The Fitness Effects of Codon Composition of the Horizontally Transferred Antibiotic Resistance Genes Intensify at Sub-lethal Antibiotic Levels

Michael Shaferman,[1] Melis Gencel,[2] Noga Alon,[1] Khawla Alasad,[1] Barak Rotblat,[1,3] Adrian W.R. Serohijos,[2,4] Lital Alfonta,[1,5] and Shimon Bershtein*[,1]

[1]Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

[2]Departement de Biochimie, Universite de Montreal, Montreal, Quebec, Canada

[3]Ben-Gurion University of the Negev, National Institute for Biotechnology in the Negev (NIBN), Beer-Sheva, Israel

[4]Centre Robert-Cedergren en Bio-informatique et Genomique, Universite de Montreal, Montreal, Quebec, Canada

[5]Department of Chemistry, Ben-Gurion University of the Negev, Beer-Sheva, Israel

***Corresponding author:** E-mail: shimonb@bgu.ac.il.

Associate editor: Miriam Barlow

Abstract

The rampant variability in codon bias existing between bacterial genomes is expected to interfere with horizontal gene transfer (HGT), a phenomenon that drives bacterial adaptation. However, delineating the constraints imposed by codon bias on functional integration of the transferred genes is complicated by multiple genomic and functional barriers controlling HGT, and by the dependence of the evolutionary outcomes of HGT on the host's environment. Here, we designed an experimental system in which codon composition of the transferred genes is the only variable triggering fitness change of the host. We replaced *Escherichia coli*'s chromosomal *folA* gene encoding dihydrofolate reductase, an essential enzyme that constitutes a target for trimethoprim, with combinatorial libraries of synonymous codons of *folA* genes from trimethoprim-sensitive *Listeria grayi* and trimethoprim-resistant *Neisseria sicca*. The resulting populations underwent selection at a range of trimethoprim concentrations, and the ensuing changes in variant frequencies were used to infer the fitness effects of the individual combinations of codons. We found that when HGT causes overstabilization of the 5′-end mRNA, the fitness contribution of mRNA folding stability dominates over that of codon optimality. The 5′-end overstabilization can also lead to mRNA accumulation outside of the polysome, thus preventing the decay of the foreign transcripts despite the codon composition-driven reduction in translation efficiency. Importantly, the fitness effects of mRNA stability or codon optimality become apparent only at sub-lethal levels of trimethoprim individually tailored for each library, emphasizing the central role of the host's environment in shaping the codon bias compatibility of horizontally transferred genes.

Key words: **codon bias, horizontal gene transfer, bacterial adaptation, codon optimality, mRNA folding stability.**

Introduction

The genetic code is redundant. With the exception of methionine and tryptophan, each amino acid is encoded by two to six codons. The synonymous codons are not used with equal frequencies, resulting in codon usage bias (Grantham, Gautier and Gouy 1980; Grantham, Gautier, Gouy, et al. 1980; Gouy and Gautier 1982; Sharp and Li 1987; Chen et al. 2004; Tuller, Carmi, et al. 2010; Goodman et al. 2013; Pechmann and Frydman 2013; Chaney et al. 2017; Jacobs and Shakhnovich 2017). The codon bias observed among different genomes is thought to be driven predominantly by genome-wide mutational (i.e., neutral) processes that shape the global nucleotide composition, rather than by selection on individual coding sequences (Knight et al. 2001; Chen et al. 2004). In contrast, codon bias found *within* genomes cannot arise from neutral processes alone, and requires the involvement of selection (Hershberg and Petrov 2008). The selectionist view of the intragenomic codon bias is corroborated by three robust phenomena detected across multiple prokaryotic and eukaryotic species. The first phenomenon is the universal reduction in the folding stability of the 5′-end mRNA secondary structure found in nearly all species tested, which is interpreted to be a result of selection pressure to optimize translation initiation (Kudla et al. 2009; Gu et al. 2010; Tuller, Waldman, et al. 2010; Bentele et al. 2013; Goodman et al. 2013; Bhattacharyya et al. 2018). The second phenomenon is the match between frequently used codons and abundant tRNA molecules (Ikemura 1985; Yamao et al. 1991; Moriyama and Powell 1997; Kanaya et al. 1999, 2001; Duret 2000; Rocha 2004). Selection may favor more frequent codons because optimizing the speed and accuracy of translation elongation

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

Article

increases the availability of free ribosomes and reduces the frequency of translational mutations (Gouy and Gautier 1982; Duret and Mouchiroud 1999; Ghaemmaghami et al. 2003; Goetz and Fuglsang 2005; Plotkin and Kudla 2011). The third phenomenon is the association of non-optimal (less frequent) codons with the formation of co-translational folding intermediates (Pechmann and Frydman 2013; Jacobs and Shakhnovich 2017). Reduction in the rates of translation elongation induced by clusters of non-optimal codons is thought to be beneficial for promoting the formation of native protein structures on the ribosome (Komar 2009; Schlesinger et al. 2017).

An important outcome of the neutral and adaptive forces operating on synonymous codons is the emergence of distinct patterns of codon bias in the genomes of different organisms (Deng et al. 2020). Bacterial genomes, in particular, exhibit stark differences in both genome-wide usage of codons and in non-random distribution of codons within individual genes (Bentele et al. 2013). Bacterial genomes are also continuously reshaped via gene loss and gene acquisition (Popa and Dagan 2011; Soucy et al. 2015). Genes are acquired by bacteria in a process dubbed "horizontal gene transfer" (HGT). HGT has the potential to induce rapid and profound phenotypic changes and is considered to be a major driving force of bacterial adaptation and speciation (Ochman et al. 2000; Nakamura et al. 2004). HGT, however, may be hindered by codon bias variability among bacterial genomes. Indeed, when a foreign gene integrates into the genome of a bacterial host, the codon incompatibility between the donor and the recipient may cause an inadequate expression of the newly acquired gene and/or incur a fitness cost on the host (Amoros-Moya et al. 2010; Lind et al. 2010; Bershtein, Serohijos, et al. 2015). It is, therefore, imperative to define the impact of codon bias on the establishment and maintenance of successful HGT events in bacteria.

The task of delineating the role of codon bias in HGT is complicated by multiple genomic and functional barriers that constrain HGT (Thomas and Nielsen 2005; Lind et al. 2010; Popa and Dagan 2011; Bershtein, Serohijos, et al. 2015; Kacar et al. 2017), and by dependence of the evolutionary outcomes of HGT on a particular environment of the host (Amoros-Moya et al. 2010). Computational studies that analyzed HGT events among bacterial genomes revealed that HGT frequency positively and strongly correlates with the similarity of tRNA pools between donors and acceptors, suggesting that the compatibility between codon usage of the donor and tRNA pools of the host, that is, codon optimality, constitutes an important codon usage-related requirement to successful HGTs (Medrano-Soto et al. 2004; Tuller et al. 2011). Such analyses, however, do not reveal the molecular mechanisms underlying failed instances of HGT. Indeed, Kudla et al. have demonstrated that only 5% of the variability in fluorescence of synthetic GFP genes with diversified synonymous codons expressed in *Escherichia coli* can be explained by codon optimality measured by codon adaptation index (CAI), whereas the folding stability of mRNA secondary structure near the translation start codon explained almost 60% of the fluorescence variability (Kudla et al. 2009). Conversely, higher CAI values of GFP variants correlated significantly with faster bacterial growth, possibly by sequestering less ribosomes than variants carrying low-frequency codons, whereas no correlation was observed between the 5′-end mRNA folding stability and growth. These findings provided a clear first indication that both mRNA folding stability at the starts of genes and codon optimality constitute barriers to the heterologous gene expression. However, mRNA stability and codon optimality cannot be correlated to the fitness of cells carrying individual GFP variants since GFP function is irrelevant to organismal physiology. A direct link between the function of the transferred genes and fitness of the host was established in another experimental study that measured the effects of 200 diverse antibiotic resistance-conferring plasmid genes on *E. coli* growth (Porse et al. 2018). Surprisingly, no significant contribution of the mRNA folding stability or codon usage preferences to the compatibility of the horizontally transferred genes was detected. Instead, resistance mechanisms were deemed to be major factors that control the compatibility of the heterologous antibiotic resistance genes (Porse et al. 2018). It is reasonable to conclude that the very high antibiotic concentrations used in this study (up to 30-fold the minimal inhibitory concentration [MIC]) potentially minimized the effect of sequence composition on fitness. This notion is particularly relevant, since bacteria in nature and in clinic are often exposed to non-lethal (sub-MIC) levels of antibiotics (Gullberg et al. 2011; Hughes and Andersson 2012). Further, the inferring of the contribution of codon bias of the various antibiotic resistance genes to *E. coli*'s growth could have been masked by mechanisms affecting the abundance and function of the protein products of the transferred genes downstream to transcription and translation steps, and, therefore, being unrelated to codon bias. These mechanisms may include contrasting sensitivities of transferred proteins to degradation and aggregation within the milieu of the host cell (Bershtein, Serohijos, et al. 2015), uneven efficiencies in protein transport to the periplasm or differences in fitness cost associated with such a transfer (Rajer and Sandegren 2022), or the formation of spurious interactions with cellular counterparts of the host (Bhattacharyya et al. 2016).

We reasoned that the incongruities found between the studies discussed above can be settled by establishing a clear link between codon bias of the transferred genes and fitness of the host. To this end, we used *folA* gene encoding dihydrofolate reductase (DHFR), an essential enzyme that constitutes a target for trimethoprim (TMP), as a model for HGT. We replaced the open reading frame of *E. coli*'s chromosomal copy of the *folA* gene (while preserving the endogenous regulatory region) with combinatorial libraries of synonymous codons of *folA* genes from TMP-sensitive *Listeria grayi* and TMP-resistant *Neisseria sicca*. We found that both mRNA stability at the translation initiation regions and codon optimality impact the

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

fitness of the host; however, this effect is prominent only within a relatively narrow range of sub-MIC TMP concentrations individually tailored for each xenologue, underlining the importance of the host's environment in shaping the codon bias compatibility of horizontally transferred genes.

Results

folA Gene as a Model of HGT

To tease apart the constraints imposed by codon bias on the functional integration of horizontally transferred genes, we devised the following experimental criteria. First, physiologically relevant changes in the function and abundance of the protein products of transferred genes should produce fitness effects in the host. Second, factors affecting the abundance and activity of foreign proteins downstream to the transcription and translation steps must not mask the changes in protein abundances caused by codon bias. Third, the fitness effects of the transferred genes must be measured at a range of suitable environmental conditions. To comply with these criteria, we performed an artificial HGT by replacing the open reading frame of the chromosomal *E. coli folA* gene encoding DHFR with various bacterial xenologoues. The first criterion is met in such an experimental setup through the essentiality of DHFR in *E. coli*. DHFR catalyzes NADPH-dependent reduction of dihydrofolate (DHF) to tetrahydrofolate (supplementary fig. S1, Supplementary Material online). The latter is an essential carrier for the one-carbon transfer reactions required for nucleic acids and amino acids metabolism (Schnell et al. 2004). Changes in functional DHFR levels, including reduction in DHFR abundance induced by the diversification of synonymous codons (Bhattacharyya et al. 2018), are known to directly control bacterial growth (Bershtein et al. 2013; Bershtein, Choi, et al. 2015; Bershtein, Serohijos, et al. 2015; Bhattacharyya et al. 2016; Rodrigues et al. 2016). When the product of DHFR abundance and catalytic activity ($k_{cat}/K_{M(DHF)}$) drops below the basal level, which in *E. coli* constitutes 40–100 DHFR molecules per cell (Taniguchi et al. 2010), the decrease in bacterial growth rate follows Michaelis–Menten-like kinetics (supplementary fig. S1, Supplementary Material online and ref. (Bershtein et al. 2013; Bershtein, Choi, et al. 2015; Rodrigues et al. 2016; Bhattacharyya et al. 2018)). The drop in functional DHFR levels causes a profound imbalance in the metabolic pools (Kwon et al. 2008; Bershtein, Choi, et al. 2015; Bhattacharyya et al. 2021), which, in turn, leads to an upregulation of *folA* transcription via a negative metabolic regulation loop operating through binding of TyrR transcription activator to two TyrR boxes located in *folA* regulatory region [supplementary fig. S1, Supplementary Material online and Yang et al. (2007)]. Furthermore, the increase in DHFR abundance, well above the basal level, results in a decline in the growth rate of *E. coli*, owing to the formation of transient protein–protein interactions that trigger toxic metabolic imbalance [supplementary fig. S1, Supplementary Material online, and Bhattacharyya et al. (2016)].

Upon the horizontal transfer of xenologous *folA* genes, changes in DHFR levels can occur not only as a result of codon bias but also due to the differential sensitivities of the xenologous DHFR proteins to aggregation and degradation in the cytoplasm of *E. coli* (Bershtein et al. 2013; Bershtein, Serohijos, et al. 2015). Moreover, the xenologous *folA* strains may differ in pleiotropic effects caused by the spurious interactions between their protein products and the cellular constituents of the host (Kramer 2015; Bhattacharyya et al. 2016). Pleotropic effects can be manifested in reshaping the metabolic and protein–protein interaction networks as well as in invoking a stress response (Bershtein, Choi, et al. 2015; Bhattacharyya et al. 2016), phenomena that have a direct effect on bacterial growth. Thus, to comply with the second criterion, we measured and compared changes in fitness effects caused by codon usage diversification within each individual xenologous *folA* strain, rather than comparing fitness effects across strains.

To fulfill the third criterion, we took advantage of the DHFR sensitivity to TMP, a widely-used antibiotic that competitively binds to active sites of bacterial DHFRs (Burchall and Hitchings 1965). We measured the fitness effects incurred by codon bias of xenologous DHFR proteins on *E. coli* growth at a range of TMP concentrations, including sub-MIC levels.

Lastly, comparative genomics studies have demonstrated that HGT plays an important role in the evolution of folate metabolic pathway in bacteria, including the spread of antifolate resistance (de Crécy-Lagard, et al. 2007; Stern et al. 2010), making DHFR a suitable target to study the constraints imposed by codon bias on HGT.

We assembled a collection of diverse xenologous *folA* genes from *Bacillus subtilis*, *Lactococcus lactis*, *Leuconostoc mesenteroides*, *N. sicca*, *Methyloccocus capsulatus*, *Listeria grayi*, and *Pseudomonas putida* that differ with respect to their sequence composition (GC content ranges from 34% to 66%), evolutionary distance (between 31% and 49% amino acid sequence identity relative to *E. coli*'s DHFR), and catalytic efficiency of their protein products ($k_{cat}/K_{M(DHF)}$ values of DHFRs range from 3.9 to 71.7 $\mu/M^{-1}\,sec^{-1}$) (table 1, supplementary table S1, Supplementary Material online).

Using CAI and tAI codon usage metrics of the donor bacteria and *E. coli* (Sharp and Li 1987; Tuller et al. 2011) (supplementary table S2, Supplementary Material online), we calculated how the CAI and tAI values of the xenologous *folA* genes, CAIg, and tAIg (defined as the geometric mean of the CAI or tAI of all codons of a gene) would change upon horizontal transfer to *E. coli*. CAI and tAI scales estimate how codons affect translational efficiencies. CAI assumes that frequent codons are used more often in highly expressed genes (Sharp and Li 1987). tAI relies on relative tRNA abundance and on constraints imposed by

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

codon–tRNA wobble interactions (dos Reis et al. 2004). We determined that the transfer of the xenologous *folA* genes to the chromosome of *E. coli* would be accompanied by a drop in CAIg and tAIg values for most xenologues (supplementary fig. S2*A* and *B*, Supplementary Material online). To estimate how representative this finding is with respect to other genes, we obtained distributions of CAIg and tAIg values of ten randomly selected orthologous genes encoding central metabolic enzymes (table S3 and fig. S3*A* and *B*, Supplementary Material online). We found that drop in means of these distribution upon changing either the CAI or tAI metrics of the donor species to those of *E. coli* significantly correlates with the corresponding drop in the individual CAIg and tAIg values of *folA* genes (Pearson $R = 0.83$, $P = 0.02$ for CAIg metrics; $R = 0.99$, $P = 0.000019$ for tAIg metrics), suggesting that *E. coli* features a rather different codon usage pattern than the chosen donor species (fig. S3*C* and *D*, and table S3, Supplementary Material online, and **Materials and Methods**).

As mentioned above (Gu et al. 2010), genes are universally subject to selection pressure to reduce mRNA secondary structure near the translation start codon. However, if a horizontally transferred gene is acquired without the original promoter, it must be inserted near host promoter to be expressed. The combination of the 5′-untranslated regulatory region of the host with the synonymous codons near the start codon of the donor gene may induce changes in the folding stability of the secondary mRNA structure of a xenologue, and, potentially, affect the efficiency of translation. We, therefore, calculated the expected changes in the folding stability profiles (Gibbs free energy difference between unfolded and folded states of mRNA, ΔG) of the 5′-end mRNA transcripts of the xenologous *folA* genes in a scenario in which the original sequence immediately upstream to the protein coding region, as found in the donor organism, is replaced with the *folA* regulatory sequence of the host upon horizontal transfer to *E. coli*. Since *folA* transcription in *E. coli* commences at nucleotide −25, the mRNA folding stability profiles were calculated for a 30-nt long fragment, the length that roughly corresponds to a ribosomal footprint size (Martens et al. 2015), from nucleotide −25 of the upstream sequence and up to nucleotide +5 within the coding sequence (see **Materials and Methods**). As anticipated, we found a wide range of stability effects near the translation start site (supplementary fig. S2*C*, Supplementary Material online). In some xenologues, mRNA transcripts exhibited a 1–3 kcal/mol increase in folding stability relative to the original organism, whereas in others, loss of folding stability ranged from 1 to 11 kcal/mol. It is worth noting, however, that in some donor organisms, *folA* genes are transcribed as part of a larger polycistronic transcript, often immediately downstream to the coding sequences of other genes (supplementary fig. S4, Supplementary Material online). Therefore, the possible contribution of the 5′-end *folA* mRNA folding stability to translation efficiency in these bacteria may differ markedly from other organisms where, like in *E. coli*, *folA* is transcribed as a single gene from a unique promoter.

Previously, the replacement of rare synonymous codons of the endogenous *folA* gene in *E. coli* with the most frequent codons produced a reduction in bacterial growth (Bhattacharyya et al. 2018). We, therefore, assumed that a similar replacement within xenologous *folA* genes will also cause changes in the growth of the host. To test this assumption, we synthesized two sets of protein coding sequences for each of the seven chosen xenologous *folA* genes. The first set is composed of the original sequences as they appear in the donor organisms (from here on "ORG" genes) (supplementary table S1, Supplementary Material online), and the second set consists of modified sequences, in which the original synonymous codons were replaced with the most frequent codons in the *folA* gene of *E. coli* ("MOD" genes) (fig. S5 and table S4, Supplementary Material online). Using the same tactic, we also generated a MOD sequence for the endogenous *folA* gene from *E. coli*. As expected, codon replacement led to an increase in CAIg and tAIg values (calculated using *E. coli*'s codon usage metrics, supplementary table S2, Supplementary Material online) in all of the modified sequences as well as to an increase in the GC content in sequences whose original GC content was <60% (supplementary fig. S6, Supplementary Material online and table 1). Codon replacement also induced profound changes in the mRNA folding stability profiles of MOD sequences. mRNA stability was calculated using a 30-nucleotide long sliding window starting from the mRNA transcription start (nucleotide −25) in steps of 1 nt for a total of 100 windows (see **Materials and Methods**). As a rule, MOD sequences were found to be more stable than their ORG counterparts (supplementary fig. S7, Supplementary Material online). However, the extent of the change varied between strains and between different locations within the same mRNA transcript. Lastly, we replaced the open reading frame of the chromosomal copy of the *folA* gene of *E. coli* with each of the synthesized genes, while preserving the endogenous upstream regulatory region of *folA*, and measured the ensuing effects on the *folA* promoter activity, mRNA abundance, intracellular DHFR levels, and growth rates of the host either in the absence or presence of TMP.

The Original to Frequent Codon Replacements Within Xenologous *folA* Genes Affects the strains' Sensitivity to Sub-MIC Trimethoprim Levels

The introduction of ORG sequences into the *E. coli* host produced a distribution of growth effects ranging from a 10% to 20% drop (for five out of seven strains) and up to a 10% increase (for one strain) in the growth rates (fig. 1*A*). We found that the functional capacity of xenologous DHFRs (the product of DHFR abundance and catalytic activity ($k_{cat}/K_{M(DHF)}$) is not a good predictor of the obtained distribution of growth effects (fig. 1*C* and supplementary fig. S8*A*, Supplementary Material online). Indeed, some xenologous strains with functional capacity

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

Table 1. Orthologous *folA* Genes Used in the Study.

#	Origin of *folA* gene		$GC^{a}_{genome,\%}$	$GC^{b}_{folA,\%}$	$AA^{c}_{identity,\%}$	$K_{M(DHF),}{}^{d}\mu M$	$k_{cat}/K_{M(DHF)}{}^{d}{}_{,}/\mu Msec^{-1}$
	Species	Class					
1	*Bacillus subtilis (B.su)*	Bacilli	44	43	44	1.8	30.8
2	*Lactococcus lactis (L.la)*	Bacilli	35	34	31	0.8	44.2
3	*Leuconostoc mesenteroides (L.me)*	Bacilli	38	42	32	1.6	28.1
4	*Neisseria sicca (N.si)*	Betaproteobacteria	51	58	44	0.2	62.5
5	*Methylococcus capsulatus (M.ca)*	Gammaproteobacteria	64	66	49	1.5	3.9
6	*Listeria grayi (L.gr)*	Bacilli	42	45	37	0.9	19.5
7	*Pseudomonas putida (P.pu)*	Bacilli	60	63	42	0.9	71.7
8	*Escherichia coli (E.co)*	Gammaproteobacteria	51	53	100	0.8	8.1

[a]GC content of the original genome (Bentele et al. 2013).
[b]GC content of *folA* genes (see supplementary table S1, Supplementary Material online).
[c]% amino acid identity relatively to *E. coli*'s DHFR.
[d]from Bershtein et al. (2013).

comparable to or even higher than that of *E. coli* (e.g., *B. subtilis*, *L. lactis*, *N. sicca*, and *L. sicca*) nonetheless exhibit a reduction in bacterial growth, whereas growth of others (e.g., *P. putida* and *M. capsulatus*) remains unperturbed or even improved (fig. 1C). This finding is further supported by only a mild change in *folA* promoter activity with the introduction of ORG sequences (measured using a reporter plasmid in which the endogenous *folA* promoter is fused to GFP (Zaslaver et al. 2006)), indicating that DHFR functional capacity of most ORG strains is not lacking and, therefore, cannot drive the observed variability in growth effects (fig. 1D, *left panel* and supplementary fig. S8*B* and table S5, Supplementary Material online).

As mentioned above, the differences in growth effects between xenologous strains are complex and involve diverse mechanisms, many of which operate downstream to transcription and translation steps (i.e., at the level of functional DHFR protein). Thus, to delineate the contribution of codon usage of xenologous genes to fitness of the host, we compared the effects of codon replacements from ORG to MOD sequences for each individual DHFR xenologue, rather than conducting comparisons across ORG strains. In the absence of TMP, codon replacement resulted in a 5–20% decrease in the growth rates of four out of seven xenologous strains (*L. mesenteroides*, *N. sicca*, *M. capsulatus*, and *P. putida*), a trend similar to that observed upon rare to frequent codon replacement in the endogenous *folA* coding sequence in *E. coli* [fig. 1A and Bhattacharyya et al. (2018)]. In these strains, the drop in growth was typically accompanied by a decrease in the intracellular soluble DHFR fraction (fig. 1B), and, consequently, in the DHFR functional capacity (fig. 1C). In the strain carrying *L. grayi*, however, a drop in protein abundance upon codon replacement was accompanied by a 30% *increase* in growth, suggesting that higher DHFR levels produced by the original sequence are toxic to *E. coli* (fig. 1A and *B*). In *L. lactis* strain, a drop in DHFR levels produced almost no effect on growth. No change in either growth rate or protein abundance was observed in the strain carrying modified *B. subtilis folA* gene (fig. 1A). The variability in the growth effects did not correlate with changes in CAIg/tAIg values or the GC content of the modified sequences. However, changes in CAIg/tAIg did correlate significantly with changes in *folA* promoter activity (Pearson's correlation $R = 0.87$, $P = 0.005$ for tAIg; $R = 0.8$, $P = 0.02$ for CAIg) (fig. 1D and supplementary fig. S9*A* and *B*, Supplementary Material online). The correlation is inverse, meaning that higher CAIg/tAIg values correlate with less active promoter. CAIg/tAIg values are measures of codon optimality, and an increase in these values is expected to be associated with more efficient translation.

The correlation between a decrease in *folA* promoter activity, which is indicative of an increase in the functional DHFR levels (Bollenbach et al. 2009; Bershtein, Choi, et al. 2015), with an increase in tAIg/CAIg values, indeed indicates an increase in translation efficiency. Using a 30-nucleotide long sliding window starting from the mRNA transcription start (25 nt upstream of the translation start codon) and moving 1 nt at a time, we also calculated the difference in the mRNA folding stability between the ORG and MOD *folA* xenologs, $\Delta\Delta G$. Changes in the promoter activity upon codon replacement correlated weakly, yet significantly, with mRNA $\Delta\Delta G$ values near the translation start codons (windows 1–5) (supplementary fig. S9*E*, Supplementary Material online). An increase in mRNA stability was accompanied by a stronger promoter activation, indicating that overstabilization of the 5′-end of mRNA structure interfered with translation and led to a reduction in functional DHFR levels (Spearman's correlation $r = 0.79$, $P = 0.035$) (supplementary fig. S9*E*, Supplementary Material online). However, no significant correlation between mRNA $\Delta\Delta G$ values and changes in growth was observed.

To estimate the fitness effects of codon diversification in the presence of TMP, we exposed the xenologous strains to a wide range of TMP concentrations (below and up to MIC levels) and determined the concentrations that caused 50% drop in growth (IC50). With the exception of the *L. lactis* xenologous strain, which is highly resistant to TMP, growth differences between ORG and MOD strains became pronounced in the presence of sub-MIC levels of TMP, with ORG over MOD ratios of IC50 concentrations ranging from 0.44- to 5.8-fold for the xenologous strains and 18-fold for the endogenous *folA* from *E. coli*

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

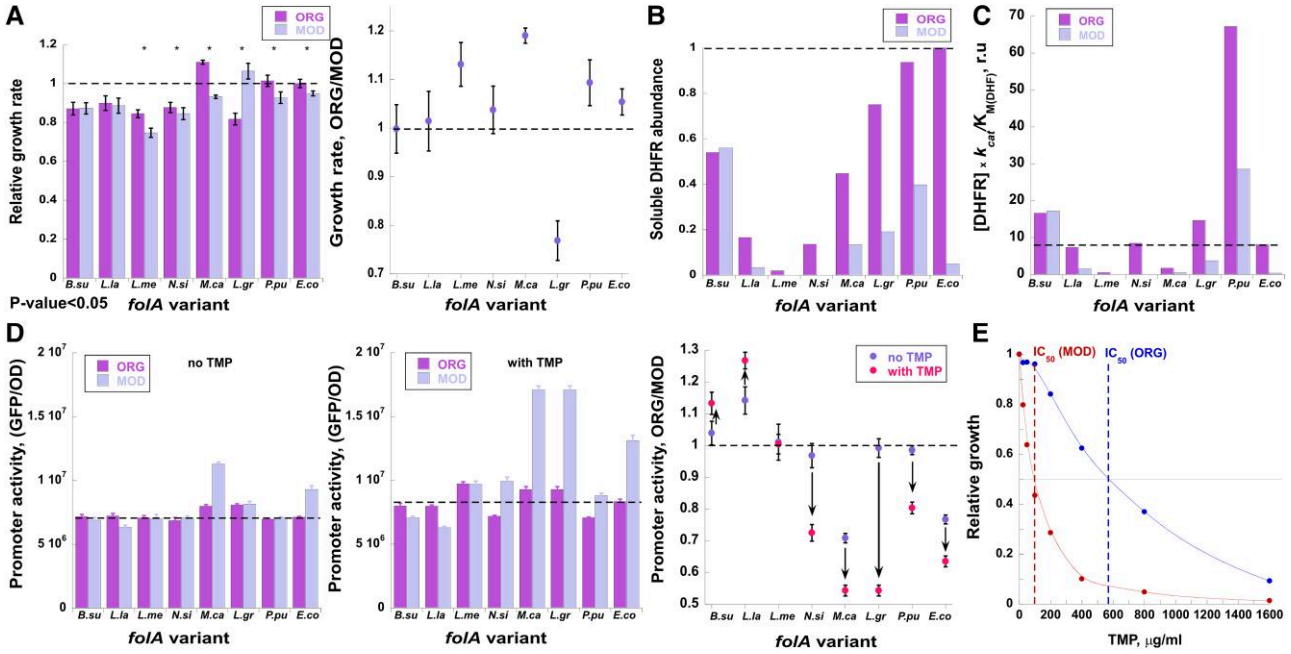


Fig. 1. Codon replacements within *folA* xenologoues induce phenotypic effects in *E. coli* host. *folA* variants are marked as in table 1. (*A*) *Left panel*. Growth rates of strains carrying original (ORG) and modified (MOD) *folA* sequences relative to growth of WT *E. coli* strain (dashed line). Asterisks indicate variants exhibiting significant differences in growth rates (two-tailed test, $P < 0.05$). *Right panel*. Growth ratios between ORG and MOD *folA* variants. Variants above the dashed line exhibit better growth with original *folA* sequence. (*B*) Intracellular DHFR abundance in strains carrying ORG and MOD *folA* sequences relative to DHFR abundance in WT *E. coli* (dashed line). (*C*) Functional capacity (a product of the relative intracellular soluble abundance of ORG and MOD xenologous DHFR proteins and their catalytic activity, k_{cat}/K_M(DHF)) *r.u.*, relative units. Dashed line indicates functional capacity of the endogenous DHFR in *E. coli*. (*D*) Promoter activity of ORG and MOD strains (presented as a ratio between GFP fluorescence generated by the reporter plasmid and optical density of bacterial culture at 600 nm) in the absence (*left panel*) or presence (*middle panel*) of IC50 amounts of TMP (see table 2). Dashed line corresponds to *folA* promoter activity in WT *E. coli*. *Right panel*. Ratios of promoter activity between strains carrying ORG and MOD *folA* sequences are plotted for each variant in the absence of TMP (*blue*) and in the presence of IC50 amounts of TMP (*red*). The direction of change upon TMP addition is indicated with black arrows. In variants found below the dashed line, the *folA* promoter tends to be more active with MOD *folA* sequences. Error bars were calculated from four independent measurements. (*E*) Normalized growth of *E. coli* strain carrying ORG (*blue*) or MOD (*red*) *folA* gene from *N. sicca* is plotted as a function of TMP concentrations. Dashed line intersections with the black line indicate TMP concentrations at which growth of each variant reaches 50% of its growth in the absence of TMP (IC50).

(fig. 1*E*, supplementary fig. S10, Supplementary Material online, and table 2). In contrast, at TMP levels approaching a MIC, the difference in growth between strains carrying ORG and MOD *folA* sequences largely dwindled (fig. 1*E*, supplementary fig. S10, Supplementary Material online, and table 2).

As a general trend, the replacement of original to frequent codons resulted in an increased sensitivity to sub-MIC levels of TMP. This was true also for the *L. lactis* strain, which, in the absence of TMP, showed an increase in growth of the strain carrying MOD sequence (fig. 1*A*). It is plausible that in the presence of TMP, a competitive inhibitor of DHFR, higher *L. lactis* DHFR abundance confers a fitness advantage. Conversely, in case of *B. subtilis* xenologous strain, the MOD sequence exhibited a reduced sensitivity to TMP (table 2 and supplementary fig. S10, Supplementary Material online). Interestingly, in contrast to other xenlogues, the 5′-mRNA region of the MOD sequences of *B. subtilis folA* exhibited almost 5 kcal/mol overstabilization compared with the ORG sequence (supplementary fig. S7*A* Supplementary Material online). Changes in CAIg/tAIg values or in GC content were not predictive of the growth effects observed in the host grown in the presence of TMP, but the observed changes in TMP sensitivity were largely captured by the response of the *folA* endogenous promoter fused to a GFP in a plasmid reporter (fig. 1*D*). Addition of TMP at IC50 levels led to an increase in the promoter activity in *N. sicca*, *M. capsulatus*, *L. grayi*, and *P. putida* xenologous strains carrying MOD *folA* sequences. Conversely, in the case of *B. subtilis* and *L. lactis*, the promoter activity was markedly increased in strains carrying the ORG sequences, indicating that in these genes, codon replacement led to a reduction in TMP sensitivity. Overall, a significant inverse correlation was observed between changes in tAIg values and changes in promoter activity (Pearson's correlation $R = 0.88$, $P = 0.004$) (supplementary fig. S9*C*, Supplementary Material online). Following a similar, albeit insignificant, trend, an inverse correlation was detected between changes in CAIg values and promoter activity (supplementary fig. S9*D*, Supplementary Material online). The increase in promoter activity upon codon replacement in the presence of sub-MIC TMP was also significantly correlated with mRNA thermodynamic stabilization near the translation start

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

Table 2. MIC and IC50 Concentrations of TMP for Xenologous and Endogenous *E. coli* Strains Carrying Original (ORG) and Modified (MOD) *folA* Sequences.

#	Origin of *folA* gene	ORG[a]		MOD[b]		IC_{50}[e]
		MIC[c]	IC_{50}[d]	MIC	IC_{50}	ORG/MOD
1	*Bacillus subtilis*	>10	3.1	>10	7.1	0.44
2	*Lactococcus lactis*	>2,000	1,750	>2,000	1,750	1
3	*Leuconostoc mesenteroides*	>1,600	380	>1,600	120	3.2
4	*Neisseria sicca*	>1,600	580	>1,600	100	5.8
5	*Methylococcus capsulatus*	>10	5.8	>10	3	1.9
6	*Listeria grayi*	>2.5	0.37	>2.5	0.1	3.7
7	*Pseudomonas putida*	>1,600	1,200	1,600	590	2
8	*Escherichia coli*	>10	1.8	2	0.1	18

[a]Strains carrying original *folA* sequences.
[b]Strains carrying original *folA* sequences.
[c]Minimal Inhibitory concentration of TMP, μg/ml.
[d]TMP concentration inhibiting 50% of growth, μg/ml.
[e]Ratio of IC50 TMP concentrations between strains carrying ORG and MOD *folA* genes.

codon, similar to what was observed in the absence of TMP (Pearson's correlation $R = 0.89$, $P = 0.007$) (supplementary fig. S9*E*, Supplementary Material online).

Collectively, these findings indicate that codon bias-related fitness effects vary substantially between the xenologues. Importantly, these effects are highly dependent on TMP concentrations. In fact, for each xenologue, a relatively narrow range of sub-MIC TMP concentrations exists in which the growth effects of codon replacement become pronounced, revealing the high sensitivity of codon bias-related fitness effects of the horizontally transferred antibiotic resistance genes to the environmental levels of antibiotics. The changes in CAIg/tAIg values, GC content, or 5′-end mRNA folding stability upon codon replacement were not predictive of the changes in growth of the xenologous strains. However, changes in both codon optimality (CAIg/tAIg values) and 5′-end mRNA thermodynamic stability were predictive of the changes in *folA* promoter activity.

Competition among Xenologous *folA* Variants with Diversified Codons Intensifies at Sub-lethal TMP Levels

Although it is clear from the obtained data that ORG to MOD codon replacements within xenologous *folA* strains directly affect *folA* promoter activity, intracellular DHFR abundance, host growth, and TMP sensitivity, it was not possible to reliably infer the changes in bacterial growth from changes in codon composition (CAIg/tAIg values or GC content) or mRNA stability, most plausibly because of the small number of xenologous sequences and a rather limited variety of codon replacements (only a single combination of replaced codons was tested). To address this problem, we chose to focus on *folA* genes from *N. sicca* and *L. grayi*. These xenologues differ substantially in their sensitivity to TMP: MIC of TMP in the *E. coli* strain carrying ORG *N. sicca folA* is around 1,600-μg/ml TMP, but only 2.5-μg/ml TMP in case of ORG *L. grayi folA* (table 2, fig. 1*E*, and supplementary fig. S10*D*, Supplementary Material online). To identify the areas in the coding sequences of both genes that are responsible for the most pronounced codon bias-related fitness effects, we performed replacements of the fully modified sequences (MOD) back to ORG codons in stretches of 30 codons targeting the beginning, middle, and end of both genes. The resulting chimeric, original, and fully modified strains were grown in the presence of sub-MIC TMP concentrations, and their fitness was compared. We determined that ≥80% of codon bias-driven fitness effects are localized to codons 1–30 (supplementary fig. S11*A* and *B*, Supplementary Material online). Similar results were obtained when the MOD endogenous *E. coli folA* sequence was restored back to the ORG state (supplementary fig. S11*C*, Supplementary Material online). Having determined that the observed codon bias-driven fitness effects are localized primarily within the first 15–30 codons, we generated six chimeric libraries, two for each of the *folA* sequences from *N. sicca*, *L. grayi*, and *E. coli*. To this end, the synonymous codons found within codons 1–15, or codons 16–30 of the fully modified *folA* sequences (MOD) were diversified back to the original codons in a combinatorial manner. The resulting chimeric gene libraries were used to generate libraries of chimeric xenologous strains by replacing the chromosomal *folA* coding sequence of *E. coli* without perturbing the endogenous regulatory region upstream to the coding sequence (fig. 2*A* and *B*). Deep sequencing revealed that the generated chimeric strains covered between 55% and 98% of the theoretically possible combinations of codons (fig. 2*C*). To limit the possibility for occurrence of de novo mutations, the obtained libraries were subjected to only two serial passages, each 16 h long, at a range of sub-MIC TMP concentrations. At each time point (day0, day1, and day2), the libraries were analyzed by deep sequencing of the *folA* gene (fig. 2*A* and *C*). From changes in the relative frequencies of the individual library variants that occurred during the two consecutive passages (from day0 to day1, and from day1 to day2) (supplementary table S6, Supplementary Material online), we calculated the distribution of fitness effects as well as the changes

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

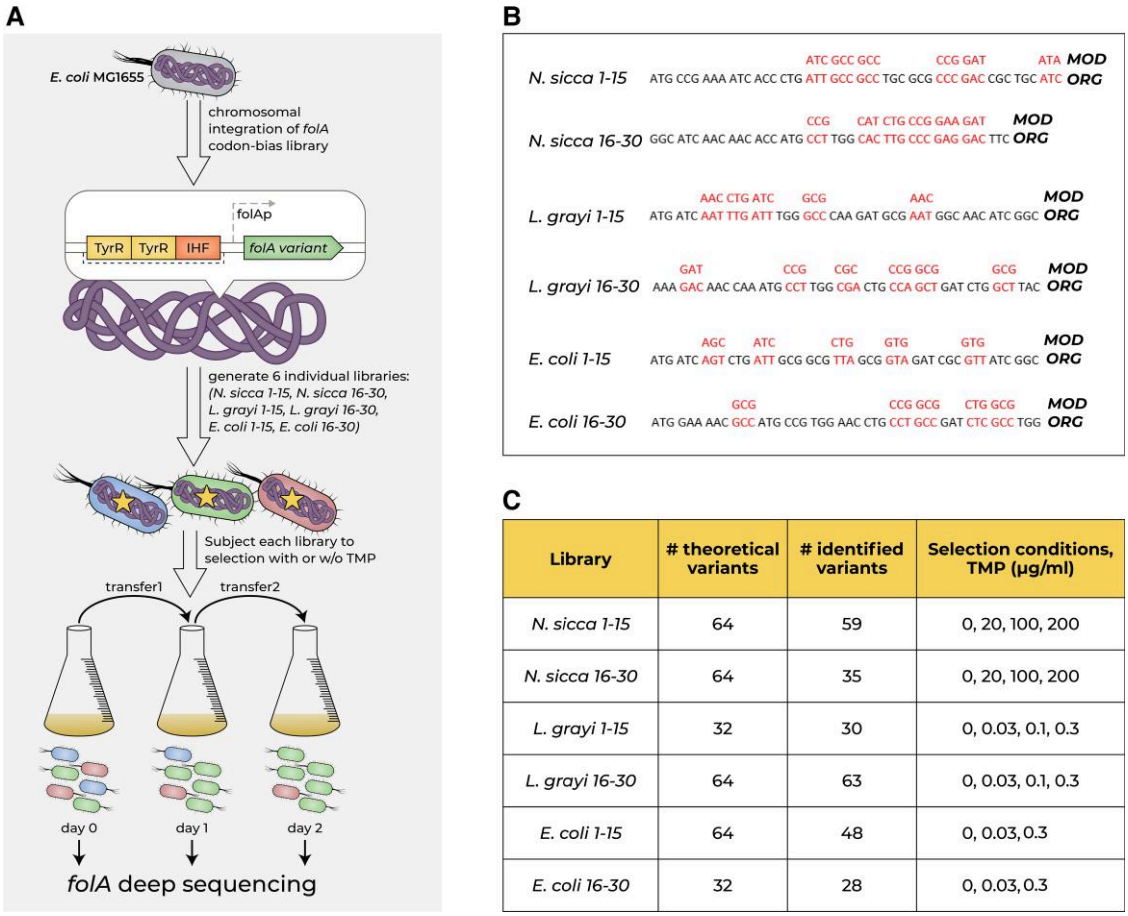
in the population average fitness and population Shannon diversity (see **Materials and Methods**) (fig. 3 *upper panels*). Analysis of the distribution of fitness effects within libraries subjected to selection, calculated as $\log_2$ of fold of change in normalized frequencies of individual variants [$\log_2$(FC)] within a population (see **Materials and Methods** for details), reveals that variants carrying the full combination of the original codons within the diversified areas of *folA* genes were not necessarily the fittest (i.e., reaching the highest frequency increase upon selection), or even more fit than the fully modified variants, which indicates that selection on synonymous codons of the transferred genes has the potential to further improve fitness of the host (fig. 3 *lower panels*). Furthermore, in *L. grayi* and *N. sicca* libraries, selection with increasing amounts of TMP produced a gradual increase in the population average fitness and a decrease in the population diversity after only a single selection passage (from day0 to day1) (fig. 3*A*, *B*, *D*, and *E*). The gradual increase in selection stringency with the increasing TMP levels was apparent from the changes in the distribution of fitness effects of the individual library variants, a higher fraction of which became disadvantageous [i.e., $\log_2$(FC) drops below 0] at higher TMP levels (fig. 3*A*, *B*, *D*, and *E*).

Interestingly, the impact of selection was generally more pronounced in *L. grayi* and *N. sicca* libraries diversified within codons 16–30 than within codons 1–15. For instance, in *N. sicca* libraries 1–15, the average population fitness has increased by 2.3-fold with day0 to day1 passage in the presence of 200 μg/ml TMP versus 6.2-fold in libraries 16–30 (fig. 3*A* and *D*). Such differences can be explained by a stronger clonal interference between variants that comprise libraries 1–15. Clonal interference between multiple variants with comparable fitness effects delays the increase in the population average fitness by suppressing the selection sweeps by individual variants. Indeed, in libraries 1–15 of *N. sicca*, the relative frequency of the most fit variant in the library after a single passage selection in the presence of 200 μg/ml TMP increased to only 19.7% (from the initial frequency of 3.5% at day0). The relative frequency of this variant increased to 30% after the second passage (supplementary table S6, Supplementary Material online). Conversely, under the same conditions, the most fit variant in libraries 16–30 reached a relative frequency of 85% on day1 (from initial frequency of 11.4% at day0). On day2, this variant almost fully swept the population, reaching over 98% in frequency (supplementary table S6, Supplementary Material online). These dynamics are also captured by the changes in population diversity. The normalized Shannon diversity index of *N. sicca* libraries 1–15 (relative to the diversity of the naive library that was set as 1) dropped to 0.63 after a single passage selection with 200-μg/ml TMP (fig. 3*A*). For comparison, the diversity of the *N. sicca* libraries 16–30 under the same conditions dropped to 0.22, nearly a 3-fold decrease (fig. 3*D*). In contrast to *N. sicca* and *L. grayi* libraries, *E. coli* libraries 1–15 and 16–30 did not respond to a single passage selection in the presence of increasing TMP concentrations in a gradual manner, showing only modest changes in the population average fitness and comparable distributions of fitness effects in the absence and presence of TMP (fig. 3*C* and *F*). Both *E. coli* libraries (1–15 and 16–30) also maintained a high Shannon diversity index, even after two selection passages, suggesting that they are subjected to a much higher impact of clonal interference than the xenologous libraries (fig. 3*C* and *F*).

To validate the reproducibility of the reported fitness measurements in the combinatorial libraries, we undertook the following approaches. First, we tested the correlation between fold of change in the normalized frequencies of individual variants in populations subjected to selection under most comparable conditions. Specifically, we compared *L. grayi* libraries 1–15 subjected to selection at 0 versus 0.03 mg/ml TMP, 0.03 mg/ml versus 0.1 mg/ml TMP, and 0.1 mg/ml versus 0.3 mg/ml, and *N. sicca* libraries 1–15 subjected to selection at 0 versus 20 mg/ml TMP, 20 mg/ml versus 100 mg/ml TMP, and 100 mg/ml versus 200 mg/ml. We found significant correlations in all pairs in both *L. grayi* and *N. sicca* libraries (Pearson R ranges from 0.64 to 0.89; $P \leq 0.00013$) (supplementary fig. S12, Supplementary Material online). Importantly, the significance of the correlation increased gradually with the increase in TMP levels. This phenomenon is anticipated, since selection at higher TMP concentration is more stringent, and, therefore, its outcome is more deterministic. Second, we repeated the selection on *L. grayi* libraries 1–15 at 0.03-mg/ml TMP and found a significant correlation in the distribution of $\log_2$(FC) values of the individual library variants between two biological repeats (Pearson $R = 0.58$, $P = 0.0082$) (supplementary fig. S13*A*, Supplementary Material online). Collectively, these findings indicate high reproducibility of the obtained fitness values.

Since the shift in frequencies of competing variants was measured after a relatively short period of selection (the effect of selection became apparent in the xenologous libraries already after a single 16-h long passage), the calculated fitness effects could be, in theory, affected by the bias in the variant frequencies existing within the naive libraries prior to selection. Thus, to estimate the role of the initial variant frequencies on the outcome of selection and to confirm the contribution of clonal interference to the observed evolutionary dynamics, we performed evolutionary simulations using SodaPop (Gauthier et al. 2019) (see **Materials and Methods**). SodaPop is a forward-time simulator of the evolution of large asexual populations that entails a user-defined fitness landscape. The simulations were run using either the experimentally determined initial frequencies of *N. sicca* and *L. grayi* libraries 1–15 and 16–30, or, alternatively, with equalized initial frequencies. We found that the dynamics of changes in the average population fitness and the Shannon diversity of the simulated populations largely recaptured the experimentally obtained dynamics (supplementary fig. S14, Supplementary Material online). The simulation outcomes were found to be insensitive to the initial frequencies of the individual

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024



Library	# theoretical variants	# identified variants	Selection conditions, TMP (μg/ml)
N. sicca 1-15	64	59	0, 20, 100, 200
N. sicca 16-30	64	35	0, 20, 100, 200
L. grayi 1-15	32	30	0, 0.03, 0.1, 0.3
L. grayi 16-30	64	63	0, 0.03, 0.1, 0.3
E. coli 1-15	64	48	0, 0.03, 0.3
E. coli 16-30	32	28	0, 0.03, 0.3

Fig. 2. The competition experiment. (*A*) The experimental setup. (*B*) Codons diversified in each library (marked in *red*). (*C*) The theoretical number of expected codons, total number of the identified codons, and a range of TMP concentrations used in selection are shown for each library.

variants, thus validating the experimental results. Furthermore, the contribution of clonal interference to the observed dynamics of fitness changes in simulated populations was evident from the inverse linear correlation between the average population fitness and the diversity of the evolving populations (supplementary fig. S15, Supplementary Material online). At a higher selection stringency (higher TMP levels), the increase in the average population fitness was accompanied by a steeper decrease in population diversity, as expected under a clonal interference regime (supplementary fig. S15, Supplementary Material online).

These findings show that in the xenologous *folA* libraries, there exist two markedly different evolutionary dynamics, one for variants diversified within codons proximal to the translation start (codons 1–15) and another for variants diversified within codons 16–30. The difference in dynamics stems from the co-existence of multiple variants with comparable fitness effects in libraries 1–15, which delays selection sweeps by individual variants, as opposed to libraries 16–30 that comprised several variants with high fitness values that rapidly take over the populations. The clonal interference in libraries 1–15 also explains the gradual change in the evolutionary dynamics in response to an increase in TMP levels. The fitness effects of individual variants become more dissimilar with the elevation of stringency of selection. It is evident that in *E. coli folA* libraries, clonal interference is overall stronger, although being equally pronounced in both the 1–15 and 16–30 libraries.

mRNA Stability is the Major Determinant of Codon Bias-induced Fitness Effects upon HGT

Next, we determined whether the observed changes in fitness effects of the individual variants upon selection can be explained by their mRNA thermodynamic stability, CAIg/tAIg values, or GC content (supplementary table S7, Supplementary Material online). We dismissed the possibility that the codon bias of the library variants affected fitness by controlling the co-translation folding of DHFR, since the diversified codons in our libraries cover only the first 30 amino acids. The diversified part of the nascent polypeptide chains in its entirety is therefore expected to fill the ribosome tunnel, known to accommodate 30–40 residues (Jacobs and Shakhnovich 2017). In all the libraries diversified within codons 1–15 (*N. sicca* libraries 1–15, *L. grayi* libraries 1–15, and *E. coli* libraries 1–15), we

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

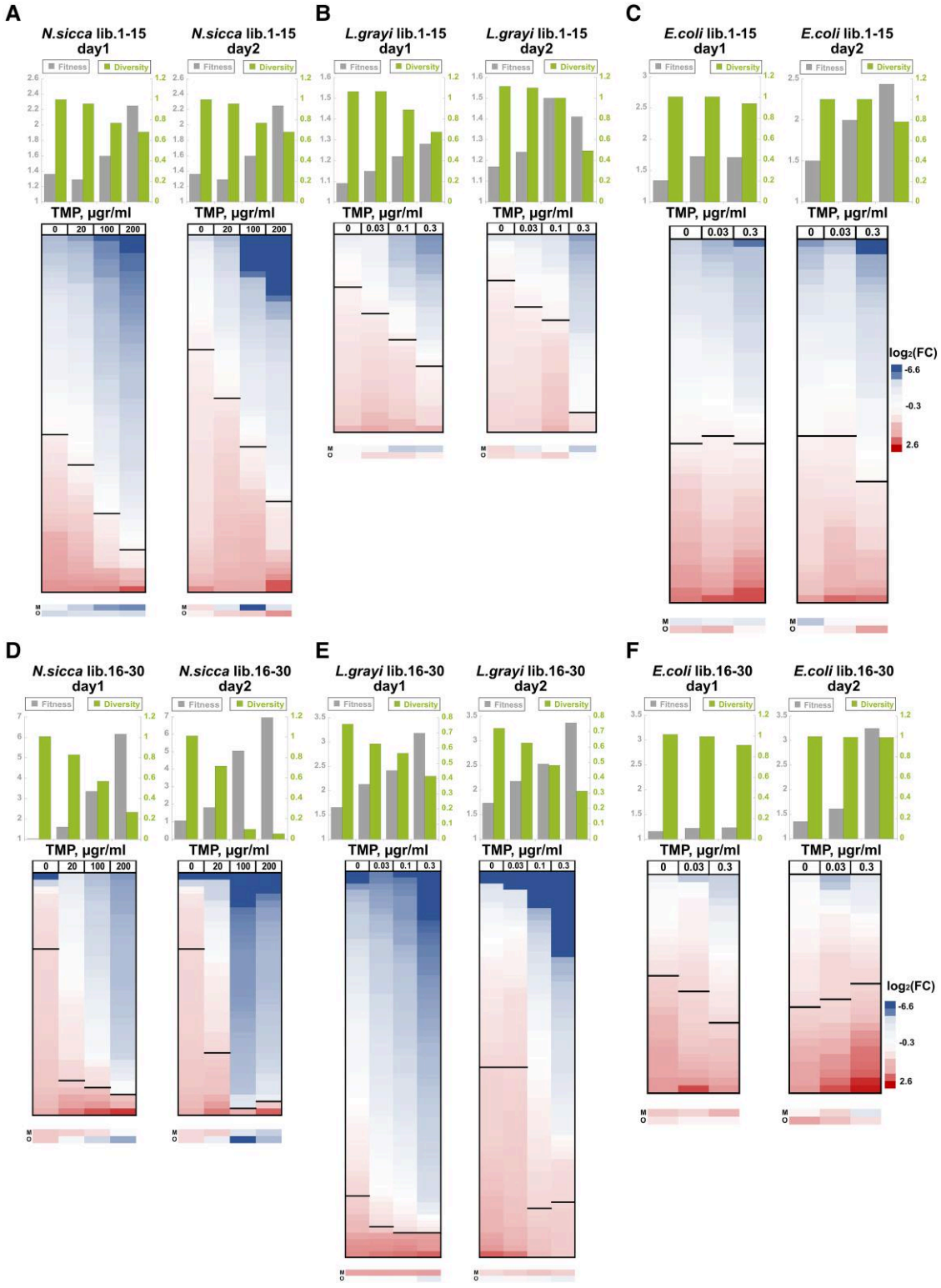


FIG. 3. Codon bias drives population levels effects upon selection on *N. sicca*, *l. grayi*, *E. coli* libraries 1–15 (*A*–*C*) and libraries16–30 (*D*–*F*). *Upper panels*. Changes in the population average fitness (*gray*) and Shannon diversity (*green*) after day1 and day2 of selection in the presence of 0, 20, 100, and 200 µg/ml of TMP in the case of *N. sicca*, 0, 0.03, 0.1, and 0.3 µg/ml of TMP in the case of *L.grayi*, and 0, 0.03, and 0.3 µg/ml of TMP in the case of *E. coli* libraries. *Lower panels*. The log2 of fold of change in the normalized variant frequencies upon selection ($\log_2$(FC)), a proxy to variant fitness, was sorted from lowest (*blue*) to highest (*red*) independently at each selection regime and for each library and presented as heat maps. (Note that variants at each row are not necessarily identical, as their enrichment may change upon change in the selection regime). Black bars separate variants that were depleted upon selection ($\log_2(FC) < 0$) from those that were enriched ($\log_2(FC) > 0$). M, corresponds to variants with fully MOD sequence. O, corresponds to variants in which all synonymous codons encoding amino acids 1–15 are represented by ORG codons. The heat map ruler indicates $\log_2$(FC) of the individual variants.

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

established a strong correlation between the fitness of individual variants and their mRNA stability. The mRNA folding stability was calculated using either a single 145-nt long fragment covering the diversifications in codons 1–30 (from nucleotide −25, the beginning of transcription, and up to nucleotide +120) or via a 30-nucleotide long sliding window starting from the mRNA transcription start site (nucleotide −25), in steps of 1 nt for a total of 110 windows (see **Materials and Methods**). Regardless of the method used for calculating mRNA folding stability, in all three libraries diversified within codons 1–15, selection unequivocally favored variants with less stable mRNA (fig. 4A, supplementary figs. S16A, S17A, S19A, S20A, and S22A, Supplementary Material online). Furthermore, the correlation between the folding stability of mRNA transcripts and the fitness effects was highly dependent on TMP concentration. For example in *N. sicca*, up to 47% in fitness variance could be explained by mRNA stability at 200 μg/ml TMP upon day1 of selection (Spearman's correlation $r = 0.68$; $P = 2.54{\cdot}10^{-9}$), whereas no correlation between mRNA stability and fitness was detected at 0 TMP (fig. 4A). Similar trends to those observed after day1 of selection were also detected when fitness effects were calculated for day2 (fig. 4A, supplementary fig. S18A and S21A, Supplementary Material online).

In addition to mRNA stability, the changes in fitness effects in libraries 1–15 also correlated, albeit much less significantly, with CAIg values in *N. sicca* (Spearman's correlation $r = -0.4$; $P = 0.002$; 200 μg/ml TMP, day1) (fig. 4B), with CAIg/tAIg values and GC content in *L. grayi* (Spearman's correlation $r = -0.42$; $P = 0.02$; 0.3 μg/ml TMP, day1 for tAIg-fitness correlation) (supplementary fig. S18A, Supplementary Material online), and with GC content in *E. coli* (Spearman's correlation $r = -0.42$; $P = 0.03$; 0.03 μg/ml TMP, day1) (supplementary fig. S20A, Supplementary Material online).

Which of these parameters, mRNA stability, codon optimality (expressed in either CAI or tAI scales), or GC content, then constitute the primary driver of fitness effects observed in libraries 1–15? To address this question, we first note that the correlation between fitness and CAIg/tAIg values is inverse, indicating that variants with more optimal codons are less fit, which contradicts the assumptions that codon optimality should produce more efficient translation and, therefore, higher TMP resistance (fig. 4B, supplementary figs. S17A, S18A, S20A and S21A, Supplementary Material online). Second, we note that prior to selection (day0), there exists a significant inherent inverse correlation between mRNA stability and either CAIg or tAIg values among variants of the naive libraries (fig. 4C, supplementary figs. S17C and S20C, Supplementary Material online) so that less stable mRNA variants have lower CAIg/tAIg values. The same inherent linkage between codon optimality and mRNA stability at gene starts was demonstrated for most bacterial strains with GC content higher than 50% (Bentele et al. 2013) and is explained by the simple fact that mRNA stability increases with increasing GC content. Consequently, selection to reduce secondary mRNA structure favors the incorporation of AU-rich codons. Such codons also happen to be rare (i.e., less optimal) in bacteria with %GC >50. Thus, synonymous codons found in the vicinity of gene starts are sub-optimal primarily because they control translation by reducing the secondary structure of mRNA transcripts, and not because low codon optimality per se is advantageous at the starts of genes (Bentele et al. 2013; Goodman et al. 2013). This statement is supported by a strong inverse correlation between mRNA stability and GC content and/or by strong positive correlation between CAIg/tAIg values and GC content among naive variants of libraries 1–15 (supplementary figs. S17C–E and S20C–E, Supplementary Material online). The similarity in correlation trends between the CAIg and tAIg values is also explained by the strong inherent correlation between them in most naive libraries (fig. 4D, supplementary figs. S17F and S20F, Supplementary Material online). It is, therefore, plausible that the observed selection on codon bias in libraries 1–15 is driven primarily by the thermodynamic stability of mRNA transcripts, specifically by favoring variants with reduced secondary structure. The accompanying inverse correlation of fitness effects with codon optimality and GC content most probably also stems from the inherent correlation between stability, codon optimality, and GC content of the library variants prior to selection (fig. 4E). To validate the reproducibility of our findings, we determined that both biological repeats of selection on naive *L. grayi* libraries 1–15 in the presence of 0.03 μg/ml TMP up to 50% in variability of fitness effects among individual variants can be explained by their mRNA stability near the translation start. Moreover, we found a strikingly similar pattern of distribution of correlations between variant fitness and mRNA folding stability between both biological repeats (Spearman $R = 0.96$, $P = 0$) (supplementary fig. S13B, Supplementary Material online). To further validate that our conclusions are not a result of the bias in the variant frequencies in the naive libraries, we introduced a bias of up to 10-fold in the initial frequency of variants in *N. sicca* libraries 1–15 (supplementary fig. S18A, Supplementary Material online) and repeated the selection at 0 and 100 μg/ml TMP (supplementary fig. S23B, Supplementary Material online). Similar to the results obtained with the unperturbed library, we found that the observed fitness effects in the presence of TMP were best explained by 5′-mRNA stability, with less stable variants being favored by selection (Spearman's correlation $r = 0.47$; $P = 0.62{\cdot}10^{-5}$; 100 μg/ml TMP, day1) (supplementary fig. S23C, Supplementary Material online). Furthermore, fitness effects also correlated with CAIg values, but as previously observed, the correlation was inverse and less significant than the one obtained with mRNA stability (Spearman's correlation $r = -0.32$; $P = 0.02$; 100 μg/ml TMP, day1) (supplementary fig. S18C, Supplementary Material online).

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

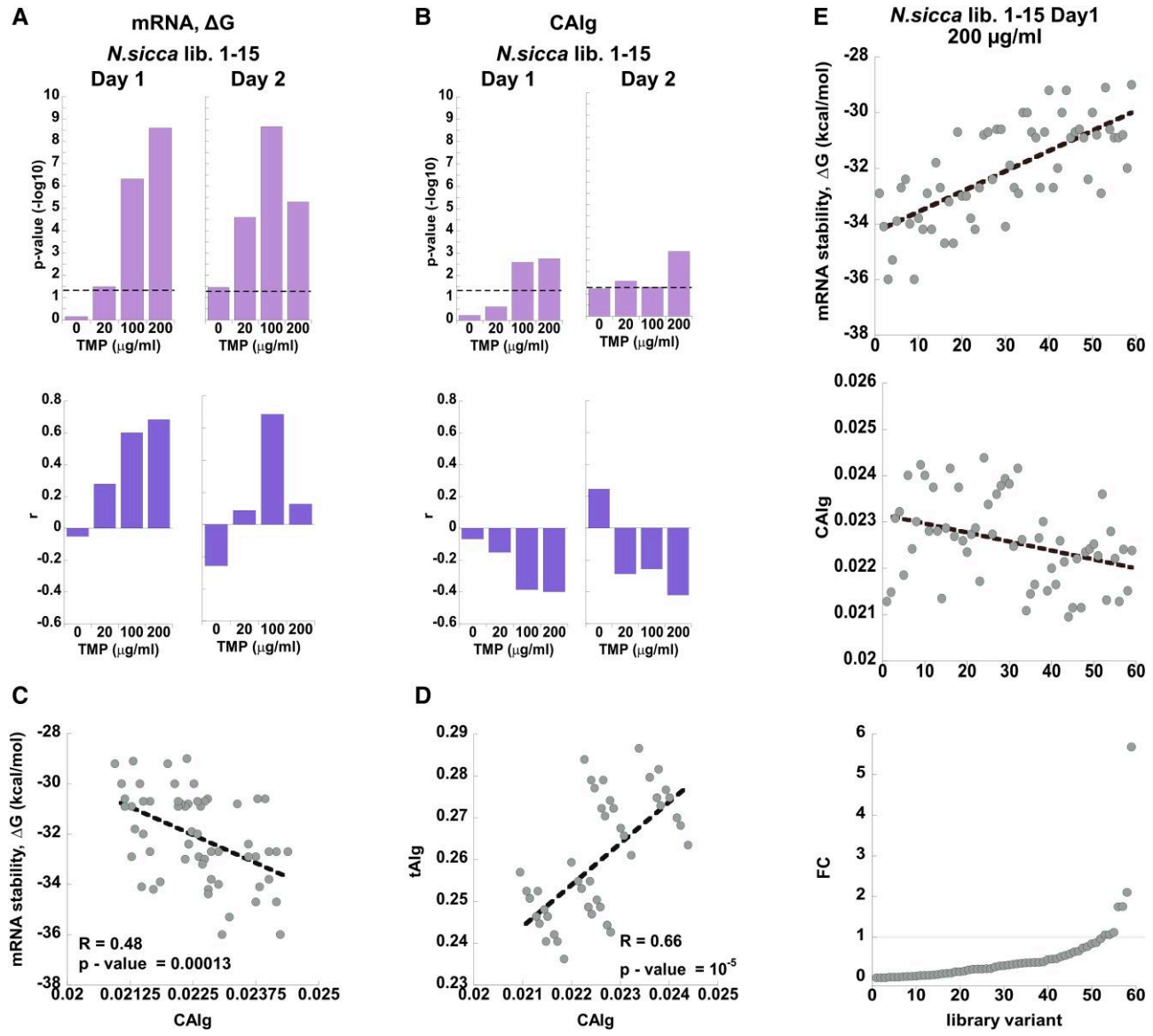


FIG. 4. The selection on *N. sicca* libraries 1–15 is driven by 5′-mRNA stability. (*A*, *B*) *Upper panels*. *P*-values for Spearman non-parametric test for the association between the fitness effects and (*A*) the mRNA stability (calculated for a single fragment spanning 114 nucleotides from nucleotide −25, the beginning of transcription, and up to nucleotide +90), and the fitness effects and (*B*) the codon optimality (CAIg). *Lower panels*. Corresponding Spearman's correlation coefficients, *r*. (*C*, *D*) Pearson's correlations between (*C*) the mRNA stability and codon optimality (CAIg) and (*D*) between CAIg and tAIg values for variants comprising naive libraries (day0). Pearson's *P*-values and *R*-values are indicated. (*E*) *Lower panel*. All library variants are sorted according to fold of change (FC) in normalized frequency (from lowest to highest values) upon selection at 200-μg/ml TMP, day1. *Middle panel*. CAIg values of variants are shown according to their order in the lower panel. *Upper panel*. mRNA stability values of variants are shown according to their order in the lower panel.

Analysis of the libraries diversified with codons 16–30 revealed that in *L. grayi* and *E. coli*, mRNA stability remains the most dominant force underlying the observed fitness effects (supplementary figs. S15*B*–S22*B*, Supplementary Material online). However, in both libraries, the significance of the mRNA stability–fitness correlation was substantially reduced in comparison to that of libraries 1–15. Depending on the day of selection, the correlation passed the significance threshold only when mRNA stability was measured using a 30-nt-long sliding window, and only in some of the selection conditions (supplementary figs. S19*B* and S22*B*, Supplementary Material online). Similar to dynamics observed in libraries 1–15, changes in the fitness effects in *L. grayi* and *E. coli* libraries 16–30 were inversely correlated with CAIg, tAIg, and/or GC content on either day1 or day2 of selection (supplementary figs. S17*B*, S18*B*, S19*B*, and S21*B*, Supplementary Material online). As discussed above, this behavior can be explained by the inherent correlations that exist among the naive library variants, namely those between mRNA stability and CAIg values (supplementary figs. S12*G* and S15*G*, Supplementary Material online), mRNA stability and GC content (supplementary figs. S12*H* and S15*H*, Supplementary Material online), CAIg and GC content (supplementary

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

figs. S12*I* and S15*I*, Supplementary Material online), and tAIg and CAIg values (supplementary figs. S12*J* and S15*J*, Supplementary Material online).

Analysis of the *N. sicca* libraries 16–30 produced markedly opposite trend to that obtained with *L. grayi* and *E. coli* libraries 16–30. In *N. sicca*, no apparent correlation between mRNA destabilization and improved fitness could be observed neither on day1 nor on day2 of selection (fig. 5*A*). In fact, when mRNA folding stability was calculated using a 30-nt sliding window, more stable variants appeared to be weakly favored by selection (windows #40–50 on day1, and windows #70–80 on day2, supplementary fig. S11*B*, Supplementary Material online). In addition, fitness effects obtained upon selection at 20 μg/ml TMP on day1 and at 100 and/or 200 μg/ml TMP on day2 significantly correlated with codon optimality. As evidenced by tAIg and/or CAIg metrics, or by GC content, variants that carry more optimal codons or those that exhibit a higher GC content were found more fit (Spearman's correlation $r = 0.25$, $P = 0.009$ for tAIg-fitness correlation; $r = 0.55$ $P = 0.0007$ for GC content-fitness correlation on day1 and $r = 0.5$, $P = 0.003$ for CAIg-fitness correlation; $r = 0.05$, $P = 0.002$ for tAIg-fitness correlation; $r = 0.45$ $P = 0.006$ for GC content-fitness correlation on day2) (fig. 5*B*, supplementary fig. S24*A* and *B*, Supplementary Material online). The positive correlation between codon optimality and fitness effects directly supports the notion that codon optimality controls translation efficiency and, therefore, is favored by selection. The similarity between correlation trends obtained with CAIg and tAIg values is not surprising, given that both metrics are highly correlated among the naive variants (fig. 5*D*). Notably, no inherent correlation is observed between mRNA stability and codon optimality in variants that make up the naive library (fig. 5*C*). It, therefore, appears that in *N. sicca* libraries 16–30, selection towards codon optimality was the driving force behind the observed changes in fitness effects (fig. 5*E*). The correlation between fitness effects and GC content is also most likely driven by codon optimality due to the significant positive correlation observed between the tAIg values and GC content among the naive variants comprising the library (supplementary fig. S24*C*, Supplementary Material online).

Collectively, the presented findings suggest that both mRNA stability and codon optimality contribute to the compatibility of the transferred *folA* genes. However, their impact is highly dependent on the environmental conditions, that is, the concentration of TMP. Although *N. sicca* and *L. grayi* DHFR differ substantially in TMP sensitivity, there exists a range of sub-MIC TMP concentrations for each xenologous strain whereby codon bias-related fitness effects may be predicted directly from sequence changes. mRNA stability dominates the codon bias-related cause of fitness changes. When HGT leads to 5′-end overstabilization of mRNA transcripts, mRNA stability outweighs the fitness contribution of codon optimality.

5′-end Overstabilization of mRNA Transcripts Can Cause mRNA Accumulation outside the Polysome

Having established the dominant role of 5′-end mRNA folding stability in codon bias compatibility of the transferred genes, we were interested to see whether changes in mRNA stability can affect mRNA abundance. To this end, we measured the changes in the steady-state mRNA abundance upon synonymous codon replacement in all *folA* xenologous strains (fig. 6*A*). In the absence of TMP, the effect of codon replacement on mRNA levels was highly variable. Some xenologues genes exhibited a severe to mild decrease in mRNA levels, if any change was observed at all (*N. sicca*, *M. capsulatus*, *P. putida*, and *M. mesenteroides*). Other xenologues exhibited an increase in mRNA levels (*B. subtilis*, *L. lactis*, and *L. grayi*). The addition of TMP had no effect on this trend (supplementary fig. S25*A*, Supplementary Material online). Surprisingly, the increase in the steady-state levels of mRNA transcripts upon codon replacement was not necessarily matched by a corresponding increase in protein levels (fig. 1*B* and supplementary fig. S26, Supplementary Material online). For instance, app. 5-fold increase in mRNA abundance of *folA* from *L. lactis* and *L. grayi* was accompanied by a 4- to 6-fold decrease in the protein levels (figs. 1*B* and 6*A*).

How can such an inconsistency be explained? Since the inefficient ribosome translation elongation triggered by codon usage-incompatible mRNA is expected to cause transcript decay (Presnyak, et al. 2015), most plausibly through the involvement of Rho factor (Richardson 1991), we hypothesized that the lack of mRNA decay observed on the background of inefficient translation is a result of the reduced engagement of mRNA transcripts with translating ribosomes. In other words, mRNA transcripts escape degradation and accumulate due to the reduced participation in translation. Thermodynamic overstabilization of the endogenous *folA* mRNA structure near the translation start codon was indeed shown to impede translation initiation by hindering the accessibility to Shine–Dalgarno binding site in *E. coli* (Bhattacharyya et al. 2018). Although in this example, the mRNA transcripts unengaged in translation were efficiently degraded, it is plausible that the overstabilization of the 5′-end of some of the xenologous *folA* transcripts not only limits the accessibility of the ribosomal binding site but also conceals potential cleavage sites recognized by endonucleases, thus reducing the degradation propensity of the transcript (Rauhut and Klug 1999). To test this hypothesis, we extracted the total and polysomal mRNA fractions from mid-exponentially growing MOD and ORG *L. grayi* strains and quantified the ratios of the polysomal and the total *folA* mRNA steady-state levels in both strains (see supplementary fig. S25*B*, Supplementary Material online and **Materials and Methods**). As hypothesized, we observed a significant shift of the MOD *L. gayi folA* mRNA transcripts from the polysomal to the total fraction, indicating that the accumulated transcripts indeed avoid translation (fig. 6*B*). The aggregation resistance of the MOD

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

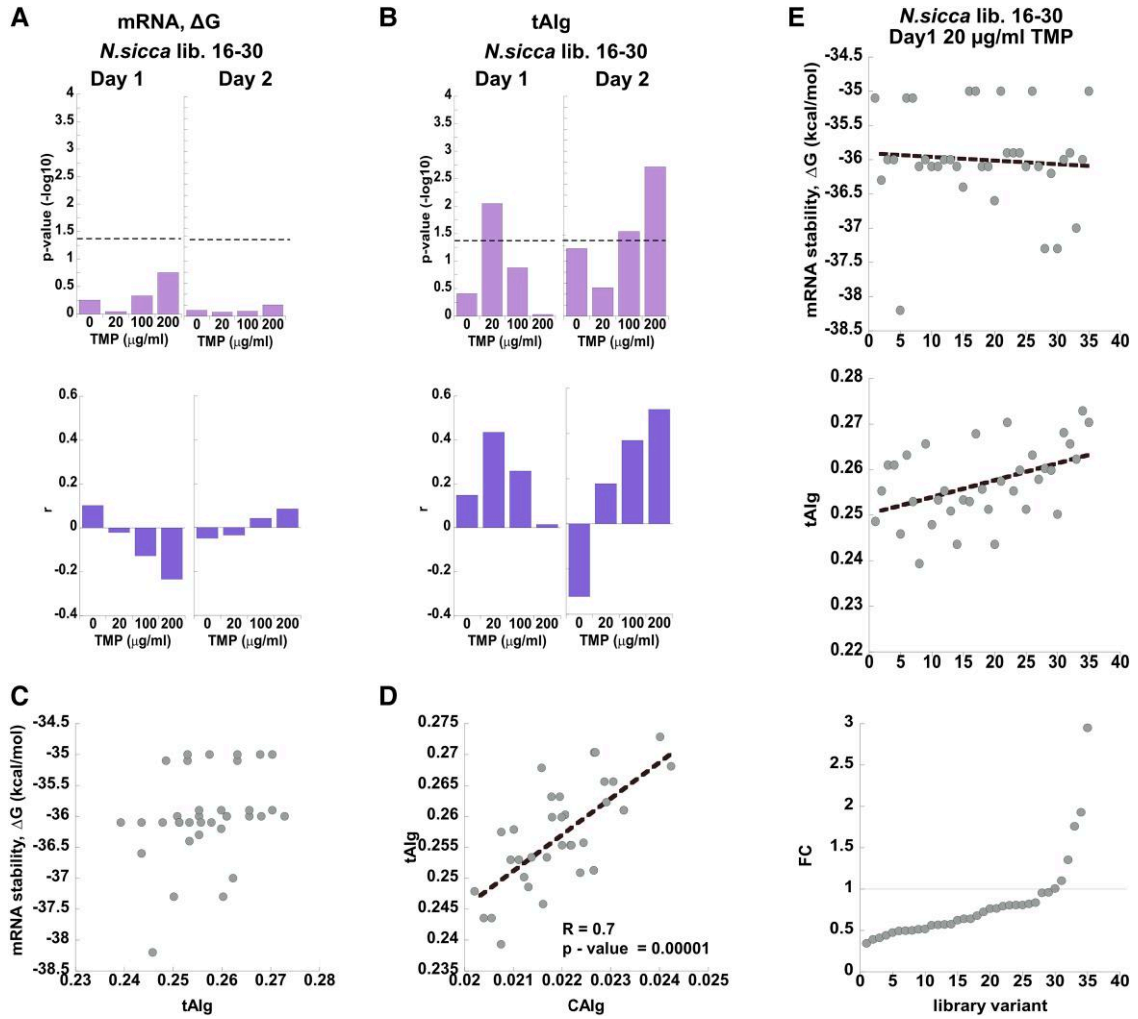


Fig. 5. The selection on *N. sicca* libraries 16–30 is driven by codon optimality. (*A*, *B*) *Upper panels*. *P*-values for Spearman non-parametric test for association between fitness effects and (*A*) mRNA stability (calculated for a single fragment spanning 114 nucleotides from nucleotide −25, the beginning of transcription, and up to nucleotide +90), and fitness effects and (*B*) codon optimality (tAIg). *Lower panels*. Corresponding Spearman's correlation coefficients, *r*. (*C*, *D*) Correlations between (*C*) mRNA stability and codon optimality (CAIg) and (*D*) between CAIg and tAIg values for variants comprising naive libraries (day0). Pearson's *P*-values and *R* values are indicated only for the significant correlation ($P < 0.05$). (*E*) *Lower panel*. All library variants are sorted according to fold of change (FC) in normalized frequency (from lowest to highest values) upon selection at 20-μg/ml TMP, day1. *Middle panel*. tAIg values of variants are shown according to their order in the lower panel. *Upper panel*. mRNA stability values of variants are shown according to their order in the lower panel.

transcript may result from the overstabilization of its 5′-end by 5 kcal/mol (supplementary fig. S7*F*, Supplementary Material online).

Our findings suggest that mRNA decay is not solely the immutable outcome of the codon bias-induced reduction in translation efficiency of the horizontally transferred genes, as was previously thought. Inefficient translation and transcript accumulation can occur simultaneously, resulting from the poor engagement of transcripts by translating ribosomes. Thus, steady-state mRNA abundance cannot serve as a reliable predictor of translation efficiency in the case of HGT.

Discussion

Previous works by Kudla et al. (2009) and Goodman et al. (2013) have demonstrated that an increase in mRNA folding stability at the starts of heterologous genes has a detrimental effect on protein abundance but not fitness of *E. coli* cells carrying the genes. Conversely, codon optimality of the heterologous genes was shown to have a profound impact on fitness of the host but did not correlate with the abundance of their protein products. The importance of codon optimality in HGT was also established from the analyses of compatibility between codon usage of the donor and tRNA pools of the host (Medrano-Soto et al. 2004; Tuller et al. 2011). These pioneering studies, however, left unanswered the following important questions. First, does mRNA folding stability control the success of HGT events? Addressing this question is particularly important in light of the finding that mRNA stability of most genes is significantly reduced near the start codons in all tested species (Gu et al. 2010). Consequently, the 5′-end mRNA stability of genes that are being transferred horizontally

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

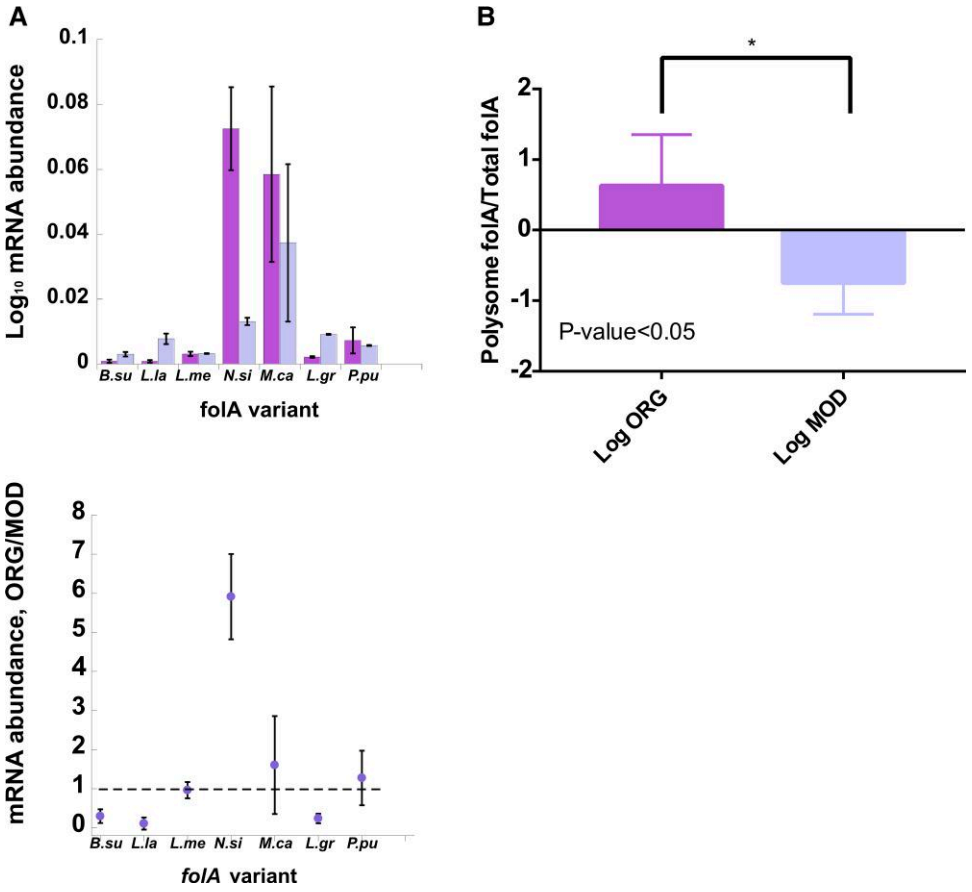


FIG. 6. The effect of codon usage of *folA* xenologoues on mRNA abundance in *E. coli* host. *folA* variants are marked as in table 1. (*A*) *Upper panel*. Intracellular steady-state levels of original (ORG) and modified (MOD) *folA* mRNA. *Lower panel*. Ratios of *folA* mRNA abundance between ORG and MOD *folA* sequences. (*B*) Log ratios between *L. grayi folA* mRNA abundance found within the polysomal and total fractions calculated for ORG and MOD sequences. Error bars were calculated from four independent measurements.

is expected to be reduced as well, hinting that its contribution to HGT might be insignificant. Second, if both mRNA stability and codon optimality are nonetheless at play, can we understand their individual roles in HGT? By using an experimental system that establishes a direct link between the variability in codon usage of the transferred genes and fitness of the host, we were able to provide definitive answers to both these questions. We found that despite the universal reduction in mRNA stability near gene starts, the combination of the 5′-untranslated regulatory region of the host with synonymous codons near the start codon of the horizontally transferred genes may cause a substantial overstabilization of the resulting chimeric mRNA structures. When this occurs, mRNA stability overrides the fitness contribution of codon optimality and becomes the most dominant cause of the codon bias-related fitness effects in HGT. Specifically, the fitness effects of codon bias within codons 1–15 of combinatorial *folA* libraries were dominated by mRNA stability—selection strongly favored destabilized mRNA variants. Since mRNA destabilization is achieved by codons with reduced GC content, and because such codons also happen to be rare in bacteria with GC content >50%, selection of variants with less stable 5′-end mRNA structures was accompanied by an apparent enrichment of variants composed of less optimal codons. Importantly, in *N. sicca* and *L. grayi* libraries 1–15, the selection of less stable variants was acutely dependent on the environmental conditions, that is, levels of TMP in the growth media. Indeed, in both libraries, no enrichment of less stable mRNA was apparent in the absence of TMP. The selection intensified, however, following only one day of exposure to sub-lethal levels of TMP. Since *N. sicca* and *L. grayi* DHFRs differ in TMP sensitivity, the range of TMP concentrations that triggered selection for less stable mRNA variants did not overlap between the libraries (0.03–0.3 μg/ml TMP in the case of *L. grayi* vs. 20–200 μg/ml TMP in the case of *N. sicca*).

Conversely, we found that when the contribution of mRNA stability to fitness is limited, codon optimality assumes the leading role. Thus, for the first time, we were able to demonstrate a clear hierarchy existing between the two parameters. Indeed, the dominating role of mRNA stability was challenged by codon optimality in *folA* libraries diversified within codons 16–30. Whereas in *L. grayi* libraries 16–30, mRNA stability remained the main cause of codon bias fitness effects, in *N. sicca* libraries 16–30, the variability in fitness effects was best explained by the contribution of optimal codons, defined either as CAI or tRNA adaptation index (tAI), with variants composed of optimal codons being favored by selection. As with mRNA stability, the contribution of codon optimality to codon bias-related fitness effects was apparent only within a relatively narrow range of TMP concentrations (fig. 5*B*). The change in TMP sensitivity upon codon replacement from original to frequent codons performed throughout the entire sequence of *folA* xenologoues also

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

supports the dominating role of the 5′-end of mRNA stability over codon optimality. Indeed, although codon replacement led to an increase in codon optimality (either CAIg or tAIg values) in all seven xenologous strains (supplementary fig. S6A and B, Supplementary Material online), only in the *E. coli* strain carrying MOD *B. subtilis* such a replacement was accompanied by an increase in resistance to TMP (supplementary fig. S10A, Supplementary Material online and table 2). This xenologue also happened to be the only one that exhibited a substantial destabilization (around 5 kcal/mol) of its 5′-end mRNA structure upon codon replacement (supplementary fig. S7A, Supplementary Material online). Thus, unless 5′-end mRNA is destabilized, the contribution of codon optimality to fitness in the xenologous strains is subdued.

Quite intriguingly, we also found that 5′-end mRNA overstabilization of the transferred genes in some instances can lead to an accumulation of mRNA transcripts outside of the polysome, suggesting that codon bias-induced reduction in translation efficiency is not always accompanied by mRNA decay. Thus, mRNA abundance cannot serve as a reliable proxy to translation efficiency of foreign mRNA transcripts.

Our findings suggest that although selection operates universally to reduce the thermodynamic stability at gene starts in all donor organisms, HGT may nonetheless result in the overstabilization at the beginning of mRNA transcripts and, consequently, in reduced production of foreign proteins within the host bacteria. In this scenario, the mRNA stability, which controls translation initiation, becomes the most dominant codon bias-related effect in HGT, superseding the impact of codon optimality on controlling translation elongation. This conclusion supports the idea that translation initiation constitutes the rate-limiting step of protein synthesis by the ribosome.

Materials and Methods

Chromosomal Replacements of the *folA* Xenologues

The chosen xenologous *folA* gene sequences fused to a tag encoding six histidines (table 1, supplementary tables S1 and S4, Supplementary Material online) were synthesized (Integrated DNA Technologies), cloned into vector pKD13 specifically designed for λ Red homologous recombination of *folA* genes (Bershtein et al. 2012), and integrated into *E. coli* chromosome by replacing the endogenous *folA* coding sequence but preserving the upstream endogenous regulatory region. Recombinant strains were selected on LB agar plates supplemented with 25 μg/ml chloramphenicol and 35 μg/ml kanamycin, and chromosomal integration was validated by Sanger sequences, essentially as described (Bershtein et al. 2012).

Media and Growth Conditions

All variants were grown overnight from a single colony in 5-ml M9 minimal media salts supplemented with 0.2% glucose, 0.1% casamino acids, 1-mM $MgSO_4$, 0.5-μg/ml thiamine at 37 °C. Bacterial cultures were then diluted 1/100 and grown at 37 °C for 12 h in 96-well microtiter plates (16 wells for each strain) in the absence or presence of a range of concentrations of trimethoprim (Sigma). A total of 600-nm OD data were collected at 10-min intervals. The resulting growth curves were fit to a bacterial growth model to obtain growth rates parameters (Zwietering et al. 1990).

IC_{50} Determination

Growth was quantified by integration of the area under the growth curve (OD vs. time) between 0 and 15 h, as described in Rodrigues et al. (2016). Growth integrals determined for a given variant were normalized in respect to the corresponding growth of that mutant measured in the absence of TMP. IC_{50} values were determined from the fit of a logistic equation to plots of growth versus TMP concentrations.

Intracellular Protein Abundance Measurements

All variants were grown overnight from a single colony in 5-ml supplemented M9 minimal media at 37 °C. Bacterial cultures were then diluted 1/100 and grown at 37 °C for about 4 h until the cells reached OD ~ 0.5. Cells were lysed with BugBuster (Novagen), and the total protein concentration within soluble fractions was determined by BCA (CYANAGEN). Lysates were then normalized by total protein abundance, loaded on Ni-NTA 0.2-ml column (G-BIOSCIENCES), and DHFR proteins were isolated using manufacturer's instructions. The intracellular DHFR amounts within soluble fractions were eventually determined by SDS-PAGE followed by standard Western Blot protocol using mouse anti-His monoclonal antibodies (abm) and goat anti-mouse HRP antibodies (abm).

Intracellular mRNA Abundance Measurements

All variants were grown overnight from a single colony in 5-ml supplemented M9 minimal media at 37 °C. Bacterial cultures were then diluted 1/100 and grown at 37 °C for about 4 h, until cells reached OD ~ 0.6. A total of 1.5 ml of the cells were treated with RNAprotect Bacteria Reagent (QIAGEN). Total RNA was extracted using GeneJET RNA purification kit (Thermo scientific), followed by DNAse I digestion (Thermo scientific). Total RNA concentration was determined by nanodrop (DeNovix). A total of 400 ng of the purified RNA was subjected to iScript cDNA synthesis (BIO-RAD). Abundance of *folA* mRNA was estimated through quantitative PCR using SYBR green kit (KAPABIOSYSTEMS) (supplementary table S5, Supplementary Material online, primers# 1–2). 16S rDNA gene was used as an internal reference (supplementary table S5, Supplementary Material online, primers# 3–4).

Promoter Activity Measurements

MG1655 strains were transformed with pUA-66 plasmid containing *folA* promoter fused to green fluorescent protein-coding sequence (Zaslaver et al. 2006). The strains

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

were grown overnight in 5-ml M9 minimal media supplemented with 0.2% Glucose, 1-mM $MgSO_4$, 0.5-μg/ml thiamine, 0.1% casamino acids and, diluted 1/100 and grown at 37 °C for 12 h in 96-well microtiter plates (16 wells for each strain) in the absence or presence of IC50 levels of trimethoprim (see table 2). A total of 600-nm OD and GFP fluorescent signal (excitation 485 nm and emission 517 nm) data were collected at 10-min intervals. The ratio between the fluorescent signal and biomass production (OD) was defined as promoter activity (supplementary table S5, Supplementary Material online).

mRNA Folding Stability Calculation

mRNA folding stability (Gibbs free energy difference between folded and folded state, ΔG, kcal/mol) was calculated for either a single fragment spanning 145 nucleotides (from nucleotide −25, the beginning of transcription, and up to nucleotide +120) to cover the diversified codons 1–30, or using a 30-nucleotide long sliding window (to reflect the size of a ribosome footprint) starting from the mRNA transcription start (nucleotide −25) in steps of 1 nt. mRNA stability calculations were performed with unafold software (Markham and Zuker 2008).

Generation of Libraries With Codon Diversification

Fully modified (MOD) *folA* sequences from *E.coli*, *L.grayi* and *N.sicca* were diversified back to the original codons (ORG) within codons 1–15 or 16–30 in combinatorial manner in positions indicated in figure 2*B*. The diversification was carried out by PCR performed with a mixture of primers diversified in the designated positions (primers #5–15, supplementary table S8, Supplementary Material online). The products of PCR for each library were then amplified by primers # 17–18 (supplementary table S8, Supplementary Material online), cloned into vector pKD13 specifically designed for λ Red homologous recombination of *folA* genes (Bershtein et al. 2012), and integrated into *E. coli* chromosome by replacing the endogenous *folA* coding sequence while preserving the upstream endogenous regulatory region, essentially as described (Bershtein et al. 2012). Integrants for each library were selected on LB agar plates supplemented with 25-μg/ml chloramphenicol and 35-μg/ml kanamycin. In total, between 685 and 2,800 individual colonies were obtained for each library. The colonies were collected from the LB agar plates by scraping in the presence of 10-ml LB and cryopreserved at −80°C in the presence of 50% glycerol.

Laboratory Evolution Experiment

The evolution experiment was performed for 16 h over two consecutive days at a range of TMP concentrations (fig. 2C). The cryopreserved libraries (designated as day0) were defrosted, diluted 1/100 in 50 ml of supplemented M9 media, and grown for 16 h at 37 °C, 220RPM in a 250-ml Erlenmeyer flask (designated as day1 of selection). This was followed by another 1/100 dilution in a fresh 50 ml of supplemented M9 media and growth for additional 16 h at 37 °C, 220RPM in a 250-ml Erlenmeyer flask (designated as day2 of selection). Aliquots from each time point were collected and cryopreserved at −80 °C with addition of 50% glycerol. The frequencies of each variant in the libraries were then analyzed by deep sequencing in three time points (day 0—initial frequencies, day1—after 16 h of evolution, and day 2—after 32 h of evolution).

Deep Sequencing

Sample preparation for the deep sequencing had three steps: first, *folA* gene from the evolved and naive populations was amplified by PCR. Ten microliters of each aliquot were diluted in 100-μl ddw, and 1 μl of the template was amplified in 50-μl PCR reactions supplemented with 4 μl 2.5 mM dNTPs, 1 μl (250 U) PrimeStarGXL (TAKARA), and 5 μl of 10 μmol/μl "FWD 2 Deep seq" and "Rev Deep seq" primers (supplementary table S8, Supplementary Material online). Second, the PCR products were run and purified from 1% agarose gel (Machary-Nagel) and used as template for the next amplification reactions performed in 50 μl with 5 μl i5 and i7 from Nextera XT DNA library preparation kit (Illumina) (see supplementary table S9, Supplementary Material online for Nextera primers used), 1 μl (250 U) PrimeSTAR GXL DNA polymerase (TAKARA), 4 μl 2.5 mM dNTPs, and 55 ng DNA template DNA. The PCR program was: 94 °C for 5 min (one cycle), followed by 12 cycles of 95 °C for 10 s, 55 °C for 15 s and 68 °C for 45 s, and a single cycle of 68 °C for 5 min. Beckman Coulter Life Sciences is the manufacturer of the kit used for PCR purification. The deep sequencing was performed by Nextera Miseq reagent micro kit v2 by MiSeq platform (Illumina). Raw reads (FASTQ files) were trimmed using fastp tool (Chen et al. 2018). Next, any sequences in the mutagenized library that were shorter than the expected length were discarded. The generated fastq files were then analyzed with Enrich2 software (Rubin et al. 2017) to determine the read count for each variant using the "count only" scoring option and the "wild type" normalization parameter.

Population Analysis

The frequency of each variant at a specific time point (t) was first normalized by the division of the number of reads by the total number of read at this time point. The fold of change (FC) of each variant was then calculated by dividing the variant's normalized frequency under selection by its normalized frequency before selection. Specifically, the effect of selection after first growth passage on each variant (i) is defined as the normalized frequency (f) obtained on day1 divided by the normalized frequency on day0, $FC_{day1} = f_{i(day1)}/f_{i(day0)}$. Consequently, the fold change in frequency of individual variants upon second passage of selection is defined as $FC_{day2} = f_{i(day2)}/f_{i(day1)}$. To account for the population growth between two time points, we calculated $\log_2$ of fold change, $\log_2(FC)$ independently

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

for day1 and day2. We treat the distribution of $\log_2$(FC) values as a proxy to distribution of fitness effects of individual variants in a population. A proxy to population average fitness at each time point was determined as weighted average of the enrichment scores, $\overline{w} = \sum_i f_{i,\text{day}} * \text{FC}$. The value of $\overline{w}$ at day 0 was defined as 1. The population Shannon diversity (H) was calculated as $H = -\sum_{i=1}^{s} \rho(i) ln \rho(i)$, where P is the normalized frequency of an individual variant I, and s is the number of variants.

Statistics

To estimate the typical range of distribution of the CAIg and tAIg values of the transferred genes, we selected ten random orthologous genes encoding central metabolic enzymes in donor bacteria (*aroK*, *cysE*, *pgk*, *glnA*, *zwf*, *gdhA*, *leuA*, *metE*, *adk*, and *metK*) and calculated the distributions of the CAIg and tAIg values of these genes for each of the seven organisms listed in table 1, using the CAI and tAI metrics of both donor bacteria and *E. coli* (supplementary table S2, Supplementary Material online). Distributions calculated with the metrics of donor bacteria were always shifted toward higher values in comparison to those calculated with the metric of *E. coli*. Using K–S normality test, we determined that all the obtained distributions are normal (supplementary table S3, Supplementary Material online). The two-tailed *t*-test was then used to determine that the means of distributions calculated with the metrics of donor bacteria differ significantly from the means obtained with *E. coli*'s metrics ($P \leq 0.0008$) (supplementary table S3, Supplementary Material online).

Spearman rank tests were performed between ($\log_2$(FC)) and CAIg, tAIg, mRNA folding stability, and GC content at each time point for each population (supplementary table S7, Supplementary Material online).

Polysome Extraction

Polysome extraction was done essentially as described (Liang et al. 2018; Alasad et al. 2020). *Escherichia coli* strains carrying ORG and MOD *L. grayi folA* genes were grown overnight from a single colony in 5-ml supplemented M9 media at 37 °C. Bacterial cultures were then diluted 1/100 in 50 ml of M9 media (in a 250-ml Erlenmeyer flask) and grown at 37 °C for about 4 h until the cells reached OD ~ 0.6. The cultures were then poured into 50-ml flasks (25 ml to each flask) packed loosely with crashed ice. Cultures were quickly mixed with the ice by inverting the flask. Next, the cells were pelleted down by centrifugation (8,000 RCF, 20 min, 4 °C) and resuspended in 0.5-ml lysis buffer (10-mM Tris-HCl pH8.0, 10-mM MgCl2, 1-mg/ml lysozyme 1 mg/ml). The resuspended cells were flash-frozen in liquid nitrogen and placed in an ice cooled water bath with occasional mixing. Once the cells melted, they were frozen again in liquid nitrogen and kept at −80 °C for the next step. The cells were defrosted on ice, mixed with 15-µl sodium deoxycholate, and spanned down in pre-cooled microcentrifuge (10000RCF, 10 min, 4 °C). The clarified portion was transferred to a fresh tube and mixed with 3 µl of RNaseOUT(Invitrogen). Fifty microliters were then collected, supplemented with 450 µl DDW, and 500 µl Trizol (Sigma), and stored at −80 °C as the total mRNA fraction. The rest of the clarified portion was loaded on a sucrose gradient and subjected to ultracentrifugation (229,884×g, 2.5 h, 4 °C). The polysome profile was read using a piston gradient collector (Biocomp) fitted with a UV detector (Tirax). Three polysomal fractions were collected for each growth condition, placed in Trizol (Sigma), and stored at −80 °C as the polysomal mRNA fraction. The RNA from the total and polysomal RNA fractions were extracted using manufacturer's instruction and analyzed by qPCR with primers for *folA* gene (rimers# 1–2, supplementary table S8, Supplementary Material online) and *adk* gene (primers# 21–22, supplementary table S8, Supplementary Material online) that was used as an internal control. The ratio between the threshold cycle of *folA* and *adk* was calculated (ΔCT = CT(folA)−CT(adk)) and was used to determine the relative abundance in each fraction $2^{-2\Delta Ct}$. Increased association of the *folA* gene with polysomes will increase this value. Because of the noisy nature of Polysomal extraction method, log2 transformation was applied on the results. log2 transformation of $\left(\frac{\text{Polysomal folA}}{\text{total folA}}\right)$ ratio was applied.

Simulations

Evolutionary simulations were performed with SodaPop software (Gauthieret al. 2019). The frequency (f) of each variant at a specific time point (t) was calculated by the division of the number of reads (n (t)) by the total number of read at this time point (N (t)), $\left(f(t) = \frac{n(t)}{N(t)}\right)$. The fitness at each time point (w(t)) was calculated by the ratio between the frequencies of a variant at a specific time point (f (s)) and a reference time point $\left(f(r)\right)$, $\left(w(t) = \frac{f(t=s)}{f(t=r)}\right)$. The simulations started with an initial size of 10^6 cells, partitioned into the library variants using either equalized or experimentally determined variant frequencies at $t = 0$. Then, the frequency of each variant was updated using a standard Moran process (Moran 1958), where the fitness of each cell quantitatively represents the doubling time of each cell. This fitness represents a "fitness score" estimated using $\left(\frac{f(t=s)}{f(t=r)}\right)$, starting from the experimental read counts. To mimic the fluctuation in experimental population sizes due to daily dilutions upon passaging, we let the simulated populations reach 10^8 cells over 12 generations, followed by 1/100 dilution every six generations. The dilution is implemented as a random down sampling. Similar to the experiment, in this short number of generations, we assumed that no de novo mutations occurred in the library and that all variant frequency changes are either due to its initial fitness or neutral drift (fully implemented in the explicit Moran process by SodaPop). We performed ten replicate simulations for each of the four experimental conditions corresponding to TMP concentrations.

Supplementary Material

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

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

Acknowledgments

We thank Adrian I. Levy for his feedback on the manuscript. S.B. was supported by the Israel Science Foundation personal research grant 593/21. B. R. was supported by the Israel Science Foundation personal grant 1436/19.

Data availability

Sequences of the individual original (ORG) and modified (MOD) folA genes used in the study are listed in tables S1 and S4, respectively, Supplementary Material, online. Sequences and frequencies of the individual variants within combinatorial libraries prior and after selection are summarized in table s6, Supplementary Material, online.

References

Alasad K, Voeltzke K, Levin L, Reifenberger G, Leprivier G, Rotblat B. 2020. 4EBP1/2 are active under standard cell culture conditions to regulate the translation of specific mRNAs. *Cell Death Dis*. **11**: 968.

Amoros-Moya D, Bedhomme S, Hermann M, Bravo IG. 2010. Evolution in regulatory regions rapidly compensates the cost of nonoptimal codon usage. *Mol Biol Evol*. **27**:2141–2151.

Bentele K, Saffert P, Rauscher R, Ignatova Z, Blüthgen N. 2013. Efficient translation initiation dictates codon usage at gene start. *Mol Syst Biol*. **9**:675.

Bershtein S, Choi JM, Bhattacharyya S, Budnik B, Shakhnovich E. 2015. Systems-level response to point mutations in a core metabolic enzyme modulates genotype-phenotype relationship. *Cell Rep*. **11**:645–656.

Bershtein S, Mu W, Serohijos AW, Zhou J, Shakhnovich EI. 2013. Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. *Mol Cell*. **49**: 133–144.

Bershtein S, Mu W, Shakhnovich EI. 2012. Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. *Proc Natl Acad Sci U S A*. **109**:4857–4862.

Bershtein S, Serohijos AW, Bhattacharyya S, Manhart M, Choi JM, Mu W, Zhou J, Shakhnovich EI. 2015. Protein homeostasis imposes a barrier on functional integration of horizontally transferred genes in bacteria. *PLoS Genet*. **11**:e1005612.

Bhattacharyya S, Bershtein S, Adkar BV, Woodard J, Shakhnovich EI. 2021. Metabolic response to point mutations reveals principles of modulation of in vivo enzyme activity and phenotype. *Mol Syst Biol*. **17**:e10200.

Bhattacharyya S, Bershtein S, Yan J, Argun T, Gilson AI, Trauger SA, Shakhnovich EI. 2016. Transient protein-protein interactions perturb *E. coli* metabolome and cause gene dosage toxicity. *Elife*. **5**:e20309.

Bhattacharyya S, Jacobs WM, Adkar BV, Yan J, Zhang W, Shakhnovich EI. 2018. Accessibility of the Shine-Dalgarno sequence dictates N-terminal codon bias in *E. coli*. *Mol Cell*. **70**: 894–905.e5.

Bollenbach T, Quan S, Chait R, Kishony R. 2009. Nonoptimal microbial response to antibiotics underlies suppressive drug interactions. *Cell*. **139**:707–718.

Burchall JJ, Hitchings GH. 1965. Inhibitor binding analysis of dihydrofolate reductases from various species. *Mol Pharmacol*. **1**: 126–136.

Chaney JL, Steele A, Carmichael R, Rodriguez A, Specht AT, Ngo K, Li J, Emrich S, Clark PL. 2017. Widespread position-specific conservation of synonymous rare codons within coding sequences. *PLoS Comput Biol*. **13**:e1005531.

Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. **34**:i884–i890.

Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH. 2004. Codon usage between genomes is constrained by genome-wide mutational processes. *Proc Natl Acad Sci U S A*. **101**:3480–3485.

de Crécy-Lagard V, El Yacoubi B, de la Garza RD, Noiriel A, Hanson AD. 2007. Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. *BMC Genomics*. **8**:245.

Deng Y, de Lima Hedayioglu F, Kalfon J, Chu D, von der Haar T. 2020. Hidden patterns of codon usage bias across kingdoms. *J R Soc Interface*. **17**:20190819.

dos Reis M, Savva R, Wernisch L. 2004. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res*. **32**:5036–5044.

Duret L. 2000. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet*. **16**:287–289.

Duret L, Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci U S A*. **96**:4482–4487.

Gauthier L, Di Franco R, Serohijos AWR. 2019. Sodapop: a forward simulation suite for the evolutionary dynamics of asexual populations on protein fitness landscapes. *Bioinformatics*. **35**: 4053–4062.

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. 2003. Global analysis of protein expression in yeast. *Nature*. **425**:737–741.

Goetz RM, Fuglsang A. 2005. Correlation of codon bias measures with mRNA levels: analysis of transcriptome data from *Escherichia coli*. *Biochem Biophys Res Commun*. **327**:4–7.

Goodman DB, Church GM, Kosuri S. 2013. Causes and effects of N-terminal codon bias in bacterial genes. *Science*. **342**:475–479.

Gouy M, Gautier C. 1982. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res*. **10**:7055–7074.

Grantham R, Gautier C, Gouy M. 1980. Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res*. **8**:1893–1912.

Grantham R, Gautier C, Gouy M, Mercier R, Pave A. 1980. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res*. **8**: r49–r62.

Gu W, Zhou T, Wilke CO. 2010. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput Biol*. **6**:e1000664.

Gullberg E, Cao S, Berg OG, Ilbäck C, Sandegren L, Hughes D, Andersson DI. 2011. Selection of resistant bacteria at very low antibiotic concentrations. *PLoS Pathog*. **7**:e1002158.

Hershberg R, Petrov DA. 2008. Selection on codon bias. *Annu Rev Genet*. **42**:287–299.

Hughes D, Andersson DI. 2012. Selection of resistance at lethal and non-lethal antibiotic concentrations. *Curr Opin Microbiol*. **15**: 555–560.

Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*. **2**:13–34.

Jacobs WM, Shakhnovich EI. 2017. Evidence of evolutionary selection for cotranslational folding. *Proc Natl Acad Sci U S A*. **114**: 11434–11439.

Kacar B, Garmendia E, Tuncbag N, Andersson DI, Hughes D. 2017. Functional constraints on replacing an essential gene with its ancient and modern homologs. *mBio*. **8**: e01276-17.

Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T. 2001. Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J Mol Evol*. **53**:290–298.

Kanaya S, Yamada Y, Kudo Y, Ikemura T. 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene*. **238**:143–155.

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024

Knight RD, Freeland SJ, Landweber LF. 2001. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biol*. **2**:RESEARCH0010.
Komar AA. 2009. A pause for thought along the co-translational folding pathway. *Trends Biochem Sci*. **34**:16–24.
Kramer EM. 2015. A stranger in a strange land: the utility and interpretation of heterologous expression. *Front Plant Sci*. **6**:734.
Kudla G, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*. **324**: 255–258.
Kwon YK, Lu W, Melamud E, Khanam N, Bognar A, Rabinowitz JD. 2008. A domino effect in antifolate drug action in *Escherichia coli*. *Nat Chem Biol*. **4**:602–608.
Liang S, Bellato HM, Lorent J, Lupinacci FCS, Oertlin C, van Hoef V, Andrade VP, Roffe M, Masvidal L, Hajj GNM, *et al*. 2018. Polysome-profiling in small tissue samples. *Nucleic Acids Res*. **46**:e3.
Lind PA, Tobin C, Berg OG, Kurland CG, Andersson DI. 2010. Compensatory gene amplification restores fitness after interspecies gene replacements. *Mol Microbiol*. **75**:1078–1089.
Markham NR, Zuker M. 2008. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol*. **453**:3–31.
Martens AT, Taylor J, Hilser VJ. 2015. Ribosome A and P sites revealed by length analysis of ribosome profiling data. *Nucleic Acids Res*. **43**:3680–3687.
Medrano-Soto A, Moreno-Hagelsieb G, Vinuesa P, Christen JA, Collado-Vides J. 2004. Successful lateral transfer requires codon usage compatibility between foreign genes and recipient genomes. *Mol Biol Evol*. **21**:1884–1894.
Moran PA. 1958. A general theory of the distribution of gene frequencies. II. Non-overlapping generations. *Proc R Soc Lond B Biol Sci*. **149**:113–116.
Moriyama EN, Powell JR. 1997. Codon usage bias and tRNA abundance in *Drosophila*. *J Mol Evol*. **45**:514–523.
Nakamura Y, Itoh T, Matsuda H, Gojobori T. 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet*. **36**:760–766.
Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature*. **405**:299–304.
Pechmann S, Frydman J. 2013. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol*. **20**:237–243.
Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*. **12**:32–42.
Popa O, Dagan T. 2011. Trends and barriers to lateral gene transfer in prokaryotes. *Curr Opin Microbiol*. **14**:615–623.
Porse A, Schou TS, Munck C, Ellabaan MMH, Sommer MOA. 2018. Biochemical mechanisms determine the functional compatibility of heterologous genes. *Nat Commun*. **9**:522.
Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR, *et al*. 2015. Codon optimality is a major determinant of mRNA stability. *Cell*. **160**:1111–1124.
Rajer F, Sandegren L. 2022. The role of antibiotic resistance genes in the fitness cost of multiresistance plasmids. *mBio* **13**:e0355221.
Rauhut R, Klug G. 1999. mRNA degradation in bacteria. *FEMS Microbiol Rev*. **23**:353–370.
Richardson JP. 1991. Preventing the synthesis of unused transcripts by Rho factor. *Cell*. **64**:1047–1049.
Rocha EP. 2004. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res*. **14**:2279–2286.
Rodrigues JV, Bershtein S, Li A, Lozovsky ER, Hartl DL, Shakhnovich EI. 2016. Biophysical principles predict fitness landscapes of drug resistance. *Proc Natl Acad Sci U S A*. **113**:E1470–E1478.
Rubin AF, Gelman H, Lucas N, Bajjalieh SM, Papenfuss AT, Speed TP, Fowler DM. 2017. A statistical framework for analyzing deep mutational scanning data. *Genome Biol*. **18**:150.
Schlesinger O, Chemla Y, Heltberg M, Ozer E, Marshall R, Noireaux V, Jensen MH, Alfonta L. 2017. Tuning of recombinant protein expression in *Escherichia coli* by manipulating transcription, translation initiation rates, and incorporation of noncanonical amino acids. *ACS Synth Biol*. **6**:1076–1085.
Schnell JR, Dyson HJ, Wright PE. 2004. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu Rev Biophys Biomol Struct*. **33**:119–140.
Sharp PM, Li WH. 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*. **15**:1281–1295.
Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. *Nat Rev Genet*. **16**:472–482.
Stern A, Mayrose I, Penn O, Shaul S, Gophna U, Pupko T. 2010. An evolutionary analysis of lateral gene transfer in thymidylate synthase enzymes. *Syst Biol*. **59**:212–225.
Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS. 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*. **329**:533–538.
Thomas CM, Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol*. **3**: 711–721.
Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y. 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*. **141**:344–354.
Tuller T, Girshovich Y, Sella Y, Kreimer A, Freilich S, Kupiec M, Gophna U, Ruppin E. 2011. Association between translation efficiency and horizontal gene transfer within microbial communities. *Nucleic Acids Res*. **39**:4743–4755.
Tuller T, Waldman YY, Kupiec M, Ruppin E. 2010. Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci U S A*. **107**:3645–3650.
Yamao F, Andachi Y, Muto A, Ikemura T, Osawa S. 1991. Levels of tRNAs in bacterial cells as affected by amino acid usage in proteins. *Nucleic Acids Res*. **19**:6119–6122.
Yang J, Ogawa Y, Camakaris H, Shimada T, Ishihama A, Pittard AJ. 2007. *Fola*, a new member of the TyrR regulon in *Escherichia coli* K-12. *J Bacteriol*. **189**:6080–6084.
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**:623–628.
Zwietering MH, Jongenburger I, Rombouts FM, van 't Riet K. 1990. Modeling of the bacterial growth curve. *Appl Environ Microbiol*. **56**:1875–1881.

Downloaded from https://academic.oup.com/mbe/article/40/6/msad123/7176441 by Universite de Montreal user on 04 April 2024