PNAS COMMENTARY

How selection shapes the short- and long-term dynamics of molecular evolution

Csaba Pál[a,1] and Balázs Papp[a,b,c]

The field of molecular evolution traditionally relies on comparing the genomes of a small number of related species. These studies inform us of the driving evolutionary forces that acted in the long term, during molecular divergence of species. However, much less is known about how selection acts within species and in the short term. With the advance of high-throughput sequencing technologies, it has now become feasible to study thousands of complete genomes within species, offering new insights into recent genomic evolution. In particular, with the growing interest in genomic epidemiology, genomic data on clinical and natural isolates of bacterial pathogens are growing at an accelerating rate. This gives us an excellent opportunity to analyze how selection currently acts on bacterial pathogens, possibly as a result of human interventions.

In this issue of PNAS, Vigué and Tenaillon present a new bioinformatic method called GLASS that allows investigating the short-term dynamics of selection within species at an unprecedented resolution (1). Unlike most established methods that compare the frequencies of synonymous and nonsynonymous mutations, GLASS infers selection based on the functional impact of nonsynonymous mutations. For this, the method utilizes a recent computational framework to predict the impact of amino acid changes on protein stability and function (2, 3). The authors applied GLASS and other established methods to the genomes of over 60,000 *Escherichia coli* strains and found marked differences in the evolutionary forces driving the short- and long-term dynamics of protein evolution (Fig. 1). In agreement with earlier works (4, 5), they found that the long-term dynamics of protein evolution is governed by gene expression level: Highly expressed genes generally evolve at lower rates. In the short term, however, gene essentiality is also a major driver of purifying selection, indicating especially efficient removal of harmful variants in essential proteins from the population. In sum, the overall functional importance of genes (widely thought to reflect the level of functional constraint) contributes to within-species genetic polymorphism, but not to long-term protein evolution. Why should it be so?

To answer this question, we first need to discuss why proteins evolve at vastly different rates across species. Over 45 y ago, Zuckerkandl put forth the proposition that the primary factor governing the evolutionary rate of a protein is its functional density (6). In other words, the proportion of sites within the protein that are involved in specific functions plays a crucial role. To accurately predict the impact of selection on the entire protein, we must not only consider the proportion of sites affected by selection but also the distribution of selection strength among these sites. If a protein's performance is compromised in any way, the fitness effects will be more pronounced, particularly in proteins that make a significant overall contribution to fitness. Such proteins are often deemed essential as inactivating them experimentally results in organism lethality. Hence, according to this theory, the more important a protein is, the slower it should evolve.

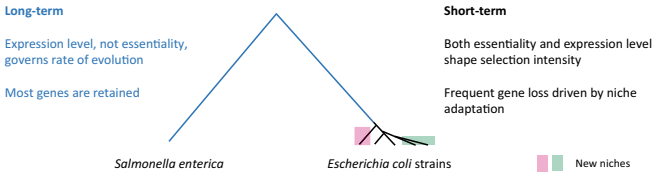

Fig. 1. Short- and long-term evolution of protein coding genes follow different rules. Schematic phylogenetic tree represents long (i.e., between-species) and short (i.e., within-species) time scales. Many genes that are under purifying selection over long time scales are recurrently lost in the short term as a result of local adaptation to new niches.

Surprisingly, however, the overall importance of proteins (estimated by systematic gene inactivation studies) seems to be a relatively poor predictor of evolutionary rate. Studies on a wide range of species indicate instead that the strongest predictor of evolutionary rate is the expression level of a protein (5). The exact mechanisms underlying the low evolutionary rate of highly expressed genes are still unclear. One controversial hypothesis claims that it results from the action of natural selection against nonheritable molecular errors during translation and other cellular processes (7). More specifically, selection against toxicity of misfolded proteins generated by translational errors may be especially strong in genes encoding highly abundant proteins. Others suggested that highly expressed proteins are under a stronger selection pressure to minimize gene expression costs, avoid protein misinteractions, or enhance mRNA folding (5). Put differently, mutations can have detrimental effects when they disrupt cellular mechanisms or introduce toxicity unrelated to the exact molecular function of the corresponding gene.

Author affiliations: [a]Synthetic and System Biology Unit, Biological Research Centre, National Laboratory of Biotechnology, Eötvös Loránd Research Network, Szeged HU-6726, Hungary; [b]Hungarian Centre of Excellence for Molecular Medicine - Biological Research Centre Metabolic Systems Biology Research Group, Szeged HU-6726, Hungary; and [c]National Laboratory for Health Security, Biological Research Centre, Eötvös Loránd Research Network, Szeged HU-6726, Hungary

Author contributions: C.P. and B.P. wrote the paper.

The authors declare no competing interest.

See companion article, "Predicting the effect of mutations to investigate recent events of selection across 60,472 *Escherichia coli* strains," 10.1073/pnas.2304177120.

[1]To whom correspondence may be addressed. Email: cpal@brc.hu.

Published August 2, 2023.

Why is the correlation between protein functional importance and evolutionary rate weak during long-term evolution? Several ideas have been put forward, which operate with confounding factors such as within-gene distribution of deleterious mutations or differences in gene essentiality in the wild and in the laboratory (8). However, as Vigué and Tenaillon demonstrated, selection against deleterious mutations within species is particularly strong in essential genes (1). The contrasting effects of gene essentiality on the long and short-term dynamics of molecular evolution could be resolved by at least two considerations. First, harmful mutations in essential genes are likely to impair fitness in most environments. In contrast, nonessential genes may typically have environment-specific functions that can be disrupted transiently during adaptation to varying environments but are maintained by selection over long time scales (see below). Thus, it is the class of nonessential genes that exhibit distinct evolutionary dynamics over short versus long time scales. Second, gene essentiality is itself an evolving trait. As gene essentiality can only be measured in extant species, these values might not be representative of the long-term evolution of the protein.

In PNAS, Vigué and Tenaillon present a new bioinformatic method called GLASS that allows investigating the short-term dynamics of selection within species at an unprecedented resolution.

Empirical data support the evolutionary lability of gene essentiality. The impact of gene inactivation on fitness shows considerable variation among closely related species and even between different isolates of the same unicellular species (9, 10). This suggests the widespread presence of compensatory mutations that can mitigate the effects of gene loss in certain populations but not in others. Interestingly, laboratory studies involving yeast and bacteria have demonstrated that when an important gene is lost, it triggers adaptive genomic changes that swiftly restore fitness (11, 12). Such compensatory mutations arise in other genes or regulatory regions, effectively compensating for the loss of the inactivated gene. These compensatory mutations may restore the normal function of the affected pathway or provide alternative mechanisms to maintain essential cellular processes. The observed phenomenon appears general as it is not limited to a single species or specific genes. It underscores the flexibility and adaptive potential of organisms to overcome genetic perturbations and maintain their fitness in the face of gene loss.

Another important finding of Vigué and Tenaillon is that loss-of-function mutations frequently target transcription factors (1). Although putatively deleterious loss-of-function alleles are prevalent in microbial populations, the underlying evolutionary mechanisms are far from being clear. There are at least three possibilities. First, population bottlenecks and genetic drift promote the accumulation of deleterious mutations, but this scenario may be unfeasible in microbes with large effective population sizes. Second, compensatory mutations may arise first in the population that later permit the rise of loss-of-function mutations without serious fitness consequences. This scenario is consistent with a recent systematic study showing that preexisting natural genetic variants frequently suppress deleterious mutations in specific natural strains of baker's yeast (13). Third, loss-of-function mutations could benefit the organism transiently when facing new stressful environments, but they remain useful in the long term. Indeed, antagonistic pleiotropy, i.e., when the same mutation is beneficial in one environment but harmful in another, is common in yeast strains with single gene deletions (14). As suggested by Vigué and Tenaillon and others (15), such loss-of-function mutations are very common and can provide rapid adaptation to specific environmental niches.

Perhaps the best evidence of adaptive loss of transcription factors comes from antibiotic adaptation studies (16). These studies convincingly demonstrated that loss-of-function mutations of specific transcriptional repressors are beneficial in times of antibiotic stress, as they yield elevated activity of multidrug efflux pumps. However, such mutations have specific associated fitness costs in antibiotic-free environments, hindering their long-term fixation in nature.

In sum, the study conducted by Vigué and Tenaillon offers new insights into the factors that shape protein evolution within species. In the future, it will be important to combine GLASS and related computational tools with deep-scan targeted mutagenesis and directed evolution. In recent years, the advent of recombineering- and CRISPR-based technologies has revolutionized the field by enabling highly efficient and targeted mutagenesis at multiple specific locations within the native genomic context (17). These groundbreaking technologies have dramatically accelerated and fine-tuned the process of introducing mutations in desired genomic loci and allow studying the phenotypic impact of mutations at unprecedented resolution.

By studying recent selection events using large genome collections, the approach developed by Vigué and Tenaillon can also be applied to infer signatures of selection in cancer genomes. Cancer is an end product of somatic evolution caused by the accumulation of specific driver mutations. Accurate assessment of positive and purifying selection in cancers is crucial to obtain a more complete catalogue of driver genes as well as genes essential for cancer growth (18). We anticipate that GLASS will complement existing tools of molecular evolution to address such questions and will yield new insights into the role of low-frequency polymorphisms in tumor formation.

ACKNOWLEDGMENTS. This work was supported by the National Laboratory of Biotechnology Grant 2022-2.1.1-NL-2022-00008 (C.P. and B.P.), the National Research, Development and Innovation Office 'Élvonal' program KKP KH125616 (B.P.), The National Laboratory for Health Security RRF-2.3.1-21-2022-00006 (B.P.), and the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 739593 (B.P.).

1. L. Vigué, O. Tenaillon, Predicting the effect of mutations to investigate recent events of selection across 60,472 *Escherichia coli* strains. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2304177120 (2023).
2. F. Morcos *et al.*, Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1293–E1301 (2011).
3. M. Figliuzzi, H. Jacquier, A. Schug, O. Tenaillon, M. Weigt, Coevolutionary landscape inference and the context-dependence of mutations in beta-lactamase TEM-1. *Mol. Biol. Evol.* **33**, 268–280 (2016).
4. C. Pál, B. Papp, L. D. Hurst, Highly expressed genes in yeast evolve slowly. *Genetics* **158**, 927–931 (2001).
5. J. Zhang, J. R. Yang, Determinants of the rate of protein sequence evolution. *Nat. Rev. Genet.* **16**, 409–420 (2015).
6. E. Zuckerkandl, Evolutionary processes and evolutionary noise at the molecular level. I. Functional density in proteins. *J. Mol. Evol.* **7**, 167–183 (1976).
7. D. A. Drummond, J. D. Bloom, C. Adami, C. O. Wilke, F. H. Arnold, Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14338–14343 (2005).
8. Z. Wang, J. Zhang, Why is the correlation between gene importance and gene evolutionary rate so weak? *PLoS Genet.* **5**, e1000329 (2009).
9. F. Rousset *et al.*, The impact of genetic diversity on gene essentiality within the *Escherichia coli* species. *Nat. Microbiol.* **6**, 301–312 (2021).
10. M. Galardini *et al.*, The impact of the genetic background on gene deletion phenotypes in *Saccharomyces cerevisiae*. *Mol. Syst. Biol.* **15**, e8831 (2019).
11. B. Szamecz *et al.*, The genomic landscape of compensatory evolution. *PLoS Biol.* **12**, e1001935 (2014).
12. D. Blank, L. Wolf, M. Ackermann, O. K. Silander, The predictability of molecular evolution during functional innovation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3044–3049 (2014).
13. L. Parts *et al.*, Natural variants suppress mutations in hundreds of essential genes. *Mol. Syst. Biol.* **17**, e10138 (2021).
14. W. Qian, D. Ma, C. Xiao, Z. Wang, J. Zhang, The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Rep.* **2**, 1399–1410 (2012).
15. A. W. Murray, Can gene-inactivating mutations lead to evolutionary novelty? *Curr. Biol.* **30**, R465–R471 (2020).
16. L. Praski Alzrigat, D. L. Huseby, G. Brandis, D. Hughes, Resistance/fitness trade-off is a barrier to the evolution of MarR inactivation mutants in *Escherichia coli*. *J. Antimicrob. Chemother.* **76**, 77–83 (2021).
17. T. M. Wannier *et al.*, Recombineering and MAGE. *Nat. Rev. Methods Primers* **1**, 7 (2021).
18. I. Martincorena *et al.*, Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041.e21 (2017).