The evolution of fitness effects during long-term adaptation in bacteria

Alejandro Couce[1,2,3*], Melanie Magnan[1], Richard E. Lenski[4,5] and Olivier Tenaillon[1*]

[1]Unité Mixte de Recherche 1137 (IAME, INSERM), Université Sorbonne Paris Nord, 75018 Paris, FRANCE
[2] Department of Life Sciences, Imperial College London, SW7 2AZ London, UK
[3]Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM), 28223 Madrid, SPAIN
[4]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824, USA
[5]Program in Ecology, Evolution, and Behavior, Michigan State University, East Lansing, Michigan 48824, USA
*Corresponding authors: a.couce@upm.es, olivier.tenaillon@inserm.fr

Abstract

Understanding the distribution of fitness effects of new mutations is central to predicting adaptive evolution. Short-term experiments provide a snapshot of this distribution, but observing how it changes as organisms adapt is challenging. Here we use saturated, genome-wide insertion libraries to quantify how the fitness effects of new mutations changed in two *E. coli* populations that adapted to a constant environment for 15,000 generations. The proportions of neutral and deleterious mutations remained constant, despite fitness gains of ~50%. In contrast, the beneficial fraction declined rapidly and became exponentially distributed, with genetic interactions profoundly reshuffling the loci subject to beneficial mutations. Despite this volatility, the ancestral distribution predicts many of the alleles that become dominant in the long-term experiment, even after the initial period of rapid adaptation. Overall, our results suggest that short-term adaptation can be idiosyncratic but empirically reproducible, and that long-term dynamics can be described by simple statistical principles.

Long-term evolution results from the accumulation of new mutations. Therefore, detailed knowledge of the proportions of mutations that are beneficial, neutral, or deleterious is important for predicting the course and outcomes of evolution. Indeed, assumptions about the distribution of fitness effects (DFE) of new mutations (DFE) lie at the core of many theories describing fundamental evolutionary phenomena, including the speed of adaptation *(1)*, fitness decay in small populations *(2)*, the maintenance of genetic variation *(3)*, the probability of parallel (*4*) versus divergent *(5)* evolution, the pace of the molecular clock *(6)*, and the evolution of sex *(7)* and mutation rates *(8)*. Driven by this interest, many experimental and comparative studies have produced estimates of this distribution in different organisms. While most studies have been small in scale *(9)* or focused on narrow genomic regions *(10)*, some important features appear similar across a variety of model systems. In particular, most mutations are neutral or deleterious, lethal mutations form a distinct class within the deleterious tail, and those rare beneficial mutations are typically (but not always) exponentially distributed *(11)*.

However, the DFE only indicates what is possible at a particular point in time, and it is unknown how long the distribution will remain relevant as evolution proceeds and especially as beneficial substitutions accumulate. Predictions on how the shape of the DFE changes are implicit in some theoretical models, most famously that the beneficial tail should approach an exponential distribution near a fitness peak *(12, 13)*. The picture is more ambiguous for the deleterious tail, with different models predicting it to become heavier (*14*) or lighter *(15)* with adaptation. In any case, these predictions address only the macroscopic form of the DFE, with little attention to the microscopic processes underlying the changes in shape. And while the shape influences the dynamics, the microscopic details—especially the sign and intensity of interactions among mutations (*i.e.*, epistasis)—determine the outcomes of adaptation *(16)*. For example, in the absence of epistasis, adaptation will shorten the beneficial tail simply by the process of sampling without replacement. In this case, a complete DFE would suffice to specify the probabilities of all possible adaptive trajectories in a given environment. At the other extreme, if sign epistasis is the norm—such that mutations go from beneficial to deleterious, and vice versa, as the genome evolves *(17)*—then new mutations will continually change the effects and rank order of the remaining mutations, rendering futile any prediction about adaptive trajectories beyond the very short term.

High-throughput insertion mutagenesis and fitness measurements

To illuminate experimentally how adaptation and epistatic interactions change the DFE, one would ideally like to measure the relative fitness of a large set of mutants at multiple time points along a broad and well-characterized adaptive trajectory. To do so, we take advantage of the long-term evolution experiment (LTEE) in which populations of *Escherichia coli* have been adapting to a glucose-limited medium for tens of thousands of generations, resulting in large fitness increases *(18)*. Examples of both weak and strong epistasis among beneficial mutations have been reported in this system *(16, 19)*. Moreover, the most important mutations driving adaptation have been identified from signatures of parallelism in whole-genome sequences *(20, 21)*, allowing the predictive capacity of a DFE at one time point to be compared with the actual fate of mutations during later adaptation. To measure the DFE, we created genome-wide libraries of insertion mutants using a transposon engineered to capture the 14-bp sequence adjacent to each insertion site, which in most cases identifies unequivocally the target locus *(22)* (Methods, fig. S1). We typically identified >100,000 different insertion mutants, which mapped to

>78% of the ancestral genome's 8424 loci including both open reading frames (ORFs) and intergenic regions (Methods).

We estimated the fitness effects of all these mutants as selection coefficients, for which we tracked the frequency trajectory of every allele during 5-day, bulk competition assays under the same conditions as in the LTEE (Fig. 1, a-c; Methods). Although transposon insertions typically cause losses of function, we also found evidence of two types of more subtle effects (Fig. 1, d). First, an insertion in the C-terminus of a gene may cause only a partial loss of function or even a change in function. This outcome is prominently revealed by the tolerance of many essential genes to insertions in that region, notably including insertions in *topA* and *serB* that show large benefits *(23*). Second, the position of many beneficial insertions, including in intergenic regions and genes upstream of known targets of adaptation in the LTEE, suggests changes in gene expression. To preserve these types of effects, while also ensuring robust fitness estimates, we divided each locus into 5 segments of equal length and then pooled all insertions in each segment. As an added benefit, fitness comparisons among segments of the same locus allow identification of outliers and provide an internal control to quantify the reproducibility of the fitness estimates (Methods, fig. S2).

Changes in the size and shape of the beneficial tail of the DFE

Using this approach, we first sought to characterize how the macroscopic shape of the DFE changed as fitness increased during the LTEE. The rate of fitness increase declined during the LTEE, such that half of the ~70% gain typically seen at 50,000 generations had already occurred by 5,000 generations *(18*). We decided therefore to create transposon libraries in three genetic backgrounds: the ancestor (which we call "Anc") and clones sampled from population Ara+2 at 2,000 ("2K") and 15,000 ("15K"), when fitness had increased by ~25% and ~50%, respectively. Despite these large fitness gains, Figure 2 shows that the overall shape of the DFEs remained similar, with one critical difference—namely, the fraction of beneficial insertion mutations is substantially larger in the ancestor than in the evolved backgrounds (7.4% for Anc versus 5.1% and 3.9% for 2K and 15K, respectively; $P < 0.004$ both cases, two-sample Kolmogorov–Smirnov [K-S] test). In contrast, the deleterious fraction is essentially constant across the three backgrounds (21.4% for Anc versus 22.6% and 21.2% for 2K and 15K, respectively; $P > 0.073$ both cases, two-sample K-S test). These patterns are consistent with analyses performed at the level of individual genes for both beneficial and deleterious mutations (Fig. 3, a-b).

We also examined whether these results depended on the particular evolutionary lineage that we chose to study. To that end, we also measured the DFEs for clones sampled at 2,000 and 15,000 generations from population Ara–1, an independently evolving lineage that accumulated different beneficial mutations along its adaptive trajectory (Methods, table S1). At least two major features distinguish the evolutionary history of this lineage from that of Ara+2. First, Ara–1 fixed a mutation in *topA* early in the LTEE, which confers the highest fitness benefit seen in this system for any single substitution (*19*). Mutations in this locus reached fixation in five of the twelve LTEE populations, but they never reached detectable frequency in Ara+2. Second, Ara–1 evolved a mutator phenotype, whereas Ara+2 retained the low ancestral mutation rate throughout the experiment (note, however, that Ara–1 became hypermutable only

after ~21,000 generations, and hence it did not affect the clones in our study). Despite their independent evolutionary histories, we obtained strikingly similar overall results for the two lineages, at both the macroscopic and microscopic levels (fig. S3).

How do our observations compare to previous studies and expectations? An influential prediction based on statistical arguments is that the effects of beneficial mutations should be exponentially distributed if the population is well-adapted to its environment (*1*, *12*). Despite some empirical support (*24–26*), the evidence remains inconclusive overall owing to a severe limitation of most studies: without detailed knowledge of a population's evolutionary history, it is difficult to evaluate its level of adaptation to a particular environment. Our data, therefore, provides a unique test of these ideas. We find that, indeed, beneficial mutations in the evolved backgrounds are well fit by an exponential distribution, whereas this fit can be rejected for the ancestor (P = 0.016 for Anc versus P = 0.716 and P = 0.984 for Ara+2 clones 2K and 15K, respectively; one-sample K-S test). We also considered other commonly used alternative distributions, but the exponential fit provides the most satisfactory distribution for the evolved backgrounds (Methods, table S2). In contrast, the beneficial tail for the ancestor was reasonably well fit by both gamma and Weibull distributions (P = 0.059 and P = 0.21, respectively; one-sample K-S test), consistent with previous reports for viral and bacterial genotypes thought to be poorly adapted to the test environment *(26, 27*). Overall, our results support the view that, after an early period of rapid adaptation to a new environment, the distribution of beneficial mutations becomes exponential. Thus, by analyzing changes in the DFE in a temporal series of genetic backgrounds becoming better adapted to their environment, we can reconcile disparate pieces of evidence and provide insights relevant to many models of adaptation.

Constancy of the deleterious tail of the DFE

The constancy of the deleterious tail, however, stands in contrast to a study that measured the DFE for 710 insertion mutations in hybrid yeast genotypes with a fitness range spanning ~20%, in which deleterious effects were significantly worse in the more-fit backgrounds *(28*). A potentially important difference is that the fitness variation among the yeast backgrounds was generated by crossing two distantly related strains, whereas we use a series of backgrounds from lineages undergoing adaptation to the same environment in which we assess the fitness effects of the new mutations. As further support for our findings, a companion study focused on the evolution of gene essentiality in the LTEE found no systematic changes in deleterious effects across all of the lineages at 50,000 generations (*29*). In any case, theoretical predictions about the tail of deleterious mutations differ substantially and have been guided mostly by plausibility arguments *(14*, *15*), and so all of these studies should help refine current models by clarifying the assumptions and narrowing the range of parameters.

Changing identity of beneficial mutations and sign epistasis

Having now described how the macroscopic structure of the DFE changed as the bacteria adapted to the LTEE environment, we next sought to understand how these macroscopic changes emerged from changes at the level of genes and mutations. Figure 3a shows that deleterious mutations generally exhibit only slight epistasis across the three focal genetic backgrounds of the Ara+2 lineage; the magnitude of

their harmful effects may vary, but without any tendency towards either aggravating or alleviating effects. Thus, deleterious mutations in the ancestor remain deleterious in the evolved backgrounds, consistent with the observed constancy of the deleterious tail. In stark contrast, beneficial mutations are largely dominated by strong, sign-epistatic interactions (fig. 3b). Only 6.1% of the mutations beneficial in the ancestor are still beneficial by 2,000 generations, with most becoming effectively neutral (73.9%) and some deleterious (20%) (fig. 3c, left panel). This pattern also holds in the reverse direction: most beneficial mutations at 2,000 generations are neutral (74.1%) or deleterious (19%) in the ancestor (fig. 3c, left panel). Similar patterns occur when comparing how fitness effects changed between 2,000 and 15,000 generations (fig. 3C, right panel). Intrigued by the transitory nature of beneficial effects, we asked whether the overall DFE of the initially beneficial mutations retains even a slightly positive tendency at the later time points. In fact, it does not. The DFE of mutations that were beneficial in the ancestor becomes indistinguishable from a random sample of the parent distribution (fig. 3d, left panel), and the same holds for the reverse scenario (fig. 3d, right panel) ($P > 0.130$ both cases; two-sample K-S test). This regression to the mean persists even when we account for measurement noise around neutrality (fig. S2).

What might explain this turnover in the identity of the beneficial mutations? In a previous study, the first five mutations to fix in one LTEE population were shown to exhibit diminishing-returns epistasis, such that their benefits declined in magnitude as the background fitness increased *(19)*. However, it was unlikely *a priori* that these five mutations would show sign epistasis because they were chosen precisely because their combination was favored by natural selection *(30)*. By contrast, another study analyzed the co-occurrence of fixed mutations across 115 lines of *E. coli* that had evolved under thermal stress, and found that sign epistasis was indeed common *(31)*. Moreover, that study found that the prevalence of different types of epistasis reflected the modular architecture of cellular traits: mutations affecting different modules tend to interact more or less additively, while mutations impacting the same module tend to be redundant. We therefore investigated the extent of modularity in our data, and we found that beneficial mutations tend to cluster together in operons, the most basic functional level (Methods, $P < 0.01$). Mutations in the same operon are likely to alter the same cellular process in similar ways, and therefore the potential for redundancy at this level provides a simple explanation for why large sets of beneficial mutations disappear, and other sets emerge, as adaptation proceeds. Even without considering these specific details, the increased prevalence of sign epistasis with adaptation has also been predicted from general properties of the genotype-to-fitness map *(32)*.

We identified a large pool of loci that can produce beneficial mutations, including some known targets for adaptation in the LTEE (e.g., *topA*, *pykF*, *nadR*) *(20)*. However, the fate of beneficial mutations is determined not only by their fitness, but also by the nature and intensity of their interactions with other beneficial mutations *(16, 17)*. As a consequence, only a fraction of all possible beneficial mutations will contribute to adaptation in an evolving population. To gain further insight into this issue, we compared our data with metagenomic data obtained by sequencing whole-population samples from the 12 LTEE populations over the course of 60,000 generations *(21)*. We see a significant, but fairly weak, correlation between our fitness estimates of mutations in the ancestor and the abundance of corresponding alleles during the LTEE ($r = 0.36$, fig. 4a), and this correlation largely disappears when using the beneficial

effects estimated in the evolved backgrounds (fig. S4). An important factor contributing to these weak correlations is that our methods involve insertion mutations, which usually, but not always, cause losses of function (fig. 1). While losses of unused functions have contributed to adaptation in the LTEE (*20*, *33*), subtle changes that typically require point mutations have also been important in refining some functions (*16*, *20*, *34*). In contrast, the abundance of alleles in the metagenomic data correlates more strongly with the target size of the locus (r = 0.72, fig. 4b) (Methods). These "quasi-neutral" patterns are consistent with intense competition among independently segregating beneficial mutations (*i.e.*, clonal interference), a phenomenon that is pervasive in the LTEE *(21*, *35*). Under intense clonal interference, the probability that different mutations occur may shape genomic evolution more than their individual fitness effects *(36*). In any case, the best linear model includes target size as the most explanatory single variable, but it also includes significant contributions from the fitness effects in both the ancestral and 2,000-generation genetic backgrounds (fig. 4c, table S3).

Predicting future beneficial mutations as adaptation proceeds

Finally, given that sign epistasis is widespread, it is natural to ask for how long the information about the particular loci in the beneficial tail of a DFE can predict the subsequent steps of adaptation. To address this question, we recorded the alleles that were nearing fixation through time using the metagenomic data, and we calculated how many of them corresponded to those loci for which we detected beneficial effects. We found that the ancestral DFE predicted most of the mutations that become dominant early in the LTEE populations; the predictive power decays rapidly but was still evident for ~15,000 generations (fig. 4d). Interestingly, this rapid decay was largely driven by the lineages that evolved hypermutability early in the LTEE; when these mutator populations are removed from the analysis, the ancestral DFE retained significant predictive power through 50,000 generations (fig. S5). In turn, the DFEs measured in the evolved backgrounds had less predictive power, and it takes longer for their predictions to materialize; the latter effect may reflect the declining rate of adaptation. These patterns are reminiscent of work showing that parallel genomic evolution was more common early in the LTEE than in later generations *(20*, *37*).

Taken together, our results shed new light on which aspects of adaptation to a novel environment are predictable, and which are not, given our current understanding of genetics and evolution. While the first stages of adaptation are not yet well-captured by theory, we have demonstrated that the major drivers of early adaptation can be predicted from high-throughput empirical fitness data. By contrast, the long-term dynamics of adaptation are well-described using simple statistical arguments. However, predicting the genetic identity of the late drivers of adaptation is more difficult, and it will remain a challenge until a general theory of epistasis has been developed.

Figures

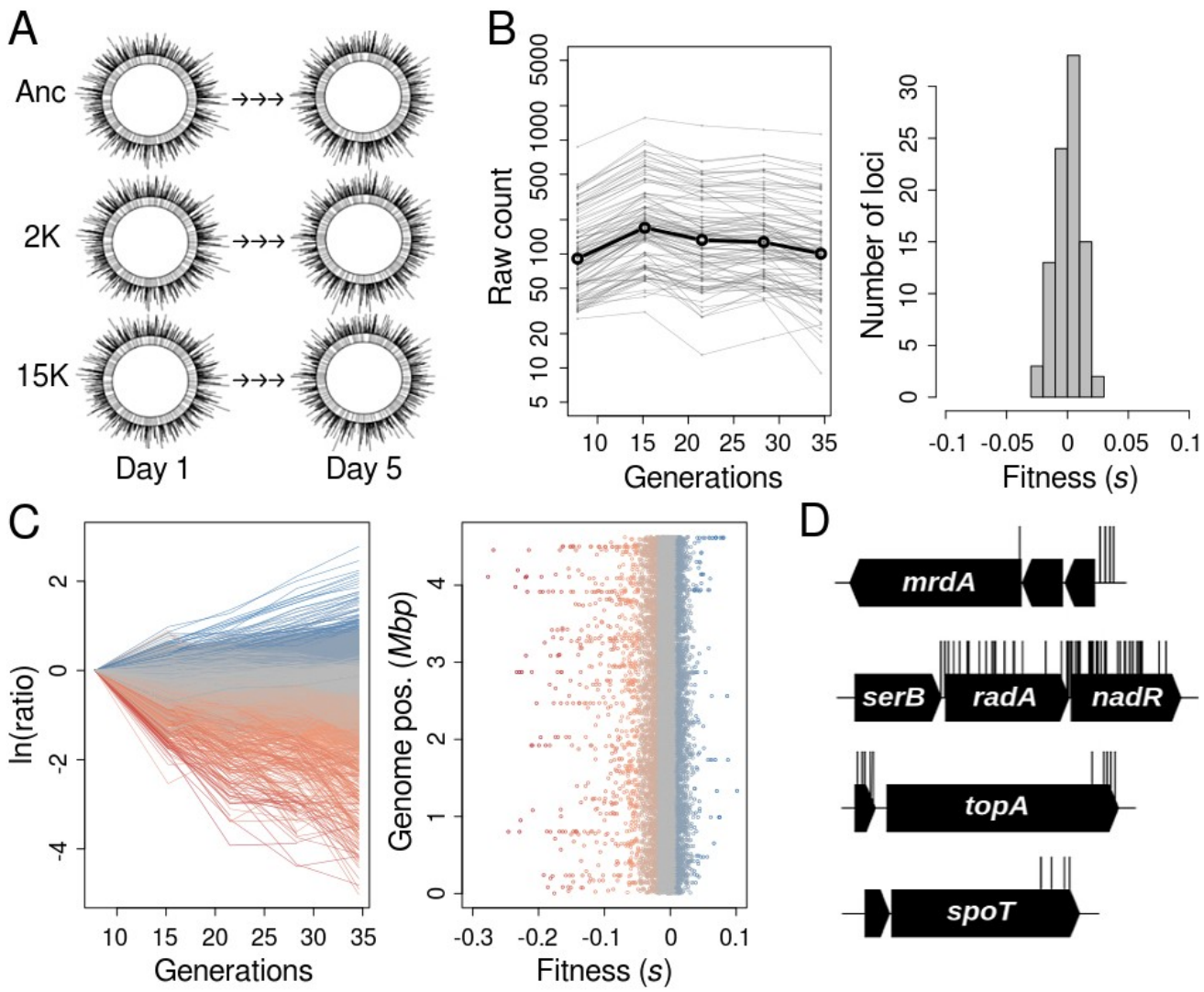

Fig. 1. Overview of experimental procedures. (A) Several saturated, genome-wide insertion libraries in the ancestor and two evolved isolates (2K and 15K generations) from each of two LTEE populations (Ara+2 and Ara–1) were subjected to bulk competition and sequencing. **(B)** The abundance trajectories of well-known neutral loci were taken as reference to estimate selection coefficients (left). The values for this neutral set were closely centered around zero (right). **(C)** Frequency trajectories of the whole mutant library in the ancestor (left), and mapping of the selection coefficient estimates along the chromosome (right). Colors indicate fitness effects, from deleterious (red) to beneficial (blue). **(D)** Examples of important sub-genic structure for known targets of selection in which we observed polar effects involving the preceding regions of the same transcription unit (*mrdA*, *nadR*, and *topA*), and showing the tolerance to insertions of the C-terminal portion of essential genes (*serB*, *topA*, and *spoT*).

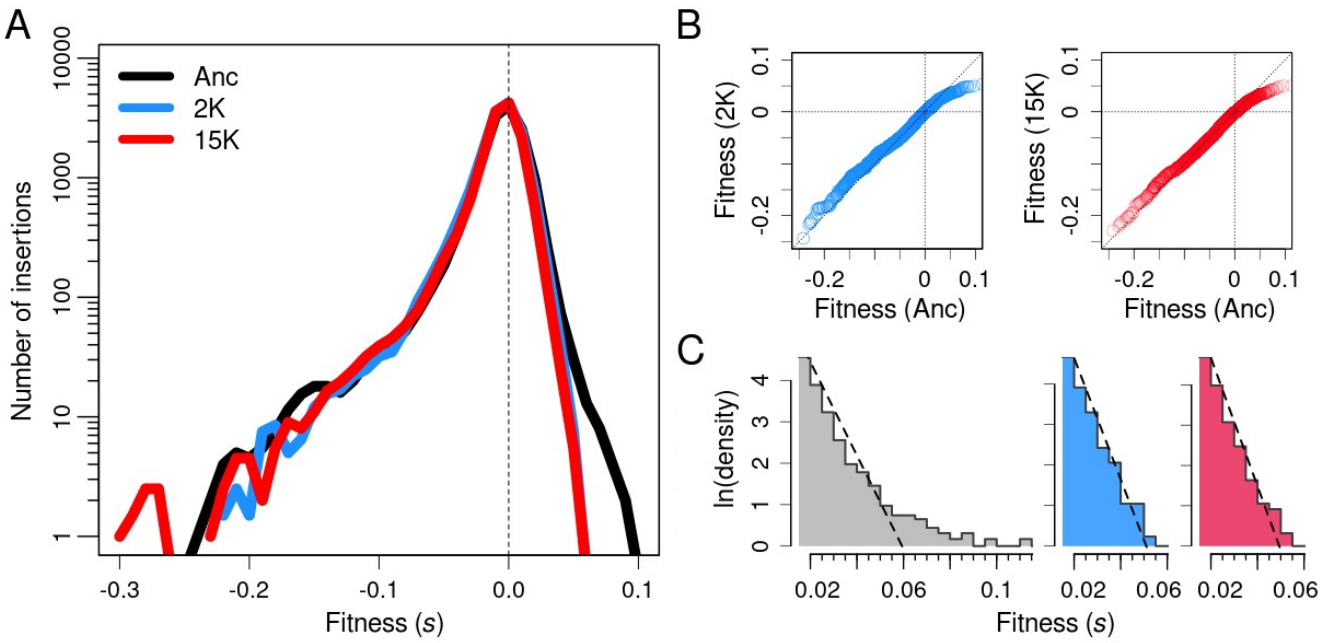

Fig. 2. Change in the DFE along a large fitness gradient. (A) DFEs in the ancestor (black), 2K (blue) and 15K (red) evolved strains from population Ara+2. Note the logarithmic scaling of the y-axis. **(B)** Deleterious tails remained unchanged during adaptation, as indicated by comparing the cumulative distributions between the ancestor and 2K (left), and the ancestor and 15K (right). **(C)** Beneficial tails are rapidly truncated and become exponentially distributed. Histograms show the best fit to an exponential distribution (dashed line) in the ancestor (gray), 2K (blue), and 15K (red) backgrounds.

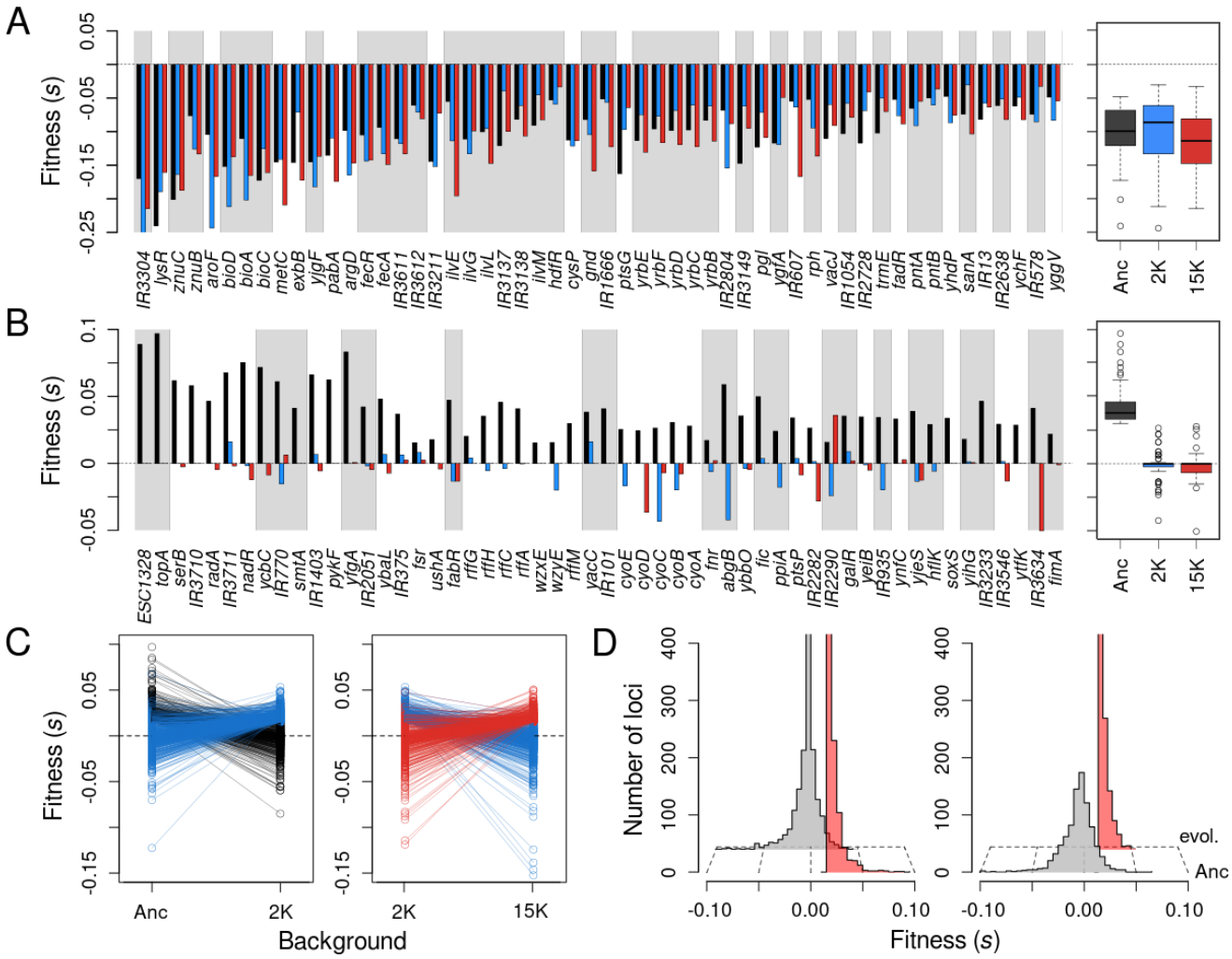

Fig. 3. Effect of genetic background on the fitness effects of mutations in specific gene targets. (A) The genes and intergenic regions subject to the most severely deleterious mutants in three backgrounds in the Ara+2 lineage. Colors indicate ancestor (black), 2K (blue), and 15K (red) evolved strains. Gray shaded areas indicate members of the same transcription unit. **(B)** The genes and intergenic regions with the most beneficial alleles in the ancestral background, and their fitness effects in the 2K (blue) and 15K (red) backgrounds. **(C)** Most beneficial mutations available to the ancestor became neutral or deleterious in the 2K background, while most beneficial mutations available in the 2K background were neutral or deleterious in the ancestor (left). The same general pattern occurs when comparing beneficial mutations in the 2K and 15K backgrounds (right). **(D)** More than 90% of initially beneficial mutations become neutral or deleterious in later generations (left). Likewise, more than 90% of beneficial mutations from later generations were neutral or deleterious in the ancestor.

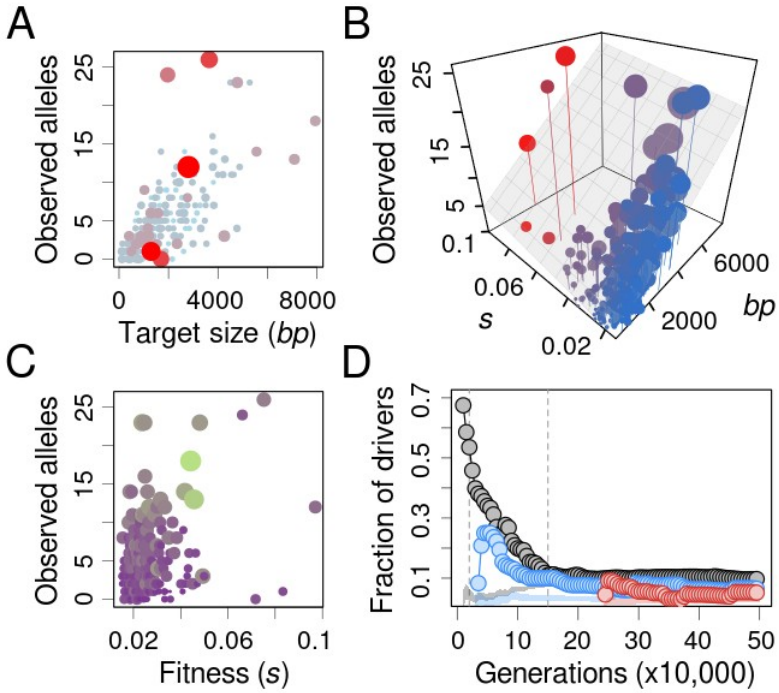

Fig. 4. Determinants of evolutionary outcomes. (A, C) The prevalence of mutations in the LTEE is better explained by mutational target size (A, area and color of dots represent fitness) than by fitness effects measured in the ancestor (C, area and color of dots represent target size). **(B)** The best linear model for mutation prevalence is strongly dependent on the mutational target size (area of dots represents size, and color represents fitness). **(D)** The predictive capacity of DFEs as a function of time in the LTEE. Colors indicate the DFE measured in the ancestor (black), 2K (blue) and 15K (red) evolved strains. Shaded areas show the null expectations based on randomly sampling neutral and deleterious mutations.

References

1. C. O. Wilke, *Genetics*. **167**, 2045–2053 (2004).
2. M. Lynch, R. Bürger, D. Butcher, W. Gabriel, *J. Hered.* **84**, 339–344 (1993).
3. D. Charlesworth, B. Charlesworth, M. T. Morgan, *Genetics*. **141**, 1619–1632 (1995).
4. L.-M. Chevin, G. Martin, T. Lenormand, *Evol. Int. J. Org. Evol.* **64**, 3213–3231 (2010).
5. J. J. Welch, A. Eyre-Walker, D. Waxman, *J. Mol. Evol.* **67**, 418–426 (2008).
6. T. Ohta, *Annu. Rev. Ecol. Syst.* **23**, 263–286 (1992).
7. J. R. Peck, G. Barreau, S. C. Heath, *Genetics*. **145**, 1171–1199 (1997).
8. O. Tenaillon, B. Toupance, H. Le Nagard, F. Taddei, B. Godelle, *Genetics*. **152**, 485–493 (1999).
9. R. Sanjuán, *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 1975–1982 (2010).
10. H. Kemble, P. Nghe, O. Tenaillon, *Evol. Appl.* **12**, 1721–1742 (2019).
11. A. Eyre-Walker, P. D. Keightley, *Nat. Rev. Genet.* **8**, 610–618 (2007).
12. J. H. Gillespie, *Evolution*. **38**, 1116–1129 (1984).
13. H. A. Orr, *Genetics*. **163**, 1519–1526 (2003).
14. D. P. Rice, B. H. Good, M. M. Desai, *Genetics*. **200**, 321–329 (2015).
15. L. Perfeito, A. Sousa, T. Bataillon, I. Gordo, *Evol. Int. J. Org. Evol.* **68**, 150–162 (2014).
16. R. J. Woods *et al.*, *Science*. **331**, 1433–1436 (2011).
17. D. M. Weinreich, R. A. Watson, L. Chao, *Evolution*. **59**, 1165–1174 (2005).
18. M. J. Wiser, N. Ribeck, R. E. Lenski, *Science*. **342**, 1364–1367 (2013).
19. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, *Science*. **332**, 1193–1196 (2011).
20. O. Tenaillon *et al.*, *Nature*. **536**, 165–170 (2016).
21. B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, M. M. Desai, *Nature*. **551**, 45–50 (2017).
22. A. L. Goodman *et al.*, *Cell Host Microbe*. **6**, 279–289 (2009).
23. A. Couce *et al.*, *Proc. Natl. Acad. Sci.* **114**, E9026–E9035 (2017).
24. R. Kassen, T. Bataillon, *Nat. Genet.* **38**, 484–488 (2006).
25. D. R. Rokyta, P. Joyce, S. B. Caudle, H. A. Wichman, *Nat. Genet.* **37**, 441–444 (2005).
26. R. Sanjuán, A. Moya, S. F. Elena, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8396–8401 (2004).
27. R. D. H. Barrett, R. Craig MacLean, G. Bell, *Biol. Lett.* **2**, 236–238 (2006).
28. M. S. Johnson, A. Martsul, S. Kryazhimskiy, M. M. Desai, *Science*. **366**, 490–493 (2019).
29. A. Limdi, S. V. Owen, C. Herren, R. E. Lenski, M. Baym, *bioRxiv* doi.org/10.1101/2022.05.17.492023 (2022).
30. F. Blanquart, G. Achaz, T. Bataillon, O. Tenaillon, *Evolution*. **68**, 3537–3554 (2014).
31. O. Tenaillon *et al.*, *Science*. **335**, 457–461 (2012).
32. D. Greene, K. Crona, *PLoS Comput. Biol.* **10**, e1003520 (2014).
33. V. S. Cooper, D. Schneider, M. Blot, R. E. Lenski, *J. Bacteriol.* **183,** 2834-2841 (2001).
34. R. Maddamsetti *et al*., *Genome Biol. Evol.* **9,** 1072-1083 (2017).
35. D. E. Deatherage, J. E. Barrick, *Cell Syst.* **12**, 1187-1200 (2021).
36. M. M. Desai, D. S. Fisher, A. W. Murray, *Curr. Biol. CB*. **17**, 385–394 (2007).
37. R. E. Lenski, *Am. Nat*. **190**: S57-S68 (2017).

Acknowledgements

We thank Michael Baym and Anurag Limdi for valuable discussions; and Adrien Launay, Alexandra Baron, Romain Fernandes and Damien Roux for technical assistance. Funding: This work was supported by the European Commission under the 7th Framework Program (ERC Grant 310944 to O.T.) and under the Horizon 2020 Framework Programme (MSCA-IF 750129 to A.C.). The work was also partly supported by the Spanish Ministry of Science, Innovation and Universities (PID2019-110992GA-I00 to A.C.), the regional government of Madrid (2019-T1/BIO-12882 to A.C.) and the National Science Foundation (DEB- 1951307 to R.E.L), and Michigan State University. Competing interests: We declare no competing interests.

Supplementary Materials for

The evolution of fitness effects during long-term adaptation in bacteria

Alejandro Couce, Melanie Magnan, Richard E. Lenski and Olivier Tenaillon

This PDF file includes:

Materials and Methods

Generation of transposon mutant libraries

We used the INSeq methodology *(1)* to create transposon mutant libraries in the ancestor and several derivatives from the *E. coli* long-term evolution experiment (LTEE) (table S1). We employed a kanamycin-resistance version of the original INSeq transposon carried in the pSAM plasmid. The ends of this transposon encode recognition sequences for the restriction enzyme *Mme*I, which cuts 20 bp away from its binding site and thus allows the capture of the 14 bp adjacent to the insertion site (Fig. S1). The pSAM plasmid also carries the Himar1C9 transposase, an ampicillin resistance marker and RP4-oriT/oriR6K, a mobilizable and *pir*-dependent origin of replication. We maintained the plasmid into a well-known, pir-gene encoding *E. coli* donor strain, MG1655 MFD*pir (2*); from which it was transferred into the different *E. coli* recipient strains by standard, agar plate conjugation methods. Mating time was limited to 3-4 h, after which we selected transconjugants by spreading the cells onto LB agar plates supplemented with streptomycin (100 mg/L) and kanamycin (100 mg/L). Libraries were assembled from at least 100,000 colonies coming from at least 10 independent conjugation mixtures. The colonies were scraped off the plates, pooled in saline solution and added glycerol (10% final) to store at -80°C.

Bulk competition experiments

We propagated aliquots of the mutant libraries in conditions replicating those of the LTEE: 1/100 daily dilutions in 50-mL flasks filled with 10 mL of DM25 and incubated at 37°C in an orbital shaker *(3*). To ensure that the initial cell density was similar to the one used in the LTEE (~ 5 x 10^7 cells / mL), we inoculated several flasks for each experiment with a different aliquot from the defrosted libraries, which next day we discarded as appropriate based on viable count estimates. The competition experiments were conducted for 5 or 8 days (Ara+2 and Ara-1 datasets, respectively), and the remaining culture after each serial passage was stored at -80°C for further analyses. Viable counts were monitored each day to verify each passage represented the expected ~ 6.6 generations of binary fission.

Generation of DNA libraries for sequencing

The defrosted stocks from the bulk competition experiments were pelleted and subjected to genomic DNA extraction using the DNeasy Blood & Tissue kit (Qiagen). Due to the relatively low cell density supported by DM25, the purified DNA was subjected to high-fidelity, unbiased whole-genome amplification using the *Phi29*-polymerase-based REPLI-g midi kit (Qiagen). The resulting DNA was digested by *Mme*I (NEB), concentrated using a vacuum concentrator and run on a 1% agarose gel. The expected size of the transposon plus the adjacent genomic DNA is 1.4 Kb, so we excised from the gel lanes the regions corresponding to 1.2-1.5 Kb using as a reference the Quick-Load 100 bp DNA ladder (NEB). To enrich the excised samples for transposons, the INSeq method takes advantage of the fact

that *Mme*I digestion results into two-nucleotide, sticky-end fragments; so that custom made, PCR adaptors can be added in a simple ligation reaction (Fig. S1, table S4). We performed the ligations using T4 DNA ligase (NEB). Ligation products contain the 14 bp adjacent to the insertion site flanked by the transposon and the custom-made adaptor. Using these known sequences for annealing, we used a 28-cycle PCR to amplify a small region containing the 14 bp of interest (Fig. S1, table S4). All PCR reactions were carried out using the HiFi DNA Polymerase (KAPA Biosystems). The resulting molecules, 116 bp and 148 bp in size (Ara+2 and Ara-1 datasets, respectively), were isolated by electrophoresis and band excision on a 2% agarose gel. To add the required indexes for multiplexed, paired-end sequencing, we subjected the samples to a secondary, 18-cycle PCR using the Nextera XT DNA Library Prep Kit (Illumina). The final Nextera libraries were purified, quantified and sent for sequencing on Illumina platforms. The Ara+2 libraries were sequenced using a HiSeq platform (Integragen, Evry, France), while the Ara-1 libraries were sequenced using a MiSeq platform (Bichat Hospital, Paris, France). After filtering (see below), the libraries showed an average of ~ 3.55 and ~ 0.39 million reads, mapping to an average of ~ 0.33 and ~ 0.27 million different insertion sites (Ara+2 and Ara-1 datasets, respectively). To control for potential technical biases, libraries for the ancestor were processed using both transposon capture and sequencing methods (Fig. S5).

Insertion location, filtering and abundance determination

We developed a series of custom scripts to map transposon insertion sites, discard low-quality cases and quantify the abundance of each insertion allele. We first used a Python script to extract the captured 14 bp genomic fragments from the Illumina sequence reads. This was accomplished using the module "regex" to identify any of the constant sequences expected to be flanking the captured fragments, up to a maximum edit distance of 1 (*i.e.*, one position indel or mismatch) (Fig. S1). Second, we used the BWA algorithm (Li et al., 2009) to map these genomic fragments onto the *E. coli* B str. REL606 reference genome (NCBI reference: NC_012967), allowing a maximum edit distance of 1. We only retained fragments for which the best hit mapped to a single genomic location, excluding therefore repeated regions from the analysis. Pairs from paired-end reads were merged except when they showed discordant results (in these cases, only the one offering the best alignment was retained). These filtered read mappings were then compiled in a list with all the unique insertion sites, the number of reads per site and the identity of the corresponding locus. As loci, we considered both open reading frames and intergenic regions but excluded repetitive and mobile genetic elements. Finally, to be able to observe sub-genic effects (Fig. 1, D), we divided each locus into equally sized segments (5 for the Ara+2, and 3 for the lower-coverage Ara-1 dataset; in both cases loci were not divided when < 100 bp in size), and calculated the abundance of all insertions mapping onto the same segment. In preliminary tests we observed that excluding insertion sites with very low counts at this pooling step reduced the number of outliers in downstream analyses. This occurs presumably because very low abundance insertions are enriched for PCR and Illumina sequencing artifacts. Therefore, for each segment, we pooled together only insertions with >10 or >3 total counts across all data points (Ara+2 and Ara-1 datasets, respectively). All the following analyses were conducted at this pooled segment level.

Fitness estimates and outlier removal

To estimate the fitness of the pooled insertion mutants, we evaluated the slope of their frequency trajectories relative to that of a set of presumed neutral loci. The natural logarithm of this slope approximates the selection coefficient from classical population genetics, and is related to the ratio of realized growth rates (the standard fitness metric used in the LTEE literature) by a scaling factor of *ln2* (Fig. S6, A) *(4)*. The set of presumed neutral loci included genes annotated as cryptic and the L-arabinose transport and catabolism operons (L-arabinose auxotrophy has extensively been used as a neutral marker in the LTEE). In addition, we only considered their internal segments, to reduce spurious deviations from neutrality linked to polar or change-of-function effects (*e.g.*, Fig. 1, D). We estimated slopes by fitting a linear regression model to the data using the lm() function in R. We weighted the regression using the counts at each time point (using the "weight" argument), in an attempt to reduce the noise associated with low count observations. We found this weighting improved estimates in the deleterious range (Fig. S6, B). Mutants with less than 3 non-zero time points were discarded.

In tests with the neutral set we observed that most of the poor quality fits can be filtered out simply by excluding mutants with low initial abundance (cutoffs for exclusion, <10 or <3 counts in the first data point; Ara+2 and Ara-1 datasets, respectively). However, when the much larger complete set of loci was analyzed, we noticed several instances of marked outliers characterized by increasingly steeper slopes. These cases probably reflected the hitch-hiking of individual insertion alleles with *de novo* beneficial mutations occurring during the competition experiments. In agreement with this interpretation, the outliers were proportionately more common in the higher-coverage Ara+2 dataset; presumably because of the increased resolution to detect rare events. Filtering out by the regression standard error eliminated these prominent outliers from both datasets (cut-offs: 0.01 and 0.015, Ara+2 and Ara-1 datasets, respectively). However, we noticed a few highly beneficial mutations in the ancestor from the Ara+2 dataset that we suspected constituted another class of more subtle outliers. Our suspicion arose from two facts: first, they were completely absent from the other ancestor data, and second, their fitness markedly deviated from the fitness measured for the other insertions in the same locus. These cases probably represented even rarer hitch-hiking events in which the insertion we tracked occurred in a genetic background harboring either a preexisting beneficial mutation or a secondary, beneficial insertion. In these cases the insertion behaves as a beneficial mutation over the whole time course, and therefore can not be discarded on the basis of a high regression standard error. Of note, these rare events could even make deleterious mutations appear neutral or beneficial. To address this issue, we systematically compared all segments for each locus, and removed the internal segments deviating more than 1% in fitness from the locus average. N- or C-terminal segments were retained on the grounds that deviations may represent polar or change-of-function effects (but to retain them when beneficial, we required other beneficial mutations were found in the same operon).

Statistical methods

All statistical tests were performed using either built-in functions or publicly available packages from the R programming language. To fit linear models we used the built-in lm() function. We used the package "Matching" to implement a bootstrap version of the Kolmogorov-Smirnov test, which corrects for the fact that the empirical distributions of fitness effects are discrete in nature *(5*). We used the package "Fitdistrplus" to find which of the distributions commonly used in the literature provided the best fit to the beneficial tails (table S2) *(6*). For all tests conducted on the deleterious or beneficial tail, we considered only the values outside the range of neutrality, as established by looking at the DFE for the set of presumed neutral loci (Fig. S7). Finally, to test whether beneficial mutations in the ancestor tended to cluster together in operons more than expected at random, we sampled without replacement an equal number of loci from the reference genome and computed the frequency of these having another member from the same transcription unit in the sample. The reported p-value is the proportion of cases out of 100 simulations that show higher or equal frequency of clustered loci than the observed data. Operons harboring mild effect mutations are expected to be underrepresented because most of these mutations may fall below the threshold of neutrality. To avoid this bias, we focused the analyses in the 50% most beneficial mutations observed. The list of transcriptional units for the ancestor was retrieved from the BioCyc database.

DFE-based predictions versus allele prevalence in the LTEE

The metagenomic time series of the LTEE were obtained from https://github.com/benjaminhgood/, and when needed genomic positions were used to standardize the annotated loci labels. For every non-synonymous allele observed in the metagenomic data, we recorded the number of lineages in which it was detected, the maximum frequency it reached, and the time point at which this maximum was attained. Six of the LTEE lineages evolved hypermutability at different generations, which had the effect of introducing a large number of "passenger" mutations (*i.e.*, non-beneficial mutations that reach high frequency due to their association with a beneficial, "driver" mutation) *(7*). Since passengers typically appear as singletons, we restricted our analyses to alleles that were observed in at least 2 independent lineages. In addition, to further ensure that these alleles represented *bona fide* targets of selection, we only considered those reaching a frequency of at least 50% at some point. When attempting to correlate the abundance of these alleles with our fitness data, we realized that working at the sublocus or even at the locus level gave rise to sparse data and therefore low statistical power to identify any patterns. To mitigate this issue, we performed the analyses at the operon level: fitness was computed as the maximum value recorded for any of its constituents, and target size as the sum of constituent lengths (excluding those without beneficial effects). We followed the same approach when we looked at how well the DFEs can predict the subsequent steps of adaptation. To compute the fraction of drivers captured by the DFE, we retrieved the alleles reaching ≥50% frequency at a given time point and counted how many of these belonged to operons present in the beneficial tail of the DFE we studied. In the analyses with the ancestor, we required the allele to reach the 50% threshold in at

least 2 lineages simultaneously. In analyses with the 2K and 15K backgrounds, we only considered alleles surpassing the threshold in their own lineage (Ara+2 or Ara-1), discarding those that were already fix in the specific strains we used (table S1). Note that despite the focus in a single lineage, these mutations still come from the filtered “driver” set, so that lineage-specific singletons were not considered. To compute a null expectation for these analyses, we ran them on random samples taken from the neutral and deleterious fractions of each DFEs. The samples were drawn without replacement and of size equal to the beneficial mutation set to which they were compared. The shaded areas shown in Figures 4, S3 and S4 represent the average of 50 simulations ± three times the standard error of the mean.

Supplementary Figures

Fig. S1. Procedure for capturing the transposon flanking regions. (A) Genomic DNA is extracted in bulk, digested with *MmeI*, and separated by gel electrophoresis. **(B)** Transposon-sized fragments are ligated with double-stranded oligonucleotide adapters. **(C)** Two rounds of PCR are performed to add indexes for Illumina sequencing.

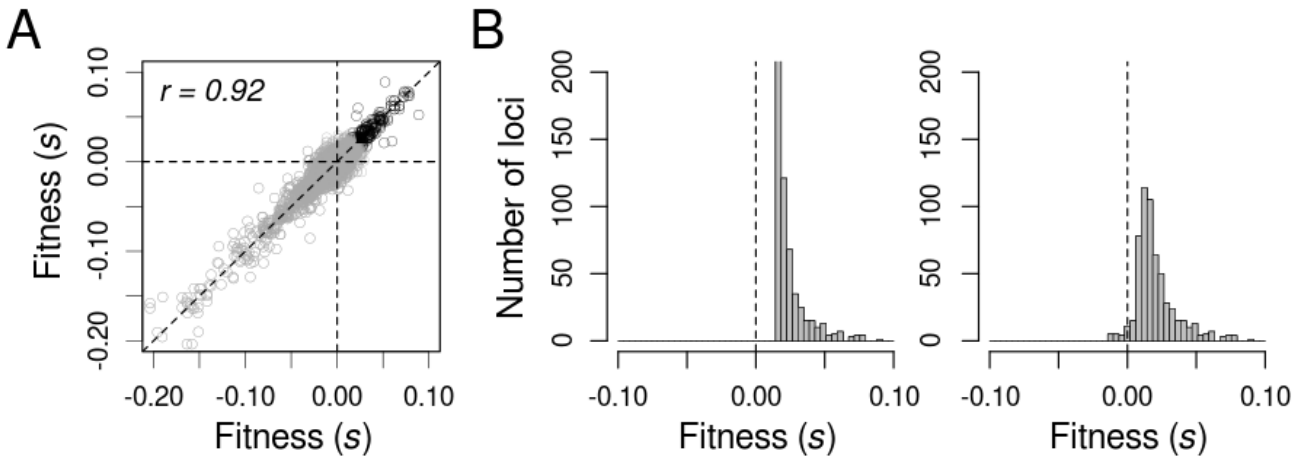

Fig. S2. Reproducibility of the fitness estimates (A) Pairwise comparisons among segments from the same locus show high reproducibility. Pearson's coefficient, *r*, is shown in upper left corner. **(B)** Despite measurement noise, the shape of the DFE for beneficial mutations (left panel) retains a marked shift towards positive values when constructed using different segments from the same locus (right panel).

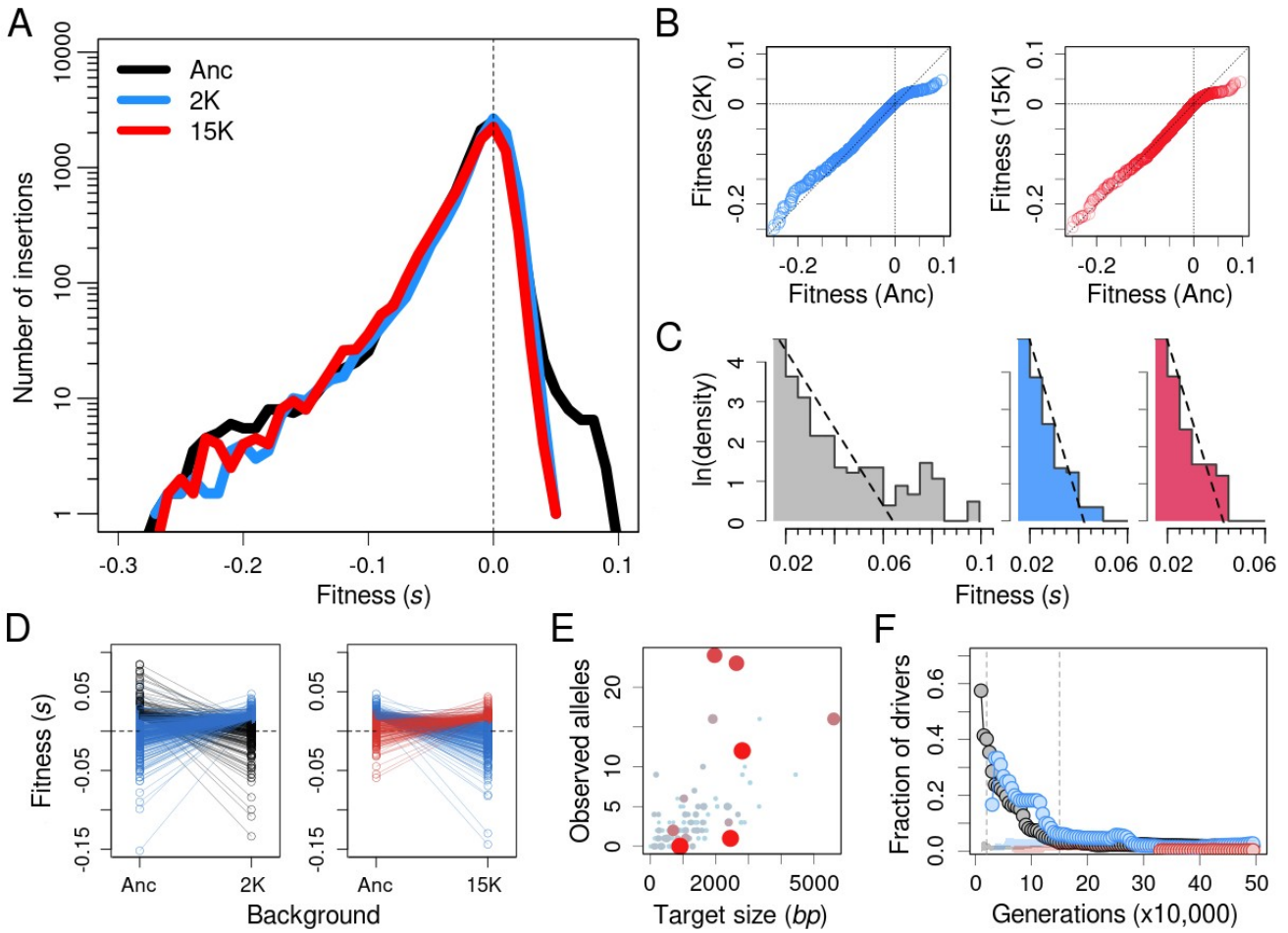

Fig. S3. Change in the DFE along a second large fitness gradient. (A) DFEs in the ancestor (black), 2K (blue) and 15K (red) evolved strains from population Ara–1. Note the logarithmic scaling of the y-axis. **(B)** Deleterious tails remained unchanged during adaptation, as indicated by comparing the cumulative distributions between the ancestor and 2K (left) and the ancestor and 15K (right). **(C)** Beneficial tails are rapidly truncated and become exponentially distributed. Histograms show the best fit to an exponential distribution (dashed line) in the ancestor (gray), 2K (blue) and 15K (red) backgrounds. **(D)** Most beneficial mutations available to the ancestor become neutral or deleterious in the 2K background, while beneficial mutations in the 2K background were neutral or deleterious in the ancestor (left). The same general pattern occurs when comparing beneficial mutations in the 2K and 15K backgrounds (right). **(E)** The prevalence of mutations in the LTEE is correlated with the mutational target size (area and color of dots represent fitness). **(F)** The predictive capacity of the DFEs as a function of time in the LTEE. Colors indicate the DFE measured in the ancestor (black), 2K (blue) and 15K (red) evolved strains. Shaded areas show the null expectations based on randomly sampling neutral and deleterious mutations; these expectations are lower than for population Ara+2 (Fig. 4D) as a consequence of the lower coverage, and hence lower resolution at the sub-genic level, of the dataset for population Ara–1 (Fig. S5). The results shown here in all panels **(A-F)** for LTEE population Ara–1 are otherwise similar to those seen for the independently evolving Ara+2 population (see Fig. 2A-C, Fig. 3C, and Fig. 4A,D).

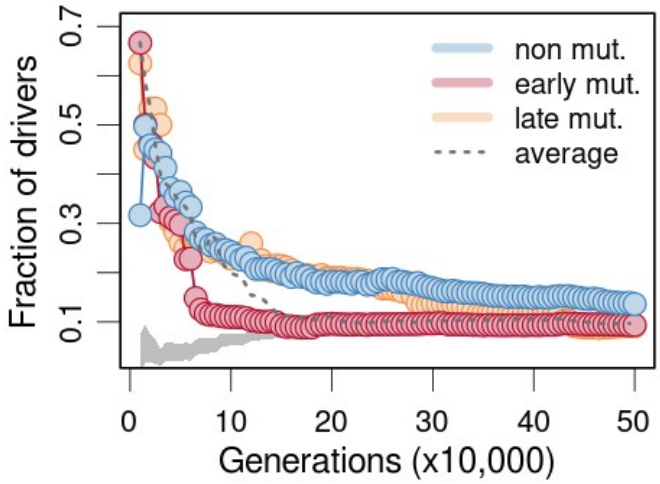

Fig. S4. Determinants of evolutionary outcomes across different sets of the LTEE lineages. The predictive capacity of DFEs as a function of time in the LTEE. Colors correspond to analyses performed separately using data from the non-mutator (blue), early mutator (orange), and late mutator lineages (red). The dashed line shows the aggregate average when pooling all lineages, as shown in Fig. 4D. The shaded area shows the null expectation based on randomly sampling neutral and deleterious mutations.

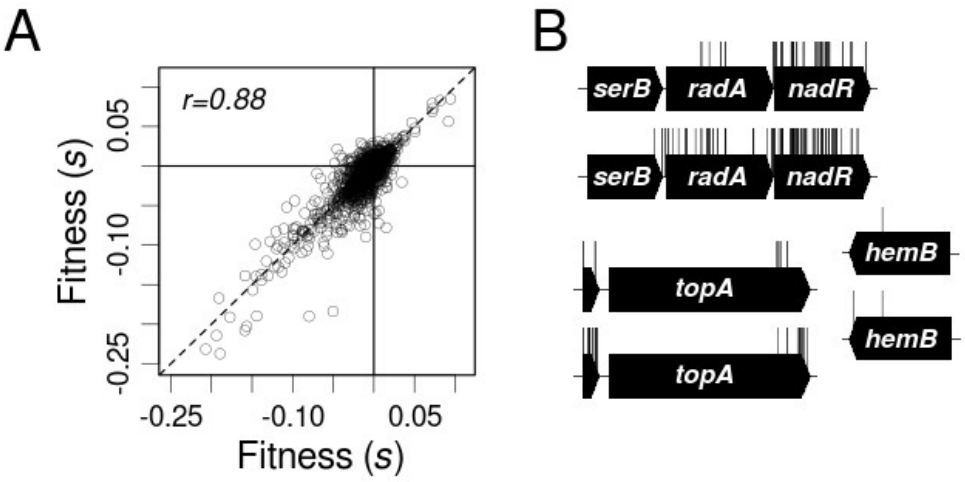

Fig. S5. Comparison of fitness estimates between two datasets. (A) Correlation between selection coefficients measured in the ancestral genetic background in the independent datasets for the LTEE lineages (Ara+2 on the x-axis; Ara–1 on the y-axis). Pearson's coefficient, *r*, is shown in the upper left corner. **(B)** The lower coverage of the Ara–1 dataset resulted in lower resolution at the sub-genic level. This difference is best illustrated by mapping insertions in essential genes for which insertions in the C-terminal end are neutral (*hemB*) or even beneficial (*serB, topA*). The diversity of insertion sites in these loci is markedly lower in Ara–1 (top alignments) than Ara+2 (bottom alignments).

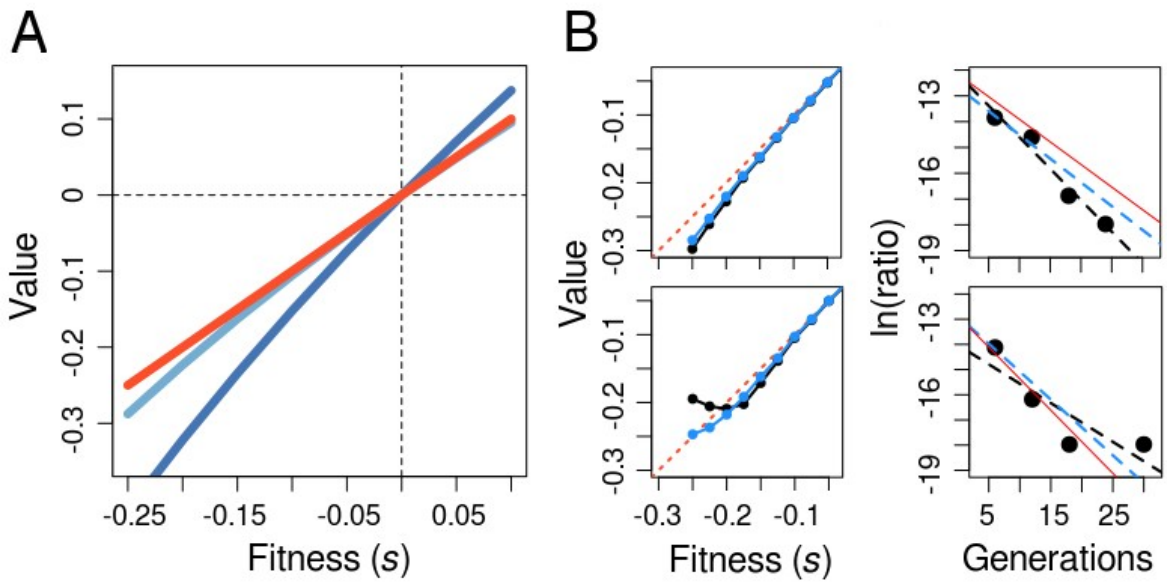

Fig. S6. Measuring fitness effects from bulk competition data. (A) Relationship between the fitness effect (red; often called the selection coefficient, *s*, in population-genetics literature) and the ratio of realized growth rates (dark blue; typically used as the measure of relative fitness, *W*, in the LTEE). Empirical estimates of *s* based on the slope of allele frequency trajectories make the approximation $s \approx \ln(1 + s)$, which equals *W* multiplied by a factor of ln(*2*) (light blue) *(4)*. Note, however, that this approximation starts to deviate at large values of *s*. **(B)** Weighting the regression by the abundance of each allele improves estimation of *s*. This improvement is most important in the lower range of *s*, when the abundance of a deleterious allele at later time points approaches zero. Lines show the true *s* (red), standard regression (black), and weighted regression (blue). Top left panel shows results from a set of simulations with a single deleterious allele starting at a frequency of 7×10^{-7}, and the top right panel shows a single run with $s = -0.18$. Bottom left panel shows results from simulations in which the deleterious allele started at a frequency of 2×10^{-7}, and the bottom right panel shows a single run with $s = -0.25$. Each dot in the two left panels shows the average of 100 simulations at each value of *s*. Simulation model adapted from *(8)*, using a total population size of 6×10^{7}.

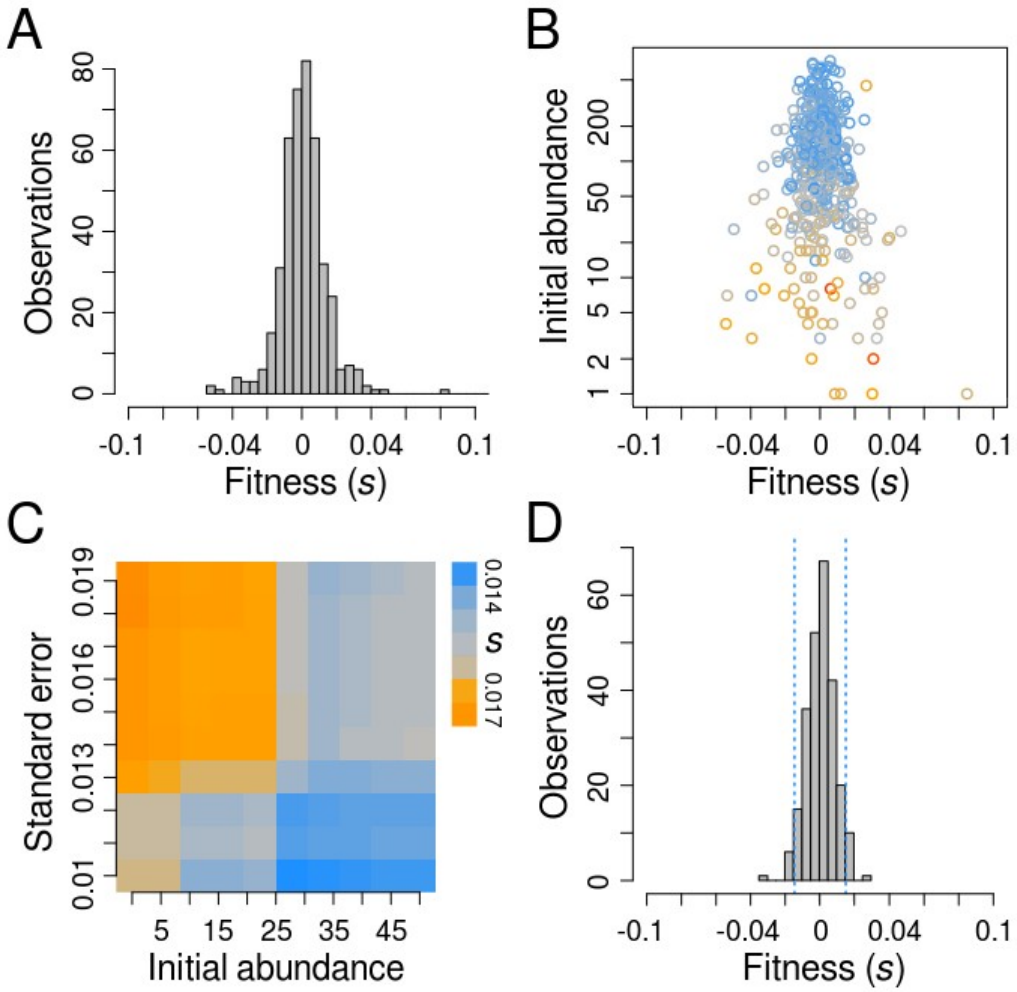

Fig. S7. Data filtering and the threshold of neutrality. (A) Raw distribution of selection coefficients for presumed neutral loci in datasets from the ancestor and both evolved Ara+2 strains. **(B)** Most small deviations from zero can be explained by sampling error, while larger deviations result from low initial abundance (y-axis, note logarithmic scaling) or poor linear fit (colors from blue to gray to orange indicate increasing standard errors). **(C)** After filtering outliers (Fig. S2), we scanned combinations of initial abundance and standard error to identify selection coefficients encompassing 97% percent of the data (colors, see scale bar on right). **(D)** Final distribution of selection coefficients for presumed neutral loci after applying all the filters. The blue dotted lines indicate the resulting threshold for approximate neutrality ($-0.015 < s < 0.015$).

Supplementary Tables

Table S1. Strains used in this study

Name	strain ID	mutated loci	source
Ancestor	REL606	None (*ara*- wild-type)	(*3*)
Ara+2 2K	REL1159A	*pykF, wcaM, infB, nfuA/gntT, pitA, spoT, hsrA, ECB_03710, rbsD-rbsB*	(*9*)
Ara+2 15K	REL7184A	*lon, ybaL, nmpC, lrp, rpsA, appC, flgB, rnb, speG, pykF, lpp-ydiM, lpp-ydiM, wzy-wcaM, wcaM, nupC/yfeA, purL, serV, yqeB, ECB_02797/ECB_02822, mreB, fis/yhdJ, spoT, rbsD-rbsA, ilvL, yihP, hslU, iclR, metH, sgcR-fimB*	(*9*)
Ara-1 2K	REL1164A	*ECB_00736/ECB_00737, topA, pykF, spoT, glmU/atpC, rbsD-yieO*	(*10*)
Ara-1 15K	REL7177A	*mokC/nhaA, pcnB, araJ, ybaL, nmpC-ECB_00513, mrdA, nagC, ompF/asnS, dhaM, ldrC/chaA, narI/ECB_01206, topA, pykF, ynjI, yedW/yedX, manB-cpsG, yegI, ECB_02816, yghJ, ebgR, infB, arcB, gltB, yhdG/fis, rpsM, malT, glpE, spoT, glmU/atpC, rbsD-yieO, hslU, pflC, iclR, fimA, nadR*	(*10*)

Table S2. Fits of the beneficial tail to common distributions.

One-sample Kolmogorov-Smirnov test, p-values						
Background	Exponential	Gamma	LogNorm	Weibull	Normal	Logistic
Anc	0.012	0.059*	$<10^{-4}$	**0.167***	$<10^{-4}$	$<10^{-4}$
2K	**0.716***	0.814*	$<10^{-4}$	0.820*	$<10^{-4}$	$<10^{-4}$
15K	**0.984***	0.939*	0.005	0.971*	$<10^{-4}$	$<10^{-4}$
Goodness-of-fit, Akaike's Information Criterion						
Background	Exponential	Gamma	LogNorm	Weibull	Normal	Logistic
Anc	-8151	-8166	-8022	**-8174***	-6804	-7186
2K	**-6114***	-6113	-5948	-6112	-5520	-5601
15K	**-4345***	-4343	-4288	-4343	-3915	-3985

Table S3. Multiple linear model for the prevalence of mutations in the LTEE.

Variable	Estimate	Std. Error	t value	P(>|t|)
(Intercept)	-0.918	0.542	-1.693	0.092
Anc fitness	53.175	18.792	2.830	**0.005***
2K fitness	55.810	21.761	2.565	**0.011***
15K fitness	37.444	22.722	1.648	0.101
target size	0.002	$< 10^{-3}$	14.764	**$<10^{-15}$***

Table S4. Primers used in this study

Name	sequence
MS-custom-adaptor-F	5'-TTCCCTACACGACGCTCTTCCGATCTNN-3'
MS-custom-adaptor-R	5'-AGATCGGAAGAGCGTCGTGTAGGGAA-3'
HS-custom-adaptor-F	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNN-3'
HS-custom-adaptor-R	5'-CTGTCTCTTATACACATCTGACGCTGCCGACGA-3'
MS-junction-PCR-F	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNTTCCCTACACGACGCTCTTCCGATCT-3
MS-junction-PCR-R	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNAGACCGGGGACTTATCATCCAACCTGT-3
HS-junction-PCR-F	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'
HS-junction-PCR-R	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNAGACCGGGGACTTATCATCCAACCTGT-3'

Supplementary References

1. A. L. Goodman *et al.*, *Cell Host Microbe*. **6**, 279–289 (2009).
2. L. Ferrières *et al.*, *J. Bacteriol.* **192**, 6418–6427 (2010).
3. R. E. Lenski, M. R. Rose, S. C. Simpson, S. C. Tadler, *Am. Nat.* **138**, 1315–1341 (1991).
4. L.-M. Chevin, *Biol. Lett.* **7**, 210–213 (2011).
5. J. S. Sekhon, *J. Stat. Softw.* **42**, 1–52 (2011).
6. M. L. Delignette-Muller, C. Dutang, *J. Stat. Softw.* **64**, 1–34 (2015).
7. A. Couce *et al.*, *Proc. Natl. Acad. Sci.* **114**, E9026–E9035 (2017).
8. A. Couce, O. Tenaillon, *Nat. Commun.* **10**, 3114 (2019).
9. O. Tenaillon *et al.*, *Nature*. **536**, 165–170 (2016).
10. J. E. Barrick *et al.*, *Nature*. **461**, 1243–1247 (2009).