

Universal Constraints on Protein Evolution in the Long-Term Evolution Experiment with *Escherichia coli*

Rohan Maddamsetti *

Department of Biomedical Engineering, Duke University, Durham, North Carolina, USA

*Corresponding author: E-mail: rohan.maddamsetti@duke.edu.

Accepted: 30 March 2021

Abstract

Although it is well known that abundant proteins evolve slowly across the tree of life, there is little consensus for why this is true. Here, I report that abundant proteins evolve slowly in the hypermutator populations of Lenski's long-term evolution experiment with *Escherichia coli* (LTEE). Specifically, the density of all observed mutations per gene, as measured in metagenomic time series covering 60,000 generations of the LTEE, significantly anticorrelates with mRNA abundance, protein abundance, and degree of protein–protein interaction. The same pattern holds for nonsynonymous mutation density. However, synonymous mutation density, measured across the LTEE hypermutator populations, positively correlates with protein abundance. These results show that universal constraints on protein evolution are visible in data spanning three decades of experimental evolution. Therefore, it should be possible to design experiments to answer why abundant proteins evolve slowly.

Key words: purifying selection, experimental evolution, protein evolution.

Significance

A universal evolutionary pattern is that highly abundant and highly interacting proteins evolve slowly. This pattern was discovered in analyses that cover millions of years' worth of sequence variation, so it is not clear how long it takes (decades, centuries, millennia) for such patterns to emerge. Here, I report that this universal evolutionary pattern emerges in metagenomic data that cover just 30 years of experimental evolution.

Introduction

One consequence of the high complexity and intricate functional organization of organisms is that most mutations are deleterious. Natural selection resists the loss of function and fitness caused by mutation accumulation over time (Leiby and Marx 2014; LaBar and Adami 2017; Grant et al. 2021). This process, called purifying selection, maintains the complexity and functional integrity of evolved organisms.

Despite its importance, purifying selection has been little studied in experimental systems (Alvarez-Ponce et al. 2016), in contrast to adaptive evolution (Barrick and Lenski 2013). In two recent papers, my colleagues and I reported evidence for purifying selection in metagenomic time series of Lenski's long-term evolution experiment with *Escherichia coli*, often called the LTEE for short (Lenski et al. 1991; Good et al. 2017). We considered the molecular evolution of the six

hypermutator LTEE populations, which have elevated mutation rates due to evolved defects in DNA repair (Tenaillon et al. 2016; Maddamsetti and Grant 2020a). These populations continue to increase in fitness due to adaptive evolution, even though genome evolution in these populations largely reflects the accumulation of nearly neutral mutations (Couce et al. 2017). In Grant et al. (2021), we reported evidence for purifying selection on aerobic- and anaerobic-specific genes in *E. coli*. In Maddamsetti and Grant (2020b), we then reported evidence for purifying selection on genes that were found to be essential in the ancestral LTEE strain, REL606, in a transposon mutagenesis screen (Couce et al. 2017).

Here, I report evidence that purifying selection in the LTEE reflects a universal constraint on protein evolution found across the tree of life, namely that highly abundant and highly interacting proteins evolve slowly (Fraser et al. 2002; Hahn

et al. 2004; Drummond et al. 2005; Hahn and Kern 2005; Drummond and Wilke 2008; Alvarez-Ponce et al. 2017). Despite the universality and simplicity of this pattern of purifying selection, its proximate causes continue to be debated (Plata et al. 2010; Plata and Vitkup 2018; Razban 2019; Usmanova et al. 2021). A number of compelling hypotheses have been proposed, but consensus has not been reached. The findings reported here will not settle this debate. Nonetheless, an important consequence of my findings is that it may be possible to resolve the causes of this universal pattern by experimental means.

Results

Rationale and Study Design

This study takes a novel approach to study the anticorrelation between protein abundance and evolutionary rates (Pál et al. 2001; Drummond et al. 2005, 2006; Drummond and Wilke 2008; Lobkovsky et al. 2010; Yang et al. 2010; Wylie and Shakhnovich 2011; Serohijos et al. 2012; Serohijos and Shakhnovich 2014). In this section, I present the logical structure of the hypotheses and predictions under consideration and explain the methods that I use (fig. 1).

I assume that the mutation rates in the hypermutator LTEE populations are high enough that the vast majority of observed mutations are nearly neutral hitchhikers, whose dynamics are driven by a relatively small number of highly beneficial mutations (Barrick and Lenski 2009; Levy et al. 2015; Maddamsetti, Lenski, et al. 2015; Tenaillon et al. 2016; Couce et al. 2017; Good et al. 2017; Ba et al. 2019; Maddamsetti and Grant 2020a). This allows us to infer information about mutation rates and biases (Couce et al. 2017; Maddamsetti and Grant 2020a) even under environmental and population-genetic conditions that favor strong positive selection. It follows that the mutations observed across the nonmutator and hypermutator LTEE populations, to a large extent, reflect different parts of the distribution of mutation fitness effects (DFE) per gene.

With this assumption in hand, I start from the hypothesis that purifying selection causes abundant proteins to evolve slowly. This means that the DFE for abundant proteins should contain more deleterious mutations than the DFE for less abundant proteins, all else being equal. It follows that highly abundant proteins should have fewer observed mutations in the hypermutator LTEE populations, because it is unlikely that highly deleterious mutations will reach observable allele frequencies in the LTEE, given the population-genetic conditions of the LTEE (Good et al. 2017). This is the logical basis for using the hypermutator LTEE populations to test for purifying selection on abundant proteins.

The key technical trick is that we do not need to calculate evolutionary rates for the LTEE—in fact, we can completely ignore the phylogenetic structure of each population. Instead,

we only need to count the number of observed mutations per gene across all hypermutator populations, and normalize by gene length (fig. 1). An additional benefit of this approach is that the effects of clonal interference and frequency-dependent selection (Maddamsetti, Lenski, et al. 2015; Good et al. 2017) can be ignored, because these phenomena do not affect the density of mutations that are ever observed in the LTEE. By contrast, clonal interference and frequency-dependent selection may have significant effects on evolutionary rates (Lang et al. 2013; Serohijos and Shakhnovich 2014; Maddamsetti, Lenski, et al. 2015; Good et al. 2017).

The great advantage of the LTEE, and other evolution experiments with microbes, is the “fossil record” of frozen samples that can be revived for comparison with later samples. The vast majority of mutations in the LTEE lie off the line of descent, but are still accessible from sequencing those frozen population samples (Good et al. 2017). By contrast, analyses of natural sequence data are largely restricted to extant within-population polymorphism and between-species fixations. The use of mutations off the line of descent in the LTEE, along with its multidecade duration, provides sufficient (and ever increasing) statistical power to discern patterns of purifying selection, such as the one discussed in this work.

Correlations between mRNA and Protein Abundance and Mutation Density per Gene in LTEE Populations

I compared the density of observed mutations in the LTEE (Good et al. 2017) with mRNA and protein abundance data for the LTEE ancestral strain, REL606, grown in DM500 media (Caglar et al. 2017). These comparisons are shown in figure 2; note that throughout this section, all significant Spearman correlation coefficients and associated *P* values are labeled on the figures. In the hypermutator LTEE populations, mRNA abundance during exponential growth significantly anticorrelates with mutation density, whereas protein abundance, at all time points, significantly anticorrelates with mutation density. The same anticorrelation holds, for all time points, when only nonsynonymous (i.e., missense and nonsense) mutations are considered (fig. 3). The significance of these anticorrelations increases when genes with no observed mutations in the metagenomic data are excluded (supplementary fig. S1, Supplementary Material online, for all mutation types; supplementary fig. S2, Supplementary Material online, for nonsynonymous mutations). By contrast, the density of synonymous mutations across the hypermutator populations shows a significant positive correlation with mRNA and protein abundance for REL606 in DM500 media, across all phases of growth (fig. 4). When genes with no mutations in the metagenomic data are excluded, significant positive correlations remain between synonymous mutation density and mRNA and protein abundance, although to a lesser degree (supplementary fig. S3, Supplementary Material online). In the nonmutator LTEE populations, both mRNA and protein

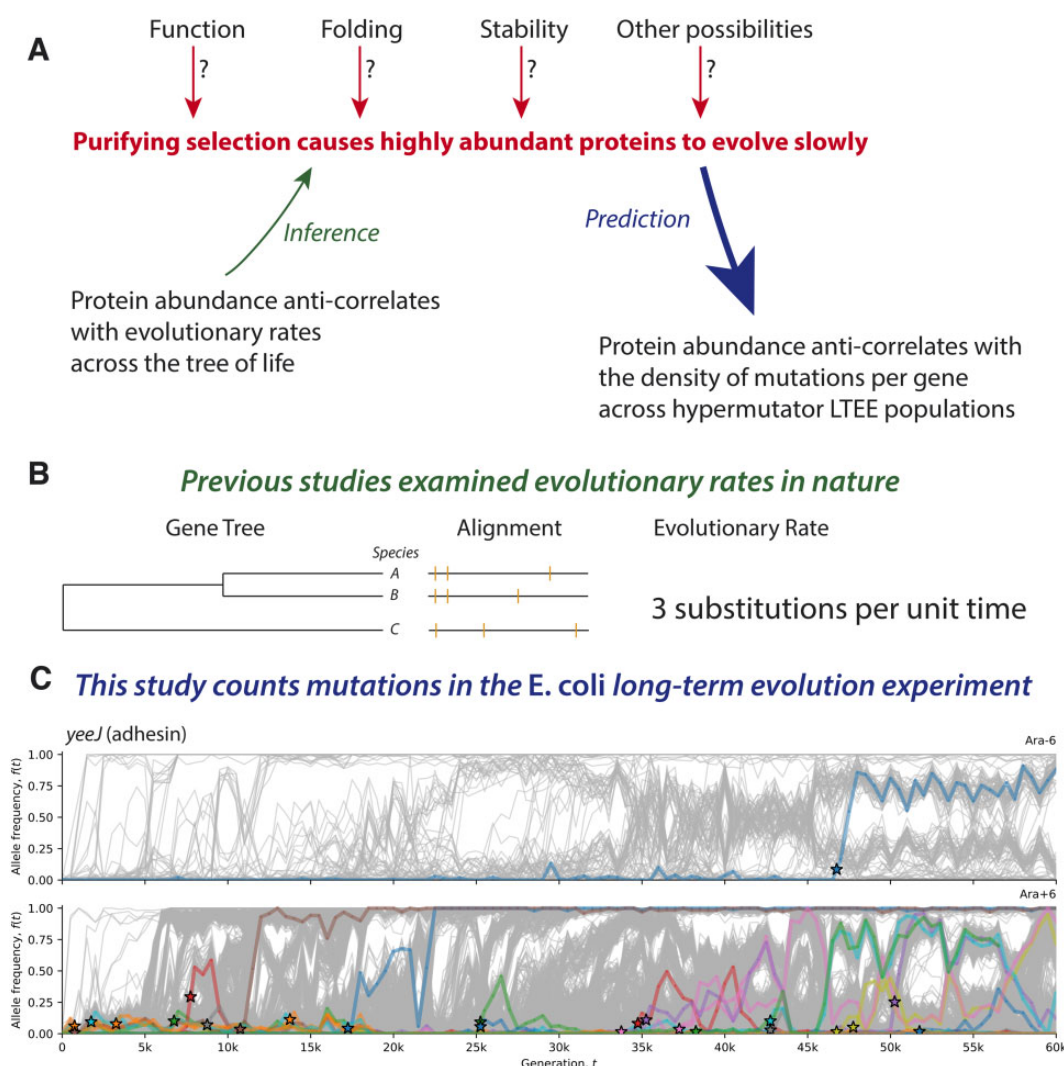


FIG. 1.—Study design. (A) Many studies have reported that highly abundant proteins evolve slowly. If this fact is caused by purifying selection, then mutations in highly abundant proteins, should be more deleterious than mutations in less abundant proteins, on average. This logic leads to the prediction that highly abundant proteins should have fewer observed mutations than less abundant proteins across the hypermutator populations of the LTEE, taking gene length into account. (B) Previous studies inferred evolutionary rates using DNA and protein sequence comparisons across species. (C) This study sums all observed mutations per gene in metagenomic time series of the long-term evolution experiment with *Escherichia coli* (LTEE), considering nonmutator and hypermutator populations separately. This approach increases statistical power over a rate-based approach and is affected by neither clonal interference nor frequency-dependent selection. To give a concrete example, the top panel in (C) shows the number of observed mutations (stars) in the adhesin gene *yeeJ* in the nonmutator population Ara – 6 over 60,000 generations. The bottom panel in (C) shows the number of observed mutations (stars) in *yeeJ* in the hypermutator population Ara + 6 over the same period. For comparison across genes, the number of observed mutations is normalized by gene length.

abundance for REL606 grown in DM500 show significant positive correlations with the density of observed mutations (supplementary fig. S4, Supplementary Material online).

I also asked whether the strength of the Spearman correlations between protein abundance and mutation density in the hypermutator populations increased over the course of the LTEE (fig. 5). In analyses of natural sequence variation, it is understood that the strength of anticorrelation between protein evolutionary rates and protein abundance increases with divergence time among the taxa under consideration

(Serohijos et al. 2012). Based on protein biophysics, Serohijos et al. (2012) additionally predicted that the strength of the anticorrelation between evolutionary rates and protein abundance would increase, but at declining rates over time. Even though the differences in measurements, units, and timescales make direct comparisons to those theoretical predictions impossible, it is striking that a similar functional form of the relationship between time and the strength of the rate-abundance anticorrelation occurs with the mutations observed across the LTEE hypermutator populations (fig. 5A

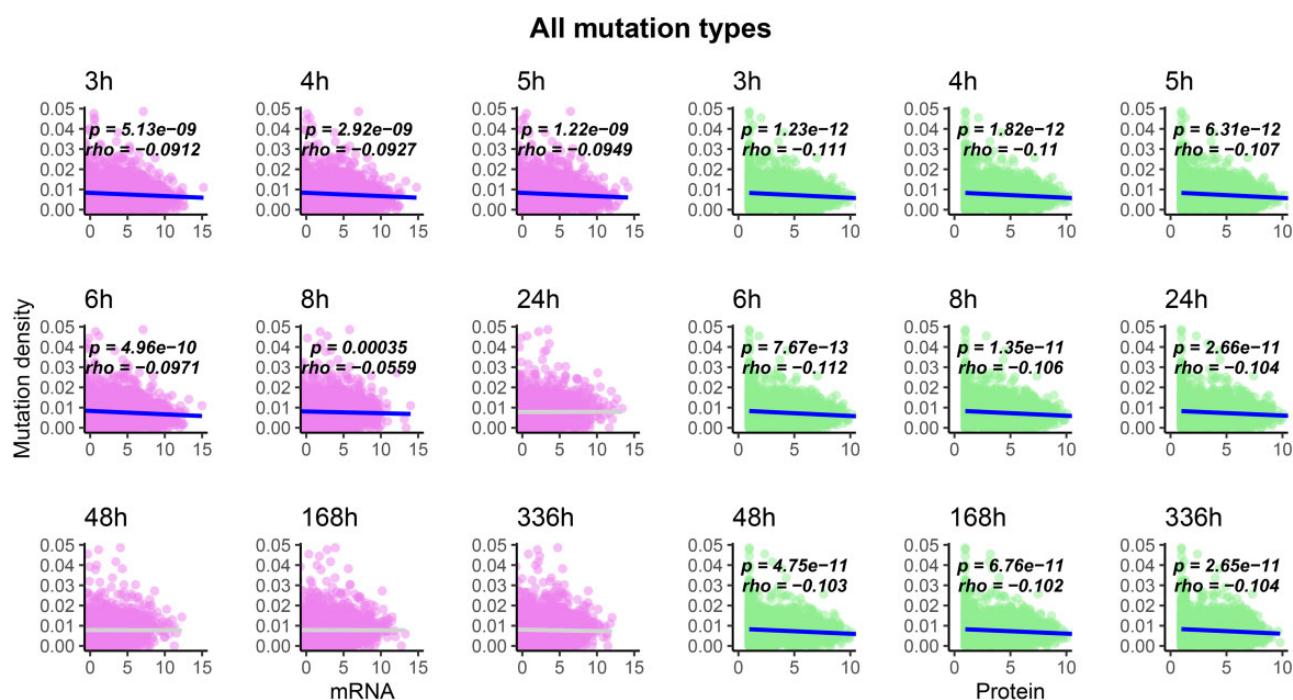


FIG. 2.—The density of observed mutations per gene across all hypermutator LTEE populations anticorrelates with mRNA abundance in exponential growth phase, and anticorrelates with protein abundance at all time points. RNA and protein abundance were measured for the ancestral LTEE clone REL606, grown in DM500 media (Caglar et al. 2017). Each point represents a protein-coding gene in the genome of the ancestral LTEE clone, *Escherichia coli* B strain REL606. The abundance of mRNA or protein expressed per gene is shown on the x axis of each plot. The density of observed mutations per gene is shown on the y axis of each plot. Comparisons to mRNA abundance are shown in purple, whereas comparisons to protein abundance are shown in green. Statistically significant correlations are shown in blue, whereas nonsignificant correlations are shown in light gray. Spearman correlation coefficients (ρ) and associated P values are shown on each panel.

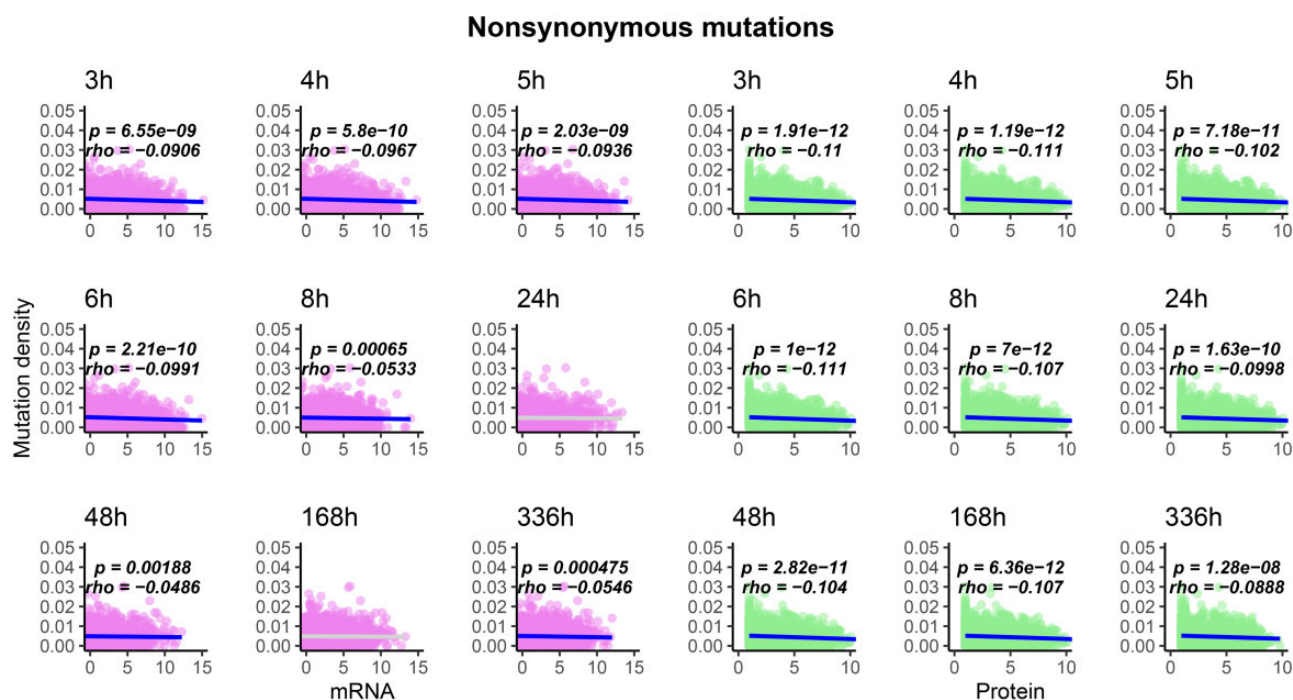


FIG. 3.—The density of observed nonsynonymous mutations per gene across all hypermutator LTEE populations anticorrelates with mRNA abundance in exponential growth phase, and anticorrelates with protein abundance at all time points (see fig. 2 legend for further details).

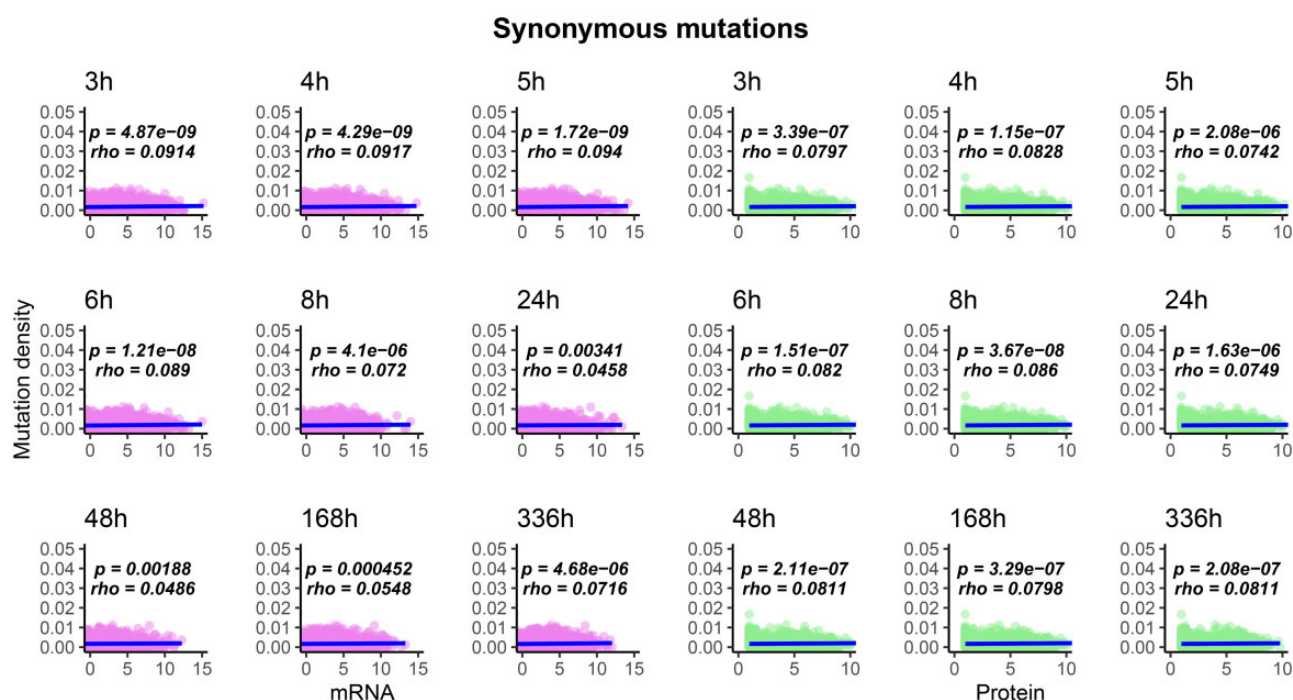


Fig. 4.—The density of observed synonymous mutations per gene across all hypermutator LTEE populations positively correlates with mRNA and protein abundance at all time points (see fig. 2 legend for further details).

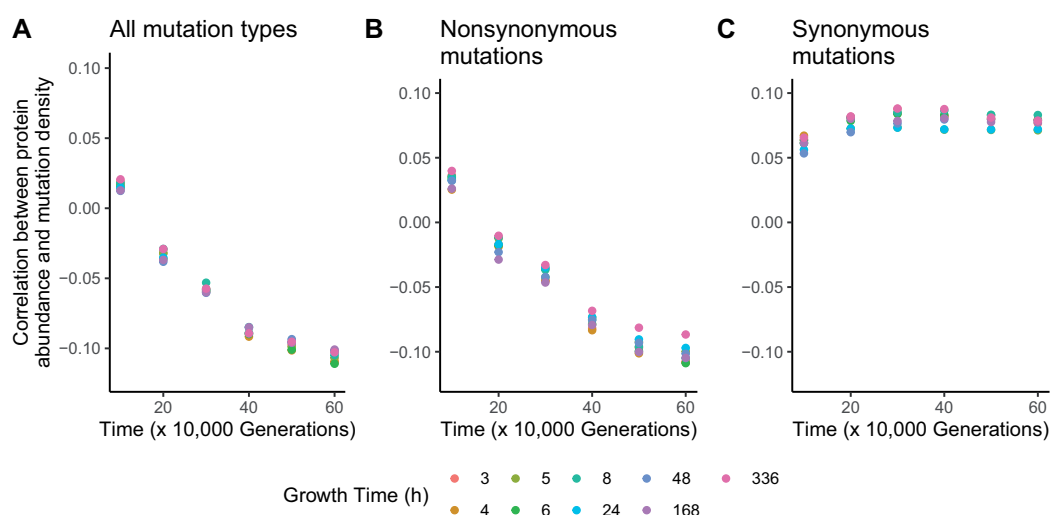


Fig. 5.—Correlations between protein abundance in REL606 and mutation density are consistent over time. Points represent Spearman correlation coefficients, calculated using the cumulative number of mutations observed by each 10,000-generation mark in the metagenomic time series for the LTEE hypermutator populations. Colors indicate the growth time at which protein abundance was sampled for REL606; the growth times correspond to the separate panels in figures 2–4. (A) Correlations between protein abundance in REL606 and mutation density across all hypermutator LTEE populations. (B) Correlations between protein abundance in REL606 and nonsynonymous mutation density across all hypermutator LTEE populations. (C) Correlations between protein abundance in REL606 and synonymous mutation density across all hypermutator LTEE populations.

and B). By contrast, the positive Spearman correlation coefficient between synonymous mutation density and protein abundance remains steady at ~ 0.075 for at least 40,000 generations, ranging from the 20,000-generation mark through 60,000 generations (fig. 5C).

A limitation of these analyses is that these RNA and protein abundance data come from the ancestral LTEE clone, REL606, and so these patterns may not hold for evolved strains. To address this limitation, I examined RNA abundance data for eleven 50,000 generation LTEE clones, grown to exponential

phase in DM4000 media (Favate et al. 2021). In every single case, the density of observed mutations per gene, measured across all hypermutator populations, significantly anticorrelates with mRNA abundance (supplementary fig. S5, Supplementary Material online). In addition, a significant anticorrelation is seen with nonsynonymous mutations for all 11 clones (supplementary fig. S6, Supplementary Material online), whereas a positive correlation is seen with synonymous mutations, again for all 11 evolved clones (supplementary fig. S7, Supplementary Material online). The density of observed mutations per gene in the nonmutator populations significantly correlates with mRNA abundance in seven out of 11 clones (supplementary fig. S8, Supplementary Material online).

As an additional check for the robustness of these correlations, I compared the density of observed mutations per gene in the LTEE with protein abundance data in the ProteomeVis database (Razban et al. 2018). Although these data only cover 664 out of 4,205 genes analyzed in the LTEE metagenomic data, they still reveal significant anticorrelations between mutation density per gene in the hypermutator populations and protein abundance, when all mutations and nonsynonymous mutations are analyzed (supplementary fig. S9, Supplementary Material online). Corresponding results for synonymous mutations in the hypermutator LTEE populations, and for all mutation types in the nonmutator LTEE populations, are not statistically significant.

Highly Interacting Proteins Evolve Slowly in Hypermutator Populations

Another universal pattern is that highly interacting proteins evolve more slowly than those with fewer interaction partners (Fraser et al. 2002; Hahn et al. 2004; Hahn and Kern 2005; Alvarez-Ponce et al. 2017). I hypothesized that highly interacting proteins would be under strong selection in the LTEE, based on those reports, as well as previous results showing that the *E. coli* core genome is under positive selection in the LTEE (Maddamsetti et al. 2017), and that global regulators of gene expression show evidence of strong positive selection in both nonmutator and hypermutator LTEE populations (Maddamsetti and Grant 2020b). In particular, I hypothesized that highly interacting proteins should evolve rapidly in the nonmutator LTEE populations due to positive selection, but should evolve slowly in the hypermutator populations during purifying selection.

I compared the number of protein–protein interactions (PPI) with the density of observed mutations across LTEE populations for every protein-coding gene in the *E. coli* genome, using three curated data sets of PPI in *E. coli* (Razban et al. 2018; Cong et al. 2019; Zitnik et al. 2019), which I refer to as the Cong data set, the Zitnik data set, and the Razban data set. These comparisons are shown in figure 6 and supplementary figure 10, Supplementary Material online. I find

significant negative correlations between mutation density and PPI degree in the hypermutators (Spearman's $\rho = -0.056$, $P = 0.00037$ for Cong data set; Spearman's $\rho = -0.11$, $P < 10^{-11}$ for Zitnik data set; Spearman's $\rho = -0.068$, $P < 10^{-4}$ for Razban data set). However, the weak positive correlations between mutation density and PPI degree in the nonmutators are not significant (supplementary fig. 10, Supplementary Material online).

Discussion

I show that a number of well-known but poorly understood correlations between mRNA abundance, protein abundance, PPI degree, and evolutionary rates across the tree of life are also found in the hypermutator populations of the LTEE. In some cases, I find significant anticorrelation between mutation densities and mRNA abundance in exponential phase, but not during stationary phase. The simplest explanation for this finding is that mRNAs decay more rapidly than the proteins they encode. Protein abundance consistently shows a negative correlation with the density of all observed mutations (fig. 2) and with nonsynonymous mutation density across all time points (fig. 3).

It is widely believed that these correlations are driven by purifying selection on universal aspects of protein evolution (Drummond et al. 2006; Drummond and Wilke 2009; Serohijos et al. 2012; Serohijos and Shakhnovich 2014), and indeed, this is the most parsimonious explanation for why similar patterns are seen in the LTEE. An intriguing difference, however, is the positive correlation that I find between synonymous mutation density across LTEE hypermutator populations and protein abundance (fig. 4)—which contrasts with the anticorrelation between the rate of synonymous mutations and gene expression seen in nature (Drummond and Wilke 2008). In part, this may be explained by the differences in the distribution of synonymous mutations observed in the LTEE, and the distribution of synonymous diversity per gene in nature (Maddamsetti, Hatcher, et al. 2015; Maddamsetti and Grant 2020a), although the causes for this difference between natural variation and experiment is still a matter for hypothesis generation (Maddamsetti 2016), data collection, hypothesis testing, and debate.

An important limitation of these results is that the protein and mRNA abundance data for LTEE strains were collected in DM500 and DM4000 media (Caglar et al. 2017; Favate et al. 2021). These media contain much more than the 25 mg/l glucose in the DM25 media used in the LTEE. This represents a technical compromise due to the fact that researchers have not yet succeeded in isolating sufficient mRNA from exponential phase cultures in DM25 for RNA-seq (Jagdish T and Grant N, personal communication). With this caveat in mind, my findings support the conclusion that highly abundant proteins evolve slowly in the hypermutator LTEE populations.

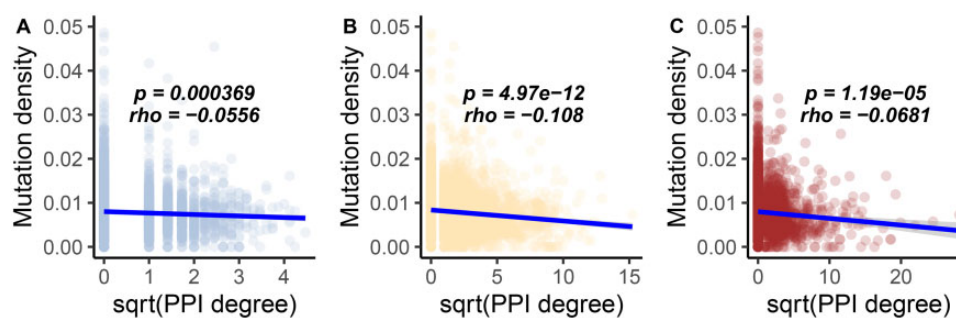


FIG. 6.—The density of observed mutations per gene across all hypermutator LTEE populations negatively correlates with PPI degree. Comparisons to the PPI data from Cong et al. (2019) are shown in light blue, comparisons to the PPI data from Zitnik et al. (2019) are shown in orange, and comparisons to the PPI data in the ProteomeVis database (Razban et al. 2018) are shown in red. Significant Spearman correlations are shown in blue. For improved visual dispersion, PPI degree is square-root transformed; the Spearman correlation is unaffected by this monotonic data transformation. (A) Proteins with more interactions in the Cong et al. (2019) data set tend to evolve more slowly than those with fewer interactions in the hypermutator LTEE populations. (B) Proteins with more interactions in the Zitnik et al. (2019) data set tend to evolve more slowly than those with fewer interactions in the hypermutator LTEE populations. (C) Proteins with more interactions in the ProteomeVis *Escherichia coli* PPI data set (Razban et al. 2018) tend to evolve more slowly than those with fewer interactions in the hypermutator LTEE populations.

The causes for why highly abundant proteins evolve slowly may emerge from a number of different, and nonmutually exclusive phenomena, so many explanations have been proposed (Razban 2019). These include the protein misfolding avoidance hypothesis (Yang et al. 2010), the protein misinteraction avoidance hypothesis (Levy et al. 2012; Yang et al. 2012), the mRNA folding hypothesis (Park et al. 2013), purifying selection on protein function (Konaté et al. 2019), folding stability (Serohijos et al. 2012; Serohijos and Shakhnovich 2014), and others (Tartaglia et al. 2007; Plata et al. 2010; Kepp and Dasmeh 2014).

Differentiating among these possibilities is difficult, because it is challenging to study the *causes* of patterns that span millions of years of protein evolution. I do not draw conclusions about the causes of these correlations. Rather, my results show that evolution experiments are reasonable model systems to study the causes of evolutionary rate variation in proteins. A concrete approach would be to recode the genomes of hypermutator strains to modulate the anticipated action of purifying selection per protein, based on the predictions of a particular explanation, and then ask whether those predictions are borne out during experimental evolution. Breakthroughs that allow for the inexpensive recoding of whole bacterial genomes may be needed, but it is plausible that such experiments will be feasible in the future.

Many other experimental directions are possible. First, a better understanding of how chaperones and other molecular mechanisms of protein quality control affect evolutionary rates and fitness (Chen et al. 2017; Alvarez-Ponce et al. 2019; Samhita et al. 2020) is needed. We also need to better understand purifying selection on synonymous mutations (Walsh et al. 2020). Second, studies on how RNA transcription error rates (Li and Lynch 2020) and RNA folding errors affect evolutionary rates would be valuable. Indeed, mRNA accessibility seems to be an important predictor of protein abundance

(Terai and Asai 2020)—and RNA chaperones buffer deleterious mutations in LTEE hypermutator strains (Rudan et al. 2015). Third, it would be interesting to experimentally test the hypothesis that protein and RNA chaperones evolve under more and more stringent purifying selection during long-term experimental evolution, which follows from the premise that hypermutator LTEE populations are affected by a mutation load that affects protein folding and stability. Studies on the existence and relevance of phenomena like evolutionary capacitance caused by the contributions that PPI make to folding stability (Dixit and Maslov 2013; Jarzab et al. 2020; Mateus et al. 2020), including cryptic genetic variation hidden by protein and RNA chaperones (Queitsch et al. 2002; Bergman and Siegal 2003; Masel 2005, 2006, 2013; Trotter et al. 2014; Geiler-Samerotte et al. 2016; Zheng et al. 2019) during experimental evolution, and the effects of such phenomena on rates of protein evolution may be especially valuable in this regard. Finally, it would be valuable to develop a better understanding of the temperature sensitivity of evolved LTEE populations (Mongold et al. 1996, 1999; Leiby and Marx 2014), and to collect data on protein evolutionary rates in long-term experiments conducted at elevated temperatures (Bennett et al. 1990; Tenaillon et al. 2012). Much remains to be explored, in regard to how evolution experiments can deepen our understanding of purifying selection on molecular and cellular organization and function.

Materials and Methods

Preprocessed LTEE metagenomic data were downloaded from: <https://github.com/benjaminhg/LTEE-metagenomic>.

Transcriptomic and proteomic data for REL606, grown in Davis minimal media with 500 mg/l glucose (DM500), were taken from the supplementary tables for Caglar et al. (2017). For robustness, I also analyzed the transcriptomic data for eleven

50,000 generation LTEE clones grown in DM4000 media (Favate et al. 2021) available at: <https://github.com/shahlab/LTEE-gene-expression>. I analyzed three different data sets of PPI in *E. coli*. First, I used the PPI network for *E. coli* K-12 MG1655 in the STRING database (Szklarczyk et al. 2021) as curated by Zitnik et al. (2019). Second, I used the data set of high confidence *E. coli* PPI interactions reported by Cong et al. (2019), which combines coevolutionary information in large protein multiple sequence alignments with gold-standard protein complexes in *E. coli* reported in the Ecocyc and Protein Databank (PDB) databases (Berman et al. 2000; Keseler et al. 2013). PPI network statistics were calculated using the SNAP toolkit (Leskovec and Sosič 2016; Zitnik et al. 2019). Third, additional data on *E. coli* PPI interactions and protein abundance were downloaded using the web interface to the ProteomeVis database (Razban et al. 2018), available at <http://proteomevis.chem.harvard.edu/>. Associated metadata for ProteomeVis were downloaded from: https://github.com/rzabnan/proteomevis/blob/master/make_database/proteomevis_inspect.csv.

All statistical analyses involve two-sided tests for Spearman correlation coefficients that are significantly different from zero, using the `cor.test` function in the R statistical programming language, version 4.0 (R Core Team 2020). Unless stated otherwise, all correlations include genes with no mutations (i.e., zeros are included).

Supplementary Material

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

Acknowledgments

I thank Richard Lenski, Lingchong You, Jeffrey Barrick, Stephanie Spielman, Tanush Jagdish, and Nkrumah Grant for valuable comments, advice, and discussions. I also thank Rostam Razban for help with the ProteomeVis database and John Favate and Premal Shah for access and discussions about their LTEE RNA-seq data. The LTEE that generated the bacteria used in this study is supported by a grant from the National Science Foundation (currently DEB-1951307) to Richard Lenski and Jeffrey Barrick.

Data Availability

The data and analysis codes underlying this article are available on the Dryad Digital Repository (doi: 10.5061/dryad.x0k6djh5). Analysis codes are also available at <https://github.com/rohanmaddamsetti/LTEE-network-analysis>.

Literature Cited

Alvarez-Ponce D, Aguilar-Rodríguez J, Fares MA. 2019. Molecular chaperones accelerate the evolution of their protein clients in yeast. *Genome Biol Evol.* 11(8):2360–2375.

- Alvarez-Ponce D, Feyertag F, Chakraborty S. 2017. Position matters: network centrality considerably impacts rates of protein evolution in the human protein–protein interaction network. *Genome Biol Evol.* 9(6):1742–1756.
- Alvarez-Ponce D, Sabater-Muñoz B, Toft C, Ruiz-González MX, Fares MA. 2016. Essentiality is a strong determinant of protein rates of evolution during mutation accumulation experiments in *Escherichia coli*. *Genome Biol Evol.* 8(9):2914–2927.
- Ba ANN, et al. 2019. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* 575(7783):494–499.
- Barrick JE, Lenski RE. 2009. Genome-wide Mutational Diversity in an Evolving Population of *Escherichia coli*. *Cold Spring Harb Symp Quant Biol.* 74:119–129.
- Barrick JE, Lenski RE. 2013. Genome dynamics during experimental evolution. *Nat Rev Genet.* 14(12):827–839.
- Bennett AF, Dao KM, Lenski RE. 1990. Rapid evolution in response to high-temperature selection. *Nature* 346(6279):79–81.
- Bergman A, Siegal ML. 2003. Evolutionary capacitance as a general feature of complex gene networks. *Nature* 424(6948):549–552.
- Berman HM, et al. 2000. The protein data bank. *Nucleic Acids Res.* 28(1):235–242.
- Caglar MU, et al. 2017. The *E. coli* molecular phenotype under different growth conditions. *Sci Rep.* 7:45303.
- Chen K, et al. 2017. Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation. *Proc Natl Acad Sci U S A.* 114(43):11548–11553.
- Cong Q, Anishchenko I, Ovchinnikov S, Baker D. 2019. Protein interaction networks revealed by proteome coevolution. *Science* 365(6449):185–189.
- Couce A, et al. 2017. Mutator genomes decay, despite sustained fitness gains, in a long-term experiment with bacteria. *Proc Natl Acad Sci U S A.* 114(43):E9026–E9035.
- Dixit PD, Maslov S. 2013. Evolutionary capacitance and control of protein stability in protein-protein interaction networks. *PLoS Comput Biol.* 9(4):e1003023.
- Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A.* 102(40):14338–14343.
- Drummond DA, Raval A, Wilke CO. 2006. A single determinant dominates the rate of yeast protein evolution. *Mol Biol Evol.* 23(2):327–337.
- Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134(2):341–352.
- Drummond DA, Wilke CO. 2009. The evolutionary consequences of erroneous protein synthesis. *Nat Rev Genet.* 10(10):715–724.
- Favate JS, Liang S, Yadavalli SS, Shah P. 2021. The landscape of transcriptional and translational changes over 22 years of bacterial adaptation. *bioRxiv*. Available from: 10.1101/2021.01.12.426406.
- Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary rate in the protein interaction network. *Science* 296(5568):750–752.
- Geiler-Samerotte KA, Zhu YO, Goulet BE, Hall DW, Siegal ML. 2016. Selection transforms the landscape of genetic variation interacting with Hsp90. *PLoS Biol.* 14(10):e2000465.
- Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017. The dynamics of molecular evolution over 60,000 generations. *Nature* 551(7678):45–50.
- Grant NA, Maddamsetti R, Lenski RE. 2021. Maintenance of metabolic plasticity despite relaxed selection in a long-term evolution experiment with *Escherichia coli*. *Am Nat.* Available from: 10.1086/714530.
- Hahn MW, Conant GC, Wagner A. 2004. Molecular evolution in large genetic networks: does connectivity equal constraint? *J Mol Evol.* 58(2):203–211.

- Hahn MW, Kern AD. 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol Biol Evol.* 22(4):803–806.
- Jarzab A, et al. 2020. Meltome atlas—thermal proteome stability across the tree of life. *Nat Methods.* 17(5):495–503.
- Kepp KP, Dasmeh P. 2014. A model of proteostatic energy cost and its use in analysis of proteome trends and sequence evolution. *PLoS One* 9(2):e90504.
- Keseler IM, et al. 2013. EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* 41(Database issue):D605–D612.
- Konaté MM, et al. 2019. Molecular function limits divergent protein evolution on planetary timescales. *eLife* 8:e39705.
- LaBar T, Adami C. 2017. Evolution of drift robustness in small populations. *Nat Commun.* 8(1):12.
- Lang GI, et al. 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* 500(7464):571–574.
- Leiby N, Marx CJ. 2014. Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. *PLoS Biol.* 12(2):e1001789.
- Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat.* 138(6):1315–1341.
- Leskovec J, Sosič R. 2016. Snap: a general-purpose network analysis and graph-mining library. *ACM Trans Intell Syst Technol.* 8(1):1–20.
- Levy ED, De S, Teichmann SA. 2012. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc Natl Acad Sci U S A.* 109(50):20461–20466.
- Levy SF, et al. 2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* 519(7542):181–186.
- Li W, Lynch M. 2020. Universally high transcript error rates in bacteria. *eLife* 9:e54898.
- Lobkovsky AE, Wolf YI, Koonin EV. 2010. Universal distribution of protein evolution rates as a consequence of protein folding physics. *Proc Natl Acad Sci U S A.* 107(7):2983–2988.
- Maddamsetti R. 2016. Gene flow in microbial communities could explain unexpected patterns of synonymous variation in the *Escherichia coli* core genome. *Mob Genet Elem.* 6(1):e1137380.
- Maddamsetti R, et al. 2017. Core genes evolve rapidly in the long-term evolution experiment with *Escherichia coli*. *Genome Biol Evol.* 9(4):1072–1083.
- Maddamsetti R, Grant NA. 2020a. Divergent evolution of mutation rates and biases in the long-term evolution experiment with *Escherichia coli*. *Genome Biol Evol.* 12(9):1591–1603.
- Maddamsetti R, Grant NA. 2020b. A simple test to infer mode of selection in metagenomics time series of evolving asexual populations. *bioRxiv.* Available from: 10.1101/2020.05.23.112508.
- Maddamsetti R, Hatcher PJ, et al. 2015. Synonymous genetic variation in natural isolates of *Escherichia coli* does not predict where synonymous substitutions occur in a long-term experiment. *Mol Biol Evol.* 32(11):2897–2904.
- Maddamsetti R, Lenski RE, Barrick JE. 2015. Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli*. *Genetics* 200(2):619–631.
- Masel J. 2005. Evolutionary capacitance may be favored by natural selection. *Genetics* 170(3):1359–1371.
- Masel J. 2006. Cryptic genetic variation is enriched for potential adaptations. *Genetics* 172(3):1985–1991.
- Masel J. 2013. Q&A: evolutionary capacitance. *BMC Biol.* 11:103.
- Mateus A, et al. 2020. The functional proteome landscape of *Escherichia coli*. *Nature* 588(7838):473–476.
- Mongold JA, Bennett AF, Lenski RE. 1996. Evolutionary adaptation to temperature. IV. Adaptation of *Escherichia coli* at a niche boundary. *Evolution* 50(1):35–43.
- Mongold JA, Bennett AF, Lenski RE. 1999. Evolutionary adaptation to temperature. VII. Extension of the upper thermal limit of *Escherichia coli*. *Evolution* 53(2):386–394.
- Pál C, Papp B, Hurst LD. 2001. Highly expressed genes in yeast evolve slowly. *Genetics* 158(2):927–931.
- Park C, Chen X, Yang J-R, Zhang J. 2013. Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A.* 110(8):E678–E686.
- Plata G, Gottesman ME, Vitkup D. 2010. The rate of the molecular clock and the cost of gratuitous protein synthesis. *Genome Biol.* 11(9):R98.
- Plata G, Vitkup D. 2018. Protein stability and avoidance of toxic misfolding do not explain the sequence constraints of highly expressed proteins. *Mol Biol Evol.* 35(3):700–703.
- Queitsch C, Sangster TA, Lindquist S. 2002. Hsp90 as a capacitor of phenotypic variation. *Nature* 417(6889):618–624.
- R Core Team. 2020. R: a language and environment for statistical computing. [Internet]. Vienna (Austria): R Foundation for Statistical Computing. Available from: <https://www.r-project.org/>.
- Razban RM. 2019. Protein melting temperature cannot fully assess whether protein folding free energy underlies the universal abundance–evolutionary rate correlation seen in proteins. *Mol Biol Evol.* 36(9):1955–1963.
- Razban RM, et al. 2018. ProteomeVis: a web app for exploration of protein properties from structure to sequence evolution across organisms' proteomes. *Bioinformatics* 34(20):3557–3565.
- Rudan M, Schneider D, Warnecke T, Krisko A. 2015. RNA chaperones buffer deleterious mutations in *E. coli*. *eLife* 4:e04745.
- Samhita L, Raval PK, Agashe D. 2020. Global mistranslation increases cell survival under stress in *Escherichia coli*. *PLoS Genet.* 16(3):e1008654.
- Serohijos AW, Rimas Z, Shakhnovich EI. 2012. Protein biophysics explains why highly abundant proteins evolve slowly. *Cell Rep.* 2(2):249–256.
- Serohijos AW, Shakhnovich EI. 2014. Contribution of selection for protein folding stability in shaping the patterns of polymorphisms in coding regions. *Mol Biol Evol.* 31(1):165–176.
- Szklarczyk D, et al. 2021. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49(D1):D605–D612.
- Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. 2007. Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends Biochem Sci.* 32(5):204–206.
- Tenaillon O, et al. 2012. The molecular diversity of adaptive convergence. *Science* 335(6067):457–461.
- Tenaillon O, et al. 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* 536(7615):165–170.
- Terai G, Asai K. 2020. Improving the prediction accuracy of protein abundance in *Escherichia coli* using mRNA accessibility. *Nucleic Acids Res.* 48(14):e81.
- Trotter MV, Weissman DB, Peterson GI, Peck KM, Masel J. 2014. Cryptic genetic variation can make “irreducible complexity” a common mode of adaptation in sexual populations. *Evolution* 68(12):3357–3367.
- Usmanova DR, Plata G, Vitkup D. 2021. The relationship between the misfolding avoidance hypothesis and protein evolutionary rates in the light of empirical evidence. *Genome Biol Evol.* 13(2): doi: 10.1093/gbe/evab006. .
- Walsh IM, Bowman MA, Santarriaga IFS, Rodriguez A, Clark PL. 2020. Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *Proc Natl Acad Sci U S A.* 117(7):3528–3534.
- Wylie CS, Shakhnovich EI. 2011. A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc Natl Acad Sci U S A.* 108(24):9916–9921.

- Yang J-R, Liao B-Y, Zhuang S-M, Zhang J. 2012. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc Natl Acad Sci U S A*. 109(14):E831–E840.
- Yang JR, Zhuang SM, Zhang J. 2010. Impact of translational error-induced and error-free misfolding on the rate of protein evolution. *Mol Syst Biol*. 6:421.
- Zheng J, Payne JL, Wagner A. 2019. Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science* 365(6451):347–353.
- Zitnik M, Sosič R, Feldman MW, Leskovec J. 2019. Evolution of resilience in protein interactomes across the tree of life. *Proc Natl Acad Sci U S A*. 116(10):4426–4433.

Associate editor: Brian Golding