Dissecting the Fitness Costs of Complex Mutations

Pablo Yubero ⓘD and Juan F. Poyatos ⓘD*

Logic of Genomic Systems Laboratory, CNB-CSIC, Madrid, Spain

*Corresponding author: E-mail: jpoyatos@cnb.csic.es.

Abstract

The fitness cost of complex pleiotropic mutations is generally difficult to assess. On the one hand, it is necessary to identify which molecular properties are directly altered by the mutation. On the other, this alteration modifies the activity of many genetic targets with uncertain consequences. Here, we examine the possibility of addressing these challenges by identifying unique predictors of these costs. To this aim, we consider mutations in the RNA polymerase (RNAP) in *Escherichia coli* as a model of complex mutations. Changes in RNAP modify the global program of transcriptional regulation, with many consequences. Among others is the difficulty to decouple the direct effect of the mutation from the response of the whole system to such mutation. A problem that we solve quantitatively with data of a set of constitutive genes, those on which the global program acts most directly. We provide a statistical framework that incorporates the direct effects and other molecular variables linked to this program as predictors, which leads to the identification that some genes are more suitable to determine costs than others. Therefore, we not only identified which molecular properties best anticipate fitness, but we also present the paradoxical result that, despite pleiotropy, specific genes serve as more solid predictors. These results have connotations for the understanding of the architecture of robustness in biological systems.

Key words: fitness costs, pleiotropy, RNAP.

Introduction

One recurrent problem in Biology is to understand the impact that mutations have on fitness (Griffiths et al. 2015). Admittedly, this topic has been the center of most recent research in Molecular Biology, with a catch. The majority of mutations, for which we have a well-defined knowledge of the underlying causes of their fitness costs, are "simple." By simple, we refer to mutations in molecular elements with a specific function, for example, an enzyme catalyzing a particular biochemical reaction or a transcription factor linked to the activation of a given gene.

We will not examine here fitness costs of simple mutations but alternatively of those considered "complex." Complex mutations can be commonly established by the pleiotropic action of the molecular agents experiencing the mutation (Dudley et al. 2005). For instance, these agents could refer to a core element of the metabolic or expression cellular machinery, whose function is recognized to be highly pleiotropic. One way to further outline this definition is to add that the said molecular element is active in different contexts (He and Zhang 2006), that is, that it presents a characteristic *environmental fitness cost map*. In this map, one represents pairs of fitness values for both the wild-type (WT) and a given mutant in a set of environmental conditions (fig. 1A). Impairment of a pleiotropic agent should lead to a proportional decrease in fitness characterized by a global scale factor compared with simple mutations that uniquely display fitness costs in specific situations (fig. 1B).

In this work, we initially exemplify these concepts using a genome-wide computational model of *Escherichia coli*'s metabolism (Feist et al. 2007). We then consider the RNA polymerase (RNAP) as experimental model. Three different mutations of the gene *rpoB*, which encodes the β subunit of the RNAP, follow the characteristic environmental fitness cost map of a complex mutation. Indeed, mutations in *rpoB*, usually obtained in response to rifamycins (Rif) (Goldstein 2014)—a class of antibiotics—have been studied in many species and they entail a long list of pleiotropic effects (Jin and Gross 1989; Tóth et al. 2003; Cai et al. 2017; Karthik et al. 2019).

Once we define these mutations as complex, we then ask what set of molecular properties could be a priori relevant to understand their cost in fitness. We thus hypothesize several features, which organize in two broad categories, linked to the global program of transcription and the alarmone (p)ppGpp, or ppGpp onwards.

The former is motivated by the ubiquitous role of the RNAP in gene expression and its coupling to the growth rate. In fact, early works attributed fitness costs to a decreased transcriptional efficiency of the RNAP in *E. coli* (Reynolds 2000), whereas subsequent studies found larger, genome-wide, transcriptional reprogramming in *Pseudomonas aeruginosa* (Qi et al. 2014), *Mycobacterium tuberculosis* (Trauner et al. 2018) and *E. coli* (Wytock et al. 2020) that was not clearly connected to these costs. Our work will enable us to reexamine these issues.
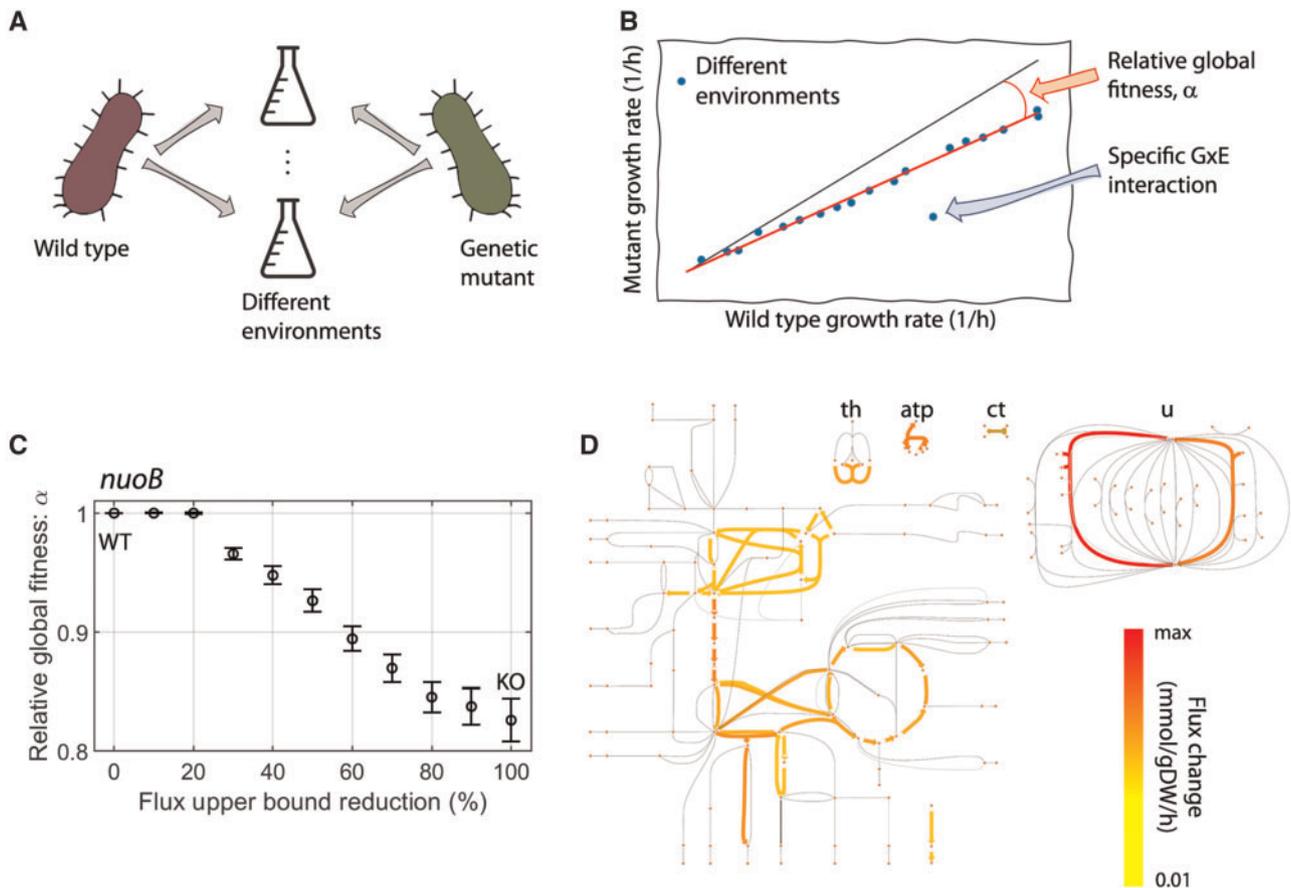
Open Access

Article

Fig. 1. Complex mutations have a characteristic *environmental fitness cost map* because they affect globally the system. (*A*) Environmental fitness cost maps are obtained by measuring, and comparing, the phenotype of a genetic mutant and its WT relative in different environments. In our case, we focus on growth rate. (*B*) Sketch of an environmental fitness cost map. It facilitates the identification of complex mutations and specific gene−environment interactions (GxE). Although the former is a rescaling of the fitness in most environments (red line, relative global fitness α), the latter are shown as outliers from this trend. (*C*) We computed the value of α for multiple mutants of *nuoB* using a computational metabolic model of *E. coli* (see Materials and Methods). Error bars denote the 95% CI of the slope after robustly fitting data to a linear trend (as in panel B; see Materials and Methods; 100% flux reduction denotes a knockout, KO). (*D*) Sketch of *E. coli*'s *nuoB* KO metabolism with the median flux change across all environments. We show only the 10% of reactions that are most affected by the mutation. Transhydrogenase (th), ATP synthase (atp), carbonate (ct), and ubiquinone reduction/oxidation (u) pathways are also shown.

The second broad category includes different features of the interaction between the RNAP and ppGpp mediated by the gene *dksA* (Paul et al. 2004; Irving and Corrigan 2018; Sanchez-Vazquez et al. 2019). Notably, the RNAP associated with *rpoB* mutants was found to work like a stringent RNAP (Zhou and Jin 1998), and an altered stringent response was held responsible for fitness costs in *E. coli* (Wytock et al. 2020). On top of all, the concentration of ppGpp tightly controls optimal resource allocation and hence, growth rate (Zhu and Dai 2019).

Finally, we quantify all these properties in a collection of constitutive genes as "reporters." These genes are useful for reading the RNAP regulatory signal since they do not present any class of specific regulation (Schaechter et al. 1958; Maaløe 1979).

Armed with this data collection, we develop a quantitative framework to predict fitness costs of several *rpoB* mutants. This leads us to reconsider earlier results. Transcriptional efficiency, that is, the rate of mRNA production, does emerge as a relevant determinant. However, comparing transcription

levels between a WT and a mutant that grows at a slower rate calls for special care. Indeed, empirical laws of resource allocation show that gene expression in general, and transcriptional promoter activity in particular, are structurally dependent on the availability of global resources, which in turn, impact growth rate (Liang et al. 1999; Klumpp and Hwa 2008; Klumpp et al. 2009). This is all captured in our results.

Note that although in this example we had some knowledge of the biology involved, in general, our approach does not necessarily need a mechanistic rationale to select a particular predictor. And, although this could seem a significant drawback, it can, in turn, serve to guide research in situations where the origin of fitness costs is unknown. The statistical model can potentially integrate any number of predictors without prior knowledge about their relevance. In such a case, however, the number of experimental points needed to distinguish spurious correlations from significant ones would quickly increase. These are common problems, of course, in other theoretical and applied areas where multiple regression analysis is applied, for example, Quantitative
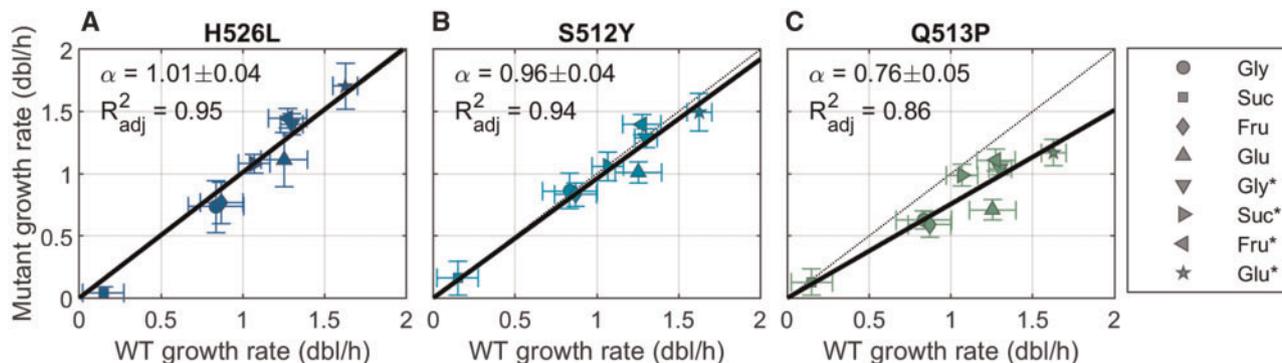
Fig. 2. Experimental *rpoB* mutants display global fitness costs. (A−C) Growth rate of the three *rpoB* mutant strains (H526L, S512Y, and Q513P) and their WT relative in eight different growth media (markers, asterisks denote the addition of casamino acids; see Materials and Methods). Their fitness is proportional to that of the WT, and hence can be described by their relative global fitness α (with its 95% CI interval). Error bars denote one standard deviation among 12 replicates.

Genetics (Falconer and Mackay 1996) or Ecology (Johnson and Omland 2004).

More broadly, our work contributes to the general program of predicting cellular phenotypes from a molecular basis by effectively decreasing the dimensionality assumed to determine such phenotypes and has implications for our comprehension of the architecture of robustness in biological systems.

Results

Complex Mutations Display Global Fitness Costs

We first explore complex mutations in silico, using a genome-scale metabolic model. Specifically, we employed one convenient model of *E. coli* that incorporates 1,260 open reading frames and 2,077 reactions (Feist et al. 2007). We simulate the effect of a mutation on a given enzyme by constraining the fluxes of the reactions in which it participates. Then, we compute the fitness of the WT and mutant strains in a minimal medium supplemented with one of 174 carbon sources (fig. 1A; see Materials and Methods). This enables us to build the environmental fitness cost map of a given mutation (fig. 1B), from which we can systematically extract a broad trend, that is, *the relative global fitness* α, and possible outliers that deviate from it (specific gene−environment, GxE, interactions; see Materials and Methods). Therefore, identifying complex mutations consists of finding those that exhibit consistent changes in many environments, which leads to an α different from 1.

Enzymes involved in the energetic regulation of the metabolism are potential candidates for complex mutations. As a case study, we examined a series of *nuoB* mutants, an oxidoreductase which is part of the respiratory chain, that spanned the entire range of the effect of a mutation: from unconstrained (WT) to null (knockout) flux. Figure 1C indicates that mutants manifest a stronger decrease in relative global fitness ($\alpha < 1$ indicating fitness costs) for larger effects of the mutation. In the limiting case, when the reactions are turned off, we obtain the relative global fitness of the *nuoB* knockout (83%). Note that the complex character of these

mutations is linked to a considerable reorganization of metabolic fluxes (fig. 1D).

Beyond mutants in *nuoB*, we systematically observed both global effects and specific GxE interactions in other enzymes (see supplementary text and fig. S1, Supplementary Material online for further examples and a comprehensive discussion of these mutations).

Overall, complex mutations manifest themselves in a multitude of different environments and are not specific to a particular external cue. This highlights the broader reach of these mutations and their coupling to core enzymes involved in cell growth.

Mutations in *rpoB* Are Complex

We next establish how RNAP mutants represent a well-grounded experimental model system for complex mutations given the RNAP's essential role during gene expression and cellular growth. Specifically, we consider the WT strain REL606 of the bacterium *E. coli* (Barrick et al. 2009) and three mutant derivatives in the *rpoB* gene (with the following amino acid substitutions: H526L, S512Y, and Q513P) that have been selected experimentally through Rif resistance (Jin and Gross 1988; Garibyan et al. 2003).

To obtain an experimental environmental fitness cost map, we measured the growth rate of the four strains in M9 minimal media with different carbon sources (see Materials and Methods). Figure 2 shows this map for the three mutants. We observe that although the derivative H526L (fig. 2A) exhibits no fitness cost, S512Y (fig. 2B), and Q513P (fig. 2C) exhibit mild 4% and large 24% costs, respectively. This global response is similar to the one produced by complex mutations in the genome-scale metabolic model in the previous section. In this case, since these mutations correspond to RNAP (localized in the *rpoB* gene), we can characterize a set of molecular features directly related to the change in transcriptional performance. Ultimately, we will assess these features as potential candidates for anticipating the fitness cost of complex mutations in a statistical model.
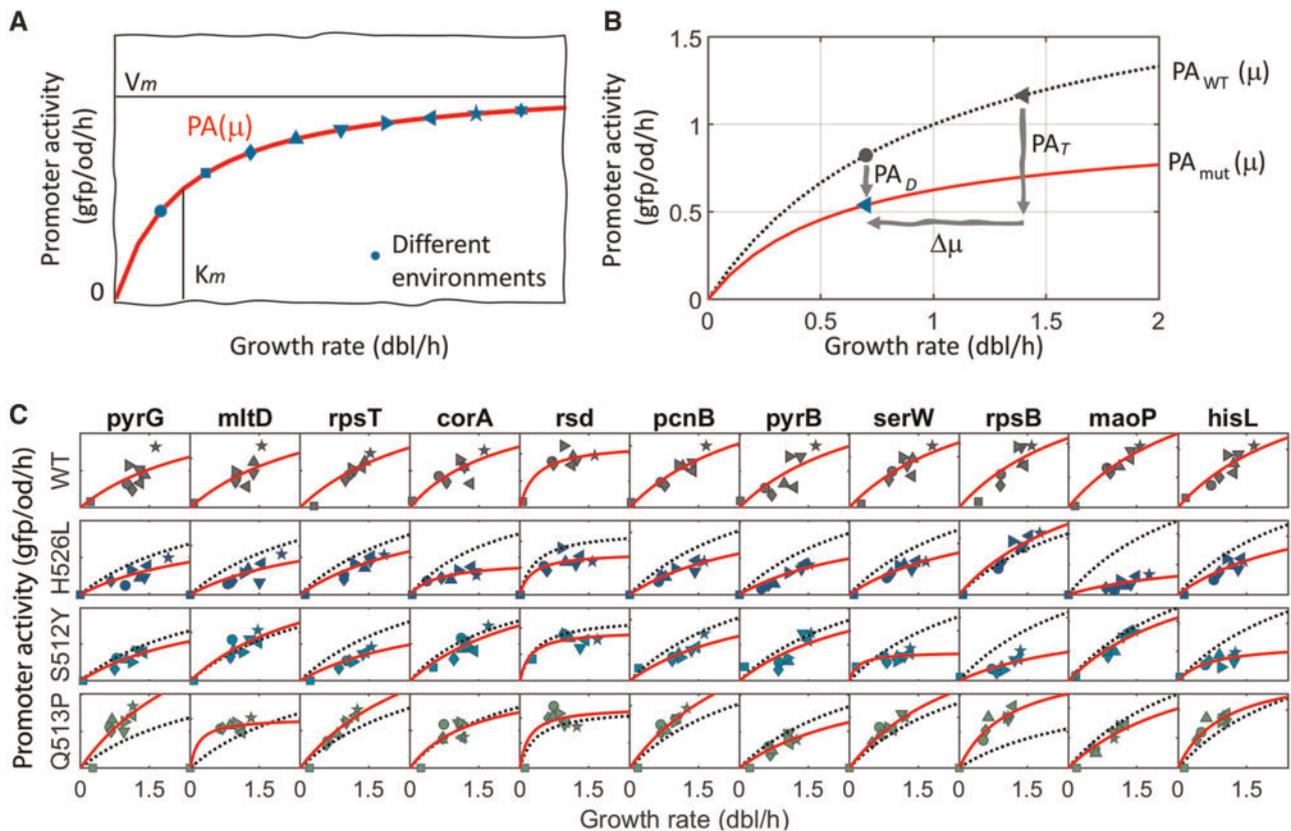
FIG. 3. Uncoupling the total and direct effects of mutations on promoter activity. (A) Sketch illustrating the typical growth-rate dependency of the promoter activity of constitutive genes (red line). These are obtained from PA and growth-rate values during balanced growth in media with different carbon sources (blue symbols) and they are characterized by V_m (the maximal PA), and K_m (growth rate at which PA is half-maximal). (B) Sketch depicting the difference between the total and direct effects of a mutation, PA_T, and PA_D, respectively. PA_T measures the change in PA between the WT and the mutant in the same environment (triangles) but at different growth rates due to fitness costs ($\Delta\mu$). Quantifying the PA(μ) profiles in the WT and mutant (black dotted, and red solid lines respectively) enables us to capture PA_D, which measures the expected change in PA when WT and mutant grow at the same rate but in different environments (different markers). (C) PA(μ) profiles (red lines) of eleven constitutive genes in an array of growth media in all four strains (markers and colors, respectively, as in fig. 2). The corresponding profile of the WT is also shown for comparability (black dotted line).

Mutations in *rpoB* Alter the Global Transcriptional Program

We quantified changes in the transcriptional activity of the RNAP by measuring the promoter activity (PA), that is, the rate of mRNA production. As gene expression is strongly dependent on the growth rate, and consequently on the availability of global resources (Liang et al. 1999; Klumpp et al. 2009), changes in PA observed in the mutants present two possible causes. One is associated with a decrease in growth rate, whereas a second is directly linked to changes in the functional activity of the mutant RNAP (Utrilla et al. 2016). To uncouple these effects, we measured PA as a function of growth rate μ during balanced growth in multiple carbon sources (fig. 3A). We introduced the notion of the *total* and *direct* promoter activity changes PA_T and PA_D, respectively (fig. 3B). Although PA_T measures the difference in PA between the WT and a mutant in a given condition (and different growth rates), PA_D is the expected change in PA between the WT and a mutant when growing at the same rate (and different environmental conditions). This second measure quantifies in this way the potential change in the

activity of the mutated RNAP controlling for changes in the availability of global resources due to fitness costs.

We experimentally measure PA in all strains as the accumulation rate of a reporter green fluorescent protein (GFP) of a selected set of promoters (see Materials and Methods). We selected 11 *constitutive* promoters available in a reporter plasmid library (Zaslaver et al. 2006). Constitutive genes are particularly suitable because their expression does not rely on the concentration of any specific transcription factor, and thus they read the availability of global resources and the performance of the pool of RNAPs (Schaechter et al. 1958; Maaløe 1979; Klumpp and Hwa 2008). We then model the growth-rate dependencies of promoter activities, PA(μ), from PA measurements during exponential growth in eight different media (see Materials and Methods).

Figure 3C shows the growth-rate dependencies of the promoter activities of the selected genes, in all strains, together with the best fit to a Michaelis–Menten equation $PA(\mu) = V_m\mu/(K_m + \mu)$ with parameters V_m, maximum expression, and K_m, growth rate at which PA is half-maximal (fig. 3A; see Materials and Methods). We recovered not only that, in

general, each promoter follows a specific profile with different parameters V_m and K_m, but also that some of them reside in the linear regime with large K_m (Liang et al. 1999; Gerosa et al. 2013; Yubero and Poyatos 2020).

Most importantly, the activity of promoters in the RNAP mutant strains still follow hyperbolic patterns although different across strains. We found a significant tendency of H526L and S512Y toward smaller values of V_m whereas Q513P displayed a general decrease in K_m (supplementary fig. S2A and B, Supplementary Material online). However, the quantitative change in these parameters is mutation- and promoter-specific. Therefore, changes in these profiles, that is, in the global transcriptional program, are candidates for predictors of fitness costs.

The availability of a predictive model of PA(μ) for all promoters in all strains enables us to distinguish between the direct effect of a mutation, PA_D to the total change in promoter activity PA_T. Interestingly, in most promoters, we observe significant direct effects. Even if RNAP mutations do not produce fitness costs, as in strain H526L, most promoter activities are significantly altered in a consistent manner across environments (>80%, supplementary fig. S2C—H, Supplementary Material online). This highlights the importance of the global transcriptional program and better quantifies the changes in PA when controlling for growth rate. Hence, apart from the total effects on PA, we also consider separately the direct effects as potential fitness predictors.

Mutations in *rpoB* Alter the Action of ppGpp-RNAP

The performance of the RNAP is strongly dependent on its interaction with the alarmone ppGpp playing a pivotal role in controlling growth rate in both minimal and rich media (Potrykus et al. 2011; Irving and Corrigan 2018; Zhu and Dai 2019) and during the stationary phase (Hirsch and Elliott 2002). Besides, changes in the concentration of ppGpp, together with the presence of *dksA*, alters the genome-wide transcriptional pattern with profound consequences in resource allocation (Paul et al. 2004; Sanchez-Vazquez et al. 2019; Zhu and Dai 2019). Since some *rpoB* mutants also display defective RNAP-ppGpp action (Zhou and Jin 1998), we posit that mutations should also impact both growth and transcription during the stringent response at the exit of the exponential phase, and during the stationary phase.

Thus, we considered the following three proxies to quantitatively assess alterations in RNAP—ppGpp interactions: The promoter activity and protein level during stationary phase and the deceleration in growth rate during the stringent response. The first assesses the transcriptional reprogramming in stationary phase. The second is a measure of the aggregate effect of PA deregulation during both balanced growth and stationary phase. Finally, the deceleration rate measures the efficiency of RNAP-ppGpp in arresting growth.

Firstly, we measured the promoter activities in stationary phase during the last 2 h of the experiment (PA$_f$, fig. 4A; note that other time windows produce qualitatively similar results). This parameter describes the appropriate ability of the pair RNAP-ppGpp to reprogram transcription when nutrients are depleted. We observe that only a subset of

4, 1, and 3 promoters in strains H526L, S512Y, and Q513P, respectively, have a significant under/overactivity in stationary phase across different growth media.

Secondly, in analogy to PA$_f$ we measure the protein level also in stationary phase (p$_f$, fig. 4E) to assess the combined effect of reduced promoter activity and growth rate. We find that these are more often altered than PA$_f$ although the responses are still mutation- and promoter-specific (fig. 4F—H). Note that the relative change of p$_f$ tends to be negative in strains H526L and S512Y as opposed to Q513P.

Finally, we used the deceleration rate as a proxy of the interaction RNAP-ppGpp at the onset of the stringent response, given its fundamental role in arresting growth at the exit of the exponential phase. We measure the deceleration rate as the slope of the linear fit to the instantaneous growth rate during 4 h after the exponential phase (fig. 4I, again, other time windows produce similar results). Unsurprisingly, across all strains we observed a strong negative linear correlation between the deceleration rate and the growth rate during balanced growth (fig. 4J). Thus, reaching a larger growth rate during exponential phase leads to a faster deceleration rate during growth arrest. Then, we searched for changes in the *normalized* deceleration rates across mutants, which controls for the respective exponential phase growth rates. Figure 4K shows that both strains with fitness costs display a significantly reduced normalized deceleration rate with respect to the WT across environments.

A Statistical Model for Complex Fitness Predictions

The characterization of all previous features equipped us with the necessary data to introduce a statistical model capable of explaining the fitness costs of three *rpoB* mutants in eight different growth media from specific molecular determinants (fig. 5A). Given the uncoordinated changes in expression observed in the previous sections, not only do we seek which determinants are best suited for fitness costs prediction but also of which reporter genes.

Specifically, we considered the following predictors related to gene expression: The total and direct promoter activity changes PA_T and PA_D, respectively; the global transcriptional program parameters V_m and K_m; the promoter activity during stationary phase PA$_f$; the protein level during stationary phase p_f; and the normalized deceleration rate during growth arrest $\partial_t \mu$. The model describes the relative growth rate of mutants as a function of the relative change of predictors. The expression for each gene, in Wilkinson notation, is:

$$\frac{\mu - \mu_{wt}}{\mu_{wt}} \sim 1 + \sum_i \frac{p_i - p_{i,wt}}{p_{i,wt}}, \qquad (1)$$

where μ is the growth rate; p_i is the ith predictor; the subscript wt denotes the WT strain and 1 refers to a constant intercept. Therefore, a positive parameter estimate implies that the relative change of the predictor correlates positively with the relative change in the growth of the mutant (supplementary fig. S3, Supplementary Material online shows all cross-correlations between variables). Each model integrates data of the three mutants during growth in the eight media, fitting
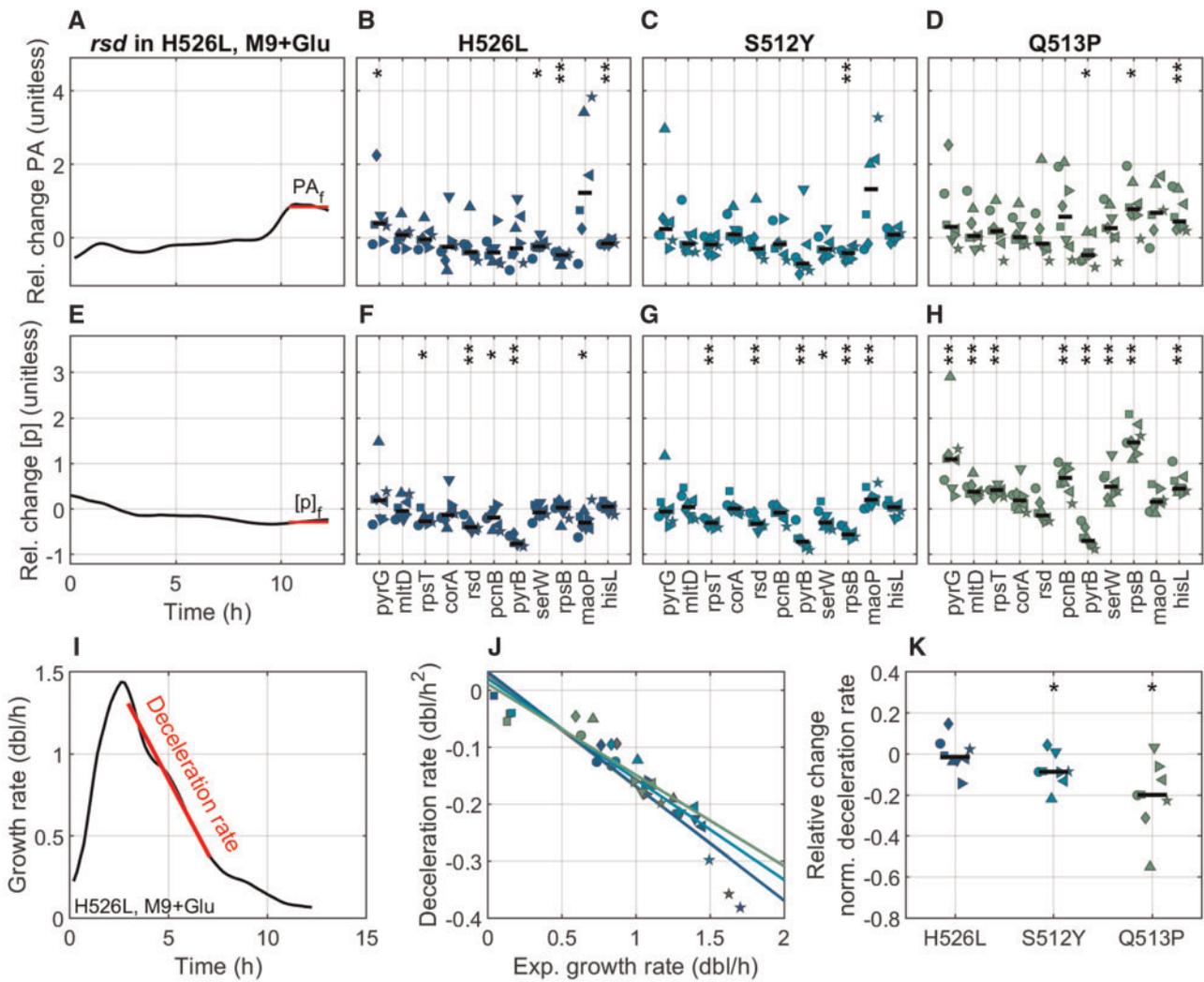
FIG. 4. Promoter activity and protein concentration during stationary phase, and the deceleration rate constitute additional potential predictors of fitness costs. (A) Relative change of the final promoter activity, $(PA_f^{mut} - PA_f^{wt})/PA_f^{wt}$, exemplified with data of *rsd* during growth in M9+glucose. (B−D) Relative change in PA_f of all promoters and mutant backgrounds (x-axis). (E) Relative change of the final protein level, $([p]_f^{mut} - [p]_f^{wt})/[p]_f^{wt}$. (F−H) Relative change in $[p]_f$ of all promoters and mutant backgrounds (x-axis). (I) deceleration rate during growth in M9 and glucose of H526L. (J) Deceleration rates correlate strongly with the exponential growth rates reached in that particular media (markers) in all strains (colors; Pearson's $\rho < -0.86$ and $P < 0.01$ in all strains). (K) Even when controlling for this correlation, the relative deceleration rates of different mutants differ significantly. Note that although the first two scores are measured during the last 2 h of the experiment when cultures are in stationary phase (red horizontal lines), the deceleration rate is computed from the change in growth rate right after the exponential phase (slope of the red line). In all panels, we tested a homogeneous response, either positive or negative, across all environments using a two-sided Wilcoxon sign rank test for medians (*$P < 0.05$ and **$P < 0.01$). Colors and markers denote strain and media composition as in figure 2.

a total of 24 points. Note that, for variables that are not environment-dependent (V_m and K_m) we consider an equal value across environments.

With the statistical model, we seek which genes best describe the fitness changes and with which combinations of predictors. To do so, we used an algorithm that starts from an initially constant model, that is, fitness costs do not depend on any variable, and that iteratively adds predictors depending on whether the Bayesian information criterion (BIC) improves. The BIC enables the comparison of different models and it decreases when the likelihood of the model increases, but it penalizes the addition of parameters to a model. Therefore, this approach favors better predictions while preventing data overfitting. In addition, note that the

final models do not necessarily contain all predictor variables. The results of the best model for each promoter are shown in figure 5 and table 1.

Figure 5B shows two measures of the goodness of fit. First, the root mean squared error is a measure of how accurately models predict the relative fitness costs. And second, the R_{adj}^2 measures how the final models are better than the initially constant model. We observe that all models reach a convenient RMSE and R_{adj}^2, with the exception of *corA*, an ion transporter; *pyrB*, part of the pyrimidine biosynthesis pathway; and *pcnB* involved in RNA polyadenylation.

In our protocol we guided the addition of predictors in the models with the BIC. Hence the change in BIC, that is, ΔBIC, quantifies how the inclusion of a variable improves the
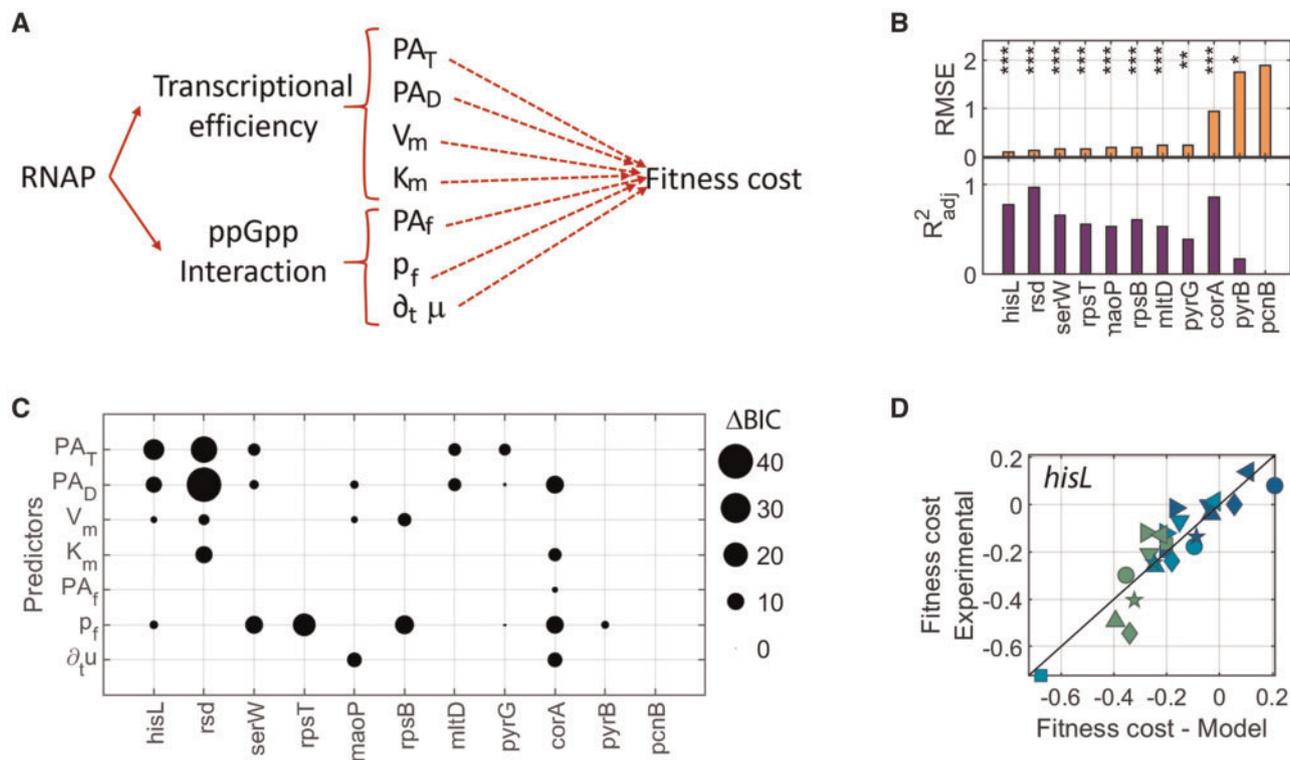
FIG. 5. Anticipating fitness costs from molecular predictors of a variety of promoters. (A) To predict the fitness costs of three RNAP mutants in different media conditions, we used a linear model with an iterative addition and subtraction of predictors following the BIC. Predictors are organized into two broad categories related to the RNAP: Transcriptional efficiency and ppGpp interaction. (B) We assess the goodness of fit of the final models with two parameters. The root mean squared error (RMSE, top; *P < 0.05, **P < 0.01 and ***P < 0.001) quantifies the errors in the predictions, while R^2_{adj} (bottom) compares the performance of the final model to the baseline constant model. (C) The change in BIC (ΔBIC) of a variable is associated to the increased performance of the corresponding model when including the variable. Note that the final models do not include all variables. (D) Fitness costs as expected from the final model of *hisL* (x-axis) compared with the measured values (y-axis; markers and colors as in fig. 2).

Table 1. Linear Models for the Anticipation of Fitness Costs.

	(Intercept)	PA_T	PA_D	V_m	K_m	PA_f	p_f	$\partial_t\mu$	RMSE
hisL	0.00(6)	0.32(8)	−0.54(8)	0.3(1)	—	—	−0.11(5)	—	0.096
rsd	−1.6(4)	0.85(4)	−0.60(4)	−2.3(5)	−3.91(4)	—	—	—	0.114
serW	−0.13(4)	0.32(8)	−0.3(1)	—	—	—	−0.26(5)	—	0.146
rpsT	−0.09(3)	—	—	—	—	—	−0.31(6)	—	0.149
maoP	−0.18(7)	—	−0.07(3)	−0.3(2)	—	—	—	−0.19(4)	0.188
rpsB	−0.11(5)	—	—	0.4(1)	—	—	−0.19(3)	—	0.189
mltD	−0.10(5)	0.5(1)	−0.6(1)	—	—	—	—	—	0.225
pyrG	−0.11(6)	0.4(1)	−0.5(1)	—	—	—	−0.10(6)	—	0.230
corA[a]	−0.5(3)	—	−3.0(5)	—	1.7(4)	−1.8(4)	3.7(6)	−2.1(7)	0.933
pyrB[a]	1.7(6)	—	—	—	—	—	2.1(9)	—	1.75
pcnB[a]	0.5(4)	—	—	—	—	—	—	—	1.89
Median	−0.10	0.40	−0.54	0	−1.1	−1.8	−0.11	−1.17	0.189

NOTE.—We show the coefficients of the predictors (columns) obtained for the data set of each promoter (rows). The number in parentheses is the standard error of the coefficient in the last decimal digit shown. The last column contains the root mean squared errors as a measure of goodness of fit. Models were selected in a step-wise manner following the BIC (see Materials and Methods).
[a]Genes with largest RMSE that fit poorly the fitness costs.

prediction of the model. Figure 5C shows the ΔBIC of each variable present in the final model for each gene. We observe a clear pattern of PA_T and PA_D as the principal predictors of fitness costs, although p_f is also present in multiple models. Although the first two are only related to transcriptional efficiency of the RNAP during balanced growth, the third partly

involves the interaction between RNAP and ppGpp during the stringent response.

Overall, we find that despite pleiotropy, a multivariate regression with as little as four (median) variables as predictors anticipates the fitness costs of different *rpoB* mutants growing in a variety of carbon sources (fig. 5D). For several genes, a few
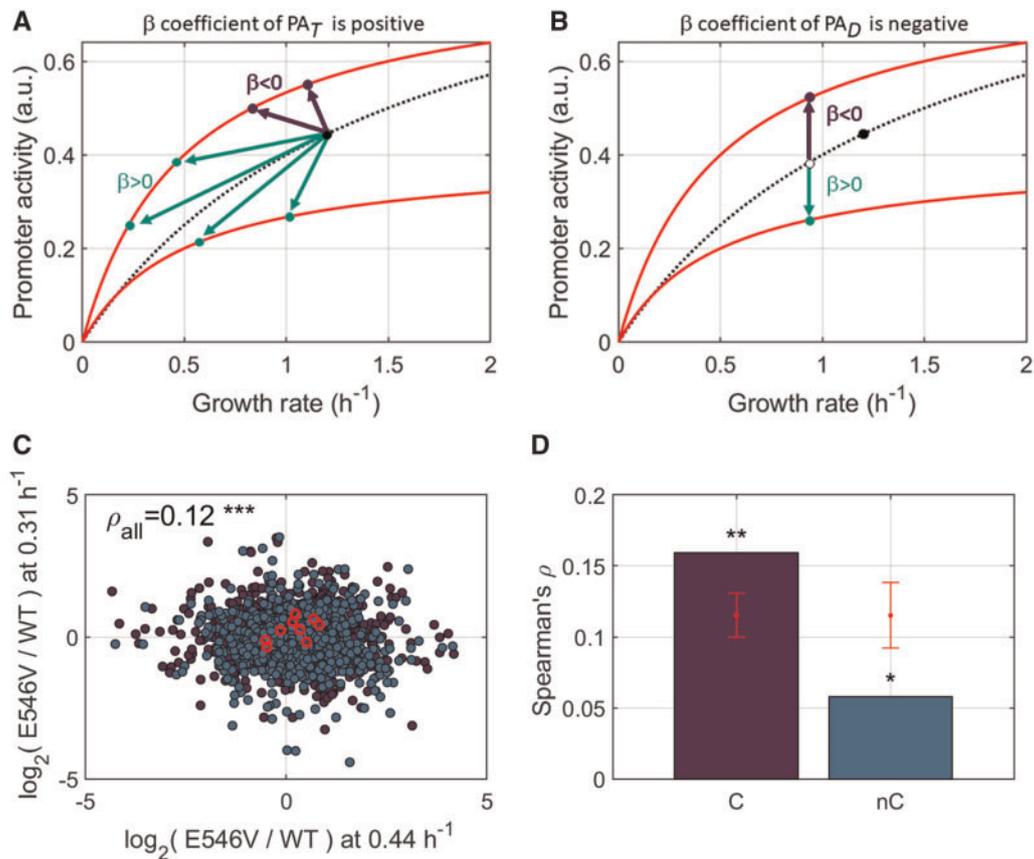
Fig. 6. Insights on the fitness costs phenotypes of RNAP mutants. Impact on PA(μ) of the coefficient sign of PA$_T$ and PA$_D$. In both panels we show the PA(μ) profiles associated to a hypothetical promoter in the WT (black dotted line) and two mutants (red solid lines) to assess the impact of the sign of the coefficient, β, obtained in the statistical models. (A) In the case of PA$_T$, a positive sign of β delimits changes in PA to those similar to the green arrows, where a decrease in growth rate (fitness costs) produce a decrease in PA$_T$. (B) In the case of PA$_D$, the sign of the coefficient is negative. Thus, fitness costs are associated to the over expression of the promoter (purple arrow) rather than under expression (green arrow) from the expected PA of the WT at the growth rate of the mutant (black empty circle). Both results, when combined give a general view of how PA(μ) profiles change due to mutations in *rpoB*. (C) Transcriptomic changes in *rpoB* mutant E546V at different growth rates correlate only slightly (Spearman's $\rho_{all} = 0.12$). Each point corresponds to a gene, and red circles indicate the genes used in this study, data for *hisL* and *serW* are not available. (D) Rank correlations of the subset of constitutive (C, purple) and regulated (nC, blue) genes. Red dots and error bars represent the mean and one standard deviation of a 10^4 permutation test. Constitutive genes correlate significantly more than regulated genes (*$P < 0.05$; **$P < 0.01$, two-tailed).

phenotypic variables appear repeatedly, indicating that they are associated with general mechanisms that, potentially, link the pleiotropic element (RNAP) to fitness.

Insights Revealed by the Analysis of the Phenotypic Effects of Complex Mutations

The previous models are able to anticipate the fitness costs of three *rpoB* mutants in different media from expression data of a particular gene. Interestingly, the coefficients for PA$_D$ and PA$_T$ have opposite signs across all promoters studied (table 1), likely highlighting a general mechanism. On the one hand, a positive coefficient of PA$_T$ implies that mutations in *rpoB* preserve the general shape of PA(μ) profiles as a monotonically increasing function (fig. 6A). On the other hand, a negative coefficient of PA$_D$ highlights that for a fixed growth rate, larger fitness costs are associated with the overexpression of constitutive promoters (fig. 6B). This effect is clear when

observing the PA(μ) profiles of the strain with the largest fitness cost (Q513P in fig. 3C).

Moreover, a comparison of the expression response of a mutant to the WT for a *fixed* growth rate could further confirm constitutive genes as the best reporters of fitness costs. To this aim, we used RNA-seq data of the *rpoB* mutant E546V and its WT ancestor (Utrilla et al. 2016). The transcriptional changes produced by E546V at two different (fixed) growth rates correlate only slightly (Spearman's $\rho = 0.12$; fig. 6C). But most importantly, we found that this correlation greatly originates from the response of constitutive rather than regulated genes (fig. 6D). Should this be a general case, it highlights not only that the transcriptional changes produced by a mutation in *rpoB* are dependent on the growth rate, but also that constitutive genes display a more coordinated response. Consequently, these genes are probable better fitness costs predictors than genes subjected to more specific regulation. In other words, the regulatory network can partially buffer the transcriptional changes produced by the mutant RNAP.

Therefore, our results show that point mutations in *rpoB* alter the transcriptional profiles at a promoter-specific level. Among all, constitutive promoters are likely most sensitive to these global phenotypic effects. Moreover, larger fitness costs are associated to a larger promoter activity than expected when controlling for growth rate, hence reducing the overall efficiency of the transcriptional machinery of the system, something that was previously suspected but that was poorly assessed in a single promoter and growth media (Reynolds 2000; Qi et al. 2014).

Discussion

One encounters three potential problems when characterizing the fitness costs of complex mutations: 1) to define which molecular elements are likely subjects of complex mutations, 2) to recognize which of the molecular features altered by these mutations are driving the costs, and 3) to identify whether some specific target elements (of the molecular agent) can act as a distinctive reporter of such modified features and, in this way, of the costs. We find an answer to the first problem with the use of the environmental fitness cost map, while providing an explicit list of molecular elements that are complex mutations in a metabolic model (supplementary text, Supplementary Material online), and suggesting how to identify them in other systems. To the second by dissecting a set of potential predictors—changes in transcriptional efficiency and interaction with ppGpp—quantified in reporter genes, that are ultimately integrated into a statistical model. By identifying patterns in the models of a variety of genes, this approach also helps us to resolve the third problem: Which targets could be most relevant to predict the fitness costs of mutations.

That we observe complex mutations in a metabolic model supports the idea that they are likely prevalent in regulatory networks and hence, in biological systems. Moreover, we verify that such perturbations are associated with fundamental organismal functions and a larger system-level reprogramming as they are apparent in all environments. The broad reach of these mutations could be connected to pleiotropic effects. Here we find not only that mutations in *E. coli*'s RNAP are complex, but also that their phenotype changes are highly unique to the mutation.

The use of RNAP as an experimental (model) system presents some advantages. First, we can select predictors with clear biological significance. These predictors are related to either the performance of RNAP or its interaction with the alarmone ppGpp. Second, we can test the validity of our approach to earlier discussions on the fitness costs of RNAP mutations. Last, we can consider constitutive genes as an appropriate set of reporters. These genes are valid reporters of direct effects on transcriptional efficiency and indirect ones on cell physiology (see below). Given that the sensitivity to the latter (the global program of transcription) is gene-dependent (Liang et al. 1999; Gerosa et al. 2013; Yubero and Poyatos 2020), we identify some genes within this class that are eventually better predictors than others through the same subset of variables to acceptable levels (but three genes

fail terribly in the task; see supplementary fig. S4, Supplementary Material online for analysis of specific molecular attributes).

We propose that genes that perform better are somehow sensitive to the growth rate. This sensitivity could be read through the changes in a set of features, for example, their expression, as in the following scenarios. First, a gene whose expression is highly robust to a mutation producing fitness costs will likely fail at predicting these costs, as even in the presence of such mutation there will be no observable change in the features. Second, a gene that is disrupted by the presence of the mutation will again be a bad predictor as its expression becomes irrelevant or unreliable. We hypothesize that in between these scenarios, there are a few genes whose predictability is maximal as they are only partially affected by the mutation. We verify this by quantifying the overall effect of a mutation on a gene as the sum of the squared relative change of the predictors included in the statistical framework (supplementary fig. S5, Supplementary Material online).

Specific Implications to the Interplay between Transcriptional Efficiency and Fitness Cost in Rif-Resistant *rpoB* Mutants

Mutations in *rpoB* are most commonly found in antibiotic resistance and adaptive evolution experiments and have been studied extensively due to their implications in tuning fitness. More specifically, mutants producing fitness costs have been traditionally correlated to changes in the transcriptional efficiency of the mutant RNAPs. However, there are several issues with the previous studies.

First, changes in transcriptional efficiency are promoter, environment and (mutant) strain-dependent. A restricted number of any of these variables limits, therefore, the generality of these results. However, alleviating this largely increases the cost and difficulties of such studies. Our data set is a compromise that allows having a broader view of the impact of mutations in *rpoB* on the transcription of different promoters across multiple growth media.

Second, there exists a core dependency between growth rate and gene expression unaccounted for in previous studies. This relationship is most evident in the $PA(\mu)$ profiles of constitutive genes as PA increases together with growth rate (fig. 3) what anticipates a decrease in transcription when cells grow at a reduced rate even in the absence of mutations. Moreover, mutations in *rpoB* directly affect the transcriptional activity of the RNAP producing fitness costs, which in turn, further constrain the efficiency of the RNAP.

For this reason, total changes in PA have a direct contribution to the mutation and what we called an *indirect* influence on the fitness cost. To dissect these effects, one can control for the same growth rate enabling the quantification of changes in PA when WT and mutant strains share an equivalent "physiological state," that is, PA_D. To our knowledge, this is the first quantitative description of how RifR mutations modify the global transcriptional program in general, and $PA(\mu)$ profiles in particular (fig. 3). That we observe the direct effect of mutations upon promoter activity, PA_D, as

an important determinant accentuates the intricate relationship between RNAP activity and fitness. Moreover, in the strain with the most visible fitness costs, there is a significant contribution to changes in PA_T from the limited availability of global resources.

General Implications

All these results show that decoupling the direct effect is fundamental for a better understanding of the transcriptional reprogramming observed in *rpoB* mutants and its impact on fitness costs. A partially similar approach was used to find a decisive shift in two other *rpoB* mutations whose RNAPs prioritize growth over hedging genes (Utrilla et al. 2016). The authors also compare the genome-wide expression between WT and mutants at a constant growth rate to control for a similar physiological state.

This is a particular example of a more general problem in which the target of a mutation and a phenotype are coupled. Conditions in which a phenotypic change is produced not only by a direct perturbation of a molecular agent, but also by the system-level adaptation to such perturbation are widespread. Some of these systems, but not only, can be found in the context of fitness costs produced by antibiotic resistance mutations when such mutations occur in the molecular target of the antibiotic. Indeed, these perturbations potentially result in complex mutations since antibiotics may impede general cellular functions vital for bacterial growth, for example, DNA replication (quinolones), protein synthesis (macrolides), or transcription (rifamycins) as in our work. But this problem also applies to more specific mutations that also cause genome-scale rewiring. Many open questions remain on whether this rewiring is limited by particular genomic mechanisms, for example, the possibility of transcriptional compensation (Kafri et al. 2005; Wong and Roth 2005), and thus signifies no fitness costs, or is eventually deleterious, and consequently involves additional costs (Kovács et al. 2021).

Finally, the result that, from a range of multiple phenotypes, we can distinguish just a few contributing to fitness proposes a model in which extended phenotypic pleiotropy and fitness-relevant modularity coexist. This model was also lately presented by Kinsler et al. (2020). But note that in their work, the authors were unable to identify the precise phenotypes. In our context, we could consider each constitutive gene as a phenotypic element ("phenotypic pleiotropy"), defined by a series of features (comprising our model, supplementary fig. S6, Supplementary Material online). We see in effect that only a few of these elements contribute to fitness ("fitness modularity") and that, interestingly, they do this in some cases through different properties. Thus, fitness as a complex trait is robust to some phenotypic changes. We need to continue studying these issues to finally discern how robust function encoded in cells shapes their response to genetic variation.

Materials and Methods

Computational Models of Complex Mutations

We used the genome-scale metabolic model of *E. coli* iAF1260 (Feist et al. 2007) together with the Cobrapy toolbox (Ebrahim et al. 2013) to compute the fitness of the WT and mutants in an array of media. We simulated mutations on an enzyme by imposing a limit in the flux of reactions in which it participates. In the knockdown example, *nuoB*, we took into consideration the logic of reaction rules. The limit is a fraction of the maximum flux observed across all media in the WT strain and it is fixed for a given mutant during growth in any media. We used minimal media supplemented with one of the 174 carbon sources found in the original study that support growth (Feist et al. 2007). The exchange rate for any carbon source was set equal to that of glucose (8 mmol/gDW/h). We compute the relative global fitness as the slope of the robust least-squares fit (bisquare method) of the fitness of the mutant relative to the WT (fig. 1B). The bisquare method iteratively weighs data points as a function of their distance to the fitted line. In this way, it is able to recognize outliers (specific GxE interactions) preventing them to negatively affect the final fit to the core trend (relative global fitness). Then, we classify outliers as having residuals more than three SMAD (scaled median absolute deviations) away from the median. Data points where the mutant is lethal are excluded from all fits. We also used the tool Escher to produce figure 1D (King et al. 2015).

Strains and Growth Conditions

We used *E. coli* Rel606 as WT, and three mutant derivatives with the following amino acid substitutions in the gene *rpoB*; H526L, S512Y, and Q513P obtained previously through rifampicin resistance. In general, strains were retrieved from −80 °C frozen stocks, plated in agar plates with selective media (when necessary), and grown overnight at 37 °C Reporter plasmids were extracted from a library (Zaslaver et al. 2006) and purified with the Qiagen Mini-prep kit following the manufacturer's protocol. Then, each strain was transformed with each reporter plasmid with TSS (Chung et al. 1989). When necessary, selective media for *rpoB* mutants was prepared with rifampicin (100 μg/ml), and for plasmid-bearing strains with kanamycin (50 μg/ml). Both antibiotics were used simultaneously when selecting *rpoB* mutants bearing the fluorescent reporter plasmid. All bacterial growth was at 30 °C unless otherwise specified. Also, cultures were grown under the shade to prevent rifampicin degradation.

Growth media consisted of M9 minimal media supplemented 1) with one of the following carbon sources at 0.5%(w/v): Glycerol, sucrose, fructose, and glucose, and 2) either with or without amino acids to a final concentration of 0.2% (w/v), thus making eight different nutrient conditions in total.

Single colonies were pre-cultured in 1 ml of M9 minimal media supplemented with glucose at 0.5% (w/v) for 3 h. Then, 96-well flat-bottom plates filled with the corresponding media were inoculated with 20 μl of preculture to a final volume of 220 μl, we then added 30 μl of mineral oil to prevent evaporation. Optical density at 600 nm, and fluorescence 490/535 nm when appropriate, were assayed in a Victor X2 (Perkin Elmer) at 5 min intervals with orbital shaking (30 s, 1 mm) for more than 12 h.

Data Processing and Promoter Activity Modeling

First, OD and GFP measurements were corrected for background levels by subtracting the value of blank wells filled with each corresponding growth media. GFP measurements were further corrected by subtracting the autofluorescence produced during the growth of the corresponding strain transformed with the pUA66 promoterless plasmid (Zaslaver et al. 2006). Only then, growth rate time series were computed as the two-point finite differences of $\log_2(OD)$, $\mu(t) = \Delta \log_2(OD)/\Delta t$ (in doublings per hour), and promoter activities were computed as the two-point finite difference in time of fluorescence per OD unit, $PA_{pl}(t) = \Delta GFP/\Delta t/OD$ (in units of GFP/OD/h). Balanced-growth data was computed from the mean time-series measurements of three technical replicates as the average value in a 1 h time-window during observable exponential growth.

Promoter activity dependence on growth rate was modeled with a Michaelis–Menten equation as $PA(\mu) = V_m \mu/(K_m + \mu)$ where V_m is the maximum promoter activity and K_m is the growth rate at which PA is half-maximal (Liang et al. 1999). Data from balanced growth were fit to this equation through robust least squares (bisquare) with an upper K_m limit of 3 dbl/h to avoid overfitting linear profiles (Yubero and Poyatos 2020).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Data Availability

Data and Code of Predicting the fitness cost of complex mutations [Data set] available at Zenodo: http://doi.org/10.5281/zenodo.4596721.

References

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461(7268):1243–1247.

Cai X-C, Xi H, Liang L, Liu J-D, Liu C-H, Xue Y-R, Yu X-Y. 2017. Rifampicin-resistance mutations in the *rpoB* gene in *Bacillus velezensis* CC09 have pleiotropic effects. *Front Microbiol*. 8:178.

Chung CT, Niemela SL, Miller RH. 1989. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A*. 86(7):2172–2175.

Dudley AM, Janse DM, Tanay A, Shamir R, Church GM. 2005. A global view of pleiotropy and phenotypically derived gene function in yeast. *Mol Syst Biol*. 1(1):2005.0001.

Ebrahim A, Lerman JA, Palsson BØ, Hyduke DR. 2013. Cobrapy: constraints-based reconstruction and analysis for python. *BMC Syst Biol*. 7(1):74.

Falconer D, Mackay T. 1996. Introduction to quantitative genetics. Harlow (UK): Prentice-Hall.

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ. 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol*. 3(1):121.

Garibyan L, Huang T, Kim M, Wolff E, Nguyen A, Nguyen T, Diep A, Hu K, Iverson A, Yang H, et al. 2003. Use of the *rpoB* gene to determine the specificity of base substitution mutations on the *Escherichia coli* chromosome. *DNA Repair (Amst)*. 2(5):593–608.

Gerosa L, Kochanowski K, Heinemann M, Sauer U. 2013. Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol Syst Biol*. 9(1):658.

Goldstein BP. 2014. Resistance to rifampicin: a review. *J Antibiot (Tokyo)* 67(9):625–630.

Griffiths AJ, Wessler SR, Carrol SB, Doebley J. 2015. Introduction to genetic analysis. 11th ed. New York: W. H. Freeman & Co.

He X, Zhang J. 2006. Toward a molecular understanding of pleiotropy. *Genetics* 173(4):1885–1891.

Hirsch M, Elliott T. 2002. Role of ppGpp in *rpoS* stationary-phase regulation in *Escherichia coli*. *J Bacteriol*. 184(18):5077–5087.

Irving SE, Corrigan RM. 2018. Triggering the stringent response: signals responsible for activating (p)ppGpp synthesis in bacteria. *Microbiology (Reading)* 164(3):268–276.

Jin DJ, Gross CA. 1988. Mapping and sequencing of mutations in the *Escherichia coli rpoB* gene that lead to rifampicin resistance. *J Mol Biol*. 202(1):45–58.

Jin DJ, Gross CA. 1989. Characterization of the pleiotropic phenotypes of rifampin-resistant *rpoB* mutants of *Escherichia coli*. *J Bacteriol*. 171(9):5229–5231.

Johnson JB, Omland KS. 2004. Model selection in ecology and evolution. *Trends Ecol Evol*. 19(2):101–108.

Kafri R, Bar-Even A, Pilpel Y. 2005. Transcription control reprogramming in genetic backup circuits. *Nat Genet*. 37(3):295–299.

Karthik M, Meenakshi S, Munavar M. 2019. Unveiling the molecular basis for pleiotropy in selected rif mutants of *Escherichia coli*: possible role for tyrosine in the rif binding pocket and fast movement of RNA polymerase. *Gene* 713:143951.

King ZA, Dräger A, Ebrahim A, Sonnenschein N, Lewis NE, Palsson BØ. 2015. Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLOS Comput Biol*. 11(8):e1004321.

Kinsler G, Geiler-Samerotte K, Petrov DA. 2020. Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *eLife* 9:e61271.

Klumpp S, Hwa T. 2008. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc Natl Acad Sci U S A*. 105(51):20245–20250.

Klumpp S, Zhang Z, Hwa T. 2009. Growth rate-dependent global effects on gene expression in bacteria. *Cell* 139(7):1366–1375.

Kovács K, Farkas Z, Bajić D, Kalapis D, Daraba A, Almási K, Kintses B, Bódi Z, Notebaart RA, Poyatos JF, et al. 2021. Suboptimal global transcriptional response increases the harmful effects of loss-of-function mutations. *Mol Biol Evol*. 38(3):1137–1150.

Liang S-T, Bipatnath M, Xu Y-C, Chen S-L, Dennis P, Ehrenberg M, Bremer H. 1999. Activities of constitutive promoters in *Escherichia coli*. *J Mol Biol*. 292(1):19–37.

Maaløe O. 1979. Regulation of the protein-synthesizing machinery—ribosomes, tRNA, factors, and so on. Boston (MA): Springer US, p. 487–542.

Paul BJ, Barker MM, Ross W, Schneider DA, Webb C, Foster JW, Gourse RL. 2004. DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* 118(3):311–322.

Potrykus K, Murphy H, Philippe N, Cashel M. 2011. ppGpp is the major source of growth rate control in *E. coli*. *Environ Microbiol*. 13(3):563–575.

Qi Q, Preston GM, MacLean RC. 2014. Linking system-wide impacts of RNA polymerase mutations to the fitness cost of rifampin resistance in *Pseudomonas aeruginosa*. *mBio* 5(6):e01562–14.

Reynolds MG. 2000. Compensatory evolution in rifampin-resistant *Escherichia coli*. *Genetics* 156(4):1471–1481.

Sanchez-Vazquez P, Dewey CN, Kitten N, Ross W, Gourse RL. 2019. Genome-wide effects on *Escherichia coli* transcription from ppGpp binding to its two sites on RNA polymerase. *Proc Natl Acad Sci U S A.* 116(17):8310–8319.

Schaechter M, Maaløe O, Kjeldgaard N. 1958. Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *J Gen Microbiol.* 19(3):592–606.

Tóth I, Csík M, Emódy L. 2003. Spontaneous antibiotic resistance mutation associated pleiotropic changes in *Escherichia coli* O157: h 7. *Acta Vet Hung.* 51(1):29–44.

Trauner A, Banaei-Esfahani A, Gygli SM, Warmer P, Feldmann J, Shafieechashmi S, Eschbach K, Zampieri M, Borrell S, Collins BC, et al. 2018. Resource misallocation as a mediator of fitness costs in antibiotic resistance. *bioRxiv*.

Utrilla J, O'Brien EJ, Chen K, McCloskey D, Cheung J, Wang H, Armenta-Medina D, Feist AM, Palsson BO. 2016. Global rebalancing of cellular resources by pleiotropic point mutations illustrates a multi-scale mechanism of adaptive evolution. *Cell Syst.* 2(4):260–271.

Wong SL, Roth FP. 2005. Transcriptional compensation for gene loss plays a minor role in maintaining genetic robustness in *Saccharomyces cerevisiae*. *Genetics* 171(2):829–833.

Wytock TP, Zhang M, Jinich A, Fiebig A, Crosson S, Motter AE. 2020. Extreme antagonism arising from gene-environment interactions. *Biophys J.* 119(10):2074–2086.

Yubero P, Poyatos JF. 2020. The impact of global transcriptional regulation on bacterial gene order. *iScience* 23(4):101029.

Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, Liebermeister W, Surette MG, Alon U. 2006. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat Methods* 3(8):623–628.

Zhou YN, Jin DJ. 1998. The *rpoB* mutants destabilizing initiation complexes at stringently controlled promoters behave like "stringent" RNA polymerases in *Escherichia coli*. *Proc Natl Acad Sci U S A.* 95(6):2908–2913.

Zhu M, Dai X. 2019. Growth suppression by altered (p)ppGpp levels results from non-optimal resource allocation in *Escherichia coli*. *Nucleic Acids Res.* 47(9):4684–4693.