Article

Current Biology

Rediversification following ecotype isolation reveals hidden adaptive potential

Highlights

- Related microbial communities rapidly rediversify following ecotype removal
- Competitions unveil changes in ecological dynamics and community interactions
- Rediversified ecotypes evolve distinct growth/survival traits that enable coexistence

Authors

Joao A. Ascensao, Jonas Denk, Kristen Lok, QinQin Yu, Kelly M. Wetmore, Oskar Hallatschek

Correspondence

ohallats@berkeley.edu

In brief

Ascensao et al. explore the capacity of simple microbial communities to regenerate ecological diversity after ecotype removal. Rediversified communities show major shifts in ecological dynamics and interactions. This study highlights how community structure is shaped by adaptive pathways that only become accessible after a community perturbation.

Ascensao et al., 2024, Current Biology *34*, 855–867
February 26, 2024
https://doi.org/10.1016/j.cub.2024.01.029

CellPress

Current Biology

CellPress

Article

Rediversification following ecotype isolation reveals hidden adaptive potential

Joao A. Ascensao,[1] Jonas Denk,[2,3] Kristen Lok,[1,6] QinQin Yu,[2,7] Kelly M. Wetmore,[4] and Oskar Hallatschek[2,3,5,10,*]
[1]Department of Bioengineering, University of California, Berkeley, Berkeley, CA, USA
[2]Department of Physics, University of California, Berkeley, Berkeley, CA, USA
[3]Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, USA
[4]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[5]Peter Debye Institute for Soft Matter Physics, Leipzig University, 04103 Leipzig, Germany
[6]Present address: Department of Biomedical Engineering, Duke University, Durham, NC, USA
[7]Present address: Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[10]Lead contact
*Correspondence: ohallats@berkeley.edu
https://doi.org/10.1016/j.cub.2024.01.029

SUMMARY

Microbial communities play a critical role in ecological processes, and their diversity is key to their functioning. However, little is known about whether communities can regenerate ecological diversity following ecotype removal or extinction and how the rediversified communities would compare to the original ones. Here, we show that simple two-ecotype communities from the *E. coli* long-term evolution experiment (LTEE) consistently rediversified into two ecotypes following the isolation of one of the ecotypes, coexisting via negative frequency-dependent selection. Communities separated by more than 30,000 generations of evolutionary time rediversify in similar ways. The rediversified ecotype appears to share a number of growth traits with the ecotype it replaces. However, the rediversified community is also different from the original community in ways relevant to the mechanism of ecotype coexistence—for example, in stationary phase response and survival. We found substantial variation in the transcriptional states between the two original ecotypes, whereas the differences within the rediversified community were comparatively smaller, although the rediversified community showed unique patterns of differential expression. Our results suggest that evolution may leave room for alternative diversification processes even in a maximally reduced community of only two strains. We hypothesize that the presence of alternative evolutionary pathways may be even more pronounced in communities of many species where there are even more potential niches, highlighting an important role for perturbations, such as species removal, in evolving ecological communities.

INTRODUCTION

Ecological diversification refers to the evolution of a population or community of organisms to occupy distinct ecological niches or habitats within an ecosystem.[1] Such diversification can manifest through various mechanisms, including the evolution of unique physical or behavioral traits that enable individuals to utilize diverse resources or withstand varied environmental conditions.[2–4] The propensity for ecological diversification in a community is influenced by factors like environmental conditions, prevailing biodiversity, and the interactions among the species present.[4–7] The potential for diversification can be modulated by the presence of unoccupied niches, or "ecological opportunities."[5,8] These opportunities may diminish as diversity increases and niches become occupied. However, existing communities can also generate new niches, facilitating the introduction of novel ecotypes. An example is cross-feeding, where species produce metabolites that pave the way for the rise of new ecotypes by forming exploitable niches.[9–13]

Microbial communities offer a valuable model for investigating the intertwined evolutionary and ecological processes driving diversification due to their rapid reproductive and evolutionary rates.[11,13–21] In both natural settings[22–25] and experimental systems, swift ecological diversification in microbial communities has been documented, typically propelled by mechanisms such as cross-feeding,[26–30] resource partitioning,[9,11,31–33] spatial niche differentiation,[15,34–36] and potentially other ecological trade-offs.[37,38] Interactions within microbial communities can either inhibit,[17,18] promote,[11,13] or have mixed impacts[39] on diversification. The enduring coexistence of a new ecotype with its immediate ancestor is not assured and may hinge on community characteristics, such as metabolic trade-offs.[12,40] In experimental contexts, ecological differentiation of a diversified ecotype is often indicated when an ecotype's fitness inversely correlates with its frequency, i.e., displaying negative frequency-dependent fitness effects. Stable coexistence between the diversified ecotype and its ancestor is implied if the former can invade when rare but not when abundant.

Even when ecotypes can stably coexist, it does not guarantee that they will coexist indefinitely or at all locations. Ecotypes can migrate to new territories, potentially without other community members, or some ecotypes within the community may spontaneously go extinct (e.g., due to demographic stochasticity or environmental fluctuations). In either case, the community becomes perturbed, losing one or more members and potentially leaving ecological niches unfilled. Environmental disturbances that cause ecotype loss are prevalent across diverse types of microbial ecosystems, including aquatic, soil, and human-associated environments.[41–47] Oftentimes, local ecotype/species extinction is not benign—loss of microbial taxa has been associated with deterioration of ecosystem functioning in natural systems.[48,49] It has long been noted among biologists that newly isolated species, and species extinctions, can open up ecological opportunities and lead to rapid diversification events.[50–53]

Theoretical models suggest that perturbed communities may respond with a combination of ecological and evolutionary changes.[12,54,55,56] These evolutionary changes may include both directional and diversifying selection,[56] with newly evolved variants either replacing existing community members or coexisting alongside them. However, it remains unclear which communities have the potential to rediversify. When rediversification does occur after ecotype removal, there are two possible scenarios: (1) the perturbed community rediversifies and eventually returns to a state similar to the original community before the disturbance, or (2) the perturbed community rediversifies and forms a community that is qualitatively different from the original one.

Here, we investigate the aforementioned questions surrounding rediversification using a minimal microbial model community of only two, naturally diversified *E. coli* strains. Specifically, we employ two strains derived from the *E. coli* long-term evolution experiment (LTEE), which was started by Dr. Richard Lenski and has been running for over 30 years or more than 70,000 generations.[57] An initially isogenic strain of *E. coli* was split into 12 replicate populations and propagated through daily dilutions in glucose minimal media (Davis minimal media [DM]25). At the outset of the LTEE around 6,500 generations, it was found that one lineage, ara-2, spontaneously diversified into two lineages—small (*S*) and large (*L*)—that coexist via negative frequency dependence.[16] The ecotypes were named for the sizes of their colonies on certain agar plates, either *S* or *L*. The *S* and *L* lineages inhabit distinct temporal and metabolic niches in the LTEE environment. During exponential phase, *L* grows more quickly on glucose, while *S* specializes in stationary phase survival and utilizes acetate, a byproduct of overflow metabolism.[27,58] Since their diversification, the lineages have persisted and evolved over time, exhibiting genetic, transcriptional, and metabolic divergence.[16,27,58–63] The LTEE-derived communities are ideal for our plan to investigate the possibility and potential patterns of rediversification over evolutionary time. We can revive the *S*-*L* community at 6,500 generations to probe rediversification right after emergence of the community and compare with rediversification at later stages of the evolution experiment.

We found that when we isolated the *S* ecotype under certain conditions, it would spontaneously rediversify, giving rise to a new big colony ecotype S_B, even if we used *S* clones separated by more than 30,000 generations of evolutionary time. The new ecotype, S_B, displays hallmarks of ecological differentiation, including negative frequency-dependent fitness effects when in coculture with its ancestral *S* clone. We dissected the new, rediversified community, and found that while S_B shares a number of traits with both *L* and *S*, it also behaves in entirely new ways. Our findings suggest that even in a maximally reduced community of only two strains, evolution may leave room for alternative diversification processes, suggesting a hidden adaptive potential only revealed by ecotype removal. This raises the possibility that perturbations, such as ecotype removal, could play an important role in evolving ecological communities by creating opportunities for alternative evolutionary pathways.

RESULTS

S can quickly diversify into a new ecotype

The ability of the *S* ecotype to emerge and coexist with the *L* ecotype in the LTEE has been attributed to its proficiency in scavenging acetate released from overflow metabolism during glycolysis, as well as its ability to survive and thrive during stationary phase.[27,58] It has been proposed that the *L*-*S* and similar polymorphisms may arise because of a fundamental, hard-to-break trade-off between glucose and acetate growth rates in *E. coli*.[12,55] Based on these explanations, one may suspect that after removing either *L* or *S* in the two-ecotype community the community may eventually rediversify and will eventually approach a two-ecotype community similar to the original *L*-*S* community.

We performed a simple experiment where we cultured an *S* clone isolated around 6,500 generations, immediately after the ara-2 lineage diversified into *S* and *L*, in glucose minimal media (DM25) for approximately 60 generations (9 days), with 12 biological replicates. To visualize colony morphologies of the resulting cultures, we plated the cultures on tetrazolium arabinose (TA) agar plates. Surprisingly, 2 of the independent cultures displayed a mixture of large and small colonies (Figure 1A).

After eliminating contamination possibilities by sequencing several diagnostic genetic loci, we examined whether the large colony phenotype was heritable. We isolated several large and small colonies and propagated them in DM25 for around 30 generations (5 days). The phenotype appeared to be stably heritable for all selected colonies. To avoid prematurely associating the larger colony phenotype with the *L* type, we referred to the emerging type in our experiments as S_B, due to its large (big) colonies and its ancestor *S*.

To gain insights into the robustness of the observed rediversification after isolation of *S* over evolutionary timescales, we isolated *S* from later generations, spanning more than 30,000 generations of evolution. We repeated the same experiment with *S* clones from 17,000 and 40,000 generations with 24 independent cultures each; however, we did not see any noticeable emergence of big colonies after 60 generations (screened about 200 colonies per plate). It is unclear why we did not see any big colonies; one possible explanation may be that the rate at which *S* morphs transition to S_B morphs may be low enough that we would need to have many more replicate cultures to observe rediversification (as in the 6,500 *S* clones). We previously noticed that 6,500 S_B clones (labeled 1 and 6) grew much better in lysogeny broth (LB) liquid media compared with *S* clones

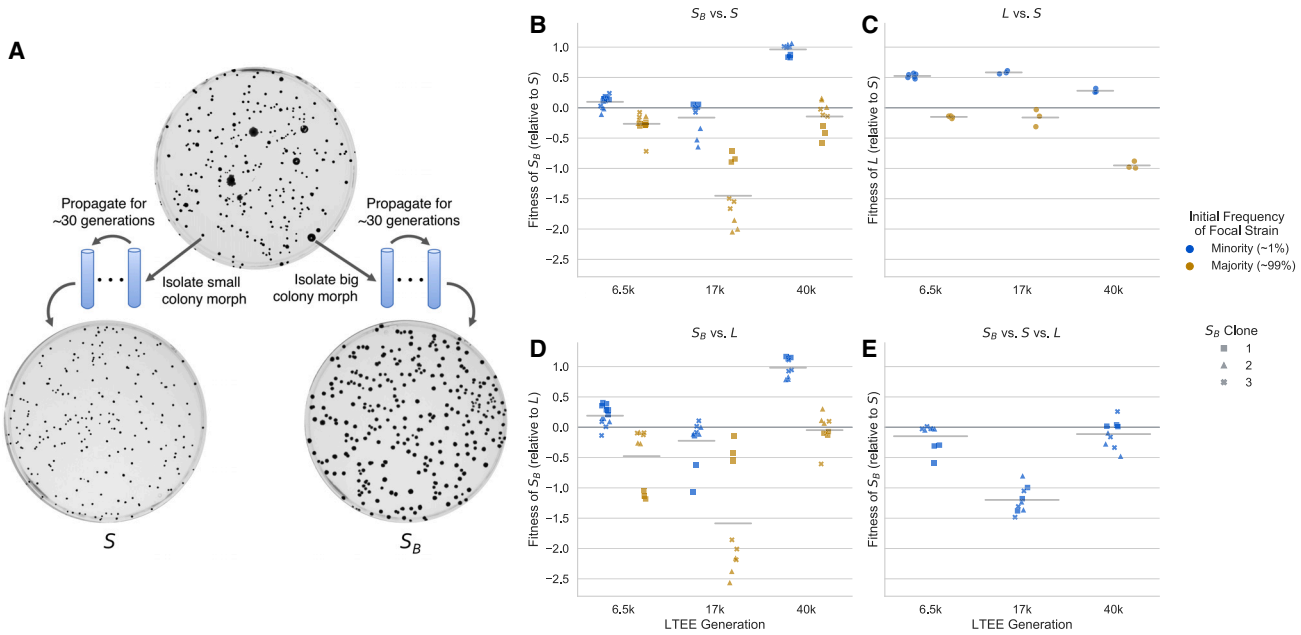

Figure 1. Emergence of the stably heritable S_B morph and frequency-dependent fitness effects

(A) Big colony morphs can arise in *S* cultures derived from 3 different LTEE time points, separated by more than 30,000 generations of evolution (6,500 clones are shown here as an example). When both small and big colonies are isolated and propagated in liquid DM25 culture for about 30 generations, then plated on TA agar plates, we see that the colony size is heritable.

(B–D) Reciprocal invasion experiments, measuring relative fitness of clones when they are in the minority of the population (approximately 1%) or in the majority (approximately 99%). Each point represents a biological replicate, and horizontal lines represent mean across all points. Competitions between (B) S_B clones and *S*, (C) *S* and *L*, and (D) S_B and *L*. We generally see negative frequency-dependent fitness effects across all strains and competitions.

(E) Triple competition between S_B, *S*, and *L*, where *L* and S_B are near their equilibrium frequencies and S_B in the minority (around 1%). See also Figures S1–S3.

(potentially accounting for their bigger colonies sizes on similar agar plates). Thus, we sought to see whether we could enrich for the appearance of S_B by growing 6,500, 17,000, and 40,000 *S* clones in LB liquid culture. Under these growing conditions, we indeed saw that S_B colonies appeared rapidly, within 1–3 days, in nearly all of the independent *S* cultures across the three LTEE time points (Figure S3). We attributed this to the higher fitness of S_B in LB, relative to *S*. The new S_B clones were again stably heritable for at least 30 generations.

The big colony phenotype S_B bears at least a superficial resemblance to *L*, which begs the question: do S_B and *S* represent genuinely different ecotypes, occupying different ecological niches, with the potential to coexist with each other? To answer this question, we performed reciprocal invasion experiments, where we mixed *S* and S_B clones at high and low frequencies, and tracked how their frequencies change via flow cytometry over the course of three growth cycles (see STAR Methods), to estimate their relative, frequency-dependent fitness effects (Figure 1B). While relative frequencies of LTEE strains are typically measured by colony counting, we found significant bias (Figures S1G and S1H) in frequency measurements of *S*/*L* when measured via colony-forming units (CFUs). In contrast, we see that flow cytometry provides unbiased frequency measurements (Figures S1A–S1F). We thus chose to use flow cytometry for all further measurements instead of CFUs, owing to its minimal bias and reduced measurement noise (Figure S1). The introduction of genomically integrated fluorescent proteins does not have a measurable impact on fitness (Figure S2C).

We found that most S_B clones had significant negative-frequency-dependent fitness differences when in competition with their parental *S* clone, a hallmark of ecological differentiation ($p < 0.05$ for all clones except 6,500 S_B 3). These data suggest that many of the S_B clones can coexist with *S*, because relative fitness is greater than 0 at low frequencies and less than 0 at high frequencies. However, it is not clear if this is the case for all of the isolated S_B clones, as some have a relative fitness near or less than 0 at low frequencies. This may be because the aforementioned S_B clones either genuinely do not coexist with *S*, or perhaps they coexist at a frequency around or lower than the one where we took the measurements.

The frequency-dependent fitness differences between S_B and *S* were similar in magnitude to the fitness differences between *L* and *S* (Figure 1C), which were all significant at $p < 0.01$. We also competed S_B against *L* (Figure 1D) and again found significant frequency-dependent fitness differences for most clones ($p < 0.01$ for all clones except 6,500 S_B 3 and 17,000 S_B 1). However, if at least some S_B clones can invade both *S* and *L* when rare, why has the S_B morph not appeared in the ara-2 population of the LTEE, where *L* and *S* have been coexisting and coevolving for tens of thousands of generations? We hypothesized that S_B could not invade an already "full" community and could only have the chance to invade when one of the ecotypes is removed. We performed a triple competition experiment, with *L* and *S* near their equilibrium frequency and S_B in the minority (Figure 1E). We found that most S_B clones had a significantly lower fitness compared with when it was in the minority with either *S* or *L* alone

($p < 0.05$ except for 6,500 S_B clones 2 and 3 compared to when competed against S alone, and 17,000 S_B 1 6,500 S_B 3 when competed against L alone).

While we have shown that S_B spontaneously emerges from a monoclonal population of S and occupies a distinct ecological niche, it is not yet clear how S_B compares with S and L. In particular, we want to understand if S_B simply fills the same niche that L had occupied before removal, making it somewhat functionally equivalent to L. The negative frequency-dependent selection between L and S_B suggests that they must be different to some degree, but it is still unclear if S_B represents a sort of intermediate between S and L, or if it shows novel traits. In the following, we will show that while S_B resembles L in some of its growth properties, it also shows clear differences that are critical for its coexistence with S.

Within-cycle growth dynamics of cocultures

To better understand how ecological differentiation arises in the S_B-S and L-S systems, we measured the within-cycle growth dynamics of S_B, L, and S in coculture with each other via flow cytometry. The LTEE environment is a seasonal one,[27,64]—every 24 h, cultures are transferred 1:100 into fresh glucose minimal media. The populations spend the first part of the day in exponential phase; the remaining time, more than 2/3 of the day, is spent transitioning out of exponential phase and in stationary phase. It has been previously shown that L and S occupy different temporal niches from one another, where L specializes on exponential growth on glucose, and S specializes on stationary phase survival and growth on acetate. Thus, it is natural to ask how temporal variations in growth are similar or different in the S_B-S system.

To perform the experiments, we propagated S, S_B, and L separately in monocultures for 2 days, before mixing S with S_B and S with L, both at high and low frequencies. We mixed strains with their partners from the same LTEE generation. For simplicity, we only used S_B clone 1 for all experiments and LTEE generations. We propagated the cocultures for one more cycle to allow the populations to physiologically adapt to the new environment. At the end of the 24-h cycle, we took a flow cytometry measurement of the culture, then split the cultures into biological replicates and diluted the cocultures 1:100 into fresh media. Afterward, we took flow cytometry measurements from the cocultures approximately every hour for about 8 h, then we took one last measurement at the end of the 24-h cycle (Figures 2 and S4). We chose this design because the fastest dynamics occur during and right after exponential phase—the first 8 h—while dynamics in stationary phase are much slower. The cultures were grown in a 37°C shaking water bath. We corrected the cell counts measured in flow cytometry by the total dilution rate.

We initially focus on the dynamics of strains from 6,500 generations (Figure 2). Overall, it is immediately clear that there are larger differences in dynamics in the L-S cocultures compared with the S_B-S cocultures. When S is in both the majority and minority, L has a long, 2-h lag time, while S starts growing much more quickly (Figures 2C and 2D), causing a large upward spike in S frequency. We fit a generalized logistic model to the growth curves to more precisely extract the lag times (Figure S5), and we see that 6,500 L has a longer lag time than S ($p < 0.01$, in both cases). When S is cocultured with S_B, we do not see any noticeable lag time; however, when S is in the minority, S "wakes up" more quickly than S_B ($p = 7 \cdot 10^{-8}$), leading to a small spike in S frequency at the beginning of the time course. We see similar patterns in the cocultures from 17,00 and 40,000 generations—both L and S_B appear to have growth rates very close to 0 at the beginning of the time course, but S consistently has a larger initial growth rate (Figure S4). The initially faster growth of S only occurs when S is in the minority for both L and S_B strains ($p < 0.02$, across all comparisons); there is no longer a noticeable difference when S is in the majority ($p > 0.1$, across all comparisons).

When 6,500 L starts growing, it has a significantly larger growth rate than S, pushing the frequency of S back down. The magnitude of this growth rate difference is similar regardless of the relative frequency of the ecotypes (Figures 2E and 2F). In contrast, the differences between S_B and S are much smaller. At both starting frequencies, S_B may have a small growth rate advantage compared with S early in exponential phase, then S appears to grow faster in late exponential phase.

In contrast to the dynamics in lag and exponential phase, the later "stationary phase" dynamics are highly dependent on which ecotype is in the majority. While most conditions show non-zero growth rates after about 8 h of growth, we still refer to this period as stationary phase, because the growth rates are small. When S is cocultured with L, S grows better than L under both conditions, but the absolute growth rates differ between the conditions (insets in Figures 2E and 2F). When S is in the minority with L, both S and L have net positive growth in stationary phase, although it is higher for S ($p = 9 \cdot 10^{-4}$), potentially pointing to the favorable conditions of L-dominated stationary phase and the putatively large amount of excreted acetate available for exploitation. In contrast, when S is in the majority with L, S has a smaller, albeit still positive, net growth rate, while L has a net negative growth rate in stationary phase ($p = 5 \cdot 10^{-4}$). Concordantly, these patterns suggest that S-dominated stationary phase is much less hospitable to both S and L.

We see different stationary phase patterns when S_B and S are in coculture, where S_B now performs consistently better than S (insets in Figures 2E and 2F). The ecotype growth rates are significantly different ($p = 0.016$) when S_B is in the majority with S—S_B has a moderately positive net growth rate, while S has essentially a net 0 growth rate in stationary phase. Then, when S_B is in the minority, both S_B and S have net negative growth rates, but S declines more than S_B, although the difference is non-significant ($p = 0.16$). If S_B were more similar to L, i.e., an exponential phase specialist that secretes a substantial amount of acetate, we would have expected that S_B-S and L-S cocultures would have similar behavior in stationary phase. Instead, S_B appears to have enhanced survival in stationary phase and decreases the survival prospects of S, perhaps because of the reduced availability of acetate. Thus, while S_B does not have a significant advantage over S in exponential phase like L has, it compensates with a clear advantage over S in stationary phase, essential for coexistence of S_B with S.

The results show differences in stationary phase behavior across generations, as well as several conserved features (Figure S4). Similar to the 6,500 strains, when 17,000 S are in the minority with L, S has a large positive growth rate during

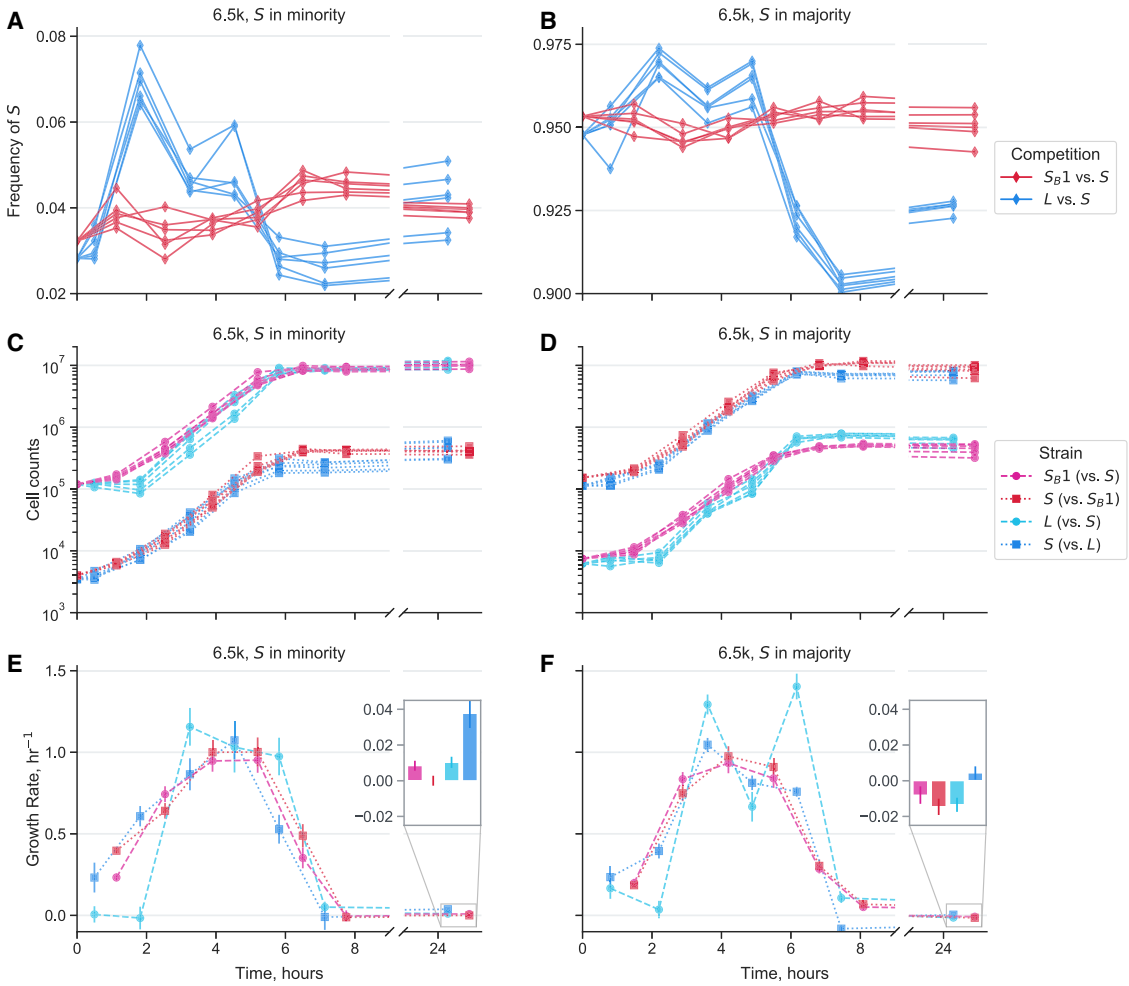

Figure 2. Growth dynamics of cocultures over the course of one 24-h growth cycle

Measurements were taken approximately every hour via flow cytometry for the first 8 h after transfer into new media. An additional measurement was taken approximately 24 h after the start of the cycle. Mixed S_B 1 with *S* along with *L* with *S*, all from 6,500 generations, where ecotypes were mixed both in the majority and minority of the population. Different lines represent biological replicates.

(A and B) Frequency dynamics of *S* against S_B and against *L*.

(C and D) Total cell count dynamics, separated by each strain in the cocultures.

(E and F) Empirically measured growth rates over time for each strain in the cocultures, calculated as the slope of log-transformed abundance between adjacent time points, using the second time point as the x axis location. Insets show growth rates during stationary phase, from around 8 to 24 h, on the y axis—presented to provide a more fine-grained view of the slow changes in abundance during stationary phase. Error bars represent standard errors. See also Figures S4 and S5.

stationary phase, whereas *L* does not grow. However, when *S* is in the majority with *L*, its growth rate is comparable to that of *L*. The 40,000 *S* and *L* strains show different patterns, where *L* generally has a higher stationary phase growth rate. However, this appears to be offset by a large growth advantage of *S* right at the end of exponential phase/beginning of stationary phase; this growth advantage is much larger when *S* is in the minority compared with when it is in the majority. This indicates that the growth advantage of 40,000 *S* has shifted earlier, potentially because it has adapted to consume the acetate secreted by *L* much more quickly.

Again, the stationary phase behavior when 17,000 and 40,000 *S* and S_B are grown in coculture is noticeably distinct from the behavior of *L*-*S* coculture. Similarly, 17,000 *S* also does not grow well in S_B-dominated stationary phase. And 17,000 S_B actually has a large positive stationary phase growth rate when *S* is setting the environment, suggesting that S_B has more to gain from stationary phase when it is in the minority compared with vice versa. The picture shifts again with the 40,000 strains—S_B benefits very little from being in stationary phase, but in contrast, *S* grows well in stationary phase, especially when dominated by S_B. This is quite different from the behavior of 40,000 *S*-*L* cocultures, albeit in a different direction than the strains from the earlier generations. Thus, in 40,000 cultures, it appears that S_B-*S* cocultures act more like *L*-*S* cocultures from earlier generations, where S_B is the clear exponential phase specialist and *S* is the stationary phase specialist.

Together, these results show that growth traits of *L*-*S* cocultures change over evolutionary time, and S_B-*S* cocultures are

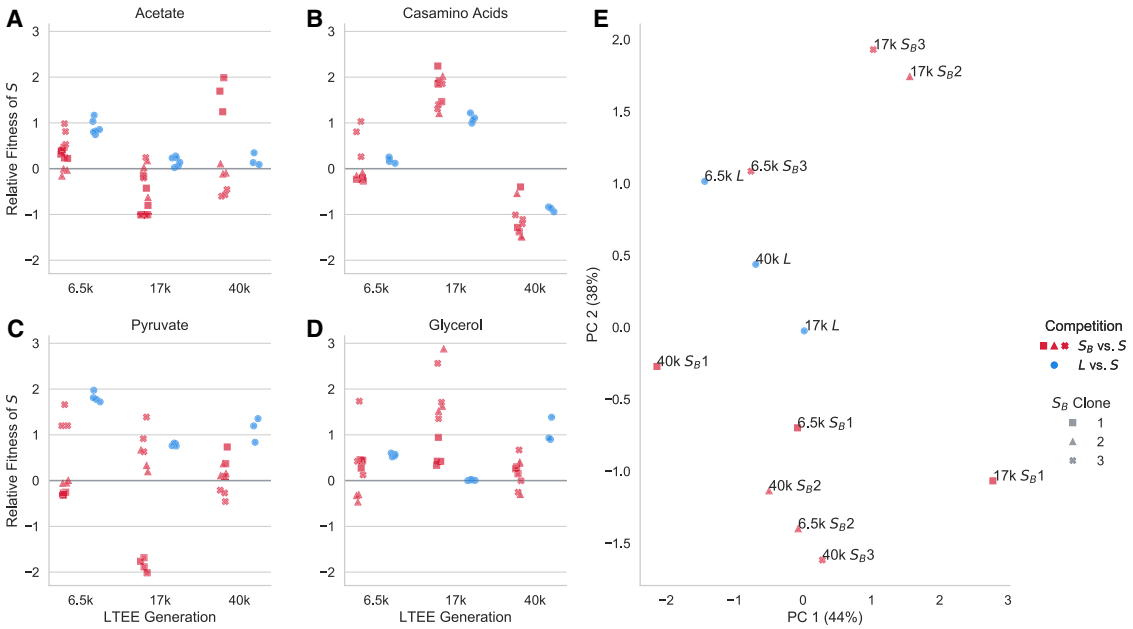

Figure 3. Competition of S_B and L against S in novel environments

(A–D) Red and blue points represent the relative fitness of S in competition with S_B and L clones from the same LTEE time point, respectively, where different symbols represent different clones. Competitions performed in exponential phase in the same media base (DM) supplemented with different carbon sources: (A) 200 mg/L acetate, (B) 1 mg/mL casamino acids, (C) 20 mM pyruvate, (D) 20 mM glycerol.

(E) Principal-component analysis, using relative fitness in each environment as features. Percentages in parentheses represent percent variance explained by each principal component. See also Table S4.

similar in important ways (e.g., initial growth rates), but also show departures from the original community (e.g., stationary phase behavior) that reveal how the ecological dynamics have shifted with the new, rediversified ecotype.

Growth traits in novel environments

While we have shown that S_B has distinct growth traits when in coculture with S in the evolutionary condition, does S_B also behave differently compared with S in novel environments that neither have been in contact with before? If S and S_B mostly behave similarly in novel environments, then perhaps the underlying change between the two morphs is targeted only toward traits relevant to the mechanism of ecological differentiation. Other newly diversified ecotypes have previously shown targeted changes to niche adaptation, such as acetate-specialist *E. coli* ecotypes that evolve due to specific mutations in the main acetate-scavenging gene, *acs*.[29,30] On the other hand, if pleiotropic effects are widespread, then the underlying metabolic/physiological shift in S_B may involve global, rather than targeted changes.

To this end, we competed S_B clones against S clones for each LTEE time point in the same minimal media base as the evolutionary condition (DM), supplemented with different carbon sources (Figure 3). For comparison, we also competed S against L clones for each LTEE time point in each of the conditions. We chose four different carbon sources that support the growth of S, S_B, and L clones from all time points and that enter into central metabolism at different points,[65] potentially allowing us to gain insight into global changes in physiology and metabolism. After growing cocultures together for 2 days in DM25, we diluted them 1:100 in each different media. We kept the cultures in exponential phase and took two-ecotype frequency measurements via flow cytometry: one right before transfer into the new media, and one at the end of exponential phase. As usual, relative fitness was computed as the change in logit frequency.

We see that for most S_B clones, across most conditions, S_B is noticeably non-neutral relative to S (false discovery rate [FDR]-corrected two-sided t tests; Table S4). Consistent with previous experiments,[60] we see that L is also usually non-neutral relative to S across the different carbon sources. The relative fitness of S_B and L clones varies considerably across time points and carbon sources. There is not a clear relationship between the fitness of S_B (relative to S) and the fitness of L (relative to S) from the same time point, also visible in the principal-component analysis (PCA) representation of the data (Figure 3E). In the PCA, it appears that S_B clones within time points generally cluster together (but not completely), not with the L clones from their time point; however, leveraging a modified permutational multivariate analysis of variance (PERMANOVA) test (see STAR Methods), we could not reject the null hypothesis that S_B clones within a time point cluster together more than the L clone within the time point ($p > 0.15$ for all time points).

Again, there is some variation between different S_B clones. For example, the 17,000 S_B 1 clone behaves noticeably differently compared with the 17,000 S_B 2 and S_B 3 clones especially in the pyruvate ($p < 0.001$, both clones) and glycerol ($p < 0.02$, both clones) conditions, while the three clones cluster together in the acetate condition. The 17,000 S_B 2 and S_B 3 clones also appear to cluster together, away from the 17,000 S_B 1 clone in the PCA plot (Figure 3E). The 17,000 S_B 1 clone also behaved

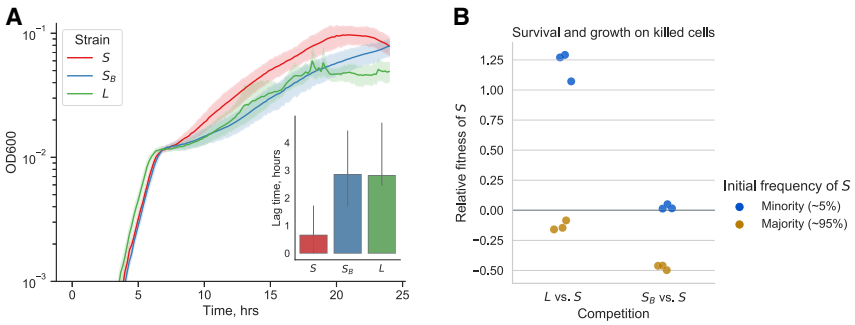

Figure 4. Traits affecting coexistence

(A) Growth dynamics of 6,500 strains in glucose/acetate media. All strains were grown in monoculture, and prior to approximately 6 h, they are growing exponentially on glucose. Afterward, a diauxic shift if visible in all growth curves, where the strains begin to switch to using acetate. Error bars are standard deviations, over 8 biological replicates for each culture. Inset represents lag time from glucose to acetate growth; error bars are 95% confidence intervals (CIs).

(B) Survival and growth of strains on heat-killed cells provides evidence of differential ability to grow and survive on cell debris. After 24 h of growth, we thoroughly washed a 6,500 *S* culture (to remove any extracellular metabolites), resuspended the culture in blank media, and heat-killed the culture. We then washed cultures of 6,500 strains *L* 1, S_B 1, and *S* 1, mixed the strains, took flow cytometry measurements, and transferred the cultures 1:1,000 to the heat-killed culture. We then allowed the cultures to incubate for 16 h before taking another measurement.

differently compared with the other two in the reciprocal invasion experiment against 17,000 *L*, where 17,000 S_B 1 did not show noticeable frequency dependence (Figure 1D). The 6,500 S_B 3 and 40,000 S_B 1 clones also cluster away from the other two clones within their time point. The conditions where these "outlier" clones diverge from the other clones varies between time points-6,500 S_B 3 is different when grown in in pyruvate and casamino acids ($p < 0.03$, all comparisons), and 40,000 S_B 1 is primarily different in the acetate condition ($p < 0.03$, all comparisons).

Across all three time points, we see that *S* is better at growth in acetate compared with *L*. The evidence for this is stronger for strains from 6,500 generations ($p = 5 \cdot 10^{-4}$) and 17,000 generations ($p = 0.017$), compared with those from 40,000 generations ($p = 0.093$). This is consistent with prior findings,[58] and the notion that *S* represents a consistent acetate-scavenging specialist over evolutionary time. In contrast, the behavior of S_B in acetate is more variable, both across time points and between different S_B clones. Two of three 6,500 S_B clones have a fitness disadvantage in acetate ($p < 0.03$, in both cases) relative to *S* (albeit less pronounced compared with *L*), whereas at least one S_B clone from both 17,000 and 40,000 generations have a fitness advantage in acetate ($p < 0.02$, in both cases).

We found the fitness disadvantage of 6,500 S_B in acetate puzzling, because it appears to generally perform better in stationary phase compared with *S* (Figure 2). At least in the *S*-*L* community, the advantage of *S* in stationary phase appears to be mostly driven by decreased lag time to acetate growth,[66] increased acetate growth rate,[58] and increased ability to scavenge dead cells.[27] We began by testing if S_B has a smaller glucose-to-acetate lag time compared with *S*, which could explain its stationary phase advantage. By growing S_B, *S*, and *L* strains on media containing both glucose and acetate, we could observe how they transition from glucose to acetate growth (Figure 4A). Contrary to expectations, S_B actually has a longer lag time than *S*, more similar to that of *L*, which should give it a disadvantage in stationary phase. So we moved on to testing whether S_B could scavenge cell debris more efficiently than *S*. To do so, we competed mixtures of S_B and *S*, and *L* and *S*, together in heat-killed cultures, resuspended in blank media to eliminate the presence of extracellular metabolites (Figure 4B). Consistent with prior data,[27] we see that *S* is generally more efficient than *L* at growing and surviving in an environment of dead cells, albeit with a frequency-dependent effect. In contrast, S_B is able to scavenge dead cells better than *S*, which may explain its ability to perform better in stationary phase.

Together, this is another sign that S_B is occupying a genuinely different ecological niche compared with *L*, which may be shifting over evolutionary time.

Transcriptional differences between ecotypes

Given the strong heritability of the S_B phenotype, and multiple traits that differ with respect to *S*, we reasoned that the S_B phenotype may have an underlying genetic cause. Thus, we performed whole-genome shotgun sequencing of several *S* and S_B clones with both short-read sequencing (Illumina) and long-read sequencing (Nanopore) (see STAR Methods). After reference-based assembly, we saw that all S_B clones had several mutations relative to their ancestor, and all *S* clones from the same LTEE generation also had several mutations relative to each other. The mutations were a mix of synonymous and non-synonymous point mutations, insertions and deletions, and several large genomic rearrangements (Table S3). However, none of the mutations differentiated *S* and S_B—there were no consistent mutations in specific genes or operons. The large number of mutations separating S_B clones from their *S* ancestor is not surprising; the ara-2 lineage fixed a hypermutator allele before the *S* and *L* lineages split, such that the germline mutation rate is about 100× higher than that of the LTEE ancestor.[67] This makes it likely that many of the mutations are likely (nearly) neutral hitchhikers, or otherwise were not affected by selection. We attempted to determine whether there was any parallelism on the level of Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations instead of genes, but again, we did not detect anything. Thus, because of the combination of the high mutational background and lack of detectable genetic parallelism, we cannot determine whether the S_B phenotype has a genetic cause, or what the causative mutation(s) would be. If the S_B phenotype is caused by some genetic change, it is likely that many different mutations are able to cause a similar phenotype.

To further understand the underlying causes of the S_B phenotype, we turned to measuring transcriptional differences between *L*, *S*, and S_B from 6,500 generations using RNA sequencing (RNA-seq). We chose to focus on 6,500 strains because this is the LTEE time point immediately after the *S* and *L* lineages diversified, allowing us to focus on the "minimal" differences between *S* and *L*, rather than after extensive evolution and divergence. We cultured two biological replicates of two independent clones of

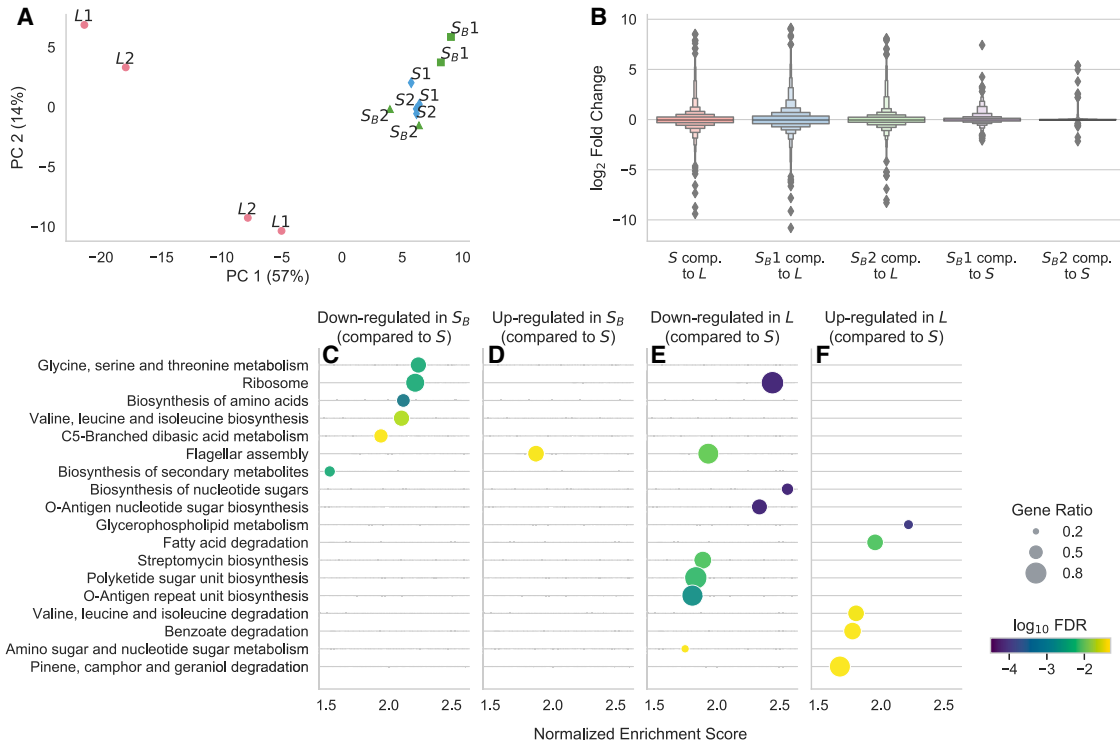

Figure 5. Results from RNA-seq of *L*, *S*, and S_B clones from 6,500 generations

(A) Principal-component analysis of RNA-seq data, after processing. Samples with the same name represent biological replicates of the same clone; the 1 and 2 labels are to indicate which clone the samples come from.

(B) Distributions of $\log_2$ fold changes in gene expression across all genes, comparing different strains to each other.

(C–F) Results of a KEGG gene set enrichment analysis to identify pathways with coordinated changes in gene expression between ecotypes, where (C) and (D) are comparing S_B with *S* and (E) and (F) are comparing *L* with *S*. Only pathways that are called as significant at $p < 0.05$ after an FDR correction are included; points are colored by FDR-corrected $\log_{10}$ p value. Pathways are ordered by normalized enrichment score, which is roughly a measure of the extent to which pathway-associated genes are overrepresented at the top or bottom of the entire list of genes, ranked by fold expression change. The size of the points is proportional to the "gene ratio," which is the ratio of core enrichment genes to the total number of genes in the pathway, i.e., the fraction of genes in the pathway that show differential expression. See also Table S3 and Figure S6.

each *L*, *S*, and S_B from 6,500 generations in glucose minimal media, and collected samples in mid-exponential phase (see STAR Methods), in line with previous, similar transcriptomic measurements.[60,68] For a broad overview of the data, we first performed a PCA, using (normalized, transformed) expression for each gene as the features (Figure 5A). We see that the first principal component already captures more than half of the variance between samples, which primarily serves to separate the *L* clones from the *S* and S_B clones. The *S* clones appear to cluster together strongly, with the S_B clones flanking them. Hierarchical clustering also reveals that the *S* clones cluster together, with the S_B 2 clone as the next most similar, and the S_B 1 clone as the outer-most member of the cluster (Figure S6A). This suggests that there are more differences between *S* and S_B than there are between the two *S* clones, but there are stronger differences comparing both *S* and S_B with the *L* clones. The same picture emerges if we look at the distribution of $\log_2$ fold expression changes between different ecotypes (Figures 5B and S6D). Comparing *S* and S_B with *L*, there are many genes with a large range of expression changes, both increasing and decreasing in expression. In contrast, there are generally smaller differences between the two S_B clones and *S*. Again, there are larger and more differences between S_B 1 and *S*, compared with S_B 2 and *S*, suggesting variability between the two S_B clones.

Given that there are noticeable differences between S_B and *S*, we next sought to understand what those differences represent. Are there identifiable pathways with coordinated expression changes? How do they compare with the differences between *L* and *S*? To this end, we performed gene set enrichment analyses to identify differentially expressed KEGG pathways.[69] We first compared S_B to *S* and *L* to *S*, and only look at pathways that are significantly enriched at $p < 0.05$ after a multiple-testing correction (Figures 5C–5F). We see that there are a number of pathways significantly downregulated in S_B compared with *S*, and only one pathway significantly upregulated (Figures 5C and 5D). Most of the downregulated pathways are related to different aspects of amino acid metabolism. We also separately compared S_B 1 and S_B 2 against *S* to better understand the variability between the two clones (Figure S6C). As expected, most of the terms identified in the pooled analysis (e.g., ribosomal proteins, amino acid metabolism terms) appeared as the top terms when we analyzed the clones separately, albeit at a lower significance level than when the data from the two clones are pooled together. There are potentially a handful of differences in enriched

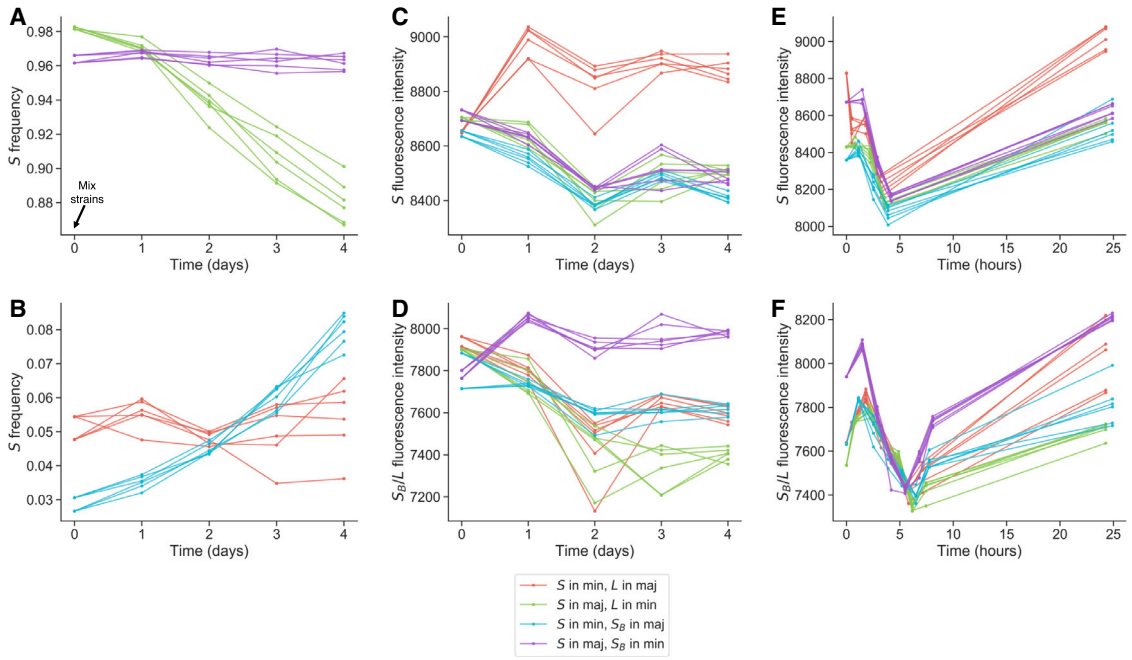

Figure 6. Dynamics of σ^{70} promoter activity revealed by fluorescence intensity measurements

(A–D) We performed an experiment where we mixed 6,500 *S* 1 and either 6,500 *L* 1 or S_B 1 at high and low frequencies at day 0 after 3 days of growth in monoculture, and measured (A and B) frequencies and (C and D) YFP/BFP fluorescence intensities via flow cytometry. The YFP and BFP genes are under the control of a constitutive σ^{70} promoter. We propagated the strains in coculture thereafter. We see large, consistent, community composition-dependent shifts in fluorescence intensity for several conditions upon introduction to the community context, especially noticeable when *S* is in the minority with *L*, and when S_B is in the minority with *S*.

(E and F) We quantified the within-cycle fluorescence dynamics (same dataset as shown in Figure 2), where we see substantial changes over the course of the growth cycle.

pathways between the two clones. For example, terms related with O-antigen biosynthesis (e.g., biosynthesis of nucleotide sugars, O-antigen nucleotide sugar biosynthesis) may be upregulated in S_B 1, but not S_B 2. The differentially expressed pathways between *L* and *S* are mostly different, there are no terms related to amino acid biosynthesis, and many terms related to lipid metabolism and O-antigen biosynthesis (Figures 5E and 5F). Differentially expressed pathways in *L* tend to not be differentially expressed in S_B, and vice versa (Figure S6B).

There are two pathways that are enriched in both comparisons: flagellar assembly and ribosomal proteins. The changes to flagellar assembly expression are in the opposite direction for S_B and *L*, where it is upregulated in S_B but downregulated in *L*, suggesting that gene expression for this pathway is ordered $L < S < S_B$. In contrast, expression of ribosomal proteins is downregulated in both *L* and S_B, perhaps indicating some degree of parallelism involving a fundamental aspect of cell physiology between the two ecotypes. However, overall, with the exception of the downregulation of ribosomal proteins, it appears that the transcriptional changes that differentiate S_B and *L* from *S* are quite distinct.

Our RNA-seq dataset was restricted to monoculture mid-exponential phase culture, so we sought to elucidate gene expression changes throughout the growth cycle, while varying community composition, by leveraging the fluorescent reporters we introduced into the ecotypes, all inserted into the same genomic location. Fluorescence intensity is commonly used to measure protein concentration and promoter activity.[70–73] The genes encoding the fluorescent proteins are under the control of a constitutive σ^{70} promoter, giving a read-out of σ^{70} activity. The net activity of σ^{70} is a useful measure of the global transcriptional state because it is highly responsive to changes in environmental conditions and growth phases. It plays a crucial role in coordinating the expression of genes in response to environmental stresses, nutrient availability, and other external stimuli.[74–77] By monitoring σ^{70} activity, we hope to gain insights into how the cell responds and adapts to different conditions, providing a snapshot of the global transcriptional state.

To this end, we performed an experiment where we mixed 6,500 *S* 1 with either 6,500 *L* 1 or S_B 1 at high and low frequencies, after 3 days of growth in monoculture and measured their frequencies over time and population-averaged fluorescence intensity at the end of each cycle (Figures 6A–6D). We see large, consistent, community composition-dependent shifts in fluorescence intensity for several conditions upon introduction to the community context, especially noticeable when *S* is in the minority with *L*, and when S_B is in the minority with *S*. Specifically, in both cases, the fluorescence intensity sharply increases after 1 day of coculture growth, a shift that is maintained for the duration of the experiment. The parallel changes in putative σ^{70} activity across S_B and *S* reveal that both actively change their transcriptional programs in response to community composition. To a less noticeable degree, we see that the fluorescence intensity of *L* is lower when it is in the minority with *S*, again pointing to a degree of composition-dependence.

We measured the fluorescence dynamics over the course of the growth cycle to investigate how the difference arise (Figures 6E and 6F). We see substantial, parallel changes throughout the cycle, with fluorescence intensity dropping during exponential phase, before increasing again during stationary phase. These data reveal that increased fluorescence intensity at the end of the growth cycle seen when *S* and S_B are in the minority with *L* and *S* respectively accumulates during stationary phase, with the difference mostly disappearing during exponential phase. Together, these data provide evidence that S_B has a transcriptional reaction to the presence of *S* in the majority that is parallel to how *S* reacts to the presence of *L* in the majority.

DISCUSSION

Our study explores the capacity of an evolved microbial community to quickly regenerate ecological diversity following the removal of an ecotype. Our results suggest that even in the case of a community composed of only two strains in a minimal environment, evolution can leave room for alternative diversification processes.

The rediversified ecotype, S_B, demonstrates the robustness of microbial communities to perturbations by sharing several growth traits with the ecotype it replaces, *L*. For instance, both S_B and *L* exhibit slower initial growth or longer lag times compared with *S* across all LTEE time points, which may be involved in a trade-off allowing for higher exponential growth rates, as observed in other systems.[78] However, differences between the rediversified and original communities suggest that the mechanism of ecotype coexistence has shifted. Notably, we observe variations in stationary phase responses, survival, and ability to scavenge dead cells, as well as distinct patterns of gene expression. At least in the 6,500 strains, coexistence between *S* and S_B appears primarily driven by a trade-off between glucose growth in late exponential phase (where *S* does better), and the ability to survive in stationary phase, owing to the advantage in scavenging dead cells (where S_B does better). Together, these findings indicate that ecological rediversification in the *S*-*L* system may be influenced by a combination of constraints and opportunities. While some traits may evolve nearly deterministically due to strong ecological or physiological constraints, other trait values may be more unconstrained. The interplay between contingency and determinism mirrors patterns observed in various other evolving systems, including the LTEE.[79–81] Dissecting why some traits are more evolutionarily constrained during diversification compared with others could be a fruitful avenue for future investigation.

We attempted to determine a potential genetic origin of the S_B phenotype. However, we did not find any consistent mutations shared between the independent S_B clones, relative to their *S* ancestor. Thus, the S_B phenotype likely either has a large target size, such that many different mutations can cause the same phenotype,[82,83] or it is caused by a non-genetic heritable change. Despite the fact that we did not find any shared mutations, the transcriptional changes of two S_B clones were targeted to the same handful of pathways, predominantly related to amino acid metabolism. This points to parallelism among independent S_B clones, at least on the transcriptional level, if not on the genetic level. The downregulation of amino acid biosynthesis may be related to the scavenging lifestyle of S_B, where its amino acid needs can be met by consuming dead cells instead of through *de novo* synthesis. Additionally, while the differentially expressed pathways in S_B and *L* relative to *S* were generally different, we saw decreased expression of ribosomal proteins in both ecotypes. The fraction of the proteome devoted to ribosomes is known to control many growth traits in bacteria,[84,85] so the similar changes in *L* and S_B may help to explain the handful of observed similarities in growth traits. One might expect that ribosome expression should be lower in *S*, due to its slower exponential growth rate[86,87]; so the fact that this is not the case may suggest that S_B and *L* are both allocating their proteome not just to optimize exponential growth rate, but also other growth traits as well.

While we saw that *S* could rediversify following isolation, we did not see any obvious ecological or phenotypic diversification when *L* was isolated. There may be several reasons for this. (1) S may have some amount of physiological/genetic/metabolic plasticity that allows it to diversify that *L* lacks. (2) Diversification of *L* may happen slowly or rarely, or more quickly only under certain environmental conditions. (3) Perhaps *L* can rapidly diversify, but cryptically, where no phenotypic changes are obvious without more extensive phenotyping. It is certainly the case that we would not have found S_B without the obvious changes in colony size. It could be that rediversification is much more common than currently appreciated, but simply not detected. Sequencing technologies, including metagenomic[62] and DNA-barcoding-based methods,[88] could help to better reveal the full extent of rediversification across microbial communities. In fact, through metagenomic sequencing, we now know that ecological diversification is much more common in the LTEE than previously thought.[62]

Our study contributes to the understanding of the ecological consequences of ecotype removal or extinction, which often occurs in natural microbial communities due to sudden environmental shifts.[41–47] The ability of these communities to recover their diversity after such disruptions might be key to maintaining their functions and stability over time. Contrary to the notion that evolutionary processes are too slow to influence ecological recovery, our findings underscore the importance of evolution in the rebound of communities after a disturbance. We used simplified two-ecotype communities, which are ideal for such studies because they are well documented and amenable to experimental manipulation.[16,27,58–63] The methods we developed to investigate rediversification in this simple model could serve as a framework for understanding this process in more complex ecosystems. The presence of alternative eco-evolutionary pathways, even in a maximally reduced community of only two strains, hints at more complex dynamics in richer ecosystems. Ultimately, our work sheds light on the resilience of microbial communities, their ability to recover ecological diversity, and their adaptability to environmental changes. Future research on the processes that control these dynamics is essential for a comprehensive understanding of microbial community function and stability, especially in the face of environmental shifts.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2024.01.029.

ACKNOWLEDGMENTS

We thank Adam Arkin, Morgan Price, Benjamin Good, Tanush Jagdish, Michael Desai, Jeff Barrick, Dominique Schneider, and all members of the Hallatschek lab (past and present) for helpful comments and advice on the project. We thank Richard Lenski for sending us the LTEE-derived strains and populations, along with experimental advice and feedback. Research reported in this publication was supported by a National Science Foundation CAREER award (1555330). This work was supported by the National Institute of General Medical Sciences of the NIH under award R01GM115851 and by a Humboldt Professorship of the Alexander von Humboldt Foundation. J.A.A. acknowledges support from an NSF graduate research fellowship, a Berkeley fellowship (from UC Berkeley), and Lloyd and Brodie scholarships (from UC Berkeley Dept of Bioengineering). We thank Mary West of the Cell and Tissue Analysis Facility (CTAF) at UC Berkeley. This work was performed in part in the QB3 CTAF that provided the ThermoFisher Attune Flow Cytometer (2017 model). RNA sequencing and processing was performed by SeqCenter, LLC. Nanopore library preparation and genomic sequencing along with Illumina sequencing was performed by the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation grant.

AUTHOR CONTRIBUTIONS

Conceptualization, J.A.A., J.D., and O.H.; methodology, J.A.A., J.D., and O.H.; investigation, J.A.A., J.D., K.L., Q.Y., and K.M.W.; writing – original draft, J.A.A., J.D., and O.H.; writing – review & editing, J.A.A. and O.H.; funding acquisition, O.H.; resources, K.M.W. and O.H.; supervision, J.A.A. and O.H.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 10, 2023
Revised: November 9, 2023
Accepted: January 10, 2024
Published: February 6, 2024

REFERENCES

1. Benton, M.J. (1995). Diversification and extinction in the history of life. Science *268*, 52–58.
2. Wellborn, G.A., and Langerhans, R.B. (2015). Ecological opportunity and the adaptive diversification of lineages. Ecol. Evol. *5*, 176–195.
3. Burress, E.D. (2015). Cichlid fishes as models of ecological diversification: patterns, mechanisms, and consequences. Hydrobiologia *748*, 7–27.
4. Rabosky, D.L. (2013). Diversity-dependence, ecological speciation, and the role of competition in macroevolution. Annu. Rev. Ecol. Evol. Syst. *44*, 481–502.
5. Losos, J.B. (2010). Adaptive radiation, ecological opportunity, and evolutionary determinism: American society of naturalists eo wilson award address. Am. Nat. *175*, 623–639.
6. Stroud, J.T., and Losos, J.B. (2016). Ecological opportunity and adaptive radiation. Annu. Rev. Ecol. Evol. Syst. *47*, 507–532.
7. Whittaker, R.H. (1972). Evolution and measurement of species diversity. Taxon *21*, 213–251.
8. Schluter, D. (2000). The Ecology of Adaptive Radiation (OUP).
9. Friesen, M.L., Saxer, G., Travisano, M., and Doebeli, M. (2004). Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in escherichia coli. Evolution *58*, 245–260.
10. Spencer, C.C., Bertrand, M., Travisano, M., and Doebeli, M. (2007). Adaptive diversification in genes that regulate resource use in escherichia coli. PLoS Genet. *3*, e15.
11. Herron, M.D., and Doebeli, M. (2013). Parallel evolutionary dynamics of adaptive diversification in escherichia coli. PLoS Biol. *11*, e1001490.
12. Dieckmann, U., and Doebeli, M. (1999). On the origin of species by sympatric speciation. Nature *400*, 354–357.
13. Estrela, S., Diaz-Colunga, J., Vila, J.C., Sanchez-Gorostiaga, A., and Sanchez, A. (2022). Diversity begets diversity under microbial niche construction. Preprint at bioRxiv.
14. Helling, R.B., Vargas, C.N., and Adams, J. (1987). Evolution of escherichia coli during growth in a constant environment. Genetics *116*, 349–358.
15. Rainey, P.B., and Travisano, M. (1998). Adaptive radiation in a heterogeneous environment. Nature *394*, 69–72.
16. Rozen, D.E., and Lenski, R.E. (2000). Long-term experimental evolution in escherichia coli. viii. dynamics of a balanced polymorphism. Am. Nat. *155*, 24–35.
17. Brockhurst, M.A., Colegrave, N., Hodgson, D.J., and Buckling, A. (2007). Niche occupation limits adaptive radiation in experimental microcosms. PLoS One *2*, e193.
18. Gómez, P., and Buckling, A. (2013). Real-time microbial adaptive diversification in soil. Ecol. Lett. *16*, 650–655.
19. Schick, A., and Kassen, R. (2018). Rapid diversification of pseudomonas aeruginosa in cystic fibrosis lung-like conditions. Proc. Natl. Acad. Sci. USA *115*, 10714–10719.
20. Meroz, N., Tovi, N., Sorokin, Y., and Friedman, J. (2021). Community composition of microbial microcosms follows simple assembly rules at evolutionary timescales. Nat. Commun. *12*, 1–9.
21. Good, B.H., and Rosenfeld, L.B. (2023). Eco-evolutionary feedbacks in the human gut microbiome. Nat. Commun. *14*, 7146.
22. Gómez, P., and Buckling, A. (2013). Real-time microbial adaptive diversification in soil. Ecol. Lett. *16*, 650–655.
23. Folkesson, A., Jelsbak, L., Yang, L., Johansen, H.K., Ciofu, O., Høiby, N., and Molin, S. (2012). Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: An evolutionary perspective. Nat. Rev. Microbiol. *10*, 841–851.
24. Zhao, S., Lieberman, T.D., Poyet, M., Kauffman, K.M., Gibbons, S.M., Groussin, M., Xavier, R.J., and Alm, E.J. (2019). Adaptive Evolution within Gut Microbiomes of Healthy People. Cell Host Microbe *25*, 656–667.e8.
25. Sousa, A., Ramiro, R.S., Barroso-Batista, J., Güleresi, D., Lourenço, M., and Gordo, I. (2017). Recurrent Reverse Evolution Maintains Polymorphism after Strong Bottlenecks in Commensal Gut Bacteria. Mol. Biol. Evol. *34*, 2879–2892.
26. Kinnersley, M., Wenger, J., Kroll, E., Adams, J., Sherlock, G., and Rosenzweig, F. (2014). Ex uno plures: clonal reinforcement drives evolution of a simple microbial community. PLoS Genet. *10*, e1004430.

27. Rozen, D.E., Philippe, N., Arjan de Visser, J., Lenski, R.E., and Schneider, D. (2009). Death and cannibalism in a seasonal environment facilitate bacterial coexistence. Ecol. Lett. *12*, 34–44.

28. Treves, D.S., Manning, S., and Adams, J. (1998). Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of Escherichia coli. Mol. Biol. Evol. *15*, 789–797.

29. Rosenzweig, R.F., Sharp, R.R., Treves, D.S., and Adams, J. (1994). Microbial evolution in a simple unstructured environment: genetic differentiation in Escherichia coli. Genetics *137*, 903–917.

30. Kinnersley, M.A., Holben, W.E., and Rosenzweig, F.E. (2009). E Unibus Plurum: Genomic Analysis of an Experimentally Evolved Polymorphism in Escherichia coli. PLoS Genet. *5*, e1000713.

31. Meyer, J.R., Dobias, D.T., Medina, S.J., Servilio, L., Gupta, A., and Lenski, R.E. (2016). Ecological speciation of bacteriophage lambda in allopatry and sympatry. Science *354*, 1301–1304.

32. Tyerman, J., Havard, N., Saxer, G., Travisano, M., and Doebeli, M. (2005). Unparallel diversification in bacterial microcosms. Proc. Biol. Sci. *272*, 1393–1398.

33. Cooper, T.F., and Lenski, R.E. (2010). Experimental evolution with E. coli in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. BMC Evol. Biol. *10*, 11.

34. Frenkel, E.M., McDonald, M.J., Van Dyken, J.D., Kosheleva, K., Lang, G.I., and Desai, M.M. (2015). Crowded growth leads to the spontaneous evolution of semistable coexistence in laboratory yeast populations. Proc. Natl. Acad. Sci. USA *112*, 11306–11311.

35. Flohr, R.C.E., Blom, C.J., Rainey, P.B., and Beaumont, H.J.E. (2013). Founder niche constrains evolutionary adaptive radiation. Proc. Natl. Acad. Sci. USA *110*, 20663–20668.

36. Flynn, K.M., Dowell, G., Johnson, T.M., Koestler, B.J., Waters, C.M., and Cooper, V.S. (2016). Evolution of Ecological Diversity in Biofilms of Pseudomonas aeruginosa by Altered Cyclic Diguanylate Signaling. J. Bacteriol. *198*, 2608–2618.

37. King, T., Ishihama, A., Kori, A., and Ferenci, T. (2004). A regulatory trade-off as a source of strain variation in the species Escherichia coli. J. Bacteriol. *186*, 5614–5620.

38. Maharjan, R.P., Ferenci, T., Reeves, P.R., Li, Y., Liu, B., and Wang, L. (2012). The multiplicity of divergence mechanisms in a single evolving population. Genome Biol. *13*, R41.

39. San Roman, M., and Wagner, A. (2021). Diversity begets diversity during community assembly until ecological limits impose a diversity ceiling. Mol. Ecol. *30*, 5874–5887.

40. Fink, J.W., and Manhart, M. (2023). How do microbes grow in nature? The role of population dynamics in microbial ecology and evolution. Preprint at EcoEvoRxiv.

41. Herren, C.M., Webert, K.C., and McMahon, K.D. (2016). Environmental Disturbances Decrease the Variability of Microbial Populations within Periphyton. mSystems *1*, e00013-16.

42. Kim, M., Heo, E., Kang, H., and Adams, J. (2013). Changes in Soil Bacterial Community Structure with Increasing Disturbance Frequency. Microb. Ecol. *66*, 171–181.

43. Shade, A., Read, J.S., Welkie, D.G., Kratz, T.K., Wu, C.H., and McMahon, K.D. (2011). Resistance, resilience and recovery: aquatic bacterial dynamics after water column disturbance. Environ. Microbiol. *13*, 2752–2767.

44. Dethlefsen, L., and Relman, D.A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc. Natl. Acad. Sci. USA *108* (*Suppl 1*), 4554–4561.

45. Jernberg, C., Löfmark, S., Edlund, C., and Jansson, J.K. (2007). Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. ISME J. *1*, 56–66.

46. Jakobsson, H.E., Jernberg, C., Andersson, A.F., Sjölund-Karlsson, M., Jansson, J.K., and Engstrand, L. (2010). Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. PLoS One *5*, e9836.

47. Xue, K.S., Walton, S.J., Goldman, D.A., Morrison, M.L., Verster, A.J., Parrott, A.B., Yu, F.B., Neff, N.F., Rosenberg, N.A., Ross, B.D., et al. (2023). Prolonged delays in human microbiota transmission after a controlled antibiotic perturbation. Preprint at bioRxiv.

48. Handa, I.T., Aerts, R., Berendse, F., Berg, M.P., Bruder, A., Butenschoen, O., Chauvet, E., Gessner, M.O., Jabiol, J., Makkonen, M., et al. (2014). Consequences of biodiversity loss for litter decomposition across biomes. Nature *509*, 218–221.

49. Philippot, L., Raaijmakers, J.M., Lemanceau, P., and van der Putten, W.H. (2013). Going back to the roots: the microbial ecology of the rhizosphere. Nat. Rev. Microbiol. *11*, 789–799.

50. Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. Nature *405*, 907–913.

51. Grant, P.R., and Grant, B.R. (2002). Unpredictable Evolution in a 30-Year Study of Darwin's Finches. Science *296*, 707–711.

52. Lerner, H.L., Meyer, M., James, H.F., Hofreiter, M., and Fleischer, R.C. (2011). Multilocus Resolution of Phylogeny and Timescale in the Extant Adaptive Radiation of Hawaiian Honeycreepers. Curr. Biol. *21*, 1838–1844.

53. Genner, M.J., Seehausen, O., Lunt, D.H., Joyce, D.A., Shaw, P.W., Carvalho, G.R., and Turner, G.F. (2007). Age of Cichlids: New Dates for Ancient Lake Fish Radiations. Mol. Biol. Evol. *24*, 1269–1282.

54. Doebeli, M. (2002). A model for the evolutionary dynamics of cross-feeding polymorphisms in microorganisms. Popul. Ecol. *44*, 59–70.

55. Peng, F., Widmann, S., Wünsche, A., Duan, K., Donovan, K.A., Dobson, R.C.J., Lenski, R.E., and Cooper, T.F. (2018). Effects of Beneficial Mutations in pykF Gene Vary over Time and across Replicate Populations in a Long-Term Experiment with Bacteria. Mol. Biol. Evol. *35*, 202–210.

56. Good, B.H., Martis, S., and Hallatschek, O. (2018). Adaptation limits ecological diversification and promotes ecological tinkering during the competition for substitutable resources. Proc. Natl. Acad. Sci. USA *115*, E10407–E10416.

57. Lenski, R.E. (2017). Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. ISME J. *11*, 2181–2194.

58. Großkopf, T., Consuegra, J., Gaffé, J., Willison, J.C., Lenski, R.E., Soyer, O.S., and Schneider, D. (2016). Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment. BMC Evol. Biol. *16*, 163.

59. Rozen, D.E., Schneider, D., and Lenski, R.E. (2005). Long-Term Experimental Evolution in Escherichia coli. XIII. Phylogenetic History of a Balanced Polymorphism. J. Mol. Evol. *61*, 171–180.

60. Le Gac, M., Plucain, J., Hindré, T., Lenski, R.E., and Schneider, D. (2012). Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with Escherichia coli. Proc. Natl. Acad. Sci. USA *96*, 10242–10247.

61. Plucain, J., Hindré, T., Le Gac, M., Tenaillon, O., Cruveiller, S., Médigue, C., Leiby, N., Harcombe, W.R., Marx, C.J., Lenski, R.E., et al. (2014). Epistasis and allele specificity in the emergence of a stable polymorphism in Escherichia coli. Science *343*, 1366–1369.

62. Good, B.H., McDonald, M.J., Barrick, J.E., Lenski, R.E., and Desai, M.M. (2017). The dynamics of molecular evolution over 60,000 generations. Nature *551*, 45–50.

63. Ascensao, J.A., Wetmore, K.M., Good, B.H., Arkin, A.P., and Hallatschek, O. (2023). Quantifying the local adaptive landscape of a nascent bacterial community. Nat. Commun. *14*, 1–19.

64. Vasi, F.K., Travisano, M., and Lenski, R.E. (1994). Long-term experimental evolution in *Escherichia coli*. ii. changes in life-history traits during adaptation to a seasonal environment. Am. Nat. *144*, 432–456.

65. Holms, H. (1996). Flux analysis and control of the central metabolic pathways in Escherichia coli. FEMS Microbiol. Rev. *19*, 85–116.

66. Mukherjee, A., Ealy, J., Huang, Y., Benites, N.C., Polk, M., and Basan, M. (2023). Coexisting ecotypes in long-term evolution emerged from interacting trade-offs. Nat. Commun. *14*, 3805.

67. Sniegowski, P.D., Gerrish, P.J., and Lenski, R.E. (1997). Evolution of high mutation rates in experimental populations of E. coli. Nature *387*, 703–705.

68. Favate, J.S., Liang, S., Cope, A.L., Yadavalli, S.S., and Shah, P. (2022). The landscape of transcriptional and translational changes over 22 years of bacterial adaptation. eLife *11*, e81979.

69. Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *28*, 27–30.

70. Lo, C.-A., Kays, I., Emran, F., Lin, T., Cvetkovska, V., and Chen, B. (2015). Quantification of Protein Levels in Single Living Cells. Cell Rep. *13*, 2634–2644.

71. Kannan, S., Sams, T., Maury, J., and Workman, C.T. (2018). Reconstructing Dynamic Promoter Activity Profiles from Reporter Gene Data. ACS Synth. Biol. *7*, 832–841.

72. Kelly, J.R., Rubin, A.J., Davis, J.H., Ajo-Franklin, C.M., Cumbers, J., Czar, M.J., de Mora, K., Glieberman, A.L., Monie, D.D., and Endy, D. (2009). Measuring the activity of BioBrick promoters using an in vivo reference standard. J. Biol. Eng. *3*, 4.

73. Ducrest, A.-L., Amacker, M., Lingner, J., and Nabholz, M. (2002). Detection of promoter activity by flow cytometric analysis of GFP reporter expression. Nucleic Acids Res. *30*, e65.

74. Reppas, N.B., Wade, J.T., Church, G.M., and Struhl, K. (2006). The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. Mol. Cell *24*, 747–757.

75. Ishihama, A. (2000). Functional Modulation of Escherichia Coli RNA Polymerase. Annu. Rev. Microbiol. *54*, 499–518.

76. Sharma, U.K., and Chatterji, D. (2010). Transcriptional switching in Escherichia coli during stress and starvation by modulation of sigma activity. FEMS Microbiol. Rev. *34*, 646–657.

77. Nyström, T. (2004). MicroReview: Growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition? Mol. Microbiol. *54*, 855–862.

78. Basan, M., Honda, T., Christodoulou, D., Hörl, M., Chang, Y.F., Leoncini, E., Mukherjee, A., Okano, H., Taylor, B.R., Silverman, J.M., et al. (2020). A universal trade-off between growth and lag in fluctuating environments. Nature *584*, 470–474.

79. Blount, Z.D., Lenski, R.E., and Losos, J.B. (2018). Contingency and determinism in evolution: Replaying life's tape. Science *362*, eaam5979.

80. Beatty, J. (2006). Replaying Life's Tape. J. Philos. *103*, 336–362.

81. Gould, S.J. (1989). Wonderful Life (WW Norton).

82. Besnard, F., Picao-Osorio, J., Dubois, C., and Félix, M.A. (2020). A broad mutational target explains a fast rate of phenotypic evolution. eLife *9*, 1–70.

83. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell *169*, 1177–1186.

84. Scott, M., and Hwa, T. (2011). Bacterial growth laws and their applications. Curr. Opin. Biotechnol. *22*, 559–565.

85. Erickson, D.W., Schink, S.J., Patsalo, V., Williamson, J.R., Gerland, U., and Hwa, T. (2017). A global resource allocation strategy governs growth transition kinetics of Escherichia coli. Nature *551*, 119–123.

86. Forchhammer, J., and Lindahl, L. (1971). Growth rate of polypeptide chains as a function of the cell growth rate in a mutant of Escherichia coli 15. J. Mol. Biol. *55*, 563–568.

87. Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and gene expression: Origins and consequences. Science *330*, 1099–1102.

88. Levy, S.F., Blundell, J.R., Venkataram, S., Petrov, D.A., Fisher, D.S., and Sherlock, G. (2015). Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature *519*, 181–186.

89. Ai, H.W., Shaner, N.C., Cheng, Z., Tsien, R.Y., and Campbell, R.E. (2007). Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. Biochemistry *46*, 5904–5910.

90. Schlechter, R.O., Jun, H., Bernach, M., Oso, S., Boyd, E., Muñoz-Lintz, D.A., Dobson, R.C.J., Remus, D.M., and Remus-Emsermann, M.N.P. (2018). Chromatic Bacteria - A Broad Host-Range Plasmid and Chromosomal Insertion Toolbox for Fluorescent Protein Expression in Bacteria. Front. Microbiol. *9*, 3052.

91. parts. igem.org. (2006). Part:BBa J23119. https://parts.igem.org/Part:BBa_J23119.

92. Reis, A.C., and Salis, H.M. (2020). An automated model test system for systematic development and improvement of gene expression models. ACS Synth. Biol. *9*, 3145–3156.

93. Yurtsev, E., Friedman, J., and Gore, J. (2015). FlowCytometryTools. Version 0.4.5 (Zenodo). https://zenodo.org/records/32991.

94. Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100.

95. Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. Nat. Methods *15*, 461–468.

96. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915.

97. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

98. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

99. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141.

100. Kremers, G.J., Goedhart, J., Van Munster, E.B., and Gadella, T.W. (2006). Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Förster radius. Biochemistry *45*, 6570–6580.

101. Bagwell, C.B. (2005). Hyperlog-a flexible log-like transform for negative, zero, and positive valued data. Cytometry A *64*, 34–42.

102. Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. Austral Ecol. *26*, 32–46.

103. Zwietering, M.H., Jongenburger, I., Rombouts, F.M., and van 't Riet, K. (1990). Modeling of the Bacterial Growth Curve. Appl. Environ. Microbiol. *56*, 1875–1881.

104. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nat. Methods *17*, 261–272.

105. Deatherage, D.E., and Barrick, J.E. (2014). Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol. Biol. *1151*, 165–188.

106. Jeong, H., Barbe, V., Lee, C.H., Vallenet, D., Yu, D.S., Choi, S., Couloux, A., Lee, S., Yoon, S.H., Cattolico, L., et al. (2009). Genome Sequences of Escherichia coli B strains REL606 and BL21(DE3). J. Mol. Biol. *394*, 644–652.

107. Stephens, M. (2017). False discovery rates: a new deal. Biostatistics *18*, 275–294.

108. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Various LTEE-derived *E. coli* strains	This paper	See Table S1
Chemicals, peptides, and recombinant proteins		
RNAse-free DNAse	Invitrogen	Catalogue number: AM2222
Ready-Lyse Lysozyme Solution	Lucigen	Catalogue number: R1804M
Critical commercial assays		
Monarch Total RNA Miniprep Kit	New England BioLabs	Catalogue number: T2010S
Stranded Total RNA Prep, Ligation with Ribo-Zero Plus	Illumina	Catalogue number: 20040529
DNeasy Blood and Tissue Kit	Qiagen	Catalogue number: 69504
NEBNext DNA Library Prep kit	New England BioLabs	Catalogue number: E7645
Nanopore ligation sequencing kit	Oxford Nanopore	Catalogue number: SQK-LSK109
Deposited data		
All DNA/RNA sequencing data	This paper	NCBI BioProject accession: PRJNA970313
REL606 genome	Genbank	NCBI RefSeq assembly: GCF_000017985.1
Processed data	This paper	https://github.com/joaoascensao/Rediversification
Oligonucleotides		
Oligonucleotides for genotyping and cloning	This paper	See Table S2
Recombinant DNA		
pBad-EBFP2 (plasmid with eBFP2 gene)	Ai et al.[89]	Addgene catalog number: 14891
pMRE-Tn7-133 (miniTn7 plasmid to insert sYFP2 into genome at attTn7)	Schlechter et al.[90]	Addgene catalog number: 118551
pJA17 (miniTn7 plasmid to insert eBFP2 with strong promoter into genome at attTn7)	Designed in this paper; Backbone from pMRE-Tn7-133; eBFP2 from pBad-EBFP2; BBa_J23119 promoter[91]; RBS designed in silico[92]	N/A
pJA18 (miniTn7 plasmid to insert sYFP2 with strong promoter into genome at attTn7)	Designed in this paper; Backbone from pMRE-Tn7-133; BBa_J23119 promoter[91]; RBS designed in silico[92]	N/A
Software and algorithms		
Code	This paper	https://github.com/joaoascensao/Rediversification
flowcytometrytools (v0.4.5)	Yurtsev et al.[93]	https://eyurtsev.github.io/FlowCytometryTools/index.html
minimap2 (v2.26)	Li[94]	https://github.com/lh3/minimap2
Sniffles (v2.2)	Sedlazeck et al.[95]	https://github.com/fritzsedlazeck/Sniffles
HISAT2 (v2.2.0)	Kim et al.[96]	https://daehwankimlab.github.io/hisat2/
featureCounts (v2.0.1)	Liao et al.[97]	https://subread.sourceforge.net/featureCounts.html
DESeq2 (v1.38.3)	Love et al.[98]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
clusterProfiler (v4.6.2)	Wu et al.[99]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
Other		
Attune Flow Cytometer (2017 model)	ThermoFisher	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SpectraMax 190 (shaking plate reader)	Molecular Devices	N/A
Pippin Prep	Sage Science	N/A
Femto pulse system	Aligent	N/A

RESOURCE AVAILABILITY

Lead contact

Further information regarding the manuscript and requests for reagents may be directed to, and will be fulfilled by the lead contact, Oskar Hallatschek (ohallats@berkeley.edu).

Materials availability

All newly constructed strains and plasmids presented in this paper are available upon request.

Data and code availability

All raw genomic and transcriptomic data has been deposited at the NCBI Sequence Read Archive (SRA), accession numbers are listed in the key resources table. Code and processed data are available at https://github.com/joaoascensao/Rediversification. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Most of the experiments presented here were performed in Davis Minimal Media (DM) base [5.36 g/L potassium phosphate (dibasic), 2g/L potassium phosphate (monobasic), 1g/L ammonium sulfate, 0.5g/L sodium citrate, 0.01% Magnesium sulfate, 0.0002% Thiamine HCl]. The media used in the LTEE and the competitions shown in Figures 1 and 2 is DM25, that is DM supplement with 25mg/L glucose. The strains used in this work were all isolated from the *E. coli* Long-Term Evolution Experiment (LTEE), and listed in Table S1.

For competition experiments, generally we first inoculated the strain into 1mL LB + 0.2% glucose + 20mM pyruvate (which we found prevented the emergence of the S_B while allowing for robust growth). After overnight growth, we washed the culture 3 times in DM0 (DM without a carbon source added) by centrifuging it at 2500xg for 3 minutes, aspirating the supernatant, and resuspending in DM0. We transferred the washed culture 1:1000 into DM25 in a glass tube. If a strain was isolated directly from a colony, we would instead directly resuspend the colony in DM25. Generally, we grew 1mL cultures in a glass 96 well plate (Thomas Scientific 6977B05). We then grew the culture for 24 hours at 37°C in a shaking incubator. The next day, we transferred all the cultures 1:100 again into 1mL DM25. After another 24 hours of growth under the same conditions, we would mix selected cultures at desired frequencies, then transfer the mixture 1:100 to DM25. After another 24 hours of growth under the same conditions, we would transfer the culture 1:100 to a desired media and start taking flow cytometry measurements–in the competitions of Figures 1 and 2, the media is DM25, for the competitions of Figure 3, the media is DM supplemented with 200mg/L acetate, 1mg/mL casamino acids, 20mM pyruvate, or 20mM glycerol. For the competitions of Figure 1, we took measurements for 3-4 total days, doing 1:100 serial transfers every 24 hours in DM25; for Figure 2 we took measurements approximately every hour for 8 hours, then another measurement at 24 hours; for Figure 3 we took a second measurement after 8 hours, when the cultures were still in exponential phase. For growth in glucose/acetate media, we grew all strains in DM + 250mg/L acetate + 250mg/L glucose, after three cycles of growth on DM25, measuring OD600 absorbance in a shaking plate reader (SpectraMax 190; Molecular Devices) over the course of 24 hours. For the competitions in the heat-killed cultures, we grew 6.5k *S* 1, *L* 1, and S_B 1 cultures overnight in DM2000. We washed all the cultures in DM0 3x, as described above, to eliminate the presence of extracellular metabolites. We then heat-killed a portion of the *S* 1 culture by incubating it at 70°C for 45 minutes. We then mixed 6.5k *S* 1, *L* 1, and S_B 1 cultures appropriately, took a flow cytometry measurement, and resuspended them 1:1000 in 1mL of the heat-killed culture. We allowed the cultures to grow at 37°C for 16 hours (the approximate length of stationary phase), and then took another measurement.

METHOD DETAILS

Integration of fluorescent proteins

We sought to use flow cytometry to quantify ecotype abundances, which would necessitate that we could differentiate the strains via fluorescence. We decided to integrate fluorescent proteins into a neutral genomic location of our various strains rather than using plasmids, because plasmids can carry a significant metabolic burden, and it is often necessary to add antibiotics to the media to select against plasmid loss. We used a system based on that of Schlechter et al.[90] to integrate fluorescent proteins with miniTn7, a transposon that inserts cargo at a putatively neutral intergenic site downstream of *glmS*. Briefly, the system works by mating the recipient strain-of-interest with a donor strain, harboring a plasmid with the miniTn7 proteins, an ampicillin-resistance gene, a temperature-dependent origin of replication, and the cargo flanked by the left and right Tn7 recognition sites. In this case, the cargo

consists of a fluorescent protein, under the control of a broad host-range promoter, and a chloramphenicol resistance gene, for selection of integration.

Our protocol for integration proceeded as follows. First, we grew the donor strain with the desired plasmid in LB + 100μg/mL carbenicillin + 10μg/mL chloramphenicol at 30°C shaken, overnight. We also grew the recipient strain overnight in DM2000 media at 37°C, directly from glycerol stock. The next day, we washed the donor culture by centrifuging it at 2500xg for 3 minutes, aspirating the supernatant, and resuspending in DM0. We then measured the optical density (OD) of both cultures, and mixed about 1 $OD \cdot mL$ of each culture on a 20mL LB/agar plate supplemented with 0.2% glucose + 20mM pyruvate. The cultures were allowed to grow into a lawn overnight at 30°C, allowing the donor strain to conjugate with the recipient. Afterwards, we scraped up the lawn and resuspended it in 3mL DM0. We washed the resuspended culture 3 times, as previously described, and then streaked out the culture on a DM2000 + 10μg/mL chloramphenicol + agar plate, then allowing the plates to incubate overnight at 37°C. This step simultaneously selects against the presence of the donor (the donor is a proline auxotroph), against the Tn7 plasmid (it has a temperature-sensitive origin of replication), and for integration of the Tn7 cargo (via the chloramphenicol resistance gene). After two days of growth, we restreaked a number of colonies that appeared on DM2000/agar plates for isolation. We then tested for integration of the Tn7 cargo by amplifying and sanger sequencing the junction between the genome and the fluorescent protein insertion (see Table S2 for oligonucleotide sequences), and by looking for fluorescence via fluorescence microscopy. We confirmed that the plasmid was not present in the colony by testing resistance against carbenicillin. We ensured that the colony was not the donor or a contaminant by checking colony morphologies on tetrazolium -maltose (TM), -arabinose (TA), and -xylose (TX) agar plates. We further confirmed identity by sanger sequencing the *arcA* and *aspS* loci of the clones we moved forward with (see Table S2 for oligonucleotide sequences).

We found that the fluorescence provided by the plasmids designed in Schlechter et al.[90] were insufficiently strong for our purposes. We also needed two different fluorescent proteins with non-overlapping fluorescence profiles so that we could distinguish the two in our flow cytometer. We decided to use the fluorescent proteins sYFP2[100] and eBFP2[89] because they share the same ancestor and are highly homologous, and are thus likely to have the same or similar physiological effects on their host, and they have sufficiently different fluorescence profiles that are compatible with our flow cytometer. Thus, we sought to increase the expression levels of the fluorescent proteins, and add in BFP, by constructing new plasmids. We chose to use the strong, constitutive σ^{70} BBa_J23119 promoter[91] and a ribosome binding site (RBS) designed *in silico* with the Salis lab "RBS calculator"[92], placing them immediately upstream of the fluorescent protein sequences. We used Gibson assembly to construct the plasmids by ordering compatible oligonucleotides with the promoter and RBS sequences on them, and then using the backbone of pMRE-Tn7-133 from Schlechter et al.[90] and the eBFP2 gene from pBad-EBFP2[89] for the BFP plasmid. Final plasmid sequences were confirmed via sanger sequencing.

Flow cytometry

For all population measurements taken with flow cytometry, we used the ThermoFisher Attune Flow Cytometer (2017 model) at the UC Berkeley QB3 Cell and Tissue Analysis Facility (CTAF). For every measurement, we loaded the samples into a round bottom 96 well plate, for use with the autosampler. Typically we diluted the samples 1:5 in DM0, but we changed the dilution rate over the course of the 8 hour within-cycle timecourse. We set the flow cytometer to perform one washing and mixing cycle before each measurement, and ran 50μL of bleach through the autosampler in between each measurement to ensure that there was no cross-contamination between wells. We used the "VL1" channel to detect eBFP2 fluorescence, which uses a 405nm laser and a 440/50nm bandpass emission filter. We used the "BL1" channel to detect sYFP2 fluorescence, which uses a 488nm laser and a 530/30nm bandpass emission filter. For the triple competitions shown in Figure 1E, we used a BFP-tagged *S*, a YFP-tagged S_B, and a non -fluorescent *L* strain. To estimate the frequency of *L*, we added 5 μM of Syto62 red fluorescent dye (ThermoFisher S11344) to the sample immediately before measurement. We used the "RL1" channel to detect Syto62 fluorescence, which uses a 637nm laser and a 670/14nm bandpass emission filter. We always used a sample flow rate of 25μL/min.

We use the package flowcytometrytools (v0.4.5)[93] to analyze the flow cytometry data. We first perform a hyperlog transform[101] and then created threshold gates to sufficiently separate the "noise cloud" (nonfluorescent particles present even when running blank media) from particles with clear fluorescence. We noticed that in addition to seeing single positive BFP^+ and YFP^+ particles, we also see some particles called as fluorescent in both channels (Figure S2A). We observed that the proportion of double positive events decreased as a function of fluid flow rate and dilution rate (Figure S2B), suggesting that sometimes multiple cells end up in front of the flow cytometry laser at the same time, and are counted as one event. Thus, we sought to correct for this effect. We assume that the probability of a cell ending up in front of the laser is constant per unit time, and uncorrelated in time, i.e. that it is a Poisson process. Thus, for any given window of time, the probability of observing some number of events is distributed as a Poisson distribution. So under this model, the observed BFP or YFP ‘‘clouds’’ will consist of single cells, double cells, triple cells, and so on. Similarly, there are many combinations of BFP/YFP cells that can end up in the double positive cloud. So, in order to get the expectation of the observed frequencies, we add up the contributions of singlets, doublets, triplets, etc by considering the probability of *n* cells passing in front of the laser together times the probability of all *n* cells being the same color,

$$f_i^{obs} = \sum_{n=1}^{\infty} p(n \text{ cells}) f_i^n \qquad \text{(Equation 1)}$$

where $i \in \{1,2\}$. As previously mentioned, $p(n \text{ cells})$ will follow a Poisson distribution, but as we do not observe the case when zero cells pass in front of the laser, we will use a zero-truncated Poisson.

$$f_i^{obs} = \sum_{n=1}^{\infty} \frac{\lambda^n}{n!(e^\lambda - 1)} f_i^n = \frac{e^{\lambda f_i} - 1}{e^\lambda - 1} \quad \text{(Equation 2)}$$

Where λ is the average number of cells per event. We have two equations (for f_1^{obs} and f_2^{obs}) and two unknowns (λ and f_1), so we can solve for the real frequencies, which we solve for via numerical root-solving. The total cell count N also must be corrected, where $N_{corrected} = N_{observed}\lambda e^\lambda/(e^\lambda - 1)$. The post-correction frequencies appear to be well-reflective of frequencies measured with colony counting (Figure S1). The primary reason why we chose to use a mathematical correction rather than diluting the samples to the point where $\lambda \approx 0$ was for time and efficiency. We found that in order to reduce the number of mixed events to near zero, we would have to run a much larger volume through the flow cytometer, which takes much more time. This is especially problematic for the growth curve experiments in Figure 2, where the dynamics are quite rapid, and long times spent in the flow cytometer would likely distort the data.

QUANTIFICATION AND STATISTICAL ANALYSIS

Once we obtained estimates of strain frequencies and total abundances, we can calculate several downstream metrics of the dynamics. Throughout the manuscript, we estimate the fitness effect of a strain s from the dynamics of ecotype frequencies, $f(t)$, using the model

$$\text{logit } f(t) = st + \text{logit } f_0 + \varepsilon_t \,. \quad \text{(Equation 3)}$$

We measure time t in units of 24-hour growth cycles, and thus fitness effects are measured in units of 1/cycle. We fit the model using ordinary least squares, jointly estimating s and the intercept f_0. We calculated p-values for differences in fitness effects as a function of initial ecotype frequency (Figures 1B–1D) using a standard two-sample t-test, then performed a standard Benjamini-Hochberg FDR correction, pooled across all of the comparisons.

Once we obtained fitness data to measure growth traits in novel environments (Figure 3), we were first interested in testing how often fitness effects are significantly non-neutral relative to S (i.e. deviate from 0). Thus, we performed standard two-sided, one-sample t-tests on all environment-strain conditions, then corrected the p-values with a Benjamini-Hochberg FDR correction (Table S4). To test if S_B clones (within a timepoint) had significantly different fitness effects in each condition, we used two-sided, two-sample t-tests, then again corrected with a Benjamini-Hochberg FDR correction. Then, we were interested in testing the hypothesis that the S_B clones within a timepoint cluster together more strongly than with the L clone within the timepoint. Thus, we turned to using a slightly modified version of permutational multivariate analysis of variance (PERMANOVA),[102] to better reflect the structure of our data, where we have multiple biological replicates for each measurement. To calculate the F statistic, we first computed the mean relative fitness effect of each strain in each environment, across biological replicates, $\bar{s}$. We used total Euclidean distance between mean fitness effects as the metric, i.e. the total squared distance is the sum over all environments of the squared difference in fitness between two strains. We only computed the distance for the four strains (S_B clones 1-3, L clone) within a timepoint. As previously described,[102] and without modification, we used the total sum-of-squares and the within groups sum-of-squares to compute the F statistic, using the distance metric. Then, we needed to estimate a null distribution to compare the F statistic. If we were to estimate the null distribution with the standard method, we would only get 3 values (treating either S_B clone 1, 2, or 3 as the outgroup), which is not sufficient to construct a null distribution. However, the estimated mean fitness effects are calculated from a finite number of biological replicates; with resampling, the mean will change. Thus, we use a parametric resampling scheme to model the variability from sampling. We resampled each mean fitness effect 10,000 times using a Student's t distribution, $s_{rs} = \bar{s} - t\sqrt{\text{vars}/n}$, where s_{rs} is the resampled fitness effect, $\bar{s}$ is the empirical mean fitness effect, vars is the empirical (unbiased) variance across biological replicates, n is the number of replicates, and $t \sim t_{n-1}$ is drawn from a t-distribution with $n - 1$ degrees of freedom. We get similar results if we resample using a gaussian distribution instead of a t distribution. We then treat either S_B clone 1, 2, or 3 as the outgroup (instead of L), and repeat the multivariate ANOVA procedure to get F statistics for each resampled sample. We concatenate the samples from cases where each of the S_B clones is treated as the outgroup, so that the final size of the null distribution is 30,000 samples. We calculate the p-value as the fraction of values in the null distribution that are larger than the F statistic.

We calculated empirical growth rates (Figures 2 and S4) over the course of the 24 hour cycle using the total ecotype abundance data, $n(t)$. For each pair of adjacent time points in a cycle, we used ordinary least squares to extract estimates of the growth rate (r) and its standard error, using the model $\log n(t) = rt + n_0 + \varepsilon_t$. Growth rate estimates were compared using Wald's test, and p-values were corrected via a Benjamini-Hochberg FDR correction.

We fit generalized logistic models (Richard's curves) to the 24 hours growth curve abundance data to extract estimates of lag times for each strain and condition (Figures 4A and S5).[103] For the OD600 data in Figure 4A, we used the timecourse after the end of glucose exponential phase. Denoting the fitted abundance as $\hat{n}(t)$, we use the following form for the generalized logistic model:

$$\hat{n}(t) = a\left(1 + Te^{-k(t-t_m)}\right)^{-1/T} + b, \quad \text{(Equation 4)}$$

We jointly infer the parameters a, T, b, k, t_m by using inverse-variance weighted least squares, i.e. min $\sum_i \sum_j (n_j(t_i) - \hat{n}_j(t_i))^2 / v_{i,j}$, where i labels time points and j labels biological replicates. We use $v_{i,j}$ as the variance of the measurement error, which is taken to be the variance of a Poisson random variable. We implement the global minimization by using the differential evolution optimization algorithm implemented in scipy.[104] We used standard bootstrapping to estimate the sampling distributions of the lag times, L, with 1000 resamples. We removed outliers from the bootstrapped distributions by first robustly estimating the standard deviation, $\hat{\sigma}$, of the distribution via the median absolute deviation (MAD), $\hat{\sigma} = \text{med}|L - \text{med}\, L|/0.67449$. Then we counted a bootstrapped replicate as an outlier, and discarded it, if it was more than 3 standard deviations away from the median, $3\hat{\sigma} > |L - \text{med}\, L|$. We then used the bootstrapped distributions to compute confidence intervals and p-values.

Whole genome sequencing

To perform short-read sequencing of S_B and S clones (see SI), we first grew the clones overnight in 1mL of DM2000, then pelleted the cultures and extracted genomic DNA with the DNeasy Blood and Tissue Kit (Qiagen 69504). We prepared the sample libraries with NEBNext DNA Library Prep kit for Illumina according to the manufacturer's protocol (New England Biolabs E7645). We sequenced the samples with the Illumina 4000 HiSeq 150PE. We used breseq (v0.33.2)[105] to compare raw reads to the REL606 genome[106] (GenBank: CP000819.1) and to the S ancestor of each S_B, and then call genetic variants. Read coverage was around 100x across the genome, for all samples. We used default parameters for the breseq pipeline, which uses a bayesian model to call single nucleotide polymorphisms, incorporating information from the FASTQ phred quality score from each read.[105]

We attempted to determine if there was any parallelism between 6.5k S_B clone mutations on the level of KEGG annotations, focusing on nonsynonymous and indel mutations. We excluded 6.5k S_B clone 4 from the analysis, as it is a sister to clone 2. We first compiled all KEGG annotations of all genes with nonsynonymous and indel mutations across the 6.5k S_B clones, and computed how many times a gene mapped to a given annotation. For each annotation that appeared, only one gene in our set mapped to it. We then expanded our list to include genes immediately adjacent to intergenic mutations, as well as pseudogenes affected by mutations. In the expanded gene set, we see two genes each map to three different annotations (carbon metabolism, exopolysaccharide biosynthesis, sulfur metabolism). We had planned to implement the multiplicity test to detect parallelism presented in Good et al.,[62] however, they recommend focusing on set items with 3 or more hits to avoid false positives from low counts. We thus do not believe that there is any parallelism between mutations on the level of KEGG annotations.

To perform long-read sequencing of S_B and S clones (see SI), we again grew the clones overnight in 1mL of DM2000, then pelleted the cultures. High-molecular weight DNA extraction was performed via a standard phenol-chloroform extraction and isopropanol precipitation. Distribution of DNA fragment sizes were obtained using the Agilent Femto Pulse System. Fragment size selection was performed using Pippin Prep (Sage Science). The samples were prepared for sequencing with the Nanopore ligation sequencing kit (Oxford Nanopore, SQK-LSK109). The libraries were then sequenced on an Oxford Nanopore MinION. We used minimap2 (v2.26)[94] and sniffles (v2.2)[95] with default parameters to detect structural variants.

RNA sequencing

6.5k S and L clones 1 and 2 were isolated from REL11555 and REL11556 respectively; 6.5k S_B clones 1 and 2 were the same clones as previously described. Cultures of 6.5k S, S_B, and L clones 1 and 2 were started directly from glycerol stock into 1ml LB + 2g/L glucose + 20mM pyruvate, as a pre-culture. We started two independent cultures for each clone as biological replicates. After overnight growth, the cultures were washed by centrifuging the cultures at 2500xg for 3 minutes, aspirating the supernatant, and resuspending in DM0, repeated three times. Then, the cultures were diluted $1:10^{-4}$ into 1mL fresh DM media supplemented with 4g/L glucose, in glass tubes. After approximately four hours of growth at 37°C, the cultures were again diluted 1 : 50 in 1mL of the same media in glass tubes. The cultures were grown shaken at 37°C. The cultures were grown to mid-exponential phase, i.e. until $OD \sim 0.4$, then the entire culture was immediately centrifuged at 2500xg for 3 minutes to pellet. Immediately after centrifugation, we resuspended the pellets in 25μL TES buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA, and 100 mM NaCl) and then lysed the pelleted cultures with 250U/μL lysozyme (Ready-Lyse Lysozyme Solution; Lucigen R1804M) at room temperature for 5 minutes. For all subsequent steps, we used Monarch Total RNA Miniprep Kit (New England BioLabs T2010S) according to the standard given protocol for gram-negative bacteria. Samples were eluted in 30μl nuclease-free water, and stored at -80°C. The concentration and purity of all RNA samples was quantified using Qubit.

RNAse-free DNAse (Invitrogen AM2222) was used to treat the samples for DNA removal. The library preparation was conducted using Illumina's Stranded Total RNA Prep Ligation with Ribo-Zero Plus kit and 10bp IDT for Illumina indices. Subsequently, the samples were sequenced using NextSeq2000, resulting in 2x51bp reads. The process of demultiplexing, quality control, and adapter trimming was carried out using bcl-convert (v3.9.3) and bcl2fast (v2.20.0.445) (both are proprietary Illumina software for the conversion of bcl files to basecalls). HISAT2 (v2.2.0)[96] was used for read mapping. Reads were mapped to the REL606 genome[106] (GenBank: CP000819.1). The read quantification was performed using the functionality of featureCounts (v2.0.1) in Subread.[97] All of the above steps in the pipeline were performed with default parameters, the last two steps also were run with –very-sensitive and -Q 20 tags, respectively. All sequencing and pre-processing steps were performed by SeqCenter, LLC.

After pre-processing, we obtained a matrix of read counts for each gene for each sample. With this table, we used DESeq2 (v1.38.3)[98] to compute fold change in expression between strains and variance-stabilized relative expression values for each gene across samples (blindly with respect to the design matrix), all with default parameters. We used the variance-stabilized relative

expression values for the principal components analysis (PCA). We used the ashr method (v2.2)[107] with default parameters to shrink and regularize the $\log_2$ fold changes in expression. We computed $\log_2$ fold change in expression between samples in two ways, (1) treating the S_B clones as one "strain", and (2) treating the S_B clones as separate, so that we get different fold changes in expression for each clone. Otherwise, for *S* and *L*, we pooled data across the two clones and biological replicates when computing fold change in expression. We used the package clusterProfiler (v4.6.2)[99] to perform the KEGG gene set enrichment analysis (GSEA).[108] We used the previously computed $\log_2$ fold change in expression as the metric to pre-sort the list of genes. We used the gseKEGG method along with the parameters organism="ebr", nPerm=1000000, minGSSize=3, maxGSSize=800, eps=1e-20 to perform the analysis.

Current Biology, Volume *34*

Supplemental Information

Rediversification following ecotype isolation reveals hidden adaptive potential

Joao A. Ascensao, Jonas Denk, Kristen Lok, QinQin Yu, Kelly M. Wetmore, and Oskar Hallatschek

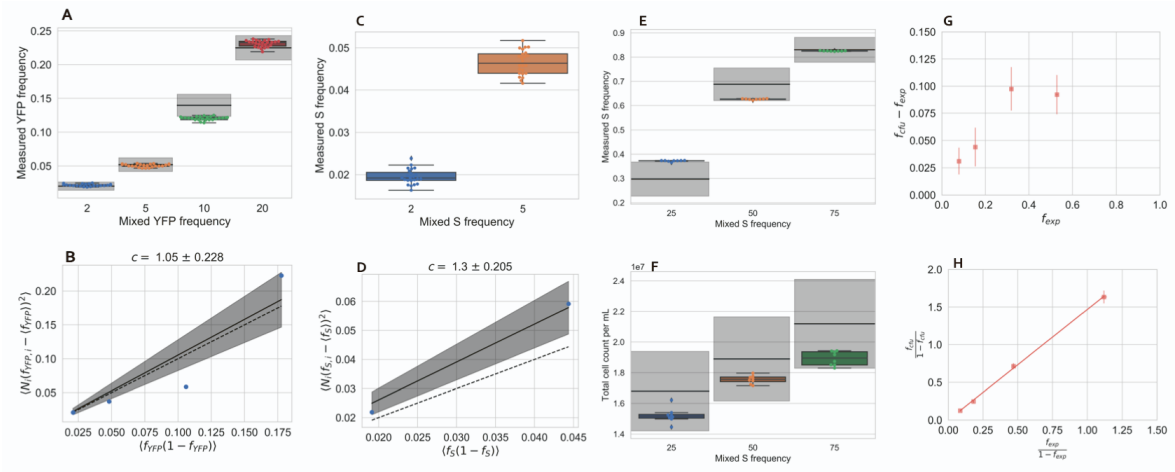

Figure S1: Accuracy of flow cytometer in recovering frequencies of fluorescently-labeled strains in a population. Related to Figure 1. (**A**) We cultured YFP- and BFP-tagged versions of REL606 in DM25, then mixed the cultures at four different frequencies. After culturing for another serial dilution round, we measured each culture 24 times in the flow cytometer (i.e. technical replicates). We also plated the cultures on LB/agar plates and counted the number of blue/yellow colonies for each culture to get CFUs. The colored points represent the flow cytometry measurements. The black lines represent frequencies from CFUs for each of the four cultures, ± standard error. We see that the CFU and flow cytometry measurements mostly match up. The x-axis labels refer to the initial volumetric mixing fraction of the two strains. (**B**) A plot showing the (scaled) mean-variance relationship, where the variables are transformed such that we would expect a linear relationship with a slope of 1 if the error associated with measuring frequencies followed a binomial, and ≥ 1 if it was overdispersed. $f_{YFP,i}$ represents the relative frequency of YFP cells in the population, for replicate i, and N_i represents the total number of events from the flow cytometer. The blue points are measurements from each experiment, the solid line is the fit to the points, and the dashed line is the binomial expectation where $x = y$. Indeed, we see that error looks approximately binomial. We fit the regression line with ordinary least squares (without and intercept) and obtained the standard error using standard bootstrapping on the replicates. (**C,D**) We cultured YFP- and BFP-tagged versions of 6.5k L 1 and S 1 respectively in DM25, then mixed the cultures at two different frequencies. After culturing for another serial dilution round, we measured each culture 24 times in the flow cytometer (i.e. technical replicates). (**C**) The colored points represent the flow cytometry measurements. The x-axis labels refer to the initial volumetric mixing fraction of the two strains. (**D**) A plot showing the (scaled) mean-variance relationship, similar to panel B. The blue points are measurements from each experiment, the solid line is the fit to the points, and the dashed line is the binomial expectation where $x = y$. Again, we see that measurement error is consistent with a binomial. We fit the regression line with ordinary least squares (without and intercept) and obtained the standard

error using standard bootstrapping on the replicates. (**E,F**) In a separate set of experiments, we again cultured YFP- and BFP-tagged versions of 6.5k L 1 and S 1 respectively in DM25, then mixed the cultures at three different frequencies. We also plated the cultures on standard LB agar plates, and counted the number of BFP^+ and YFP^+ colonies on the plate after overnight growth. The colored points show flow cytometry technical replicates for each conditions, and the black lines show the CFU-based measurements (shaded region is 95% CI). The (**E**) frequencies and (**F**) total cell counts measured via flow cytometry and CFUs on LB plates mostly match up. The cell counts measured by flow cytometry lie within the CFU confidence intervals, they may be slightly biased downwards ($p = 0.11$; linear model without intercept) by a factor of about 0.9. The frequency measurements do not appear to be biased. (**G,H**) In another separate experiment, we cultivated four 6.5k L 1 and S 1 cultures at different frequencies, and measured frequencies both by taking a single flow cytometry measurement and by plating the cultures on tetrazolium maltose (TM) plates and counting for CFUs (3 independent plates per culture). (**G**) We compared the S frequencies from the flow cytometer (f_{exp}) to the average S frequencies from CFUs (f_{cfu}) and see that S is consistently over-represented in the CFU measurements. (**H**) We hypothesized that there might be a multiplicative bias affecting frequencies. If this was true, then the true frequency would be $f_S = n_S/(n_S + n_L)$ where n_i is the count of ecotype i; the biased frequency would be $f_S^{bias} = bn_S/(bn_S + n_L)$ where b is the multiplicative bias. Thus, the odds ratio of the true and biased frequencies should be related linearly, $bf_S/(1 - f_S) = f_S^{bias}/(1 - f_S^{bias})$. We indeed see a linear relationship between the odds ratio of the CFU and flow cytometry frequencies, where $\hat{b} \approx 1.5$. If we trust the flow cytometry measurements more, at least because it more directly measures cell counts and it is unbiased with the LTEE ancestor REL606, then this would imply that measuring S/L frequencies with CFUs on TM plates incurs a significant multiplicative bias.

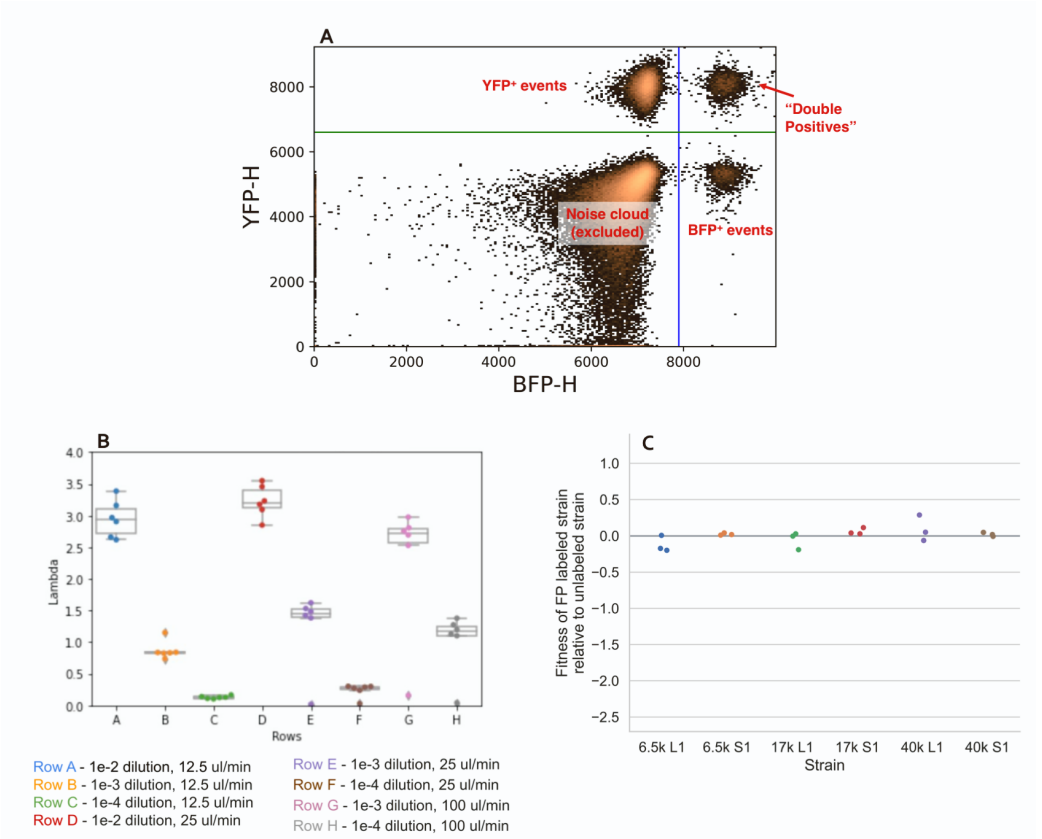

Figure S2: Flow cytometry and fluorescent protein labeling. Related to Figure 1. (**A**) Example of raw flow cytometry data and gates, shown as a 2D histogram of events. The lower left quadrant represents "noise" that is present even when running blank media, and is thus excluded from further analysis. The lower right and upper left quadrants represent events that were called as either BFP or YFP positive, respectively. The upper right quadrant is composed of events that are both BFP and YFP fluorescent; it appears that these events happen when both a BFP and YFP positive cell pass in front of the flow cytometry lasers and are counted as one event. This process appears to be well-described as a poisson process; when we account for this, we appear to recover the correct frequencies (Figure S1). The process by which we account for "double positive" events is detailed in the Methods section. (**B**) The effect of varying dilution rate and flow rate in the flow cytometer on the average number of cells that end up in front of the laser, i.e. the poisson mean λ. The center line represents the median of the data, the box represents the inter-quartile range (IQR), and the whiskers extend out to 1.5 times the IQR. (**C**) Fitness effects of fluorescent protein labels. We competed fluorescently labeled versions of S and L clones we used in our study against unlabeled versions, and measured the fitness effects via flow cytometry, with three biological replicates each. We do not see any consistent fitness effects of the fluorescent protein label across strains. Figure is plotted on same scale as Figure 1B-E, to facilitate comparison.

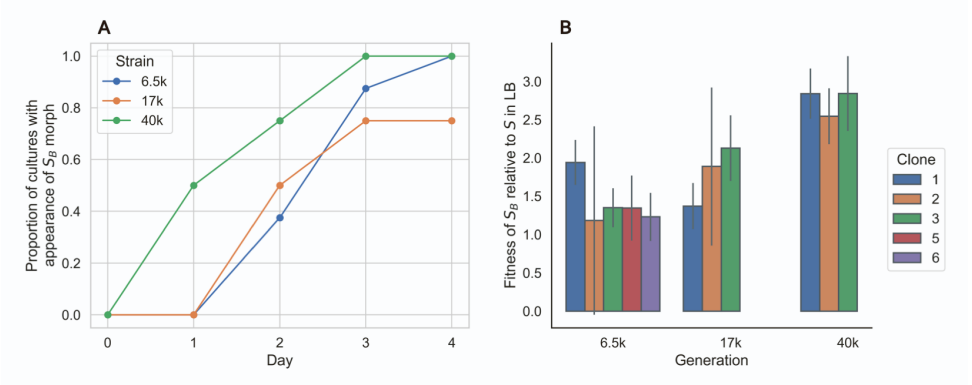

Figure S3: Rediversification of S into S_B in LB, and fitness in LB. Related to Figure 1. (**A**) At day 0, we started S clones (picked from DM2000 plates) in 8 replicate 1mL LB cultures, each for S clones from 6.5k, 17k, and 40k LTEE generations. Every day thereafter, we diluted the culture 1:100 into 1mL fresh LB, and we plated the cultures on TA plates (to a final density of around 100 colonies). We then recorded if we saw the appearance of the S_B morph in each of the cultures. We saw the appearance of S_B in most of the cultures across the S clones from the three LTEE timepoints. (**B**) Fitness of S_B clones in LB, relative to S. All S_B and S clones were propagated for one day in DM25, then mixed at approximately a 50-50 ratio, and transferred 1:100 to LB. Frequency of S_B was determined by counting colonies on TA plates. Fitness is computed as the change in logit frequency. Error bars represent standard errors.

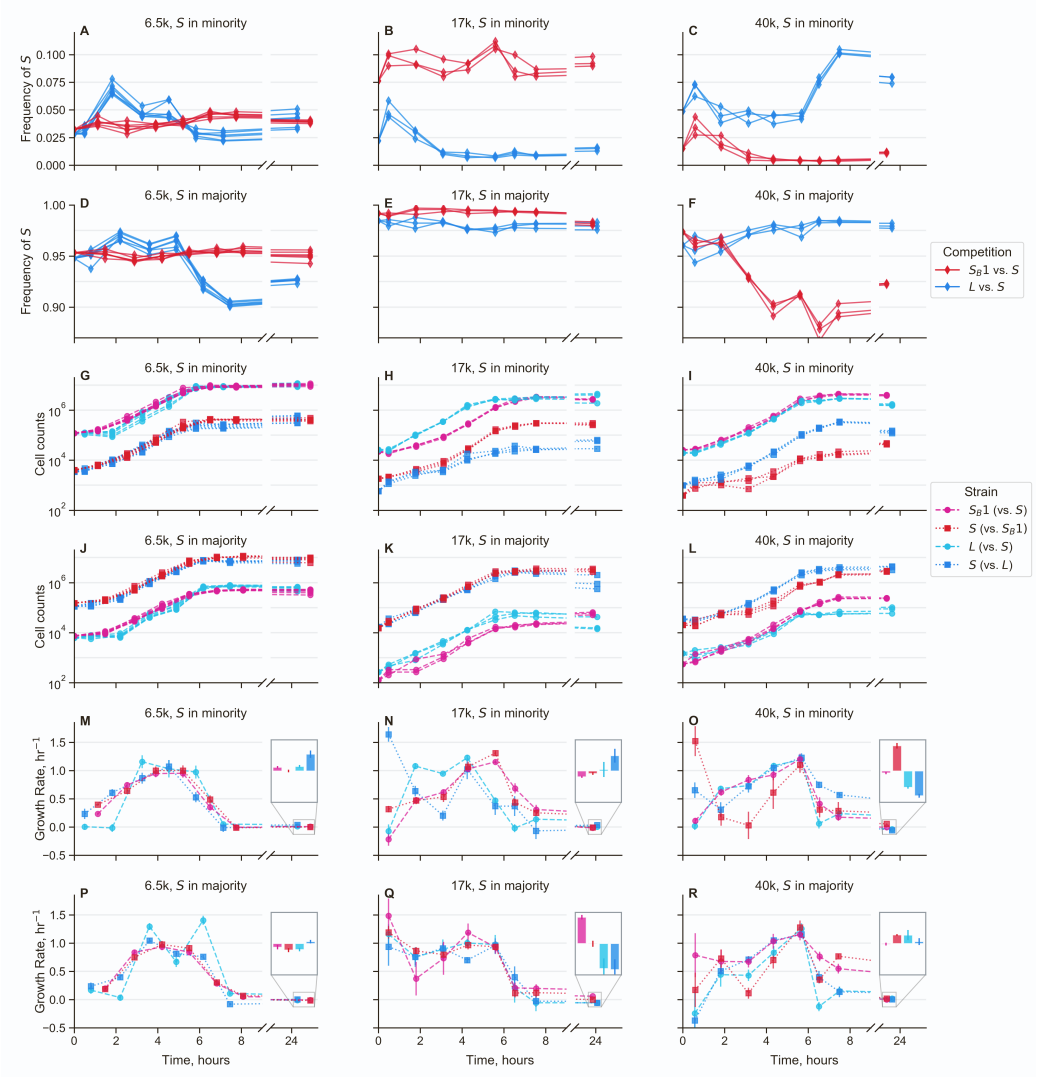

Figure S4: Growth dynamics of cocultures over the course of one twenty-four hour growth cycle. Related to Figure 2. Measurements were taken approximately every hour via flow cytometry for the first eight hours after transfer into new media. An additional measurement was taken approximately 24 hours after the start of the cycle. Mixed S_B with S along with L with S from the same LTEE generation, where ecotypes were mixed both in the majority and minority of the population. All competitions involved S_B used clone 1. Different lines represent biological replicates. Plots in left column represent experiments done with clones from 6.5k generations; 17k clones in the middle column; 40k clones in the right column. (**A-F**) Frequency dynamics of S against S_B and against L. (**G-L**) Total cell count dynamics, separated by each strain in the cocultures. (**M-R**) Growth rates over time for each strain in the cocultures, calculated as the log-slope between adjacent timepoints, using the second timepoint as the x-axis location. Insets represent growth rates in stationary phase, from around 8 to 24 hours. All insets use the same y-axis limits, set from -0.13 to 0.13 hr^{-1}. Error bars represent standard errors.

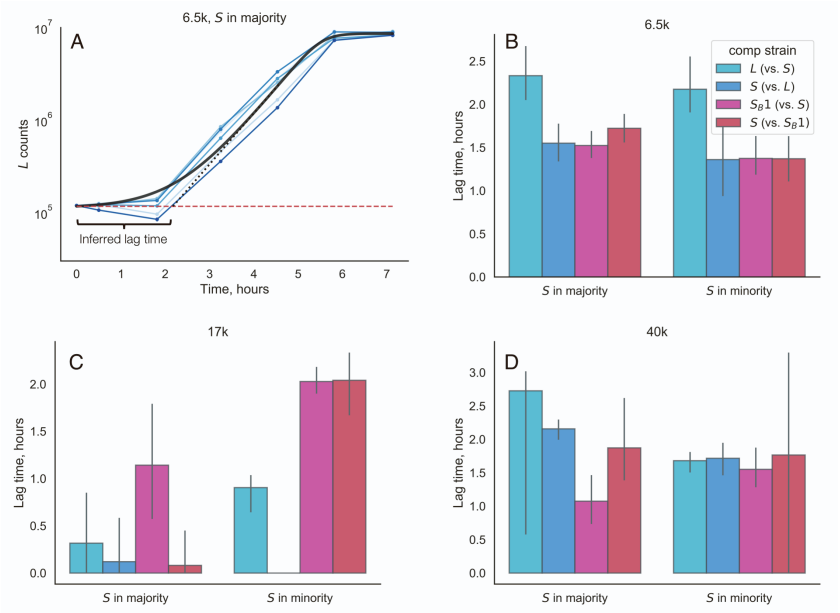

Figure S5: Fitting growth curves to extract lag time estimates. Related to Figure 2. (**A**) Example of procedure to extract lag time estimates. We used least-squares curve fitting to fit a generalized logistic curve (black line) to all the growth curves for a given strain and condition (blue lines). We then found the point of maximum growth rate (on a log scale), and extrapolated that rate back in time (black dotted line). In accordance with previous approaches,[S1] we designated the lag time as the time where the extrapolated line meets the initial abundance (red dashed line). We used this procedure to compute lag times for all conditions and strains from (**B**) 6.5k, (**C**) 17k, and (**D**) 40k generations. We computed error bars (95% CIs) via standard bootstrapping.

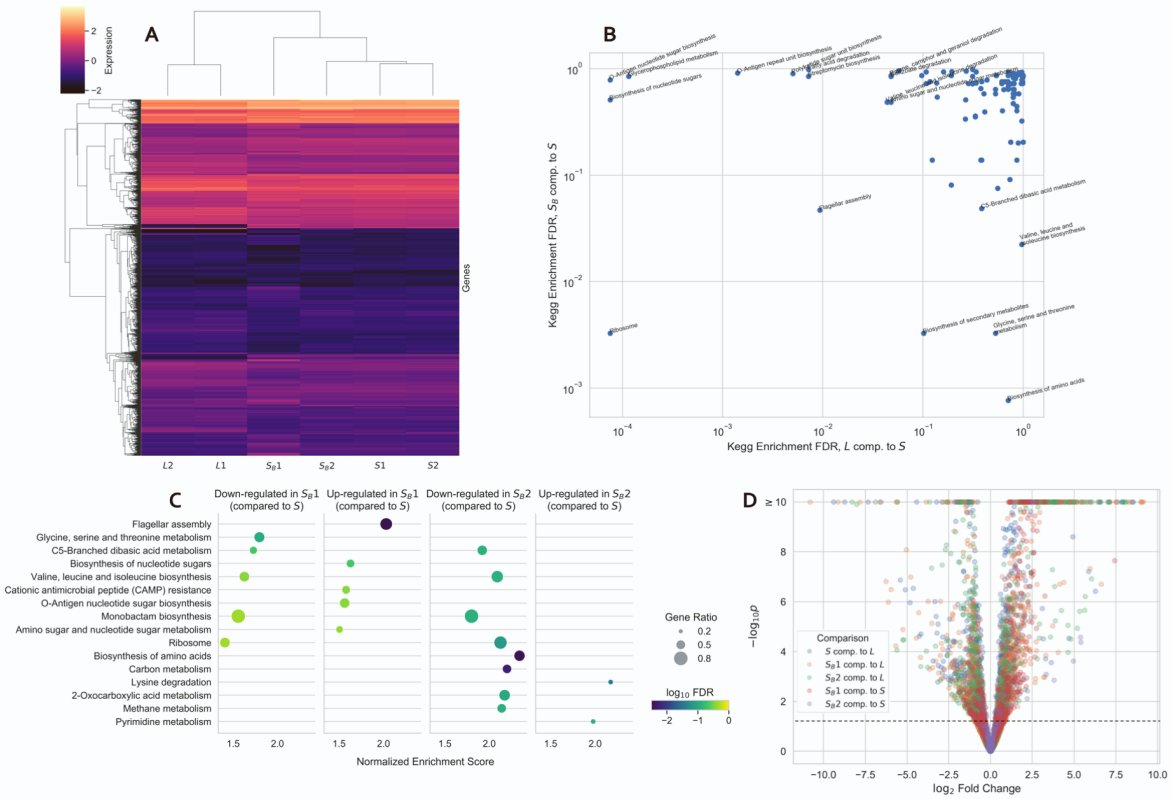

Figure S6: Further RNASeq Analysis. Related to Figure 5. (**A**) Heatmap and clustering of gene expression patterns. Expression of individual genes is clustered on the y-axis, expression of strains are clustered on the x-axis. We used the variance-stabilized gene expression values from DESeq2, then averaged the two biological replicates for each strain, and centered and scaled the values. (**B**) Comparison of differentially expressed KEGG pathways between S_B and L. Shown are the FDR-corrected p-values of each KEGG pathway, with the name shown of significantly enriched pathway in either condition. With the exception of the "Ribosome" and "Flagellar Assembly" pathways, the pathways differentially expressed in S_B and L are mostly orthogonal to each other. (**C**) KEGG gene set enrichment analysis for the two S_B clones, showing the top ten genes (sorted by FDR-corrected p-value) for each clone. (**D**) Volcano plot of differential expression, comparing different strains. The p-values on the y-axis are calculated after a Benjamini-Hochberg FDR correction. The black dashed line represents a p-value cutoff of 0.05.

Strain Name	Internal Name	Ancestor	Notes
REL606	REL606		ara- LTEE ancestor
6.5k S 1	eJA046	REL11555	subclone
6.5k S 2	eJA047	REL11555	subclone
6.5k S 3	eJA048	REL11555	subclone
6.5k L 1	eJA027	REL11556	subclone
6.5k L 2	eJA028	REL11556	subclone
6.5k S_B 1	eJA049	6.5k S 1	Rediversified in DM25
6.5k S_B 2	eJA036	6.5k S 1	Sister to 6.5k S_B 4*
6.5k S_B 3	eJA395	6.5k S 1	
6.5k S_B 4	eJA037	6.5k S 1	Sister to 6.5k S_B 2*
6.5k S_B 5	eJA034	6.5k S 2	
6.5k S_B 6	eJA035	6.5k S 3	Rediversified in DM25
17k S 1	eJA052	REL11557	
17k L 1	eJA031	REL11578	
17k S_B 1	eJA055	17k S 1	
17k S_B 2	eJA397	17k S 1	
17k S_B 3	eJA398	17k S 1	
40k S 1	eJA172	REL10927	Isolated from mixed population
40k L 1	eJA174	REL10927	Isolated from mixed population
40k S_B 1	eJA177	40k S 1	
40k S_B 2	eJA399	40k S 1	
40k S_B 3	eJA400	40k S 1	

Table S1: Table of strains. Related to STAR Methods. *6.5k S_B 2 and 4 were isolated from the same rediversification experiment culture–they come from two different colonies from the same timepoint from the same plate.

Oligo Name	Sequence	Purpose
ja35	GAAGAGGAT AAAACCGTGGA	Amplify arcA for genotyping
ja36	AGGTCAGGG ACTTTTGTAC	Amplify arcA for genotyping
ja40	TTCGAAGCG ACAGATGGC	Sanger sequence arcA for genotyping
ja37	CAGTTGTGACATA CAGCTAACGCT	Amplify aspS for genotyping
ja38	GCTCATGGGAGTT CACTCAGTTG	Amplify aspS for genotyping
ja42	ACTCTAACC ACGTCAACACC	Sanger sequence aspS for genotyping
ja200	ATTCGCTAA ACTGTgctagcatta tacctaggactgagct agctgtcaagctgtc cataaaaccgccc	Gibson assembly primer to make pJA17/18
ja201	atgctagcACAGTTT AGCGAATACGT CATAGAGCATTA AGGAGGTCATAG atggtgagcaagggcgag	Gibson assembly primer to make pJA17/18
ja183	tccaagctcagctaattacttg tacagctcgtccatgc	Gibson assembly primer to make pJA17
ja184	cgagctgtacaagtaattagct gagcttggactcc	Gibson assembly primer to make pJA17
ja190	cttcgccgatcaggatgcg	Downstream of glmS - amplify attTn7 junction with FP
ja191	tcgccctcgaacttcacctc	In pJA17/18 - amplify attTn7 junction with FP
ja170	caaaatcggttacggttgag	Downstream of glmS - sequence attTn7 junction with FP

Table S2: Table of oligonucleotides. Related to STAR Methods.

strain	position	mutation	annotation	gene
6.5k S_B 1	242,204	(G)9→10	pseudogene (194/373 nt)	ECB_00212 →
6.5k S_B 1	942,720	(C)7→8	intergenic (+464/+235)	clpA → / ← serW
6.5k S_B 1	1,292,775	A→G	intergenic (-48/-556)	hns ← / → tdk
6.5k S_B 1	1,368,422	(T)8→9	intergenic (+40/-172)	pspE → / → ycjM
6.5k S_B 1	1,444,298	C→T	R136R (CGC→CGT)	ydbC →
6.5k S_B 1	1,829,779	G→A	intergenic (+11/+82)	pncA → / ← ydjE
6.5k S_B 1	2,034,188	A→G	Y26Y (TAT→TAC)	manC ←
6.5k S_B 1	2,165,552	(C)5→6	pseudogene (306/1685 nt)	yehU ←
6.5k S_B 1	2,704,655	T→C	F41F (TTT→TTC)	ygaX →
6.5k S_B 1	2,771,495	T→C	T51A (ACC→GCC)	cysN ←
6.5k S_B 1	2,998,766	G→A	intergenic (+40/-66)	yqgA → / → pheV
6.5k S_B 1	3,937,360	A→G	D43G (GAC→GGC)	rffT →
6.5k S_B 2	752,223	C→T	intergenic (+723/-124)	ybgG → / → cydA
6.5k S_B 2	942,720	(C)7→8	intergenic (+464/+235)	clpA → / ← serW
6.5k S_B 2	1,186,311	G→A	L754F (CTT→TTT)	mfd ←
6.5k S_B 2	2,195,273	(C)9→8	coding (110/837 nt)	yeiG →
6.5k S_B 2	2,762,201	+T	coding (875/1365 nt)	ygbN →
6.5k S_B 4	752,223	C→T	intergenic (+723/-124)	ybgG → / → cydA
6.5k S_B 4	942,720	(C)7→8	intergenic (+464/+235)	clpA → / ← serW
6.5k S_B 4	1,186,311	G→A	L754F (CTT→TTT)	mfd ←
6.5k S_B 4	1,995,685	(C)8→9	intergenic (+103/+351)	yeeN → / ← asnW
6.5k S_B 4	2,004,021	-1258 bp	1258 bp deletion	
6.5k S_B 4	3,213,279	A→G	A20A (GCA→GCG)	yhaV →
6.5k S_B 4	4,598,479	T→C	intergenic (+28/+279)	nadR → / ← yjjK
6.5k S_B 5	242,204	(G)9→10	pseudogene (194/373 nt)	ECB_00212 →
6.5k S_B 5	1,422,705	INV	179,806 bp inversion	
6.5k S_B 5	2,567,537	T→C	L658L (CTA→CTG)	pbpC ←
6.5k S_B 5	4,606,996	C→T	A151V (GCC→GTC)	creC →
6.5k S_B 6	508,855	A→G	G32G (GGA→GGG)	glxR →
6.5k S_B 6	751,910	IS1 (–) +8 bp	intergenic (+410/-430)	ybgG → / → cydA
6.5k S_B 6	1,038,663	A→G	V141V (GTA→GTG)	yccR →
6.5k S_B 6	1,140,567	+771 bp	771 bp insertion	
6.5k S_B 6	1,422,704	INV	179,807 bp inversion	
6.5k S_B 6	1,184,045	A→G	R312R (CGT→CGC)	ycfS ←
6.5k S_B 6	1,480,910	(A)8→7	coding (786/1407 nt)	ydcR →
6.5k S_B 6	2,178,109	G→A	S140L (TCG→TTG)	yohF ←
6.5k S_B 6	3,726,210	T→C	intergenic (-70/+10)	cysE ← / ← gpsA
6.5k S_B 6	4,288,802	A→G	intergenic (-24/+35)	alsB ← / ← rpiR
17k S_B 1	794,722	(G)10→9	pseudogene (282/462 nt)	ECB_00735 →
17k S_B 1	1,292,373	IS1 (–) +9 bp	coding (346-354/414 nt)	hns ←
17k S_B 1	1,603,015	T→C	I90V (ATC→GTC)	ECB_01505 ←
17k S_B 1	1,680,075	+770 bp	770 bp insertion	
17k S_B 1	1,796,607	(C)10→9	intergenic (-53/+55)	celF ← / ← celD
17k S_B 1	2,032,376	G→A	T405T (ACC→ACT)	manB ←
17k S_B 1	2,104,764	(CCAG)21→20	pseudogene (238-241/272 nt)	ECB_01992 →
17k S_B 1	4,447,544	+34 bp	pseudogene (636/1388 nt)	treB ←

Table S3: Mutations in S_B clones relative to their ancestor. Related to Figure 5. All comparison tables generated with (1) `minimap2`[S2] and `sniffles`[S3] for structural variants, and (2) `breseq`[S4] for all other variants. Please note that the tables only document mutations of S_B clones relative to their immediate S ancestor. The various S clones also differ between one another, e.g. both 6.5k S 2 and 3 have (C)7→8 mutations at position 942,720. Illumina sequencing sometimes has issues with homopolymer tracts,[S5] so expansions and contractions of homopolymer tracts may not reflect the presence of true mutations.

Generation	Media	S_B clone or L	Mean fitness effect, s	FDR, $s \neq 0$	Significant
6.5k	Acetate	1	0.33941992	0.00123886	y
6.5k	Acetate	2	-0.0678909	0.16952222	
6.5k	Acetate	3	0.77471854	0.02828481	y
6.5k	Acetate	L	0.90360717	0.00047829	y
6.5k	CasAA	1	-0.2227881	0.00302507	y
6.5k	CasAA	2	-0.1521415	0.04057528	y
6.5k	CasAA	3	0.699906	0.06634847	
6.5k	CasAA	L	0.1856107	0.01344624	y
6.5k	Glycerol	1	0.39714395	0.00491058	y
6.5k	Glycerol	2	-0.3678152	0.02107772	y
6.5k	Glycerol	3	0.76054915	0.1582816	
6.5k	Glycerol	L	0.55867873	0.00047829	y
6.5k	Pyruvate	1	-0.2905477	0.00123886	y
6.5k	Pyruvate	2	-0.0361562	0.13061226	
6.5k	Pyruvate	3	1.35323686	0.01668868	y
6.5k	Pyruvate	L	1.81943367	0.00047829	y
17k	Acetate	1	-0.7378141	0.00700758	y
17k	Acetate	2	-0.1450888	0.34500569	
17k	Acetate	3	-0.0288927	0.44826275	
17k	Acetate	L	0.16207076	0.01668868	y
17k	CasAA	1	1.87108396	0.00307499	y
17k	CasAA	2	1.69034966	0.02196037	y
17k	CasAA	3	1.52119526	0.01668868	y
17k	CasAA	L	1.09350998	0.00075765	y
17k	Glycerol	1	0.5302828	0.03111336	y
17k	Glycerol	2	2.00514665	0.03854516	y
17k	Glycerol	3	1.87593382	0.03111336	y
17k	Glycerol	L	0.01028681	0.10462645	
17k	Pyruvate	1	-1.8367082	0.00066473	y
17k	Pyruvate	2	0.39755245	0.0746602	
17k	Pyruvate	3	0.97834684	0.03971706	y
17k	Pyruvate	L	0.78560192	0.00047079	y
40k	Acetate	1	1.64176166	0.02107772	y
40k	Acetate	2	-0.0352703	0.35777264	
40k	Acetate	3	-0.5421621	0.01119657	y
40k	Acetate	L	0.18952022	0.09300927	
40k	CasAA	1	-1.0206621	0.06088943	
40k	CasAA	2	-1.171821	0.0496925	y
40k	CasAA	3	-1.1075316	0.00491058	y
40k	CasAA	L	-0.8815546	0.00307499	y
40k	Glycerol	1	0.23601119	0.03111336	y
40k	Glycerol	2	0.15847985	0.32312438	
40k	Glycerol	3	0.13744615	0.35777264	
40k	Glycerol	L	1.07114232	0.02196037	y
40k	Pyruvate	1	0.39831846	0.10462645	
40k	Pyruvate	2	0.20900129	0.08263866	
40k	Pyruvate	3	-0.312443	0.04122104	y
40k	Pyruvate	L	1.12839991	0.02107772	y

Table S4: Growth traits in novel environments. Related to Figure 3.

Supplemental references

S1. Zwietering, M.H., Jongenburger, I., Rombouts, F.M., and van't Riet, K. (1990). Modeling of the Bacterial Growth Curve. Appl. Env. Microbiol. *56*, 1875–1881.

S2. Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinform. *34*, 3094–3100.

S3. Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., Von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. Nat. Meth. *15*, 461–468.

S4. Deatherage, D.E., and Barrick, J.E. (2014). Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Meth. Mol. Bio. *1151*, 165–188.

S5. Stoler, N., and Nekrutenko, A. (2021). Sequencing error profiles of Illumina sequencing instruments. NAR Genom. Bioinform. *3*, lqab019.