COMMENTARY

Refining a key metabolic innovation in *Escherichia coli*

Sergey Kryazhimskiy and Michael M. Desai[1]

Department of Organismic and Evolutionary Biology, Department of Physics, and FAS Center for Systems Biology, Harvard University, Cambridge, MA 02138

The origins of evolutionary novelty are notoriously difficult to predict in advance or even to pinpoint in retrospect. Qualitatively new traits can arise due to complex mutational events, such as regulatory changes (1), or exaptation of genes coding for existing functions (2). This often involves multiple mutations that are all required to produce a novel function, making evolutionary innovation historically contingent on the chance order of random mutational events (3). Lenski and colleagues recently discovered a striking example of this process in their long-term evolution experiment (LTEE) in *Escherichia coli*: after more than 31,000 generations of evolution, 1 of 12 replicate lines evolved the ability to metabolize citrate under oxic conditions (4). Inability to grow aerobically on citrate is often used to define *E. coli* as a species (5), so the spontaneous appearance of this citrate-metabolizing phenotype offers an unprecedented opportunity to study the mechanisms underlying a major evolutionary innovation. The work by Quandt et al. (6) in PNAS sheds new light on the genetic basis of this novel trait.

Previous studies have shown that the ability to robustly metabolize citrate under aerobic conditions evolved in the LTEE via at least three distinct mutational events. First, an as-yet undetermined potentiating mutation or mutations created a genetic background in which the citrate-metabolizing phenotype was possible (4). Next, a tandem duplication captured the aerobically expressed *rnk* promotor to drive expression of the citrate transporter CitT (7). This *rnk-citG* actualizing mutation allowed *E. coli* to metabolize citrate, but only inefficiently, creating a weak Cit^{+} phenotype that persisted at low frequency in the population. Finally, subsequent refinement mutations converted the weak Cit^{+} phenotype into a strong Cit^{++} phenotype that quickly spread to high frequency and led to a dramatic increase in population density (7). This sequence of events leads to a natural all-or-none epistasis hypothesis, in which all three mutations (potentiation, actualization, and refinement) are necessary to confer the Cit^{++} phenotype.

With this picture in mind, Quandt et al. (6) set out to identify the potentiating and refinement mutations. This task is complicated by the need to distinguish these mutations among more than 70 other mutations present in early Cit^{++} clones. To solve this problem, Quandt et al. introduce a technique, recursive genomewide recombination and sequencing (REGRES), to map the mutations involved in the Cit^{++} phenotype. They first backcross an evolved Cit^{++} clone (previously isolated from generation 33,000 of the LTEE) (7) with its Cit^{-} ancestor. This generates a pool of recombinant strains, each with some subset of the mutations present in the evolved clone. The authors screen these recombinant strains to isolate several that have the Cit^{++} phenotype and repeat the backcrossing procedure several times. They then sequence the resulting backcrossed Cit^{++} clones to identify a minimal set of mutations required to confer the Cit^{++} phenotype (Fig. 1*A*).

This minimal set of mutations turned out to be surprisingly small. As expected, Quandt et al. found the actualizing *rnk-citG* mutation, along with a known refinement mutation, *dsbG–insA-9*. The only other mutation on the list was *dctA**, a mutation in the promoter region of *dctA* that appeared in the LTEE after the *rnk-citG* mutation. No potentiating mutations were found. To exclude the possibility that some mutations were missed at the sequencing stage, Quandt et al. (6) show that the combination of just two mutations, *rnk-citG* and *dctA**, is indeed necessary and sufficient to convert the ancestral Cit^{-} strain into a Cit^{++} strain. This contradicts the all-or-none epistasis hypothesis and suggests a different scenario for the evolution of the Cit^{++} phenotype. Even though potentiating mutation(s)

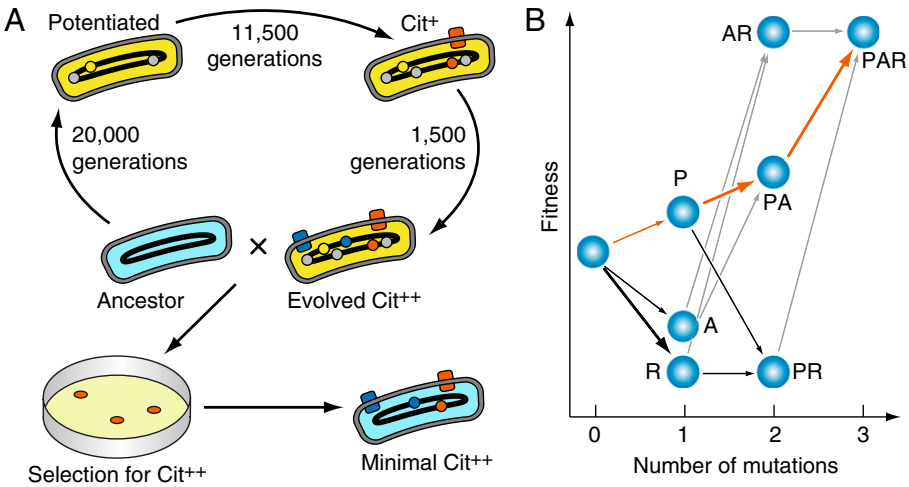

Fig. 1. (*A*) One or more mutations (yellow) that potentiated the evolution of citrate utilization arose by generation 20,000 of the LTEE. This was followed by the *rnk-citG* actualizing mutation (orange) that conferred a weak Cit^{+} phenotype, which was later refined (blue) into the strong Cit^{++} phenotype. More than 70 other mutations (gray) also accumulated during this time. Backcrossing, selection, and sequencing allowed Quandt et al. to isolate mutations necessary for the Cit^{++} phenotype; this did not include the original potentiating mutation(s). (*B*) A possible fitness landscape for the evolution of Cit^{++} phenotype. Thick and thin arrows show confirmed and hypothesized fitness effects of potentiating (P), actualizing (A), and refinement (R) mutations; orange arrows show the actual path taken by the evolving population.

Author contributions: S.K. and M.M.D. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 2217.

[1]To whom correspondence should be addressed. E-mail: mdesai@oeb.harvard.edu.

were required for the actualizing *rnk-citG* mutation to become beneficial and were therefore essential to the original evolution of the citrate metabolism, they are rendered obsolete by the subsequent *dctA** refinement mutation.

The physiological effect of the *dctA** mutation suggests the mechanistic basis for this interesting type of epistasis. Quandt et al. (6) find that the *dctA** mutation increases the expression of DctA, a C_4-dicarboxylate transporter (8). They hypothesize that the *rnk-citG* actualizing mutation is not initially beneficial in the LTEE ancestor because intracellular concentrations of succinate and other C_4-dicarboxylates are too low to power citrate import via the CitT antiporter. The unknown potentiation mutation(s) somehow increase the intracellular supply of C_4-dicarboxylates, enabling the *rnk-citG* mutation to power citrate import and hence confer the weak Cit^+ phenotype. The *dctA** refinement mutation then allows for the recovery of exported C_4-dicarboxylates, enabling much more efficient import of citrate by the CitT antiporter. This makes the original potentiating mutation(s) superfluous for the Cit^{++} phenotype. However, because succinate and other C_4-dicarboxylates are otherwise rare in the media, the *dctA** mutation is unlikely to be beneficial in the absence of *rnk-citG*.

These results highlight the importance of historical contingency in the evolution of this metabolic innovation. Why did the *E. coli* of Lenski and colleagues not take the most direct evolutionary path from Cit^- to Cit^{++} via just two mutations, *dctA** and *rnk-citG*? The answer probably lies in the shape of the "fitness landscape" on which these populations evolve (Fig. 1*B*). Only the *rnk-citG* and *dctA** mutations are required for the Cit^{++} phenotype, but because both are likely to be neutral or deleterious on their own, evolution cannot easily find this solution. Instead, the potentiating mutation(s) provide a way around this dynamical barrier. This is similar in spirit to the conclusions of another recent experimental study, which found that adaptation of budding yeast to high temperature proceeded initially by whole-chromosome duplications that were later replaced by more refined adaptive solutions (9). In both of these systems, the dynamics of the evolutionary process, which proceeds by tinkering rather than by design, are key to understanding the mechanisms underlying novel phenotypes.

The dynamics of the evolutionary process, which proceeds by tinkering rather than by design, are key to understanding the mechanisms underlying novel phenotypes.

It is interesting to note that a purely retrospective analysis could easily have concluded that only the *rnk-citG* and *dctA** mutations were involved in the evolution of the Cit^{++} phenotype. This would imply an adaptive trajectory in which evolution crossed a fitness valley or plateau. However, the earlier evidence (4, 7) for the existence of potentiating mutation(s) shows that this is not necessarily the case. This argues for a degree of caution in interpreting other retrospective analyses of epistasis among mutations that underlie the evolution of any given trait. Later refining mutations can mask the effects of earlier steps, making them difficult to detect and leading to a qualitatively incorrect picture of the evolutionary history. Even in this system, where the existence of potentiating mutation(s) is already suspected, the masking effect of the *dctA** mutation makes them difficult to find. Thus far, their identity remains an intriguing mystery.

1 True JR, Carroll SB (2002) Gene co-option in physiological and morphological evolution. *Annu Rev Cell Dev Biol* 18:53–80.

2 Gould SJ, Vrba ES (1982) Exaptation: A missing term in the science of form. *Paleobiology* 8(1):4–15.

3 Gould SJ (1989) *Wonderful Life* (Norton, New York).

4 Blount ZD, Borland CZ, Lenski RE (2008) Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc Natl Acad Sci USA* 105(23):7899–7906.

5 Scheutz F, Strockbine NA (2005) *Bergey's Manual of Systematic Bacteriology, Volume Two*, eds Garrity GM, et al. (Springer, New York).

6 Quandt EM, Deatherage DE, Ellington AD, Georgiou G, Barrick JE (2014) Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in *Escherichia coli*. *Proc Natl Acad Sci USA* 111:2217–2222.

7 Blount ZD, Barrick JE, Davidson CJ, Lenski RE (2012) Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature* 489(7417):513–518.

8 Davies SJ, et al. (1999) Inactivation and regulation of the aerobic C(4)-dicarboxylate transport (dctA) gene of Escherichia coli. *J Bacteriol* 181(18):5624–5635.

9 Yona AH, et al. (2012) Chromosomal duplication is a transient evolutionary solution to stress. *Proc Natl Acad Sci USA* 109(51):21010–21015.