Universal constraints on protein evolution in the long-term evolution experiment with *Escherichia coli*

Rohan Maddamsetti[1]

[1]Department of Biomedical Engineering, Duke University, Durham NC, USA
E-mail: rohan.maddamsetti@duke.edu

Abstract

Although it is well known that highly expressed and highly interacting proteins evolve slowly across the tree of life, there is little consensus for why this is true. Here, I report that highly abundant and highly interacting proteins evolve slowly in the hypermutator populations of Lenski's long-term evolution experiment with *E. coli* (LTEE). Specifically, the density of observed mutations per gene, as measured in metagenomic time series covering 60,000 generations of the LTEE, strongly anti-correlates with mRNA abundance, protein abundance, and degree of protein-protein interaction. Weaker positive correlations between protein thermostability and mutation density are observed in the hypermutator populations, counterbalanced by negative correlations between protein thermostability and mRNA and 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 highly expressed and highly interacting proteins evolve slowly.

Key words: purifying selection, experimental evolution, protein 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. 2020). 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. (2020), we reported evidence for purifying selection on aerobic-specific 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 expressed 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. 2020). 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 that is that it may be possible to resolve the causes of this universal pattern by experimental means.

Results

Strong anti-correlation between mRNA and protein abundance and mutation density per gene in hypermutators

I compared the density of observed mutations in the LTEE (Good, et al. 2017) to mRNA and protein abundance data for the LTEE ancestral strain, REL606 (Caglar, et al. 2017). These

comparisons are shown in Figure 1. In the hypermutator LTEE populations, mRNA abundance during exponential growth strongly anticorrelates with mutation density, while protein abundance, at all timepoints, strongly anticorrelates with mutation density (Pearson's $r < -0.09$, Bonferroni-corrected $p < 0.05/9$ in all cases). mRNA abundance during stationary phase positively correlates with mutation density in the nonmutator LTEE populations (Supplementary Figure 1). Therefore, highly expressed genes evolve slowly in the hypermutator populations of the LTEE.

Highly interacting proteins evolve rapidly in nonmutator populations and 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 to purifying selection.

I compared the number of protein-protein interactions (PPI) to the density of observed mutations for every protein-coding gene in the *E. coli* genome, using two distinct, high quality datasets of protein-protein interactions in *E. coli* (Cong, et al. 2019; Zitnik, et al. 2019), which I refer to as the Cong dataset and the Zitnik dataset. These comparisons are shown in Figure 2. As expected, I find positive correlations between mutation density and protein-protein interaction degree (PPI degree) in the non-mutators (Pearson's $r = 0.08$, $p < 0.0001$ for Zitnik dataset; Pearson's $r < 0.07$, $p = 0.0127$ for Cong dataset) and negative correlations between mutation density and PPI degree in the hypermutators (Pearson's $r = -0.11$, $p < 10^{-8}$ for Zitnik dataset; Pearson's $r = -0.01$, $p = 0.0378$ for Cong dataset).

Protein contact density anti-correlates with mutation density in both hypermutator populations

Another factor that has been reported to correlate with the rate of protein evolution is intra-protein contact density (Bloom, et al. 2006; Razban, et al. 2018). Contact density, when defined as the average number of contacts per residue in the Protein Data Bank (PDB) structure, has been reported to positively correlate with evolutionary rates (Razban, et al. 2018). I compared the contact density values reported in the ProteomeVis database (Razban, et al. 2018) to mutation densities per gene in the nonmutator and hypermutator populations of the LTEE. These results are shown in Supplementary Figure 2. In contrast to previously reported results, I find that mutation density negatively correlates with contact density in both nonmutator (Pearson's $r = -0.058$, $p = 0.0390$) and hypermutator populations (Pearson's $r = -0.068$, $p = 0.0156$).

Protein thermostability positively correlates with mutation density in hypermutators, but negatively correlates with mRNA and protein abundance

Recent technical advances have allowed researchers to quantitatively measure the thermal stability of entire proteomes (Leuenberger, et al. 2017; Jarzab, et al. 2020). An ongoing debate in the field is whether these data are informative for understanding why highly expressed proteins evolve slowly, especially in the context of assessing the misfolding avoidance hypothesis (Drummond and Wilke 2008; Plata, et al. 2010; Plata and Vitkup 2018; Razban 2019; Usmanova, et al. 2020). In light of this debate, I compared *E. coli* thermostability data from the Meltome Atlas (Jarzab, et al. 2020) to the LTEE metagenomics data (Good, et al. 2017) as well as mRNA and protein abundance data for the LTEE ancestor strain, REL606 (Caglar, et al. 2017).

Protein melting temperature positively correlates with mutation density in the hypermutators (Figure 3A). Excluding 488 proteins that did not melt in the Meltome Atlas data ("nonmelters"), I find a positive correlation between protein melting temperature and mutation density (Pearson's product-moment correlation: $r = 0.054$, $p = 0.0031$). Furthermore, nonmelters have a higher mutation density on average than proteins with a measured melting point (Wilcoxon rank sum test with continuity correction: $p = 0.0028$).

On the other hand, protein melting temperature negatively correlates with mRNA and protein abundance, at least during exponential growth phase (Figure 3B). During exponential growth (time points before 8 hours), the third of mRNA and proteins with the lowest melting temperature are most abundant (Pearson's $r < -0.05$, Bonferroni-corrected $p < 0.05/9$ in all cases).

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. 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), and indeed, this is the most parsimonious explanation for why similar patterns are seen in the LTEE.

These universal correlations may emerge from a number of different, and non-mutually 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), 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). 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 or cryptic genetic variation hidden by protein and RNA chaperones (Queitsch, et al. 2002; Bergman and Siegal 2003; Masel 2005, 2006; Dixit and Maslov 2013; Masel 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

Pre-processed LTEE metagenomic data was downloaded from: https://github.com/benjaminhgood/LTEE-metagenomic. Genome-wide transcriptomic and proteomic data for REL606, grown in glucose-limited minimal media, was taken from the supplementary tables for Caglar et al. (2017). For robustness, I analyzed two different datasets of protein-protein interactions (PPI) in *E. coli.* First, I used the PPI network for *E. coli* K-12 MG1655 in the STRING database (Szklarczyk, et al. 2016) as curated by Zitnik et al. (2019). Second, I used the dataset of high confidence *E. coli* PPI interactions reported by Cong et al. (2019), which combines co-evolutionary 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). Intra-protein contact density data was downloaded using the web interface to the ProteomeVis database (Razban, et al. 2018), available at http://proteomevis.chem.harvard.edu/. Associated metadata for ProteomeVis was downloaded from: https://github.com/rrazban/proteomevis/blob/master/make_database/proteomevis_inspect.csv. Thermostability data for the *E. coli* proteome was downloaded using the web interface to the Cross-species Meltome Atlas (Jarzab, et al. 2020), available at http://meltomeatlas.proteomics.wzw.tum.de:5003/.

Analysis codes are available at: https://github.com/rohanmaddamsetti/LTEE-network-analysis. Datasets and analysis codes to replicate the findings and figures in this paper are available on the Dryad Digital Repository (DOI pending publication).

Acknowledgements

I thank Richard Lenski, Lingchong You, Jeffrey Barrick, and Nkrumah Grant for valuable comments, advice, and discussions, and I thank Rostam Razban for help with the ProteomeVis database. 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.

References

Alvarez-Ponce D, Aguilar-Rodríguez J, Fares MA. 2019. Molecular chaperones accelerate the evolution of their protein clients in yeast. Genome biology and evolution 11:2360-2375.

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 biology and evolution 9:1742-1756.

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 biology and evolution 8:2914-2927.

Barrick JE, Lenski RE. 2013. Genome dynamics during experimental evolution. Nature Reviews Genetics 14:827-839.

Bennett AF, Dao KM, Lenski RE. 1990. Rapid evolution in response to high-temperature selection. Nature 346:79-81.

Bergman A, Siegal ML. 2003. Evolutionary capacitance as a general feature of complex gene networks. Nature 424:549-552.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The protein data bank. Nucleic Acids Research 28:235-242.

Bloom JD, Drummond DA, Arnold FH, Wilke CO. 2006. Structural determinants of the rate of protein evolution in yeast. Molecular biology and evolution 23:1751-1761.

Caglar MU, Houser JR, Barnhart CS, Boutz DR, Carroll SM, Dasgupta A, Lenoir WF, Smith BL, Sridhara V, Sydykova DK, et al. 2017. The E. coli molecular phenotype under different growth conditions. Scientific reports 7:45303.

Chen K, Gao Y, Mih N, O'Brien EJ, Yang L, Palsson BO. 2017. Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation. Proceedings of the National Academy of Sciences 114:11548-11553.

Cong Q, Anishchenko I, Ovchinnikov S, Baker D. 2019. Protein interaction networks revealed by proteome coevolution. Science 365:185-189.

Couce A, Caudwell LV, Feinauer C, Hindré T, Feugeas J-P, Weigt M, Lenski RE, Schneider D, Tenaillon O. 2017. Mutator genomes decay, despite sustained fitness gains, in a long-term experiment with bacteria. Proceedings of the National Academy of Sciences 114:E9026-E9035.

Dixit PD, Maslov S. 2013. Evolutionary capacitance and control of protein stability in protein-protein interaction networks. PLoS Comput Biol 9:e1003023.

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. Proceedings of the National Academy of Sciences 102:14338-14343.

Drummond DA, Raval A, Wilke CO. 2006. A single determinant dominates the rate of yeast protein evolution. Molecular biology and evolution 23:327-337.

Drummond DA, Wilke CO. 2009. The evolutionary consequences of erroneous protein synthesis. Nature Reviews Genetics 10:715-724.

Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 134:341-352.

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary rate in the protein interaction network. Science 296:750-752.

Geiler-Samerotte KA, Zhu YO, Goulet BE, Hall DW, Siegal ML. 2016. Selection transforms the landscape of genetic variation interacting with Hsp90. PLoS biology 14:e2000465.

Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017. The dynamics of molecular evolution over 60,000 generations. Nature 551:45-50.

Grant NA, Maddamsetti R, Lenski RE. 2020. Maintenance of metabolic plasticity despite relaxed selection in a long-term evolution experiment with Escherichia coli. bioRxiv.

Hahn MW, Conant GC, Wagner A. 2004. Molecular evolution in large genetic networks: does connectivity equal constraint? Journal of molecular evolution 58:203-211.

Hahn MW, Kern AD. 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Molecular biology and evolution 22:803-806.

Jarzab A, Kurzawa N, Hopf T, Moerch M, Zecha J, Leijten N, Bian Y, Musiol E, Maschberger M, Stoehr G. 2020. Meltome atlas—thermal proteome stability across the tree of life. Nature methods 17:495-503.

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:e90504.

Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martínez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M. 2013. EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Research 41:D605-D612.

Konaté MM, Plata G, Park J, Usmanova DR, Wang H, Vitkup D. 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. Nature Communications 8:1-12.

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: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. The American Naturalist 138:1315-1341.

Leskovec J, Sosič R. 2016. Snap: A general-purpose network analysis and graph-mining library. ACM Transactions on Intelligent Systems and Technology (TIST) 8:1-20.

Leuenberger P, Ganscha S, Kahraman A, Cappelletti V, Boersema PJ, von Mering C, Claassen M, Picotti P. 2017. Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. Science 355.

Levy ED, De S, Teichmann SA. 2012. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Proceedings of the National Academy of Sciences 109:20461-20466.

Li W, Lynch M. 2020. Universally high transcript error rates in bacteria. eLife 9:e54898.

Maddamsetti R, Grant NA. 2020a. Divergent Evolution of Mutation Rates and Biases in the Long-Term Evolution Experiment with Escherichia coli. Genome biology and evolution 12:1591-1603.

Maddamsetti R, Grant NA. 2020b. A simple test to infer mode of selection in metagenomics time series of evolving asexual populations. bioRxiv.

Maddamsetti R, Hatcher PJ, Green AG, Williams BL, Marks DS, Lenski RE. 2017. Core genes evolve rapidly in the long-term evolution experiment with Escherichia coli. Genome biology and evolution 9:1072-1083.

Masel J. 2006. Cryptic genetic variation is enriched for potential adaptations. Genetics 172:1985-1991.

Masel J. 2005. Evolutionary capacitance may be favored by natural selection. Genetics 170:1359-1371.

Masel J. 2013. Q&A: evolutionary capacitance. BMC biology 11:103.

Mongold JA, Bennett AF, Lenski RE. 1996. Evolutionary adaptation to temperature. IV. Adaptation of Escherichia coli at a niche boundary. Evolution 50:35-43.

Mongold JA, Bennett AF, Lenski RE. 1999. Evolutionary adaptation to temperature. VII. Extension of the upper thermal limit of Escherichia coli. Evolution 53:386-394.

Park C, Chen X, Yang J-R, Zhang J. 2013. Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly. Proceedings of the National Academy of Sciences 110:E678-E686.

Plata G, Gottesman ME, Vitkup D. 2010. The rate of the molecular clock and the cost of gratuitous protein synthesis. Genome biology 11:1-10.

Plata G, Vitkup D. 2018. Protein stability and avoidance of toxic misfolding do not explain the sequence constraints of highly expressed proteins. Molecular biology and evolution 35:700-703.

Queitsch C, Sangster TA, Lindquist S. 2002. Hsp90 as a capacitor of phenotypic variation. Nature 417:618-624.

Razban RM. 2019. Protein melting temperature cannot fully assess whether protein folding free energy underlies the universal abundance–evolutionary rate correlation seen in proteins. Molecular biology and evolution 36:1955-1963.

Razban RM, Gilson AI, Durfee N, Strobelt H, Dinkla K, Choi J-M, Pfister H, Shakhnovich EI. 2018. ProteomeVis: a web app for exploration of protein properties from structure to sequence evolution across organisms' proteomes. Bioinformatics 34:3557-3565.

Samhita L, Raval PK, Agashe D. 2020. Global mistranslation increases cell survival under stress in Escherichia coli. PLoS genetics 16:e1008654.

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P. 2016. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Research:gkw937.

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 in biochemical sciences 32:204-206.

Tenaillon O, Barrick JE, Ribeck N, Deatherage DE, Blanchard JL, Dasgupta A, Wu GC, Wielgoss S, Cruveiller S, Médigue C. 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. Nature 536:165-170.

Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. 2012. The molecular diversity of adaptive convergence. Science 335:457-461.

Terai G, Asai K. 2020. Improving the prediction accuracy of protein abundance in Escherichia coli using mRNA accessibility. Nucleic Acids Research.

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 68:3357-3367.

Usmanova DR, Plata G, Vitkup D. 2020. Ability of the misfolding avoidance hypothesis to explain protein evolutionary rates in the light of empirical evidence. bioRxiv:736280.

Walsh IM, Bowman MA, Santarriaga IFS, Rodriguez A, Clark PL. 2020. Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. Proceedings of the National Academy of Sciences 117:3528-3534.

Yang J-R, Liao B-Y, Zhuang S-M, Zhang J. 2012. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. Proceedings of the National Academy of Sciences 109:E831-E840.

Yang JR, Zhuang SM, Zhang J. 2010. Impact of translational error-induced and error-free misfolding on the rate of protein evolution. Molecular systems biology 6:421.

Zheng J, Payne JL, Wagner A. 2019. Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. Science 365:347-353.

Zitnik M, Feldman MW, Leskovec J. 2019. Evolution of resilience in protein interactomes across the tree of life. Proceedings of the National Academy of Sciences 116:4426-4433.

TABLES AND FIGURES

Figure 1. The density of observed mutations per gene in hypermutator LTEE populations anti-correlates with mRNA abundance in exponential growth phase, and anti-correlates with protein abundance at all time points. Each point represents a protein-coding gene in the genome of the *E. 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, while comparisons to protein abundance are shown in green. Statistically significant linear regressions are shown in blue, while non-significant regressions are shown in light gray.

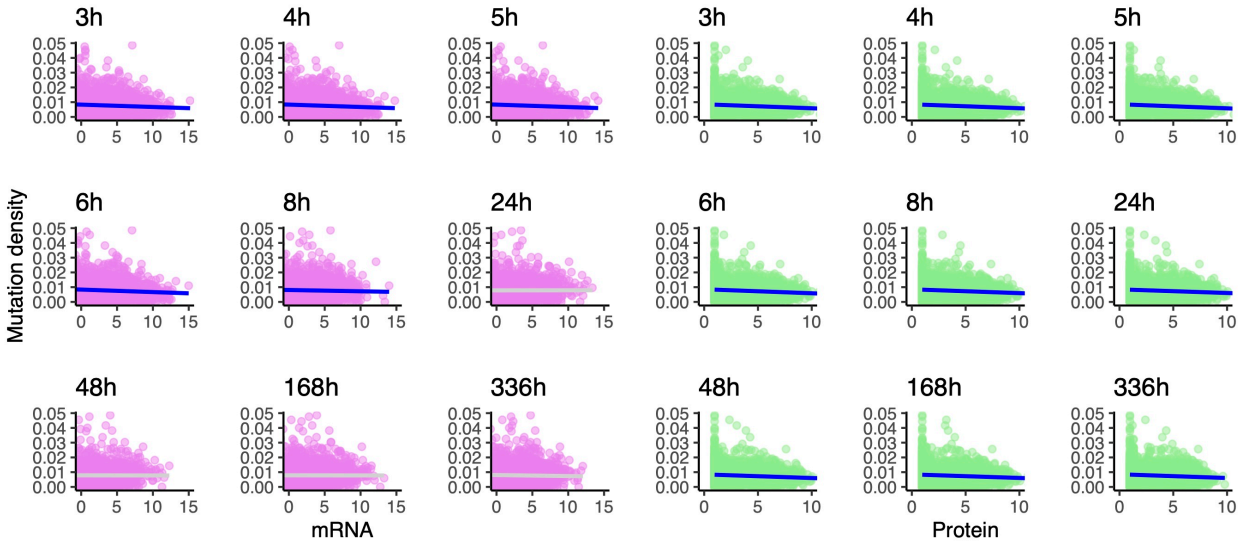


Figure 2. The density of observed mutations per gene in nonmutator LTEE populations positively correlates with PPI degree, while the density of observed mutations per gene in hypermutator LTEE populations negatively correlates with PPI degree. Comparisons to the PPI data from Zitnik et al. (2019) are shown in light blue, while comparisons to the PPI data from Cong et al. (2019) are shown in orange. A linear regression is shown in blue. A) Proteins with more interactions in the Zitnik et al. (2019) dataset tend to evolve faster than those with fewer interactions in the nonmutator LTEE populations. B) Proteins with more interactions in the Cong et al. (2019) dataset tend to evolve faster than those with fewer interactions in the nonmutator LTEE populations. C) Proteins with more interactions in the Zitnik et al. (2019) dataset tend to evolve more slowly than those with fewer interactions in the hypermutator LTEE populations. D) Proteins with more interactions in the Cong et al. (2019) dataset tend to evolve more slowly than those with fewer interactions in the hypermutator LTEE populations.

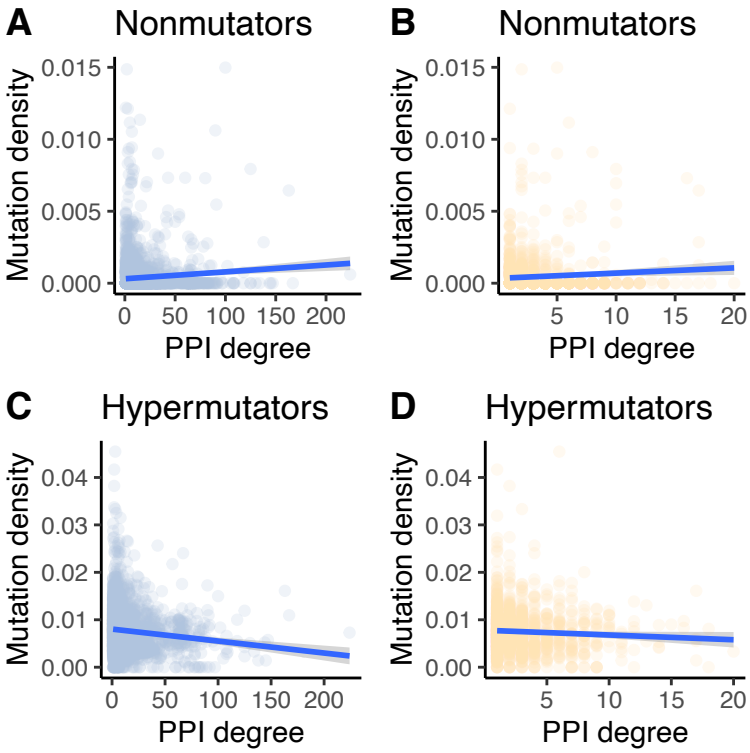
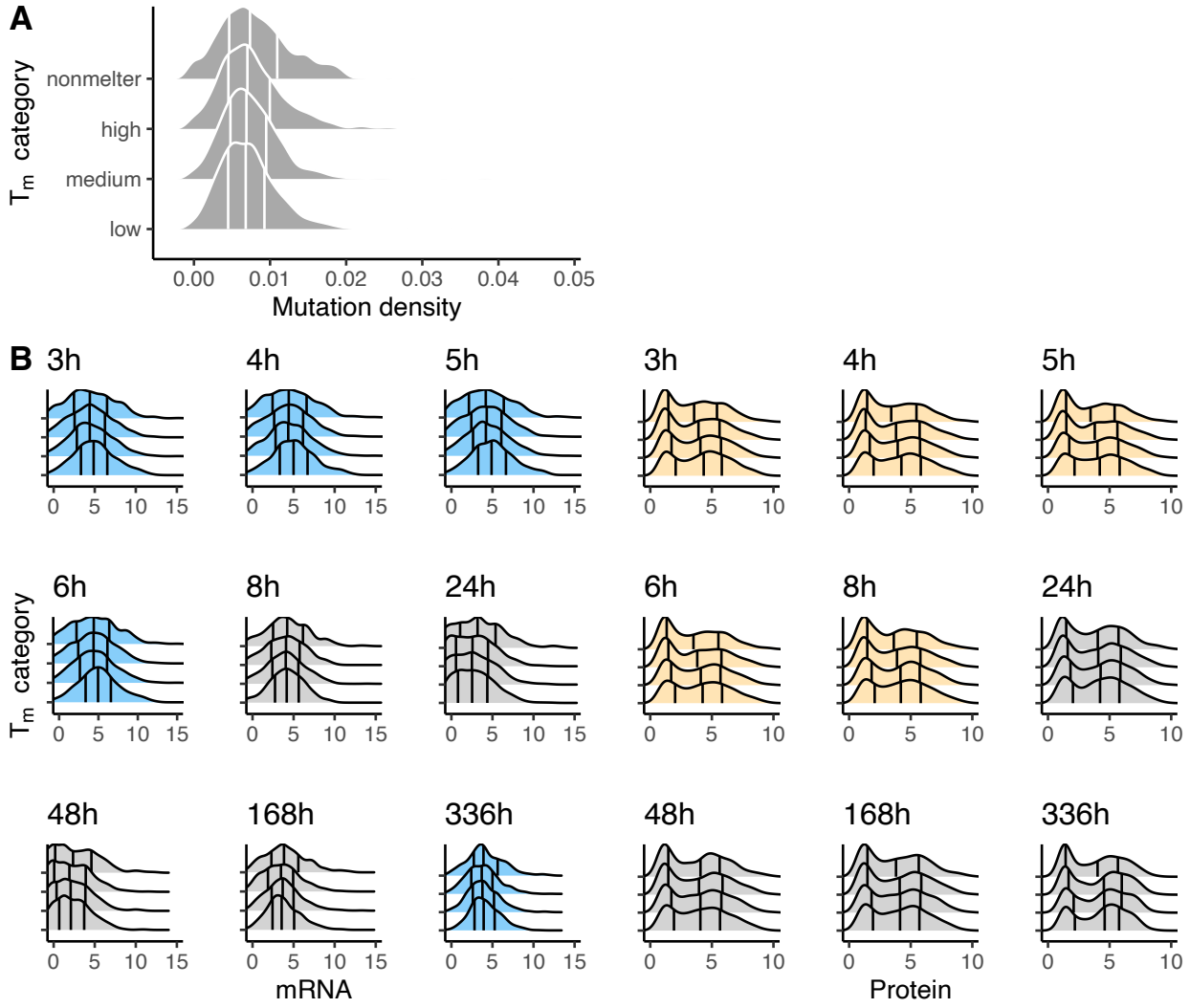
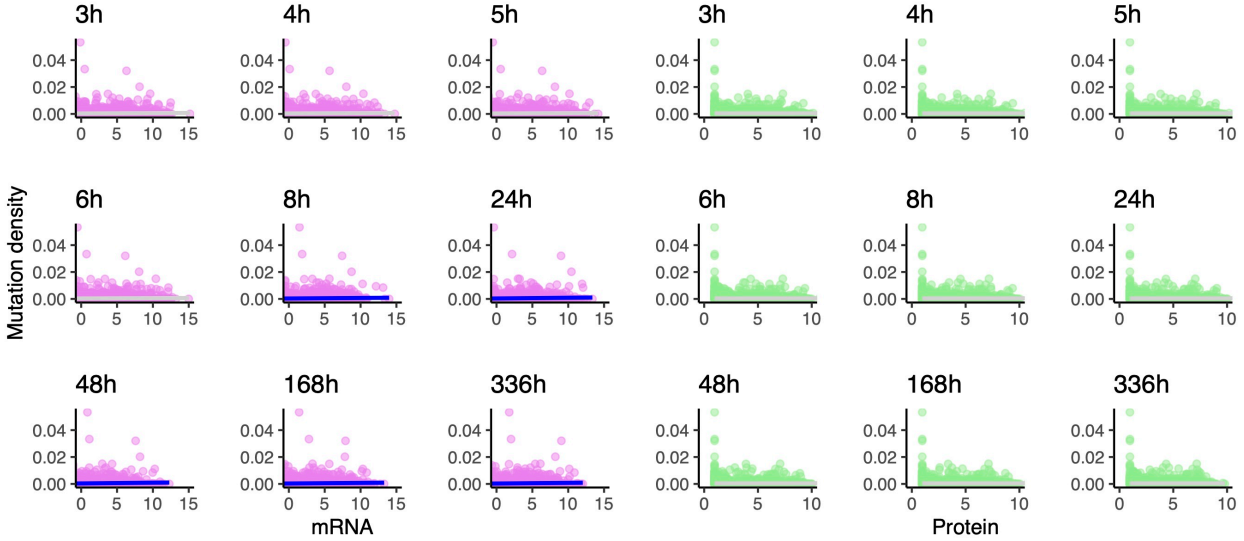


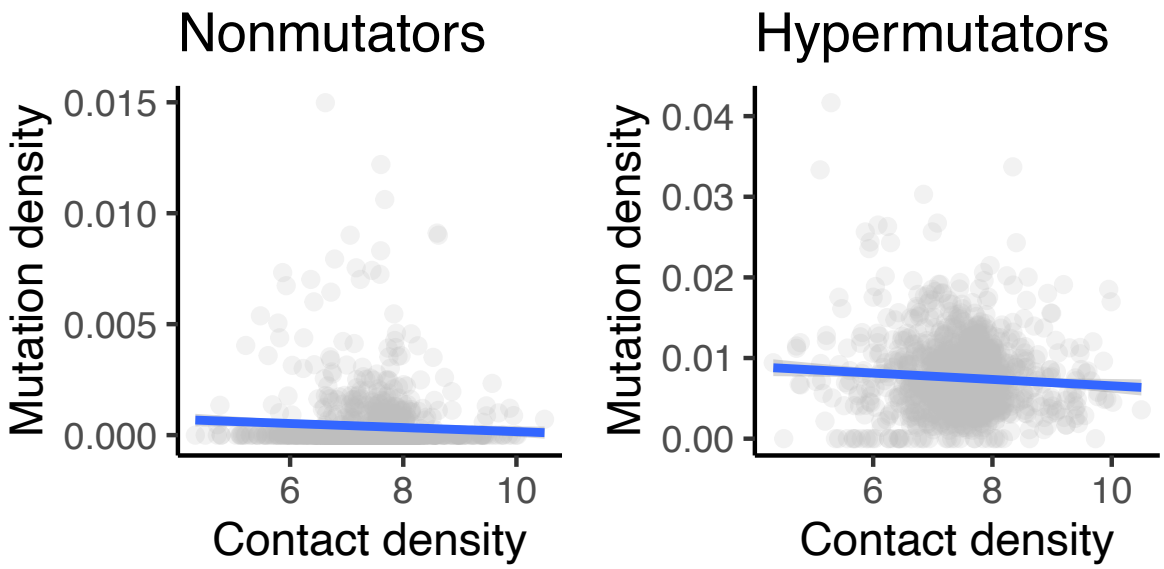
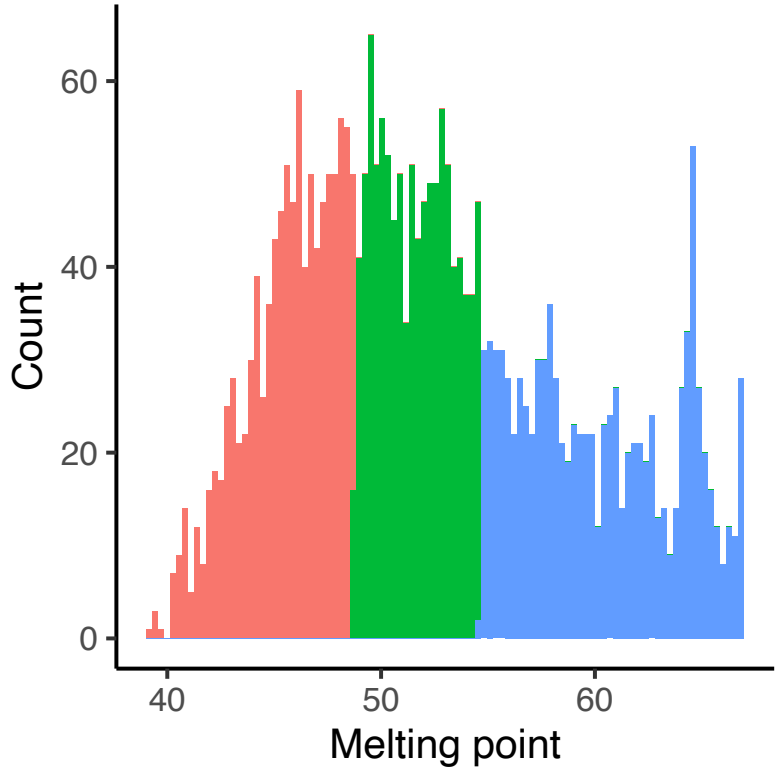
Figure 3. Comparison of protein melting temperatures (T_m) from the Meltome Atlas (Jarzab et al. 2020) to mutation density in LTEE hypermutator populations and mRNA and protein abundances in in the LTEE ancestral strain, REL606. A) T_m positively correlates with mutation density in LTEE hypermutator populations. B) T_m negatively correlates with mRNA and protein abundance during exponential growth phase. T_m categories are ordered as in panel (A).



Supplementary Figure S1. The density of observed mutations per gene in nonmutator LTEE populations positively correlates with mRNA abundance during stationary phase. Each point represents a protein-coding gene in the genome of the *E. 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, while comparisons to protein abundance are shown in green. Statistically significant linear regressions are shown in blue, while non-significant regressions are shown in light gray.



Supplementary Figure S2. The density of observed mutations per gene anti-correlates with intra-protein contact density, as reported in the ProteomeVis database, in both nonmutator and hypermutator LTEE populations.



Supplementary Figure S3. Classification of proteins into thirds (tertiles) by melting temperature (low, medium, and high), following the binning method used in the analysis of the Meltome Atlas reported by Jarzab et al. (2020).