Title: **Bayesian analysis of longitudinal RB-TnSeq resolves the fitness seascape in fluctuating environments**

Authors: Carl J. Stone[a,b] and Megan G. Behringer[a,b,c,d]#

Affiliations:

[a]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37232

[b]Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN 37232

[c]Vanderbilt Institute for Infection, Immunology and Inflammation, Nashville, TN 37232

[d]Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN 37232

#Corresponding Author:

Megan G. Behringer

Megan.g.behringer@vanderbilt.edu

Vanderbilt University

Department of Biological Sciences

VU Station B, Box 35-1634

Nashville, TN 37235-1634

Key words: starvation, quantitative genetics, trade-offs, *Escherichia coli*

ABSTRACT

Temporally structured environments are ubiquitous in nature, but time-dependent fitness effects are difficult to measure and thus understudied. To resolve temporal fitness structure at genome scale, we developed a Bayesian multilevel framework for longitudinal randomly barcoded transposon sequencing (RB-TnSeq) that stabilizes noisy mutant abundance trajectories into time-resolved selection-rate estimates with interpretable uncertainty. Applying this approach to a feast–famine starvation regime in *Escherichia coli*, we uncovered distinct fitness trajectories underpinned by shared molecular strategies across growth-curve phase, consistent with shifting constraints and antagonistic pleiotropy. Fitness during initial growth strongly constrained cumulative success, such that later advantages under stress could not rescue mutants that were initially deleterious. We then compressed these dynamics with a one-dimensional Fisher's geometric "seascape" model that orders mutants along a latent axis aligning with generalist–specialist and growth–survival trade-offs, providing a compact quantitative description of genome-wide constraints in a fluctuating environment. Finally, our longitudinal estimates and inferred seascape are predictive of both the identity and timing of mutational targets in an extended evolution experiment under similar repeated feast–famine conditions, linking short-term competitive fitness effects, with their related trade-offs and constraints, to long-term adaptive outcomes.

INTRODUCTION

Understanding how genetic variation shapes organismal fitness in fluctuating environments is a central challenge in evolutionary biology (Abdul-Rahman et al. 2021; Yamamichi et al. 2023). In natural environments, adaptation rarely occurs under constant conditions. These dynamic conditions are exemplified by microbial populations in the gut, soil, aquatic ecosystems, and host-associated niches which encounter repeated shifts in nutrient availability, stressors, and ecological interactions (Dethlefsen and Relman 2011; Thaiss et al. 2015; Woodhouse et al. 2016; Kram et al. 2017; Smits et al. 2017; Meisner et al. 2021; Behringer et al. 2022; Greenwald and Wolfgang 2022). Under such dynamics,

variants that are beneficial in one condition may be neutral or deleterious in another, and their transient effects can determine the fate of lineages and the predictability of evolutionary trajectories (Nguyen et al. 2021; Behringer et al. 2024). Resolving these fine-scale, time-dependent contributions of individual mutations is therefore critical for linking genotypes to evolutionary outcomes.

High-throughput assays such as pooled CRISPR perturbation screens, barcoded lineage tracking, and transposon-insertion sequencing (TnSeq) have transformed the study of fitness landscapes by enabling genome-wide measurement of fitness effects for thousands of mutations in parallel (van Opijnen et al. 2009; van Opijnen and Camilli 2013; Shalem et al. 2014; Levy et al. 2015). These approaches have yielded genome-wide maps of genotype–phenotype relationships, quantified the role of clonal interference and epistasis, and illuminated the genetic basis of traits such as antibiotic resistance and host colonization (Goodman et al. 2009; van Opijnen et al. 2009; Gallagher et al. 2011; Levy et al. 2015; Venkataram et al. 2016; Li et al. 2023; Abreu et al. 2024; Couce et al. 2024). Yet, despite their power, these methods are largely restricted to static or endpoint measurements. The intrinsic noise in sequencing-based assays makes it especially difficult to resolve small or transient fitness effects with statistical confidence, leading to a methodological trade-off between genome-wide scale and temporal resolution (Chao et al. 2016). As a result, many fitness effects that may be central to adaptation in realistic, dynamic environments remain undetected.

Bayesian statistical methods provide a powerful framework to overcome these limitations. Bayesian inference can incorporate prior knowledge of neutral expectations and sequencing count distributions, producing posterior probability distributions that propagate uncertainty across time points. Historically, applying Bayesian inference at genome scale has been limited by computational difficulties due to the high-dimensional parameter space of high-throughput fitness assays; however, recent advances in statistical computing have lowered these barriers, making genome-wide Bayesian inference increasingly practical for noisy pooled assays (Zhang et al. 2022). As such, Bayesian approaches have been applied in similar high-throughput and longitudinal contexts, including assessing the fitness of novel variants in high-throughput lineage tracking assays and longitudinal analyses of metagenomic

data (Sweeny et al. 2023; Razo-Mejia et al. 2024; Kim et al. 2025). However, Bayesian longitudinal inference has not yet been applied to Tn-Seq to resolve genome-scale, time-dependent fitness effects within fluctuating environments.

Bacterial starvation provides an ideal model for studying fitness in a fluctuating environment. In *Escherichia coli*, the transitions from exponential growth through stationary phase, death phase, and long-term stationary phase (LTSP) encompass reproducible shifts in nutrient availability, pH, resource scavenging, and stress responses (Zambrano et al. 1993; Finkel and Kolter 1999; Zinser and Kolter 2000; Behringer et al. 2018; Behringer et al. 2022). These dynamic selection pressures across growth phases, coupled with the inherent noise in barcode-based measurements, have limited the ability to use pooled mutant libraries to resolve transient or subtle fitness effects. As such, investigation of the genes that contribute to fitness across the microbial growth curve has largely focused on large-effect global regulatory pathways—such as sigma factor regulons, the stringent response, and catabolite repression—but the combined effect of transient contributions of individual genes across the growth curve remain underexplored (Zambrano et al. 1993; Zinser and Kolter 2000; Gosset et al. 2004; Nair and Finkel 2004; Kram et al. 2020; Irving et al. 2021).

Here, we develop a Bayesian framework for longitudinal inference of mutant fitness from randomly barcoded TnSeq (RB-TnSeq) data. Applying this approach to an *E. coli* library sampled over a 10-day starvation regime, we obtain accurate, interpretable confidence intervals for thousands of selection-rate estimates. This allows us to detect subtle, condition-specific fitness effects that are obscured by standard endpoint analyses. We identify distinct temporal trajectories associated with stress-response regulation, biofilm modulation, and resource scavenging. Then, we compress these time-resolved fitness dynamics into a one-dimensional Fisher's geometric "seascape" model, providing a compact description of genome-wide trade-offs across growth, death, and LTSP. Finally, we show that our results parallel patterns observed in long-term evolution experiments under similar selective regimes (Kram et al. 2017; Behringer et al. 2018; Katz et al. 2021; Behringer et al. 2024). Together, these

findings demonstrate how rigorous statistical modeling can unlock fine-scale maps of gene–fitness relationships in fluctuating environments, advancing both functional genomics and evolutionary biology.

RESULTS

Bayesian modeling improves the accuracy and interpretability of longitudinal RB-TnSeq

To obtain time-resolved fitness estimates from longitudinal RB-TnSeq while retaining interpretable uncertainty, we developed a Bayesian framework that infers gene-by-interval selection rates from barcode trajectories and returns posterior distributions for each estimate. We validated this approach by benchmarking it against a widely-used *t*-test—based analysis method (BarSeq) on simulated longitudinal data (**Fig. S1**) (Wetmore et al. 2015). Simulations incorporated varying parameters for likely sources of experimental noise: count variability between different transposon insertions in the same gene, random measurement error (a single parameter for multiple possible stochastic processes like amplification bias, differences between replicates, etc.), and directional batch effects. We chose noise parameter ranges similar to the magnitude of within-gene, between-replicate, and between-batch variability we have observed in RB-TnSeq measurements (**Fig. S2**). Across all simulations, our model achieved accuracy comparable to BarSeq in estimating true selection rates (**Fig. S1A**; our model: median RMSE 0.41, 95% CI [0.39, 0.44]; BarSeq: median RMSE 0.40, 95% CI [0.38, 0.43]).

Along with similar accuracy to BarSeq, our model consistently exhibited higher coverage probability, defined as the fraction of credible intervals containing the true value (**Fig. S1B**; our model: median coverage 53%, 95% CI [49%, 57%]; BarSeq: median coverage 39%, 95% CI [33%, 42%]). This advantage was most pronounced under high within-gene variability and high measurement error, where coverage increased to 58% (95% CI [53%, 65%]) compared to 23% (95% CI [18%, 29%]) for BarSeq (**Fig. S1D**). Large batch effects reduced coverage probability in both methods. Overall, these simulations indicate that our Bayesian model preserves accuracy while substantially improving uncertainty quantification in longitudinal RB-TnSeq experiments, especially under the noisy conditions characteristic of sequencing-based assays like Tn-Seq.

Longitudinal fitness analysis resolves shifts in adaptive pressure during the entry into long-term stationary phase

After validating our model, we applied it to an RB-TnSeq library in *E. coli* to quantify genome-wide, gene-specific fitness effects across the extended microbial growth curve. For each interval between sampling points, the model estimates a posterior distribution for the selection rate of each gene relative to an empirically neutral gene set, defined as genes whose mutants have consistently near-zero selection rates across a large compendium of growth and stress conditions (**Table S1**) (Price et al. 2018). These interval-specific selection rate estimates (denoted for each interval $t \rightarrow t+1$ as $s_{t \rightarrow t+1}$) are inferred simultaneously across all three intervals which can then be combined as a time-weighted average to calculate the net selection rate over the full experiment (s_{net}; **Fig. 1A**). Because inference produces posterior distributions rather than point estimates, uncertainty can easily be propagated into downstream analyses.

We applied this framework to the *E. coli* transposon library grown for 10 days in LB medium. The barcoded transposon-mutant library was sampled immediately after inoculation (day 0) and again at days 1, 4, and 10, providing a time series that captured mutant abundance trajectories across the major physiological transitions of the extended growth curve: rapid growth, stationary phase, death phase, and lastly long-term stationary phase (Zambrano et al. 1993; Finkel 2006). Selection rates were estimated for each time interval ($s_{0 \rightarrow 1}$, $s_{1 \rightarrow 4}$, and $s_{4 \rightarrow 10}$) relative to the neutral gene set. We identified 822 genes that were annotated by Wetmore *et al.* as present in the RB-TnSeq library but were not detected in our samples (**Table S2**) (2015). These genes are essential or highly deleterious when disrupted and were likely purified from the library before we acquired it or lost during preparatory cultures (Wetmore et al. 2015). Functional enrichment analysis revealed that these genes are largely associated with ribosome biogenesis, aminoacyl-tRNA synthesis, lipid metabolism, DNA replication and repair, and one-carbon metabolism (**Fig. S3**).

Out of the 3668 assessed genes present in our library, we then identified 1067 genes whose disruption conferred confidently positive or negative fitness effects in at least one time interval (**Table 1**, **Fig. 1B**). Mutants were classified as non-neutral if their posterior probability distributions indicate at least a 90% probability that their selection rate is above or below zero. Fitness distributions varied considerably across time intervals (**Table 1**, **Fig. S4, Fig. S5**). Mutations were slightly biased toward deleterious effects during the interval 0→1 d, then were showed more beneficial effects from 1→4 d and more deleterious effects from 4→10 d. These shifts illustrate that selective pressures change dramatically as environmental conditions fluctuate over the course of starvation. We then averaged interval-specific selection rates to calculate s_{net}, the cumulative selection rate of each mutant over the 10-day experiment. Most gene disruptions (733 genes) were cumulatively deleterious, although a subset (361 genes) were cumulatively beneficial across the 10 days (**Table 1**, **Dataset S1, Fig. 1B**, bottom panel; Fisher-Freeman-Halton test: $p = 5.0 \times 10^{-6}$, s_{net} χ^2 goodness-of-fit with BH correction: $p = 5.4 \times 10^{-22}$). Using a stricter probability cutoff of greater than 97.5% reduced the number of confidently non-zero mutant fitness effects but did not qualitatively change any of the temporal or cumulative fitness trends (**Table S3**).

Gene clusters reveal distinct temporal fitness strategies

Next, we grouped non-neutral genes by their selection rate trajectories to identify temporal fitness trade-offs. Principal component analysis (PCA) of selection rates across the three intervals revealed that the first principal component captured 80% of the variance and aligned with the combined eigenvector of $s_{0\to1}$ and $s_{4\to10}$ while being strongly anticorrelated with $s_{1\to4}$, which spans death phase, during which culture cell counts decrease by ~10-100—fold (Pearson correlation of selection rates: $r_{0\to1:1\to4} = -0.90$, $p < 2.2\text{x}10^{-16}$; $r_{1\to4:4\to10} = -0.76$, $p < 2.2\text{x}10^{-16}$) (Finkel 2006). This anticorrelation between adjacent time intervals is also present in the raw selection rates that were inputs to our Bayesian model ($r_{0\to1:1\to4} = -0.41$, $p < 2.2\text{x}10^{-16}$; $r_{1\to4:4\to10} = -0.44$, $p < 2.2\text{x}10^{-16}$). The second principal component (20% of variance) primarily separated exponential-phase effects from LTSP effects, distinguishing mutants

who were more beneficial in $0 \rightarrow 1$d from those that were beneficial in $4 \rightarrow 10$d. We then grouped mutants into distinct fitness trajectories using k-means clustering on each mutant's three interval selection rates (k = 4, **Fig. S6**). Mutants within clusters represent distinct survival strategies across long-term culture (**Fig. 2A**). The resulting clusters had similar sizes ranging from 329 genes in cluster IV to 432 genes in cluster I (**Table S4**). Each cluster was organized by the sign of $s_{0\rightarrow1}$ and $s_{4\rightarrow10}$ to facilitate biological interpretation (**Fig. 2B-C**). Interestingly, only clusters I and IV, both of which included mutants with positive $s_{0\rightarrow1}$, contained genes with positive s_{net}. In contrast, although cluster II contained mutants with beneficial effects at day 10, these effects were not strong enough to overcome initial deleterious effects during the 0→1 d interval. This suggests that early fitness exerts a dominant influence on cumulative success over this culture length.

To investigate the biological functions associated with each fitness trajectory, we performed functional enrichment analysis using gene ontology terms, KEGG pathways, and EcoCyc pathway annotations (**Fig. 2C, Table S5**) (Ashburner et al. 2000; Kanehisa and Goto 2000; Kanehisa 2019; Kanehisa et al. 2025; Karp et al. 2025; The Gene Ontology Consortium 2026). Cluster I, which showed consistent benefits early and late in culture, was enriched for genes encoding efflux systems, particularly those associated with detoxification (**Fig. 2D**). Cluster II, which gained fitness late in culture, contained mutants related to the biosynthesis of enterobacterial common antigen, secretion of surface structural components such as pili and flagella, and multiple methyltransferases, including *dam* and several involved in rRNA and tRNA modification. These genes are important for growth, metabolism, and membrane homeostasis, making them deleterious to mutate early but dispensable, and possibly energetically favorable to lose, at later times.

Cluster III genes were consistently deleterious when disrupted, suggesting roles in critical processes. This group included genes involved in central metabolism, outer membrane integrity, nucleoid organization, and iron storage. In contrast, cluster IV mutants were initially beneficial but had neutral or negative $s_{4\rightarrow10}$ and were enriched for genes involved in conditional stress responses, such as

amino acid transport, two-component systems, and siderophore synthesis. The loss of these genes may confer short-term energetic advantages that become costly as the environment changes.

In addition to identifying functional enrichment across clusters, we examined the mutants with the largest fitness effects to identify major physiological strategies that shape survival and competition during prolonged culture. Efflux systems showed particularly strong and coherent patterns in s_{net} (**Fig. S7A**). Most individual efflux pumps were slightly beneficial when inactivated, whereas disruption of *mdtABC* was mildly deleterious, and disruption of *tolC*, required for all major efflux systems, was strongly deleterious (Zgurskaya and Nikaido 2000; Tikhonova et al. 2011). These patterns indicate that while pumps are partially redundant, a degree of efflux capacity remains essential during prolonged culture. The strongest positive s_{net} observed in this experiment was disruption of the repressor *cytR* (**Fig. S7B**) (Shin et al. 2001; Browning and Busby 2004). Consistent with *cytR* disruption being strongly beneficial, disruptions of genes in the *cytR* regulon were on average deleterious, and the regulon had more deleterious genes than the null expectation from size-matched random gene sets, where the expected number is calculated as the sum of each gene's probability of being deleterious (Pr[*cytR* regulon > null] = 0.93; **Fig. S8**). Together these data suggest that derepression of the *cytR* regulon may provide a net advantage during starvation.

Biofilm-associated mutants also exhibited strong and time-specific fitness effects. Type I fimbriae mutants showed concerted and contrasting selection effects: activator *fimB* was deleterious and inactivator *fimE* was beneficial across both early and late intervals, indicating strong selection on fimbrial switching dynamics (**Fig. S7C**) (Klemm 1986). Curli fiber genes exhibited transient fitness effects: structural and export gene mutants (*csgA*, *csgE*) were deleterious early, while folding and polymerization gene mutants (*csgC*, *csgF*) were beneficial in the same time interval, but all curli-associated mutants were neutral at the later interval (Chapman et al. 2002). Similarly, the phase-variable adhesin Ag43 (*flu*) mutant had strong early benefits but became neutral later (Henderson et al. 1997; Wallecha et al. 2002). Together, these strong pathway-level effects exemplify the major physiological strategies shaping growth and survival, including detoxification, regulatory adaptation, and modulation of surface

structures, and illustrate how time-resolved fitness measurements can identify discrete mechanistic contributors to long-term competition.

A geometric seascape model recapitulates the growth-survival trade-off

Despite these strong pathway-level effects, these examples do not explain a broader model for how selection acts on the population. Mutants in disparate pathways show coordinated gains and losses of fitness between intervals, suggesting that selection is acting on a shared physiological trade-off rather than on a specific molecular function. To summarize this structure, we fit a one-dimensional Fisher's geometric "seascape" model that assigns each transposon mutant genotype (g) a latent phenotypic value (z_g, a unitless coordinate) and estimates a separate single-peaked fitness curve for each time interval (d) (Fisher 1930; Tenaillon 2014). We call z a latent phenotypic axis because it represents a multidimensional phenotypic state that is theoretically measurable but practically must be inferred from fitness. In this model, each interval's landscape is defined by a fitness optimum at a fixed latent phenotypic value z and a steepness (or curvature) that quantifies how strongly fitness declines as mutants deviate from the optimum (**Fig. 3A**, full specification in **Supplemental Text 1**). We fit the seascape model jointly to all mutants and time intervals, incorporating uncertainty from the longitudinal model, estimating a single latent coordinate for each gene and resolving a distinct landscape for each phase of the extended growth curve.

The fitted seascape recapitulates the antagonistic selection pressures inferred from our interval-specific analyses (**Fig. 3B**). Selective pressure shifts in both magnitude and sign across our experimental intervals. The inferred fitness optima for early growth (0→1 d) and LTSP (4→10 d) lie on the far-negative end of the latent axis, whereas the death-phase (1→4 d) optimum is on the opposite, positive end. This suggests that there is a growth–survival trade-off between exponential phase and death phase that aligns LTSP more closely with growth than survival, consistent with repeated subpopulation turnover and growth on recycled resources (secreted metabolites, necromass, etc.) during LTSP (Finkel 2006). The interval-specific landscapes also vary in steepness, with selection being strongest early and

decreasing monotonically over subsequent intervals. Combining the interval-specific functions yields the integrated 0→10 d landscape, which exhibits a single optimum z^* (**Fig. 3C**). Mutants with z_g close to optimum z^* maximize cumulative growth and survival across the entire experiment. By contrast, mutants at extremely low or high z_g incur large costs during at least one stage of the extended growth curve. Altogether, this seascape allows us to order all transposon mutants in our library on the same latent axis, giving us a structured look into the interplay between genotypes, phenotypes, and selection in this shifting environment.

Using the latent axis from the seascape model, we investigated whether the z_g positions of regulators could be predicted from the z_g of mutants in their regulons. If a regulators latent coordinate primarily reflects the combined effects of its targets, then the regulator should lie near the signed sum of its targets on the z axis; in other words, the z_g of activators should equal the sum of their targets' z_g, and the z_g of repressors should equal the negative sum of their targets. This expectation holds in specific cases: for activator *ydcI*, z_{ydcI} is −2.11 [-3.32, -0.89] (median [95% CI]) and the sum of its regulon z_g is −1.29 [-3.53, 0.92]; for repressor *cytR*, z_{cytR} is -3.96 [-4.83, -3.10] while the sum of its regulon is 4.15 [1.45, 6.80] (**Fig. 3D**). This correlation is strongest for regulators with small regulons, consistent with a geometric model assumption that latent axis positions are additive; however, as regulon size increases, the relationship weakens and ultimately reverses, suggesting a heuristic boundary beyond which large, pleiotropic regulons reorganize physiology in ways that are not well-approximated by a one-dimensional model (**Fig. 3E**). These results highlight how network complexity can produce unexpected mappings between genetic perturbations and latent phenotypes.

To interpret the physiological meaning of the latent phenotypic axis, we compared seascape z_g coordinates to previously published gene-level fitness measurements across 131 conditions using the same RB-TnSeq library (**Table S6**) (Price et al. 2018). We first calculated the distance between fitness profiles for each condition then clustered conditions by similarity (**Fig. 4A**). Condition clusters split broadly into two groups: clusters 1, 2, and 3 comprised only nutritional assays where populations were

grown in minimal media with either ammonium and a varying carbon source (hereafter, “single-carbon” experiments) or glucose and a varying nitrogen source (“single-nitrogen” experiments); clusters 4, 5, and 6 included diverse chemical and physical stresses (including antibacterials, metals, solvents, general biocides, and inorganic toxins), along with a small number of single-carbon conditions that grouped with stresses, particularly L-lactate (**Fig. 4B**, **Table S7**). We then correlated gene-level selection rates with seascape coordinates for each condition using a generalized additive model (**Fig. 4C, Fig. S9**). Genes at strongly negative z_g were near-neutral across all conditions. However, selection rate diverged across condition clusters as z_g increased: high-z_g genes showed strongly positive selection rate in the most stressful conditions (cluster 6), but neutral to mildly deleterious selection rates in moderate stress (clusters 4 and 5) and strongly deleterious effects in growth on varying carbon or nitrogen sources (clusters 1, 2, and 3) (**Fig. 4D**). Thus, ordering genes along z reveals a continuum from environmentally robust, weak-effect mutants to mutants with a strong trade-off favoring stress resistance at the expense of metabolic flexibility.

Transposon mutant fitness correlates with positive selection in extended evolution experiments

To assess the ecological relevance of predicted fitness values from our Bayesian model, we compared our data to a previous experimental evolution study in which *E. coli* populations underwent repeated feast/famine cycles every 10 days for a total of 300 days (Behringer et al. 2022). Mutations of low frequency were discarded as spurious, and duplicate alleles resulting from mapping errors were manually removed. Alleles were then classified by mutation type and by the fitness of each Tn mutant in our library.

For each mutation type, a permutation test was conducted to establish a null expectation for the Tn fitness effect (**Fig. 5A**). Genes that showed high fitness upon disruption in our Tn screen exhibited elevated mutation acquisition rates across all mutation classes, including nonsynonymous, synonymous, nonsense, indel mutations, and even large insertions and deletions, though large indels had too few occurrences for reliable null distribution estimates. Genes identified as deleterious in our

Tn fitness assay exhibited nonsynonymous and nonsense mutation rates consistent with null expectations. However, both genes with beneficial and deleterious Tn mutant effects showed notably higher rates of indels than expected, with beneficial genes exhibiting the highest rates. Interestingly, synonymous mutations occurred at increased rates in beneficial genes and decreased rates in deleterious genes compared to expectations, although the magnitude of the reduction in deleterious genes was modest yet statistically significant (**Fig. 5A**).

We next examined how mutation timing related to Tn mutant fitness. We found that mutations in genes with extreme Tn fitness values (s_{net} < -0.15 or > 0.15) arose, on average, approximately 100 days earlier than mutations in genes with intermediate fitness values (**Fig. 5B**). This suggests that absolute Tn fitness effect, rather than directionality, correlates with mutation timing across all types of mutations. However, the seascape axis showed clear trends between gene z_g coordinates and the arrival time of some types of mutations (**Fig. 5C**). IS-element insertions arose substantially earlier in genes with negative z_g than in genes with positive z_g. In contrast, nonsense and nonsynonymous mutations showed the opposite pattern, with earlier arrival in genes at higher z_g. All other mutation classes showed no strong relationship between z_g and arrival time (**Fig. S10**). Together, these patterns indicate that the seascape captures temporal structure relevant to repeated adaptation in a similar growth regime, with the strongest predictive signal for the same type of alleles used to infer the seascape.

Further analysis concentrated on genes consistently targeted by mutations across multiple LTSP experimental evolution studies, including those with shorter feast/famine cycles (~7-10 days) and prolonged stationary phases (hundreds of days to years) (**Fig. 5D**) (Kram et al. 2017; Behringer et al. 2018; Katz et al. 2021; Ratib et al. 2021; Behringer et al. 2022; Behringer et al. 2024). These repeatedly mutated genes exhibited a broad range of Tn mutant fitness effects, including genes with the highest and lowest s_{net} in our library. Notably, specific temporal mutation dynamics emerged: *cytR* mutations, both nonsense and nonsynonymous, tended to sweep through populations within the first 90 days, followed by small indels that fixed around day 600 (**Fig. 5E, Fig. S11**). Contrasting mutation dynamics were observed in *dacA* and *ompC*, both highly deleterious in our Tn assay. *dacA* rapidly accumulated

IS elements, nonsense, nonsynonymous mutations, and small indels early in evolution. In contrast, *ompC* showed no fixation events but accumulated numerous nonsynonymous mutations and indels that fluctuated without fixation over time (**Fig. 5E**).

Collectively, these findings underscore that Tn mutant fitness effects can substantially differ from those of other mutation types, even presumed inactivating mutations. Nonetheless, Tn mutant fitness may serve as a meaningful indicator of adaptive targets, highlighting genes likely to accumulate mutations in highly selective environments. Moreover, mutation timing followed mutation-type-specific trends along the latent seascape axis, showing that this approach can predict adaptation in mutations similar to the seascape dataset but also captures associations in unrelated mutation types that can be investigated further. Although inactivating mutations correlated moderately with Tn fitness, the timing of mutation acquisition displayed a stronger association with the absolute magnitude of Tn fitness effects than the direction of the effect.

DISCUSSION

Here, we present a robust framework for analyzing barcoded high-throughput fitness assays in fluctuating environments. RB-TnSeq has been transformative for mapping gene-fitness relationships across diverse stresses, but technical limitations have typically limited analyses to endpoint readouts. By pairing longitudinal sampling with Bayesian inference, we have stabilized noisy fitness trajectories, obtained interpretable posterior credible intervals for time-dependent selection rates, and propagated uncertainty into downstream analyses. Applied to the extended *E. coli* growth curve, this approach resolves transient and cumulative fitness effects genome-wide, revealing structured trade-offs across the growth–death–LTSP transition (Finkel and Kolter 1999; Kram et al. 2020). Our results align late stationary-phase selective pressures with earlier growth phase and highlight loci that repeatedly shape evolutionary outcomes under feast/famine starvation regimes (Kram et al. 2017; Behringer et al. 2024). Our longitudinal model and the inferred fitness seascape thus link short-term competitive fitness effects, with their associated trade-offs and constraints, to long-term adaptive outcomes.

The extended growth curve provides a compelling model of environmental fluctuation because its major phases are well-characterized, reproducible, and biologically interpretable: processing through rapid growth, nutrient depletion with broad physiological remodeling, widespread death and release of complex cell debris, and finally prolonged competition during LTSP. Across this 10-day time series, we identified a large set of confidently non-neutral gene disruptions whose selection rates shift in both magnitude and direction between intervals, and the overall fitness distributions themselves change across phases. These distributional shifts underscore that the growth curve and starvation are not individual selective regimes but rather are temporally structured sequences of environments in which alleles can switch from beneficial to deleterious (or vice versa) as constraints change (Zambrano et al. 1993; Abreu et al. 2024). Additionally, the substantial set of missing library genes—enriched for essential processes like translation, replication, lipid metabolism, and one-carbon metabolism—highlights an orthogonal axis of selection: mutations that are strongly deleterious or lethal in these conditions may be purged before the focal experiment begins. This emphasizes the need to interpret fitness landscapes considering library composition, as we have done by calculating selection rates relative to empirically neutral genes (Wetmore et al. 2015; Price et al. 2018). Another key consequence of this temporal structure is that fitness early during growth phase can disproportionately shape cumulative outcomes over the full 10-day experiment, where early selection may act as a filter on which genetic strategies are available in later phases, consistent with previous studies in similar systems (Katz et al. 2021; Behringer et al. 2024; Patton et al. 2025).

Beyond identifying overall fitness effects, time-resolved inference also makes it possible to identify coherent temporal strategies. Clustering of interval-specific selection-rate trajectories revealed groups of mutants that share characteristic fitness patterns over the growth curve and are enriched for distinct functional modules. Several functional classes with strong fitness effects illustrate how temporally structured environments can produce similar fitness trajectories among mechanistically different mutants. Efflux-related loci highlight redundancy at the level of individual pumps contrasted with the broad importance of TolC as a shared outer-membrane channel (Zgurskaya and Nikaido 2000;

Tikhonova et al. 2011). Biofilm-associated systems contribute in strongly phase-specific ways, with adhesion and aggregation mechanisms diverging between early growth and LTSP (Danese 2000; Chapman et al. 2002; Woude 2008). Finally, regulatory perturbations such as disruption of *cytR* produce some of the strongest effects detected in this experiment; however, some of the factors with the largest regulons—like sigma factors *rpoD* and *rpoS*—were not present in our library. Collectively, these patterns suggest that extended culture selects for various physiological programs that can be reached through multiple genetic routes. These alternative strategies could be responsible for repeated emergence of distinct coexisting ecotypes within the same environment observed in long-term evolution experiments under similar starvation or fluctuation regimes (Finkel and Kolter 1999; Kram et al. 2017; Ratib et al. 2021; Behringer et al. 2022).

Enabled by accurate genome-wide fitness measurements, we further showed that diverse gene-level and pathway-level fitness trajectories can be organized by a low-dimensional latent structure. We formalized this by fitting a one-dimensional geometric "seascape" model in which each mutant is assigned a coordinate on a latent phenotypic axis, and each interval has its own quadratic fitness surface along that axis. The fitted seascape recapitulates the antagonistic selection pressures inferred from interval-specific analyses and previous empirical studies: high fitness during death phase (1→4 d) is inversely correlated with high fitness during early growth and, to a lesser extent, survival during LTSP (Maharjan and Ferenci 2013). The decreasing curvatures over time suggest that the growth/survival trade-off is strongest from exponential growth through death phase; future studies could increase temporal sampling resolution from the onset of death phase into LTSP to determine if the period from 4 d to 10 d has distinct selective pressures with their own trade-offs. Integrating interval-specific landscapes yields a unimodal cumulative fitness peak over 10 days of culture, illustrating how temporally structured environments introduce constraints as the milieu changes, which limits the accessible set of high-fitness strategies available. Importantly, the latent phenotypic axis coordinate provides a common ordering of mutants that links mechanistically disparate loci to a shared ecological interpretation; this

organizing axis is a resource that can be used as a predictor or covariate of other assays of knockout mutants.

We further interpreted this latent axis by mapping a large, published compendium of RB-TnSeq experiments onto the seascape. We have shown that genes at low negative phenotypic coordinates tend to have weak, near-neutral effects across growth conditions, whereas genes at high positive phenotypic coordinates increasingly diverge in fitness, producing a gradient from very high fitness in the most stressful conditions to very low fitness in conditions with single carbon or nitrogen sources, which require a diverse metabolic capacity to grow. This latent axis thus appears to be a continuum from environmentally robust but weak-effect alleles to antagonistically pleiotropic alleles that are specialized for stress at the expense of growth. This trade-off at high-phenotypic-coordinate alleles aligns with established growth-stress trade-off paradigms, particularly the trade-off between self-preservation and nutritional competence (SPANC) (Ferenci 2005; Phan and Ferenci 2013).

Finally, we have shown an association between our latent phenotypic axis and transposon fitness effects and the genetic targets of adaptation in an independent long-term evolution experiment under related starvation conditions. Across repeated feast/famine evolution, genes whose disruption had large-magnitude fitness effects—whether strongly beneficial or strongly deleterious—were disproportionately likely to acquire mutations in the evolution experiment, and mutations in these genes arose earlier in the experiment than those in genes with smaller effects. Moreover, transposon insertions in genes toward the negative end of the latent axis (aligned with both the optima for early growth and LTSP and with the integrated 10-day optimum) arose earlier in the long-term experiment than in genes at the positive end of the axis. These results underscore the utility of our approach in predicting evolutionary outcomes: transposon mutants with large effects, whether positive or negative, possibly signal genes with the potential to produce any large effect via mutation and thus shift genotypes into previously inaccessible regions of phenotypic space where new adaptive paths become available; however, our seascape model takes into account the relative magnitudes of temporal trade-offs and is therefore more narrowly predictive of the fitness benefits of transposon insertions under this particular

433 selective regime. Thus, this dual experimental–modeling approach indicates that transposon mutant
434 fitness can serve as a meaningful indicator of loci that repeatedly respond to and reshape the fitness
435 landscape under strong, structured selection.

436 Together, this work highlights the utility of longitudinal RB-TnSeq, paired with principled
437 Bayesian inference, as a general platform for mapping selective dynamics in temporally structured
438 environments. By quantifying time-dependent fitness effects, compressing complex gene-by-time
439 patterns into an interpretable latent axis, and testing how regulatory structure and evolutionary
440 outcomes interact with that axis, we provide a framework for connecting mechanistic gene functions to
441 systems-level constraints and long-term adaptation. While our study focused on starvation in *E. coli*,
442 this same strategy is broadly extensible to other time scales, stress sequences, and community
443 contexts, enabling a broader empirical and theoretical synthesis of how fluctuating environments shape
444 fitness landscapes and evolutionary predictability.

445

446 **METHODS**

447 **Strain and culture conditions**

448 All experiments were conducted using Escherichia coli K-12 strains. The wild-type strain was *E.*
449 *coli* K-12 substr. MG1655 (Strain PFM2 from Lee *et al.* (2012)). Transposon mutant fitness was measured
450 using the randomly barcoded transposon library KEIO_ML9, which was previously constructed by
451 random insertion of barcoded Tn5 transposons into BW25113, the parent strain of the Keio deletion
452 collection (Baba et al. 2006; Wetmore et al. 2015). The original transposon library contained 3728 unique
453 transposon insertions with a median number of 16 unique insertions per gene (Wetmore et al. 2015).
454 Unless otherwise specified, strains were cultured in LB-Miller medium (10 g/L tryptone, 5 g/L yeast
455 extract, 10 g/L NaCl) at 37 °C with orbital shaking at 175 rpm. Cultures were grown in 16 x 100 mm
456 glass culture tubes positioned upright in the shaking incubator.

457

458

459 **Longitudinal RB-TnSeq competition assay**

460 To quantify fitness dynamics across the extended growth curve, we performed longitudinal RB-
461 TnSeq assays spanning 10 days of culture. Frozen aliquots of the KEIO_ML9 library were thawed on ice
462 and inoculated into 50 mL LB with kanamycin (30 ug/mL) in a 250-mL flask, then cultured to mid-
463 exponential phase ($OD_{600} \approx 0.5$; V-1200 Spectrophotometer, VWR International, Radnor, PA, USA). In
464 parallel, wild-type MG1655 cultures were grown under the same conditions without antibiotics, starting
465 from an overnight culture diluted to $OD_{600} = 0.05$.

466 Both library and wild-type cultures were washed twice in PBS and normalized to $OD_{600} = 1.0$.
467 Co-cultures were initiated by mixing 100 μL of KEIO_ML9 with 100 μL of wild-type cells into 10 mL of
468 fresh LB medium in a culture tube. Parallel monocultures of KEIO_ML9 alone were also established
469 under identical conditions. Three 1-mL aliquots of the KEIO_ML9 library were pelleted and frozen for
470 time 0 samples.

471 Because upright test tube cultures develop spatially structured microenvironments that are
472 disrupted by sampling, we employed a pseudo-longitudinal design (Behringer et al. 2018). Independent
473 replicate cultures were established for each sampling time point using equal volume inocula from the
474 same starting cultures. For each time point, 3 biological replicate cultures were grown each for the
475 KEIO_ML9 library alone and the KEIO_ML9 library mixed with wild-type, resulting in 6 biological
476 replicates per time point that were ultimately pooled after bioinformatic processing. Samples were
477 collected immediately after inoculation (day 0) and after 1, 4, and 10 days of incubation. For day 0, three
478 1-mL aliquots of the KEIO_ML9 library were collected prior to incubation. At each subsequent time
479 point, 1 mL samples were harvested, pelleted by centrifugation, and stored at -80 °C for until DNA
480 extraction.

481

482 **DNA isolation, barcode amplification, and sequencing**

483 DNA extraction, barcode amplification, and DNA sequencing were performed similarly to
484 previously published uses of the RB-TnSeq library (Wetmore et al. 2015; Price et al. 2018). Genomic

DNA was isolated from frozen pellets using the Omega Bio-tek E.Z.N.A Bacterial DNA Kit (Product No. D3350) following the manufacturer's protocol, including the optional mechanical lysis with glass beads (Omega Bio-tek, Norcross, GA, USA). Transposon barcodes were then amplified using the BarSeq98 protocol as previously described (Wetmore et al. 2015). Each sample was amplified using a sample-specific indexed forward primer (ITXX) and a common reverse primer (P1). PCR amplification followed the BarSeq98 protocol, with an initial denaturation at 98 °C for 4 min, followed by 25 cycles at 98 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s, with a final extension step of 72 °C for 5 min. Samples were sequenced together on an Illumina NovaSeq6000 instrument at the Vanderbilt VANTAGE genomics core facilities using paired-end 150-bp chemistry; only read 1 was used for downstream analysis. Before barcode counting, reads were trimmed using Trimmomatic v. 0.39 single-end mode to remove Illumina adapter sequences and crop reads to 75 base pairs to reduce file sizes (Bolger et al. 2014).

Barcode processing and normalization

Barcode identification, initial filtering, and counting were performed using previously published scripts from the RB-TnSeq pipeline (Wetmore et al. 2015). Downstream analyses were performed using custom R scripts developed for this study.

Barcodes were first identified from reads and counted using script MultiCodes.pl, after which barcode locations were mapped to Tn insertion locations using the pool definition file and combineBarSeq.pl (Wetmore et al. 2015). We then used the "poolcount" file output from combineBarSeq.pl in downstream analysis. Barcodes were filtered to remove any barcode with fewer than three total counts across all day 0 samples. To mitigate positional sequencing biases, read counts were corrected by subtracting a sliding median calculated across 251-barcode windows. Barcode counts were then normalized by log-ratio transformation relative to an empirically defined neutral gene set. Log-ratio transformation was done using R package ALDEx2, using the clr() function with neutral genes input as the "denom" option (**Supplemental Text 1**) (Fernandes et al. 2013; Gloor et al. 2017). Neural genes were identified based on a large published RB-TnSeq compendium as genes whose

disruption exhibited minimal fitness variation across diverse environmental conditions (**Table S1**) (Price et al. 2018).

Calculation and Bayesian inference of selection rates

Selection rates were calculated from normalized barcode abundances as the change in log-ratio abundance between adjacent time points divided by the elapsed time in days. Selection rates were calculated separately for each barcode across three intervals: day 0→1 ($s_{0\to1}$), day 1→4 ($s_{1\to4}$), day 4→10 ($s_{4\to10}$). This sliding-window approach enables time-resolved inference of mutant fitness across distinct physiological phases of the extended growth curve.

Selection rates were then modeled using Bayesian multilevel piecewise linear regression. Each gene was treated as a group-level effect, with separate slopes estimated for days 1-4 and 4-10, and intercepts reflecting the day 1 selection rate. The model structure incorporated an LKJ prior on the correlation between the intercept and slopes and regularizing Student-t priors on all hyperparameters to stabilize inference given the relatively small number of barcodes per gene (Extended methods in **Supplemental Text 1**). Population-level priors were parameterized based on reanalysis of published RB-TnSeq data (Price et al. 2018).

Posterior distributions were first approximated using the Pathfinder variational inference algorithm implemented in CmdStanR, accessed through the R package brms (Hoffman and Gelman 2011; Bürkner 2017; Zhang et al. 2022; Gabry et al. 2025). This approximation allowed efficient initialization of full Hamiltonian Monte Carlo (HMC) sampling, greatly reducing computation time (Neal 2011). Final HMC sampling was performed with draws saved for downstream analysis. Model convergence and effective sample sizes were assessed to ensure accurate posterior estimation. This hierarchical Bayesian approach provided shrinkage of noisy estimates toward neutrality while preserving strong fitness effects.

Summary statistics and functional analysis

Posterior probability distributions of the expected values were generated for each gene's selection rates. From these estimates, cumulative log relative abundance was calculated as the integral of the step function of selection rate over the full 10-day culture period (**Supplemental Text 1**). For each gene, we summarized fitness effects using both point estimates (posterior medians) and probability thresholds (e.g., probability > 90% that fitness deviates from neutrality). Genes were clustered based on selection rate trajectories using principal component analysis and k-means clustering (k = 4). Function enrichment analyses (GO, KEGG, EcoCyc pathway annotations) were performed within each cluster to identify biologically coherent groups (Ashburner et al. 2000; Kanehisa and Goto 2000; Kanehisa et al. 2025; Karp et al. 2025; The Gene Ontology Consortium 2026). Functional enrichment was done in R using package clusterProfiler and the latest Bioconductor release of the *E. coli* K-12 genome annotation (v. 3.22.0) (Yu et al. 2012; Wu et al. 2021; Carlson 2024).

Benchmarking simulation

Precision and accuracy of predicted fitness values was measured by analyzing simulated genes with noise from multiple sources. Genes were simulated with latent knockout selection rate effects. As in our biological system, the selection rate and abundance of each gene was simulated forward through three time points, with selection rates of each gene correlating across time. Noise was introduced in three ways: selection rates of each barcode were sampled from a normal distribution around the latent gene selection rate (called intragenic fitness variation); noise was added to each observation to simulate measurement error with no directional bias (called measurement error); and a directional bias was added to all observations for each replicate sample simulated (called batch effect). Data were simulated with every combination of these three noise parameters generated from normal distributions with standard deviations of 0, 0.1, or 1 with units of $days^{-1}$, as in selection rate. For each noise combination, selection rate over time was estimated using a statistical inference pipeline designed for RB-TnSeq data (BarSeq, rewritten in R from equations in Wetmore et al. 2015) or with our longitudinal inference model. For each

condition, 1000 genes were simulated and root mean square error (RMSE) and coverage probability, measured as the proportion of true selection rates within inferred 90% confidence intervals, were calculated, with nonparametric bootstrapping used to estimate confidence around these two statistics. To better parse the effects of various sources of noise, RMSE and coverage were then modeled as a correlated multivariate generalized linear model with main effects and interactions of the three noise parameters as predictors, hierarchically grouped by inference method. The effects of each inference method on RMSE and coverage were then estimated conditioned on the levels of each noise parameter using conditional_effects() in brms.

Geometric seascape model of time-dependent fitness

To summarize structured changes in mutant library fitness across the three measured intervals, we fit a one-dimensional Fisher's geometric seascape model to the interval-specific gene selection-rate estimates from our longitudinal RB-TnSeq model. Briefly, the seascape assigns each gene a latent coordinate z_g on a one-dimensional phenotypic axis and estimates an interval-specific quadratic fitness surface characterized by an optimum (θ_d) and curvature (κ_d):

$$s_{g,d} = \alpha_d - \frac{1}{2}\kappa_d\left(z_g - \theta_d\right)^2,$$

where α_l is the maximum selection rate at the optimum. To encode temporal dependence without enforcing a linear or directional trend, interval optima were regularized with a discretized Ornstein-Uhlenbeck process linking adjacent intervals (Uhlenbeck and Ornstein 1930; Hansen 1997). The model was fit in Stan, using posterior summaries from the longitudinal model as inputs (carrying forward uncertainty), and posterior draws were used to compute gene coordinates, interval-specific landscapes, and the integrated 10-day landscape. Full likelihood, priors, and parameterization are provided in Supplemental Text 1.

This model was fit jointly across all genes and intervals using Bayesian inference in Stan using R package *brms* (Bürkner 2017; Gabry et al. 2025; Stan Development Team 2025). Observed gene-level selection rate posterior estimates from the longitudinal RB-TnSeq model were treated as noisy

observations of the latent seascape-predicted selection rates. Weakly informative priors were placed on all parameters to regularize inference while allowing flexibility in landscape shape and temporal dynamics.

Cumulative fitness across the full 10-day experiment was calculated by summing interval-specific fitness functions weighted by interval duration, producing an integrated seascape with a single cumulative optimum. All downstream analyses of latent phenotypic coordinates used full posterior distributions to propagate uncertainty.

Data and code availability

Sequencing data for RB-TnSeq experiments can be downloaded from SRA, BioProject PRJNA1422734. Code, count tables, and files required to reproduce this analysis are found at https://github.com/BehringerLab/Longitudinal_RBTnSeq_Paper.

Acknowledgments

We would like to thank J.B.M. for providing the KEIO_ML9 RB-TnSeq library. We would like to thank Vanderbilt Technologies for Advanced Genomics (VANTAGE) for DNA sequencing and high-performance computing resources provided by the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University. This work was supported by Army Research Office Grant W911NF-21-1-0161 (M.G.B.) and National Institutes of Health Grant R35GM150625 (M.G.B.).

Author Contributions

C.J.S. and M.G.B.: conceptualization, investigation, data visualization, writing, and editing. C.J.S.: statistical analysis, including development of the longitudinal model and seascape model. M.G.B.: funding acquisition, resources, supervision.

REFERENCES

Abdul-Rahman F, Tranchina D, Gresham D. 2021. Fluctuating Environments Maintain Genetic Diversity through Neutral Fitness Effects and Balancing Selection. *Molecular Biology and Evolution* 38:4362–4375.

Abreu CI, Mathur S, Petrov DA. 2024. Environmental memory alters the fitness effects of adaptive mutations in fluctuating environments. *Nat Ecol Evol* 8:1760–1775.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet* 25:25–29.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* 2:2006.0008.

Behringer MG, Choi BI, Miller SF, Doak TG, Karty JA, Guo W, Lynch M. 2018. Escherichia coli cultures maintain stable subpopulation structure during long-term evolution. *Proceedings of the National Academy of Sciences* 115:E4642–E4650.

Behringer MG, Ho W-C, Meraz JC, Miller SF, Boyer GF, Stone CJ, Andersen M, Lynch M. 2022. Complex ecotype dynamics evolve in response to fluctuating resources. *Mbio* 13:e03467-21.

Behringer MG, Ho W-C, Miller SF, Worthan SB, Cen Z, Stikeleather R, Lynch M. 2024. Trade-offs, trade-ups, and high mutational parallelism underlie microbial adaptation during extreme cycles of feast and famine. *Current Biology* 34:1403-1413.e5.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Browning DF, Busby SJW. 2004. The regulation of bacterial transcription initiation. *Nat Rev Microbiol* 2:57–65.

Bürkner P-C. 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80:1–28.

Carlson M. 2024. org.EcK12.eg.db: Genome wide annotation for E coli strain K12.

Chao MC, Abel S, Davis BM, Waldor MK. 2016. The design and analysis of transposon insertion sequencing experiments. *Nature Reviews Microbiology* 14:119–128.

Chapman MR, Robinson LS, Pinkner JS, Roth R, Heuser J, Hammar M, Normark S, Hultgren SJ. 2002. Role of Escherichia coli curli operons in directing amyloid fiber formation. *Science* 295:851–855.

Couce A, Limdi A, Magnan M, Owen SV, Herren CM, Lenski RE, Tenaillon O, Baym M. 2024. Changing fitness effects of mutations through long-term bacterial evolution. *Science* 383:eadd1417.

649 Danese. 2000. The outer membrane protein, antigen 43, mediates cell-to-cell interactions within
650 Escherichia coli biofilms. *Mol. Microbiol*.

651 Dethlefsen L, Relman DA. 2011. Incomplete recovery and individualized responses of the human
652 distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl. Acad. Sci. U.S.A.*
653 108:4554–4561.

654 Ferenci T. 2005. Maintaining a healthy SPANC balance through regulatory and mutational
655 adaptation: Stress resistance and metabolic capability. *Molecular Microbiology* 57:1–8.

656 Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. 2013. ANOVA-Like Differential Expression
657 (ALDEx) Analysis for Mixed Population RNA-Seq. *PLOS ONE* 8:e67019.

658 Finkel SE. 2006. Long-term survival during stationary phase: evolution and the GASP phenotype.
659 *Nat Rev Microbiol* 4:113–120.

660 Finkel SE, Kolter R. 1999. Evolution of microbial diversity during prolonged starvation. *Proceedings*
661 *of the National Academy of Sciences USA* 96:4023–4027.

662 Fisher RA. 1930. The genetical theory of natural selection. Oxford: The Clarendon Press

663 Gabry J, Češnovar R, Johnson A, Bronder S. 2025. cmdstanr: R Interface to "CmdStan." Available
664 from: https://mc-stan.org/cmdstanr/

665 Gallagher LA, Shendure J, Manoil C. 2011. Genome-scale identification of resistance functions in
666 *Pseudomonas aeruginosa* using Tn-seq. *mBio* 2:e00315-10.

667 Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. 2017. Microbiome Datasets Are
668 Compositional: And This Is Not Optional. *Front Microbiol* 8:2224.

669 Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI. 2009.
670 Identifying genetic determinants needed to establish a human gut symbiont in its habitat.
671 *Cell Host & Microbe* 6:279–289.

672 Gosset G, Zhang Z, Nayyar S, Cuevas WA, Saier MH. 2004. Transcriptome Analysis of Crp-
673 Dependent Catabolite Control of Gene Expression in Escherichia coli. *J Bacteriol*
674 186:3516–3524.

675 Greenwald MA, Wolfgang MC. 2022. The changing landscape of the cystic fibrosis lung
676 environment: From the perspective of Pseudomonas aeruginosa. *Current Opinion in*
677 *Pharmacology* 65:102262.

678 Hansen TF. 1997. STABILIZING SELECTION AND THE COMPARATIVE ANALYSIS OF ADAPTATION.
679 *Evolution* 51:1341–1351.

680 Henderson IR, Meehan M, Owen P. 1997. A novel regulatory mechanism for a novel phase-variable
681 outer membrane protein of Escherichia coli. *Adv Exp Med Biol* 412:349–355.

682 Hoffman MD, Gelman A. 2011. The No-U-Turn Sampler: Adaptively Setting Path Lengths in
683 Hamiltonian Monte Carlo. Available from: http://arxiv.org/abs/1111.4246

684 Irving SE, Choudhury NR, Corrigan RM. 2021. The stringent response and physiological roles of
685 (pp)pGpp in bacteria. *Nat Rev Microbiol* 19:256–271.

686 Kanehisa M. 2019. Toward understanding the origin and evolution of cellular organisms. *Protein*
687 *Sci* 28:1947–1951.

688 Kanehisa M, Furumichi M, Sato Y, Matsuura Y, Ishiguro-Watanabe M. 2025. KEGG: biological
689 systems database as a model of the real world. *Nucleic Acids Res* 53:D672–D677.

690 Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*
691 28:27–30.

692 Karp PD, Paley S, Caspi R, Kothari A, Krummenacker M, Midford PE, Moore LR, Subhraveti P,
693 Gama-Castro S, Tierrafria VH, et al. 2025. The EcoCyc database (2025). *EcoSal Plus*
694 13:eesp-0019-2024.

695 Katz S, Avrani S, Yavneh M, Hilau S, Gross J, Hershberg R. 2021. Dynamics of Adaptation During
696 Three Years of Evolution Under Long-Term Stationary Phase. *Molecular Biology and*
697 *Evolution* 38:2778–2790.

698 Kim Y, Worby CJ, Acharya S, van Dijk LR, Alfonsetti D, Gromko Z, Azimzadeh PN, Dodson KW,
699 Gerber GK, Hultgren SJ, et al. 2025. Longitudinal profiling of low-abundance strains in
700 microbiomes with ChronoStrain. *Nature Microbiology* 10:1184–1197.

701 Klemm P. 1986. Two regulatory fim genes, fimB and fimE, control the phase variation of type 1
702 fimbriae in Escherichia coli. *EMBO J* 5:1389–1393.

703 Kram KE, Geiger C, Ismail WM, Lee H, Tang H, Foster PL, Finkel SE. 2017. Adaptation of
704 Escherichia coli to Long-Term Serial Passage in Complex Medium: Evidence of Parallel
705 Evolution. *mSystems* 2:e00192-16.

706 Kram KE, Henderson AL, Finkel SE. 2020. Escherichia coli Has a Unique Transcriptional Program in
707 Long-Term Stationary Phase Allowing Identification of Genes Important for Survival.
708 *mSystems* 5:e00364-20.

709 Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous
710 mutations in the bacterium Escherichia coli as determined by whole-genome sequencing.
711 *Proceedings of the National Academy of Sciences* 109:E2774–E2783.

712 Levy SF, Blundell JR, Venkataram S, Fisher DS, Sherlock G, Petrov DA. 2015. Quantitative
713 evolutionary dynamics using high-resolution lineage tracking. *Nature* 519:181–186.

714 Li F, Tarkington J, Sherlock G. 2023. Fit-Seq2.0: An Improved Software for High-Throughput Fitness
715 Measurements Using Pooled Competition Assays. *J Mol Evol* 91:334–344.

716 Maharjan RP, Ferenci T. 2013. EPISTATIC INTERACTIONS DETERMINE THE MUTATIONAL
717 PATHWAYS AND COEXISTENCE OF LINEAGES IN CLONAL *ESCHERICHIA COLI*
718 POPULATIONS: BRIEF COMMUNICATION. *EVOLUTION* 67:2762–2768.

719 Meisner A, Snoek BL, Nesme J, Dent E, Jacquiod S, Classen AT, Priemé A. 2021. Soil microbial
720 legacies differ following drying-rewetting and freezing-thawing cycles. *The ISME Journal*
721 15:1207–1221.

722 Nair S, Finkel SE. 2004. Dps Protects Cells against Multiple Stresses during Stationary Phase.
723 *Journal of Bacteriology* 186:4192–4198.

724 Neal RM. 2011. MCMC using Hamiltonian dynamics. Available from:
725 http://arxiv.org/abs/1206.1901

726 Nguyen J, Lara-Gutiérrez J, Stocker R. 2021. Environmental fluctuations and their effects on
727 microbial communities, populations and individuals. *FEMS Microbiology Reviews*
728 45:fuaa068.

729 van Opijnen T, Bodi KL, Camilli A. 2009. Tn-seq: high-throughput parallel sequencing for fitness
730 and genetic interaction studies in microorganisms. *Nature Methods* 6:767–772.

731 van Opijnen T, Camilli A. 2013. Transposon insertion sequencing: a new tool for systems-level
732 analysis of microorganisms. *Nature Reviews Microbiology* 11:435–442.

733 Patton GE, Meraz JC, Yin M, Worthan SB, Williams S, Behringer MG. 2025. Historical contingency
734 limits adaptive diversification in a spatially structured environment. *Evolution*
735 *Letters*:qraf048.

736 Phan K, Ferenci T. 2013. A design-constraint trade-off underpins the diversity in ecologically
737 important traits in species *Escherichia coli*. *The ISME Journal* 7:2034–2043.

738 Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, Kuehl JV, Melnyk RA, Lamson JS, Suh
739 Y, et al. 2018. Mutant phenotypes for thousands of bacterial genes of unknown function.
740 *Nature* 557:503–509.

741 Ratib NR, Seidl F, Ehrenreich IM, Finkel SE. 2021. Evolution in Long-Term Stationary-Phase Batch
742 Culture: Emergence of Divergent Escherichia coli Lineages over 1,200 Days. *mBio*
743 12:e03337-20.

744 Razo-Mejia M, Mani M, Petrov D. 2024. Bayesian inference of relative fitness on high-throughput
745 pooled competition assays. *PLoS Computational Biology* 20:e1011937.

746 Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE,
747 Doench JG, et al. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells.
748 *Science* 343:84–87.

749 Shin M, Kang S, Hyun SJ, Fujita N, Ishihama A, Valentin-Hansen P, Choy HE. 2001. Repression of
750 deoP2 in Escherichia coli by CytR: conversion of a transcription activator into a repressor.
751 *EMBO J* 20:5392–5399.

752 Smits SA, Leach J, Sonnenburg ED, Gonzalez CG, Lichtman JS, Reid G, Knight R, Manjurano A,
753 Changalucha J, Elias JE, et al. 2017. Seasonal cycling in the gut microbiome of the Hadza
754 hunter-gatherers of Tanzania. *Science* 357:802–806.

755 Stan Development Team. 2025. Stan Reference Manual. Available from: https://mc-stan.org/

756 Sweeny AR, Lemon H, Ibrahim A, Watt KA, Wilson K, Childs DZ, Nussey DH, Free A, McNally L.
757 2023. A mixed-model approach for estimating drivers of microbiota community
758 composition and differential taxonomic abundance.Hird SM, editor. *mSystems*:e00040-23.

759 Tenaillon O. 2014. The Utility of Fisher's Geometric Model in Evolutionary Genetics. *Annu Rev Ecol*
760 *Evol Syst* 45:179–201.

761 Thaiss CA, Zeevi D, Levy M, Segal E, Elinav E. 2015. A day in the life of the meta-organism: diurnal
762 rhythms of the intestinal microbiome and its host. *Gut Microbes* 6:137–142.

763 The Gene Ontology Consortium. 2026. The Gene Ontology knowledgebase in 2026. *Nucleic Acids*
764 *Res* 54:D1779–D1792.

765 Tikhonova EB, Yamada Y, Zgurskaya HI. 2011. Sequential mechanism of assembly of multidrug
766 efflux pump AcrAB-TolC. *Chem Biol* 18:454–463.

767 Uhlenbeck GE, Ornstein LS. 1930. On the Theory of the Brownian Motion. *Phys. Rev.* 36:823–841.

768 Venkataram S, Dunn B, Li Y, Agarwala A, Chang J, Ebel ER, Geiler-Samerotte K, Hérissant L,
769 Blundell JR, Levy SF, et al. 2016. Development of a Comprehensive Genotype-to-Fitness
770 Map of Adaptation-Driving Mutations in Yeast. *Cell* 166:1585-1596.e22.

771 Wallecha A, Munster V, Correnti J, Chan T, van der Woude M. 2002. Dam- and OxyR-Dependent
772 Phase Variation of agn43: Essential Elements and Evidence for a New Role of DNA
773 Methylation. *Journal of Bacteriology* 184:3338–3347.

774 Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G,
775 Arkin AP, et al. 2015. Rapid Quantification of Mutant Fitness in Diverse Bacteria by
776 Sequencing Randomly Bar-Coded Transposons. *mBio* 6:e00306-15.

777 Woodhouse JN, Kinsela AS, Collins RN, Bowling LC, Honeyman GL, Holliday JK, Neilan BA. 2016.
778 Microbial communities reflect temporal changes in cyanobacterial composition in a
779 shallow ephemeral freshwater lake. *The ISME Journal* 10:1337–1351.

780 Woude van der. 2008. Regulation and function of Ag43 (flu). *Ann. Rev. Microbiol.*

781 Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. 2021. clusterProfiler
782 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* 2:100141.

783 Yamamichi M, Gibbs T, Levine JM. 2023. Eco-evolutionary maintenance of diversity in fluctuating
784 environments. *Ecology Letters* 26:S152–S167.

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* 16:284–287.

Zambrano MM, Siegele DA, Almirón M, Tormo A, Kolter R. 1993. Microbial Competition: Escherichia coli Mutants That Take Over Stationary Phase Cultures. *Science* 259:1757–1760.

Zgurskaya HI, Nikaido H. 2000. Multidrug resistance mechanisms: drug efflux across two membranes. *Mol Microbiol* 37:219–225.

Zhang L, Carpenter B, Gelman A, Vehtari A. 2022. Pathfinder: Parallel quasi-Newton variational inference. *Journal of Machine Learning Research* 23:1–49.

Zinser ER, Kolter R. 2000. Prolonged Stationary-Phase Incubation Selects for *lrp* Mutations in *Escherichia coli* K-12. *Journal of Bacteriology* 182:4361–4365.

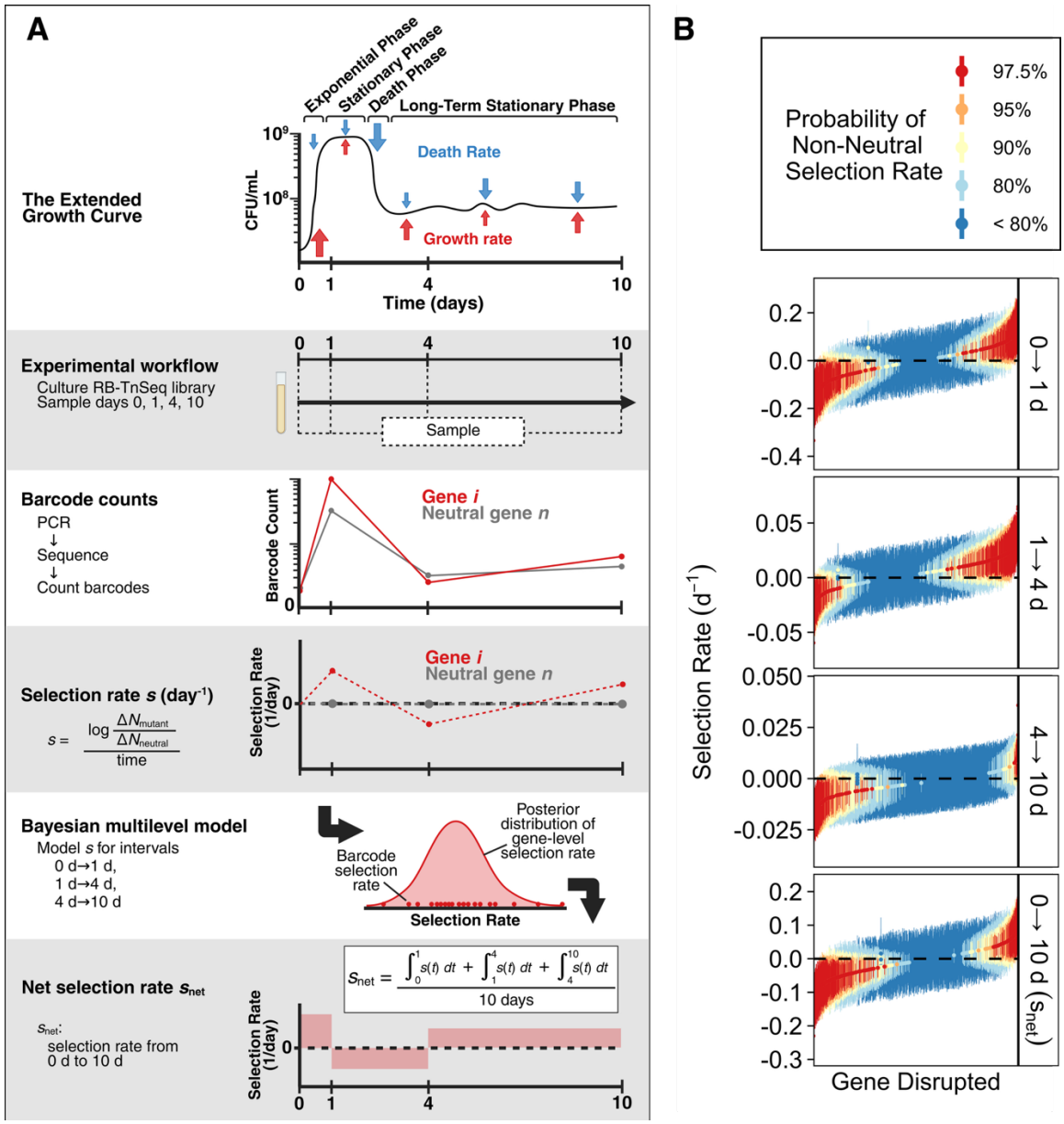


Figure 1. Inference of transposon mutant selection rates over extended culture. (A) Schematic of the experimental workflow and longitudinal Bayesian model used to infer gene-level selection rates across sequential time intervals. **(B)** Posterior summaries of gene-level selection rates for each interval (0→1 d, 1→4 d, 4→10 d) and for the average effect across the full experiment (s_{net}). Each point and vertical line shows the posterior median and 95% credible interval for one gene. Colors show the posterior probability that the selection rate is positive or negative (i.e. Pr(s > 0) or Pr(s < 0)). Genes are ranked within each panel by median selection rate, and the gene order is independent across panels.

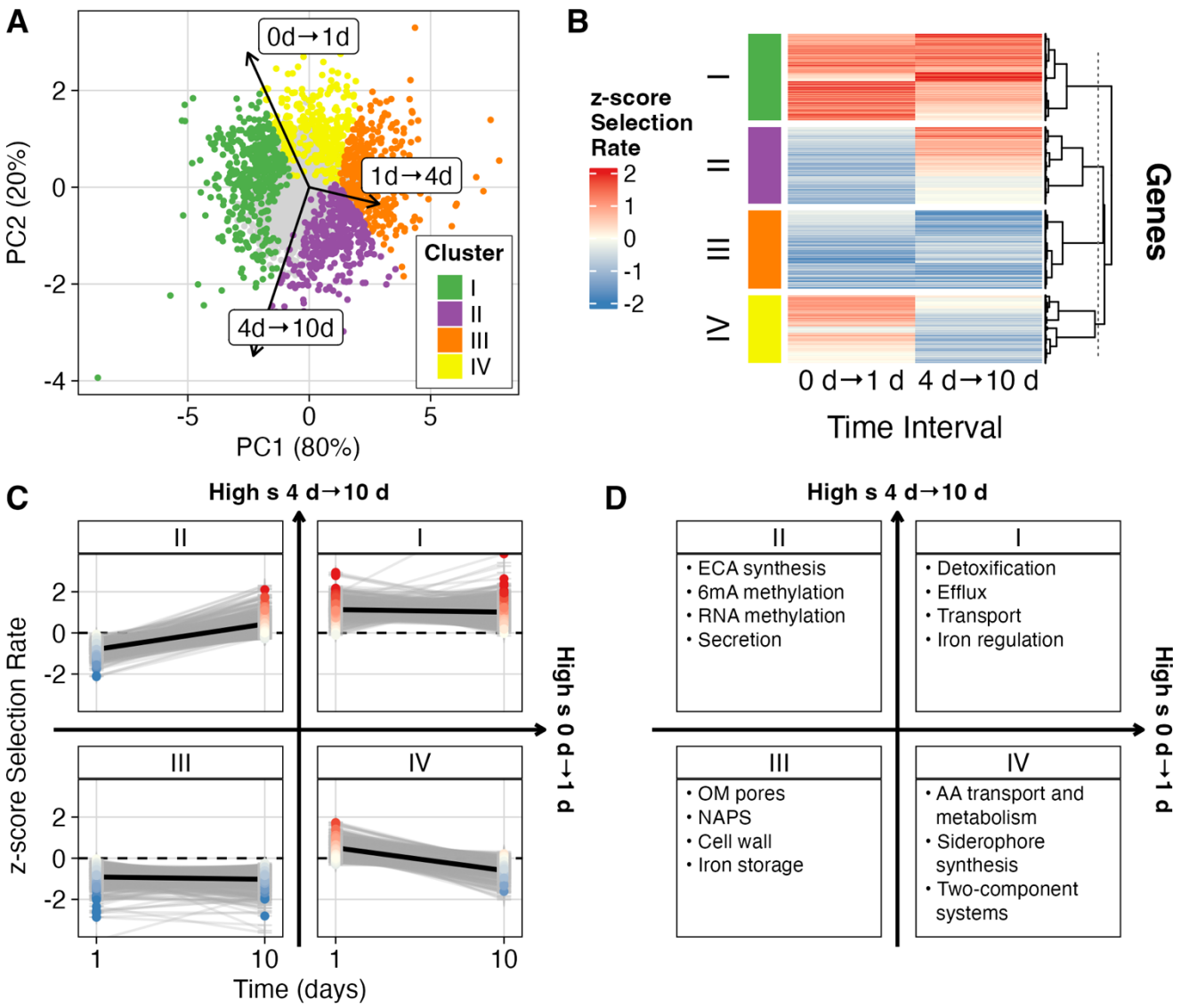


Figure 2. Tn mutants cluster into distinct temporal selective trajectories. (A) Principal component analysis of gene-level selection rate estimates across time intervals. Arrows indicate the loading of each interval on the two PCs. Each point represents one gene. Genes with non-zero selection rates in intervals 0→1 d or 4→10 d are colored, and genes with no confidently non-zero estimates are gray. Non-neutral genes were clustered by cosine distance, with cluster assignment shown by color. **(B)** Heatmap of z-scored selection rate estimates of early versus late selection (0→1 d and 4→10 d). All values outside [-2, 2] are clipped to those boundaries for visualization. Genes are clustered by selection rate similarity. **(C)** Cluster-specific temporal trajectories. Each gray line connects a single gene's scaled selection rate estimates across intervals. Panel labels indicate the cluster identity. Black lines show a robust linear fit

817 summarizing the cluster trend. Arrows between panels form a coordinate plane indicating the relative
818 order of clusters by selection rate at two intervals: horizontal axis = low 0→1 d to high 0→1 d; vertical
819 axis = low 4→10 d to high 4→10 d. **(D)** Functional categories overrepresented among genes in each
820 cluster based on annotations from Gene Ontology, KEGG, and EcoCyc pathways (FDR $q < 0.1$). As in
821 C, panel labels indicate the cluster identity, and arrows indicate the relative order of clusters by selection
822 rate at early and late intervals.
823

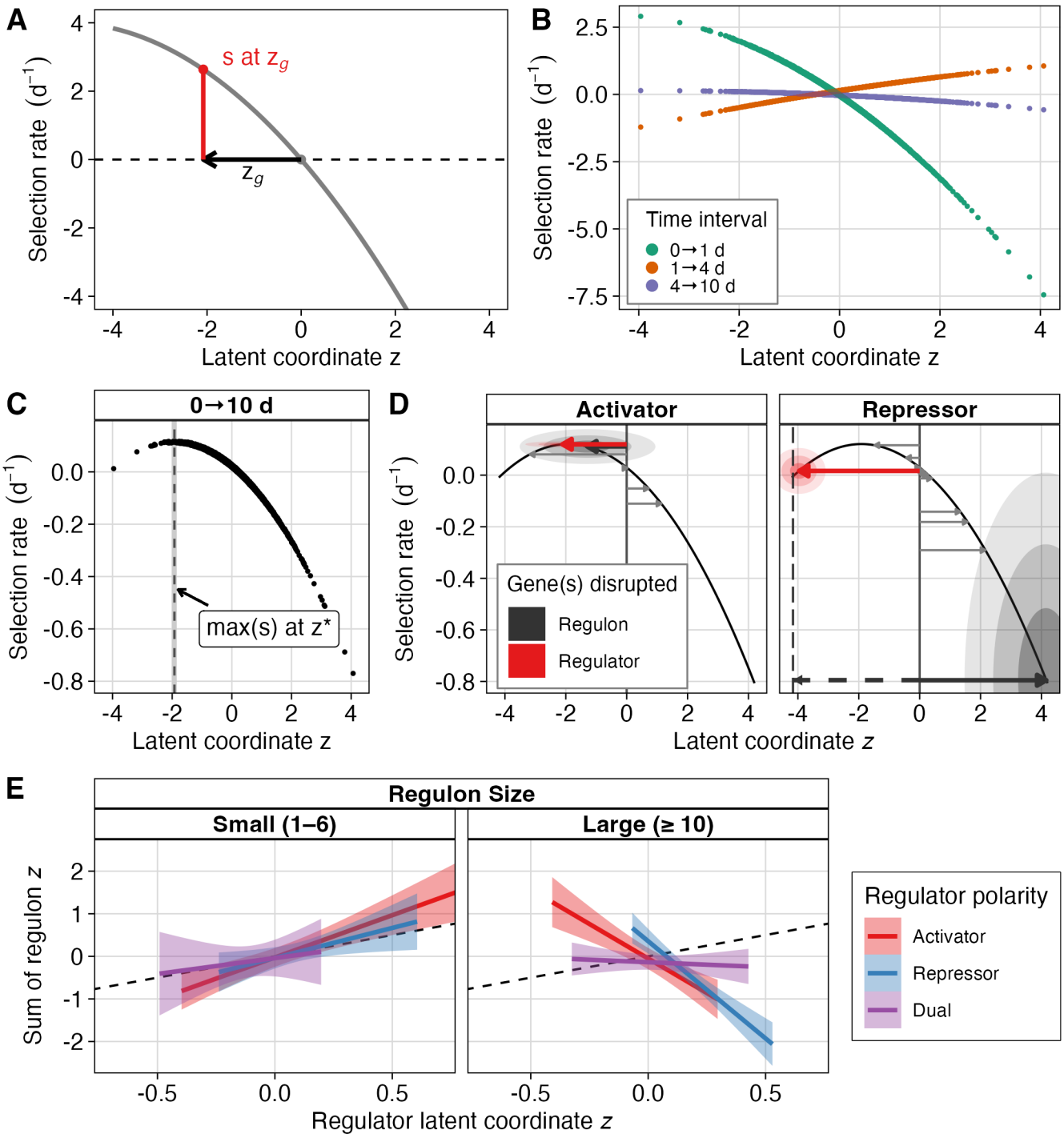


Figure 3. A geometric seascape model orders genes along a latent growth-survival axis. (A) Schematic of the one-dimensional Fisher's geometric seascape. Each gene *g* is assigned a latent phenotypic coordinate z_g, and the expected selection rate in each interval is modeled as a quadratic function of distance from the interval-specific optimum. **(B)** Interval-specific seascapes fitted to gene-level selection-rate posteriors from the longitudinal RB-TnSeq model. Points show the posterior median

of gene-level selection rate estimate as predicted by the seascape model, with colors indicating the time interval. **(C)** Time-weighted integrated seascape across 0-10 d, computed as the time-weighted average of interval-specific seascapes. The integrated optimum z^* indicates the phenotypic coordinate that maximizes cumulative expected selection rate. **(D)** Phenotypic coordinates of activator *ydcI* mutant, repressor *cytR* mutant, and their regulon mutants. Red arrows show the z_g vector of regulator mutants of *ydcI* (left panel) and *cytR* (right panel), with concentric shaded ellipses showing credible intervals (center: 50%, middle: 80%, outside: 95%) for both z and estimated selection rate. Gray arrows show the z_g vector for each mutant in the regulon. Black arrows show the signed sum of regulon mutants, with shaded ellipses showing credible intervals (50%, 80%, 95%). The dashed black arrow and vertical dashed line indicate the negative signed sum of the *cytR* regulon mutants. **(E)** The signed sum of regulon mutants correlates with regulator phenotypes in small regulons, but correlation decreases with regulon size.

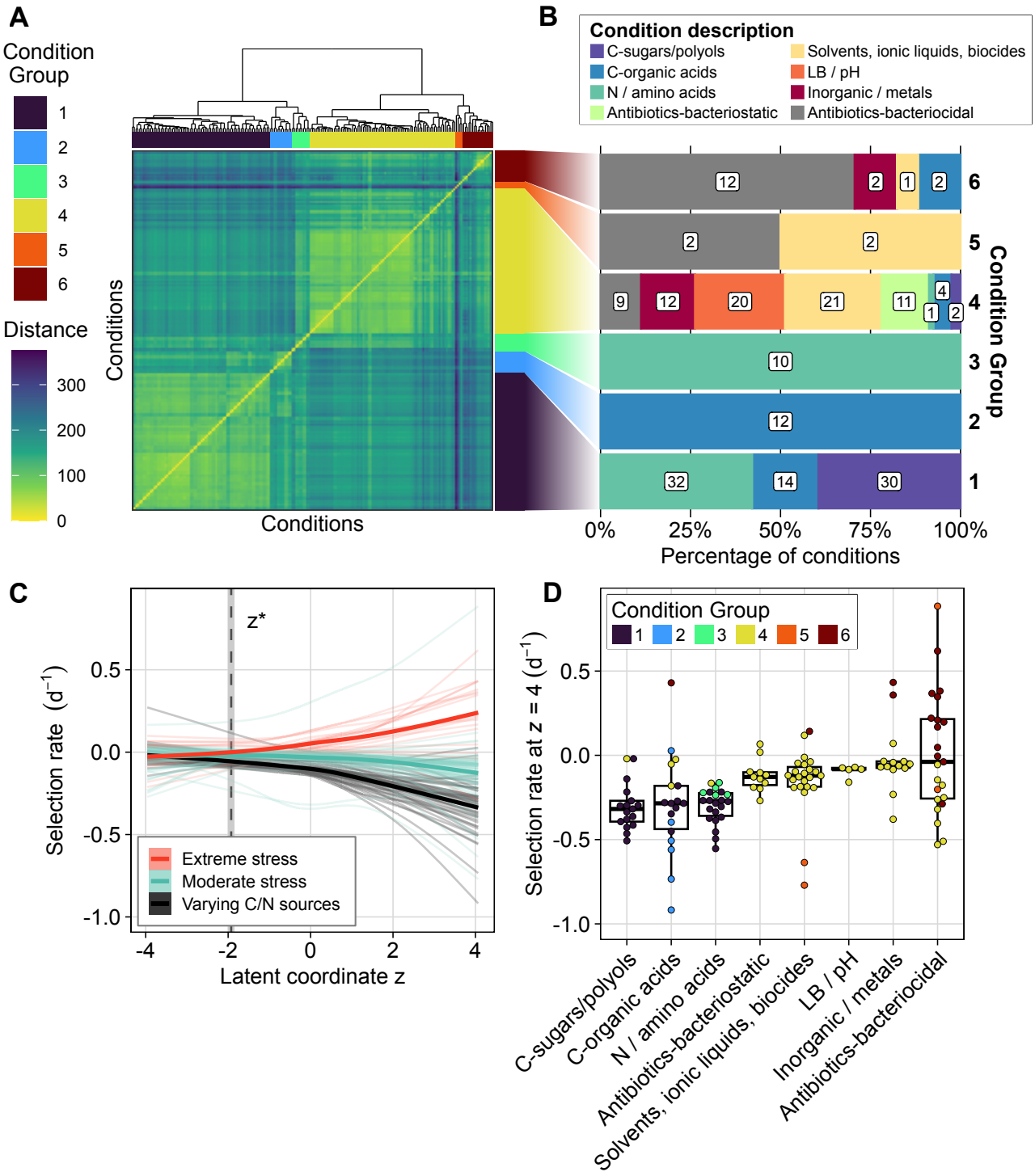
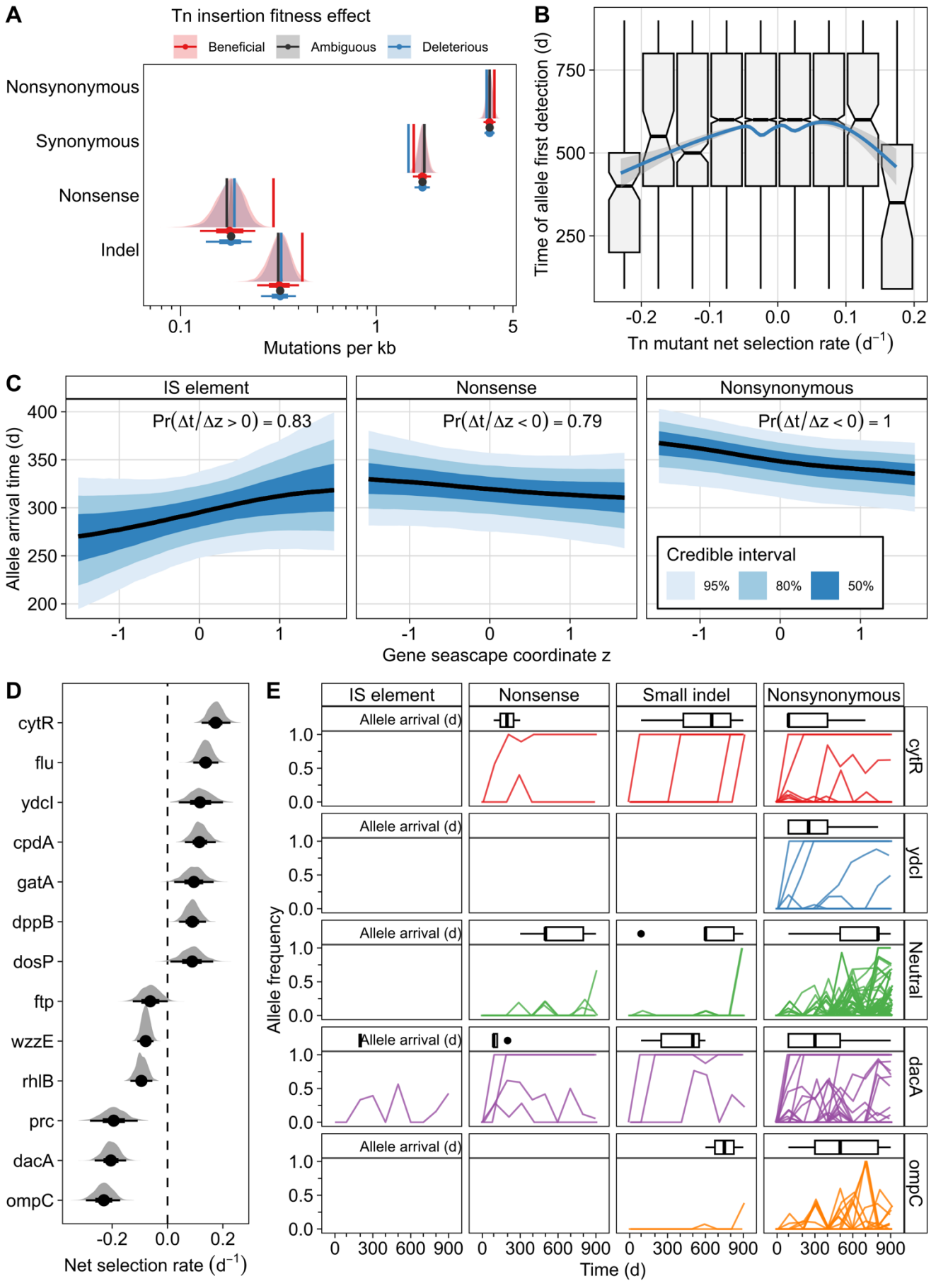


Figure 4. Mapping a published fitness compendium onto the seascape links the latent axis to growth-stress trade-offs across environments. (A) Hierarchical clustering of conditions from a published RB-TnSeq compendium using distance between genome-wide fitness profiles shows distinct groupings of experimental conditions. Conditions are shown clustered against each other on both axes,

with hierarchical clustering indicated by the dendrogram. **(B)** Composition of each condition cluster by coarse condition class shows that clusters separate nutrient limitation from diverse stresses. **(C)** Comparison of fitness across environments to the seascape phenotypic axis shows extreme trade-offs between stress resistance and nutritional competence at high phenotypic values. Each line indicates the smoothed average selection rate in one experimental condition across the latent phenotypic axis, estimated using a generalized additive model. Thick lines (one for each of the three groups) indicate the average selection rate of all experimental conditions in that group. Experimental conditions are grouped as varying C/N sources (clusters 1, 2, and 3), moderate stress (clusters 4 and 5), and extreme stress (cluster 6). **(D)** Selection rate of conditions at the positive extreme of latent axis ($z = 4$), grouped by condition type and cluster. Each point represents one experiment. Selection rate values are predicted from the generalized additive models in panel C.



860

Figure 5. Transposon mutant fitness associates with mutational targets and timing in feast/famine experimental evolution. (A) Genes with strong transposon-mutant fitness effects acquire varying numbers of mutations during a repeated 10-day feast/famine evolution experiment. Distributions and interval bars summarize the null expectation for the number of genes mutated in each fitness bin (median, 80% and 95% intervals as drawn), and vertical lines show the observed number of mutations. **(B)** Mutations in genes with large-magnitude transposon fitness effects arise earlier than genes with small effects over repeated feast/famine evolution. Arrival time is summarized across replicate populations. The relationship between arrival time and selection rate was modeled as a GAM with a penalized cubic spline (blue line with 95% confidence interval). **(C)** Seascape axis coordinates correlate with arrival time of IS element insertions, nonsense mutations, and nonsynonymous mutations in populations adapted to feast/famine. Curves and ribbons are posterior summaries from a cumulative-logit ordinal mixed model predicting allele arrival time from a penalized spline of z. Ribbons show the indicated credible intervals. For each mutation class, predictions were compared at low versus high z (2.5^{th} vs 97.5^{th} percentile), and the posterior probability of a directional difference between z values is shown. **(D)** Transposon mutant fitness values for genes repeatedly mutated across multiple long-term starvation or feast/famine evolution studies show recurring targets can have strong beneficial or deleterious transposon mutant effects. **(E)** Examples of locus-specific mutation dynamics across evolution show distinct trajectories by mutation type. Trajectories show allele frequency time series for representative genes and mutation classes. Inset summaries show the distribution of first detection times for each unique allele. The “neutral” trajectories are mutations acquired in all empirically neutral reference set genes combined. Allele frequencies are colored by gene (or neutral genes) or visualization.

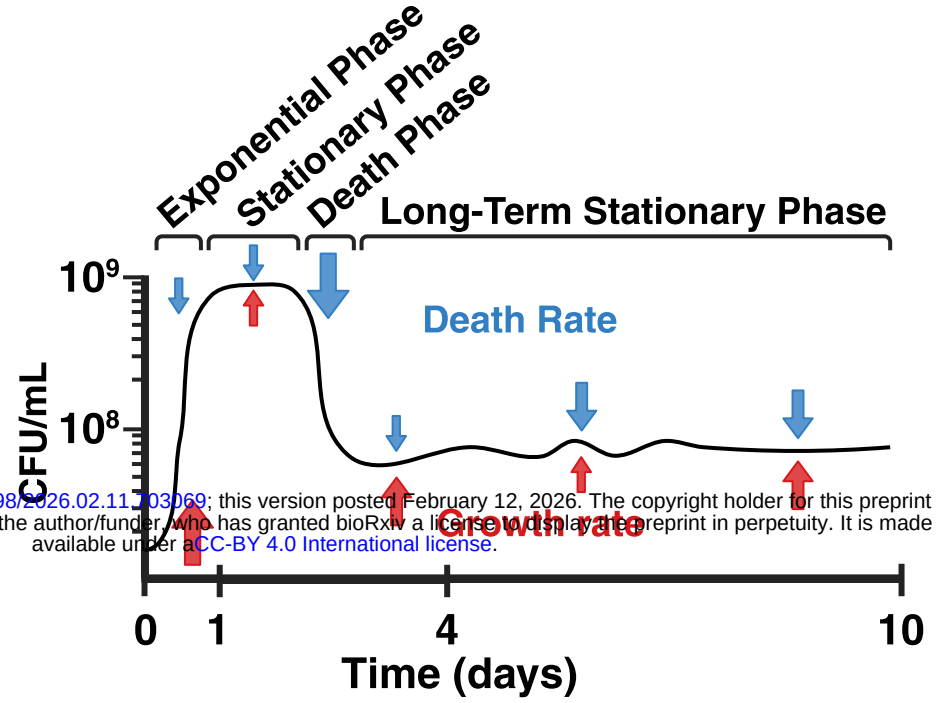
Table 1. Number of confidently non-zero transposon mutants by time interval. Genes were classified as beneficial or deleterious when the posterior probability that the gene's selection rate was above or below zero exceeded 90% (i.e., $Pr(s > 0) > 0.9$ or $Pr(s < 0) > 0.9$).

	# of Gene Disruptions			
Time Interval	Beneficial	Deleterious	Total	Beneficial-Deleterious Skew
0→1 d	503	653	1156	-0.13
1→4 d	757	297	1054	0.44
4→10 d	93	551	644	-0.71
s_{Net}	365	743	1108	-0.34

Beneficial-Deleterious Skew = (B-D)/(B+D)

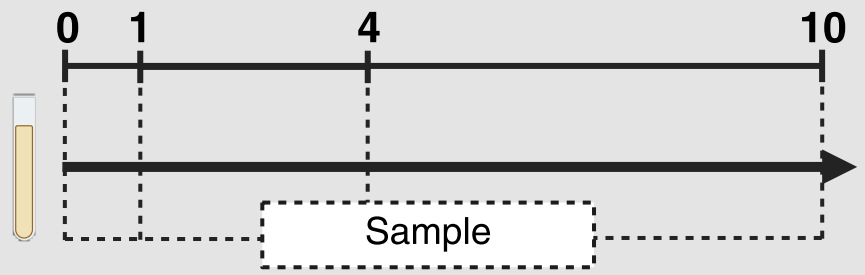
A

The Extended Growth Curve



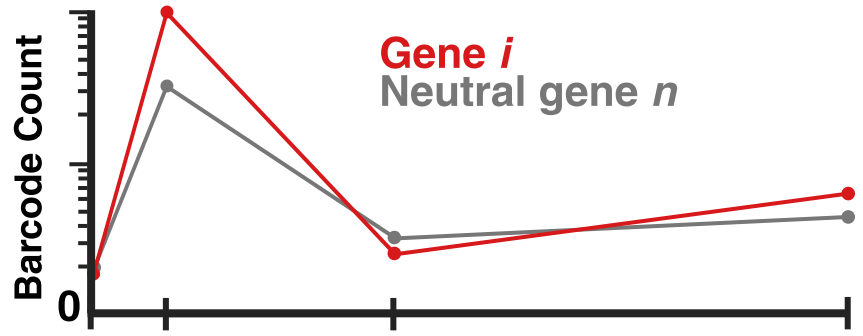
Experimental workflow

Culture RB-TnSeq library
Sample days 0, 1, 4, 10



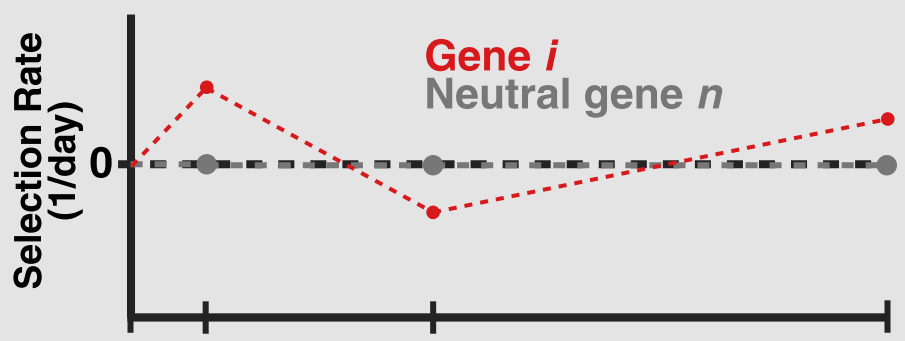
Barcode counts

PCR
↓
Sequence
↓
Count barcodes



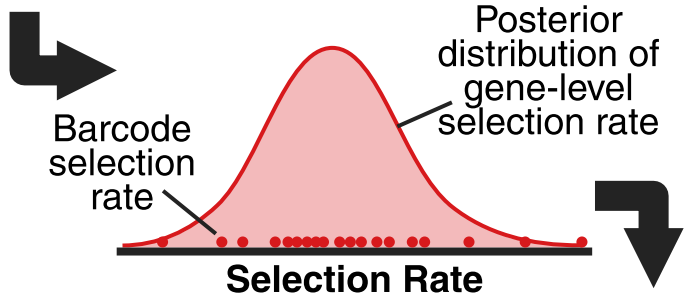
Selection rate s (day^{-1})

$$s = \frac{\log \frac{\Delta N_{mutant}}{\Delta N_{neutral}}}{time}$$



Bayesian multilevel model

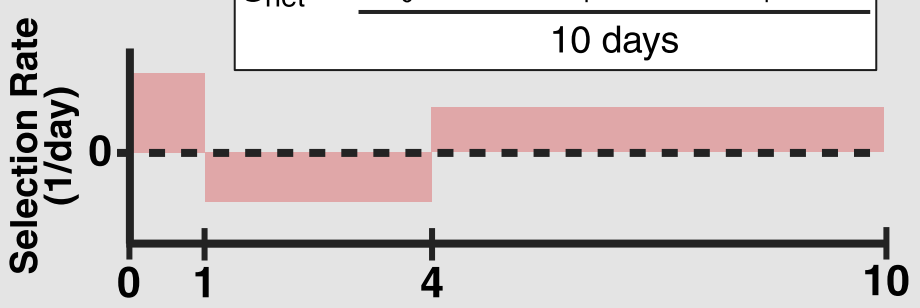
Model s for intervals
0 d→1 d,
1 d→4 d,
4 d→10 d



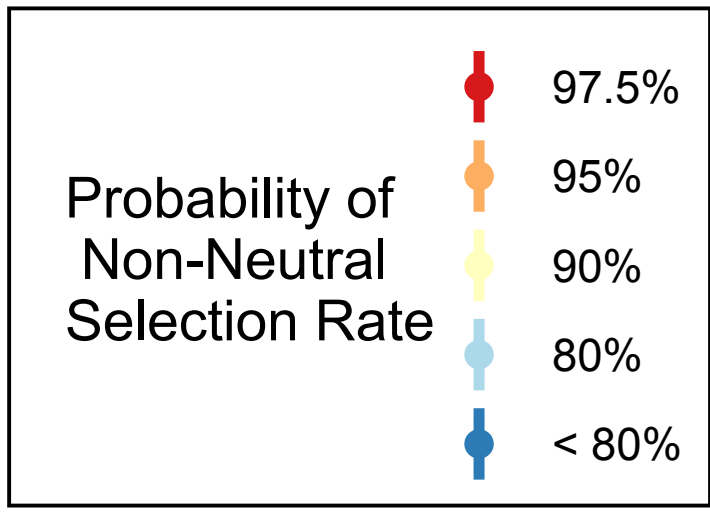
Net selection rate s_{net}

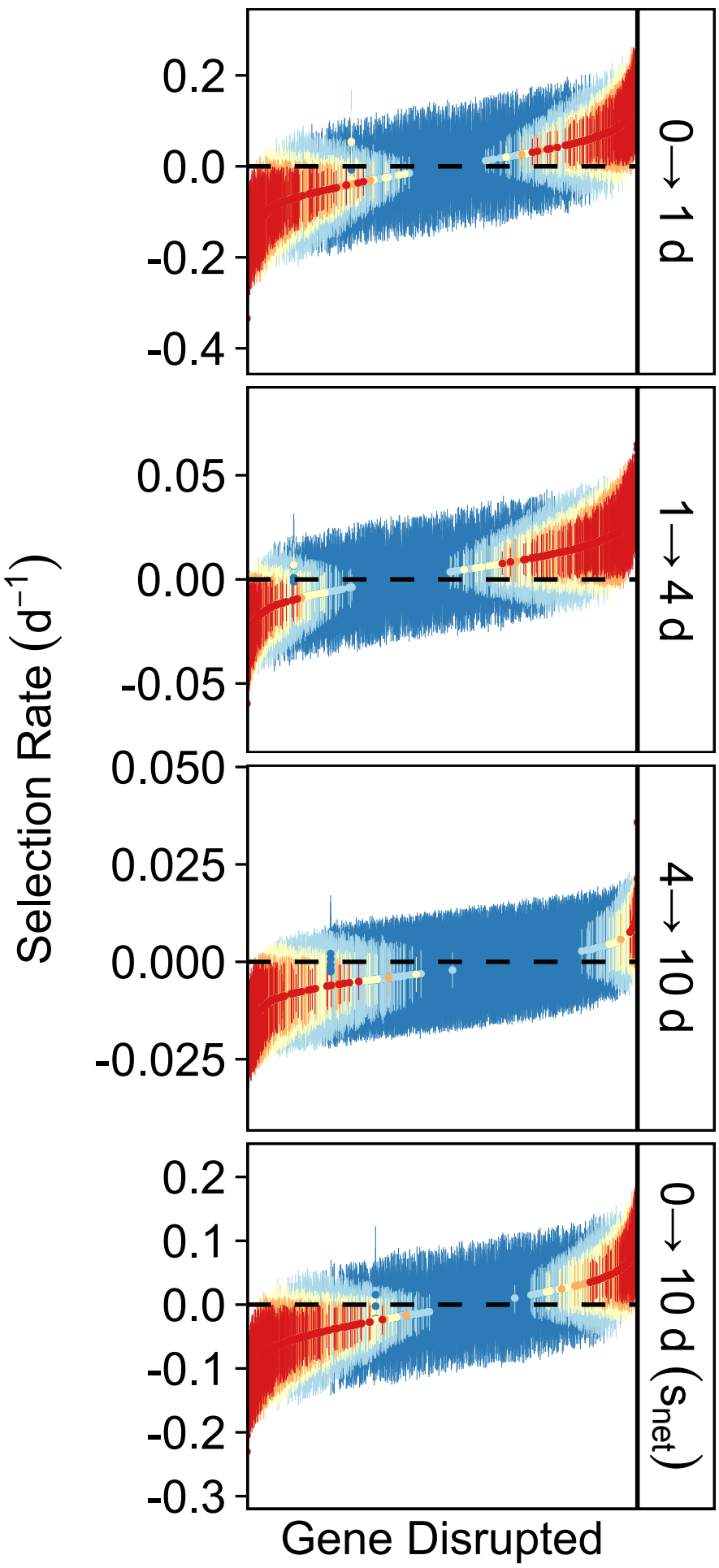
s_{net}:
selection rate from
0 d to 10 d

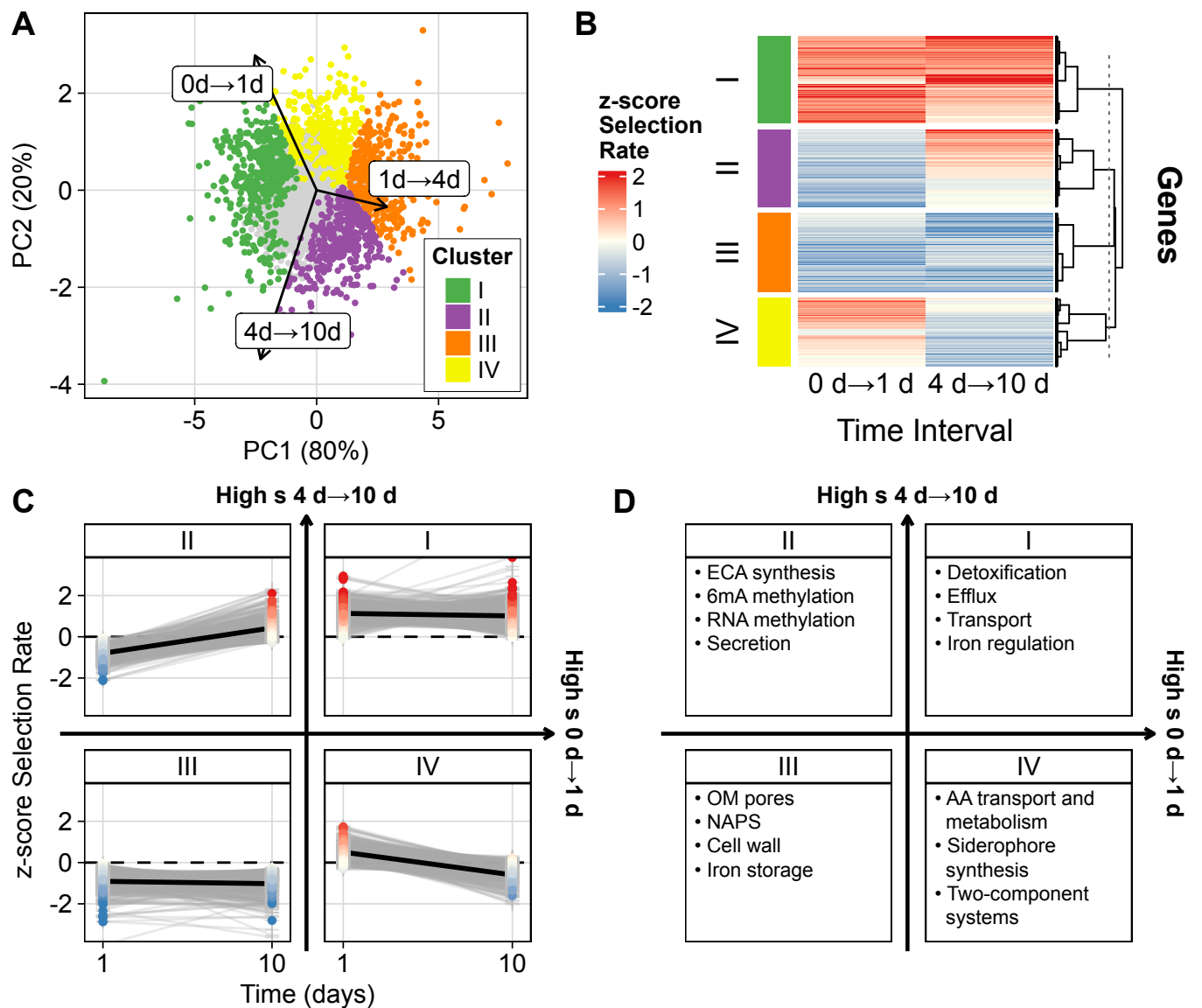
$$s_{net} = \frac{\int_0^1 s(t)\,dt + \int_1^4 s(t)\,dt + \int_4^{10} s(t)\,dt}{10 \text{ days}}$$

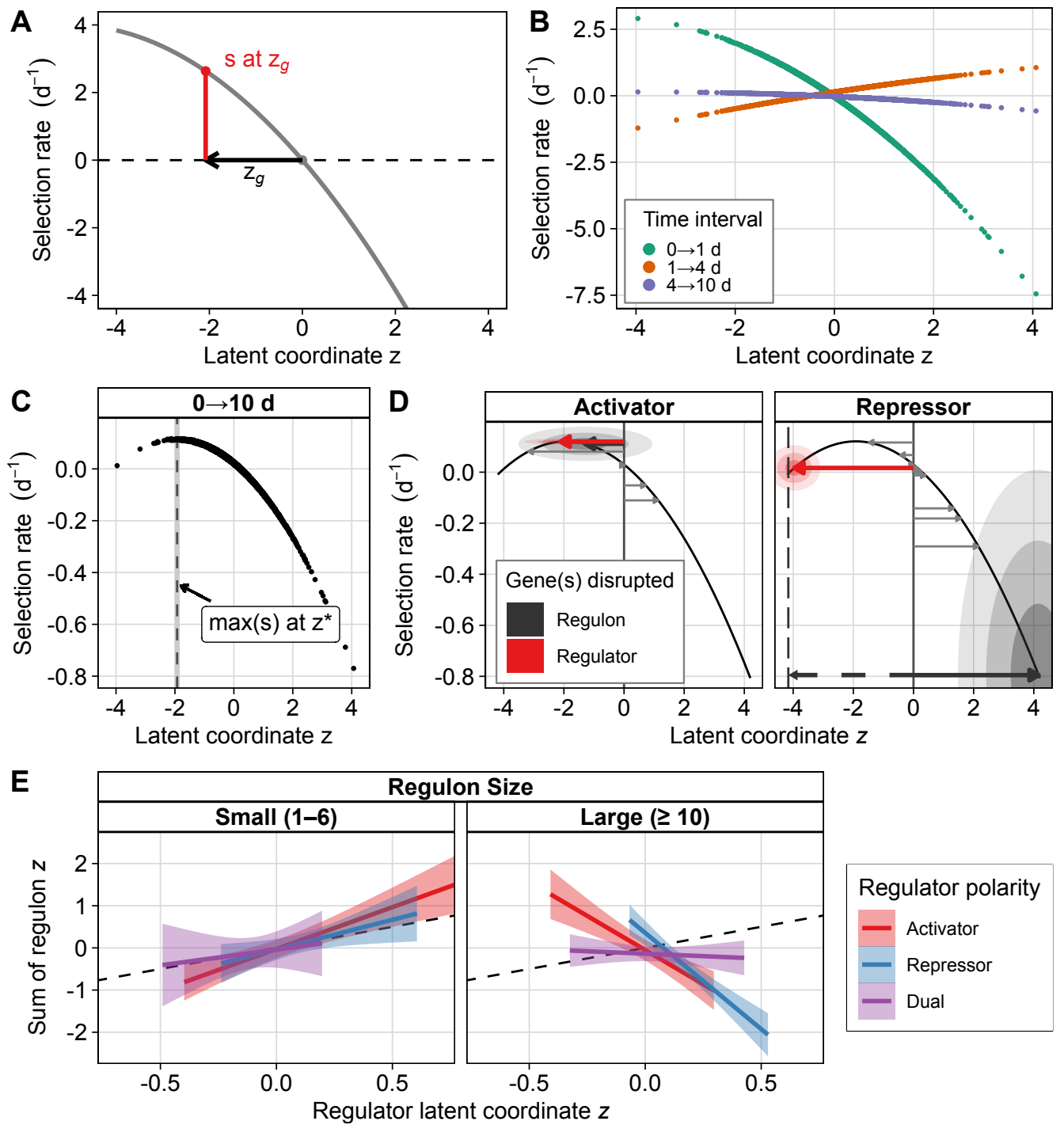


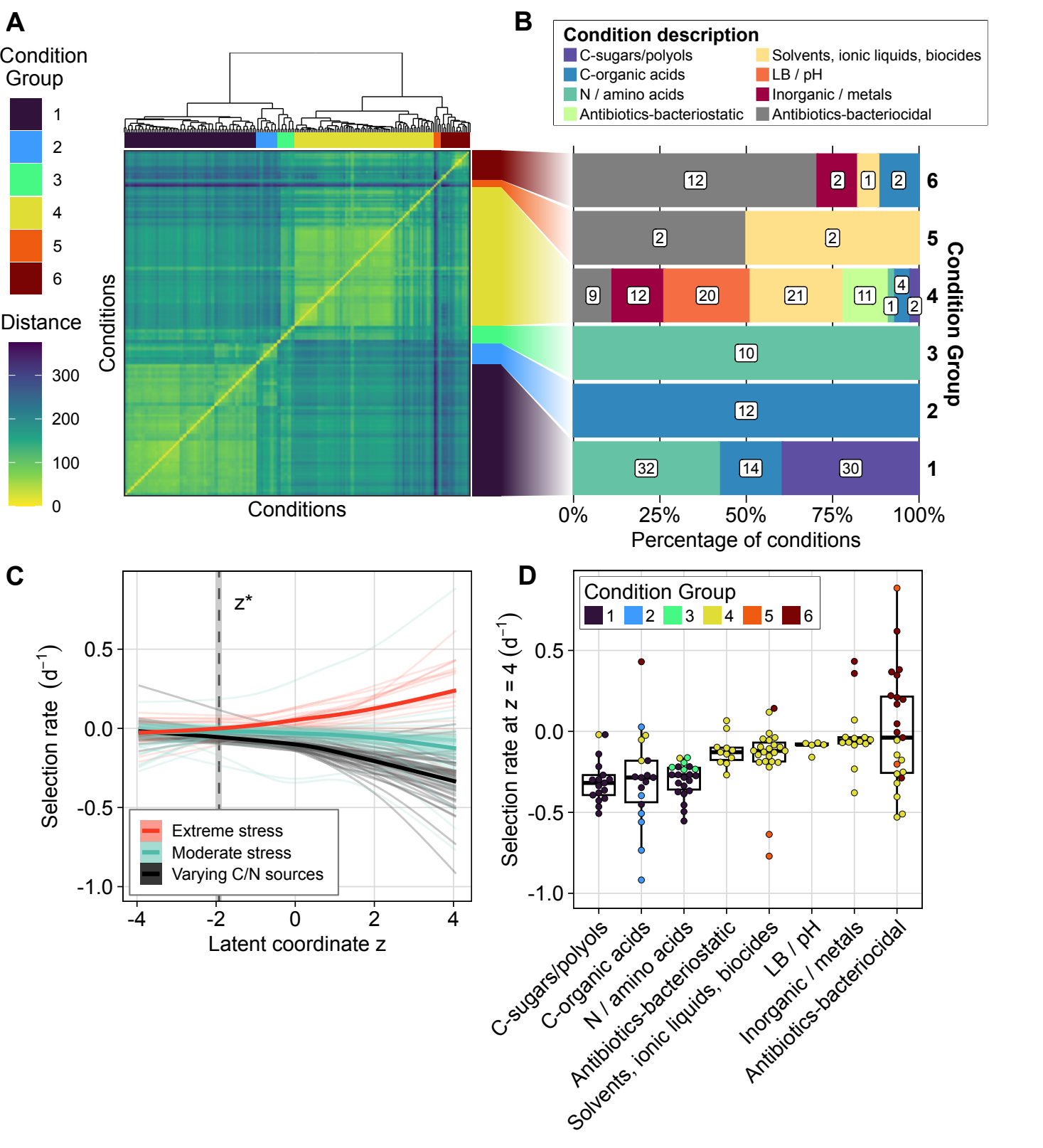
B






A
0d→1d
1d→4d
4d→10d
Cluster
I
II
III
IV
PC2 (20%)
PC1 (80%)
B
z-score
Selection
Rate
Genes
0 d→1 d
4 d→10 d
Time Interval
C
High s 4 d→10 d
High s 0 d→1 d
z-score Selection Rate
Time (days)
D
• ECA synthesis
• 6mA methylation
• RNA methylation
• Secretion
• Detoxification
• Efflux
• Transport
• Iron regulation
• OM pores
• NAPS
• Cell wall
• Iron storage
• AA transport and metabolism
• Siderophore synthesis
• Two-component systems


A
s at z_g
z_g
Selection rate (d^{-1})
Latent coordinate z
B
Time interval
0→1 d
1→4 d
4→10 d
Selection rate (d^{-1})
Latent coordinate z
C
0→10 d
max(s) at z*
Selection rate (d^{-1})
Latent coordinate z
D
Activator
Repressor
Gene(s) disrupted
Regulon
Regulator
Selection rate (d^{-1})
Latent coordinate z
E
Regulon Size
Small (1–6)
Large (≥ 10)
Sum of regulon z
Regulator latent coordinate z
Regulator polarity
Activator
Repressor
Dual


A
Condition Group
1
2
3
4
5
6
Distance
300
200
100
0
Conditions
Conditions
B
Condition description
C-sugars/polyols
C-organic acids
N / amino acids
Antibiotics-bacteriostatic
Solvents, ionic liquids, biocides
LB / pH
Inorganic / metals
Antibiotics-bacteriocidal
12
2
1
2
6
2
2
5
9
12
20
21
11
4
1
2
4
10
3
12
2
32
14
30
1
Condition Group
0%
25%
50%
75%
100%
Percentage of conditions
C
z*
0.5
0.0
-0.5
-1.0
Selection rate (d⁻¹)
Extreme stress
Moderate stress
Varying C/N sources
-4
-2
0
2
4
Latent coordinate z
D
Condition Group
1
2
3
4
5
6
Selection rate at z = 4 (d⁻¹)
0.5
0.0
-0.5
-1.0
C-sugars/polyols
C-organic acids
N / amino acids
Antibiotics-bacteriostatic
Solvents, ionic liquids, biocides
LB / pH
Inorganic / metals
Antibiotics-bacteriocidal