ARTICLE

doi:10.1038/nature11514

Genomic analysis of a key innovation in an experimental *Escherichia coli* population

Zachary D. Blount[1,2], Jeffrey E. Barrick[2,3,4], Carla J. Davidson[5] & Richard E. Lenski[1,2]

Evolutionary novelties have been important in the history of life, but their origins are usually difficult to examine in detail. We previously described the evolution of a novel trait, aerobic citrate utilization (Cit$^+$), in an experimental population of *Escherichia coli*. Here we analyse genome sequences to investigate the history and genetic basis of this trait. At least three distinct clades coexisted for more than 10,000 generations before its emergence. The Cit$^+$ trait originated in one clade by a tandem duplication that captured an aerobically expressed promoter for the expression of a previously silent citrate transporter. The clades varied in their propensity to evolve this novel trait, although genotypes able to do so existed in all three clades, implying that multiple potentiating mutations arose during the population's history. Our findings illustrate the importance of promoter capture and altered gene regulation in mediating the exaptation events that often underlie evolutionary innovations.

Evolutionary novelties are qualitatively new traits that open up ecological opportunities and thereby promote diversification[1,2]. These traits are thought to arise typically by the exaptation of genes that previously encoded other functions[2–6] through processes such as domain shuffling[7], altered regulation[8] and duplication followed by neo-functionalization[9,10]. Multiple mutations may be necessary to produce the new function[9,11], and thus its potential to evolve may be contingent on subtle differences between species, populations or even genotypes. A complete understanding of the evolution of a novel trait requires explanation of its ecological function, its physiological basis, the underlying mutations and the history of the accumulated changes[2].

Evolution experiments with microorganisms offer unparalleled opportunities to assess the evolution of novel traits[12–15]. Microbes have rapid generations and large populations, and new technologies allow discovery of mutations throughout their genomes[16–19]. Samples can be frozen and revived, allowing phylogenies and mutational histories to be constructed and analysed[18,19].

Twelve populations of *Escherichia coli* have been propagated in the long-term evolution experiment (LTEE) for over 40,000 generations in a glucose-limited minimal medium[13]. The medium also contains abundant citrate, which is present as a chelating agent[20], but *E. coli* cannot exploit citrate as a carbon and energy source in the well-aerated conditions of the experiment[21,22]. The inability to grow aerobically on citrate is a long-recognized trait that, in part, defines *E. coli* as a species[23]. Spontaneous citrate-using (Cit$^+$) mutants are extraordinarily rare[24], but a Cit$^+$ variant evolved in one population (designated Ara–3) around 31,000 generations[20]. Cit$^+$ cells became dominant after 33,000 generations, although Cit$^-$ cells persisted. This shift was accompanied by a several-fold increase in total population size owing to the high concentration of citrate relative to glucose in the medium.

The emergence of the Cit$^+$ trait was contingent upon one or more earlier mutations in the population's history. When evolution was replayed from clones isolated at various time points, clones from later generations were more likely to produce Cit$^+$ mutants than did the ancestor and other early clones[20]. This finding implied that a genetic background evolved that 'potentiated' the evolution of this trait. In principle, this effect could involve two distinct mechanisms. One possibility is that the rate of mutation, or certain types of mutation, increased such that the required event was more likely to occur in later generations. The other possibility is an epistatic interaction, such that the expression of a mutation that produced the Cit$^+$ trait required one or more preceding mutations. Here we report the results of extensive whole-genome re-sequencing that allowed us to reconstruct the population's history, identify mutations underlying the Cit$^+$ phenotype, and elucidate the physiological basis of this novel trait.

Genome sequences and phylogeny

We sequenced 29 clones sampled at various generations from population Ara–3, including nine Cit$^+$ and three Cit$^-$ clones known to be potentiated[20] (Supplementary Table 1). Mutations including single-nucleotide polymorphisms (SNPs) as well as deletions, insertions and some chromosomal rearrangements were identified (Supplementary Table 2) by comparing reads to the genome of the ancestral strain, REL606 (ref. 25). Inversions and other rearrangements that involved long sequence repeats would have escaped detection.

We reconstructed the population's phylogenetic history from presence–absence matrices of all mutations identified in the sequenced genomes. Figure 1 shows that the population was polymorphic for most of its history. Several clades, each represented by multiple clones, arose before 20,000 generations. One of them, which we call UC (unsuccessful clade), was not seen after 15,000 generations. Clades C1, C2 and C3 coexisted through the evolution of Cit$^+$ and beyond. C1 includes four clones, the earliest from generation 25,000, although a molecular clock (Fig. 1 inset) implies that C1 diverged from the ancestor of C2 and C3 before 15,000 generations. C2 and C3 had diverged by generation 20,000. C3 includes both Cit$^-$ clones and all the Cit$^+$ clones, although the first Cit$^+$ cells did not arise until ~31,000 generations. Two mechanisms may explain the prolonged coexistence of the Cit$^-$ lineages. First, they may all have acquired beneficial mutations without one gaining enough advantage to fix[26]. Alternatively, they might have filled subtly different niches, such as through the differential use of secreted metabolites[27,28].

[1]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824, USA. [2]BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, Michigan 48824, USA. [3]Department of Chemistry and Biochemistry, The University of Texas, Austin, Texas 78712, USA. [4]Institute for Cellular and Molecular Biology, The University of Texas, Austin, Texas 78712, USA. [5]Department of Microbiology, Immunology, and Infectious Diseases, University of Calgary, Calgary, Alberta T2N 4N1, Canada.

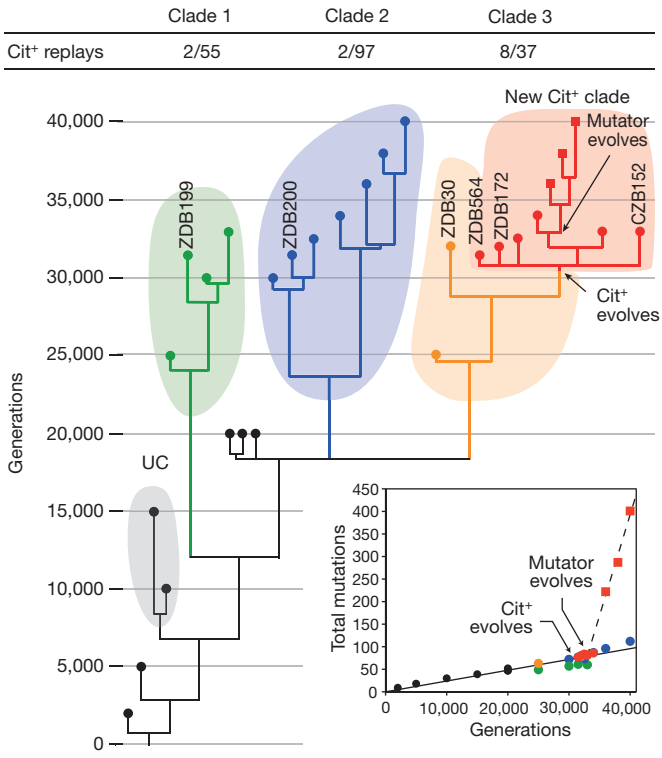

Figure 1 | Phylogeny of Ara–3 population. Symbols at branch tips mark 29 sequenced clones; labels are shown for clones mentioned in main text and figures. Shaded areas and coloured symbols identify major clades. Fractions above the tree show the number of clones belonging to the clade that yielded Cit$^+$ mutants during replay experiments (numerator) and the corresponding total used in those experiments (denominator). Inset shows number of mutations relative to the ancestor. The solid line is the least-squares linear regression of mutations in non-mutator genomes; the dashed line is the corresponding regression for mutator genomes.

The sequenced Cit$^+$ clones from generation 36,000 and later have an SNP in the *mutS* gene that produces a premature stop codon and truncates the MutS protein, thereby causing a defect in methyl-directed mismatch DNA repair[29]. We sequenced *mutS* from three Cit$^+$ clones from each of generations 33,000, 34,000, 35,000, 36,000, 37,000 and 38,000, and found this SNP in 0, 0, 2, 3 and 3 clones, respectively, indicating that the mutation arose after the origin of the Cit$^+$ lineage. Also, none of the 20 Cit$^-$ genomes has this mutation. As a consequence, Cit$^+$ genomes from later generations accumulated SNPs much faster than Cit$^-$ and early Cit$^+$ genomes (Fig. 1 inset). Mutator phenotypes evolved in several other populations in the LTEE[18,30], so this change was not unique to the Cit$^+$ lineage.

Evolution of the Cit$^+$ function

The evolution of the Cit$^+$ trait involved three successive processes: potentiation, actualization and refinement. The ancestor's rate of mutation to Cit$^+$ was immeasurably low, with an upper bound of 3.6×10^{-13} per cell-generation[20]. Potentiation refers to the evolution of a genetic background in which this function became accessible by mutation. An extremely weak Cit$^+$ variant had emerged by 31,500 generations, which represents the actualization step. The new function was then refined, which allowed the efficient exploitation of citrate, the rise of the Cit$^+$ subpopulation to numerical dominance, and the expansion of the total population size. Of these processes, actualization is the most tractable for study owing to the discrete phenotypic change, and we therefore focused on identifying and characterizing the mutational basis of actualization. With that information in hand, additional analyses then shed new light on the processes of refinement and potentiation.

Actualization of the Cit$^+$ function

One reason that *E. coli* cannot grow aerobically on citrate is its inability to transport citrate[24,31,32]. The origin of the Cit$^+$ phenotype therefore required expression of a citrate transporter. All nine sequenced Cit$^+$ genomes have two or more tandem copies of a 2,933-base-pair segment that includes part of the citrate fermentation (*cit*) operon (Fig. 2a). The amplified segment contains two genes: *rna*, which encodes RNase I[33], and *citT*, which encodes a broad-spectrum C_4-di- and tri-carboxylic acid transporter that functions in fermentation as a citrate/succinate antiporter[32]. The boundary upstream of *citT* is in the 3′ end of the *citG* gene, which encodes triphosphoribosyl-dephospho-CoA synthase, whereas the boundary downstream of *rna* is in the 5′ end of *rnk*, which encodes a regulator of nucleoside diphosphate kinase[34]. This amplification is not present in the ancestor or any of the sequenced Cit$^-$ genomes, and it is only found in population samples after the evolution of the Cit$^+$ lineage (Supplementary Table 3). Polymerase chain reaction (PCR) screens also failed to detect this segment in 27 Cit$^-$ clones from generations 33,000 through 40,000, whereas it was found in all 33 Cit$^+$ clones from the same generations (Supplementary Table 4).

Amplification mutations can alter the spatial relationship between structural genes and regulatory elements, potentially causing altered regulation and novel traits[35–38]. The structure of the *cit* amplification led us to propose that the Cit$^+$ trait arose from an amplification-mediated promoter capture (Fig. 2b and Supplementary Fig. 1). The amplification joined upstream *rnk* and downstream *citG* fragments, producing an *rnk-citG* hybrid gene expressed from the upstream *rnk* promoter. Because the *citT* and *citG* genes are normally monocistronic, the downstream copy of *citT* should therefore be co-transcribed with the hybrid gene. If the *rnk* promoter directs transcription under oxic conditions, then the new *rnk-citT* regulatory module might allow CitT expression during aerobic metabolism and thereby confer a Cit$^+$ phenotype[32].

To test this hypothesis, we first examined the capacity of the *rnk-citT* module to support *citT* expression in oxic conditions. We constructed a low-copy (one to two per cell) plasmid, pCDrnk-citTlux, with an *rnk-citT* module in which *citT* was replaced by the *luxCDABE* reporter operon. We made two other plasmids, pCDrnklux and pCDcitTlux, in which the reporter was under the control of the native upstream regulatory regions of *rnk* and *citT*, respectively. We transformed each plasmid into REL606, the ancestral strain; ZDB30, a potentiated C3 clone from generation 32,000; and ZDB172, a weakly Cit$^+$ clone from generation 32,000. We measured expression (light production) during growth and stationary phase under oxic conditions (Fig. 3). The native *citT* regulatory region showed no expression (above background) in any strain, indicating that *citT* is normally silent under oxic conditions. The native *rnk* regulatory region was expressed in all three strains, with a peak around the transition into stationary phase. Expression from the evolved *rnk-citT* module was much weaker, but there were small spikes in expression in ZDB30 and ZDB172 coincident with peak expression from the native *rnk* regulatory region. These results indicate that the *rnk-citT* module can support *citT* expression during aerobic metabolism.

This hypothesis also predicts that the introduction of the *rnk-citT* module should confer a Cit$^+$ phenotype, whereas the loss of the *cit*

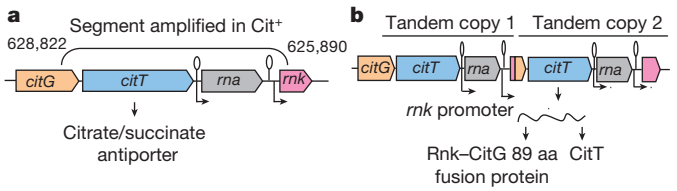

Figure 2 | Tandem amplification in Cit$^+$ genomes. **a**, Ancestral arrangement of *citG*, *citT*, *rna* and *rnk* genes. **b**, Altered spatial and regulatory relationships generated by the amplification.

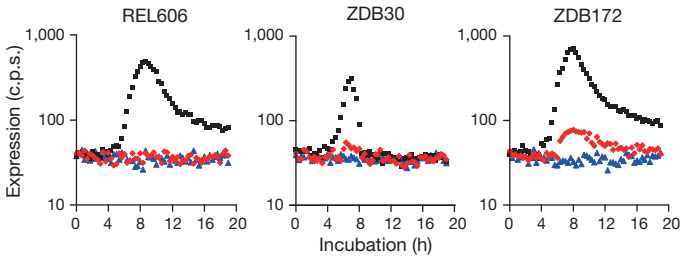

Figure 3 | Expression levels from native *citT*, native *rnk* and evolved *rnk-citT* regulatory regions during aerobic metabolism. Average time course of expression for the ancestral strain REL606 (left) and evolved clones ZDB30 (centre) and ZDB172 (right), each transformed with *lux* reporter plasmids pCDcitTlux (blue), pCDrnklux (black) and pCDrnk-citTlux (red). Expression was measured as counts per second (c.p.s.). Each curve shows the average of four replicates.

amplification should cause reversion to a Cit$^-$ state. We tested whether the *rnk-citT* module confers a Cit$^+$ phenotype by inserting a single copy of an amplification fragment containing the *rnk* promoter immediately upstream of the chromosomal copy of *citT* (Fig. 4a) in potentiated clone ZDB30. This construct, ZDB595, indeed has a Cit$^+$ phenotype, although it is extremely weak, similar to the earliest evolved Cit$^+$ variants (Fig. 4b). In the same medium used in the LTEE, ZDB595 experienced a lag of more than 60 h after glucose depletion, followed by a short period of abortive and inconsistent growth on citrate (Fig. 4b, c). These data indicate that the initial Cit$^+$ type was too weak to allow exploitation of citrate under the daily-transfer regime of the LTEE. Nonetheless, ZDB595 had a small (1.0%) but significant competitive advantage over ZDB30 in the same environment ($n = 10$, $t = 3.09$, two-tailed $P = 0.0128$). We also isolated 13 Cit$^-$ revertants of a 33,000-generation Cit$^+$ clone, CZB154. All had lost the *cit* amplification based on both PCR and Southern blot analyses (Supplementary Fig. 2), further supporting the hypothesis that the amplification event had produced the Cit$^+$ phenotype.

Refinement of the Cit$^+$ function

Given the extremely weak initial Cit$^+$ phenotype, additional mutations must have refined the new function. Refinement is an open-ended process, and we focus only on early refining mutations that led to the rise of Cit$^+$ cells to high frequency and the concurrent expansion of the population. These mutations are evidenced by the improvement on citrate of Cit$^+$ clones isolated between 31,500 and 33,000 generations (Supplementary Fig. 3). We examined the genomes of the five sequenced Cit$^+$ clones from before the population expansion, including one from generations 31,500, 32,000 and 32,500 and two from generation 33,000. We focused on the mutations on the line of descent for the Cit$^+$ subpopulation, where that line is defined by the presence of the same mutations in clones from generations 34,000, 36,000 and 38,000 (Supplementary Tables 5–8). Two SNPs and one insertion sequence (IS) element were present in only one of the 33,000-generation genomes. The IS insertion and one SNP seem unrelated to growth on citrate. The remaining SNP is in the regulatory region of *dctA*, which encodes a transporter of succinate and other C_4-di-carboxylic acids[39]. This mutation may improve recovery of succinate exported in exchange for citrate. However, the other 33,000-generation clone grew better on citrate (Supplementary Fig. 3). Thus, the *dctA* mutation does not seem to be responsible for the population expansion, although it might have been advantageous to its carriers.

These early Cit$^+$ genomes also show increases in *cit* copy number. The earliest one had a tandem duplication, whereas later genomes had a three-copy tandem array within a larger tandem duplication, a four-copy tandem array, a tandem duplication in a larger three-copy tandem array and a nine-copy tandem array (Supplementary Table 9). Changes in amplification copy number readily occur by recombination, and they have been implicated in the refinement of other weak functions[40,41]. These changes increased the number of *rnk-citT* modules relative to the earliest Cit$^+$ genome (Fig. 5a) and presumably increased the expression of CitT as well. To test whether an increased number of *rnk-citT* modules could have caused the population expansion, we cloned the module (Fig. 5b) into the high-copy plasmid pUC19 (ref. 42) and moved the resulting plasmid, pZBrnk-citT, into the potentiated clone ZDB30. The resulting strain, ZDB612, is strongly Cit$^+$, rapidly transitions from glucose to citrate and grows similarly to the 33,000-generation clone CZB152 (Fig. 5c). The increased number of *rnk-citT* modules can thus explain the refinement of the Cit$^+$ phenotype that allowed the population expansion.

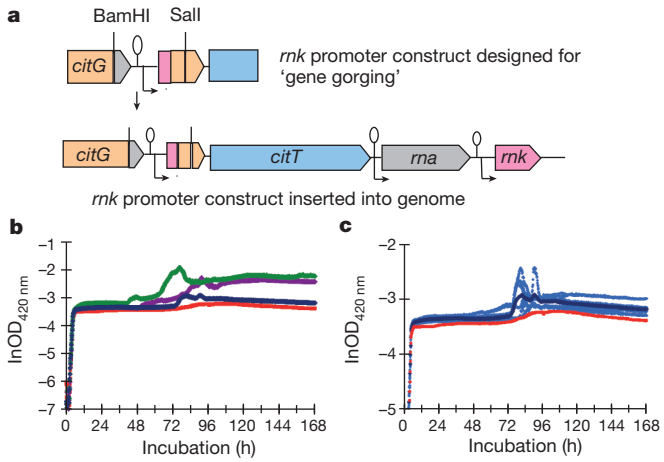

Figure 4 | New *rnk-citT* module confers Cit$^+$ phenotype in potentiated background. a, Engineered construct containing the *cit* amplification junction with the *rnk* promoter, and the arrangement following its insertion into the chromosome. **b**, Average growth trajectories in DM25 of potentiated clone ZDB30 (red), its isogenic construct with chromosomal integration of the *rnk-cit* module ZDB595 (blue), 31,500-generation Cit$^+$ clone ZDB564 (purple) and 32,000-generation Cit$^+$ clone ZDB172 (green). **c**, Trajectories of ZDB595 compared to its parent. Light blue trajectories show heterogeneity among replicate cultures of ZDB595; dark blue and red are averages for ZDB595 and its parent, ZDB30, as in panel **b** (except different scale).

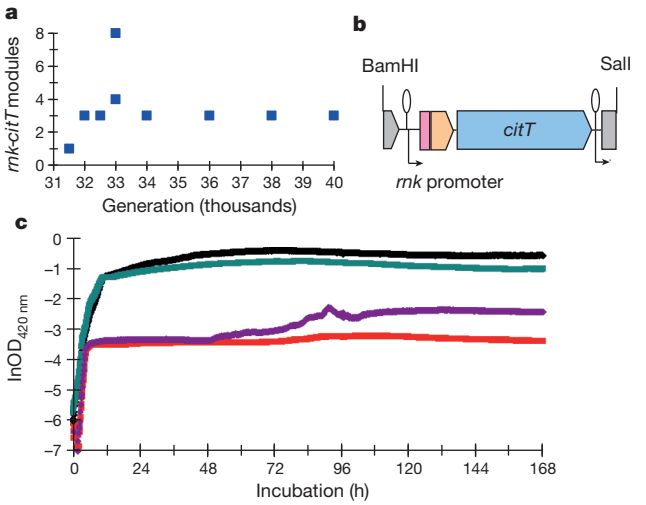

Figure 5 | Refinement of Cit$^+$ phenotype by increased number of *rnk-citT* modules. a, Change in *rnk-citT* module copy number in sequenced Cit$^+$ clones over time. **b**, Structure of *rnk-citT* module cloned into high-copy plasmid pUC19 to produce pZBrnk-citT. **c**, Growth trajectories in DM25 of potentiated Cit$^-$ clone ZDB30 (red), 31,500-generation Cit$^+$ clone ZDB564 (purple), 33,000-generation Cit$^+$ clone CZB152 (black), and ZDB612, a pZBrnk-citT transformant of ZDB30 (green). Each trajectory is the average of six replicates.

In contrast to the early variation in *cit* amplification, later Cit$^+$ genomes have four-copy tandem arrays (Fig. 5a). Amplifications tend to be unstable[40,41], and further refinement may have favoured more stable classes of mutations. The evolution of the mutator phenotype in the Cit$^+$ lineage complicates efforts to identify these later refining mutations, but some interesting candidates include SNPs in *citT* itself: *gltA*, which encodes citrate synthase; and *aceA*, which encodes isocitrate lyase.

Potentiation of Cit$^+$ evolution

Before the Cit$^+$ trait could evolve, the Ara–3 population had to evolve a genetic background in which that new function was accessible by mutation. Potentiation was demonstrated by 'replay' experiments using 270 clones sampled over the population's history[20]. The replays produced 17 Cit$^+$ mutants that derived from 13 clones, all from generation 20,000 or later. Fluctuation tests confirmed that potentiated clones had increased mutation rates to Cit$^+$, although such mutations were still extremely rare[20].

Phylogeny implies multiple potentiating mutations

The potentiating mutations are not known to confer any phenotype amenable to screening, so there is no simple way to distinguish between potentiated and non-potentiated clones. Instead, we examined the distribution of the 13 potentiated clones identified by the replay experiments using mutations (Supplementary Table 10) that differentiated clades UC, C1, C2 and C3 (Fig. 1). We also determined the distribution of the other 256 evolved clones used in the replays to assess the coverage of the clades in those experiments (Supplementary Fig. 4). Overall, 205 clones were assigned to clades, including 12 potentiated clones (Supplementary Table 11). Sixteen from generations 15,000 and earlier were in clade UC. The others came from generations 20,000 and later including 55 in C1, 97 in C2, and 37 in C3. Potentiated clones occurred in all three with 8 in C3 and 2 each in C1 and C2. Nonetheless, this distribution is highly non-random (two-tailed Fisher's exact test comparing C3 with C1 and C2 combined, $P = 0.0003$; this test shows no difference between C1 and C2, $P = 0.6206$). These data, and the absence of any Cit$^+$ mutants generated by the ancestor, indicate that potentiation involved at least two mutations, with one arising before these three clades diverged and another in C3 (Fig. 1).

Alternative hypotheses

Two distinct mechanisms might explain the potentiation effect. One is epistasis, whereby an interaction between the potentiating background and the actualizing mutation is needed to express the Cit$^+$ phenotype. The second is that the background physically promoted the final mutation; for example, a later rearrangement may require some prior genomic rearrangement. If expression of Cit$^+$ required earlier mutations, then the *rnk-citT* module should confer a weaker Cit$^+$ phenotype in a non-potentiated background than a potentiated background. Alternatively, if potentiation facilitated the amplification event itself, then that module should produce an equally strong Cit$^+$ phenotype in both potentiated and non-potentiated backgrounds.

Evidence for epistasis

To test these predictions, we first tried to move the single-copy *rnk-citT* module into the ancestral chromosome, but several attempts were unsuccessful. This outcome is not surprising under the epistasis hypothesis; a key step was screening potential recombinants for citrate use, and the single-copy module conferred an extremely weak Cit$^+$ phenotype even in a potentiated C3 clone (Fig. 4). Instead, we moved plasmid pZBrnk-citT into the ancestor and clones from clades C1, C2 and C3, and examined their growth trajectories (Fig. 6). All four transformants grew on citrate after depleting the glucose. However, transformants of the ancestor and C1 and C2 clones grew poorly on citrate, even with this high-copy plasmid, as shown by long lags while

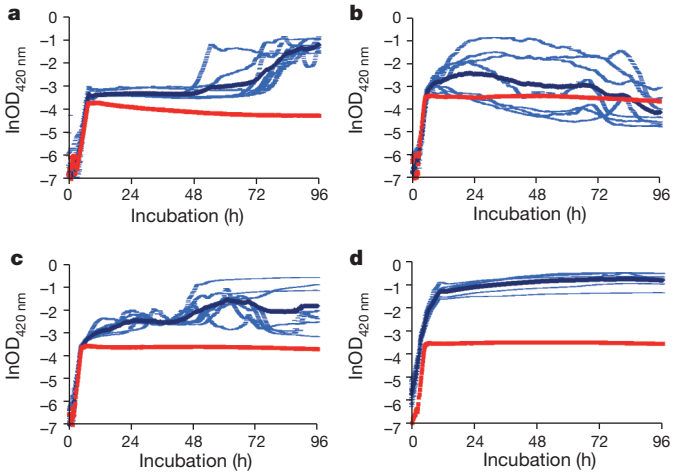

Figure 6 | Evidence for epistatic interactions in potentiation of Cit$^+$ phenotype. Growth trajectories in DM25 of diverse clones transformed with the pZBrnk-citT plasmid. **a**, Ancestral strain REL606 and its transformant ZDB611. **b**, Clade C1 clone ZDB199 and its transformant ZDB614. **c**, C2 clone ZDB200 and its transformant ZDB615. **d**, C3 clone ZDB30 and its transformant ZDB612. Red and dark blue trajectories are averages for the parent clone and its transformant; light blue trajectories show the replicates for the transformant.

transitioning from glucose to citrate, low yields, and inconsistent trajectories across replicates (Fig. 6a–c). By contrast, ZDB612, the transformant of a potentiated C3 clone, grew much faster, more extensively, and consistently across replicates (Fig. 6d). These differences demonstrate epistatic interactions between the *rnk-citT* module and mutations that distinguish the backgrounds.

These data also support the phylogenetic association between clade C3 and the strength of potentiation in the replay experiments. We examined C3 for candidate mutations that may contribute to potentiation. A mutation in *arcB*, which encodes a histidine kinase[43], is noteworthy because disabling that gene upregulates the tricarboxylic acid cycle[44]. That mutation might interact with the *rnk-citT* module by allowing efficient use of citrate that enters the cell via the CitT transporter. We tried repeatedly to move the evolved and ancestral *arcB* alleles between strains to test this hypothesis, but without success. In any case, the profound differences in growth trajectories on citrate enabled by the high-copy plasmid (Fig. 6) support the hypothesis that potentiation depends, at least partly, on epistasis between the genetic background and the amplification mutation that generated that module.

Evidence against the physical-promotion hypothesis

We examined the Cit$^+$ mutants from the replay experiments[20] for additional evidence on the nature of potentiation. The physical-promotion hypothesis predicts that these mutants should have *cit* amplifications similar or identical to the original one. If epistatic interactions enhanced *citT* expression only from the *rnk* promoter, then the prediction would be the same. However, if epistasis operated at some broader physiological level, then the replays should have diverse mutations that share only the property that they enable expression of the citrate transporter in the oxic environment of the LTEE. We examined 19 re-evolved Cit$^+$ mutants to identify the relevant mutations; the *citT* region was examined in all of them, and the genomes of six were sequenced. All have mutations affecting *citT*, and most clearly put that gene downstream of a new promoter (Supplementary Table 12). Four sequenced Cit$^+$ mutants were derived from Cit$^-$ clones that were also sequenced. Besides *citT*-related mutations, these sequenced mutants had one to three other mutations; no gene was mutated in multiple cases, and none seem related to citrate use (Supplementary Table 13), supporting the

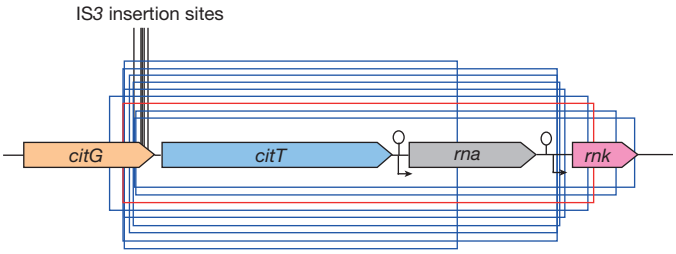

Figure 7 | Mutations that produced Cit$^+$ phenotype in 14 replay experiments. The red box shows the boundaries of the 2,933-bp amplified segment that actualized the Cit$^+$ function in the original long-term population. Blue boxes show *citT*-containing regions amplified in eight replays that produced Cit$^+$ mutants. Vertical black lines mark five locations where IS*3* insertions produced Cit$^+$ mutants in six other replays.

inference that the *citT* mutations were responsible for all of the re-evolved Cit$^+$ phenotypes.

The Cit$^+$ mutants arose by diverse mutational processes (Supplementary Table 12). Eight have *citT* duplications similar to the original one, though no two share the same boundaries (Fig. 7). In seven of these, the duplications generated alternative versions of the *rnk-citT* module; in the other, the second *citT* is downstream of the *rna* promoter. Six mutants have an IS*3* element inserted in the 3′ end of *citG* (Fig. 7). IS*3* carries outward-directed promoter elements that can activate adjacent genes[27,45]. Two mutants have large duplications encompassing all or part of the *cit* operon. One mutant has a large inversion that places most of that operon downstream of the promoter for the fimbria regulatory gene *fimB*, and another has a deletion in *citG* that presumably formed a new promoter. Also, most of these mutants have stronger phenotypes (Supplementary Fig. 5) than the earliest Cit$^+$ clones in the main experiment (Fig. 4b and Supplementary Fig. 3). In any case, this new function arose in potentiated backgrounds by a variety of mutational processes that recruited several different promoters to allow CitT expression during aerobic metabolism. Thus, these data do not support the physical-promotion hypothesis, whereas the strain-specific differences in growth on citrate conferred by the *rnk-citT* module provide clear and compelling evidence for epistasis (Fig. 6). However, these hypotheses are not mutually exclusive, and we cannot reject the possibility that some mutation rendered the genome (or the affected region) more prone to physical rearrangements (including mobile-element insertions) and thereby also contributed to the overall potentiation effect.

Perspective

The evolution of citrate use in an experimental *E. coli* population provided an unusual opportunity to study the multi-step origin of a key innovation. Comparative studies have shown that gene duplications have an important creative role in evolution by generating redundancies that allow neo-functionalization[5,6,8–10]. Our findings highlight the less-appreciated capacity of duplications to produce new functions by promoter capture events that change gene regulatory networks[38]. The evolution of citrate use also highlights that such actualizing mutations are only part of the process by which novelties arise. Before a new function can arise, it may be essential for a lineage to evolve a potentiating genetic background that allows the actualizing mutation to occur or the new function to be expressed. Finally, novel functions often emerge in rudimentary forms that must be refined to exploit the ecological opportunities. This three-step process—in which potentiation makes a trait possible, actualization makes the trait manifest, and refinement makes it effective—is probably typical of many new functions.

METHODS SUMMARY

Evolution experiment. Twelve *E. coli* B populations were propagated for over 20 years by daily 100-fold transfers into a minimal medium, DM25, containing 139 μM glucose and 1,700 μM citrate[13,46]. Population samples were frozen every 500 generations. The focus of this study was population Ara–3, in which the ability to grow on citrate under oxic conditions evolved by 31,500 generations[20].

Genome sequencing. Clones were sequenced on Illumina GA, GA II and GA IIx instruments. The reads were deposited in the NCBI SRA database (SRA026813). Mutations were predicted from read alignments to the ancestral genome (REL606)[25] using the *breseq* pipeline[47] and refined by further analyses for specific mutations.

Phylogenetic analysis. The most parsimonious tree was calculated using the irreversible Camin–Sokal model[48] with presence-absence data in each genome for all mutational events. Branch lengths were calculated by a maximum likelihood procedure with additional parameters for ancestral and mutator rates of Poisson-distributed mutation accumulation. Other clones were put in clades based on Sanger sequencing to screen for 10 phylogenetically informative mutations.

Expression experiments. Promoter regions for *citT*, *rnk* and the *rnk-citT* module were cloned into the *lux* reporter plasmid pCS26-pac[49], and transformed into strains REL606, ZDB30 and ZDB172. Light emission was measured as a proxy for luciferase expression for 19 h in DM25 in 96-well plates with intermittent shaking. Assays were replicated fourfold for each construct.

Primers and clones. Primer pairs used in genetic analyses and strain constructions are shown in Supplementary Table 14. Supplementary Table 15 describes the clones used in growth-trajectory experiments.

Full Methods and any associated references are available in the online version of the paper.

Received 18 June; accepted 22 August 2012.
Published online 19 September 2012.

1. Mayr, E. In *Evolution after Darwin* (ed. Tax, S.) Vol. 1, 349–380 (Univ. Chicago Press, 1960).
2. Pigliucci, M. What, if anything, is an evolutionary novelty? *Philos. Sci.* **75,** 887–898 (2008).
3. Jacob, F. Evolution and tinkering. *Science* **196,** 1161–1166 (1977).
4. Jacob, F. *The Possible and the Actual.* (Univ. Washington Press, 1982).
5. Gould, S. J. & Vrba, E. S. Exaptation—a missing term in the science of form. *Paleobiol.* **8,** 4–15 (1982).
6. Taylor, J. S. & Raes, J. Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* **38,** 615–643 (2004).
7. Patthy, L. Genome evolution and the evolution of exon-shuffling—a review. *Gene* **238,** 103–114 (1999).
8. True, J. R. & Carroll, S. B. Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* **18,** 53–80 (2002).
9. Zhang, J. Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18,** 292–298 (2003).
10. Bergthorsson, U., Andersson, D. I. & Roth, J. R. Ohno's dilemma: evolution of new genes under continuous selection. *Proc. Natl Acad. Sci. USA* **104,** 17004–17009 (2007).
11. Lenski, R. E., Ofria, C., Pennock, R. T. & Adami, C. The evolutionary origin of complex features. *Nature* **423,** 139–144 (2003).
12. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Rev. Genet.* **4,** 457–469 (2003).
13. Lenski, R. E. Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. *Plant Breed. Rev.* **24,** 225–265 (2004).
14. Beaumont, H. J. E., Gallie, J., Kost, K., Ferguson, G. C. & Rainey, P. B. Experimental evolution of bet hedging. *Nature* **462,** 90–93 (2009).
15. Meyer, J. R., Dobias, D. T., Weitz, J. S., Barrick, J. E. & Lenski, R. E. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **335,** 428–432 (2012).
16. Bentley, D. R. Whole-genome resequencing. *Curr. Opin. Genet. Dev.* **16,** 545–552 (2006).
17. Hegreness, M. & Kishony, R. Analysis of genetic systems using experimental evolution and whole-genome sequencing. *Genome Biol.* **8,** 201 (2007).
18. Barrick, J. E. *et al.* Genome evolution and adaptation in a long-term experiment with *E. coli*. *Nature* **461,** 1243–1247 (2009).
19. Barrick, J. E. & Lenski, R. E. Genome-wide mutational diversity in an evolving population of *Escherichia coli*. *Cold Spring Harb. Symp. Quant. Biol.* **74,** 119–129 (2009).
20. Blount, Z. D., Borland, C. Z. & Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **105,** 7899–7906 (2008).
21. Koser, S. A. Correlation of citrate-utilization by members of the colon-aerogenes group with other differential characteristics and with habitat. *J. Bacteriol.* **9,** 59–77 (1924).
22. Lütgens, M. & Gottschalk, G. Why a co-substrate is required for anaerobic growth of *Escherichia coli* on citrate. *J. Gen. Microbiol.* **119,** 63–70 (1980).
23. Scheutz, F. & Strockbine, N. A. In *Bergey's Manual of Systematic Bacteriology, Volume 2: The Proteobacteria* (eds Garrity, G. M., Brenner, D. J., Kreig, N. R. & Staley, J. R.) 607–624 (Springer, 2005).

24. Hall, B. G. Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. *J. Bacteriol.* **151,** 269–273 (1982).
25. Jeong, H. *et al.* Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J. Mol. Biol.* **394,** 644–652 (2009).
26. Fogle, C. A., Nagle, J. L. & Desai, M. M. Clonal interference, multiple mutations and adaptation in large asexual populations. *Genetics* **180,** 2163–2173 (2008).
27. Treves, D. S., Manning, S. & Adams, J. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Mol. Biol. Evol.* **15,** 789–797 (1998).
28. Rozen, D. E., Schneider, D. & Lenski, R. E. Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. *J. Mol. Evol.* **61,** 171–180 (2005).
29. Glickman, B. W. & Radman, M. *Escherichia coli* mutator mutants deficient in methyl-instructed DNA mismatch correction. *Proc. Natl Acad. Sci. USA* **77,** 1063–1067 (1980).
30. Sniegowski, P. D., Gerrish, P. J. & Lenski, R. E. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **387,** 703–705 (1997).
31. Lara, F. J. S. & Stokes, J. L. Oxidation of citrate by *Escherichia coli*. *J. Bacteriol.* **63,** 415–420 (1952).
32. Pos, K. M., Dimroth, P. & Bott, M. The *Escherichia coli* citrate carrier CitT: a member of a novel eubacterial transporter family related to the 2-oxoglutarate/malate translocator from spinach chloroplasts. *J. Bacteriol.* **180,** 4160–4165 (1998).
33. Zhu, L. & Deutsher, M. P. The *Escherichia coli rna* gene encoding RNase I: sequence and unusual promoter structure. *Gene* **119,** 101–106 (1992).
34. Shankar, S., Schlictman, D. & Chakrabarty, A. M. Regulation of nucleoside diphosphate kinase and an alternative kinase in *Escherichia coli*: role of the *sspA* and *rnk* genes in nucleoside triphosphate formation. *Mol. Microbiol.* **17,** 935–943 (1995).
35. Usakin, L. A., Kogan, G. L., Kalmykova, A. I. & Gvozdev, V. A. An alien promoter capture as a primary step of the evolution of testes-expressed repeats in the *Drosophila melanogaster* genome. *Mol. Biol. Evol.* **22,** 1555–1560 (2005).
36. Adam, D., Dimitrijevic, N. & Schartl, M. Tumor suppression in *Xiphophorus* by an accidentally acquired promoter. *Science* **259,** 816–819 (1993).
37. Bock, R. & Timmis, J. N. Reconstructing evolution: gene transfer from plastids to the nucleus. *Bioessays* **30,** 556–566 (2008).
38. Whoriskey, S. K., Nghiem, V., Leong, P., Masson, J. & Miller, J. H. Genetic rearrangements and gene amplification in *Escherichia coli*: DNA sequences at the junctures of amplified gene fusions. *Genes Dev.* **1,** 227–237 (1987).
39. Janausch, I. G., Zientz, E., Tran, Q. H., Kroger, A. & Unden, G. C_4-dicarboxylate carriers and sensors in bacteria. *Biochim. Biophys. Acta* **1553,** 39–56 (2002).
40. Andersson, D. I., Slechta, E. S. & Roth, J. R. Evidence that gene amplification underlies adaptive mutability of the bacterial *lac* operon. *Science* **282,** 1133–1135 (1998).
41. Reams, A. B., Kofoid, E., Savageau, M. & Roth, J. R. Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. *Genetics* **184,** 1077–1094 (2010).
42. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* **33,** 103–119 (1985).
43. Gunsalus, R. P. & Park, S. J. Aerobic-anaerobic gene regulation in *Escherichia coli*: control by the ArcAB and Fnr regulons. *Res. Microbiol.* **145,** 437–450 (1994).
44. Nizam, S. A., Zhu, J., Ho, P. Y. & Shimizu, K. Effects of *arcA* and *arcB* genes knockout on the metabolism in *Escherichia coli* under aerobic condition. *Biochem. Eng. J.* **44,** 240–250 (2009).
45. Charlier, D., Piette, J. & Glansdorff, N. IS3 can function as a mobile promoter. *Nucleic Acids Res.* **10,** 5935–5948 (1982).
46. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am. Nat.* **138,** 1315–1341 (1991).
47. Barrick, J. E. & Knoester, D. B. breseq. http://barricklab.org/breseq (2010).
48. Camin, J. H. & Sokal, R. R. A method for deducing branching sequences in phylogeny. *Evolution* **19,** 311–326 (1965).
49. Bjarnason, J., Southward, C. M. & Surette, M. G. Genomic profiling of iron-responsive genes in *Salmonella enterica* serovar Typhimurium by high-throughput screening of a random promoter library. *J. Bacteriol.* **185,** 4973–4982 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank N. Hajela, M. Kauth and S. Sleight for assistance, J. Meyer and J. Plucain for discussion, and C. Turner for comments on the paper. Sequencing services were provided by the MSU Research Technology Support Facility. We acknowledge support from the US National Science Foundation (DEB-1019989 to R.E.L.) including the BEACON Center for the Study of Evolution in Action (DBI-0939454), the US National Institutes of Health (K99-GM087550 to J.E.B.), the Defense Advanced Research Projects Agency (HR0011-09-1-0055 to R.E.L.), a Rudolf Hugh Fellowship (to Z.D.B.), a DuVall Family Award (to Z.D.B.), a Ronald M. and Sharon Rogowski Fellowship (to Z.D.B.) and a Barnett Rosenberg Fellowship (to Z.D.B).

Author Contributions J.E.B. and Z.D.B performed genome sequencing and analyses. J.E.B. performed phylogenetic analyses and developed code for sequence analyses. Z.D.B. performed growth curves, molecular experiments and sequenced specific genes. C.J.D. performed gene expression experiments. R.E.L. conceived and directs the long-term experiment. Z.D.B., J.E.B., C.J.D. and R.E.L. analysed data, wrote the paper and prepared figures.

Author Information All genome data have been deposited in the NCBI Sequence Read Archive database (SRA026813). Other data have been deposited in the DRYAD database (http://dx.doi.org/10.5061/dryad.8q6n4). R.E.L. will make strains available to qualified recipients, subject to a material transfer agreement that can be found at http://www.technologies.msu.edu/inventors/mta-cda/mta/mta-forms. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.D.B. (blountza@msu.edu) or R.E.L. (lenski@msu.edu).

METHODS

Evolution experiment. The long-term experiment is described in detail elsewhere[13,46]. In brief, twelve populations of *E. coli* B were started in 1988 and have evolved since under conditions of daily 100-fold dilutions in a minimal medium, DM25, containing 139 μM glucose and 1,700 μM citrate[46]. The populations undergo ~6.64 generations per day and had been evolving for 40,000 generations when this study began. Every 500 generations, samples were frozen at –80 °C with glycerol as a cryoprotectant. The focus of this study was population Ara–3, in which the ability to grow aerobically on citrate evolved by 31,500 generations[20].

Genomic DNA isolation. Clones were revived from frozen stocks by overnight growth in LB medium at 37 °C with aeration. DNA was extracted and purified using the Qiagen Genomic-tip 100/G kit (Qiagen).

Whole-genome sequencing and mutation detection. Clones were sequenced on Illumina GA, GA II and GA IIx instruments. The resulting reads were deposited in the NCBI SRA database (SRA026813). Most data sets contain single-end reads only, but additional mate-paired libraries were obtained for ZDB30 and ZDB172. Reads were mapped to the reference genome of the ancestral strain (REL606)[25], and mutations were predicted using the breseq computational pipeline[47]. This pipeline detects point mutations, deletions and new sequence junctions that may indicate IS-element insertions or other rearrangements, as described in its online documentation. Large duplications and amplifications were predicted manually by examining the depth of read coverage across each genome.

Mutation lists (Supplementary Table 2) were further refined by manually reconstructing the most plausible series of events generating the observed differences. This procedure involved: (1) splitting predicted changes into multiple mutations based on phylogenetic relationships, (2) assigning mutations to genomes when a subsequent mutational event prevented their detection (for example, an SNP within a later deletion), and (3) correcting false positives for a handful of difficult-to-predict mutations by examining alignments and coverage in all clones. After these procedures, 17 homoplasies remained in the phylogeny, most of which seem to indicate mutational hot spots: 9 are IS-element insertions at specific sites, 7 are insertions or deletions at the boundaries of IS elements, and only 1 is a single-base substitution not associated with an IS element. Given these signatures, it is likely that these mutations occurred independently in multiple lineages.

The copy number and configuration of the *citT* module were estimated from the number of reads overlapping the new *rnk-citT* sequence junction relative to the two original flanking-sequence junctions, and from the average read-depth in this region relative to a single-copy region (Supplementary Table 9).

Phylogenetic analysis. An initial parsimony-based tree was calculated using presence-absence data for all mutational events in each genome using the *dnapars* program from PHYLIP. Branch lengths were recalculated using an irreversible Camin–Sokal model[48], because the ancestral states are known and reversions unlikely given the genome size and number of mutations observed. A maximum-likelihood model was then used to estimate branching times and two mutation rates, one for non-mutator branches and one for mutator branches. The model fixed the generation when each clone was sampled and assumed that mutations accumulated on branches according to a Poisson process.

Ten mutations in nine genes (*ybaL*, *nadR*, *hemE*, *cspC*, *yaaH*, *leuA*, *tolR*, *arcB* and *gltA*) were identified as phylogenetically informative based on their association with particular clades in the Ara–3 population (Supplementary Table 10). Sanger sequencing of PCR-amplified gene fragments was used to determine the presence or absence of these mutations in replay clones, which were then mapped onto the phylogeny using the keys in Supplementary Fig. 4. Owing to the large number of clones and genes under consideration, not all genes were sequenced for all clones. Supplementary Table 11 shows the data and phylogenetic assignments. Supplementary Table 14 shows the primer pairs used to amplify each locus.

PCR screens for *cit* amplifications. The *cit* amplification was detected in population samples and clones by PCR amplification across the *rnk-citG* junction using outward-directed primers specific to *citT* (Supplementary Table 14). When screening population samples, three reactions were run for each time point, and the template was a 1:10 dilution of the frozen sample for that generation.

Expression experiments. Expression was measured using luciferase-based reporter constructs. The complete upstream regions for the native *citT* and *rnk* genes were PCR-amplified from the cognate reporters from the *E. coli* transcriptional library[50] using the primers pZE05 and pZE07. The intergenic region of the evolved *rnk-citT* module was amplified from Cit$^+$ clone ZDB172 using primers nctForward and nctReverse (Supplementary Table 14). The PCR products were cloned into the low-copy (1–2) plasmid pCS26-pac, which contains a kanamycin resistance gene and the luciferase operon (*luxCDABE*)[49]. Each plasmid was transformed into clones REL606, ZDB30 and ZDB172.

Before the expression assays, strains were grown in a 96-well plate (BD Biosciences) in 200 μl per well of DM25 supplemented with 50 μg ml^{-1} kanamycin, with constant shaking at 37 °C for two 24-h cycles for acclimation. Fresh overnight cultures were then diluted 100-fold into a black, clear-bottomed 96-well plate (9520 Costar, Corning), with 150 μl per well of DM25, and covered with a breathable sealing membrane (Nunc) to prevent evaporation. Light emission was measured using a Wallac Victor^2 plate reader (Perkin Elmer Life Sciences) every 20 min for 19 h with 90 s of 2-mm orbital shaking before each reading. Assays were run in quadruplicate in the same plate.

Isogenic strain construction. A single-copy chromosomal *rnk-citT* module was placed in a Cit$^-$ background using the 'gene gorging' method[51]. Owing to problems inherent to the manipulation of amplified genes, we did not attempt to move an entire *citT* amplification segment. Instead, we engineered a *cit* amplification junction containing the *rnk* promoter from three smaller fragments that were PCR-amplified from the 32,000-generation Cit$^+$ clone ZDB172. The first fragment contained the *cit* amplification junction (*cit*AmpJ), including the *rnk* promoter region, and was PCR-amplified using the primers *citT*AmpJ F and *citT*AmpJ R (Supplementary Table 14). The second fragment contained the *citT-citG* junction (*citT-citG*) and was PCR-amplified using the primers *citT-citG* F and *citT-citG* R. The third fragment contained sequence internal to the *citG* gene (*citG*frag) and was PCR-amplified using *citG*frag F and *citG*frag R. Each primer pair was designed with restriction sites that allowed ligation of the fragments to a hybrid construct in which the *cit* amplification junction, including the *rnk* promoter, was embedded within >500 bp of the *citG* flanking sequence. The assembled module was then PCR-amplified using *citT-citG* Gorge F and *citG*frag R (Supplementary Table 14). The forward primer incorporated an *I-SceI* restriction site required for the gene-gorging procedure, which was then performed as described elsewhere[51]. We screened for putative transformants by performing PCR and testing for a Cit$^+$ phenotype based on a positive reaction on Cristensen's Citrate Agar[52]. Successful constructs were confirmed by Sanger sequencing.

Growth trajectories. Strains of interest (Supplementary Table 15) were revived from frozen stocks by growing them in LB, and they were then acclimated by two 24-h culture cycles in DM25. Next, 100 μl of each culture was diluted into 9.9 ml of DM25, and 200 μl aliquots were placed into randomly assigned wells in a 96-well plate. For all pZBrnk-citT transformants, the medium was supplemented with 100 μg ml^{-1} ampicillin to ensure plasmid retention. Growth trajectories were replicated six- to eightfold for each strain. To limit evaporation, strains were grown in the innermost 60 wells, whereas the outermost 36 wells were filled with 300 μl of saline buffer. When assays ran longer than 96 h, the buffer was replenished after 96 h. $OD_{420\,nm}$ was measured every 10 min by a VersaMax automated plate reader (Molecular Devices). Plates were shaken orbitally for 5 s before each measurement, but were otherwise stationary.

Fitness assays. Relative fitness was measured in competition experiments described elsewhere[46]. In brief, we inoculated 0.05 ml each of acclimated cultures of ZDB595 and ZDB63, an Ara$^+$ mutant of ZDB30, into 9.9 ml of DM25 with tenfold replication. Initial densities were measured by dilution plating on tetrazolium arabinose (TA) plates, on which ZDB595 and ZDB63 made red and white colonies, respectively. The cultures were then propagated for three daily transfer cycles, and the final densities were measured using TA plates. Relative fitness was calculated as the ratio of the realized growth rates of the competitors over the course of the experiment[46].

Isolation of Cit$^-$ revertants. Three 10-ml LB broth cultures were inoculated with 15 μl from the frozen stock of 33,000-generation Cit$^+$ clone CZB154. After overnight growth at 37 °C, the cultures were diluted 10^6-fold, and 100 μl was transferred to 30 flasks, ten containing 9.9 ml of DM25, ten containing 9.9 ml of DM25 without citrate, and ten containing 9.9 ml of M9 broth (a citrate-free minimal medium) with 25 μg ml^{-1} glucose. The 30 cultures were incubated at 37 °C, and propagated through six daily 1:100 serial dilutions. Each culture was then diluted 10^4-fold, and 100 μl were spread onto each of eight TA plates. The plates were incubated for ~24 h at 37 °C, and colonies were patched onto TA and Minimal Citrate (MC) plates. Clones that grew on TA plates, but not on MC plates, were then streaked onto Christensen's agar. Those clones that did not produce a Cit$^+$ reaction on Christensen's agar were retained as Cit$^-$ revertants. Up to 512 colonies per replicate culture were tested for loss of the Cit$^+$ phenotype, but only one Cit$^-$ revertant was retained from any culture. Cit$^-$ revertants were found in 8/10 citrate-free DM25 cultures and in 5/10 cultures containing M9 medium. No revertants were isolated from any cultures in normal DM25 medium, presumably because of the advantage that Cit$^+$ cells have when citrate is present. The 13 Cit$^-$ revertants were tested for the presence or absence of the *cit* amplification by performing both PCR with the *citT*out F/R primer pair and *citT*-specific Southern blotting with EcoRV-digested genomic DNA (Supplementary Fig. 2 and Supplementary Table 14).

Plasmid construction. A DNA fragment containing the complete *rnk-citT* module was PCR-amplified from Cit$^+$ clone ZDB172 using primers *citT*AmpX F and *citT*AmpX R (Supplementary Table 14). The fragment was then inserted into the cloning site of pUC19 (NEB), and sequencing confirmed that the resulting plasmid, pZBrnk-citT, had the corresponding region. The plasmid was transformed into strains REL606, ZDB30, ZDB199 and ZDB200. Cit$^+$ transformants were identified by positive reactions on Cristensen's Citrate Agar[52].

Identification of mutations in Cit$^+$ replays. Nineteen Cit$^+$ mutants isolated in replay and related experiments[20] were analysed for mutations. The mutants were first checked for large changes in the *citT* region by Southern hybridization with *citT*-specific probes. Genomic DNA was digested with EcoRV (NEB); fragments were separated on 0.8% agarose gels with a 1-kb ladder (NEB), then transferred to nylon membranes. Hybridizations were performed at 68 °C. The *citT*-specific probe was an internal fragment amplified by PCR using primers in Supplementary Table 14, purified using an Illustra GFX PCR DNA purification kit (GE Healthcare), and labelled using the DIG DNA labelling and detection kit (Roche). Most, but not all, mutants had enlarged *citT* bands. We tried to PCR-amplify across possible amplification boundaries of each mutant using the same outward-directed *citT* primers used to screen for the original amplification in the Ara–3 population (Supplementary Table 14). PCR products were thus obtained for 8 clones. Sanger sequencing of the products showed novel junctions consistent with amplifications similar but not identical to the original case. For 7 other clones, PCR products of altered size were obtained when the region immediately upstream of *citT* was amplified, and sequencing showed that each alteration was caused by either a small deletion or an IS*3* insertion in *citG*. To identify mutations at other loci, as well as mutations affecting the *cit* region, we sequenced the genomes of six mutants, as described above. These genomes included three of the four mutants for which the mutations conferring the Cit$^+$ trait were not resolved using the approaches described above.

50. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nature Methods* **3,** 623–628 (2006).
51. Herring, C. D., Glasner, J. D. & Blattner, F. R. Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. *Gene* **311,** 153–163 (2003).
52. Christensen, W. B. Hydrogen sulfide production and citrate utilization in the differentiating of enteric pathogens and coliform bacteria. *Res. Bull. Weld County Health Dept. Greeley Colo.* **1,** 3–16 (1949).