

Benefit of transferred mutations is better predicted by the fitness of recipients than by their ecological or genetic relatedness

Yinhua Wang^{a,1}, Carolina Diaz Arenas^{a,2}, Daniel M. Stoeberl^{a,3}, Kenneth Flynn^{b,4}, Ethan Knapp^c, Marcus M. Dillon^b, Andrea Wünsche^a, Philip J. Hatcher^d, Francisco B.-G. Moore^c, Vaughn S. Cooper^{b,e}, and Tim F. Cooper^{a,5}

^aDepartment of Biology and Biochemistry, University of Houston, Houston, TX 77204; ^bDepartment of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, NH 03824; ^cDepartment of Biology, University of Akron, Akron, OH 44325; ^dDepartment of Computer Science, University of New Hampshire, Durham, NH 03824; and ^eDepartment of Microbiology and Molecular Genetics, University of Pittsburgh, Pittsburgh, PA 15219

Edited by Nancy A. Moran, University of Texas at Austin, Austin, TX, and approved March 22, 2016 (received for review December 17, 2015)

The effect of a mutation depends on its interaction with the genetic background in which it is assessed. Studies in experimental systems have demonstrated that such interactions are common among beneficial mutations and often follow a pattern consistent with declining evolvability of more fit genotypes. However, these studies generally examine the consequences of interactions between a small number of focal mutations. It is not clear, therefore, that findings can be extrapolated to natural populations, where new mutations may be transferred between genetically divergent backgrounds. We build on work that examined interactions between four beneficial mutations selected in a laboratory-evolved population of *Escherichia coli* to test how they interact with the genomes of diverse natural isolates of the same species. We find that the fitness effect of transferred mutations depends weakly on the genetic and ecological similarity of recipient strains relative to the donor strain in which the mutations were selected. By contrast, mutation effects were strongly inversely correlated to the initial fitness of the recipient strain. That is, there was a pattern of diminishing returns whereby fit strains benefited proportionally less from an added mutation. Our results strengthen the view that the fitness of a strain can be a major determinant of its ability to adapt. They also support a role for barriers of transmission, rather than differential selection of transferred DNA, as an explanation of observed phylogenetically determined patterns of restricted recombination among *E. coli* strains.

epistasis | mutation-by-genetic background interactions | beneficial mutations | horizontal gene transfer

Mutations can interact with one another and with their broader genetic background to affect fitness, a phenomenon known as epistasis (1). Epistasis plays a key role in many aspects of biology, including theories of speciation (2, 3), the evolution and maintenance of sex (4, 5), adaptation (6–8), and evolutionary contingency (9, 10). Whereas early experimental studies focused on interactions between deletion or other knock-out mutations, advances in genomic technologies now allow direct tests of interactions between spontaneously occurring beneficial mutations (11–16). Studies that manipulate beneficial mutations have the potential to identify general patterns that may underlie some degree of predictability in adaptive evolutionary outcomes. For example, interactions between beneficial mutations often follow a pattern of diminishing returns epistasis, such that the marginal benefit of additional mutations declines with the fitness of the recipient genotype (11–16). Diminishing returns epistasis reflects a global interaction between mutations that interact at the level of fitness (16). This pattern is consistent with the frequent observation of decelerating fitness trajectories as populations adapt to constant environments and the observation of a negative relationship between the starting fitness of a population and its initial rate of fitness increase (17–19). In apparent contrast, some theoretical work predicts an excess of positive interactions along

adaptive trajectories, highlighting the need for continued research in this area (20–22).

Most studies that have examined interactions that affect beneficial mutations have focused on interactions between mutations arising in a single population or in replicate populations evolved from a common ancestor (11, 12, 14, 23, 24). In these cases, there are relatively few mutational differences separating different genotypes, limiting our ability to extrapolate findings to an understanding of the influence of mutation interactions on adaptive evolution in genetically diverse natural populations. Indeed, introgression experiments indicate the importance of the broader genetic background in determining the effect of specific genetic regions (reviewed in ref. 25), and even the interaction between two focal mutations depends strongly and perhaps unpredictably on the genetic background in which they are measured (15).

A better understanding of the interaction between beneficial mutations and diverse genetic backgrounds will allow us to address whether mutations selected in one background will tend to have similar effects across different backgrounds. This question

Significance

The fitness effect of many mutations depends on the genotype of the individual in which they occur. Are these dependencies predictable? Do dependencies build on existing variation between individuals to promote divergence, or do they act to favor genetic cohesion? We examine these questions by measuring the fitness effect of mutations that conferred a benefit in a laboratory-evolved population when transferred into genetically and phenotypically diverse natural isolates of the same species. We found that fitness effects were predicted by the fitness of the strain to which they were added but not by the genetic or ecological relationship of the recipient strains. This pattern extends findings that the current fitness of strain is a major predictor of its ability to adapt.

Author contributions: F.B.-G.M., V.S.C., and T.F.C. designed research; Y.W., C.D.A., D.M.S., K.F., E.K., M.M.D., A.W., and P.J.H. performed research; Y.W., C.D.A., K.F., P.J.H., and T.F.C. analyzed data; and Y.W., F.B.-G.M., V.S.C., and T.F.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Dryad Digital Repository database, datadryad.org (doi: 10.5061/dryad.t60n5).

¹Present address: Tianjin Key Laboratory of Animal and Plant Resistance, College of Life Sciences, Tianjin Normal University, Tianjin 300387, China.

²Present address: Nucleic Acid Chemistry and Engineering, Okinawa Institute of Science and Technology, Okinawa, 904-0495, Japan.

³Present address: Department of Biology, Harvey-Mudd College, Claremont, CA 91711.

⁴Present address: RedOwl Analytics, Baltimore, MD 21230.

⁵To whom correspondence should be addressed. Email: tfcooper@uh.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1524988113/-DCSupplemental.

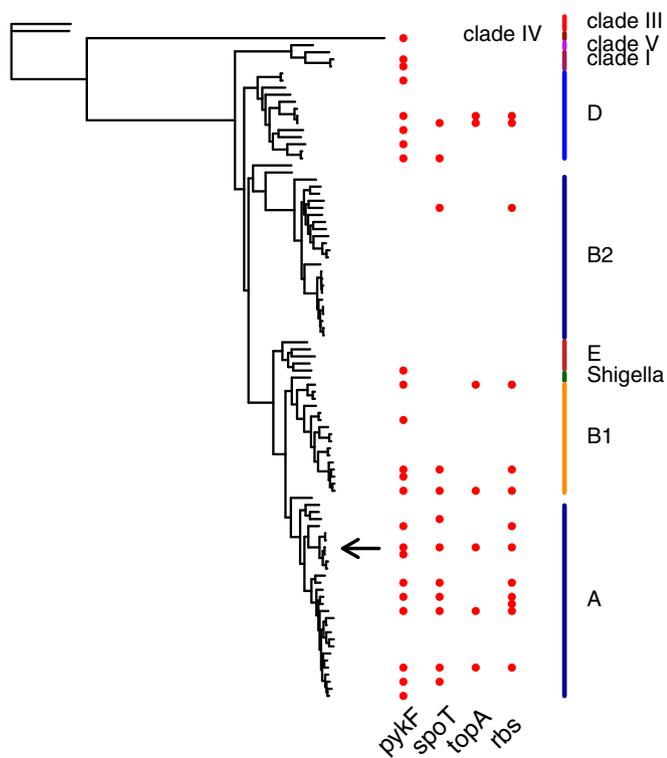


Fig. 1. Phylogeny of 96 *E. coli* strains based on their core genome (see *SI Materials and Methods* for details). Recipients of beneficial mutations are indicated (red symbols). The arrow indicates strain REL606, the ancestral strain in which mutations were initially selected as part of a laboratory evolution experiment (53). Clade and phylogroup assignments are based on previous classifications (57–60).

is particularly important in considering the evolution of bacterial populations in which the horizontal transfer and integration of short homologous DNA sequences—typically on the order of ~50–500 bp in *Escherichia coli* (26, 27)—means that the fate of new beneficial mutations will depend strongly on their interaction with the background of recipient strains. If the influence of interactions tends to be small, the effect of a beneficial mutation will be mostly independent of its genetic background, and it can spread broadly. If interactions are common, they may prevent the spread of potentially beneficial mutations between lineages. This kind of barrier has been proposed as a component of a bacterial species concept (28).

To the extent that mutation effects differ across genetic backgrounds, it is of interest to identify attributes of those backgrounds that might explain those differences. At least three candidate mechanisms have been presented: (i) The potential for differences in the specific genetic interactions that influence the effect of a new mutation increases with genetic distance. For this reason, closely related strains are expected, on average, to respond more similarly to the same new mutation than are divergent strains (2). (ii) Genetically divergent lineages can convergently evolve similar underlying genetic architectures through, for example, selection in similar ecological niches (29). For example, a comparison of bacterial metabolic networks revealed a correlation between metabolic network architecture and ecological profile that is not explained by phylogenetic relatedness (30, 31). To the extent that genetic architecture influences the effect of mutations, organisms that have similar architectures may respond similarly to an introduced mutation even if they are not genetically closely related. (iii) The relative benefit of a mutation transferred into different lineages is determined by the initial fitness of that lineage. This possibility stems from the frequent finding of globally determined fitness-dependent negative interactions between beneficial mutations

(11, 32, 33). Finally, interactions between mutations and the broader genetic background may be essentially unpredictable, perhaps depending on relatively small numbers of large effect interactions.

To examine these possibilities, we transferred four beneficial mutations that fixed in a long-term laboratory-evolved *E. coli* population to a diverse set of *E. coli* natural isolate strains (12). By measuring the fitness effect of many mutation–strain combinations, we determined the effect of interactions between the focal mutations and recipient genetic backgrounds. We also examined whether the fitness effects of the beneficial mutations were correlated with phylogenetic relatedness and/or similarity of diet niche profile—which we use as a proxy for ecological similarity—between the original and recipient genetic backgrounds. We found that the fitness of all mutations depended strongly on genetic background. This effect was not explained by differences in the genetic relatedness of the recipients and was only weakly correlated with their ecological similarity. In contrast, there was a significant relationship between mutational effects and the initial fitness of the recipient strains. This relationship followed a pattern of diminishing returns whereby more fit strains benefited less from addition of the mutations.

Results

Mutation Effects Depend Strongly on Genetic Background. We added each of four laboratory-selected mutations—in the genes or gene regions *pykF*, *rbs*, *spoT*, and *topA*—into a series of recipient naturally isolated strains of *E. coli* and measured their effect on strain fitness in the environment in which they were originally selected. Recipient strains are presented in Fig. 1 and Fig. S1 in the context of a broader phylogeny of *E. coli* derived from genome sequences of 96 diverse *E. coli* strains (Tables S1 and S2, and see *SI Materials and Methods* for details). For all mutations, recipient strains represent a random sample with respect to the complete strain set (Fig. S2).

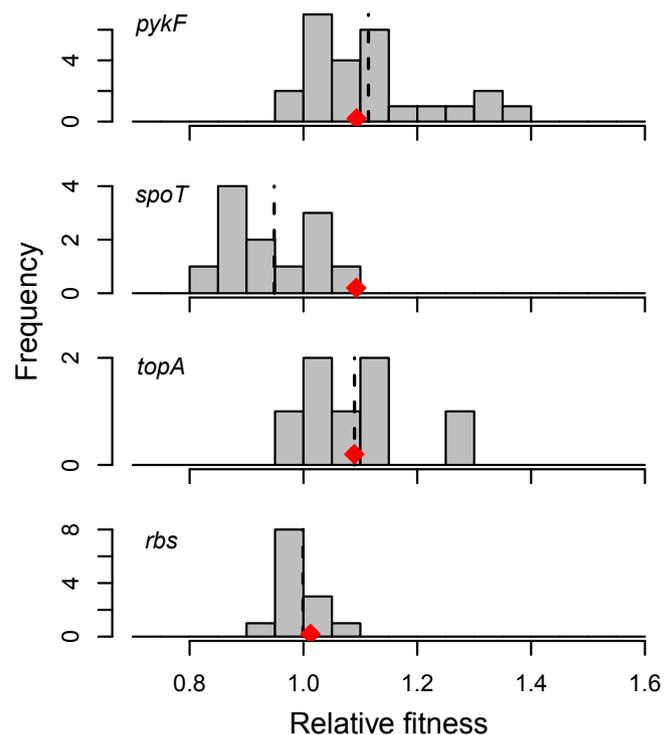


Fig. 2. Histograms of fitness effects of transferred mutations in different strains. Fitness of each mutation in the strain in which it was originally selected is indicated by the red symbol. Dashed lines indicate the mean effect of each mutation.

Table 1. Summary of differences in effect of each transferred mutation across recipient strains

Mutation	Range	F	df*	P
<i>rbs</i>	0.94–1.07	3.436	11,76	<0.001
<i>topA</i>	0.96–1.25	21.95	5,75	<0.001
<i>spoT</i>	0.84–1.09	11.29	10,103	<0.001
<i>pykF</i>	0.97–1.29	23.98	22,323	<0.001

*Degrees of freedom differ between mutations because they were introduced into overlapping but not identical sets of strains (Fig. 1 and see [SI Materials and Methods](#) for details).

We initially asked if a given mutation conferred the same effect in different recipient strains. In other words, do mutation-by-genetic background interactions influence the fitness effect of a potentially beneficial mutation? In all cases, these interactions were statistically significant, revealing a biologically meaningful range of background-dependent fitness effects (Fig. 2, Table 1, and [Table S1](#)). For example, fitness effects conferred by the *pykF* mutation, which was introduced into 23 recipient strains, ranged from –3% to 29%, compared with an effect of 9% in the strain in which it originally evolved. Only *spoT* had significantly deleterious effects in any recipient strain. Although it conferred a fitness benefit of 9% in the strain in which it was originally selected, it was deleterious in seven, neutral in three, and beneficial in only one recipient strain ([Table S3](#)).

Next we asked if different mutations tended to confer similar relative effects when they were added to the same strain. Few strains received all four mutations, so to maximize the number of recipient strains that could be considered, we tested each pair of mutations separately. We found that for four of six mutation pairs (all except *topA–pykF* and *spoT–pykF*), the relative benefit conferred by transferred mutations was mostly determined by the specific recipient strain rather than by the identity of the mutation or by the interaction between strain and mutation (Fig. 3 and [Fig. S3](#)). This result is consistent with the existence of a general strain-dependent mechanism that determines how much a strain can benefit from a transferred mutation.

Mutational Effects Are Not Well Explained by Phylogeny. We sought to test candidate factors that might explain some of the variation in mutation effect across recipient backgrounds. Specifically, we tested for a relationship between ecological, genetic, and growth attributes of recipient natural isolate strains and the effect of each introduced mutation. Where appropriate, we did this by taking into account the phylogeny of the recipient strains.

The hypothesis that closely related strains are more likely to share genotype-by-mutation interactions, and thus respond similarly to introduction of a new mutation, leads to two testable predictions. First, there will be some phylogenetic signal in the distribution of a mutation's fitness effect across multiple recipient strains. Second, there will be a negative relationship between the fitness effects of transferred mutations and the genetic distance of recipient strains relative to the strain in which they were originally selected.

To test for a general influence of a phylogenetic signal on fitness effects, we estimated Pagel's λ , a measure of phylogenetic signal in a response variable, considering the fitness effect of each mutation in the context of phylogenies created based on both core and accessory genomes (34). Values of λ greater than 0 indicate some amount of phylogenetic signal such that the fitness effect of a transferred mutation is more similar in related strains than expected by chance alone. Only for the fitness effects of the *pykF* mutation mapped onto the core genome phylogeny did the lower bound of the maximum likelihood estimate of λ allow us to reject our null hypothesis of there being no phylogenetic signal (core genome phylogeny: *pykF*, $P < 0.01$; *rbs*, $P = 0.69$; *spoT*, $P = 0.15$; *topA*, $P = 0.99$; accessory genome phylogeny: *pykF*, $P = 0.99$; *rbs*, $P = 0.99$; *spoT*, $P = 0.99$; *topA*, $P = 0.37$). Use of alternative measures of phylogenetic

signal gave qualitatively consistent results except that the phylogenetic signal of the *pykF* mutation was lost and the effects of the *spoT* and *topA* mutations were associated with the phylogeny by some metrics ([Table S4](#)). We consider these results to indicate a weak effect of phylogeny in determining the fitness effect of mutations transferred to our recipient strains.

Of course, the fact that we cannot reject the null hypothesis that phylogeny generally does not determine the effect of introduced mutations does not mean that some kind of phylogenetic signal does not exist. For example, it could be that mutational effects tend to be similar in strains closely related to the strain in which they were selected but are randomly distributed among more distantly related strains. To evaluate this possibility, we examined the relationship between the genetic distance of each recipient strain to the donor strain and the fitness effect of each mutation. In no case did we find any linear relationship between these variables, whether mutations were considered individually or in combination (Fig. 4 *A* and *B* and [Fig. S4](#)). Quadratic and exponential models, which can accommodate a baseline of neutral mutation effects in increasingly distantly related strains, did not provide any substantial improvement in fit ([Fig. S4](#)). We conclude that the fitness effects of transferred mutations are not well predicted by the genetic similarity of donor and recipient strains.

Mutational Effects Are Not Well Explained by Ecological Similarity.

Genotypes adapted to a similar ecological niche may be more likely to have similar underlying genetic architectures and therefore respond similarly to a new mutation, even if they are not genetically closely related. We tested this possibility in two ways. First, we examined the relationship between a proxy for ecological similarity—based on Biolog profiles—and the fitness effect of each introduced mutation. We found a general trend for more ecologically distant strains to benefit less from transferred mutations, but the relationship was weak. When mutations were considered together, there was a marginally significant negative relationship between mutational effect and ecological distance of the recipient relative to the donor strain (omitting the donor strain: $r = -0.27$, $P = 0.06$; $\rho = -0.33$, $P = 0.02$; [Fig. 4C](#)), but the relationship was not significant for any individual mutation ([Fig. S5](#)). To account for the fact that recipient strains can have the same overall similarity relative to the progenitor but be different from one another, we also tested for a signal of mutation effect in the context of a phylogeny based on Biolog profiles. In no case was any significant phylogenetic signal observed ([Table S4](#)). Together, these results indicate that current ecological similarity, as assessed by Biolog respiration profiles, does not explain differences in the fitness effect of any of the transferred mutations.

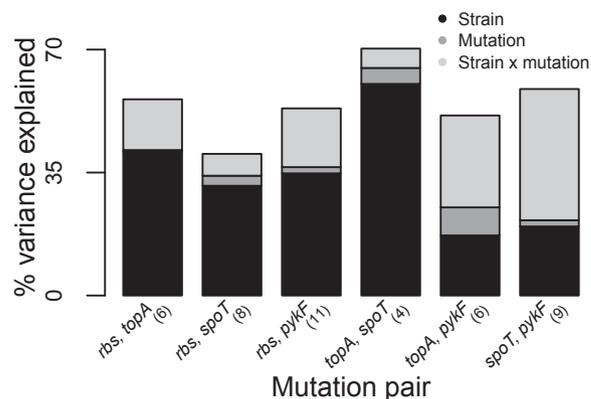


Fig. 3. Components determining the fitness effects of each mutation pair. ANOVA was used to determine variance in fitness effects explained by recipient strain, mutation, and their interaction considering those strains into which each of two mutations were added (number of strains is indicated in parentheses).

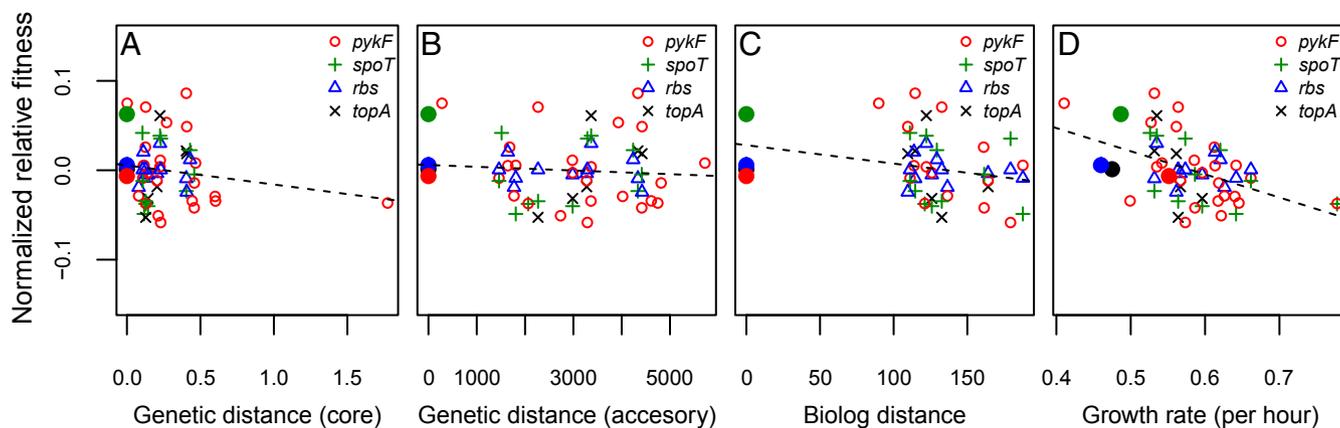


Fig. 4. Relationship between recipient strain attributes and fitness effect of added mutations. Fitness effect of added mutations is compared against core and accessory genome distance of recipients relative to the donor strain in which mutations were originally selected (A and B, respectively); ecological similarity, measured using Biolog profiles, of recipient strains against the donor strain (C); and strain growth rate (D). Solid points indicate strains in which mutations were initially selected. Fitness effects are presented as the log ratio of fitness effect of a mutation in a given strain relative to its effect in the original strain normalized so that all mutations have a mean of zero. Dashed line indicates best fit linear correlation.

Fitter Recipient Strains Benefit Less from Introduced Mutations.

Recent studies have found that epistatic interactions between beneficial mutations and their genetic backgrounds tend to become increasingly negative as the fitness of the genetic background increases (11–16). If this relationship also holds across diverse recipient strains, we predict a negative relationship between the absolute fitness of a strain (measured as growth rate) and the benefit conferred by addition of a mutation. We found that the fitness effect of two transferred mutations—in *spoT* and *pykF*—tended to decrease as the growth rate of the recipient strain increased (*spoT*: $r = -0.6$, $P = 0.04$; $\rho = -0.66$, $P = 0.02$; *pykF*: $r = -0.54$, $P = 0.01$; $\rho = -0.53$, $P = 0.01$) (Fig. S6). Negative, but nonsignificant, relationships were also found for the *rbs* and *topA* mutations (Fig. S6). As judged by comparison of Akaike information criterion (AIC) scores, no tested nonlinear relationship gave a substantially improved fit (Fig. S6). To increase our power to detect a relationship between growth rate and mutational effect, we also considered all mutations together (see *SI Materials and Methods* for details on the normalization procedure). When we do this, we find a strong overall signal of dependence of fitness effects of a transferred mutation and the growth rate of the recipient strain ($r = -0.48$, $P < 0.001$; $\rho = -0.48$, $P < 0.001$) (Fig. 4D). This relationship remains significant when two outlying strains are omitted from the analysis ($r = -0.38$, $P < 0.001$; $\rho = -0.41$, $P < 0.001$) (Fig. S6).

To test if taking phylogeny into account might influence relationships between growth rate and mutation fitness effects, we performed phylogenetic generalized least squares (GLS) analyses comparing several different models of trait evolution (Table S5). In four of eight cases, the best-fitting models did not account for the phylogenetic relationships between recipient strains. Considering the core genome, phylogeny improved the model fit for the *spoT* mutation, but the relationship between growth rate and mutation fitness effect remained significant ($P = 0.038$). In fact, growth rates of recipient strains exhibited some phylogenetic signal with respect to the core genome phylogeny ($\lambda = 0.43$, $P = 0.024$), but not the accessory genome or Biolog phylogenies, which might explain the phylogenetic signal observed for some mutation–core genome phylogeny combinations.

Interactions Between Genotype, Ecological Similarity, and Growth Rate in Explaining Fitness Effects.

Even when a factor does not individually explain a significant proportion of the variation in a response, it may still contribute in combination with other factors. To test for interactions between initial fitness, genetic and ecological similarity, and the effect of each introduced mutation, we performed partial least squares (PLS) regressions. Whereas

principal component regression seeks to determine orthogonal combinations of variables to maximize the amount of variation explained in those variables, PLS determines combinations of variables that maximize the amount of variation in the response explained. This approach is robust to having few observations relative to the number of predictor variables, as was the case especially for the *topA* mutation datasets considered here. For the *pykF*, *spoT*, and *topA* mutations, growth rate explained the largest proportion of variance in fitness effects (Fig. 5).

Discussion

Studies that have examined the effect of mutation interactions on potentially beneficial mutations have typically examined interactions that occur between a small set of focal mutations in the context of the same strain background (11, 12, 14, but see refs. 15, 35). We have extended this approach by estimating the effect of interactions between the broader genetic background and four mutations selected because they conferred a benefit in a population that was part of a long-term evolution experiment. We find that epistasis has a major influence on mutational effects. These effects depended strongly on the fitness of a recipient strain, following a pattern of diminishing returns epistasis, where the benefit of introducing a mutation declined with the fitness of the strain to which it was added. By contrast, mutational effects depend only weakly on genetic and ecological relationships between recipient strains.

We did not find a consistent phylogenetic signal in the effect of the mutations we considered, indicating that the genetic basis

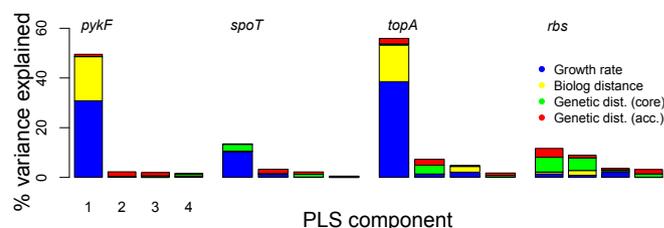


Fig. 5. PLS regression to determine contribution of growth rate, core and accessory genome distance, and biolog distance to fitness effects of each transferred mutation. The original strains are excluded from this analysis because they have a disproportionate influence on variation of genetic and ecological distance measures. Color of bars indicates the contribution of each strain attribute to that component. Only the first four components are shown.

of interactions between mutations and recipient genetic backgrounds was not well predicted by the genetic relationships used to construct our core or accessory genome phylogenies. One explanation for this finding is that there may be a small number of genes, or perhaps just sites in those genes, that interact with the transferred mutations and similarity of these genes/sites is not represented by genome-level phylogenies (36). However, our finding that mutation effects can be explained by the fitness of recipient strains, which itself does not have a strong phylogenetic signal, suggests an alternative explanation: that mutation effects are determined in large part by interactions occurring at the scale of global processes.

Mutational interactions clearly affect the specific genetic and fitness trajectory a population can follow, and recent findings of general patterns of interactions suggest that they may cause aspects of adaptive trajectories to be predictable (16, 37, 38). Several studies have found a trend toward diminishing returns epistasis between focal beneficial mutations (11, 12, 23). That is, the benefit conferred by a mutation tends to decline as it is added to more fit backgrounds. This kind of interaction can explain patterns of fitness increase and mutation accumulation in a well-studied laboratory-evolved population (16, 37, 38). Our finding of diminishing returns epistasis in the effect of selected mutations added to divergent natural isolate strains supports the idea that mutation effects often depend on some global attribute of recipient strains, rather than being determined idiosyncratically by their specific genetic background.

What could account for a strong relationship between genotype growth rate and the benefit conferred by a transferred mutation? Several studies have proposed a mechanistic basis of mutation interactions as they affect a specific enzyme or biochemical pathway (11, 32, 33, 39, 40). The bases of the beneficial effects of the four mutations we consider are not known, so we cannot offer any specific explanation that takes into account the particular mutations and backgrounds used in our study. Nevertheless, it is worthwhile to consider a general explanation that follows from the principles of metabolic control theory: that a key target of selection has a saturating form such that further improvements confer diminishing benefits. A candidate target process is translation, which constitutes a major energetic investment for bacterial cells. Cells must balance the need for sufficient translational capacity to allow expression of necessary genes, while minimizing wasteful investment in unused capacity (41). Translation provides a means to mediate interactions between seemingly disparate mutations, for example, considering three of the mutations examined in this work: Deletion of *rbs* genes might provide an advantage by reducing energy expenditure and ribosome allocation to unnecessary gene expression; *topA* impacts gene expression by altering DNA supercoiling, thereby altering promoter activity and reallocating translational resources; and *spoT* changes gene expression patterns in a way consistent with an effect on translational activity (42).

Interactions between mutations and genetic backgrounds also impact the adaptation of populations by influencing their ability to benefit from horizontally transferred mutations. Transfer of DNA that replaces a corresponding piece of the recipient genome through homologous recombination is common in many bacteria (43), including *E. coli* (26, 27, 44), and can be a major determinant of adaptation (45, 46). Horizontal transfer disconnects a beneficial mutation from the particular genetic background in which it is initially selected, so that epistatic interactions can determine both its own fate and that of recipient lineages. This process is clearly relevant to the nature of bacterial species. One bacterial species concept proposes that ecotypes—groups of genotypes that have similar ecological distribution and capability—can be recognized by being subject to purifying selective sweeps (46). Clonal selective sweeps of entire haplotypes occurring within ecotypes can create a genetic cohesion that identifies a species grouping. Genetic interactions between horizontally transferred beneficial mutations and recipient genetic backgrounds can complicate this expectation. If beneficial mutations that arise in a focal

strain tend to have similar effects in the strains that make up an ecotype, their transfer might slow clonal selective sweeps within ecotypes, instead allowing local genetic sweeps in the vicinity of the mutation and making it harder to recognize an ecotype as a genotypic cluster (47, 48). If a pattern of diminishing returns epistasis similar to the one found here applies in natural environments, this effect could be exaggerated, as less fit lineages benefit more than more fit lineages from transferred mutations, slowing clonal selective sweeps and at least transiently acting to increase ecotype genetic diversity. If a transferred mutation has beneficial effects across multiple ecotypes, its increase in frequency could produce a pattern of population-level genetic change that is hard to interpret with reference to ecotypes (49).

E. coli does appear to have some incipient ecotype structure, being composed of at least six clades (known as phylogroups) that are associated with a degree of environmental specialization (26, 50). The fact that these groups did not explain any of the variation in the fitness effects of our transferred mutations indicates the potential for gene transfer and subsequent selection to homogenize phylogroups. However, genetic diversity between sequenced strains is consistent with horizontal gene transfer being greater within than between phylogroups, and strongly selected lineages appear to have disproportionately low levels of gene transfer from other strains (27, 51). Explanations for the apparent discrepancy between potential and realized mutation transfer include a lower rate of successful gene transmission, perhaps due to lower rates of homologous recombination between more divergent strains, or ecologically mediated physical isolation of strains belonging to different phylogroups.

Our results provide direct evidence that the effects of beneficial mutations vary dramatically between divergent strains of the same bacterial species. These effects are more strongly predicted by the initial fitness of a recipient strain than by genetic or ecological similarity. This result supports an accumulating body of work consistent with an important role of global attributes, rather than specific genetic interactions, in mediating mutational effects, whereby more fit strains benefit less from a new beneficial mutation than do less fit strains. Our findings strengthen the hope of identifying general patterns of mutation interactions that lead to some predictability of adaptive evolution.

Materials and Methods

Beneficial Mutations and Strain Construction. As well as the four beneficial mutations, a mutation conferring an Ara⁻ was added into each natural isolate strain. In competition experiments, this marker allowed us to distinguish progenitor from constructed strains on tetrazolium arabinose (TA) indicator media. Details of strain constructions are presented in *SI Materials and Methods*.

Bacterial Strains and Fitness Estimates. Natural isolate recipient strains were chosen from a collection of strains sequenced as part of a Broad Institute project and from strains used in a previous study (www.broadinstitute.org/) (52) (Table S1). We sought to focus on intergene epistasis by choosing recipients that had the same amino acid sequence at the focal transferred gene as the donor REL606 strain did. For this reason, the identity of potential recipient strains differed for the mutations in *spoT* and *topA*. The *rbs* and *pykF* mutations are large deletions that result in loss of function of target gene(s). In these cases, intragene interactions are not relevant. The strains that were successfully constructed are a subset of the potential recipients because we were unable to successfully add either the focal beneficial mutation or the *araA* marker mutation into some target strains, or secondary mutations were detected during control screens. Inability to introduce a focal mutation to a strain could indicate a lethal genetic interaction, but this was the case for only one strain-recipient strain combination, so any potential bias is unlikely to be strong. Fitness effects were estimated as the ratio of Malthusian parameters (53). An alternative recommended fitness estimate, the selection rate constant calculated as the difference of Malthusian parameters, is strongly correlated (54). Raw competition counts are available from the associated Dryad Repository.

Genetic and Diet Profile Phylogeny Construction. Core and accessory genomes were estimated using whole genes and 250 base windows, identified using

PANSEQ (55), as the unit of comparison. These approaches gave similar results, and PANSEQ results are used here (Fig. S7). Core and accessory genomes were used to build phylogenies that were then used to test for a phylogenetic signal in the effect of mutations introduced to the different recipient strains. PhyML was used to build a maximum likelihood tree of the core genome. For the accessory genome, a binary input file indicating the presence/absence of each accessory gene in each strain was analyzed using default parameters of PARS in PHYLIP (56). The diet profile of strains was

measured using Biolog PM1 plates. These plates allow estimation of a combination of growth rate and respiratory activity of each strain on each of 95 distinct substrates (Biolog). A neighbor-joining tree was constructed using Biolog data using the program Neighbor in PHYLIP (56).

ACKNOWLEDGMENTS. This work was funded by National Science Foundation Grants DEB-0844355 (to T.F.C., V.S.C., and F.B.-G.M.) and DEB-1253650 (to T.F.C.).

- de Visser JAGM, Cooper TF, Elena SF (2011) The causes of epistasis. *Proc Biol Sci* 278(1725):3617–3624.
- Orr HA (1995) The population genetics of speciation: The evolution of hybrid incompatibilities. *Genetics* 139(4):1805–1813.
- Schluter D (2009) Evidence for ecological speciation and its alternative. *Science* 323(5915):737–741.
- de Visser JAGM, Elena SF (2007) The evolution of sex: Empirical insights into the roles of epistasis and drift. *Nat Rev Genet* 8(2):139–149.
- Azevedo RBR, Lohaus R, Srinivasan S, Dang KK, Burch CL (2006) Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature* 440(7080):87–90.
- Kondrashov FA, Kondrashov AS (2001) Multidimensional epistasis and the disadvantage of sex. *Proc Natl Acad Sci USA* 98(21):12089–12092.
- Woods RJ, et al. (2011) Second-order selection for evolvability in a large *Escherichia coli* population. *Science* 331(6023):1433–1436.
- Hayden EJ, Ferrada E, Wagner A (2011) Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* 474(7349):92–95.
- Weinreich DM, Delaney NF, DePristo MA, Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312(5770):111–114.
- Salverda MLM, et al. (2011) Initial mutations direct alternative pathways of protein evolution. *PLoS Genet* 7(3):e1001321.
- Chou HH, Chiu HC, Delaney NF, Segrè D, Marx CJ (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332(6034):1190–1192.
- Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332(6034):1193–1196.
- MacLean RC, Perron GG, Gardner A (2010) Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. *Genetics* 186(4):1345–1354.
- Rokyta DR, et al. (2011) Epistasis between beneficial mutations and the phenotype–fitness map for a ssDNA virus. *PLoS Genet* 7(6):e1002075.
- Wang Y, Arenas CD, Stoebel DM, Cooper TF (2013) Genetic background affects epistatic interactions between two beneficial mutations. *Biol Lett* 9(1):20120328.
- Kryazhinskiy S, Rice DP, Jerison ER, Desai MM (2014) Microbial evolution: Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* 344(6191):1519–1522.
- Couce A, Tenaillon OA (2015) The rule of declining adaptability in microbial evolution experiments. *Front Genet* 6:99.
- Barrick JE, Kauth MR, Strelhoff CC, Lenski RE (2010) *Escherichia coli* *ropB* mutants have increased evolvability in proportion to their fitness defects. *Mol Biol Evol* 27(6):1338–1347.
- Perfeito L, Sousa A, Bataillon T, Gordo I (2014) Rates of fitness decline and rebound suggest pervasive epistasis. *Evolution* 68(1):150–162.
- Draghi JA, Plotkin JB (2013) Selection biases the prevalence and type of epistasis along adaptive trajectories. *Evolution* 67(11):3120–3131.
- Greene D, Crona K (2014) The changing geometry of a fitness landscape along an adaptive walk. *PLoS Comput Biol* 10(5):e1003520.
- Blanquart F, Achaz G, Bataillon T, Tenaillon O (2014) Properties of selected mutations and genotypic landscapes under Fisher's geometric model. *Evolution* 68(12):3537–3554.
- Hall AR, MacLean RC (2011) Epistasis buffers the fitness effects of rifampicin-resistance mutations in *Pseudomonas aeruginosa*. *Evolution* 65(8):2370–2379.
- Trindade S, et al. (2009) Positive epistasis drives the acquisition of multidrug resistance. *PLoS Genet* 5(7):e1000578.
- Chandler CH, Chari S, Dworkin I (2013) Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends Genet* 29(6):358–366.
- Touchon M, et al. (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5(1):e1000344.
- Didelot X, Méric G, Falush D, Darling AE (2012) Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. *BMC Genomics* 13:256.
- Gevers D, et al. (2005) Opinion: Re-evaluating prokaryotic species. *Nat Rev Microbiol* 3(9):733–739.
- Spor A, et al. (2009) Niche-driven evolution of metabolic and life-history strategies in natural and domesticated populations of *Saccharomyces cerevisiae*. *BMC Evol Biol* 9(1):296.
- Kreimer A, Borenstein E, Gophna U, Ruppin E (2008) The evolution of modularity in bacterial metabolic networks. *Proc Natl Acad Sci USA* 105(19):6976–6981.
- Parter M, Kashtan N, Alon U (2007) Environmental variability and modularity of bacterial metabolic networks. *BMC Evol Biol* 7:169.
- MacLean RC (2010) Predicting epistasis: An experimental test of metabolic control theory with bacterial transcription and translation. *J Evol Biol* 23(3):488–493.
- Chou H-H, Delaney NF, Draghi JA, Marx CJ (2014) Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. *PLoS Genet* 10(2):e1004149.
- Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature* 401(6756):877–884.
- Pearson VM, Miller CR, Rokytka DR (2012) The consistency of beneficial fitness effects of mutations across diverse genetic backgrounds. *PLoS One* 7(8):e43864.
- Michener JK, Vuilleumier S, Bringel F, Marx CJ (2014) Phylogeny poorly predicts the utility of a challenging horizontally transferred gene in *Methylobacterium* strains. *J Bacteriol* 196(11):2101–2107.
- Wiser MJ, Ribick N, Lenski RE (2013) Long-term dynamics of adaptation in asexual populations. *Science* 342(6164):1364–1367.
- Kryazhinskiy S, Tkacik G, Plotkin JB (2009) The dynamics of adaptation on correlated fitness landscapes. *Proc Natl Acad Sci USA* 106(44):18638–18643.
- Lunzer M, Miller SP, Felsheim R, Dean AM (2005) The biochemical architecture of an ancient adaptive landscape. *Science* 310(5747):499–501.
- Dykhuizen DE, Dean AM, Hartl DL (1987) Metabolic flux and fitness. *Genetics* 115(1):25–31.
- Maitra A, Dill KA (2015) Bacterial growth laws reflect the evolutionary importance of energy efficiency. *Proc Natl Acad Sci USA* 112(2):406–411.
- Cooper TF, Rozen DE, Lenski RE (2003) Parallel changes in gene expression after 20,000 generations of evolution in *Escherichiacoli*. *Proc Natl Acad Sci USA* 100(3):1072–1077.
- Dagan T, Artzy-Randrup Y, Martin W (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc Natl Acad Sci USA* 105(29):10039–10044.
- Dixit PD, Pang TY, Studier FW, Maslov S (2015) Recombinant transfer in the basic genome of *Escherichia coli*. *Proc Natl Acad Sci USA* 112(29):9070–9075.
- Cooper TF (2007) Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biol* 5(9):e225.
- Wiedenbeck J, Cohan FM (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev* 35(5):957–976.
- Shapiro BJ, et al. (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science* 336(6077):48–51.
- Vos M (2011) A species concept for bacteria based on adaptive divergence. *Trends Microbiol* 19(1):1–7.
- Bendall ML, et al. (2016) Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J*, 10.1038/ismej.2015.241.
- Walk ST, Alm EW, Calhoun LM, Mladonicky JM, Whittam TS (2007) Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. *Environ Microbiol* 9(9):2274–2288.
- McNally A, Cheng L, Harris SR, Corander J (2013) The evolutionary path to extra-intestinal pathogenic, drug-resistant *Escherichia coli* is marked by drastic reduction in detectable recombination within the core genome. *Genome Biol Evol* 5(4):699–710.
- Moore F, Woods R (2006) Tempo and constraint of adaptive evolution in *Escherichia coli* (Enterobacteriaceae, Enterobacteriales). *Biol J Linn Soc Lond* 88:403–411.
- Lenski R, Rose M, Simpson S (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat* 138(6):1315–1341.
- Chevin L-M (2011) On measuring selection in experimental evolution. *Biol Lett* 7(2):210–213.
- Laing C, et al. (2010) Pan-genome sequence analysis using Panseq: An online tool for the rapid analysis of core and accessory genomic regions. *BMC Bioinformatics* 11(1):461.
- Felsenstein J (1989) PHYLIP—Phylogeny Inference Package (version 3.2). *Cladistics* 5:164–166.
- Snyder E, Gordon DM, Stoebel DM (2012) *Escherichia coli* lacking Rpoσ are rare in natural populations of non-pathogens. *G3 (Bethesda)* 2(11):1341–1344.
- Luo C, et al. (2011) Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proc Natl Acad Sci USA* 108(17):7200–7205.
- Gordon DM, Cowling A (2003) The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: Host and geographic effects. *Microbiology* 149(Pt 12):3575–3586.
- Walk ST, et al. (2009) Cryptic lineages of the genus *Escherichia*. *Appl Environ Microbiol* 75(20):6534–6544.
- Schneider D, Duperchy E, Coursange E, Lenski RE, Blot M (2000) Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements. *Genetics* 156(2):477–488.
- Baym M, et al. (2015) Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* 10(5):e0128036.
- Andrews S (2010) *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available at www.bioinformatics.babraham.ac.uk/projects/fastqc/.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
- Edwards DJ, Holt KE (2013) Beginner's guide to comparative bacterial genome analysis using next-generation sequencing data. *Microb Inform Exp* 3(1):2.
- Cooper VS, Vohr SH, Wrocklage SC, Hatcher PJ (2010) Why genes evolve faster on secondary chromosomes in bacteria. *PLoS Comput Biol* 6(4):e1000732.
- R Core Team (2015) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria). Available at www.R-project.org/.
- Jeong H, et al. (2009) Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J Mol Biol* 394(4):644–652.

Supporting Information

Wang et al. 10.1073/pnas.1524988113

SI Materials and Methods

Beneficial Mutations and Strain Construction. Five beneficial mutations—occurring in the genes or gene regions *rbs*, *topA*, *spoT*, *glmUS*, and *pykF*—were identified as the first to fix in a long-term evolving population of *E. coli* REL606 (12). We individually introduced four of these mutations into a series of natural isolate recipient strains (Table S1), omitting the mutation upstream of *glmUS* because of low transmission efficiency. Mutations were introduced into the chromosome of natural isolate strains using a suicide vector-based approach (12). As well as the four beneficial mutations, a mutation conferring an Ara[−] phenotype (AraA 92D; the same mutation that distinguishes REL606 from its Ara⁺ derivative, REL607) was added into each natural isolate strain. In competition experiments, this marker allowed us to distinguish progenitor from constructed strains on TA indicator media.

The evolved mutation in *pykF* involved the insertion of an *IS150* element (61). Multiple copies of this element exist in many *E. coli* genomes, making this a difficult mutation to transfer through the recombination-based approach that we used. For this reason, we instead used an in-frame 603-bp deletion *pykF* allele. These insertion and deletion mutations have the same effect on fitness in the genetic background in which the insertion allele actually occurred (12).

To control for the possibility that confounding secondary mutations occurred during strain constructions, we isolated a control strain that did not contain that replacement allele from the same construction lineage as the one that produced the successful allele replacement strain. This strain will have any mutation that occurred during construction up until the last step of the process: excision of the introduced suicide plasmid from the recipient chromosome, leaving behind the introduced (in constructed strains) or original (control strains) target allele. A strain containing the introduced allele was only used in this study if the fitness of its paired control strain was unchanged from the original progenitor strain.

Bacterial Strains. Natural isolate strains used as recipients for introduced beneficial mutations were chosen from a collection of 96 strains collected and sequenced as part of a Broad Institute project and obtained from the Michigan State University Shiga Toxin-Producing *Escherichia coli* (STEC) Center, and from strains obtained and used in a previous study by one of the authors (www.broadinstitute.org/) (52) (Table S2). Genome sequences of strains were downloaded from the Broad Institute (<https://olive.broadinstitute.org/projects/Escherichia%2520coli%2520Antibiotic%2520Resistance>) or obtained by de novo Illumina sequencing (see *Genome Sequencing*). The recipient strains used in this study are detailed in Table S1.

We sought to focus on intergene epistasis by choosing recipient natural isolate strains that had the same amino acid sequence at the focal transferred gene as the donor REL606 strain. For this reason, the identity of potential recipient strains differed for the mutations in *spoT* and *topA*. The *rbs* and *pykF* mutations are large deletions that result in loss of function of target gene(s). In these cases, intragenic interactions are not relevant, so these mutations were transferred into strains without regard to the original allele. The strains that were successfully constructed are a subset of the potential recipients because we were unable to successfully add either the focal beneficial mutation or the *araA* marker mutation into some target strains.

Fitness Competition Experiments. The fitness of constructed strains was measured using direct competition assays in the same medium used in the evolution experiment in which the mutations manipulated in this work were originally selected (12, 53). All competitions were carried out between strains with distinct Ara markers, which allowed the two types to be distinguished on TA indicator media (53). Except for control competitions, the two strains otherwise differed only by the presence or absence of an introduced mutation. Competitions were carried out in Davis Mingioli medium supplemented with 25 μg/mL glucose (DM25). Before each fitness assay, the two competitors were acclimated to the competition environment by growing them separately for two 1-d growth cycles in the same environmental conditions used in the competition. Each competitor was then diluted 200-fold into fresh DM25 medium and a sample immediately plated on TA agar to estimate the initial densities of the competing strains. At the end of 1 d of competition, a further sample was plated on TA agar to obtain the final density of each competitor. The fitness of the evolved strain relative to the ancestor was calculated as $\ln[N_C(f)/N_C(i)]/\ln[N_{WT}(f)/N_{WT}(i)]$, where $N_C(i)$ and $N_C(f)$ represent the initial and final density of the constructed clone, respectively, and $N_{WT}(i)$ and $N_{WT}(f)$ represent the initial and final density of the original wild-type clone, respectively. All assays were carried out with at least threefold replication. In some natural isolate strains, the introduced Ara[−] marker was not neutral. In these cases, we normalize relative fitness estimates to account for the marker effect. Control competitions to account for marker cost were carried out at the same time as competitions to measure the effect of the introduced mutation.

Genome Sequencing. Genomic DNA was purified using Qiagen DNeasy Blood and Tissue and Promega Wizard Genomic DNA Purification kits. Genomic DNA was eluted in nuclease-free water. The concentration of dsDNA was determined using PicoGreen. Quantified DNA was prepared for multiplexed sequencing, using Illumina Nextera kits following a modified protocol (62). In most cases, raw reads were processed using Trimmomatic to remove Nextera PE adapter sequences, and quality control was performed on all sequences with fastQC (63, 64). Reads were assembled into contigs using VELVET and contigs ordered using MAUVE following the suggested approaches outlined in ref. 65.

Phylogeny Construction. Core genes (“panorthologous” genes shared across all recipient strains) and accessory genes (genes shared among a subset of recipient strains) were identified using a previously described pipeline (66). Briefly, the pipeline first identifies putative core genes using National Center for Biotechnology Information BLASTP (release 2.2.16) to analyze the total pool of genes for sequence similarity. Homologs were identified as those gene pairs that had BLAST hits in both directions within a bit score threshold scaled by the bit score of the self-hit of the query gene. Pairs of homologous genes were grouped into families and panorthologs identified as genes from homolog families with exactly one gene from each genome. We scanned bit score thresholds of 0.1–0.9, choosing a value of 0.5, which maximized the number of panorthologs that were identified. This analysis identified a total of 1,648 panorthologs across the 27 genomes that were considered. A total of 7,017 accessory genes were identified as being present in at least one, but not all, strains. Adding subsequent stringency filters to the panortholog gene set to require a limit of <20 or <5 amino acid differences

from the consensus sequence in the trimmed alignments of each gene across all genomes reduced the size of the core genome to 1,560 or 1,442 genes, respectively, but did not qualitatively affect the outcome of any subsequent analysis.

An independent approach to identify core and accessory genome regions on the basis of shared DNA sequence windows was implemented in PANSEQ (55). Core regions were defined as regions of 250 bp present in an arbitrary reference strain that were present at a match of >80% identity in all other strains. A phylogeny was built from the core genome by concatenating core regions for each strain and performing a multiple sequence alignment. Variable sites in this alignment were extracted as a SNP file. We used PANSEQ to identify core and accessory genomes for both the 28 strains used in our experiments (27 recipients and 1 donor) and for the broader collection of 99 strains described in Bacterial Strains. Results of the PANSEQ analysis resulted in relationships between strains that were largely unchanged from those determined by the whole gene analyses (55) (Fig. S7). Core and accessory genomes determined using PANSEQ were used for all analyses presented here.

Core and accessory genomes were used to build phylogenies that were used to test for a phylogenetic signal in determining the effect of mutations introduced to the different recipient strains. PhyML was used to build a maximum likelihood tree of the core genome. For the accessory genome, a binary input file indicating the presence/absence of each accessory gene in each strain was analyzed using default parameters of PARS in PHYLIP (56).

Biolog Respiration Profile. The respiration profile of strains was measured using Biolog PM1 plates. These plates allow estimation of the respiratory activity of each strain on each of 95 distinct substrates (Biolog). Before growth in Biolog plates, each strain was grown overnight in lysogeny broth, then concentrated by centrifugation, and resuspended in PBS to an OD₆₀₀ of 0.372. An aliquot of these cells was mixed with inoculating fluid containing a dye and transferred to each well of a PM1 plate. The plate was incubated at 37 °C, and its OD₅₆₂ was measured at 2-h intervals until 12 h, 4-h intervals until 24 h, and then at 48 h. Respiration in each substrate was quantified as the area under the curve. A neighbor-joining tree was constructed using Biolog data using the program Neighbor in PHYLIP (56).

Growth Rate Analysis. The maximum growth rate of each progenitor strain was estimated by growing strains in 96-well plates containing DM medium supplemented with 500 µg/mL glucose and measuring changes in OD₄₅₀ in a VersaMax plate reader (Molecular Dynamics). All strains were preconditioned in the

assay media for two growth cycles before estimation of their growth rate. The higher concentration of glucose used in these assays compared with the competition assays was necessary for strains to achieve sufficient growth for reliable spectrophotometric detection. A custom R script was used to estimate the maximum growth rate of each strain.

Statistical Analysis. All analyses were carried out in R (version 3.1.1) (67). The nlme package was used to perform nonlinear regressions. Recipient strains and introduced mutations were treated as fixed effects because our focus was on identifying and explaining mutation–genetic background interactions between the specific strains and mutations considered here. We use PLS regression to account for correlation between combinations of variables tested to explain the effect of mutations introduced into recipient strains. Whereas principal component regression seeks to determine orthogonal combinations of variables to maximize the amount of variation explained in those variables, PLS determines combinations of variables that maximize the amount of variation in the response explained. PLS regression was implemented using the package Pcr. Phylogenetic general least squares regression analyses were performed using the functions gls and nlme in the packages Ape and Nlme, respectively. Four phylogenetic correlation matrix models were considered: (i) Brownian, which assumes that correlation of trait values is determined by shared evolutionary history; (ii) Pagel, a derivative of the Brownian model where trait correlations are scaled by a constant between 0 and 1, indicating the influence of shared, relative to independent, evolutionary history; (iii) Ornstein–Uhlenbeck, a derivative of the Brownian expectation where traits are “elastic,” having a tendency to return to a mean value; and (iv) Blomberg, a derivative of the Brownian model where the rate of trait evolution can change over time. Tests for phylogenetic signal were performed using the following functions: multiPhylosignal in the package Picante (Blomberg’s K), phylosig in the package Phytools (Pagel’s λ and Blomberg’s K), and abouheif.moran in the package Adephylo (Moran’s I and Cmean). The functions pd.calc and pd.bootstrap in the package Caper were used to compare the distance separating the strains used here to a distribution of distances between 1,000 randomly chosen sets of the same number of strains from the 96 sequenced strains contained in our overall phylogeny. For each mutation, we use the strain in which it occurred as the reference point for genetic and ecological distance regression analyses (12) (Table S1—*rbs*, REL606; *topA*, TC720; *spoT*, TC960; *pykF*, TC941). All attributes of recipient strains are presented in Table S3.

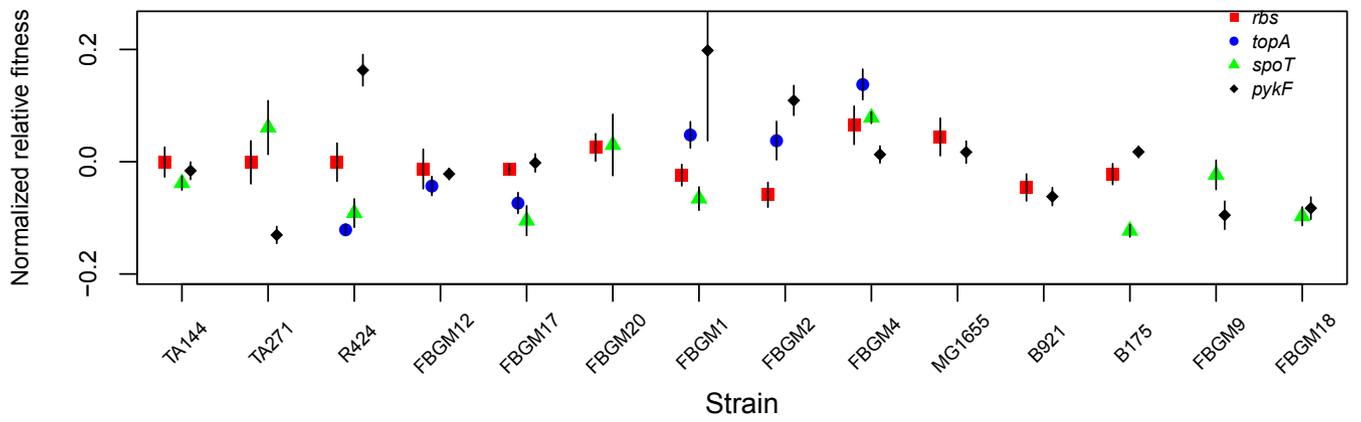


Fig. S3. Comparison of mutational effects in recipient strains recipient to multiple mutations. For each mutation–strain combination, normalized relative fitness is calculated as the log ratio of the fitness effect of a mutation to relative fitness of the same mutation in the background in which it was initially selected. These ratios are then further normalized to the same mean for each mutation. Symbols represent these estimates, and error bars indicate 95% confidence intervals.

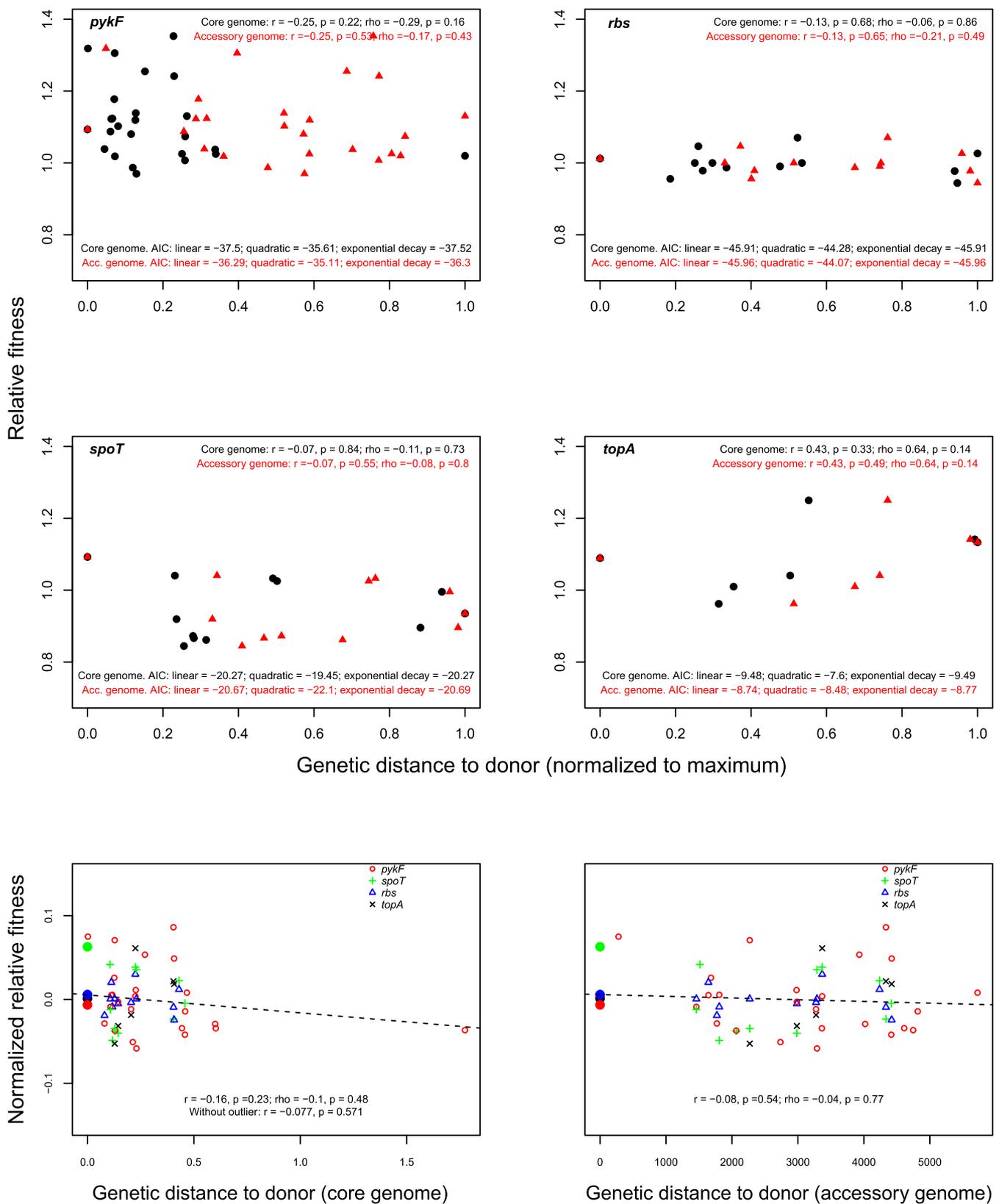


Fig. S4. Relationship between genetic distance and fitness effects of added mutations. The four *Top* panels show the relationship between the genetic distance of a recipient strain relative to the strain in which the mutation was selected (the donor strain), based on core (black symbols) and accessory (red symbols) phylogenies, and the fitness effect of the indicated mutation when added to the recipient strain. The *Bottom* two panels show the relationship between core (*Bottom Left*) and accessory (*Bottom Right*) genetic distance of recipient strains relative to the donor strain and the normalized fitness effects of all added mutations. Summary statistics of each relationship are shown in each panel.

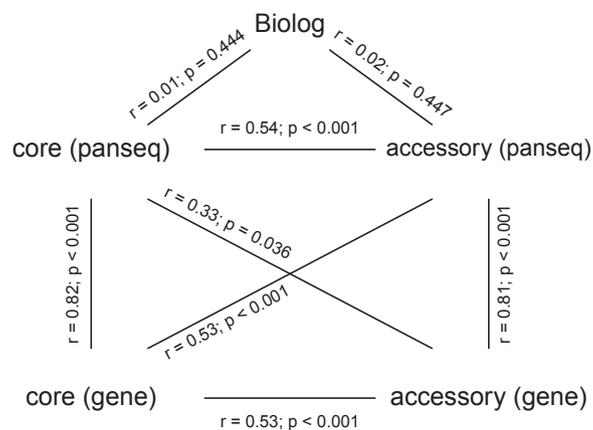


Fig. S7. Comparison of phylogenies constructed from different approaches (“gene,” whole gene, and “panseq,” DNA windows of size 500 bp) and different datasets (“core” and “accessory” genomes). Approaches and datasets are described in *SI Materials and Methods*. Mantel tests were used to compare phylogenies. Significant correlations indicate that phylogenies have a higher degree of congruence than expected by chance.

Table S1. Recipient strains used in this study

Strain ID*	Alternative ID	Source	Genome sequence
B175	TW15935	Broad Institute via MSU STEC center	Broad [†]
B706	TW15943	Broad Institute via MSU STEC center	Broad
B921	TW15945	Broad Institute via MSU STEC center	Broad
BL21		American Type Culture Collection (ATCC)	
E101	TW15947	Broad Institute via MSU STEC center	Broad
E1118	TW15949	Broad Institute via MSU STEC center	Broad
E560	TW15955	Broad Institute via MSU STEC center	Broad
FBGM1	VS-151	F.B.-G.M.	Illumina HiSeq. 2000, this study
FBGM2	VS-126	F.B.-G.M.	Illumina HiSeq. 2000, this study
FBGM4	TA135	F.B.-G.M.	Illumina HiSeq. 2000, this study
FBGM9	TA260	F.B.-G.M.	Illumina HiSeq. 2000, this study
FBGM12	VS820	F.B.-G.M.	Illumina HiSeq. 2000, this study
FBGM17	ECOR1	F.B.-G.M.	Illumina HiSeq. 2000, this study
FBGM18	ECOR11	F.B.-G.M.	Illumina HiSeq. 2000, this study
MG1655		Coli Genetic Stock Center (CGSC)	
H260	TW15964	Broad Institute via MSU STEC center	Illumina NextSeq. 500, this study
H442	TW15976	Broad Institute via MSU STEC center	Broad
M114	TW15991	Broad Institute via MSU STEC center	Broad
R424	TW15997	Broad Institute via MSU STEC center	Broad
REL606		(12)	(68)
TC720 [‡]			
TC960 [‡]			
TC941 [‡]			
TA141	TW16010	Broad Institute via MSU STEC center	Broad
TA144	TW16012	Broad Institute via MSU STEC center	Broad
TA271	TW16017	Broad Institute via MSU STEC center	Broad
TA280	TW16018	Broad Institute via MSU STEC center	Broad
TW10509	TW16023	Broad Institute via MSU STEC center	Broad

*ID is from ref. 52 or Michigan State University STEC center documentation.

[†]Broad genome sequences downloaded from <https://olive.broadinstitute.org/projects/Escherichia%2520coli%2520Antibiotic%2520Resistance>.

[‡]Strains are derived from REL606 and differ by the presence of combinations of beneficial mutations (TC720, REL606 *rhs*; TC960, REL606 *rhs*, *topA*; TC941, REL606 *rhs*, *topA*, *spoT*, *glmUS*, where genes refer to evolved alleles occurring in the Ara-1 line of the Lenski long-term evolution experiment) (12).

Table S2. List of all strains used in full *E. coli* phylogeny

Strain ID		
B088	<i>FBGM9</i>	M056
B093	H001	M114
B108	H120	M605
B175	H185	M646
B185	H218	M718
B354	H220	M863
B367	H223	M919
B574	H252	MG1655
B671	H260	R424
B706	H263	R527
B799	H288	R529
B921	H296	REL606
BL21	H299	T408
E1002	H305	T426
E101	H378	TA004
E1114	H383	TA007
E1118	H386	TA008
E1167	H397	TA014
E1492	H413	TA054
E1520	H420	TA103
E267	H442	TA141
E482	H454	TA143
E560	H461	TA144
E704	H489	TA206
<i>FBGM1</i>	H504	TA249
<i>FBGM12</i>	H588	TA255
<i>FBGM17</i>	H593	TA271
<i>FBGM18</i>	H605	TA280
<i>FBGM2</i>	H617	TA435
<i>FBGM20</i>	H660	TA447
<i>FBGM3</i>	H730	TA464
<i>FBGM4</i>	H736	TW10509

Strains in bold were obtained from the Michigan State University STEC Center, and strains in italics are described in ref. 52.

