Natural *Escherichia coli* isolates rapidly acquire genetic changes upon laboratory domestication

Bin Liu,1,2† Gustavo Eydallin,3† Ram P. Maharjan,3 Lu Feng,1,2 Lei Wang1,2 and Thomas Ferenci3,*

**Abstract**

The adaptation of environmental bacteria to laboratory conditions was analysed through the exploration of genomic changes in four strains of *Escherichia coli* freshly isolated from their natural habitats and belonging to different taxonomic clusters. Up to 25 mutations were present in all cultures of natural isolates within 10 days of transfer in rich media or with a single growth cycle involving an extended stationary phase. Among numerous individual mutations, two genes were affected in parallel in distinct backgrounds. Mutations in *rpoS* (encoding sigma factor RpoS), altering a multiplication-survival trade-off in *E. coli*, were present in isolates derived from all four different ancestors. More surprisingly, two different natural isolates acquired mutations in *mutL*, affecting DNA mismatch repair, and a third also involved higher mutation rates. The elevated mutation rates in these isolates indicate the danger of increased genetic instability arising from laboratory domestication. Neither *rpoS* nor mutator mutations were detected in the already-acclimatized MG1655 laboratory strain; only one or no new mutations were present in the laboratory strain under the same culture conditions. Our results indicate rapid adaptation to the laboratory environment. Ancestor-specific responses also arise in the laboratory and mutational events are also sensitive to culture conditions such as extended stationary phase. To maintain natural isolates in a stable state, our data suggest that the transition of strains to the laboratory should minimize culture cycles and extended stationary phase.

**INTRODUCTION**

Laboratory-based studies of bacteria invariably involve isolation and storage in environments foreign to free-living bacteria [1]. The likelihood is that bacteria adapt genetically in the new habitat; this is what we term laboratory domestication [2]. Laboratory investigations of bacteria from natural habitats often involve multiple rounds of cultivation in nutrient-rich medium either as a liquid batch culture or on plates. Culturing in batch culture or on plates does not itself present constant environments, and rapid multiplication in laboratory media is followed by a variable period of stationary phase after a few hours of exponential growth. These periods present different selection conditions for domestication and we recently found that the same bacterium with periods present different selection conditions for domestication. Laboratory investigations of bacteria from natural habitats often involve multiple rounds of cultivation in nutrient-rich medium either as a liquid batch culture or on plates. Culturing in batch culture or on plates does not itself present constant environments, and rapid multiplication in laboratory media is followed by a variable period of stationary phase after a few hours of exponential growth. These periods present different selection conditions for domestication and we recently found that the same bacterium with periods present different selection conditions for domestication.

Most previous observations on the change of bacteria in the laboratory were on strains already partially or fully adapted to the laboratory environment. A variety of reports are available on the acquisition of mutations upon laboratory culture and storage. *E. coli* changes genetically during handling for short periods in laboratory media and mutants are common within 3 days of culture [3, 4]. Also, bacteria maintained in long-term stab cultures were found to change in several aspects [5–8]. Extended incubation in stationary phase over several days in rich medium gave rise to population heterogeneity in the laboratory [9] or even during shipping stab cultures between laboratories [10]. Laboratory passage effects are not restricted to *E. coli* and were also observed in diverse bacteria including pathogens [11–18].

*Correspondence: Thomas Ferenci, tom.ferenci@sydney.edu.au*

**Keywords:** Laboratory adaptation; comparative genomics; mutations acquired in the laboratory; multiplication–survival trade-off.

**Abbreviation:** GASP, growth advantage in stationary-phase. 

†These authors contributed equally to this work.

The GenBank/EMBL/DDBJ accession numbers of genomes sequenced in this work are CP012376–CP012380.

Five supplementary tables are available with the online Supplementary Material.
The phenotypic adaptations of undomesticated, freshly isolated strains of E. coli exhibited a range of domestication effects in the strains [2]. The changes of replicate cultures were observed over short periods in laboratory media. There was also a lack of similarity between domestication changes in taxonomically different isolates of the same species or with different conditions. These results did not entirely answer whether domestication involved similar mutations in species E. coli because phenotypic effects may be different with the same mutations in different genotypes. Here we show by a genomic analysis whether domestication is a uniform process in this species.

In this study, we studied four independent isolates of E. coli derived from distinct environments with minimal transfers and processing [2]. The choice of the four strains acting as natural ancestors was based on their different phenotypic and genotypic characteristics. As shown below, the genomes of these strains are also taxonomically divergent. We sub-cultured these four strains in rich medium under controlled conditions. The rich laboratory environments are themselves complex in having multiple carbon and nitrogen sources with sequential utilization of resources [19, 20]. Also adding to the complexity of the laboratory environment is that feast and famine states upon entry into stationary phase occur within a single 24 h period [21]; the length of this stationary phase is also a factor in adaptation and mutation acquisition, as seen in the GASP (growth advantage in stationary-phase) phenomenon [2, 22, 23].

A possible explanation for domestication is that bacteria confront challenges in nature through a wide spectrum of costly adaptive responses and resistance mechanisms that are not essential or even beneficial in laboratory conditions [24–26]. Does domestication involve a limited number of mutations removing some common, expensive resource allocation costs? It is not obvious what the likeliest target(s) of selection are in domestication. An increased maximal growth rate has been proposed for E. coli [27] but our findings do not generally show faster growth in domesticated isolates [2]. The laboratory cultures of the four ancestors studied exhibited changes with variations in phenotypes [2]. To identify the mutational events and targets of adaptation, we here report the genomics of domestication. We sequenced isolates from each ancestor in each of two laboratory environments.

The analysis of whole genomes in domesticated isolates showed multiple, distinct changes confirming that several mutations occur during a few days of laboratory exposure of undomesticated strains. We also found that the target of selection is not entirely uniform in different strains; some of the mutations were present in an ancestor-specific manner. Changes in few genes were common to the four distinct backgrounds studied, but even these were not in all isolates. Thus, our results are beginning to unravel the characteristics of domestication and its complex rules.

**METHODS**

**E. coli isolates**

The GE3 and MEM isolates were from human faeces of different individuals, PAR was from cockatoo faeces and WAT was from park pond water [2]. MG1655 was a laboratory stock of a commonly used K-12 strain. For the detailed protocol of E. coli isolations, see [2]. All E. coli isolates were resuspended in Luria–Bertani (LB) broth plus 15% v/v glycerol and were saved at −80°C.

**Media and growth conditions**

Bacteria were generally grown at 37°C in peptone yeast extract broth (PYE: 10 g l⁻¹ peptone, 5 g l⁻¹ yeast extract and 5 g l⁻¹ NaCl) [28]. Briefly, the LE1 and LE2 domestication experiments involved a 100 µl aliquot of frozen E. coli strains containing >10⁻⁶ cells inoculated into 100 ml of PYE medium and incubated at 37°C with agitation. For LE1, the incubation continued for 10 days. For LE2, after 24 h, 10 µl of culture was transferred into 100 ml fresh medium and incubated at 37°C with agitation; this was repeated for a total of 10 days, providing approximately 13.3 generations of growth per day. More details are given by Eydallin et al. [2].

**Phenotypic characterization of strains**

Glycogen staining with iodine is indicative of the level of RpoS in E. coli [29]. For ancestor strains, 2 µl sample from a glycerol stock containing approximately 10⁷ cells ml⁻¹ was directly spotted on an LB plate. For the five replicate populations derived from each ancestor, unfraccionated 10 day cultures were used to show the majority characteristic in each population. A 10 µl sample originating from glycerol stock (approximately 4×10⁸ cells ml⁻¹) was 10-fold diluted in PYE and 2 µl of this suspension was plated as shown in Fig. 2 on an LB agar plate and grown overnight at 37°C. Plates were then transferred to 4°C, and kept for 16–18 h before staining with iodine solution for 5 min.

**Measurement of mutation frequencies upon domestication**

Acquisition of resistance to rifampicin from rifampicin-sensitive E. coli was used for measurement of mutation frequencies. A loopful of frozen glycerol stock of each strain containing ~10⁸ bacteria was transferred into 5 ml nutrient broth in McCartney bottles and allowed to grow overnight at 37°C with orbital shaking at 200 r.p.m. Bacteria from the overnight grown cultures (1 ml) were then harvested by centrifugation (5000 g, 2 min), resuspended with 100 µl of sterile PBS and plated on LB agar plates containing 100 µg ml⁻¹ rifampicin. The plates were then incubated at 37°C for 24 h to detect RifR mutants. For total c.f.u. counts, aliquots of appropriately diluted overnight cultures were plated on LB agar plates. The mutation frequencies were then calculated from number of RifR colonies and the total c.f.u. based on six independent cultures of each population.
Genome DNA methods

We obtained the complete genomes for five ancestral *E. coli* strains (MG1655, GE3, MEM, PAR and WAT) by a combination of PacBio and Solexa sequencing technologies. The genomic DNA was purified using the MOBIO PowerClean DNA CleanUp kit (MOBIO), followed by 10 kb library construction for P6-C4 chemistry without shearing. Aliquots of libraries were sequenced after BluePippin (Sage Science) size selection at 7 kb to obtain long libraries crossing over identical sequences. Two SMRT cells of each strain were sequenced by PacBio RS II with a 240 min movie. *De novo* assembly was conducted by using a hierarchical genome assembly process 3 workflow, including Quiver consensus polishing [30]. We also performed genome sequencing for each ancestral strain with Solexa pair-end sequencing technology [31] using Solexa Genome Analyzer Ix (Illumina), which gave 130–175-fold coverage. The Solexa reads generated were culled of duplicates with identical reads which are presumed to be from replicates arising in the PCR step, and they were then mapped to corresponding complete genomes generated by PacBio sequencing for correction of potential sequencing errors.

The 20 domesticated strains were sequenced by Solexa pair-end sequencing technology, which gave 107–178-fold coverage. The generated reads were mapped to the corresponding ancestral genome as described above. SNPs, insertions, deletions and duplications in these genomes of domesticated strains were detected as described previously [32, 33].

Annotation and genome sequence analyses for ancestors and laboratory-adapted strains

ORFs from 30 amino acids in length were predicted using Glimmer 3.0. Function predictions were based on BLASTp similarity searches in the UniProtKB, GenBank and SwissProt protein databases, as well as the Clusters of Orthologous Groups (COG) database (www.ncbi.nlm.nih.gov/COG). Transfer RNA and ribosomal RNA genes were predicted using tRNAscan-SE. Artemis [34] was used to collate data and facilitate annotation.

The phylogenetic tree of the available *E. coli* core genomes was constructed from the concatenated alignments of the 1870 genes in the *E. coli* core genome using the method described previously [35]. The sequence typing of the ancestors was based on the sequence of seven housekeeping genes (*adk, fumC, gyrB, icd,-mdh, purA* and *recA*) [36]. The GenBank accession numbers of genomes sequenced in this work are CP012376–CP012380.

RESULTS

Genome characteristics of the four ancestral undomesticated strains

The four minimally processed natural isolates (the ancestors of domesticated strains) were distinct in taxonomic, phenotypic and phylogenetic groupings (B1 type for GE3 and WAT, B2 type for PAR and A type for MEM) [2]. Characteristic of *E. coli*, the sizes of the circular chromosomes of GE3, MEM, PAR and WAT were found to be 4 909 965, 4 928 082, 4 859 743 and 4 847 481 bp, respectively. The strains differed in sequence type (ST) and GE3 is ST56, WAT is ST155, PAR is ST1858 and MEM is ST607. Both PAR and WAT contained a plasmid, and the sizes of plasmids were 7987 and 109 373 bp, respectively. There was no plasmid found in GE3 and MEM. In total, 4717, 4824, 4686 and 4669 genes were identified in GE3, MEM, PAR and WAT, respectively. There were also 76–87 tRNA genes and 19–22 rRNA operons in the genome of these four strains. In the phylogenetic tree based on the genomes (Fig. 1), the genome of PAR is close to those of uropathogenic *E. coli* strains, such as strain 536. Six pathogenicity islands have been found in strain 536 [37]. However, only a genomic island similar to PAI IV in strain 536, which contains a yersiniabactin biosynthesis determinant, was detected in strain PAR. The genomes of strains WAT and GE3 are related in Fig. 1 with those of enterogaergative *E. coli* strains, such as strain 55989 and O104:H4. However, no virulence factor typical of enterogaergative *E. coli*, such as *astA, aggR* and *aggA* [38], can be found in the genome or plasmid of WAT and GE3. The genome of MEM is more similar to that of strain P12b, which is a non-pathogenic commensal *E. coli* strain. As shown in Fig. 1, the four strains GE3, MEM, PAR and WAT had genome sequences that were widely dispersed across the tree containing genome-based relationships of available sequenced *E. coli* strains and were also separate from the commensal K-12 MG1655 strain commonly used in laboratory experiments. So the domestication changes discussed below represent adaptations of different genotypes across the species to see if there are common, *E. coli*-specific change(s).

Changes in selected isolates after 10 days in laboratory culture

The environmental regime used for laboratory domestication involved a single cycle of growth in rich medium and a period of stationary phase for a total of 10 days (condition LE1). A variation (LE2) was to have 10 serial transfers in rich medium, each day for 10 days [2]. The procedures were applied to the four strains GE3, PAR, MEM and WAT isolated with minimal processing to test for domestication changes. For comparison, we also studied the reference *E. coli* K-12 MG1655 strain which has been in the laboratory for more than 50 years [39], although even this strain continues to diverge in different laboratories [40].

We sequenced 20 isolates in total. Two independent clones A1 and A5 were from one population of each ancestor subjected to the LE1 treatment. Two further isolates, A1 and A5, were from one population of each ancestor subjected to LE2. The phenotypes of these individual strains were previously characterized and the chosen strains from each ancestor were typical of those found multiple times in the replicate populations [2]. The total number of mutations in re-sequenced clones differed between ancestral sources, but there was also some diversity in the number of mutations in the different conditions with the same strain (Fig. 2). Strains
derived from ancestor GE3 and especially the already domesticated MG1655 strain had the fewest mutations; strikingly, some other clones had over 20 mutations.

The observed difference in the total number of mutations in laboratory-adapted isolates could have been due to a difference in mutation rates between ancestors or through the spread of mutator strains in the evolving populations. The basal mutation rate varies up to 100-fold between natural isolates of *E. coli* [41] so we tested whether the difference in the number of mutations between the isolates in Fig. 2 was due to differences in ancestral mutation frequencies. As shown in Fig. 3, the laboratory MG1655 strain had a mutation frequency of $4.3 \times 10^{-8}$ (95% CI, 2.7–6.0 × 10$^{-8}$) when tested for mutations to rifampicin resistance. This value is an order of magnitude lower than the four natural isolates ($P=0.03$ by one-way ANOVA). The frequency for MG1655, and indeed the other isolates, still falls within the variable range observed with rifampicin resistance frequencies in species *E. coli* [41]. The isolates were more similar among themselves ($P=0.18$ by one-way ANOVA) but their mutational frequencies do not completely correlate with, for example, the lower number of mutations in GE3 isolates in Fig. 2. Individual explanations are needed for the number of mutations in each clone, as discussed below. In Table 1, we highlight what we consider to be the most interesting mutations, including the mutator mutations which result in high mutation frequencies, as we next discuss for each ancestral background.

### Mutational changes in domesticated cultures – strain MG1655

MG1655 cultures contained the lowest total number of mutations in the four clones sequenced, including isolates with no mutations. The lack of mutations was consistent with the absence of phenotypic changes in MG1655 after 10 days domestication in either the LE1 or LE2 environment [2]. These results are consistent with a low mutation frequency in this strain, but an alternative explanation is that MG1655 is the most domesticated strain and the least selected for changes after 10 days in the laboratory medium. Mutations in LE1 included *ydcI* mutations in both isolates. Interestingly, two different alleles of *ydcI* were found in the same population (Table S1). This is unlikely by chance and suggested a parallel beneficial effect through these mutations. The benefit may be as in GASP mutants, but the role of *ydcI* mutations is unclear from the gene’s documented properties (as a putative regulatory gene [42]).

### Mutational changes in domesticated cultures – strain GE3

This human faecal isolate obtained the next fewest number of new mutations after domestication. GE3 acquired *rpoS* mutations in both LE1 and LE2 conditions. Despite the elevated mutation frequency in GE3 relative to MG1655 (Fig. 3), the low number of mutations in isolates suggests that not many changes are rapidly selected. Importantly, sequenced isolates each contained different alleles of *rpoS* (Table S2), which suggests that these were independently enriched through selection from the ancestral culture. Each of these mutations resulted in a loss of function, as indicated by the reduced staining of isolates with iodine and Congo Red [2], which are phenotypic tests sensitive to RpoS protein function in cells [29]. This altered staining was common in all replicate cultures of GE3, as observed by Eydallin et al. [2].
The parallel emergence of such a high proportion of independent rpoS mutations suggested that the loss of RpoS activity is beneficial in laboratory environments in this strain. High levels of RpoS are deleterious to rapid growth \([29, 43]\) and rpoS mutations are of benefit in GASP \([22]\) so it is likely that reduction of RpoS levels is the source of the benefit in the rpoS mutations.

Aside from rpoS mutations, only one other mutation in hpaH was present in one isolate (see Table S2). There is no clear indication that this mutation is beneficial rather than random.

### Mutational changes in domesticated cultures – strain MEM

The most notable result in the LE1 condition was that both isolates contained the same mutation in mutL, which encodes the MutL component of the mismatch repair system of DNA repair. Mutations in mutL result in a mutator phenotype but elevated MutL reduces mutation rate \([44]\). In *E. coli* K-12, a mutL loss results in a 233-fold increase in mutation rate \([45]\). As shown in Fig. 3, the assay of the frequency of spontaneous rifampicin-resistant mutations indicated that the mutation rate after 10 days in the LE1 culture was elevated over 50-fold, from \(1.2 \times 10^{-6}\) to \(6.3 \times 10^{-5}\). Indeed, the very high number of genomic changes in the LE1 clones (14–25 mutations as shown in Fig. 2; see Table S3 for a list of all mutations) is consistent with a mutator effect. The genomic deletion in *mutL* was identical in both LE1 isolates, indicating a clonal expansion, but the pattern of other mutations was slightly different.

RpoS changes were seen in all MEM isolates. Both LE1 and LE2 conditions resulted in rpoS mutations so these were strongly selected in this genotype. This notion is reinforced by two different alleles of rpoS arising in the LE2 culture (Table S3).

In addition to rpoS, a MEM mutation recurrently found in three of four isolates and in both LE1 and LE2 cultures was in corA. This change is unique to the MEM background and suggests that the change of the CorA magnesium/nickel/cobalt transporter \([42]\) is beneficial for this genome to adapt to laboratory culture. The corA mutations in both LE1 and LE2 cultures included the generation of stop codons (Table S3) so the expression of the normally constitutively expressed corA gene is presumably a strain-specific problem for the MEM strain.

Some of the large numbers of mutations in the MEM mutator clones affect other functions (see Table S3 for the full list). It is not clear though whether any of these are selected in domestication or have randomly occurred in the mutator background.
Mutational changes in domesticated cultures – strain PAR

The genomes of the PAR strains (especially LE1 α5, Fig. 2) acquired a larger number of mutations than, for example, GE3 but not as many as the mutL mutator strains in the MEM LE1 condition. The basal mutation frequency of the ancestral GE3 strain and PAR strains was not appreciably different (Fig. 3) so that is not the explanation. The LE1 culture indeed exhibited an elevated mutation frequency above that of PAR ancestor (Fig. 3); however, no mutations were detected in previously identified mutator genes [46] in PAR LE1 α5 (Table S4). Despite the indeterminate cause(s) of elevated mutation rates in PAR LE1, there is, as with MEM and WAT, a danger of genetic instability in domesticated strains.

There is a curious dichotomy between the LE1 and LE2 effects in PAR (Table 1). An rpoS mutation was present in LE2 isolate A1, but not in the LE1 isolates. Conversely, LE1 strains contained two different alleles of proQ. The independent alleles of proQ in PAR isolates suggest parallel adaptation in the LE1 culture, indicative of an adaptive advantage beneficial in domestication. Also interesting is that the PAR proQ strains all had reduced Congo Red staining [12], which is also found in rpoS mutants and is indicative of altered RpoS-related regulation [2]. Additionally, a proQ

![Fig. 3. Mutation frequency changes in domesticated populations. Mutation frequencies were obtained from plating known numbers of bacteria from each ancestor culture as well as the 10 day LE1 and LE2 populations onto plates containing rifampicin. Error bars represents standard deviations of six independent cultures.](image)

**Table 1. Mutations in E. coli K-12 and four distinct natural isolates after 10 days in laboratory culture**

The LE1 and LE2 conditions of domestication are defined in the text. Independent colonies from each condition (α1 and α5 from LE1; A1 and A5 from LE2) were sequenced from each ancestor. The complete lists of mutations are shown in Tables S1–S5 for each strain; these tables also show the detailed properties and sites of the mutations. Not all mutations are included in Table 1; only strain-specific mutations and those shared by more than one ancestor are shown. Different colours indicate particular alleles and identical shadings indicate identical sequence changes within a gene. The MP designation is for mutations in prophage genes in WAT and the numbers in brackets indicate the number of prophage genes affected.

<table>
<thead>
<tr>
<th>Culture condition</th>
<th>Isolate label</th>
<th>MG1655</th>
<th>GE3</th>
<th>MEM</th>
<th>PAR</th>
<th>WAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>LE1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α1</td>
<td>ydcI</td>
<td>rpoS</td>
<td>mutL</td>
<td>rpoS</td>
<td>corA</td>
<td>proQ</td>
</tr>
<tr>
<td>α5</td>
<td>ydcI</td>
<td>rpoS</td>
<td>mutL</td>
<td>rpoS</td>
<td>corA</td>
<td>proQ</td>
</tr>
<tr>
<td>LE2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>rpoS</td>
<td>rpoS</td>
<td>corA</td>
<td>rpoS</td>
<td>rpoS</td>
<td>tufA</td>
</tr>
<tr>
<td>A5</td>
<td>rpoS</td>
<td>rpoS</td>
<td>corA</td>
<td>rpoS</td>
<td>rpoS</td>
<td>tufA</td>
</tr>
</tbody>
</table>
mutation was also found in the short-term laboratory evolution of MG1655 in lactate minimal media [47]. The benefit of the proQ change is difficult to predict however; ProQ has been interpreted as a proline transporter structural element [42]. However, it may also have a regulatory role [48] and ProQ may be an RNA chaperone that affects ProP proline transporter levels. Further studies are needed to test whether proQ and rpoS mutations provide alternative solutions to the same selection environment during domestication and whether they are epistatic.

Unique to PAR, isolates in both LE1 and LE2 acquired mutation in ycgK, encoding a putative protein of unknown function. Several different alleles (see Table S4) were obtained, again suggesting parallel evolution indicative of selection. Here again, the benefit is unknown, but this particular mutation is unique to PAR adaptation. Likewise, the tufA mutations affecting translation elongation factor Tu are different alleles in Table 1 and occurred in both LE1 and LE2 isolates so they provide a strain-specific outcome. So in PAR there are two strain-unique adaptations common to domestication, but very different in function to the strain-specific corA changes in MEM.

The acrA mutations in PAR are interesting in that the WAT cultures also acquired mutations in acrA in the LE1 condition. Thus, changes in efflux may be also a factor in domestication, perhaps because of a lesser cost in a less toxic environment in the laboratory than the one that the ancestors thrived in.

Mutation changes in domesticated cultures – strain WAT

The WAT strain isolates also acquired mutL mutations in both clones in the LE1 environment through clonal expansion. The WAT mutL mutation was a missense mutation that resulted in a less than 10-fold increase in mutation frequencies in Fig. 3, but there was nevertheless a significant increase in the number of detected mutations in WAT LE1, close to the numbers seen with MEM (Fig. 2). The elevated genomic mutation rate in WAT LE1 isolates is also consistent with the significant difference in rate detected by rifampicin mutation assay (P=0.001 by two-tailed t-test assuming unequal variance). The number of mutations in the non-mutator LE2 isolates was also relatively high compared to the other strains even though the ancestral mutation rate was not significantly different (Fig. 3). A contributing factor particularly in domesticated WAT isolates was the significant number of PM changes (prophage-associated mutations) in Table 1. We found that five to nine distinct changes were present in the strains, as documented in more detail in Table S5. Furthermore, these affected more than one prophage. The significance of these events can only be speculated upon and a possibility is that the prophage functions involve a cost in WAT that can be eliminated when growing in the laboratory.

Mutations in rpoS in the WAT ancestor were only in the LE2 culture, although this could also be due to culture-to-culture variation. In this ancestor, fewer RpoS-associated staining differences were seen than with the other ancestor cultures in the previously studied replicate cultures [2]. The epistatic context of rpoS probably differs from the other natural isolates, consistent with an extremely diverse regulatory environment in natural isolates [49, 50].

WAT isolates also acquired an ancestor-specific domestication mutation, this time in metG (Table 1). Three different metG mutations were present in the LE1 and LE2 environments. So WAT seems to benefit from some change in metG to adapt to the laboratory. This is surprising because metG is considered to be an essential gene and encodes a methionyl-tRNA synthetase [42]. As with the other strain-specific changes, it is not clear how the mutations impart a strain-specific benefit.

DISCUSSION

The transition of four different isolates of E. coli from nature to laboratory culture resulted in phenotypic changes [2]. These changes were underpinned by one or more mutations in E. coli isolates that spent under 140 generations in the laboratory. So it is a rule rather than an exception that the transfer to the laboratory from nature results in changes, often multiple changes, in bacteria.

The commonest target of domestication affecting all natural ancestors was rpoS; one other E. coli strain isolated recently from nature [51] also acquired an rpoS mutation in the laboratory. The phenotypes in the isolates also changed in line with a loss of RpoS function in terms of iodine and rdar (Congo Red) staining [2]. Changes in rpoS as a function of the laboratory environment were already documented in earlier studies [52]; the explanation is the removal of the expensive resource allocation to stress responses, hence improving growth [29, 53]. Mutations in rpoS are also selected in stationary phase [23], nutrient limitation is recognized as a selection condition for loss of RpoS [52, 54] and stressful environments select for increased RpoS levels [55]. This instability in rpoS and resulting differences in RpoS level is common in natural isolates [49] and increases after periods in laboratory collections and storage [56]. Instability in RpoS levels is also seen in nature in pathogen populations [57]. Deleterious mutational changes in rpoS result from a selection environment in which fitness does not need effective stress resistance through RpoS.

In the unstressed laboratory environment, there is a greater benefit in the allocation of transcriptional resources to vegetative and metabolic genes.

As expected, the old-established stock K-12 strain MG1655 was less affected by laboratory culture than isolates without previous exposure to the laboratory. The four freshly isolated strains acquired more mutations, but the number of mutations seemed to be strain and culture specific. Mutations were present in both LE1 and LE2 environments and the targets of selection seemed to overlap. The ancestor-specific mutations (in corA, ycgK and metG in MEM, PAR and WAT, respectively) and rpoS mutations were present in both LE1 and LE2 cultures and so were presumably generally beneficial in the laboratory environment.
The number of acquired mutations over 10 days did not correlate with the mutation frequencies of the ancestors of the laboratory-adapted strains in Fig. 3, which were largely comparable. There is also a deviation between the LE1 and LE2 conditions, probably because the extended stationary phase is more stressful and leads to expression of error-prone polymerases that increase mutation rate [22]. The stressful LE1 condition also prompted the secondary selection of mutator mutations and an explanation of the acquired mutation numbers was the presence of mutations in mutL, affecting the mismatch system of DNA repair in two backgrounds. Mutations in mutL result in a mutator phenotype [44]. Such mutator mutations have been noted in populations, both in the laboratory and in the environment, where multiple mutations are needed in a stressful condition [58]. The mutator mutations spread through hitchhiking with beneficial mutations [59, 60], suggesting that the domestication process involves a strong selective environment for laboratory-beneficial mutations. A second possible explanation of the differing numbers of mutations is that strains like GE3 and MEM in the LE2 condition (and indeed MG1655) are already more adapted to the laboratory environment, so fewer mutations are selected.

The limited sampling in this study prevents definitive conclusions on the universality of domestication changes or systematic differences between the LE1 and LE2 environments. The various patterns of phenotypic change in cultures of the four natural ancestors [2] are also not fully consistent with the genotyping; for example, several differences detected in Biolog phenotyping were not obviously due to identifiable metabolic mutations in this study. Conversely, some similar phenotypic changes [2] were due to different mutations. Most notably, the phenotype ascribed to rpoS change (reduced iodine and Congo Red staining) in PAR LE1 occurred in the absence of an rpoS mutation. Hence, there is no totally predictable means of identifying domesticated strains by their phenotypes and there were unique elements to the mutational changes of each of the four natural strains studied.

In conclusion, the transfer of a bacterium to a laboratory medium and standard growth environments provides a selection condition for mutations and changes in properties. Interestingly, taxonomic relatedness is a poor indicator of domestication effects and the strains GE3 and WAT that were the closest taxonomically (Fig. 1) show considerable differences in domestication changes. Importantly, from a purely practical point of view, these results reinforce the importance of keeping bacteria in the laboratory without handling cultures for extended periods before making stocks or experiments. This is especially critical for the comparison of culture collections in different laboratories.

Funding information
The project was supported by the U.S. Army Research Office Grant W911NF1010045 and FA5002-16-P-0098 ITC-PAC Fund Support to T.F., a National 973 Program of China Grant (2013CB733904) to L.F., a National Natural Science Foundation of China Key Program Grant (31530083) to L.W. and National Natural Science Foundation of China General Program Grants (31270133 and 31470194) to L.F.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References