Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution

Bryan C. Dickinson, Aaron M. Leconte, Benjamin Allen, Kevin M. Esvelt, and David R. Liu

To what extent are evolutionary outcomes determined by a population's recent environment, and to what extent do they depend on historical contingency and random chance? Here we apply a unique experimental system to investigate evolutionary reproduction and path dependence at the protein level. We combined phage-assisted continuous evolution with high-throughput sequencing to analyze evolving protein populations as they adapted to divergent and then convergent selection pressures over hundreds of generations. Independent populations of T7 RNA polymerase genes were subjected to one of two selection histories (“pathways”) demanding recognition of distinct intermediate promoters followed by a common final promoter. We observed distinct classes of solutions with unequal phenotypic activity and evolutionary potential evolve from the two pathways, as well as from replicate populations exposed to identical selection conditions. Mutational analysis revealed specific epistatic interactions that explained the observed path dependence and irreproducibility. Our results reveal in molecular detail how protein adaptation to different environments, as well as stochasticity among populations evolved in the same environment, can both generate evolutionary outcomes that preclude subsequent convergence.

Stephen J. Gould famously hypothesized that if the “tape of life”—the long evolutionary trajectory that has led to present life on earth—were rewound and played again, the outcomes would be very different (1). Different evolutionary outcomes could arise from mutational stochasticity (random chance) or from differences in past selection environments (selection history). Adaptation to a common environment can theoretically restore the similarity of evolutionary outcomes by consistently enriching for a subset of these mutations, resulting in evolutionary convergence. Several studies have investigated the reproducibility of evolution by evolving parallel populations from an identical ancestral state. Although phenotypic outcomes are often similar, the underlying genetic changes frequently differ across populations (2–12). Identical genetic outcomes occur more frequently when fitness hinges on the performance of very few genes, as with the evolution of small phages with only a handful of genes (13, 14), cellular traits determined by a single gene (15, 16), and especially for single proteins evolved in vitro (11). Indeed, biochemical explorations of all hypothetical evolutionary trajectories from a single starting sequence to a known evolutionary endpoint (11, 12, 17–20) have demonstrated that there are many more accessible paths to genotypes involving mutations scattered across the genome (21) than those with a similar number of mutations concentrated in a single gene (8). These results have led to suggestions that replaying the tape of life for protein-encoding genes might be surprisingly repetitive (8).

If parallel protein evolution frequently yields similar or identical outcomes, what conditions are sufficient to cause distinct ancestral populations to converge on similar solutions? Despite the importance of this question for the predictability of evolutionary outcomes in common environments, only a handful of experiments have directly or indirectly examined the ability of adaptation to overcome historical differences. These studies have observed high phenotypic convergence for Escherichia coli populations initially separated by genetic drift (2), limited genetic convergence among two ribozyme populations initially evolved with or without a denaturant (10), and no genetic convergence among moderately related phages adapted to high temperature (14). No experiments have directly examined the extent to which closely related protein-encoding genes can undergo convergent genetic and phenotypic evolution. Systematically investigating evolutionary convergence will require a method capable of generating protein populations with a desired level of divergence, then subjecting them to convergent selection pressures over hundreds of generations.

We hypothesized that our recently developed phage-assisted continuous evolution (PACE) system (22) could serve as an experimental platform for the investigation of protein evolutionary convergence and reproducibility in a continuous format, without concern for secondary fitness effects caused by host genome mutations. During PACE, host E. coli cells continuously dilute an evolving population of $10^{10}$ filamentous bacteriophages in a fixed-volume vessel (a “lagoon”). Dilution occurs faster than cell division but slower than phage replication, ensuring that only the phage can accumulate mutations. Each phage carries a protein-encoding gene to be evolved instead of a phage gene (gene III) that is required for infection. Phage encoding active variants trigger host-cell expression of gene III in proportion to the desired activity and consequently produce infectious progeny, but phage encoding less-active variants produce less infectious progeny that are diluted out of the lagoon.

Because PACE allows protein populations to be evolved in parallel over hundreds of generations under controlled mutation and selection conditions, it can be used to systematically investigate the evolutionary convergence and reproducibility of protein-encoding genes in a manner previously restricted to entire genomes. We recently demonstrated that PACE can experimentally explore the effects of mutation rate and selection stringency on evolutionary outcomes (23). In this work, we used PACE to experimentally address the following questions: (i) If initially identical enzyme populations are subjected to distinct selection pressures before converging toward a common evolutionary goal (Fig. 4A), will they evolve a common set of amino acids?


The authors declare no conflict of interest.

Edited* by Gerald F. Joyce, The Scripps Research Institute, La Jolla, CA, and approved April 15, 2013 (received for review November 28, 2012)

acid changes? (ii) How reproducible are the evolved similarities or differences?

Results

Design of Two Selection Pathways. To study the incidence of evolutionary convergence during protein evolution, we designed two selection pathways that subject T7 RNA polymerase (T7 RNAP) to selection pressure schedules with a common beginning and ending but that are otherwise distinct. Both pathways begin with phage encoding the wild-type T7 RNAP gene, which recognizes the T7 promoter with a high degree of sequence specificity (24), and diverge by demanding recognition of either the T3 or SP6 promoter, both of which are orthogonal to one another and to the T7 promoter in nature (25). Each pathway proceeds through a series of evolutionary “stepping stones” that introduce several nucleotide changes at a time (Fig. 1 B and C). The two initial stepping-stone promoters were designed such that wild-type T7 RNAP has enough activity on the T7/T3 and the T7/SP6 promoters to support robust plaque formation. We found that wild-type T7 RNAP retained ∼20% activity and supported phage propagation on each of two hybrid promoters in which all of bases were changed to the T3 promoter except −11G, or to the SP6 promoter except for −8C and −9T (SI Appendix, Fig. S1). These three nucleotides have previously been identified as the primary determinants of T7 vs. T3 (25–28) and T7 vs. SP6 promoter specificity (23, 29–31). The selection schedule for both pathways begins with 24 h of selection using host cells that demand recognition of the first hybrid promoter, 24 h of selection on a 1:1 mixture of host cells containing the first hybrid promoter and host cells containing the intermediate target (either T3 or SP6), and 48 h of selection on the first intermediate target (Fig. 1B).

After 96 h of evolution (∼100 phase generations), the two separate pathways then converge by requiring recognition of the same “final” promoter, a hybrid of the T3 and SP6 promoters in which 12 of 23 positions are altered relative to the starting T7 promoter. Starting with a sample from each of the 96-h populations, we performed PACE for 24 h on host cells that require recognition of the second hybrid promoter (either T3/ final or SP6/ final, which again contain all changes except those at either the −11 or the −8/−9 positions), then for 24 h on a mixture of host cells harboring the second hybrid promoter and the final promoter, and finally for 48 h on host cells that contain only the final promoter. To evaluate evolutionary stochasticity and reproducibility, we performed evolution in four replicate populations for each pathway. Replicate populations were housed in separate lagoon vessels that were diluted with the same host cell culture at a flow rate of 2 volumes per hour, ensuring that the selection histories of sibling populations were as similar as possible. Throughout the experiment, all eight populations were subjected to a mutation rate that is ∼100-fold higher than the basal E. coli mutation rate by inducing a mutagenesis plasmid with arabinose (22). In total, each of the eight populations was continuously evolved for 192 h, representing ∼200 phase generations for the average surviving phage in each population (22).

Fig. 1. Design of two evolutionary pathways. (A) Schematic overview of this study. An enzyme is guided through two separate evolutionary pathways before undergoing convergent evolution towards the same final target. (B) Independent populations of phage-encoded T7 RNAP were continuously evolved over 192 h (∼200 generations) to recognize one of two distinct intermediate promoters (T3 and SP6) followed by a common “final” promoter. Each pathway included two hybrid “stepping stone” promoters (T7/T3, T7/ SP6, T3/ final, and SP6/ final) preceding and following each intermediate. Arrows represent times during which phage populations were fed host cells bearing the indicated promoter. Overlapping arrows represent mixtures of host cell cultures. Evolution was simultaneously performed in four replicate populations for each pathway (eight populations total). (C) Promoter sequences for each target and intermediate promoter, with changed positions from the T3 (red) or SP6 (blue) promoters colored and the transcriptional start site indicated by a dash. Critical contacts at position −11 for T3 and position −8 and −9 for SP6 promoter recognition are underlined.

Evolved Population Phenotypes. At the end of each step of the evolution, lagoons contained phage populations of 107 to 109 pfu/mL. To measure the phenotypic fitness of evolved clones at the middle and end of each experiment, we subcloned T7 RNAP genes from ∼20 randomly chosen phage from each population into an E. coli expression vector after 96 and 192 h of evolution and quantified transcriptional activity on different promoters using a luciferase assay. Evolved RNAP populations at each time point possessed 13–250% average activity on their respective target promoters relative to that of wild-type T7 RNAP on its native T7 promoter, which is 100% by definition (Fig. 2 A–C). Wild-type T7 RNAP has no detectable activity (<1%) on the T3, SP6, or final promoters in this assay.

We selected a subset of individual clones from 96 h and 192 h that span the range of observed activities to assay on the T7, T3, SP6, and final promoters. At 96 h, clones from each pathway were very active on their respective T3 promoter (16–260% average activity) (Fig. 2A) or SP6 promoter (66–113% average activity) (Fig. 2B), but exhibited minimal or no detectable activity on the promoter of the other pathway (Fig. 2 A and B; and SI Appendix, Figs. S2 C and D), demonstrating strongly divergent evolved phenotypes at 96 h.

By 192 h, clones from both pathways exhibit >10% average activity on the T7, SP6, and final promoters, but lower activity on the T3 promoter (Fig. 2C, and SI Appendix, Figs. S2 C and D), suggesting that recognition of the T3 promoter and the final promoter are mutually exclusive. Variants from the T3 pathway lost most of their ability to recognize the T3 promoter as they evolved activity on the final promoter, and six of the eight assayed 192-h T3 pathway variants also gained significant SP6 promoter activity even though these clones were never explicitly selected to recognize the SP6 promoter (Fig. 2C). The two variants from the T3 pathway assayed that did not gain SP6 activity also lost T7 activity and exhibited robust activity only on the final promoter (SI Appendix, Fig. S2C, variants T3-192-2-3 and T3-192-3-14). Variants from the SP6 pathway maintained their ability to recognize the SP6 and T7 promoters while acquiring final promoter activity (Fig. 2C).

Significantly, the 192-h evolved clones from the T3 pathway exhibited an average of ∼three- to fourfold lower average activity on the final promoter than those from the SP6 pathway (Fig. 2D).
This activity difference suggests that the two pathways differed in their ability to evolve high levels of final promoter activity. Moreover, the average activity of assayed clones evolved within sibling populations that experienced identical selection histories also varied by up to 11-fold (e.g., SP6 population 2 vs. SP6 population 3) (Fig. 2C), indicating that even within the same pathway, evolutionary stochasticity was a strong determinant of phenotypic outcome.

Additional Evolution Does Not Resolve Differences in Evolved Activity Levels. To test whether these activity differences between pathways and within each pathway reflected populations that were still evolving, we subjected all eight populations to an additional 24 h of PACE on the final promoter under increased selection pressure at a high flow rate of ~3.5 volumes per hour, corresponding to ~40 additional generations per population. This additional evolution allowed SP6 population 1 to evolve final promoter activity levels comparable to SP6 populations 3 and 4, but did not significantly alter the average final promoter activities of the T3 pathway populations (Fig. 2D and SI Appendix, Fig. S3). These results suggest that some populations reached local fitness maxima by 192 h and indicate that pathway-specific differences in phenotypic outcome persist even after many generations of convergent selection pressure.

Evolved Population Genotypes. We sequenced five to eight complete clones from each of the eight populations at both the 96-h and 192-h time points, including those that had been assayed on the full panel of promoters (SI Appendix, Figs. S4–S8). Additionally, we submitted each population to high-throughput sequencing (HTS) at a coverage level sufficient to identify mutations at frequencies of 2.5% or greater to obtain a more comprehensive picture of sequence variation during evolution (SI Appendix, Figs. S9 and Dataset S1). We used the HTS data to calculate the average diversity of the populations and the number of unique and average mutational compositions within each pathway (SI Appendix, Figs. S10A and B), all of which tended to increase over the course of the evolution. We also analyzed the HTS data using FST, a widely applied measure of population differentiation that estimates the variation between populations (32) (SI Appendix, Fig. S10C). Finally, we constructed phylogenetic trees using the single clone data (SI Appendix, Figs. S11 and S12).

At 96 h, the four populations from the T3 pathway evolved a variety of mutations that we previously observed (22) to confer T3 promoter-recognition activity, including E222K, G542V, V685A, and N748D. The SP6 pathway at 96 h evolved a different set of predominant mutations, including V685A and Q758K/R, in addition to E222K.

At 192 h, the T3 pathway populations evolved a variety of additional mutations not observed at 96 h, including E643K (two of four populations), R756C (three of four populations), Q758K/R (four of four populations), and H772R (three of four populations), and all four SP6 pathway populations enriched R756C (four of four populations). Although N748D, a mutation known to facilitate recognition of the −11 nucleotide in the T3 promoter (26), was highly enriched during the T3 pathway evolution, this mutation did not significantly enrich in the four populations from the SP6 pathway at 192 h, even though the final promoter contains the same nucleotide at position −11, strongly suggesting that N748D is preferred during the T3 pathway. R756 and Q758 interact directly with the template DNA strand at the −8 and −9 positions and are thought to account for the inability of T7 RNAP to facilitate recognition of the −11 nucleotide in the T3 promoter.

To test whether these activity differences between pathways were due to differences in the set of predominant mutations, we calculated the accumulation of shared mutations, defined as the number of shared mutations present in both 96-h and 192-h populations, for each population pair (SI Appendix, Figs. S9 and Dataset S1). We then estimated the variation between populations (32) (Fig. S11). Finally, we constructed phylogenetic trees using the single clone data (SI Appendix, Figs. S11 and S12).

Fig. 2. Phenotypes of evolved RNA polymerases. (A) T3 promoter activity of the four populations on the T3 pathway after 96 h of continuous evolution. Each “+” represents the luciferase activity of a single randomly chosen clone on the T3 promoter luciferase reporter in E. coli cells, normalized to wild-type T7 RNAP on the T3 promoter (100%). Gray, red, blue and green bars represent the average activity of all of the assayed clones from one population and yellow bars represent the background signal with no exogenous RNA polymerase present. The activity of a subset of clones from all T3 pathway populations on the full panel of promoters is also shown (SI Appendix, Fig. S3A). (B) SP6 promoter activity of the four populations on the SP6 pathway after 96 h of continuous evolution. The activity of a subset of clones from all SP6 pathway populations on the full panel of promoters is also shown (SI Appendix, Fig. S3B). (C) Final promoter activity of all eight populations after 192 h of continuous evolution. The activity of a subset of clones from all populations on the full panel of promoters is also shown (SI Appendix, Fig. S3C and D). (D) Average final promoter activity of each pathway at 192 and 216 h. The average final promoter activity of the four populations from each pathway is shown. Error bars represent the SE of the four populations. (E) Crystal structure of the initiation complex of T7 RNAP (25) highlighting some of the key mutations identified by HTS and reversion analysis. The green nucleotides denote positions changed in the final promoter. Red and blue residues show sites of T3-pathway and SP6-pathway mutations, respectively. PDB Structure: 1QLN.
to recognize the SP6 promoter (25). Collectively, these results reveal that despite many generations of evolution on the same final promoter, each pathway by 192 h evolved sets of mutations that were distinct from those that evolved in the other pathway (Fig. 2E). We also observed striking genotypic differences between sibling lagoons in the same pathway. For example, F646L was present at an abundance of 97% of SP6 pathway population 3 by 192 h, but was found in ≤3% abundance in the other three SP6 pathway populations. Similarly, V574A was present in 74% of T3 pathway population 2 at 192 h, but virtually absent from the other three T3 pathway populations at 192 h (Fig. 3A). These observations establish that stochasticity can strongly limit evolutionary reproducibility, even among populations surviving many generations of evolution under identical selection histories.

### Functional Dissection of Key Mutations

To understand the functional significance of some of the most highly enriched mutations from the two pathways, we chose four representative clones from the 192-h time point (two from each pathway) and analyzed the role of the most abundant mutations. First, we incorporated each of the mutations (E222K, V685A, F646L, N748D, R756C, and Q758K) into wild-type T7 RNAP. No single mutation significantly altered the promoter specificity profile of T7 RNAP, which exhibits no significant activity on the T3, SP6, or final promoters, although some mutations (Q758K) gave rise to a modest increase in T3 promoter activity (SI Appendix, Fig. S13). We then reverted each mutation in the four evolved clones back to the wild-type amino acid and assayed the resulting clone’s complete promoter activity profile (SI Appendix, Fig. S14).

Reversion of E222K, present in all four clones from both pathways, results in a global loss of activity across all promoters assayed, regardless of the genetic background, demonstrating that this pathway-independent mutation is required to maintain high activity levels. Consistent with its high enrichment at the end of the evolution in both pathways, reversion of R756C results in decreased activity on the final promoter, but either negligible or increased activities on the T7, SP6, and T3 promoters. Reversion of N748D, the T3 pathway-preferred mutation that contacts the −11 position of the promoter, in a T3 variant that lacks R756C results in a loss of final promoter activity and a gain of SP6 promoter activity. N748D largely excludes recognition of the SP6 promoter, but is critical for final promoter activity in the absence of R756C. Surprisingly, reversion of Q758K/R in either T3 or SP6 pathway variants results in a complete loss of both SP6 and final promoter activities as well as a gain in T3 promoter activity. This result is intriguing because it not only demonstrates that Q758K/R is the mutation responsible for the loss of T3 activity in the T3 pathway, but it shows that the variants from the SP6 pathway are only one mutation away from robust T3 activity.

That different selection pathways and populations within each pathway gave rise to genetic and phenotypic differences following convergent selection suggests that epistatic interactions may have precluded certain populations from achieving high final promoter activity. We therefore examined the phenotypic effects of individual mutations for evidence of epistasis.

### Pathway Dependence Arising from Negative Epistasis Between R756C and N748D

The four T3 pathway populations stochastically enriched either N748D or R756C by 192 h. In contrast, the four SP6 pathway populations predominantly enriched R756C, and little or no N748D, by 192 h. We sought to understand the molecular basis of this striking path-dependent outcome.

The abundance of R756C was strongly anticorrelated with the abundance of N748D (Fig. 3A). Previous biochemical studies on R756C, which contacts N748 in the crystal structure (25) (Fig. 3C), suggested that mutations at this residue might reposition N748 (33). The epistasis of R756C and N748D is strongly supported by HTS data. If randomly distributed, these mutations should be present in the same clone at a frequency of 20% in T3 population 3 based on the individual abundance of each mutation. Instead, the abundance of R756C was strongly anticorrelated with the abundance of N748D (Fig. 3A). Previous biochemical studies on R756C, which contacts N748 in the crystal structure (25) (Fig. 3C), suggested that mutations at this residue might reposition N748 (33). The epistasis of R756C and N748D is strongly supported by HTS data. If randomly distributed, these mutations should be present in the same clone at a frequency of 20% in T3 population 3 based on the individual abundance of each mutation. Instead,
both mutations occurred together in only 6% of 6,661 individual T3 population 3 sequencing reads covering both positions. This epistatic interaction between N748D and R756C has important phenotypic consequences for clones that bear the T7 promoter recognition activity and do not recognize the SP6 promoter are the clones that contain N748D instead of R756C (SI Appendix, Fig. S2C, clones T3-192-2-3 and T3-192-3-14), whereas all of the 14 assayed clones containing the R756C maintain T7 promoter activity and obtain SP6 promoter activity as well. Reversion of R756C in a clone from the SP6 pathway that contains this mutation results in a sevenfold loss of final promoter activity (Fig. 3B). Replacement of N748D in a clone from the T3 pathway also results in a loss of final promoter activity (fourfold), as well as a 15-fold increase in SP6 activity (Fig. 3B). Introduction of R756C into a clone containing N748D indeed results in an almost complete loss of enzyme activity on all promoters, confirming the strong negative epistasis between these two mutations (Fig. 3B).

These results collectively provide a detailed explanation for the observed path-dependent “choice” between N748D and R756C. Although N748D and R756C both substantially increase final promoter activity, N748D strongly decreases SP6 promoter recognition. Therefore, SP6 pathway clones evolved R756C and not N748D because their histories necessarily avoided N748D. Even after selection pressure shifts entirely to transcription of the final promoter, epistasis enforces a fitness valley between these two mutations that prevented SP6 pathway clones from acquiring the N748D mutation.

**Evolutionary Irreproducibility Arising from Sign Epistasis at F646L.** We next sought to address how SP6 population 3 was able to achieve the highest activity on the final promoter. The most notable genetic differences between sibling SP6 populations is the predominance of F646L in SP6 pathway population 3, compared with its virtual absence in any other population (Fig. 3C). This choice of mutations is critical to clones isolated from SP6 population 3, this mutation is detrimental when added to clones isolated from all other populations from both 96- and 192-h time points (Fig. 4). These results indicate that the ability to benefit from F646L has already been lost by 96 h from all populations other than SP6 population 3, and therefore represents an example of “sign epistasis” (17), because F646L results in either increased or decreased activity, depending on genetic context. All four SP6 populations had equivalent SP6 activity levels at 96 h and all shared a common set of core mutations, but clones in SP6 population 3 were uniquely able to use F646L for future productive evolution. Interestingly, F646L increased the activity of wild-type T7 RNAP on the T7 promoter (SI Appendix, Fig. S13), indicating that seven of the eight populations first acquired mutations that exhibit negative epistasis with F646L. These findings collectively provide a molecular explanation for the basis of the nonreproducibility of evolutionary outcomes even among sibling populations subjected to identical selection histories.

**Discussion**

Our experimental investigation of evolutionary convergence points to the existence of at least two clusters of local fitness peaks, accessed by the different selection pathways, on the final transcriptional activity fitness landscape (SI Appendix, Fig. S15). Populations subjected to the T3 pathway were unable to access the higher-activity cluster discovered by populations exposed to the SP6 pathway, despite being subjected to many generations of selection on the same final promoter, establishing the importance of selection history in this system and the limitations of convergent selection pressure. Similarly, populations occupying lower fitness peaks within one cluster were unable to colonize higher peaks in the same cluster, directly demonstrating the limits of evolutionary reproducibility, which in the cases studied here was a result of epistasis.

Although we observed key mutations that consistently arise and contribute to overall activity, a result consistent with previous single-gene protein studies (8), we also observed stochastic outcomes between sibling populations subjected to identical selection histories, both phenotypically (such as SP6 populations 2 vs. 3, which evolved 13% vs. 150% average activity on the final promoter, respectively) and genotypically (such as T3 population 2 that enriched N748D vs. T3 population 1, which enriched R756C instead). These observations more closely parallel previous results in RNA-based evolutionary systems, which featured population
sizes and total generations (34–37) more similar to those explored by PACE than did previous protein evolution experiments.

Taken together, our results have implications for future laboratory evolution efforts. If independent populations and divergent pathways are important outcome determinants during evolution, it may be more effective to subdivide one large evolving population into several isolated subpopulations and to guide those populations through alternative evolutionary stepping-stones. Multiple isolated populations, especially if occasionally subjected to differing selection pressures, may be more likely to avoid local fitness peak traps than a single large population, a result consistent with Sewall Wright’s predictions in his original formulation of the fitness landscape (38).

Methods

For experimental methods see the SI Appendix.

Phage-Assisted Continuous Evolution. PACE was performed as previously described (22). Briefly, during the single-promoter stages of the evolution the lagoon volumes were fixed at 40 mL and during mixing stages of the evolution the lagoon volumes were raised to 80 mL to keep the dilution rate constant.

ACKNOWLEDGMENTS

We thank Irene Chen, Jacob Carlson, David Yang, and Donald Golden for technical assistance. This research was supported by Defense Advanced Research Planning Agency Grants HR0011-11-2-0003 and N66001-12-C-4207; the Howard Hughes Medical Institute; National Science Foundation Grant DMR-1121053; National Science Foundation Grant CNS-0963016; National Institutes of Health National Research Service Award Fellowship F32GM095028 (to A.M.L.); and the John Templeton Foundation (B.A.). B.C.D. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research.


Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution

Bryan C. Dickinson, Aaron M. Leconte, Benjamin Allen, Kevin M. Esvelt, and David R. Liu

Supporting Information

Supporting Materials and Methods

**General methods.** All PCR reactions were performed with HotStartPhusion II polymerase (Thermo Scientific). Water was purified using a MilliQ water purification system (Millipore, Billerica MA). All vectors were constructed by isothermal assembly cloning (New England Biolabs). Single point mutants and reversions were constructed using the Quikchange II Site-Directed Mutagenesis kit (Agilent). All DNA cloning was performed with NEB Turbo cells (New England Biolabs). Plaque assays and PACE experiments were performed using E. coli S109 cells derived from DH10B as previously described (1). Luciferase assays were performed in NEB 10-beta cells (New England Biolabs).

**Phage-Assisted Continuous Evolution (PACE).** Host cell cultures, lagoons, media, and the PACE apparatus were as previously described (1). The accessory plasmids (APs) were constructed by modifying the previously described (1) low-copy AP with an SC101 origin to include a promoter of interest driving a tandem gIII-luciferase cassette, which were then also used for activity assays on isolated clones. To induce mutagenesis, 1% arabinose (wt/vol) was added to all populations throughout all PACE experiments (1). During the single-promoter stages of the evolution the lagoon volumes were fixed at 40 mL and during mixing stages of the evolutions the lagoon volumes were raised to 80 mL to keep the dilution rate constant. For the additional 24-h selection
on the final promoter, host cells flowed through each 40-mL lagoon at 3.5 volumes per hour. Lagoon samples were collected every 24 hours during the course of the evolution.

**Phage pre-optimization.** To minimize the potential fitness advantages of mutations to the phage genome, a previously described VCM13 helper phage with T7 RNAP (HP-T7RNAP) (1) was pre-optimized using PACE. Briefly, HP-T7RNAP was propagated for 6 days using a high-copy AP in which gene III expression is driven by a T7 promoter. Wild-type T7 RNAP was then subcloned into a randomly chosen phage backbone clone from the pre-optimization population and sequenced to ensure correct cloning of the T7 RNAP gene. All evolutions began with 40 µL of $10^9$ pfu/mL of this wild-type T7 RNAP phage.

**Mutagenesis during PACE.** The basal mutation rate of replicating filamentous phage (5.3x10^{-7} substitutions/bp) (2) is sufficient to generate all possible single but not double mutants of a given gene in a 100 mL lagoon following one generation of phage replication. For a target gene 1,000 base-pairs in length, a basal mutation rate of 5x10^{-7} applied to 5x10^{10} copies of the gene in a 100 mL lagoon yields 2.5x10^{7} base substitutions, easily enough to cover all 3,000 single point mutants but not all double mutants. Arabinose induction of the MP can increase the mutation rate to ~5x10^{-5}, yielding ~2.5x10^{9} mutations spread over 5x10^{10} copies of the gene after one generation. The vast majority of these are single mutants which together comprise a target area of ~2.5x10^{12} base pairs. Approximately 1.2x10^{8} mutations should arise in a sequence of this length, sufficient to cover all 9x10^{6} possible double mutants.

**Plaque assays.** S109 cells co-transformed with the MP and an AP of interest were grown in LB media to an OD_{600} of 0.8-1.0. 75 µL of cells were then added to 25 µL dilutions of phage pre-filtered with a 0.2 µm syringe filter. After incubation for 5 min at 22 °C, 750 µL of warm top agar (7 g/L agarose in LB) was added to the phage/cell
mixture, mixed by pipetting up and down twice, and plated onto quartered plates that had been previously poured with 4 mL of bottom agar (16 g/L agarose) in each quadrant. The plates were then grown overnight at 37 °C before plaques were counted.

**T7 RNAP subcloning into expression vectors.** Phage DNA was isolated from lagoon aliquots using a miniprep kit (Qiagen). The T7 RNAP genes were then amplified by PCR using primers that installed Xho I and BamH I restriction sites. The previously described “expression plasmid” (EP) (1) was also amplified with primers that installed these restriction sites and, following PCR, was digested with Dpn I to remove remaining EP plasmid. The resulting PCR products were purified using a PCR purification kit (Qiagen), digested with Xho I and BamH I, and purified by gel electrophoresis on a 0.7% agarose gel. 150 ng of each T7 RNAP library product was then ligated with 50 ng of each EP product using T4 DNA ligase. After PCR purification, the resulting libraries were transformed into S109 cells and plated on agar plates containing spectinomycin (50 µg/mL). After overnight growth, single colonies were picked from the plate, grown overnight, and processed using a miniprep kit to isolate EPs encoding individual T7 RNAP library members.

PCR primers for amplification of T7 RNAP genes:

5’-CAACATTCAAGGATCCACGGAATACCACAAAAGAATGGCATG-3’

5’-AATCATCACACTCGAGCGGCGCAACTATCGGTATCAAGC-3’

Luciferase assays. EPs were co-transformed with an AP of interest into NEB 10-beta cells and plated onto agar plates (16 g/L in LB) with 50 µg/mL carbenicillin and 50
µg/mL spectinomycin. After overnight growth at 37 °C, each well of a 96-well deep well plate containing 1 mL of LB with 50 µg/mL carbenicillin and 50 µg/mL spectinomycin was inoculated with a single colony. After growth with shaking at 37 °C for 3 hours, 150 µL of each culture was transferred to a 96-well black wall, clear bottom plate (Costar). 1.2 µL of 10% decanal in ethanol (vol/vol) was then added to each well, the plate was shaken at room temperature for 1 minute, and then luminescence was measured on a Top Count NXT (Perkin Elmer) or Infinite M1000 Pro microplate reader (Tecan). The OD<sub>600</sub> of each well was measured on a Spectramax M5 microplate reader (Molecular Devices) or an Infinite M1000 Pro microplate reader (Tecan).

The OD<sub>600</sub> of a well containing only media was subtracted from all sample wells to obtain a corrected OD<sub>600</sub> value for each well. The raw luminescence value for each well was then divided by that well’s corrected OD<sub>600</sub> value to obtain the luminescence value normalized to cell density. For the population level assays (Figs. 2A, 2B, 2C, S3), each single clone was analyzed in duplicate from two separately picked bacterial colonies and the resulting values were averaged to obtain a single point on the scatter plots. For all other assays, each sample was measured in triplicate, from three separate bacterial colonies, and the error bars shown are the standard error of those three independent measurements. For normalization, each plate contained triplicate samples of wild-type T7 RNAP co-transformed with T7-AP along with triplicate samples of an empty vector (without any T7 RNAP gene) co-transformed with the T7-AP and other APs of interest. The resulting average values for each sample were divided by the average value for the wild-type T7 RNAP acting on the T7-AP and multiplied by 100 to obtain the percent transcriptional activity relative to wild-type T7 RNAP on the T7 promoter. We validated that picking single clones and growing for 3 hours gives the same result as inoculating with an overnight culture (Fig. S16).

**High-throughput sequencing (HTS) sample preparation.** Lagoon aliquots were processed by Miniprep kit to isolate SP samples (Qiagen). The T7 RNAP genes, along with flanking DNA sequence both upstream and downstream, were amplified by PCR using the following primers:
5’-GGAGCAGGTCGGATTTCG-3’

5’-GTCAAAAATGAAAATAGCAGCCTTTACAGAGAATAACATAAA-3’

The resulting PCR products were purified by gel electrophoresis on a 1% agarose gel and prepared for HTS using a Nextera kit (Illumina) and a slightly modified procedure. Briefly, 4 µL of DNA (2.5 ng/µL), 5 µL TD buffer, and 1 µL TDE1 were mixed together and then heated to 55 °C for 5 min. After purification (Zymo DNA purification kit), the resultant “tagmented” DNA samples were amplified with Illumina-supplied primers using the manufacturer’s protocol. The resulting PCR products were then purified using AMPure XP beads and the final concentration of DNA was quantified using PicoGreen (Invitrogen) and qPCR. The samples were sequenced on a MiSeq Sequencer (Illumina) in 2x150 paired-end runs using the manufacturer’s reagents following the manufacturer’s protocols.

High-throughput sequencing data analysis. A custom MATLAB script (available upon request) was used to align MiSeq reads to the wild-type sequence and count the nucleotide and amino acid positions from which the experimental sample deviates from the wild-type sequence. To compensate for systemic sample preparation and sequencing errors, the observed fraction of mutations at each nucleotide or amino acid position of the wild-type T7-RNAP reference gene was subtracted from the fraction of mutations in a given experimental sample to result in the “corrected fraction mutated” (3). Mutations were defined as amino acid positions with a corrected fraction mutation that is both ≥ 2.5% and at least five standard deviations higher than the corrected fraction mutation of the wild-type reference sequence.

Calculation of Inverse Simpson Index (ISI). To quantify the diversity within each population, we used the inverse Simpson index (ISI) (4). At a single locus, this index is calculated as:
\[ 2D = \frac{1}{\sum_{\text{alleles}} j p_j^2}. \]

This formula can be extended to multiple loci by taking the harmonic mean over the single-locus values of \(2D\). The harmonic mean is more mathematically natural than the arithmetic mean for the ISI (5).

The ISI is an “effective number” (5, 6). In our case, \(2D\) is the effective number of alleles at a locus (in the single-locus version), or the average effective number of alleles across multiple loci (when the harmonic mean is taken over these loci). The number is “effective” because it also takes the evenness of allele frequency distributions into account. For example, consider a locus with two alleles of frequency 99% and 1%. Though the actual number of alleles is two, the ISI gives an effective number of alleles of only \(2D \approx 1.02\), since one allele comprises the vast majority of the frequency. At the other extreme, if there are \(m\) equally abundant alleles, we have \(2D = m\).

In the case of our experiment, the ISI takes values between 1 and 2, since we consider only two alleles (wild-type and mutant) at each locus.

To obtain an unbiased estimate of the ISI from high-throughput sequencing measurements, we first compute an “effective sample size” \(\bar{N}\), as \(\bar{N} = 1/(4\epsilon^2)\), where \(\epsilon\) is the average error per measurement per locus. With this effective sample size, the standard deviation due to sampling error in estimating an allele frequency is \(2\epsilon \sqrt{p(1-p)}\), where \(p\) is the true frequency; this quantity has a maximum value of \(\epsilon\) at \(p = 1/2\). We then use Nei and Chesser’s (7) unbiased estimator for the convention Simpson index:

\[ \bar{H} = 2\bar{N} - \frac{\sum_{\text{alleles}} p_i^2}{2\bar{N} - 1}, \]

where \(p_i\) is the measured frequency of allele \(i\). Finally, we transform this into an unbiased estimator of the ISI:

\[ 2\bar{D} = \frac{1}{1 - \bar{H}}. \]
For a collection of populations (in our case, populations), $\alpha$-diversity is the average diversity across populations. Specifically, we take the harmonic mean of the values of $D$ across each population. The harmonic mean has better mathematical properties than the arithmetic mean for the ISI (5).

**Calculation of $F_{ST}$**. For a single locus with two alleles (wild-type and mutant), $F_{ST}$ (8) is defined by the formula

$$F_{ST} = \frac{\text{Var}(p)}{\bar{p}(1-\bar{p})},$$

where $\bar{p}$ is the mean frequency of the mutant allele across subpopulations, and Var($p$) is the variance of this frequency across subpopulations. $F_{ST}$ takes values between 0 and 1, with 0 indicating that all populations are identical in allele frequency. Large values of $F_{ST}$ indicate greater divergence between subpopulations. We caution, however, that most theoretical work on $F_{ST}$ concerns the case of weak selection, nonzero migration, and many generations (9,10). More modeling work is needed to understand the behavior of $F_{ST}$ under strong selection, zero migration, and relatively few generations, as is the case for this experiment.

We calculated $F_{ST}$ using the $\theta$ estimator of Weir and Cockerham (11). In referring to $\theta$ as an estimator of $F_{ST}$, we adopt Weir and Cockerham’s view that the $n$ subpopulations under consideration can be regarded as a sample from a much larger collection of subpopulations. In our case, this means that the population populations we observed are a subset of the infinite number of populations that could theoretically be created.

The $\theta$ estimator applies to the situation of estimating $F_{ST}$ from a sample of size $N$ taken from a large population. To relate this situation to ours, we calculated an “effective sample size” that would produce a sampling error of the same magnitude as the measurement error in the high-throughput sequencer. Specifically, we calculated the effective sample size as $N_e = 1/(4\epsilon^2)$, where $\epsilon$ is the average error per measurement per locus. With this effective sample size, the standard deviation due to sampling error
in estimating an allele frequency is $2\epsilon \sqrt{p(1-p)}$, where $p$ is the true frequency; this quantity has a maximum value of $\epsilon$ at $p = 1/2$.

To compute $\theta$, suppose that, at a specific locus, the frequencies of the mutant allele in the $n$ subpopulations are measured as $p_1, \ldots, p_n$. Then $\theta$ at this locus is given by (11):

$$\theta = \frac{s^2 - \frac{1}{n-1} \left( \bar{p}(1 - \bar{p}) - \frac{n-1}{n} s^2 \right)}{\bar{p}(1 - \bar{p}) + \frac{s^2}{n}}.$$

Above, $\bar{p}$ and $s^2$ are, respectively, the sample mean and sample variance of the frequency of the mutant allele across subpopulations:

$$\bar{p} = \frac{1}{n} \sum_{i=1}^{n} p_i,$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (p_i - \bar{p})^2.$$

In our calculations, each position in the protein sequence was regarded as a single locus. We obtained a multilocus $F_{ST}$ value by summing the numerator and denominator in the formula for $\theta$ across loci before dividing, as suggested (11).

We estimated the error of $F_{ST}$ due to measurement using the standard formula

$$\sigma = \epsilon \sqrt{\sum_{i=1}^{n} \left( \frac{\partial f}{\partial p_i} \right)^2}.$$

Above, $\sigma$ is the standard deviation due to measurement error in $f$, and $\epsilon$ is the estimated standard deviation due to error in the measurement of each allele frequency $p_i$.

**Phylogenetic analysis.** Several clones from each lagoon were selected at random for Sanger sequencing at 96 h and 192 h (Figs. S4-S7). A phylogenetic tree was constructed from the amino acid sequences at each time point using Bayesian analysis (MrBayes 3.2.1) (12). The evolutionary model specified a single rate across sites, with
gamma-distributed rate variation with a proportion of invariant sites. Satisfactory convergence was assumed if the average standard deviation of split frequencies was <0.02. At least four independent analyses were run and compared for each data set to confirm the tree topology for clades assigned with high confidence. Consensus trees were visualized in FigTree 1.4.

Supporting Discussion

Population genotypic characterization. To statistically evaluate the total genetic divergence between the evolved populations, we used $F_{ST}$, a widely applied measure of population differentiation that estimates the proportion of variation between populations that is beyond the variation seen within populations (13). Compared populations are more related as $F_{ST}$ approaches 0 and are more divergent as $F_{ST}$ approaches 1. We compared $F_{ST}$ values for sibling populations (those subjected to identical selection conditions) with the $F_{ST}$ value for all populations taken together, which represents the total genetic diversity of the system at a given time point. At 96 h, the T3 and SP6 populations exhibited significantly lower $F_{ST}$ than the total $F_{ST}$ for all eight populations (Fig. S10C), indicating that the populations have divergent, pathway-dependent genotypes. From 96 to 144 h the SP6 populations diverged, while the T3 populations converged, indicating that the T3 populations experienced a decrease in divergence as they adapted to the T3/final promoter stepping stone. Instead of converging once the full final promoter selection began at 144 h, the T3 populations diverged again, exhibiting internal divergence by 216 h comparable to the total divergence of all populations, which remained nearly constant throughout the convergent part of the selection. These results reveal that comparisons of population-wide genotypes are unable to account for the different phenotypic outcomes of the two pathways, and suggest that either mutational stochasticity or convergence was a primary determinant of population-wide measures of genetic evolution in this system.

Bayesian phylogenetic analysis on the complete single-clone data (Figs. S11 and S12) confirmed that variants from the same population tend to be genetically similar and
could often be identified as belonging to the same clade. However, variants from different populations following the same pathway (SP6 or T3) could not be confidently grouped with each other at 96 h or 192 h. Differences between population replicates confounded the differences between pathways, further indicating that mutational stochasticity between replicates is a primary determinant of genetic outcome.

**SUPPORTING REFERENCES**
Fig. S1. Activity of wild-type T7 RNAP on intermediate and target promoters. Normalized luminescence of an empty vector control (blue bars) and wild-type T7 RNAP (red bars) on reporter vectors driven by the T7, T7/T3 hybrid, T3, T7/SP6 hybrid, and SP6 promoter. Error bars represent standard error (n = 4).
Fig. S2. Activity profiles of single variants. (A) Activity assays of two clones from each population from the 96 h time point of the T3 pathway. (B) Activity assays of two clones from each population from the 96 h time point of the SP6 pathway. (C) Activity assays of two clones from each population from the 192 h time point of the T3 pathway. (D) Activity assays of two clones from each population from the 192 h time point of the SP6 pathway. All data is normalized to wild-type T7 RNAP acting on the T7 promoter (100% by definition). Single clones correspond to the full sequences in Figs. S4 through S7. The nomenclature for clone ID is “pathway-time point-population #-variant #”. For example T3-96-1-9 is T3 pathway, 96 h time point, population 1, variant 9.
**Fig. S3.** Phenotypes of evolved populations at 216h. Normalized luminescence at 216h of the T3 and SP6 populations on the final reporter vector. Each point represents the activity of a single randomly isolated clone.
**Fig. S4.** Fully sequenced genes of clones from the 96-h time point of the T3 pathway. The clone ID is shown at the top of each column and the mutations present in that clone listed down the column. The nomenclature for clone ID is “pathway-time point-population # - variant #”. For example T3-96-1-2 is T3 pathway, 96 h time point, population 1, variant 2.
Fig. S5. Fully sequenced genes of clones from the 96-h time point of the SP6 pathway. The clone ID is shown at the top of each column and the mutations present in that clone listed down the column. The nomenclature for clone ID is “pathway-time point-population #-variant #”. For example SP6-96-1-2 is SP6 pathway, 96 h time point, population 1, variant 2.
Fig. S6. Fully sequenced genes of clones from the 192-h time point of the T3 pathway. The clone ID is shown at the top of each column and the mutations present in that clone listed down the column. The nomenclature for clone ID is “pathway-time point-population #-variant #”. For example T3-192-1-2 is T3 pathway, 192 h time point, population 1, variant 2.
**Fig. S7.** Fully sequenced genes of clones from the 192 h time point of the SP6 pathway. The clone ID is shown at the top of each column and the mutations present in that clone listed down the column. The nomenclature for clone ID is “pathway-time point-population #-variant #”. For example SP6-192-1-2 is SP6 pathway, 192 h time point, population 1, variant 2.
Fig. S8. Sequencing data for promoters and RBS of single variants. The upstream DNA sequences containing the promoter and RBS driving the expression of the subset of clones assayed were fully sequenced. Although several mutations arise in different clones, they are generally not located in the promoter or RBS and do not correlate with the activity differences of clones. The phage genome, including the promoter and RBS driving T7 RNAP activity, were pre-evolved for 6 days, during which time these sequences had the opportunity to be optimized.
Fig. S9. Mutational frequency of evolved RNA polymerase populations. HTS analysis of mutation frequency at 96 h (A) and 192 h (B). Only mutations present at ≥ 10% at any time point are shown. Red bars represent the four T3 pathway populations while blue bars represent the four SP6 pathway populations.
Fig. S10. Time course of genetic diversity and divergence of populations using HTS data. (A) Inverse Simpson index (ISI) diversity, averaged across sequence positions and populations, during the course of the evolution. (B) Average (solid lines) number of unique mutations in each population and total number of unique mutations across all populations (dotted lines) present in ≥ 2.5% abundance during the course of the evolution. Red lines represent the T3 pathway, blue lines represent the SP6 pathway, and black lines represent all populations taken together. Error bars represent the standard deviation among the four populations. (C) $F_{ST}$ values during the course of the evolution. Red populations represent T3 pathway replicates, blue populations represent the SP6 pathway replicates, and black populations represent all evolved populations taken together. Error bars reflect calculated error from high-throughput sequencing (see Supplemental Methods).
**Fig. S11.** Phylogenetic analysis of amino acid sequences from several clones from each population at 96 h (see Figs. S4 and S5 for genotypes). Bayesian posterior probabilities are given at the corresponding clade in italics; a polytomy is shown if the posterior probability was < 0.5. Clones are named with the following convention: \textbf{W}_X_Y_Z, where \textbf{W} is the number of hours of evolution (96 or 192 h), \textbf{X} is the pathway (T3 or SP6), \textbf{Y} is the lagoon replicate, and \textbf{Z} is a clone designation. The ancestral sequence (T7) was used as the outgroup. Lineages from the T3 pathway are shown in red; those from the SP6 pathway are shown in black. Scale bar units are substitutions per site.
**Fig. S12.** Phylogenetic analysis of amino acid sequences from several clones from each lagoon at 192 h (genotypes shown in Figs. S6 and S7). Bayesian posterior probabilities are given at the corresponding clade in italics; a polytomy is shown if the posterior probability was < 0.5. Clones are named with the following convention: **W_X_Y_Z**, where **W** is the number of hours of evolution (96 or 192 h), **X** is the pathway (T3 or SP6), **Y** is the lagoon replicate, and **Z** is a clone designation. The ancestral sequence (T7) was used as the outgroup. Lineages from the T3 pathway are shown in red; those from the SP6 pathway are shown in black. Scale bar units are substitutions per site.
**Fig. S13.** Activity assays on full panel of promoters of single mutations in T7 RNAP. Full activity profiles for the T7, T3, SP6, and final promoters by in vivo luciferase expression driven by each promoter are shown for wild-type T7 RNAP and wild-type T7 RNAP with each of the single mutations.
**Fig. S14.** Activity assays on full panel of promoters of variants from both the T3 pathway and the SP6 pathway with key mutations reverted. Activity profile of two clones from the 192 h time point from the T3 pathway, T3-192-1-12 (A) and T3-192-2-3 (B), and two clones from the 192 h time point from the SP6 pathway, SP6-192-1-9 (C) and SP6-192-3-9 (D), are shown with their full panel of activity assays. Various mutations in each gene were reverted to their wild-type amino acid for each variant and the resulting revertant was assayed on the full panel of promoters.
Fig. S15. Diagrammatic representation of examples of fitness peaks and epistasis emerging from both pathways as suggested by genotype and phenotype data. Red arrows represent evolution in T3 pathway populations; blue arrows represent evolution in SP6 pathway populations. Solid peaks represent fitness maxima. Dashed lines represent fitness valleys connecting or surrounding fitness maxima.
Fig. S16. Comparison of luminescence/OD<sub>600</sub> measurements of single colonies vs. re-inoculated overnight cultures. (A) Single colonies of transformed cells were picked, grown for 3 h, and then assayed for luciferase activity. (B) Single colonies of transformed cells were picked, grown overnight, then used to re-inoculate 1 mL of media (10 µL of overnight culture added), grown for 3 h, and then assayed for luciferase activity.