.* 64: 83–93.

Lovett ST, Drapkin PT, Sutera VA Jr, Gluckman-Peskind TJ. 1993. A sister-strand exchange mechanism for *recA*-independent deletion of repeated DNA sequences in *Escherichia coli*. *Genetics* 135: 631–642.

Mao EF, Lane L, Lee J, Miller JH. 1997. Proliferation of mutators in A cell population. *J Bacteriol.* 179:417–422.

Matic I, Radman M, Taddei F, Picard B, Doit C, Bingen E, Denamur E, Elion J. 1997. Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science* 277:1833–1834.

Matic I, Rayssiguier C, Radman M. 1995. Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. *Cell* 80:507–515.

McKenzie GJ, Craig NL. 2006. Fast, easy and efficient: site-specific insertion of transgenes into enterobacterial chromosomes using Tn7 without need for selection of the insertion event. *BMC Microbiol.* 6:39.

Noller AC, McEllistrem MC, Stine OC, Morris JG Jr, Boxrud DJ, Dixon B, Harrison LH. 2003. Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. *J Clin Microbiol.* 41:675–679.

Oliver A, Baquero F, Blazquez J. 2002. The mismatch repair system (mutS, mutL and uvrD genes) in *Pseudomonas aeruginosa*: molecular characterization of naturally occurring mutants. *Mol Microbiol.* 43:1641–1650.

Oliver A, Canton R, Campo P, Baquero F, Blazquez J. 2000. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* 288:1251–1254.

Paul S, Linardopoulou EV, Billig M, Tchesnokova V, Price LB, Johnson JR, Chattopadhyay S, Sokurenko EV. 2013. Role of homologous recombination in adaptive diversification of extraintestinal *Escherichia coli*. *J Bacteriol.* 195: 231–242.

Picard B, Garcia JS, Gouriou S, Duriez P, Brahimi N, Bingen E, Elion J, Denamur E. 1999. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun.* 67:546–553.

Quenee L, Lamotte D, Polack B. 2005. Combined sacB-based negative selection and cre-lox antibiotic marker recycling for efficient gene deletion in *Pseudomonas aeruginosa*. *Biotechniques* 38:63–67.

Ramaswamy S, Musser JM. 1998. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tuber Lung Dis.* 79:3–29.

Schubert S, Darlu P, Clermont O, Wieser A, Magistro G, Hoffmann C, Weinert K, Tenaillon O, Matic I, Denamur E. 2009. Role of intraspecies recombination in the spread of pathogenicity islands within the *Escherichia coli* species. *PLoS Pathog.* 5:e1000257.

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

Spratt BG, Hanage WP, Feil EJ. 2001. The relative contributions of recombination and point mutation to the diversification of bacterial clones. *Curr Opin Microbiol.* 4:602–606.

Stecher B, Maier L, Hardt WD. 2013. "Blooming" in the gut: how dysbiosis might contribute to pathogen evolution. *Nat Rev Microbiol.* 11:277–284.

Taddei F, Radman M, Maynard-Smith J, Toupance B, Gouyon PH, Godelle B. 1997. Role of mutator alleles in adaptive evolution. *Nature* 387:700–702.

Tenaillon O, Denamur E, Matic I. 2004. Evolutionary significance of stress-induced mutagenesis in bacteria. *Trends Microbiol.* 12: 264–270.

Tenaillon O, Le Nagard H, Godelle B, Taddei F. 2000. Mutators and sex in bacteria: conflict between adaptive strategies. *Proc Natl Acad Sci U S A.* 97:10465–10470.

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

Tenaillon O, Toupance B, Le Nagard H, Taddei F, Godelle B. 1999. Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics* 152:485–493.

Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, et al. 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* 5:e1000344.

Trobner W, Piechocki R. 1984. Selection against hypermutability in *Escherichia coli* during long term evolution. *Mol Gen Genet.* 198: 177–178.

Wertman KF, Wyman AR, Botstein D. 1986. Host/vector interactions which affect the viability of recombinant phage lambda clones. *Gene* 49:253–262.

Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, Cruveiller S, Chane-Woon-Ming B, Medigue C, Lenski RE, Schneider D. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci U S A.* 110: 222–227.

Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, et al. 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol.* 60: 1136–1151.