nature portfolio

Peer Review File

A slight mismatch between a gene's codon usage and the cellular tRNA supply is beneficial

Corresponding Author: Professor Jian-Rong (Philip) Yang

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This is an increasingly important line of investigation. They report that their hypothesis challenges the conventional assumption that codon usage bias (CUB) is evolutionarily selected to represent the tRNA supply pool. The authors state that while matching the CUB to the tRNA pool increases the expression levels of the gene, it depletes the tRNA pool and impacts the cell's fitness. In essence, the authors claim that the degree of the CUB-tRNA match depends on the functional benefit vs fitness trade-off arising from the role of the gene in the cell. The more important the role of the gene in cell function/survival, the better the CUB-tRNA match. To this end, the authors create an Euclidean-distance metric "D" representing the difference between a CUB and the representative tRNA pool. The closer "D" is to zero the more identical a given gene is coded based on the cellular tRNA pool.

The authors synthesized 3 groups of small, medium, and large "D" scored sequences of synonymous gentamicin resistance genes, GmR. They fused mCherry to GmR and measured the functional payoff of the gene as the expression levels of GmR gene as a function of mCherry expression. They quantified changes in a secondary YFP gene to represent translation cost and changes in the exponential phase doubling time of the cells to represent overall fitness cost.

I have serious concerns about the assertions, assumptions, choice of experimental system, and data interpretation. The biggest flaws are:
1) The authors do not quantify tRNA pools, which makes the title & premise of the paper completely misleading.
2) The authors use gentamycin, a translational inhibitor in a study to understand translation elongation. Much of their data is not interpretable since gentamycin alters the translation that they are claiming to evaluate.
3) The authors keep claiming that the conventionally accepted paradigm is the selective pressure on genes to decrease the mismatch between their demand and tRNA pools unidirectionally. No citation is ever provided. Further, this statement is a red herring since this view is not widely accepted and therefore there is nothing to disprove. There is no singular selective pressure for all genes. Different genes have different selective pressures on them, depending on their function, when they are expressed, and in what abundance they are needed. A heat shock protein, which is needed under highly specific conditions will not have the same selective pressure on it as a glycolytic enzyme. Thus, the basis of their paper in disproving a conventionally held view is completely unfounded.

The specifics are below:

Major Comments:

Line 77-94: Why was Gentamicin antibiotic selected for this study? This is an interesting choice of antibiotic, given that gentamicin is known to exacerbate codon-tRNA misreads and lead to the incorporation of the wrong amino acids as well as slow down elongation. It would be good to discuss how reducing the stringency of tRNA-codon interactions may change the results when studying codon influence on gene expression. Could the use of gentamicin create non-linear effects in the data, given that an optimal match to tRNA becomes more important in the presence of an antibiotic that exacerbates errors with mismatched codon-tRNA? Do you have data to support or refute this? This is a near-fatal flaw of the paper. How do you interpret the role of translation elongation when the reporter system is based on perturbing translation elongation? A reporter system like ampicillin or another antibiotic that does not affect transcription or translation should have been used. It's almost impossible to interpret the results in light of the fact that gentamycin is a translation inhibitor.

Line 100: The authors did not measure tRNA supply at all. They only used CUB of highly expressed genes! The title and premise of this paper are completely misleading and is not what supported by the data presented in the paper. Based on the definition of D, it seems what the authors are calling tRNA supply is similar to the traditional codon adaptation index (CAI) for highly expressed genes. There are numerous lines of evidence that CAI is not a measure of tRNA pools, nor is there are very strong correlation between the two. tRNA levels have previously been reported for E. coli (10.1006/jmbi.1996.0428, 10.1080/15476286.2020.1790871). The authors should either measure tRNA levels directly to show it matches CAI or remove any mention of the metric they are comparing with D is tRNA pool. The use of CAI as a proxy for tRNA pools is completely unacceptable.

Line 111: Placing the YFP and GmR genes on the same plasmid as a bidirectional promoter system (convergent system) can lead to confounding effects during unwinding and transcription. This implies that the transcription of one gene can impact that of another. Did the authors ensure/prove that their GmR constructs only have trans effects on the YFP translation? Did they prove that the GmR constructs do not cause cis effects on the transcription of YFP at the DNA level? It's possible for instance that re-codes can affect plasmid structure and thus gene expression of other ORFs on the plasmid, as well as supercoiling, etc. The authors need to include sufficient controls (e.g., no RBS or moving YFP to the genome/different plasmid) to show that their results are indeed trans effects and not cis effects.

Line 358-373: A key issue with this paper concerns the calculation of the distance “D” metric.
The authors do not use a metric that relates to tRNA supply. Their metric D is the difference between demand by highly expressed genes (Y, a genomic frequency of a specific codon used in a sequence) and supply (X, the frequency of the same codon only in highly expressed genes). Effectively, they create a weight for every codon by subtracting the two frequencies above (Y-X) and then they calculate the geometric mean of all the codons and their weights. The authors claim that X or the “tRNA” supply is a representation of “effective” codon supply encompassing tRNA expression levels, tRNA aminoacylation rate, etc. But I disagree. They practically created another Codon Adaptation Index in a different way. They did not calibrate their “D” score meaning that lower D levels may not really be representing the lower tRNA supply. Effectively, for “D” to be useful it has to show that it represents the entire range of tRNA supply somewhat linearly for the authors to be able to claim that lower “D” levels close to zero indeed corresponds to the maximum concordance of demand (codon being read) to supply (tRNA level). I am pessimistic that their metric represents tRNA supply the same way across its entire range.

Line 49-51: “the prevailing hypothesis …” This sentence is poorly written and hard to understand.
Authors need to provide references for the prevailing hypotheses asserting that “translational selection is towards an optimal CUB”, “Deviations are due to mutational bias or genetic drift”.

Line 57-60: It would be easier to understand the author’s claims if they provided a supplementary figure for this assertion instead of citing some papers.

Line 72-76: The authors’ definition of the “conventional model” is not supported. This statement is a red herring giving the authors the opportunity to disprove a hypothesis that is already quite ridiculous. In no way is the prevailing/conventional wisdom that genes drift toward high CAI. Indeed, if that were the case, there would be no codon usage bias among different genes within an organism. The drift toward or away from matching tRNA supply would be highly dependent on 1) how important the protein is to the cell’s survival, 2) what concentration of the protein is required for cell function and 3) under what conditions is the protein essential. To state that drift toward a low mismatch between tRNA supply and demand is ridiculous. The entire premise of this manuscript is based on a red herring.

Line 73-76: There is unlikely to be a single answer for this since codon bias toward high and low D exists of different endogenous genes. Again, this depends on the cellular need for the specific function under the conditions tested. There is not likely to be a single answer. This general premise is highly flawed. The selective pressure is different for different genes.

Line 109-117: Authors are not controlling for changes in 5' region that are translation initiation effects. This confounds results. It's not clear whether the effects are due to translation initiation (controlled by mRNA structure & ribosome binding) or elongation as the authors want to claim. Currently, the authors do not prove that the effects are not from transcription which can misrepresent the translation results. The authors need to show that the results of expression level changes cannot be explained by translation initiation changes.

Similarly, the authors do not control for transcription rates. In bacteria, transcription and translation are coupled. The authors should show that their data is not explained simply by changes in mRNA levels of their reporter if they want to claim that it is all translation elongation effects.

Line 192-195: Authors do not show the data for the KanR gene replacement experiments that is necessary for proving their claim. Needs more information on how functional importance was quantified. Furthermore, Spearman’s correlation of -0.08 is extremely weak almost zero. Authors need to show the data spread used for their correlations. Just providing a correlation is unacceptable.

Line 207-213: This statement is not novel, is intuitive and expected, and is not a contribution of this work. It is already well-accepted that high transcription and translation rates are highly burdensome to cells. See PMID: 27989436.

Line 217-219: What is the threshold authors set for expression and why? The authors need to show the expression data distribution used for obtaining their correlational data in Figure 5.

Line 243 and 282-284: E. coli is a gut microbe. This means that evolutionary pressure for optimizing its genes has largely been driven by the gut microenvironment to match its demands. The authors culture their E. coli in LB media. This means that there will likely be a mismatch in the mutational accumulation experiments relevant to D-increasing or D-decreasing effects. It is incorrect to assume that the selective pressure in lab conditions in LB reflects true selective pressure for E. coli.

Line 303: Authors need to clarify what they mean by “waste of tRNAs for unpreferred codons”.

Regarding Figure 3.b
- Why does at high Gm concentrations the small group (theoretically with less GmR protein than the minimum group) exhibit a higher functional payoff?
- Why is everything normalized to the large group? This implies you are assuming unidirectionality in codon-match and functional payoff.
- The data contradicts the authors’ assumption that as the importance of a gene increases, D (tRNA-CUB supply-demand differential) decreases. At concentrations above 140 ug/ml we see that the small group is preferred. (In text line 154 you say that only 68.9% of the variations are explained by this assumption).

Figure 3.c-e
- It is unclear what the data points here represent. Are they average or individual single recodes? Why is there only 12 points here? Shouldn’t you have the minimum and small groups here (38 points)? Authors need to provide information about their statistical test in the figure legend.

Figure 3.f
- Relevant to result interpretations of line 160-163:
- I disagree that at concentrations above 120 ng/ul minimum D is selected. Mainly because the statistic is weakly probable. From 140 ng/ul all groups most likely have the same relative fitness. This contradicts your statement that minimum D is best at the highest functional importance (highest concentrations).

Figure 4
- This correlation does not support any direct like to the hypothesis that genes with smaller D scores have a higher functional payoff. Because the function is related to the direct role of the gene and not how it is coded. In order to support your hypothesis, you need to have direct fitness data where you measure the fitness cost associated with recoding a specific vital gene and replacing it with the native gene to determine the link between functional payoff and translation cost.
- With regards to line 211-213 An important gene does not have to be highly expressed. You cannot establish the importance of genes by only looking at highly expressed genes and measuring expression indirectly by the D index.

Figure 5
- In regards to line 216: Because D and expression level are anticorrelated does not imply anything about translational cost. You are only looking at functional payoff. Translation cost can only be measured experimentally.

Question on the methods:
- How did the authors control for the sample size reduction of highly expressed genes? Correlation scores tend to shrink in absolute values as sample sizes grow. Conversely, a strong correlation at a small sample size may not be true when the sample of the correlation is increased.

Figure 6
- While this evidence does support the authors' claim that a slight mismatch with the tRNA supply is selectively maintained, the authors do not quantify the strength of the D-changing mutations. In other words, what is the distribution of D score change in each D-increasing/-decreasing mutation group?

Another note regarding the authors' quantification of cellular tRNA:

Is the index for quantifying the cellular tRNA pool linear in its entire range? Perhaps if this index becomes non-linear in either boundary it may falsely inform that D decreasing mutations are deleterious at highly expressed genes.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The manuscript submitted by Chen, Liu, Zhou and co-authors is dedicated to the study of translational selection. Specifically, the authors propose a new model to explain the evolution of codon usage bias (abbreviated CUB; this can be roughly defined as the departure from a random distribution among synonymous codons), according to which the optimal codon usage does not (always) perfectly match the abundance of tRNAs in the cell. According to this model, a mismatch between the codon usage bias of a given gene and the tRNA supply can be beneficial because it reduces the cost imposed by the translation of the gene. In other words, a highly expressed and highly translated gene would be costly because its translation would hamper the translation of other genes.

To support this model, the authors mainly rely on genome editing experiments performed in E. coli. Specifically, they insert cassettes carrying a gentamicin-resistance gene (GmR) with different codon usage biases. The GmR gene is fused to a mCherry coding sequence, which enables the authors to measure the expression level of the GmR-mCherry fusion through fluorescence measurements. The cassettes also contain a YFP gene in the opposite orientation. The expression level of the YFP gene, again estimated through fluorescence measurements, enables the authors to estimate the translational cost of the CUB of the GmR gene as the reduction in YFP protein levels. The authors perform these experiments at different concentrations of gentamicin and they also measure the growth rate of the different E. coli clones, thus obtaining a more direct estimation of the fitness effect of the different constructs. In addition to these main experiments, the authors also analyze previously published data for mutation accumulation or gene replacement experiments in E. coli and yeast, as well as codon usage data for various species.

I find the model proposed by the authors to be interesting and reasonable. It brings an alternative to the classical model which posits (even in species where translational selection is efficient, that is, species with high effective population size such as E. coli, yeast or fruitfly) translational selection cannot optimize the codon usage bias to perfectly match the tRNA abundance of the cell because of genetic drift. However, it must be stated that the classical model already incorporates the idea of a difference between gene categories (the translational selection coefficient is expectely to vary among genes, depending on the optimum protein quantity produced by each gene, and for some genes the translational selection coefficient is indeed expected to be very low). More importantly, I am concerned about the validity of several assumptions made in this article and of several analyses performed in this article.

First, the authors state that the D metric that they define is a measure of the mismatch between the CUB of a gene and the tRNA supply of the cell. It is in fact a measure of the mismatch between the CUB of a given gene and the CUB of the highly expressed genes (Methods, lines 358-373). The authors assume that the CUB of highly expressed genes has been optimized to match the tRNA supply, although they then affirm that the optimal CUB should not perfectly match the tRNA supply, even for highly expressed genes. This is circular and illogical. There are better ways to approximate the tRNAs supply without assuming that the CUB for highly expressed genes has been optimize to match it. If the authors are unable to measure directly tRNA abundance, they could use an approximation based on tRNA gene copy number, taking into account wobble rules etc. As this affirmation is the basis of the new model proposed by the authors, it is very important to make sure that D indeed measures the distance compared to the CUB that would perfectly match the tRNA supply of the cell.

Second, the authors measure the translational cost of the CUB of the GmR artificial constructs by measuring the fluorescence level of the exogenous YFP present in the construct. However, they do not ascertain that variations in YFP fluorescence levels are indeed to variations in YFP translation, rather than for example in variations in YFP gene transcription. They also do not show that the variations in YFP fluorescence are causally connected to the translation of the GmR-mCherry gene. The authors and others have shown before that variations in CUB can also affect the mRNA level of genes, mainly due to mRNA stability. Additional effects ca be imagined at the transcriptional level, given the overall differences in base composition of the CUB variants. As the YFP gene is inserted in close proximity to the GmR-mCherry gene, on the reverse strand, transcriptional interference can occur between the two genes. Thus, their claim of a translational cost is not sufficiently supported. Furthermore, the translational cost should be evaluated on endogenous genes. As this is an important claim of the manuscript, it should be properly verified, by evaluating mRNA levels and translation rates in the different constructs (for example with RNA sequencing and riboosome profiling assays).

Third, the fact that the different E. coli strains that they used could display variations in plasmid copy number is quite concerning. I am not convinced that the lack of correlation between the copy number and D is sufficient to ensure that the results are not biased. The authors should verify that the plasmid copy number variation is not a confounding factor for fitness analysis and for translational cost analyses.

Fourth, the fact that the authors observe a positive correlation between D and expression levels among the most highly expressed genes (top 10%, top 5% and top 3%) in all 4 analyzed species (Figure 4) is puzzling. In human, translational selection is highly inefficient, unlike in the other 3 species. The fact that the same effect is observed in human, E. coli, yeast and Drosophila rather points to a more general effect, potentially a confounding effect of base composition. This should be examined carefully.

Fifth, it seems that the mutation accumulation analysis was performed using both synonymous and nonsynonymous mutations (only nonsense mutations seem to have been excluded, Methods, line 464 to 466). This does not make sense - if one wants to measure the effect of the D, only synonymous mutations should be considered. It is also not clear why this analysis was only performed on the 10% most highly expressed genes (especially given that D measures the distance to the codon usage of the most highly expressed genes). Here, a comparison between gene expression classes would be useful, to test the hypothesis that the CUB of the highly expressed genes has indeed been optimized. Here again, D should be computed with respect to the codon usage that perfectly matches the tRNA abundance (not to the codon usage of highly expressed genes).

Finally, the authors might be interested in this preprint, which shows that translational selection is surprisingly rare in metazoans: https://www.biorxiv.org/content/10.1101/2024.07.22.604600v1 .

(Remarks on code availability)
I haven't yet reviewed the code but I plan to do so for a revised version of the manuscript.

Reviewer #3

(Remarks to the Author)
The paper titled “A slight mismatch between a gene’s codon usage and the cellular
tRNA supply is beneficial” describes a very thorough work with a thought-provoking notion. Unlike the current dogma in the field, which assumes that optimal translation efficiency is achieved when codon usage perfectly matches the tRNA pool of the cell, a slight mismatch might be optimal. A mechanism based on feedbacl on the tRNA pool is suggested. A proposed mechanism is based on an alleged trans-regulation and tRNA depletion.
While previously there were some indications for this effect in unneeded genes, now they look for genes with fitness effect, either a drug resistance gene or natural genes. This is a major step forward as it allows them to balance costs against functional payoff.

This is an excellent work. It is thought-provoking, very well done and deep in it evolutionary approach. It is also very well written and technically sound. I would ultimately recommend publication in NC. Yet, if I am contribute to the robustness of the conclusion and to the discussion that this paper will surely evoke, here are some important points.
I’m accepting to sign the report with my name: Yitzhak (Tzachi) Pilpel.
1. A critical experiment could be to manipulate the tRNA pool and then test, what I believe is a prediction of their theory. That for different tRNA pools, alternative members of the collection of 53 synonymous variants would optimal. This is a clear prediction of their model since for each tRN pool other variants might be at optimal D. If they showed that I would be convinced better. If the result would show that same variants are good at different tRNA pools then their conclusion is not supported. In other words, they have manipulated so far the codons – the demand, but never the tRNA – the supply.
2. The proposed mechanism that could mediate their effect is that of tRNA depletion, “In theory, the same logic applies to endogenous genes native to the cell, since they also trans-regulates the expression of other genes via tRNA depletion” – Ref #23. First, I don’t think reference #23 shows tRNA depletion. This is a misunderstanding of what was done in Frumkin et al. The increase in tRNA supply in ref #23 was very different, first because it was done artificially by the researchers (not spontaneously by the cell), and second, because the tRNA levels themselves were not changed, but rather tRNA anticodons were mutated. So it’s incorrect to ascribe to ref#23 a trans-regulation of the tRNA pool. That said, I think we are left without a true reference or a proof, that the tRNA pool is indeed trans-regulated in response to change in codon demand. So I can see several options: 1. The authords measure tRNA pool with some existing tRNA seq protocols and show the effect; 2. Find a reference that truly shows this effect, or 3. Acknowledge that we don’t know if such effect indeed occurs at all
3. Their calculation of the D measure – the extent of deviation from the tRNA pool and codon usage is critical, especially since the whole argument is based on the difference between D that approach zero and D that is slightly above zero. My problem is with their estimation of the tRNA pool which is needed for equation in line 358. Particularly they say “Xij represents the tRNA supply of the cell, approximated by the fraction of codon j among the synonymous codons of amino acid i for the highly expressed genes in the transcriptome of the cell” ref #47. I’m not sure if and how ref #47 provides us with a tRNA pool. So a few comments on this: 1. I’d like to see their results reproduced with other, more conventional definitions of the tRNA pool. Again, they could do tRNA seq. There’s also the classical tAI – based tRNA pool deduced from tRNA gene copy number. Is their D optimized to be slightly above zero also when using these more conventional estimations of the tRNA pool? The way in which ref #47 appears to be estimating the tRNA pool is affected by codon demand, this is potentially problematic.
4. Their estimates for genes functional importance in case of natural genes is based on fitness upon replacement with Kan_R. I’m not sure if this measure is adequate because it represents a total deletion of a gene, while here we ask about reduction in its translation efficiency. I’m not sure how to solve this problem, but I’d be better convinced if they had additional measures of genes’ functional importance. Perhaps by employing bioinformatics of evolutionary conservation of genes they could generate an estimation of functional importance but not open total deletion.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors are commended on thoroughly addressing my comments.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
Following the first round of reviews, Chen and collaborators have considerably improved their manuscript entitled " A slight mismatch between a gene’s codon usage and the cellular tRNA supply is beneficial", by addressing most of reviewers' comments. I will go through the main comments below.

- In my view, they have satisfactorily addressed Reviewer 1's main objection related to the use of gentamycin, which has potential confounding effects as it affects translation. They have performed additional experiments using ampicillin and an ampicillin-resistant gene in their genetic editing work, which confirm their original observations.

- The second objection, brought forward by all three reviewers, is that their measure of distance with respect to the tRNA supply (noted D in the manuscript) is an imperfect approximation. All three reviewers highlighted the fact that D is rather a measure of distance with respect to the synonymous codon frequencies of the 10% most highly expressed genes, and not a direct measure of the distance with respect to the actual tRNA supply. The authors argue that the two should match, under the hypothesis that translational selection has optimized the codon usage bias of the 10% most highly expressed genes to match the tRNA supply. That is indeed what is expected, under the "conventional" translational selection model (which is disputed in this article), for species where translational selection is strong - that is, species with high effective population size. However, in species with low Ne, codon usage bias does not match tRNA supply even for highly expressed genes - again, see Bénitière et al. Nevertheless, the authors have followed the reviewers' suggestion and included in the revised manuscript a measure of tRNA supply based on tRNA gene expression levels. They show that their classification of GmR versions into categories (low D, medium D and high D) does not change with this definition of tRNA supply. This confirms the validity of most of their results, but not all - indeed, the analyses presented in Figure 4, 5 and 6 (which go beyond the analysis of E coli strains with different GmR versions) have not been redone with this measure of D. As these analyses are important for the authors' model, they need to be validated with this new measure of D.
On the same point, I still have trouble with the circularity of the argument: the authors rely on a measure of D that only makes sense under the "conventional" translational selection theory to refute the "conventional" translational selection theory. Under their model, codon usage bias should match the tRNA supply only for highly expressed *and essential/important* genes, for which the functional payoff is more important than the translational cost. The authors should analyze whether their analyses are still valid when defining D based on the codon usage of the most highly expressed and essential/important genes. I am not suggesting that this new D measure should replace the one originally proposed by the authors throughout the manuscript, as that would be circular as well - but rather I am asking for additional analyses, to be presented in the supplementary material.

- The authors have addressed my comment regarding the translational effects on endogenous genes by performing Ribo-seq and analyzing the decoding time of codons that are frequent in the GmR. They show that in E coli strains where the GmR has low D, there is a positive correlation between decoding time and the frequency of codons in the GmR sequence, in particularly for genes that are highly translated. I appreciate the effort that the authors have put into performing the Ribo-seq experiments and I am satisfied with the result.

- Regarding the mutation accumulation experiments, the authors have followed my suggestion to only include synonymous mutations. However, I am puzzled as to why they argue that it is not useful to look at weakly expressed genes. In fact, weakly expressed genes should provide a "control" of sorts, precisely because their CUB is not expected to have been optimized by purifying selection. I would like to see a comparison between highly expressed and weakly expressed genes for the analyses presented in Figure 6.

- Regarding the analysis presented in Figure 5 (correlation between D and expression level, among highly expressed genes), I am still not entirely convinced by the authors' interpretation, mainly because of the results observed for human. The corerlation between D and expression level is 0.15 for human and 0.13 for yeast for the top 10% most highly expressed genes, although in yeast translational selection (be it "conventional" or according to the authors' new model) should be much more efficient than in humans. Because D is defined as the distance with respect to the CUB of the 10% most highly expressed genes, and that the CUB is computed based on the transcriptome not on the genome, I wonder if part of this correlation may be artefactual (though I confess that I have trouble imagining what the confounding factor might be!). Again, here it is important to redo the analysis with the D measure that is based on tRNA abundance. Some protein-coding gene families (e.g. ribosomal proteins) are greatly over-represented among the most highly expressed genes. Could that induce a biased *amino-acid* usage, that could explain some of this correlation? Perhaps if one could compute a D value by amino-acid, and check if all amino-acid D values show the same pattern?

I am otherwise satisfied by authors' reply to my suggestions (including my questions related to YFP transcription level, plasmid copy number variations and GC content effects). Overall I believe that this is a valuable manuscript and that it should ultimately be published in Nature Communications, following an additional round of review that would address the remaining points listed above.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
I thank again the authors for their thorough reply. I remain puzzled by some of the results (in particular the apparent high translational selection levels in humans, according to the authors' observations), but overall I do not have any further

objections to publication. This manuscript deserves to be published and further examined by the scientific community.

(Remarks on code availability)

Point-by-point response to reviewers' comments

Reviewer #1:

Overall Comment:

This is an increasingly important line of investigation. They report that their hypothesis challenges the conventional assumption that codon usage bias (CUB) is evolutionarily selected to represent the tRNA supply pool. The authors state that while matching the CUB to the tRNA pool increases the expression levels of the gene, it depletes the tRNA pool and impacts the cell's fitness. In essence, the authors claim that the degree of the CUB-tRNA match depends on the functional benefit vs fitness trade-off arising from the role of the gene in the cell. The more important the role of the gene in cell function/survival, the better the CUB-tRNA match. To this end, the authors create an Euclidean-distance metric "D" representing the difference between a CUB and the representative tRNA pool. The closer "D" is to zero the more identical a given gene is coded based on the cellular tRNA pool.

The authors synthesized 3 groups of small, medium, and large "D" scored sequences of synonymous gentamicin resistance genes, GmR. They fused mCherry to GmR and measured the functional payoff of the gene as the expression levels of GmR gene as a function of mCherry expression. They quantified changes in a secondary YFP gene to represent translation cost and changes in the exponential phase doubling time of the cells to represent overall fitness cost.

I have serious concerns about the assertions, assumptions, choice of experimental system, and data interpretation. The biggest flaws are:
1) The authors do not quantify tRNA pools, which makes the title & premise of the paper completely misleading.
2) The authors use gentamycin, a translational inhibitor in a study to understand translation elongation. Much of their data is not interpretable since gentamycin alters the translation that they are claiming to evaluate.
3) The authors keep claiming that the conventionally accepted paradigm is the selective pressure on genes to decrease the mismatch between their demand and tRNA pools unidirectionally. No citation is ever provided. Further, this statement is a red herring since this view is not widely accepted and therefore there is nothing to disprove. There is no singular selective pressure for all genes. Different genes have different selective pressures on them, depending on their function, when they are expressed, and in what abundance they are needed. A heat shock protein, which is needed under highly specific

conditions will not have the same selective pressure on it as a glycolytic enzyme. Thus, the basis of their paper in disproving a conventionally held view is completely unfounded.

Response:

Please accept our sincere thanks for your time, your thoughtful reading, and very critical comments. These suggestions are taken very seriously as we believe that addressing them would significantly improve our study. Since these topics are detailed below in individual comment paragraphs, we will address them in a point-by-point manner below to avoid clumsiness and repetition. More specifically, the concern (1) is addressed in our responses to Comments 2 & 4 & 22, concern (2) in Comment 1, and concern (3) in Comment 5 & 7 & 8 below.

Comment 1

Line 77-94: Why was Gentamicin antibiotic selected for this study? This is an interesting choice of antibiotic, given that gentamicin is known to exacerbate codon-tRNA misreads and lead to the incorporation of the wrong amino acids as well as slow down elongation. It would be good to discuss how reducing the stringency of tRNA-codon interactions may change the results when studying codon influence on gene expression. Could the use of gentamicin create non-linear effects in the data, given that an optimal match to tRNA becomes more important in the presence of an antibiotic that exacerbates errors with mismatched codon-tRNA? Do you have data to support or refute this? This is a near-fatal flaw of the paper. How do you interpret the role of translation elongation when the reporter system is based on perturbing translation elongation? A reporter system like ampicillin or another antibiotic that does not affect transcription or translation should have been used. It's almost impossible to interpret the results in light of the fact that gentamycin is a translation inhibitor.

Response:

Thank you very much for your insightful comment. We would like to respond from three angles below.

1. "Why was Gentamicin antibiotic selected for this study?"

To test our model, we require a gene with modifiable functional importance in order to evaluate the relationship between functional importance and optimal codon usage. We decided to use antibiotic resistance genes, whose functional importance can be easily altered by adjusting the antibiotic concentration in the culture media. Additionally, the gene should be relatively short to reduce the cost of synthesizing its synonymous variants and minimize the complexity of the experimental procedure. The GmR gene was selected from a variety of commonly used resistance genes due to its

short sequence length of only 531 bases.

2. “Could the use of gentamicin create non-linear effects in the data, given that an optimal match to tRNA becomes more important in the presence of an antibiotic that exacerbates errors with mismatched codon-tRNA?”

While we agree that Gentamicin may enhance the importance of matching tRNA supply, the non-linearity cannot be solely attributed Gentamicin. In fact, most models of gene expression assume a nonlinear diminishing return for a gene’s functional output (e.g. PMID:20884723), and such nonlinearity is indeed observed for the fitness effect of increased expression (e.g. PMID: 27545349).

More importantly, while GmR function may explain the observed patterns of functional payoff, it cannot explain the translational costs. In particular, if we consider a perfect match between GmR codon usage and the cellular tRNA supply, its resultant increased GmR expression and enhanced effect on alleviating Gm's toxicity, then YFP expression should increase, or, in other words, lower translational cost in strains with GmRs of smaller *D*. This is the opposite of what we found for translational cost, which was generally higher in strains with GmRs of minimum *D* (Fig. 2d). This pattern of translational cost, however, can be explained by our proposed model of tRNA depletion caused by excessive expression of GmRs of minimum *D*. Hence, the data refute the possibility of a GmR-specific artifact, and they are more consistent with our proposed model.

3. “A reporter system like ampicillin or another antibiotic that does not affect transcription or translation should have been used”

Per your suggestion, we designed and synthesized 47 synonymous variants of AmpR, the ampicillin resistance gene, and constructed expression cassettes by replacing GmR in Figure 2b with one of these variants, resulting in 47 synonymous AmpR strains. These 47 strains were then tested for fitness at various ampicillin concentrations. As a result, we observed a similar pattern to that observed with GmR. That is, at low/intermediate/high concentrations of ampicillin (100/2000/4000 ug/mL), optimal fitness is achieved by cells expressing AmpR with large/small/minimum *D*. We have now added this result as **Supplementary Figure S5**, and is pasted below. Please note that the fluorescence levels were not measured for the AmpR strains since ampicillin compromises membrane integrity and causes cell content leakage, making fluorescence levels unreliable for determining the real cost and payoff.

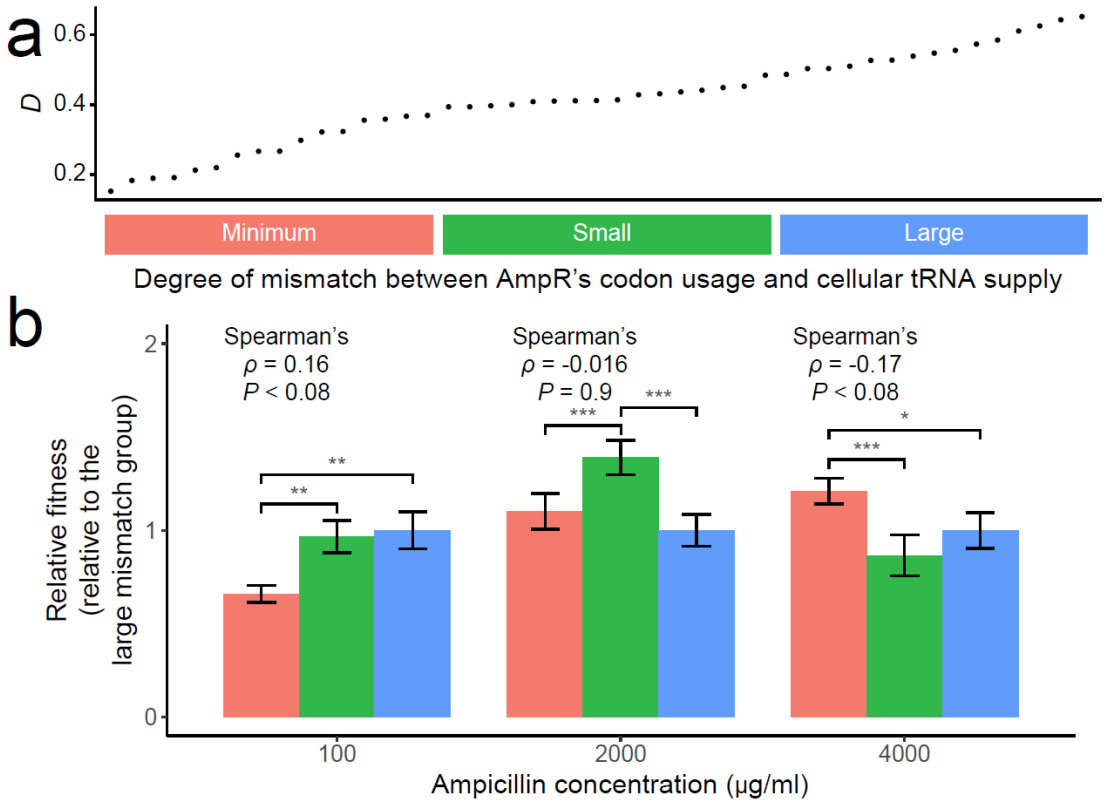


Figure S5. The optimal *D* for AmpR

(**a**) A total of 47 synonymous AmpR variants with different *D* values were designed, which were categorized into three equal-sized groups of minimum, small and large *D*. (**b**) The relative fitness of cells expressing different AmpR synonymous variants in LB media supplemented with 100, 2000, or 4000 μg/ml ampicillin. The statistical significance of two-tailed Wilcoxon Rank Sum Tests are indicated as *: $P<0.1$, **: $P < 0.05$, or ***: $P < 0.01$.

Comment 2

Line 100: The authors did not measure tRNA supply at all. They only used CUB of highly expressed genes! The title and premise of this paper are completely misleading and is not what supported by the data presented in the paper. Based on the definition of D, it seems what the authors are calling tRNA supply is similar to the traditional codon adaptation index (CAI) for highly expressed genes. There are numerous lines of evidence that CAI is not a measure of tRNA pools, nor is there are very strong correlation between the two. tRNA levels have previously been reported for E. coli (10.1006/jmbi.1996.0428, 10.1080/15476286.2020.1790871). The authors should either measure tRNA levels directly to show it matches CAI or remove any mention of the metric they are comparing with D is tRNA pool. The use of CAI as a proxy for tRNA pools is completely unacceptable.

Response:

Thank you for your comment. We will explain below the logic behind our estimation of tRNA supply, and then demonstrate that results based on tRNA quantification do not alter our conclusions. Please also see our response to the related Comment 4 below.

The logic behind our estimation of tRNA supply is twofold. **First**, the actual tRNA supply is neither the copy number of the tRNA gene nor the abundance of tRNA transcripts. Instead, it is the abundance of the ternary complex of aminoacyl-tRNA+EF-Tu+GTP (or replace EF-Tu with EF1A in eukaryotes). Currently, the abundance of ternary complexes for different types of tRNAs cannot be systematically measured. In addition, the kinetic parameters between different codon-anticodon pairs will also influence the "effective" tRNA supply, further complicating the estimation process. **Second**, because of the first point, we resorted to the assumption that the cellular tRNA supply and demand (codon usage) of the whole transcriptome have been evolutionarily optimized by natural selection to ensure a close match between supply and demand. This assumption allows us to approximate the supply of tRNA by using the codon usage (the demand) of the top 10% expressed gene (which dominate the transcriptome) in the transcriptome. It is important to note that we are using the transcriptome, not the genome, which means the codon usage has been weighted by the expression of those genes before being summed up. Furthermore, this assumption of supply-demand match has been empirically supported (PMID: 22479199 and PMCID: PMC12262485), and has been used in our previous publications (PMID: 32123323 and 34977494). It is also important to note that a supply-demand match on a transcriptome level does not necessarily imply a supply-demand match for individual genes, so this assumption does not contradict our main conclusion.

Nevertheless, per your suggestion, we obtained tRNA expression levels based on high-throughput tRNA sequencing data from a previous publication (PMID: 31353208, GEO accession GSE128812), which were then scaled by codon-anticodon affinity and used as the tRNA supply. This metric is similar to the “adaptiveness value” (w) of a codon, as used in the popular tAI metric (PMID:15448185). It was found that the grouping (minimum, small, or large D) in **Figure 3** remained unchanged, suggesting that our conclusion is not affected by the change in the method used to quantify tRNA supply. We have now included this result in **Supplementary Figure S2**, and it is pasted below.

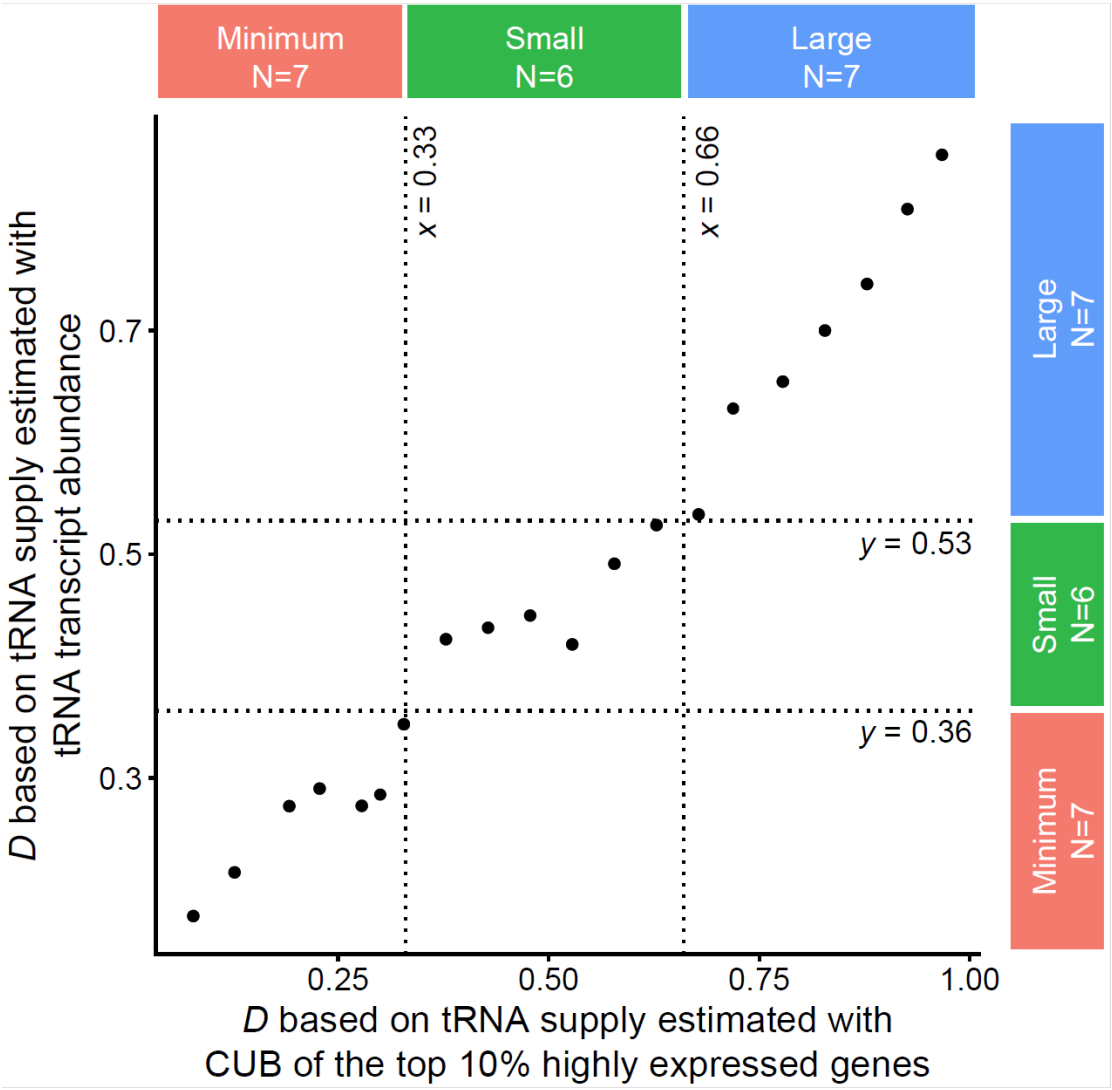


Figure S2. Grouping by D of GmR strains remains unchanged if high-throughput tRNA sequencing data are used to estimate tRNA supply

The D values of designed GmR synonymous variants were calculated with tRNA supply based on either transcriptomic codon usage of the top 10% highly expressed genes (x axis) or tRNA transcript abundance derived from high-throughput tRNA sequencing (y axis). The black lines indicate the grouping, which is identical for both types of D values. Please note that each dot has up to three unique sequences as biological replicates (see **Table S2**).

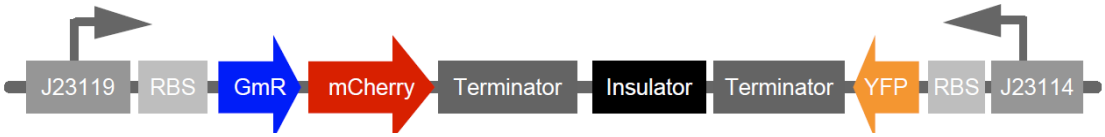
Comment 3

Line 111: Placing the YFP and GmR genes on the same plasmid as a bidirectional promoter system (convergent system) can lead to confounding effects during unwinding and transcription. This implies that the transcription of one gene can impact that of another. Did the authors ensure/prove that their GmR constructs only have trans effects

on the YFP translation? Did they prove that the GmR constructs do not cause cis effects on the transcription of YFP at the DNA level? It's possible for instance that re-codes can affect plasmid structure and thus gene expression of other ORFs on the plasmid, as well as supercoiling, etc. The authors need to include sufficient controls (e.g., no RBS or moving YFP to the genome/different plasmid) to show that their results are indeed trans effects and not cis effects.

Response:

Thank you very much for your critical comment. We have included an insulator between the terminators of YFP and GmR to prevent transcriptional interference between the two genes. The **Figure 2b** pasted below has been updated to explicitly indicate the insulator.



Furthermore, per your comment, we used quantitative RT-PCR to measure YFP mRNA expression levels in the GmR strains growing in LB. We found that YFP mRNA expression did not correlate with GmR's codon usage (*D* value). This result has been added to **Supplementary Figure S9c** and pasted below.

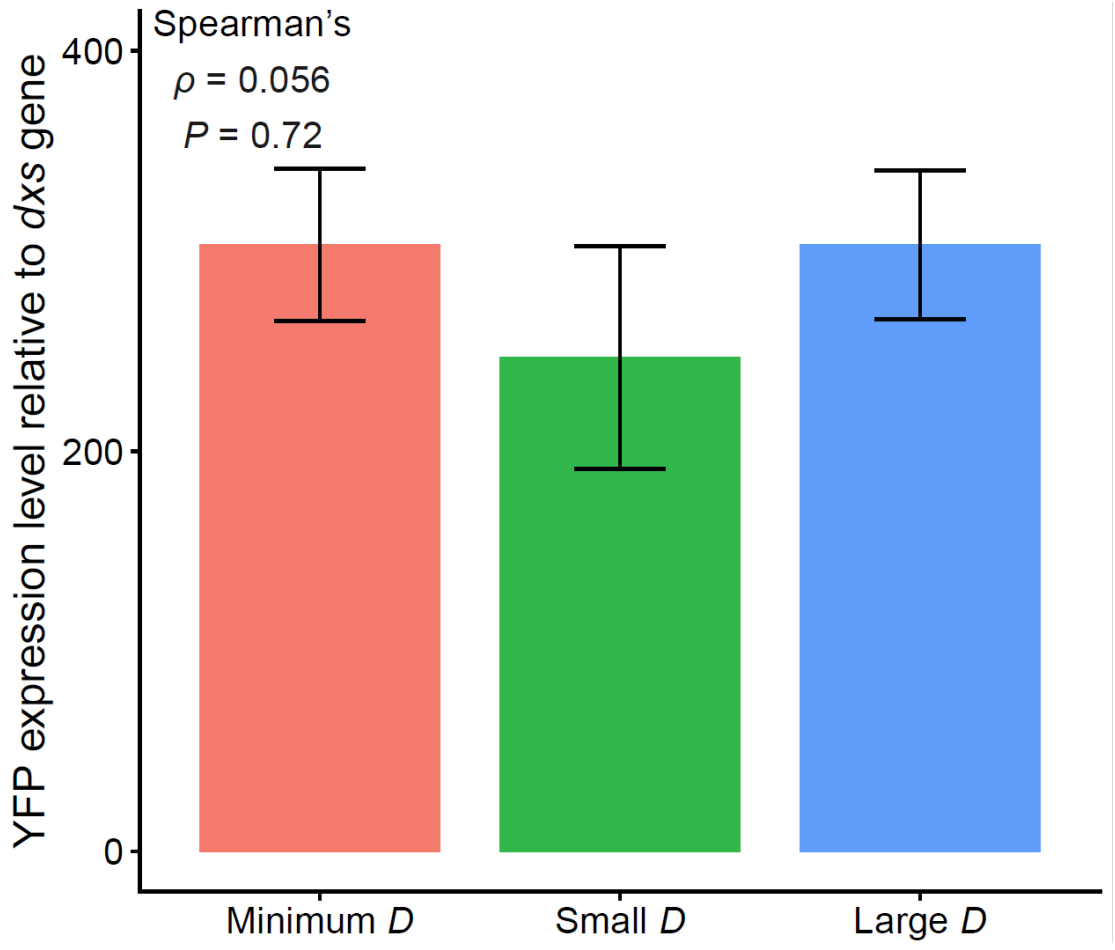


Figure S9. Correlations between *D* and functional payoff, translational cost or fitness

(**c**) The abundance of YFP mRNA in different strains, as determined by quantitative RT-PCR (see **Methods**), is not correlated with codon usage of GmR. The Spearman's Correlation Coefficient ρ of the raw data, as well as the corresponding *P* value, are indicated.

Comment 4

Line 358-373: A key issue with this paper concerns the calculation of the distance "D" metric.
The authors do not use a metric that relates to tRNA supply. Their metric D is the difference between demand by highly expressed genes (Y, a genomic frequency of a specific codon used in a sequence) and supply (X, the frequency of the same codon only in highly expressed genes). Effectively, they create a weight for every codon by subtracting the two frequencies above (Y-X) and then they calculate the geometric mean of all the codons and their weights. The authors claim that X or the "tRNA" supply is a representation of "effective" codon supply encompassing tRNA expression levels, tRNA aminoacylation rate, etc. But I disagree. They practically created another Codon Adaptation Index in a different way. They did not calibrate their "D" score meaning that lower D levels may not really be representing the lower tRNA supply. Effectively, for "D" to be useful it has to show that it represents the entire range of tRNA supply somewhat linearly for the authors to be able to claim that lower "D" levels close to zero indeed corresponds to the maximum concordance of demand (codon being read) to supply (tRNA level). I am pessimistic that their metric represents tRNA supply the same way across its entire range.

Response:

This comment raises an imporant point regarding the estimation of tRNA supply in our study, which is greatly appreciated. Our response to Comment 2 above explains the logic behind the metric *D* and demonstrates that our results are not affected by using high-throughut tRNA sequencing estimates of tRNA supply.

With regard to the more specific points in this comment, we would like to clarify that we do not claim or assume that *D* linearly represents the tRNA supply or the tRNA supply-demand mismatch. We do not make conclusions based solely on the linearity assumption, such as Pearson's correlation tests with *D* (we always accompany them with Spearman's rank-based tests). The reasons for such non-linearity can be biological (the degree of mismatch is not linear to the phenotypic consequences) and/or technical (different amino acids might have different impacts, but we are merely using their geometric average).

Moreover, the *D* metric differs from CAI in several ways. For example, increasing

CAI is selectively advantagous due to improved translational efficiency when more prefered codons are used (as assumed by its underlying model), whereas minimizing *D* is advantageous. Also, classic CAI calculation uses the genomic codon usage of the top 10% highly expressed genes without using their expression levels as a weight to sum up codon usage across genes. As for *D*, we used the transcriptomic, not the genomic, codon usage, which means the codon usage has been weighted by the expression of those genes before being summed. Moreover, CAI is calculated as the geometric mean of the relative adaptiveness of each codon site. In other words, its final calculation step invovles taking the *L*-th root of a number, where *L* is the length of the protein (or number of codons in the CDS). But for *D*, the final calculation step is usually taking 18^{-}th root, since there are only 18 groups of synonymous codons. This number of 18 may be further reduced if some amino acid is completely absent from the protein. There are other differences, such as CAI assumes the preferred codon is the best option, whereas *D* assumes perfect supply-demand balance is the best option. We hope these explanations have clarified *D*'s distinct characteristics, particularly in comparison with CAI.

Comment 5

Line 49-51: “the prevailing hypothesis …” This sentence is poorly written and hard to understand.
Authors need to provide references for the prevailing hypotheses asserting that “translational selection is towards an optimal CUB”, “Deviations are due to mutational bias or genetic drift”.

Response:

Thank you and we apologize for the confusion. We have now included two highly cited references that describe the prevailing hypothesis. For example, in [Hershberg and Petrov. Selection on Codon Bias. Annual Review Genetics. 2008] (1195 citations on Google Scholar), it is stated that

> Explanations for the existence of codon bias fall into two general classes. According to the selectionist explanation, codon bias contributes to the efficiency and/or the accuracy of protein expression and is thus generated and maintained by selection. The mutational or neutral explanation, by contrast, posits that codon bias exists because of nonrandomness in the mutational patterns.
>
> …
>
> Given the evidence that both mutational pressures and selection are involved in the phenomenon of codon bias, the current accepted model is the major codon preference model, also known as the mutation-selection-drift

balance model of codon bias. This model proposes that selection favors the major (or preferred) codons over minor codons. However, mutational pressure and genetic drift allow the minor codons to persist.

We have also cited another classic paper that proposes this prevailing model [The Selection-Mutation-Drift Theory of Synonymous Codon Usage. Michael Bulmer. Genetics. 1991] (1244 citations on Google Scholar), and also revised this sentence, which is pasted below. It is our hope that the reviewer finds this statement more understandable in explaining the current prevailing model.

Other than some special cases affecting only limited neighboring regions[5-8], the prevailing hypothesis for the widely observed CUB is the slection-mutation-drift model[9,10], which proposes that translational selection favors an optimal CUB, whereas deviations from the optimal CUB persist as a result of genetic drift and mutational bias.

Comment 6

Line 57-60: It would be easier to understand the author's claims if they provided a supplementary figure for this assertion instead of citing some papers.

Response:

Thank you and we agree with you. Our claim is summarized in **Figure 1**, which is pasted below.

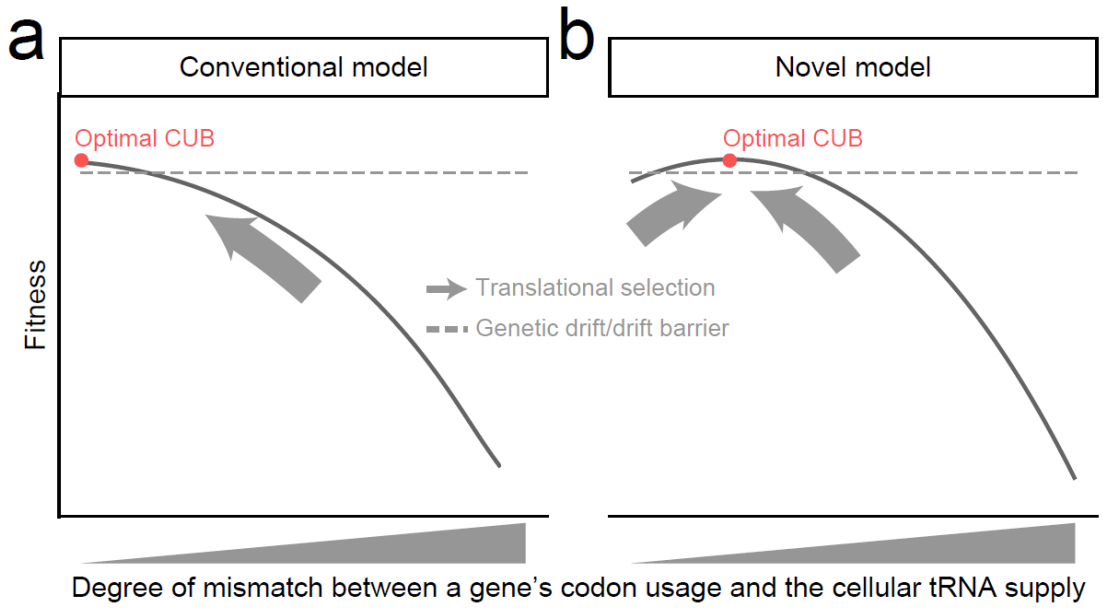


Figure 1. Two competing models for the translational selection on synonymous codon usage bias.

(a) The conventional model assumes unidirectional selection towards the minimal mismatch between a gene's codon usage and the cellular tRNA supply. (b) We propose a novel model in which the optimal CUB mismatches the cellular tRNA slightly, thereby preventing excessive translational costs. In both panels, the x axis represents D, a measure of the disparity between a gene's codon usage and the cellular tRNA supply (see Methods), and the y axis represents cellular fitness. The black curves represent the x-y relationship assumed in the two models, while the red dot, the gray arrow, and the horizontal dashed line represent the optimal CUB, the direction of translational selection, and the genetic drift/drift barrier, respectively.

Comment 7

Line 72-76: The authors' definition of the "conventional model" is not supported. This statement is a red herring giving the authors the opportunity to disprove a hypothesis that is already quite ridiculous. In no way is the prevailing/conventional wisdom that genes drift toward high CAI. Indeed, if that were the case, there would be no codon usage bias among different genes within an organism. The drift toward or away from matching tRNA supply would be highly dependent on 1) how important the protein is to the cell's survival, 2) what concentration of the protein is required for cell function and 3) under what conditions is the protein essential. To state that drift toward a low mismatch between tRNA supply and demand is ridiculous. The entire premise of this manuscript is based on a red herring.

Response:

Thank you for your comment! We do not say or imply genetic drift toward high CAI, or toward a low mismatch between tRNA supply and demand in line 72-76, or anywhere else in our manuscript. Both our proposed model and the conventional model consider genetic drift to be a counteracting force to selection (therefore constituting a "barrier" to selection). Our model differs from the conventional model in terms of the direction of selection, not direction of genetic drift. Our study are always concerned with determining what is the optimal D (= 0 or not). Genetic drift (and possibly mutation bias too) is then always viewed as a "barrier" that prevents selection from reaching the optimal D. Please see our Figure 1 and Figure 6a/b for a depiction of this logic.

Comment 8

Line 73-76: There is unlikely to be a single answer for this since codon bias toward high and low D exists of different endogenous genes. Again, this depends on the cellular need for the specific function under the conditions tested. There is not likely to be a

single answer. This general premise is highly flawed. The selective pressure is different for different genes.

Response:

Thank you ! As we have explained above, the conventional model posits that selective maximizes translational output, and drift/mutation prevents that from occurring, resulting in a different CUB (but not optimal CUB) for every gene. Our argument is that the direction of selection is not to maximize translational output, but instead to balance the functional payoff and cost of translation, resulting in different balancing/optimal CUB depending on the functional importance of the protein.

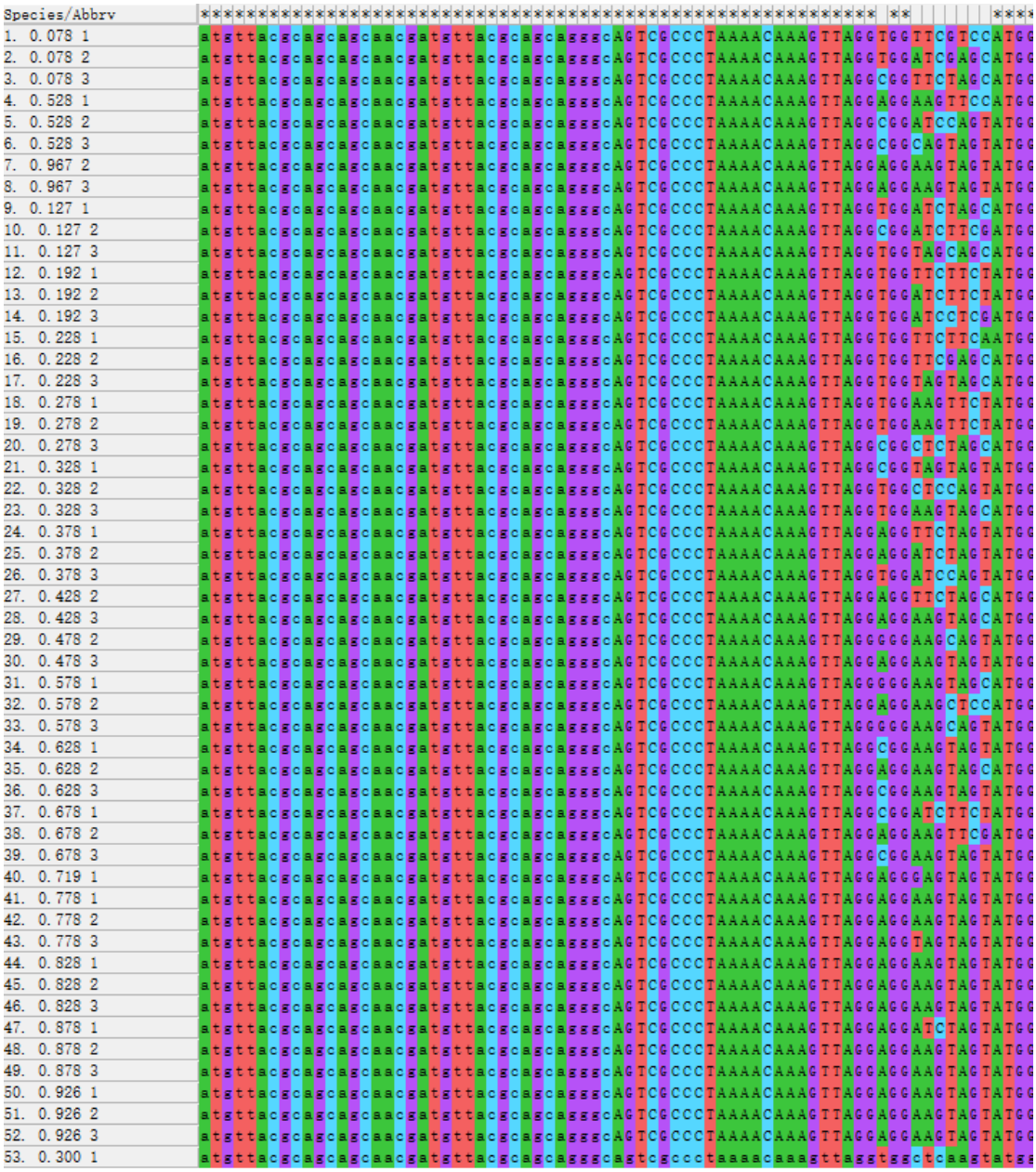
Comment 9

Line 109-117: Authors are not controlling for changes in 5' region that are translation initiation effects. This confounds results. It's not clear whether the effects are due to translation initiation (controlled by mRNA structure & ribosome binding) or elongation as the authors want to claim. Currently, the authors do not prove that the effects are not from transcription which can misrepresent the translation results. The authors need to show that the results of expression level changes cannot be explained by translation initiation changes.

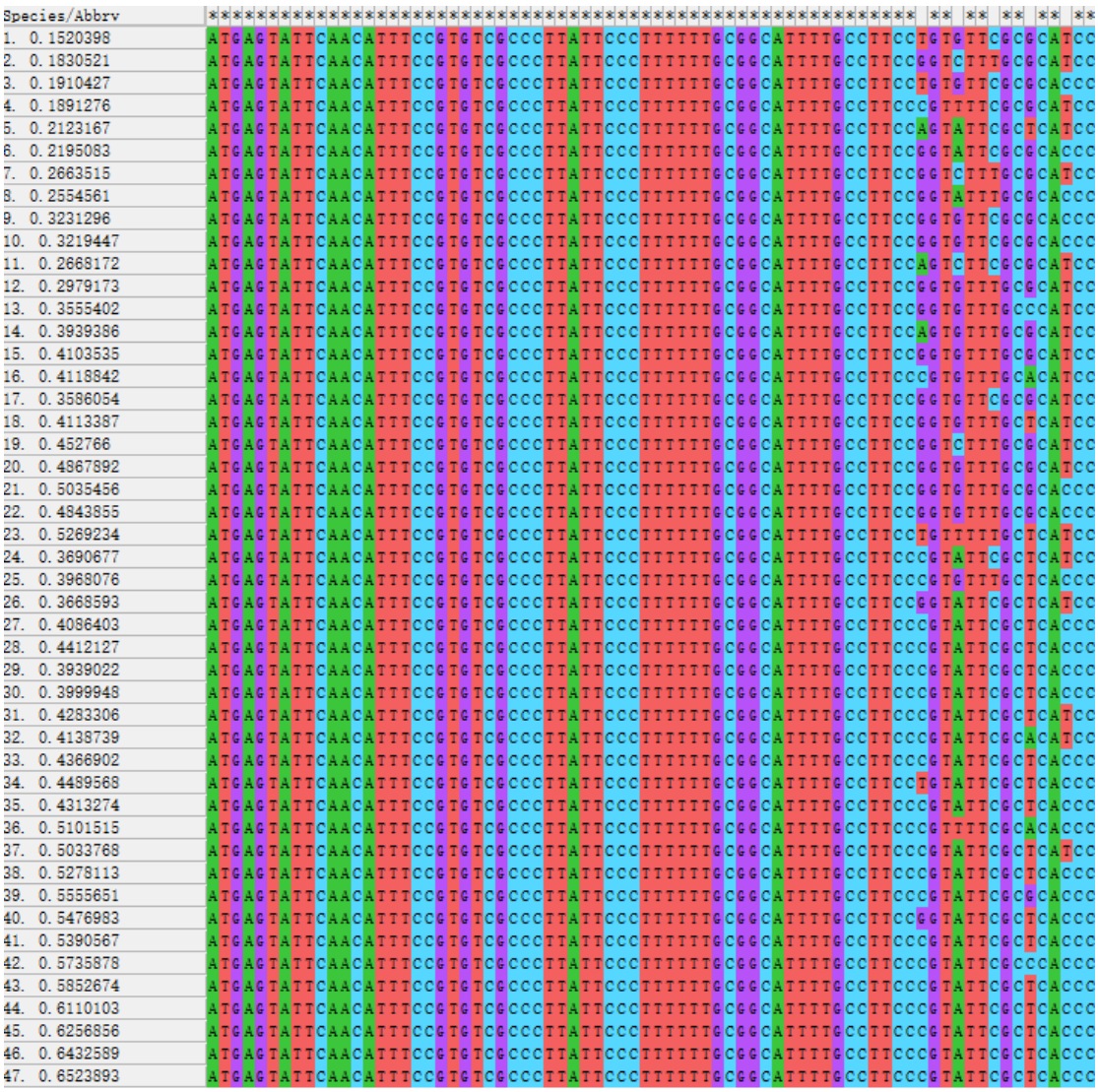
Similarly, the authors do not control for transcription rates. In bacteria, transcription and translation are coupled. The authors should show that their data is not explained simply by changes in mRNA levels of their reporter if they want to claim that it is all translation elongation effects.

Response:

Thank you for this critical point. As for the controlling of translational initiation effects, we would like to clarify that in designing the synonymous mutants for both GmR and AmpR, we deliberately left the first 19 codons of the 5′ region unmodified (details in **Table S2** and **S4**). This region of “translational ramp” has previously been found to regulate translational initiation (PMID: 20403328). Below is a multiple sequence alignment of this region in the designed synonmyous GmR variants. The sequence name to the left indicates the D value and the index of biological repeat of the variant (three different sequences for each D value, but some of them excluded because they failed construction).



Similarly, in our additional experiment with AmpR, the multiple sequence alignment of this region in the designed synonmyous AmpR variants appears below.



Regarding the possibility that our results might be confounded by transcription changes, we would like to separately discussed the issue for GmR and for YFP. As for YFP, we measured YFP mRNA expression using quantitative RT-PCR. We found no correlation between YFP transcript abundance and the *D* value of GmR, indicating that transcription rate changes cannot explain the observed translational cost. The related results have been updated in our manuscript, as shown in our response to your Comment 3.

For GmR, however, we did NOT argue for the complete exclusion of transcription level and initiation differences. This is because, for example, codon usage bias can directly affect the abundance of transcripts independent of promoter activity (PMID:28961875 and 27671647). It is also possible that codon usage downstream of the translational ramp resulted in a ribosome “traffic jam” that ultimately decreased effective initiation rate (PMID: 25148683). Consequently, the increased cost associated with GmRs with minimum *D* can be explained by (i) accelerated elongation and therefore tRNA consumption, (ii) increased effective translational initiation and therefore tRNA consumption per mRNA molecule, (iii) increased mRNA half-lives and

therefore total tRNA consumption accumulated across the entire life cycle of the mRNA, or a combination of these factors. We are, however, arguing that regardless of the specific molecular mechanisms, the cost of increased translational output of GmR due to reduced *D* is readily observable by the reduction of YFP, and also the fitness. We have cited the above related papers and explicitly mentioned these potential molecular mechanisms in line 65-68, as pasted below.

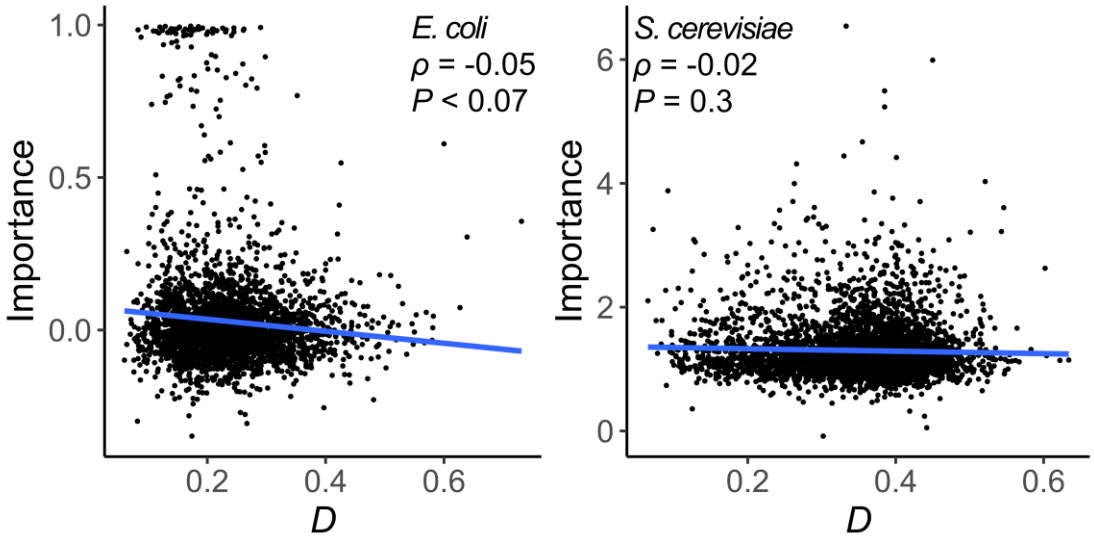
> An exogenous gene with excessively small *D* will have a very high protein expression level due to an increased abundance of mRNA[21,22] and/or more efficient translation[20,23-25], but at the same time, will impose a heavy translational cost on the cell, thus hindering the translation of other genes within the same cell[16,20].

Comment 10

Line 192-195: Authors do not show the data for the KanR gene replacement experiments that is necessary for proving their claim. Needs more information on how functional importance was quantified. Furthermore, Spearman's correlation of -0.08 is extremely weak almost zero. Authors need to show the data spread used for their correlations. Just providing a correlation is unacceptable.

Response:

Thank you for this comment. The experimental data used in this section were taken from previously published studies. Therein, the KanR gene was used to replace various target genes, and the fitness of cells after each gene's deletion was measured. The importance of each gene to the cell is calculated as 1 minus the fitness of the cell with the focal gene deleted. The scatter plot of all genes is shown below.



We fully acknowledge that the Spearman correlation between gene importance and *D*

is weak across all genes. This likely reflects the fact that lowly expressed genes (i.e. majority of the genes) are under weaker selection, so that their observed *D* cannot accurately approximate their optimal *D* (our model predicts a correlation between importance and optimal *D*, but NOT between importance and observed *D*). In contrast, highly expressed genes are subjected to stronger selection, which drives their observed *D* closer to the optimal *D*. Therefore, a correlation between importance and observed *D* is expected only for highly expressed genes. In other words, in line 192-195, the more important pattern is the strengthening of the expected correlation when only highly expressed genes are included. We have explicitly explained this logic in line 199-211, as pasted below.

> . . . our model predicts that a more important gene should have an optimal CUB that is more closely aligned with cellular tRNA supply, and therefore correspond to a smaller *D*. To gauge the optimal CUB for the native genes, we assumed that the native codon usages of endogenous genes have been optimized by translational selection to the maximum extent possible in the presence of genetic drift. Here, the relative strength of translational selection and genetic drift varies for different genes within the genome. By following previous practice[27], we presumed that genetic drift (determined by effective population size) was constant across genes, and that translational selection was determined by mRNA expression levels. Therefore, our hypothesis predicts a negative correlation between a gene's functional importance and the *D* value of its native codon usage, especially for highly expressed genes because they are under stronger translational selection that keeps their native codon usage close to their optimal CUB.

Comment 11

Line 207-213: This statement is not novel, is intuitive and expected, and is not a contribution of this work. It is already well-accepted that high transcription and translation rates are highly burdensome to cells. See PMID: 27989436.

Response:

Thank you for this comment! We agree that the existence of translational cost is not novel. The novelty here is that the optimal CUB does not maximize translational efficiency (rejects the conventional model), but rather balances the functional payoff and translational cost (supports the novel model). In other words, the conventional model on CUB evolution did not take translational cost into account. Please also note that the senior author of paper you mentioned (PMID: 27989436), Dr. Yitzhak Pilpel, praised the novelty of our proposed model in comparison to the conventional model in his signed referee report on our manuscript. His original comments on the novelty are

pasted below.

> The paper titled "A slight mismatch between a gene's codon usage and the cellular tRNA supply is beneficial" describes a very thorough work with a thought-provoking notion. Unlike the current dogma in the field, which assumes that optimal translation efficiency is achieved when codon usage perfectly matches the tRNA pool of the cell, a slight mismatch might be optimal. A mechanism based on feedbacl on the tRNA pool is suggested. A proposed mechanism is based on an alleged trans-regulation and tRNA depletion.
>
> While previously there were some indications for this effect in unneeded genes, now they look for genes with fitness effect, either a drug resistance gene or natural genes. This is a major step forward as it allows them to balance costs against functional payoff.
>
> This is an excellent work. It is thought-provoking, very well done and deep in it evolutionary approach. It is also very well written and technically sound.

Comment 12

Line 217-219: What is the threshold authors set for expression and why? The authors need to show the expression data distribution used for obtaining their correlational data in Figure 5.

Response:

Thank you for your comment. The logic behind expression thresholding is that natural selection (towards optimal *D*) should be stronger for highly expressed genes. As a result, only the highly expressed genes should show a *D* close to optimal D, but not lowly expressed genes. Nevertheless, no empirical information is available to determine the precise expression that will produce an observed *D* which is close enough (for distinguishing conventional from novel models) to the optimal *D*. Following this logic, these section and Figure 5b-e test a major prediction of our model, namely a correlation between optimal *D* and expression, which is only detectable for highly expressed genes since their observed *D* is closer to optimal *D*. So we are simply raising the expression threshold before the number of remaining genes becomes insufficient for statistical analysis.

Per your suggestion, we have now shown some scatter plots (one from each species) underlying the correlations, and compared the observed correlations for highly expressed genes with random gene sets sampled to the same size as the thresholded highly expressed genes. The updated **Figure 5b-e** are pasted below

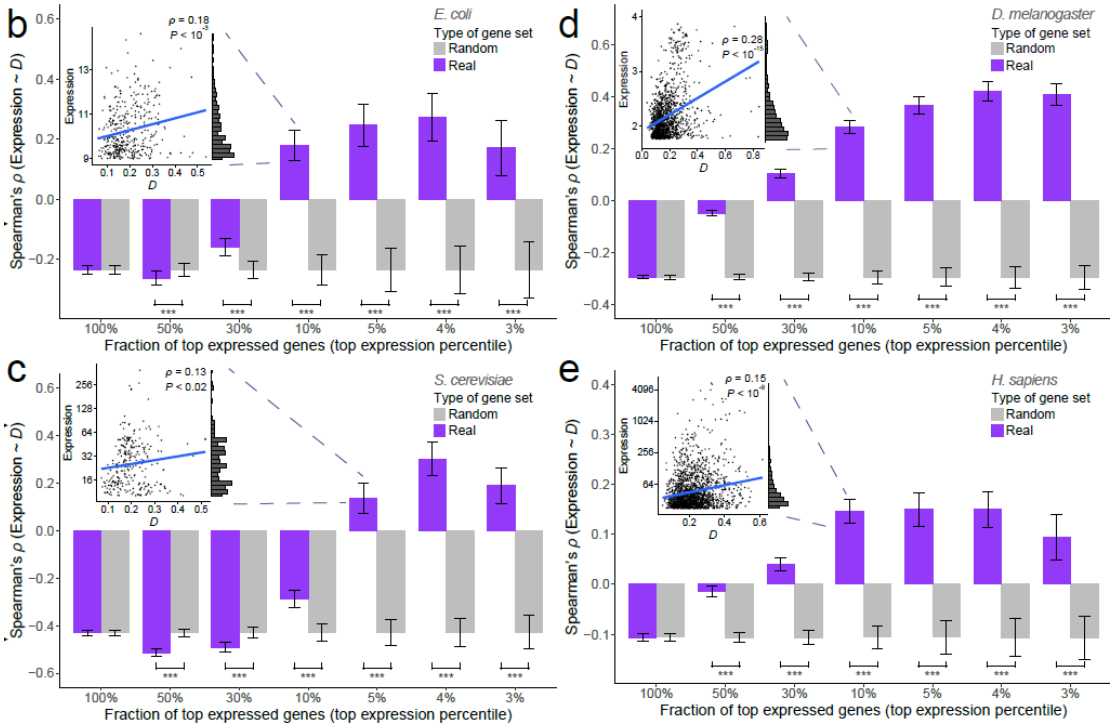


Figure 5. The effect of translational cost on native codon usage of *E. coli* endogenous genes

(**b-e**) For *E. coli* (**b**), *S. cerevisiae* (**c**), *D. melanogaster* (**d**) and *H. sapiens* (**e**), the Spearman's Rank Correlation between a gene's mRNA expression level and *D* of its native CUB (*y* axis) is estimated among endogenous genes whose expression levels exceeds a given threshold (*x* axis). Purple bars indicate results from the real set of genes above the expression threshold, while gray bars represent results from a randomly sampled set of genes consisting of the same number of genes as the real set. Error bars indicate standard deviations of the correlation assessed from 1,000 bootstraps (purple) or random samples (gray). An inset scatter plot represents one correlation from each panel, with histogram to the right indicating the distribution of gene expression. Statistical significance of Wilcox signed-rank test between real and random gene sets is indicated as *: $P<0.1$, **: $P < 0.05$, or ***: $P < 0.01$.

Comment 13

Line 243 and 282-284: E. coli is a gut microbe. This means that evolutionary pressure for optimizing its genes has largely been driven by the gut microenvironment to match its demands. The authors culture their E. coli in LB media. This means that there will likely be a mismatch in the mutational accumulation experiments relevant to D-increasing or D-decreasing effects. It is incorrect to assume that the selective pressure in lab conditions in LB reflects true selective pressure for E. coli.

Response:

We sincerely appreciate your comment. The rationale behind the mutation accumulation (MA) experiment is that there is a general absence of natural selection in MA lines. If the ancestral wildtype strain has been subjected to natural selection towards some particular phenotype, the absence of selection (in MA) could reveal the forces (genetic drift and/or mutation bias) that are counteracting natural selection. By analyzing the direction and effect of such counteracting forces as revealed in the MA lines, we can identify the type of selection that has been acting on the ancestral wildtype strain. As an analogy, when a balloon's inlet is released (absence of selection), if air rapidly exits from inside the balloon (direction of genetic drift/mutation bias), we can infer that the balloon was previously filled with compressed air (selection applied to ancestral wildtype strain).

More specifically in our study, we are not assuming any selective pressure under lab conditions in LB. Instead, we assume a lack of selection in MA lines grown in LB, so that we can determine what type of selection the wildtype *E. coli* experienced in their normal (gut) environment based on the contrasting patterns observed in MA lines (without selection) and wildtype *E.coli* cells (with selection). A similar approach has been used to determine the direction of selection for mutation rates (PMID:34193872).

Comment 14

Line 303: Authors need to clarify what they mean by “waste of tRNAs for unpreferred codons”.

Response:

We are sorry for the confusion. Here the “waste of tRNA for unpreferred codons” means that for a gene with CAI=1, all codons are preferred codons, so that the tRNAs for unpreferred codons are never used, therefore a waste. We have modified this sentence as below.

> For example, the optimal CAI (=1), in which preferred codons are maximally used, cannot in theory be considered optimal from a supply-demand perspective, since the supply of tRNA for unpreferred codons is never used.

Comment 15

Regarding Figure 3.b
- Why does at high Gm concentrations the small group (theoretically with less GmR protein than the minimum group) exhibit a higher functional payoff?

- Why is everything normalized to the large group? This implies you are assuming unidirectionality in codon-match and functional payoff.
- The data contradicts the authors' assumption that as the importance of a gene increases, D (tRNA-CUB supply-demand differential) decreases. At concentrations above 140 ug/ml we see that the small group is preferred. (In text line 154 you say that only 68.9% of the variations are explained by this assumption).

Response:

Thank you for these three questions.

For 1, the payoff difference between the small group and the minimum group is NOT statistically significant at high concentration (>140 ug/mL). The observed minor differences are likely attributable to experimental noise rather than biologically meaningful effects. Also, even in media with high gentamicin concentrations, GmR synthesis cannot exceed a physical limit (e.g., as determined by cell resource uptake rate). Functional payoffs and fitness of the group with the minimum *D* may have been affected by this limit, thus making the difference with the group of minimum/small D less statistically significant.

For 2, we are confused by your statement that we were "assuming unidirectionality in codon-match and functional payoff". We did not make that assumption. More importantly, since the normalization is only for different groups under the same gentamicin concentration, the results of the comparison between the three groups under each concentration will remain qualitatively unchanged regardless of which group we used as the standard for normalization. For example, the large group will be the fittest of the three groups at 0 μg/ml, while the small group will be the fittest at 120 μg/ml, regardless of the way these measurements are normalized under the same concentration.

For 3, we respectfully disagree. The small group is not preferred at 180 μg/ml gentamicin, as it is not significantly fitter than the minimum group (Figure 3f). The 68.9% variation explained refers to the rank variation in functional payoff of the minimum group explainable by gene importance (Figure 3c), not variation of *D* explained by gene importance. Please also see our response to the next comment 16.

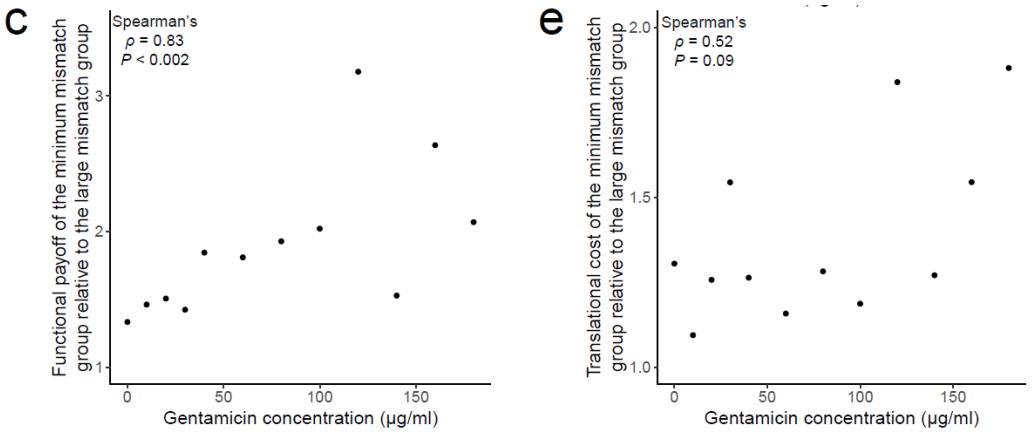
Comment 16

Figure 3.c-e
- It is unclear what the data points here represent. Are they average or individual single recodes? Why is there only 12 points here? Shouldn't you have the minimum and small groups here (38 points)? Authors need to provide information about their statistical test in the figure legend.

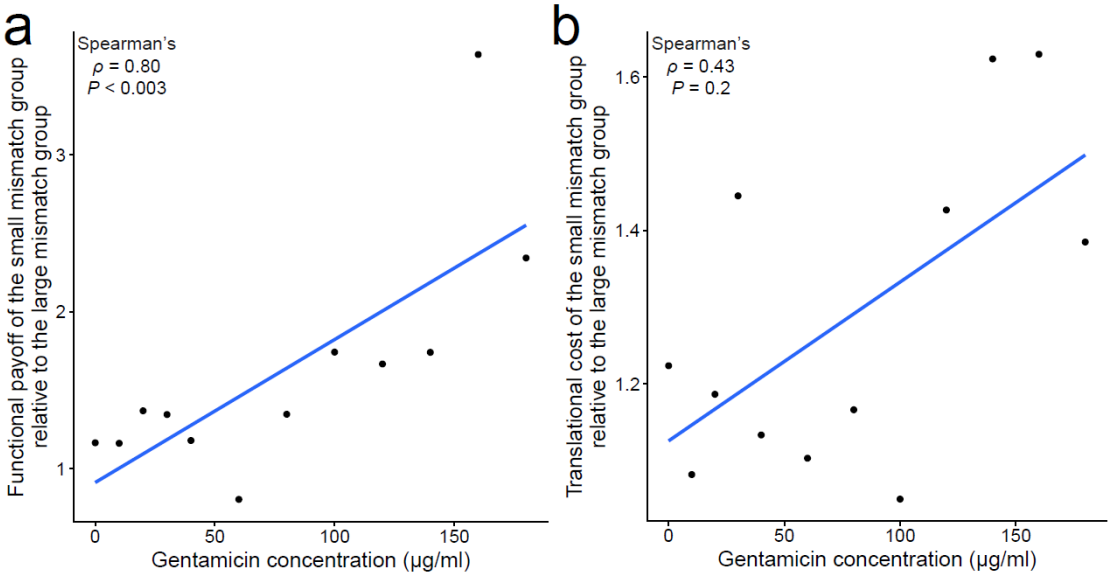
Response:

Thank you and we are sorry for the confusion. Each data point in panel c (d) represents

the functional payoff (translational cost) of the minimum group as compared to the large group, at a specific drug concentration. There are therefore only 12 points, one for each concentration. We have revised this figure to clarify this and the statistical test used, which is pasted below.



It should be noted that if the same analysis were performed using the small group instead of the minimum group, the results remained qualitatively unchanged, as pasted below. We decided to only present the results of the minimum group.



Comment 17

Figure 3.f

- Relevant to result interpretations of line 160-163:

- I disagree that at concentrations above 120 ng/ul minimum D is selected. Mainly because the statistic is weakly probable. From 140 ng/ul all groups most likely have the same relative fitness. This contradicts your statement that minimum D is best at the highest functional importance (highest concentrations).

Response:

Thank you for this question. The optimality of minimum *D* is also supported by two additional results. In Supplementary Figure S1, where at concentration 160 and 180 μg/ml, fitness is significantly negatively correlated with *D*. Thus, in the presence of high concentrations of gentamicin, enhancing GmR expression by minimizing D is most likely to be beneficial. Also, in response to your Comment 1, we have conducted additional experiments using AmpR, the ampicillin resistance gene. Consistent with our model, we observed that at high ampicillin concentration (4000 μg/ml), minimum D is the fittest group. More details on this is given in response to your Comment 1. And this result has been added to the new **Supplementary Figure S5**.

We have chosen to present the grouped result in the main figure and the correlation in the supplementary. The rationale for this is twofold. First, it seems obvious that, at high gentamicin concentrations, GmRs with a smaller D and a higher expression level should be fitter than those with a larger D and a lower expression level, unless the expression is too high for the physical limit defined by the resource uptake rate of the cell. Second, and more importantly, the most critical pattern for our argument is the small mismatch group being fitter than both the minium mismatch group and the large mismatch group at intermediate drug concentrations. The current main figure with three groups illustrates this most clearly.

Comment 18

Figure 4
- This correlation does not support any direct like to the hypothesis that genes with smaller D scores have a higher functional payoff. Because the function is related to the direct role of the gene and not how it is coded. In order to support your hypothesis, you need to have direct fitness data where you measure the fitness cost associated with recoding a specific vital gene and replacing it with the native gene to determine the link between functional payoff and translation cost.
- With regards to line 211-213 An important gene does not have to be highly expressed. You cannot establish the importance of genes by only looking at highly expressed genes and measuring expression indirectly by the D index.

Response:

Thank you for giving us the opputunity to clarify this.

For 1, we are not hypothesizing that genes with smaller *D* have a higher functional payoff (although we believe this is likely to be the case, since small D gives higher expression and, therefore, higher functional payoff). Instead, our hypothesise is that a more important gene should have a *D* value that is more closely aligned with the cellular tRNA supply. There are two critical differences between your comment and our hypothesis. First, a gene's importance is determined by its functional payoff minus its

translational cost, so a more important gene is not necessarily one with a higher functional payoff, but rather one with a lower translational cost. Second, our hypothesis posits functional importance as the cause and optimal *D* as the consequence, rather than the other way around.

We agree your suggested experiment could be useful, and your reasoning behind this design coincides quite well with our manipulative experiments using recoded antibiotic resistance genes. However, because similar experiments have not been conducted for every genes in a genome yet, we turn instead to an indirect test, which is described in the Result section, as pasted below.

> … our model predicts that a more important gene should have an optimal CUB that is more closely aligned with cellular tRNA supply, and therefore correspond to a smaller *D*. To gauge the optimal CUB for the native genes, we assumed that the native codon usages of endogenous genes have been optimized by translational selection to the maximum extent possible in the presence of genetic drift. Here, the relative strength of translational selection and genetic drift varies for different genes within the genome. By following previous practice[27], we presumed that genetic drift (determined by effective population size) was constant across genes, and that translational selection was determined by mRNA expression levels. Therefore, our hypothesis predicts a negative correlation between a gene's functional importance and the D value of its native codon usage, especially for highly expressed genes because they are under stronger translational selection that keeps their native codon usage close to their optimal CUB.

We hope this explanation helps clarify the logic behind **Figure 4**.

For 2, as you point out, and we agree, an important gene does not necessarily have to be highly expressed. We are not measuring expression by *D*, nor are we suggesting that highly expressed genes are always important. Instead, we are stating that highly expressed genes are more sensitive to reductions in *D*. This is because they, by definition, have more transcripts. A reductions in *D* in a highly expressed gene will therefore have a greater impact on the tRNA supply-demand than a same reduction in *D* in a lowly expressed gene. This then predicts a larger optimal *D* in highly expressed genes than lowly expressed genes of similar importance. Please note here, expression is the cause and optimal *D* is the consequence, not the other way around. We have clarified this point as pasted below.

> More specifically, a *D*-decreasing mutation in a gene with high transcript abundance should result in a greater increase of translational cost than that in a gene with low transcript abundance, since a greater number of RNA molecules are mutated in the former case. Consequently, highly expressed genes should be more sensitive to reductions in *D*, resulting in a larger optimal *D* in highly expressed genes than lowly expressed genes of similar importance

(**Figure 5a**).

Please also see our response to the next Comment 19.

Comment 19

Figure 5
- In regards to line 216: Because D and expression level are anticorrelated does not imply anything about translational cost. You are only looking at functional payoff. Translation cost can only be measured experimentally.

Response:

Thank you. To illustrate our logic, consider two genes with identical *D* values, but different transcriptional levels: Gene 1 has 1000 mRNA molecules, while Gene 2 has only 1. If a mutation that reduces *D* to the same level in both genes, the extra translational resources required by Gene 1 would be 1,000 times greater than those required by Gene 2. This example illustrates that, genes with higher mRNA abundance are more sensitive to reductions in *D*, therefore they tend to have a larger optimal *D*. In contrast, genes with lower expression levels should be less sensitive to reduction in *D*, thus they tend to have a smaller optimal *D* values (although their observed *D* are usually not sufficiently close to their optimal *D* sine translation selection is weak on these genes).

We would like to emphasize that this prediction of a positive correlation between gene expression and optimal *D* among highly expressed genes is unique to our novel hypothesis. If only functional payoff is considered, such as in the conventional model, the optimal *D* should be negative correlated with gene expression (because selection minimize *D* to enhance translational efficiency/accuracy in highly expressed genes). We have also illustrated this logic in **Figure 5a**.

Comment 20

Question on the methods:
- How did the authors control for the sample size reduction of highly expressed genes? Correlation scores tend to shrink in absolute values as sample sizes grow. Conversely, a strong correlation at a small sample size may not be true when the sample of the correlation is increased.

Response:

Thank you for this important methodological question. To address potential bias associated with varying sample sizes, particularly the reduction in the number of genes when the analyses is limited to highly expressed genes, we randomly down-sampled

the genes to the same size as those retained at each expression threshold. Through this approach, we are able toassess whether the observed correlations remain consistent across comparable sample sizes, thus ensuring that differences in group size do not significantly affect the strength and significance of the correlations. This additional analysis further supports our conclusions in both **Figure 4** and **Figure 5**, where subsetting by expression threshold is conducted. The revised **Figure 4** and **Figure 5** is pasted below.

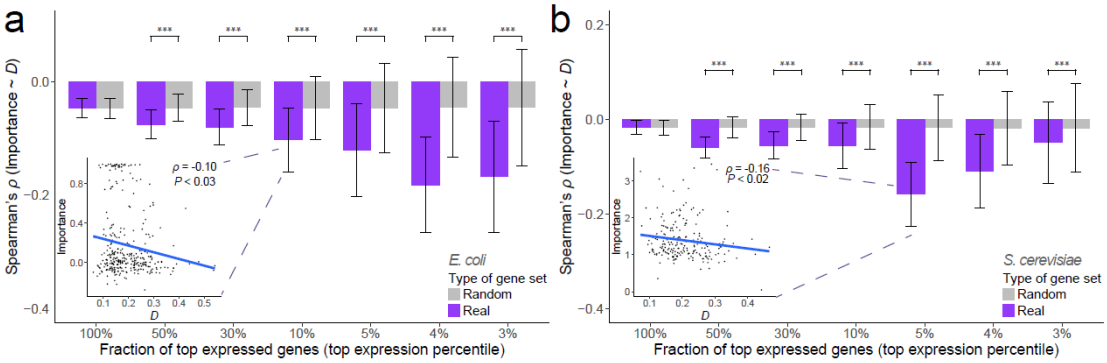


Figure 4. The effect of functional payoff on native codon usage of E. coli endogenous genes

(**a**-**b**) The Spearman's Rank Correlation between a gene's functional importance and *D* of its native CUB (*y* axis) is estimated for *E. coli* (**a**) or *S. cerevisiae* (**b**) endogenous genes whose expression levels exceed a specific threshold (*x* axis). Error bars indicate standard deviations of the correlation assessed from 1,000 bootstraps (purple) or random samples (gray). An inset scatter plot represents one correlation from each panel. Statistical significance of Wilcox signed-rank test between real and random gene sets is indicated as *: <0.1, **: $P < 0.05$, or ***: $P < 0.01$.

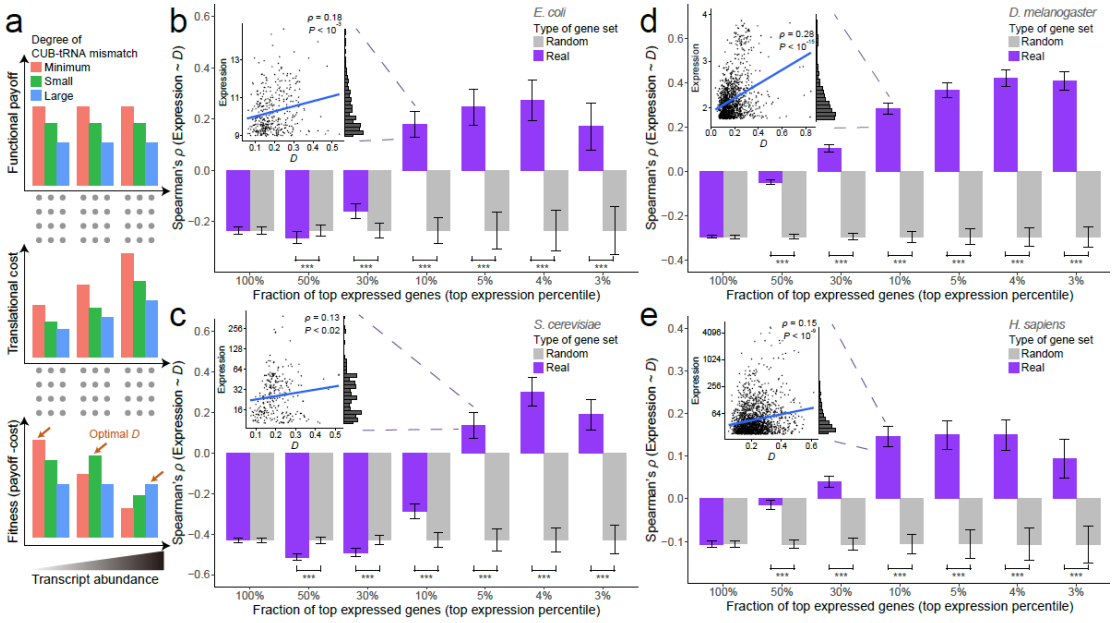


Figure 5. The effect of translational cost on native codon usage of E. coli endogenous genes

(**a**) A diagram illustrating why the optimal D would be higher for genes with greater translational costs. (**b-e**) For *E. coli* (**b**), *S. cerevisiae* (**c**), *D. melanogaster* (**d**) and *H. sapiens* (**e**), the Spearman's Rank Correlation between a gene's mRNA expression level and *D* of its native CUB (y axis) is estimated among endogenous genes whose expression levels exceeds a given threshold (x axis). Purple bars indicate results from the real set of genes above the expression threshold, while gray bars represent results from a randomly sampled set of genes consisting of the same number of genes as the real set. Error bars indicate standard deviations of the correlation assessed from 1,000 bootstraps (purple) or random samples (gray). Statistical significance of Wilcox signed-rank test between real and random gene sets is indicated as *: $P<0.1$, **: $P < 0.05$, or ***: $P < 0.01$.

Comment 21

Figure 6
- While this evidence does support the authors' claim that a slight mismatch with the tRNA supply is selectively maintained, the authors do not quantify the strength of the D-changing mutations. In other words, what is the distribution of D score change in each D-increasing/-decreasing mutation group?

Response:

Thank you for this question. As for **Figure 6c-h**, the effect size (absolute value of *D* score changes) is analyzed in **d/e/g/h**. For results in **Figure 6i** and **j**, we have now added a **Supplementary Figure S8** to show the distribution of *D* changes for each mutation group, which is pasted below. Since the effect size of *D*-decreasing and *D*-increasing mutations is similar, the observation in **Figure 6i** and **j** is not confounded by the effect size difference between the two types of mutations.

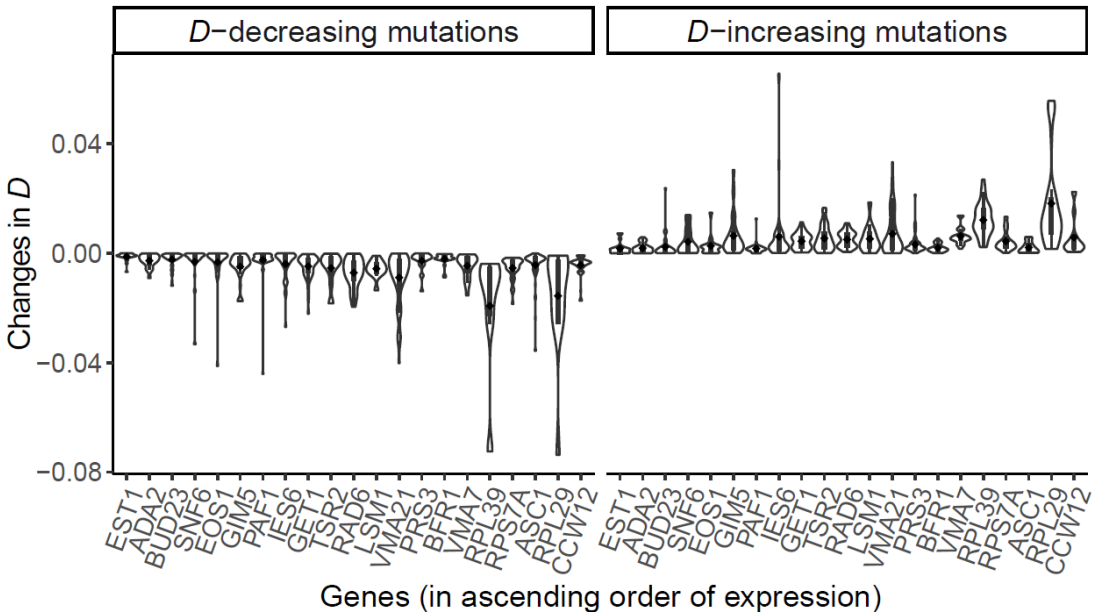


Figure S8. The distribution of *D* changes in *D*-decreasing and *D*-increasing mutations

Violin plots showing the changes in *D* due to the mutations listed in **Figure 6j**. The effect size of *D*-decreasing and *D*-increasing mutations are similar (Wilcox rank sum test $P = 0.10$).

Comment 22

Another note regarding the authors' quantification of cellular tRNA:

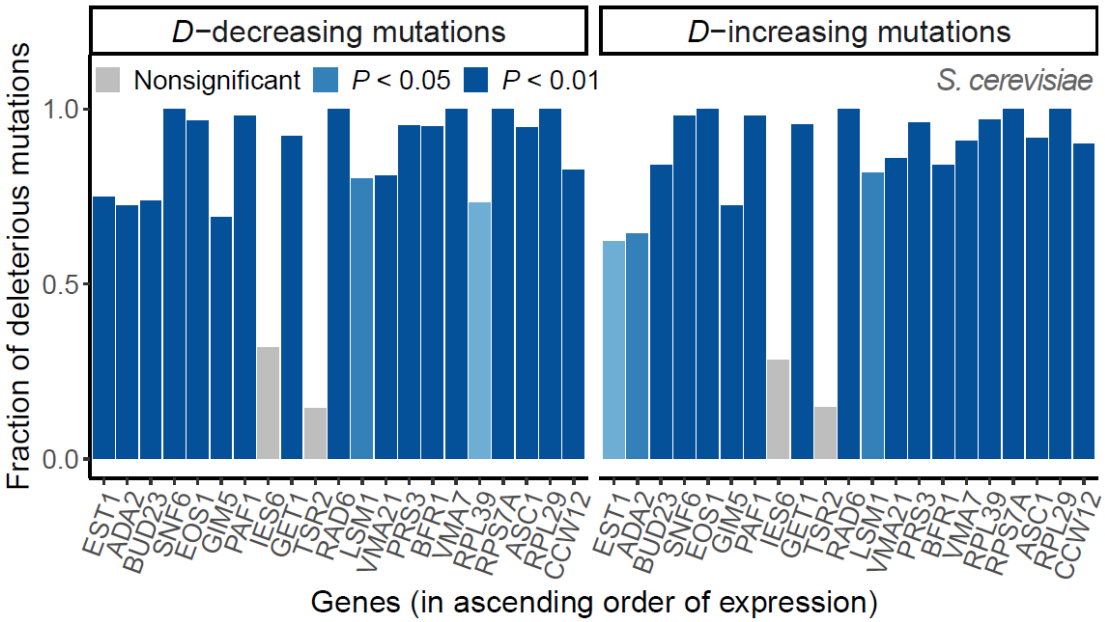
Is the index for quantifying the cellular tRNA pool linear in its entire range? Perhaps if this index becomes non-linear in either boundary it may falsely inform that D decreasing mutations are deleterious at highly expressed genes.

Response:

Thank you for this comment. We did not claim or assume that *D* linearly represents the tRNA supply or the tRNA supply-demand mismatch. We did not make conclusions based solely on the linearity assumption, such as Pearson's correlation tests with *D* (we always accompany them with Spearman's rank-based tests). The reasons for such non-linearity can be biological (the degree of mismatch is not linear to the phenotypic consequences) and/or technical (different amino acids might have different impacts, but we are merely using their geometric average).

Also, we have explained (in response to your **Comment 2**) the rationale behind estimating tRNA supply using transcriptomic codon usage of highly expressed genes, and our conclusion remained unchanged if high-throughput tRNA sequencing data is used to approximate tRNA supply instead.

More specifically for this comment (presumably you are referring to **Figure 6i**), we repeated it with *D* calculated using tRNA gene copy number in wildtype yeast (from GtRNAdb). As shown in the figure pasted below, the observation that an overwhelmingly large percentage of both *D*-decreasing and *D*-increasing mutations are deleterious remains unchanged. As a result, we exclude the possibility that the non-linearity of *D* might be responsible for this observation.



Revewer #2:

Overall Comment:

The manuscript submitted by Chen, Liu, Zhou and co-authors is dedicated to the study of translational selection. Specifically, the authors propose a new model to explain the evolution of codon usage bias (abbreviated CUB; this can be roughly defined as the departure from a random distribution among synonymous codons), according to which the optimal codon usage does not (always) perfectly match the abundance of tRNAs in the cell. According to this model, a mismatch between the codon usage bias of a given gene and the tRNA supply can be beneficial because it reduces the cost imposed by the translation of the gene. In other words, a highly expressed and highly translated gene would be costly because its translation would hamper the translation of other genes.

To support this model, the authors mainly rely on genome editing experiments performed in E. coli. Specifically, they insert cassettes carrying a gentamicin-resistance gene (GmR) with different codon usage biases. The GmR gene is fused to a mCherry coding sequence, which enables the authors to measure the expression level of the

GmR-mCherry fusion through fluorescence measurements. The cassettes also contain a YFP gene in the opposite orientation. The expression level of the YFP gene, again estimated through fluorescence measurements, enables the authors to estimate the translational cost of the CUB of the GmR gene as the reduction in YFP protein levels. The authors perform these experiments at different concentrations of gentamicin and they also measure the growth rate of the different E. coli clones, thus obtaining a more direct estimation of the fitness effect of the different constructs. In addition to these main experiments, the authors also analyze previously published data for mutation accumulation or gene replacement experiments in E. coli and yeast, as well as codon usage data for various species.

I find the model proposed by the authors to be interesting and reasonable. It brings an alternative to the classical model which posits (even in species where translational selection is efficient, that is, species with high effective population size such as E. coli, yeast or fruitfly) translational selection cannot optimize the codon usage bias to perfectly match the tRNA abundance of the cell because of genetic drift. However, it must be stated that the classical model already incorporates the idea of a difference between gene categories (the translational selection coefficient is expectely to vary among genes, depending on the optimum protein quantity produced by each gene, and for some genes the translational selection coefficient is indeed expected to be very low). More importantly, I am concerned about the validity of several assumptions made in this article and of several analyses performed in this article.

Response:

We are truly grateful for your positive assessment of our work. In accordance with your suggestions, we have conducted additional analyses and incorporated new data to address the points raised. Please find our point-to-point responses below.

Comment 1

First, the authors state that the D metric that they define is a measure of the mismatch between the CUB of a gene and the tRNA supply of the cell. It is in fact a measure of the mismatch between the CUB of a given gene and the CUB of the highly expressed genes (Methods, lines 358-373). The authors assume that the CUB of highly expressed genes has been optimized to match the tRNA supply, although they then affirm that the optimal CUB should not perfectly match the tRNA supply, even for highly expressed genes. This is circular and illogical. There are better ways to approximate the tRNAs supply without assuming that the CUB for highly expressed genes has been optimize to match it. If the authors are unable to measure directly tRNA abundance, they could use an approximation based on tRNA gene copy number, taking into account wobble rules etc. As this affirmation is the basis of the new model proposed by the authors, it is very important to make sure that D indeed measures the distance compared to the CUB that

would perfectly match the tRNA supply of the cell.

Response:

Thank you very much for your suggestion. We would like to address from two angles:
1. The assumption on CUB of highly expressed genes has been optimized to match the tRNA supply

The logic behind our estimation of tRNA supply is twofold. **First**, the actual tRNA supply is neither the copy number of the tRNA gene nor the abundance of tRNA transcripts. Instead, it is the abundance of the ternary complex of aminoacyl-tRNA+EF-Tu+GTP (or replace EF-Tu with EF1A in eukaryotes). Currently, the abundance of ternary complexes for different types of tRNAs cannot be systematically measured. In addition, the kinetic parameters between different codon-anticodon pairs will also influence the "effective" tRNA supply, further complicating the estimation process. **Second**, because of the first point, we resorted to the assumption that the cellular tRNA supply and demand (codon usage) of the whole transcriptome have been evolutionarily optimized by natural selection to ensure a close match between supply and demand. This assumption allows us to approximate the supply of tRNA by using the codon usage (the demand) of the top 10% expressed gene (which dominate the transcriptome) in the transcriptome. It is important to note that we are using the transcriptome, not the genome, which means the codon usage has been weighted by the expression of those genes before being summed up. Furthermore, this assumption of supply-demand match has been empirically supported (PMID: 22479199 and PMCID: PMC12262485), and has been used in our previous publications (PMID: 32123323 and 34977494). It is also important to note that a supply-demand match on a transcriptome level does not necessarily imply a supply-demand match for individual genes. Therefore, this assumption is neither circular nor illogical.

2. Using measurements of tRNA transcript abundance for tRNA supply does not change our conclusion.

Per your suggestion, we obtained tRNA expression levels based on high-throughput tRNA sequencing data from a previous publication (PMID: 31353208, GEO accession GSE128812), which were then scaled by codon-anticodon affinity and used as the tRNA supply. This metric is similar to the “adaptiveness value” (w) of a codon, as used in the popular tAI metric (PMID:15448185). It was found that the grouping (minimum, small, or large D) in **Figure 3** remained unchanged, suggesting that our conclusion is not affected by the change in the method used to quantify tRNA supply. We have now included this result in **Supplementary Figure S2**, and it is pasted below.

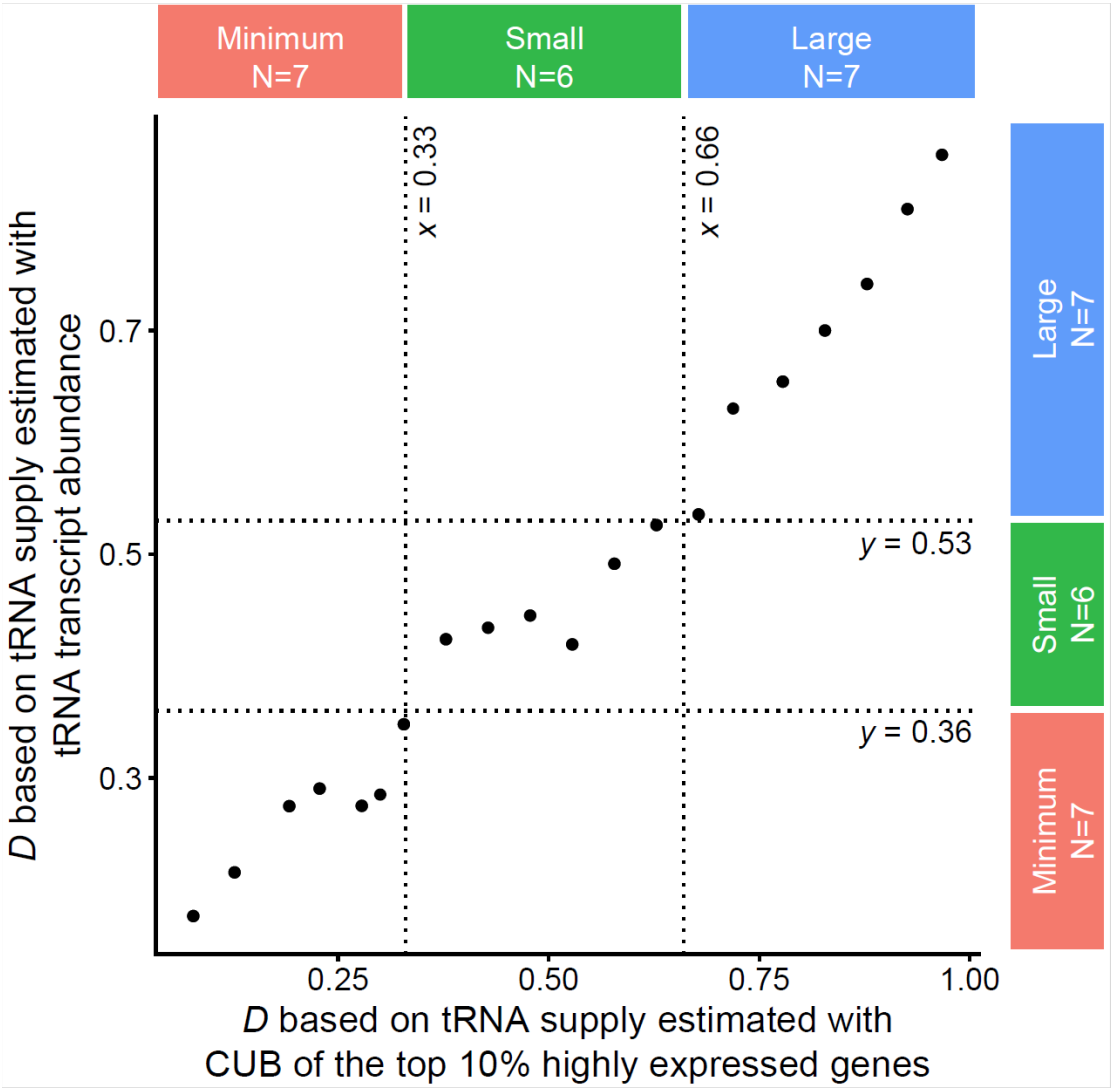


Figure S2. Grouping by D of GmR strains remains unchanged if high-throughput tRNA sequencing data are used to estimate tRNA supply

The D values of designed GmR synonymous variants were calculated with tRNA supply based on either transcriptomic codon usage of the top 10% highly expressed genes (x axis) or tRNA transcript abundance derived from high-throughput tRNA sequencing (y axis). The black lines indicate the grouping, which is identical for both types of D values. Please note that each dot has up to three unique sequences as biological replicates (see **Table S2**).

Comment 2

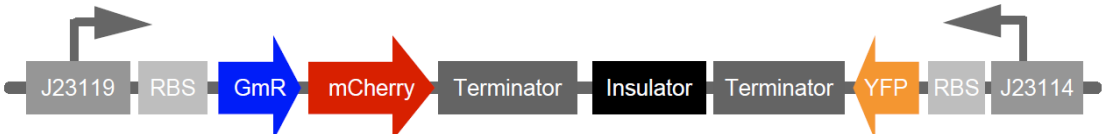
Second, the authors measure the translational cost of the CUB of the GmR artificial constructs by measuring the fluorescence level of the exogenous YFP present in the construct. However, they do not ascertain that variations in YFP fluorescence levels are indeed to variations in YFP translation, rather than for example in variations in YFP

gene transcription. They also do not show that the variations in YFP fluorescence are causally connected to the translation of the GmR-mCherry gene. The authors and others have shown before that variations in CUB can also affect the mRNA level of genes, mainly due to mRNA stability. Additional effects ca be imagined at the transcriptional level, given the overall differences in base composition of the CUB variants. As the YFP gene is inserted in close proximity to the GmR-mCherry gene, on the reverse strand, transcriptional interference can occur between the two genes. Thus, their claim of a translational cost is not sufficiently supported. Furthermore, the translational cost should be evaluated on endogenous genes. As this is an important claim of the manuscript, it should be properly verified, by evaluating mRNA levels and translation rates in the different constructs (for example with RNA sequencing and riboosome profiling assays).

Response:

Thank you very much for this insightful comment. Please allow us to address these points separately.

To prevent transcriptional interference between YFP and GmR, we have included an insulator between the terminators of YFP and GmR to prevent transcriptional interference between the two genes. The **Figure 2b** pasted below has been updated to explicitly indicate the insulator.



Furthermore, we used quantitative RT-PCR to measure YFP mRNA expression levels in the GmR strains growing in LB. We found that YFP mRNA expression did not correlate with GmR's codon usage (*D* value). This result has been added to **Supplementary Figure S9c** and pasted below.

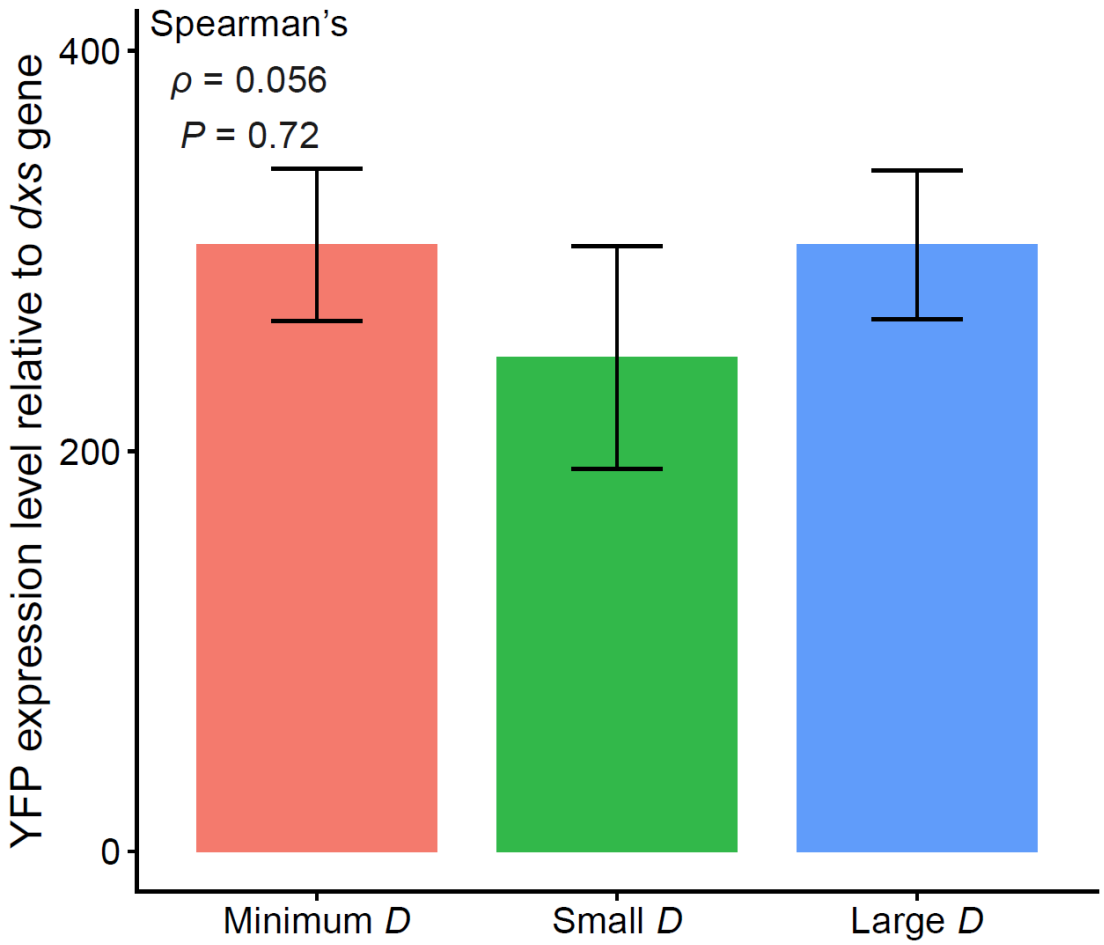


Figure S9. Correlations between *D* and functional payoff, translational cost or fitness

(**c**) The abundance of YFP mRNA in different strains, as determined by quantitative RT-PCR (see **Methods**), is not correlated with codon usage of GmR. The Spearman's Correlation Coefficient ρ of the raw data, as well as the corresponding *P* value, are indicated.

To evaluate the effect of GmR's translational cost on endogenous genes, we performed ribosome profiling (Ribo-seq) on four strains with varying *D* values (0.127, 0.278, 0.778, and 0.878) of the GmR gene. It is observed that when the GmR gene has a minimum *D* (*D* = 0.127, 0.278), the codons frequently used by GmR tend to have a longer decoding time, leading to a significant positive correlation between decoding time and GmR codon consumption among genes with sufficient Ribo-seq coverage. This observation suggests that tRNA depletion by codons frequently used in GmR is slowing down the elongation of the same codons in other endogenous genes. In contrast, strains expressing GmRs of large *D* did not show a similar correlation. This result has now been added to the Result section and as **Supplementary Figure S3**, which is pasted below. Please note that RNA-seq is not required here as the relative decoding time is always normalized within the same gene first, and regardless of the RNA-seq results, all codons within the same gene will have the same transcription level.

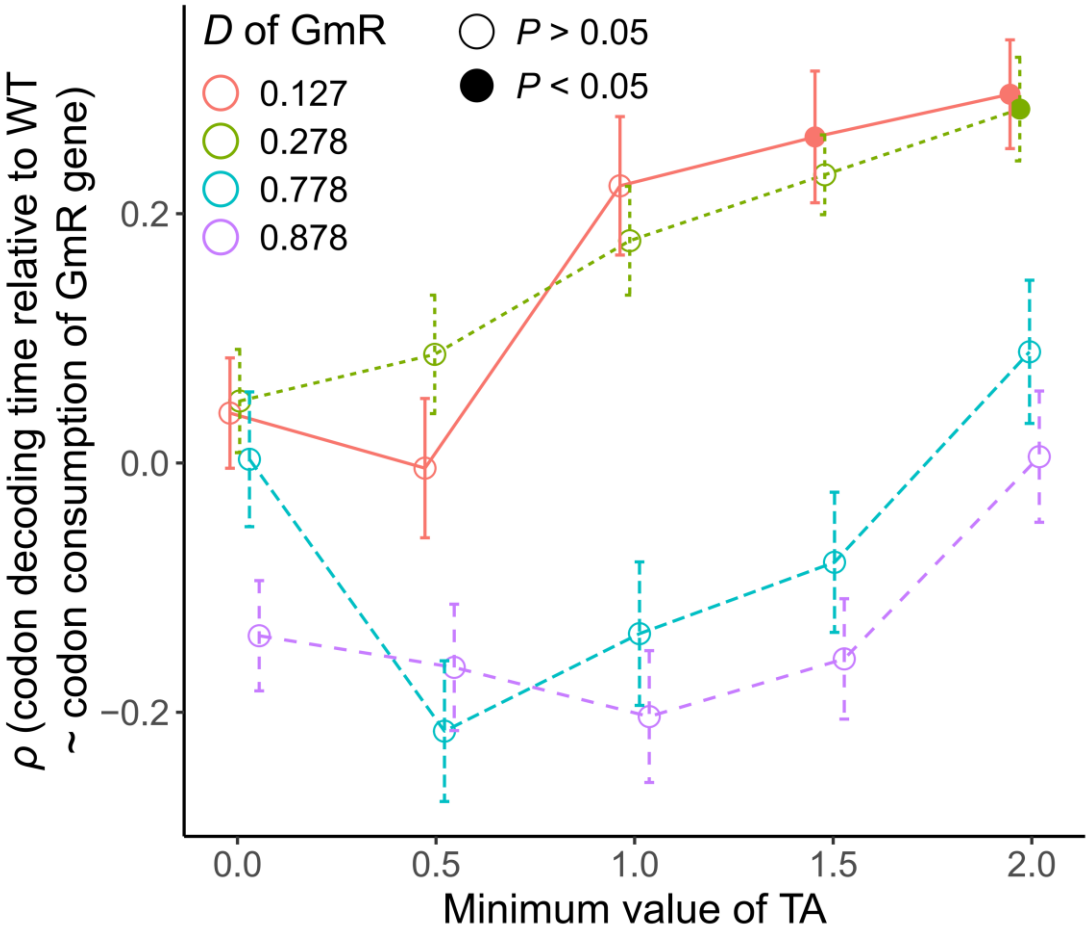


Figure S3. Slowed decoding as a result of GmR translation

The decoding time of each codon in endogenous *E. coli* genes was measured by Ribo-Seq in wildtype and four E. coli strains expressing GmRs of different *D* (see **Methods**). The Spearman's correlation coefficient between GmR's consumption of a codon and its relative decoding time as compared to the same codon in wildtype *E. coli* (*y* axis) is calculated for endogenous genes above a given minimum value of translational activity (*x* axis) (see **Methods**).

Together, these results suggest that genes with minimum *D* impose a greater burden on the translation of other endogenous genes, whereas those with large *D* contribute minimally to the overall translational burden.

Comment 3

Third, the fact that the different E. coli strains that they used could display variations in plasmid copy number is quite concerning. I am not convinced that the lack of correlation between the copy number and D is sufficient to ensure that the results are not biased. The authors should verify that the plasmid copy number variation is not a confounding factor for fitness analysis and for translational cost analyses.

Response:

Thank you for your comment. According to your suggestion, we calculated the partial Spearman's correlation between the copy number and functional payoff/translational cost/relative fitness while controlling for the plasmid copy number. The results shown in **Figure S9b**, and pasted below, are nearly identical to those in **Figure S1**, which are the same correlations without controlling for the plasmid copy number. This suggests that plasmid copy number variation is not a confounding factor in these analyses.

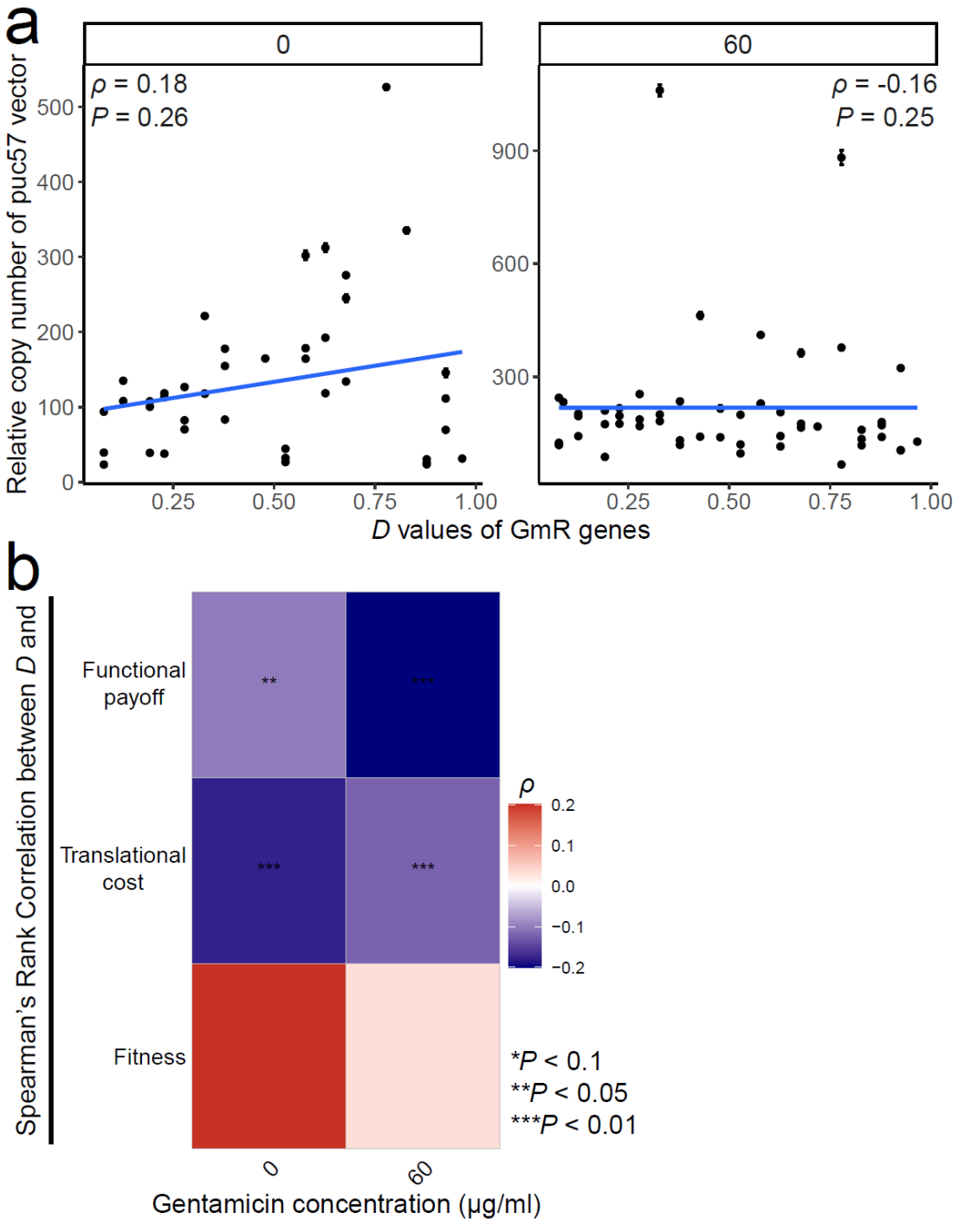


Figure S9. Copy number of GmR genes identified by quantitative real-time PCR.

(**a**) The dots and error bars represent the average and standard error of the copy number of GmR genes relative to *dxs* as measured by qPCR. The strains growing in LB medium supplemented with 0 μg/ml (left panel) or 60 μg/ml (right panel) gentamicin is analyzed. The Spearman's Rank Correlation Coefficient (ρ) and corresponding *P* value are shown at the top left corner. (**b**) The partial Spearman's correlation between *D* and functional payoff/translational cost/fitness controlling for the measured plasmid copy number.

Please also note that, we did NOT argue for the complete exclusion of translational cost contributed by increased plasmid copy number or increased GmR mRNA level. This is because, for example, codon usage bias can directly affect the abundance of transcripts independent of promoter activity (PMID:28961875 and 27671647). It is also possible that in high gentamicin concentration, strains with a high plasmid copy number exhibit growth advantage for its high baseline GmR expression. Essentially, the increased cost of GmR can be mediated by increased DNA abundance, mRNA abundance, and translational activity per mRNA. As long as these, or other potential mechanisms, simultaneously mediated the increased functional payoff AND translational cost, they are compatible with our model. We are, however, arguing that regardless of the specific molecular mechanisms, the cost of increased translational output of GmR due to reduced *D* is readily observable by the reduction of YFP, increased decoding time of codons frequently used by GmR, and also the fitness.

Comment 4

Fourth, the fact that the authors observe a positive correlation between D and expression levels among the most highly expressed genes (top 10%, top 5% and top 3%) in all 4 analyzed species (Figure 4) is puzzling. In human, translational selection is highly inefficient, unlike in the other 3 species. The fact that the same effect is observed in human, E. coli, yeast and Drosophila rather points to a more general effect, potentially a confounding effect of base composition. This should be examined carefully.

Response:

Thank you for this insightful comment. We believe you mean Figure 5, as Figure 4 is about gene importance and D, and only includes two species. We agree that human should have a less efficient translational selection, for it has a relatively small effective population size. And this is consistent with the observation that for most highly expressed genes, the positive correlation between D and expression levels are weaker in human ($\rho < 0.2$), but stronger in other species ($\rho > 0.2$).

To further exclude the potential confounding effect of base composition, we repeated the observed correlations by partial correlations that controlled A%, C%, G% T% or GC% among those genes. As shown in **Figure S6**, which is pasted below, controlling for base compositions does not affect our observation of a positive correlation between gene expression and *D* among highly expressed genes.

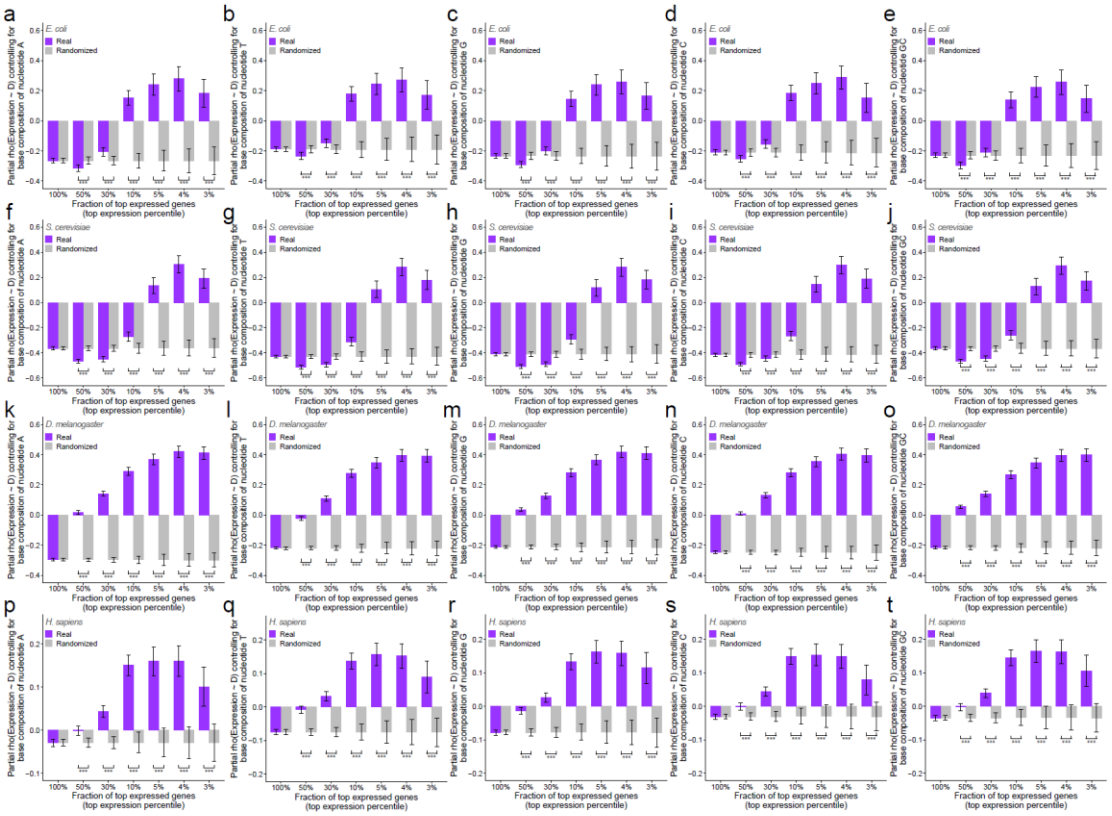


Figure S6. The correlation between *D* and expression for endogenous genes are not confounded by base composition bias

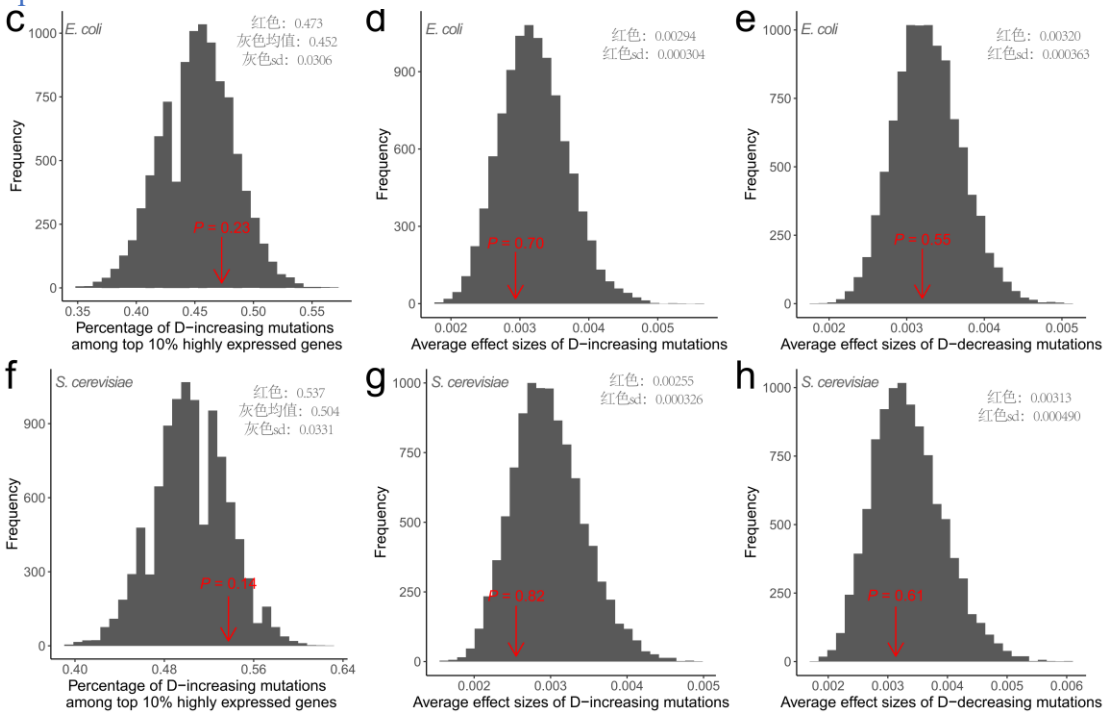
Each column of panels is similar to **Figure 5b-e**, except that the Spearman's correlation in **Figure 5b-e** is replaced with partial Spearman's correlation controlling for the percentage of Adenine (A%, in panels a/f/k/p), Thymine (T%, in panels b/g/l/q), Guanine (G%, in panels c/h/m/r), Cytosine (C%, in panels d/i/n/s) or G+C% (in panels e/j/o/t) in the sequences.

Comment 5

Fifth, it seems that the mutation accumulation analysis was performed using both synonymous and nonsynonymous mutations (only nonsense mutations seem to have been excluded, Methods, line 464 to 466). This does not make sense - if one wants to measure the effect of the D, only synonymous mutations should be considered. It is also not clear why this analysis was only performed on the 10% most highly expressed genes (especially given that D measures the distance to the codon usage of the most highly expressed genes). Here, a comparison between gene expression classes would be useful, to test the hypothesis that the CUB of the highly expressed genes has indeed been optimized. Here again, D should be computed with respect to the codon usage that perfectly matches the tRNA abundance (not to the codon usage of highly expressed genes).

Response:

Thank you very much for this comment. We now only consider synonymous mutations from MA lines. The result supports our model, and is pasted below. We have also updated the Method section to reflect this.



Also, we would like to explain why this analysis is limited to the 10% most highly expressed genes. The rationale behind the mutation accumulation (MA) experiment is that there is a general absence of natural selection in MA lines. If a gene in the ancestral wildtype strain has been subjected to natural selection towards some particular phenotype, the absence of selection (in MA) could reveal the forces (genetic drift and/or mutation bias) that are counteracting natural selection. By analyzing the direction and effect of such counteracting forces as revealed in the MA lines, we can identify the type of selection that has been acting on the ancestral wildtype strain. As an analogy, when a balloon's inlet is released (absence of selection), if air rapidly exits from inside the balloon (direction of genetic drift/mutation bias), we can infer that the balloon was previously filled with compressed air (selection applied to ancestral wildtype strain). Based on this logic, studying lowly expressed genes would be useless, since their ancestral sequences have not been subjected to strong selection on *D*, as suggested by results shown in Figures 4 and 5.

As for the question regarding the definition of *D*, please refer to our response to your Comment 1. We hope you found these explanations satisfactory.

Comment 6

Finally, the authors might be interested in this preprint, which shows that translational selection is surprisingly rare in metazoans:

https://www.biorxiv.org/content/10.1101/2024.07.22.604600v1 .

Response:

Thank you very much for letting us know about this interesting paper. We have cited this paper in the related context, which is pasted below.

> We found that *D* and expression level are positively correlated among the most highly expressed genes (**Figure 5b**, purple bars), supporting the effect of translational cost due to excessively small *D*. The positive correlation is not an artifact of the limited number of genes above the expression threshold, since randomly downsampling all genes to the same number as those retained at each expression threshold will not result in the same correlation (**Figure 5b**, gray bars). Moreover, endogenous genes in yeast, fly, and human show similar patterns (**Figure 5c-e**), and these patterns cannot be attributed to potential base composition biases (**Figure S6**). Interestingly, the maximum correlation achieved is the weakest in human compared to the other three species, which is consistent with the weak translational selection observed for metazoan species[33].

Comment 7 (Remarks on code availability

I haven't yet reviewed the code but I plan to do so for a revised version of the manuscript.

Response:

Thank you and we have updated our GitHub repo with the code for the additional analyses during this revision.

Reviewer #3:

Overall Comment:

The paper titled “A slight mismatch between a gene’s codon usage and the cellular tRNA supply is beneficial” describes a very thorough work with a thought-provoking notion. Unlike the current dogma in the field, which assumes that optimal translation efficiency is achieved when codon usage perfectly matches the tRNA pool of the cell, a slight mismatch might be optimal. A mechanism based on feedbacl on the tRNA pool is suggested. A proposed mechanism is based on an alleged trans-regulation and tRNA

depletion.
While previously there were some indications for this effect in unneeded genes, now they look for genes with fitness effect, either a drug resistance gene or natural genes. This is a major step forward as it allows them to balance costs against functional payoff.

This is an excellent work. It is thought-provoking, very well done and deep in it evolutionary approach. It is also very well written and technically sound. I would ultimately recommend publication in NC. Yet, if I am contribute to the robustness of the conclusion and to the discussion that this paper will surely evoke, here are some important points.
I'm accepting to sign the report with my name: Yitzhak (Tzachi) Pilpel.

Response:

We are profoundly grateful for your generous and encouraging evaluation of our work, especially from a respected scholar like yourself.

Comment 1

1. A critical experiment could be to manipulate the tRNA pool and then test, what I believe is a prediction of their theory. That for different tRNA pools, alternative members of the collection of 53 synonymous variants would optimal. This is a clear prediction of their model since for each tRN pool other variants might be at optimal D. If they showed that I would be convinced better. If the result would show that same variants are good at different tRNA pools then their conclusion is not supported. In other words, they have manipulated so far the codons – the demand, but never the tRNA – the supply.

Response:

Thank you very much for the critical and very constructive comment! While manipulating all tRNA genes is difficult, following your suggestion, we aimed to overexpress one tRNA by inserting it into the high copy number PUC57 vector. We assumed this would increase the supply of this specific tRNA at least twice as much as the supply of the most abundance synonymous tRNAs in the wild-type strain (I.e., set the X_{ij} of the affected codon j to twice the maximum X_{ij} of synonymous codons for amino acid i, then rescale all X_{ij} of amino acid i). We selected the argW^{CCT} tRNA gene and two specific GmR variants (D = 0.328, in the minimum D group, and D = 0.778, in the large D group), as these variants will be moved into the small D group as a result of the tRNA overexpression. Via homologous recombination, we inserted the argW^{CCT} tRNA gene, driven by its native promoter, into the respective GmR plasmids (**Figure S4a**, as pasted below), resulting in two tRNA-overexpressed strains.

As is predicted by the conventional model, a decrease in D caused by altered tRNA

supply is always beneficial (or at least not detrimental in low or intermediate gentamicin concentrations). On the contrary, our model predicts that at intermediate gentamicin concentration, both strains should be fitter after the tRNA supply change, whereas fitness changes would be opposite at low or high gentamicin concentration (**Figure S4b**). The experimental results support our model (**Figure S4c/d**), although the fitness changes observed for the strain with increased D are not statistically significant, possibly due to the relatively small change in *D*.

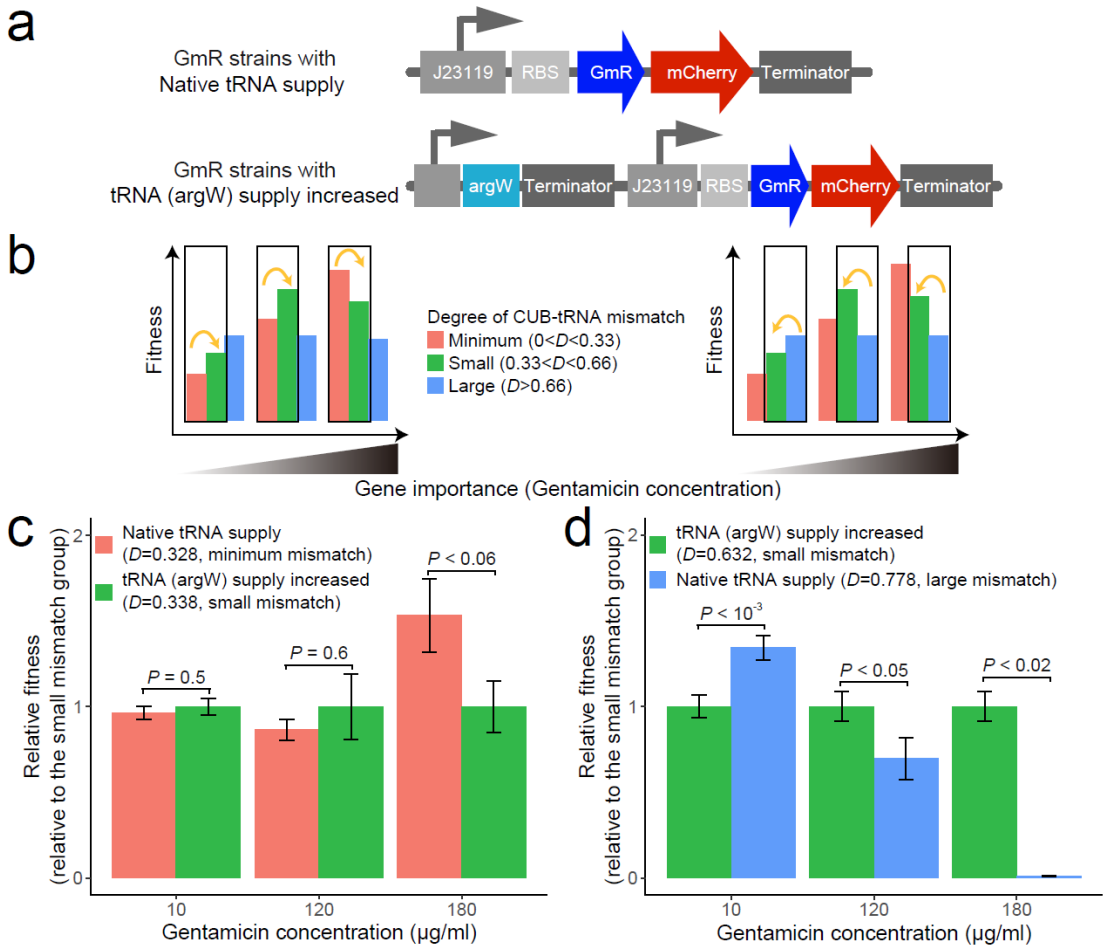


Figure S4. Fitness change upon manipulating tRNA supply

(**a**) A schematic illustration of the expression vector structure with (bottom) or without (top) the additional tRNA gene argW. (**b**) Our model's prediction for fitness changes following an increase in argW supply. This is essentially same as the bottom panel of Figure 3a. (**c** and **d**) The relative fitness of two GmR variants in cultures with 10, 120, and 180 μg/ml gentamicin before and after argW tRNA overexpression, which resulted in a *D* values change from 0.328 to 0.338 (c) and from 0.778 to 0.632 (d). The statistical significance of two-tailed Wilcoxon Rank Sum Tests are indicated. The rightmost bar in **d** indicates no observable growth for the native strain under 180 μg/ml gentamicin.

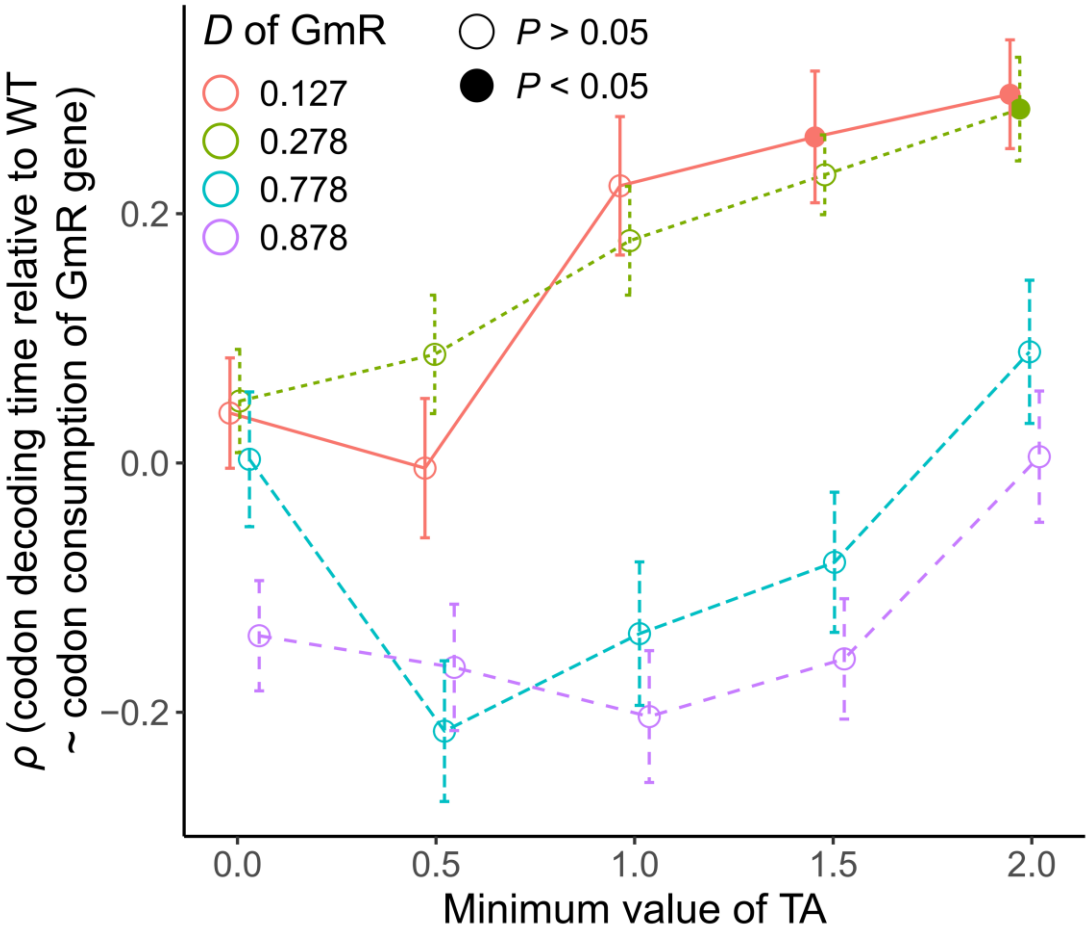
Critically, these results clearly demonstrate a case where changes in tRNA supply will also lead to a change in relative fitness in the direction predict by our model, which further strengthen our model.

Comment 2

2. The proposed mechanism that could mediate their effect is that of tRNA depletion, "In theory, the same logic applies to endogenous genes native to the cell, since they also trans-regulates the expression of other genes via tRNA depletion" – Ref #23. First, I don't think reference #23 shows tRNA depletion. This is a misunderstanding of what was done in Frumkin et al. The increase in tRNA supply in ref #23 was very different, first because it was done artificially by the researchers (not spontaneously by the cell), and second, because the tRNA levels themselves were not changed, but rather tRNA anticodons were mutated. So it's incorrect to ascribe to ref#23 a trans-regulation of the tRNA pool. That said, I think we are left without a true reference or a proof, that the tRNA pool is indeed trans-regulated in response to change in codon demand. So I can see several options: 1. The authors measure tRNA pool with some existing tRNA seq protocols and show the effect; 2. Find a reference that truly shows this effect, or 3. Acknowledge that we don't know if such effect indeed occurs at all

Response:

Thank you for this insightful comment. We agree that neither our study nor Frumkin et al. directly observed tRNA depletion. Here, we believe that measuring tRNA transcript abundance is not useful, since several factors other than the tRNA transcript abundance can affect the "effective" tRNA supply (see our response to the next Comment 3 for a more detailed explanation on this point). Instead, we shifted the focus a little bit downstream, to investigate whether the decoding time of the codons frequently used by GmR is indeed slowed down when GmR has a minimum *D*. This directly tests the translational burden exerted by the GmR on the level of individual codons, which is most parsimoniously explained by a shortage of the translation-ready tRNA. To this end, we performed ribosome profiling (Ribo-seq) on four strains with varying *D* values (0.127, 0.278, 0.778, and 0.878) of the GmR gene. It is observed that when the GmR gene has a minimum *D* (minimum mismatch with the tRNA pool, $D = 0.127$ or 0.278), the codons frequently used by GmR tend to have a longer decoding time, which gives rise to a significant positive correlation between decoding time and GmR codon consumption. On the contrary, strains expressing GmR genes of large *D* did not exhibit the same correlation. This result has now been added as Supplementary Figure S3, and pasted below. Note that RNA-seq is not necessary here since the relative decoding time is always normalized within the same gene first, and regardless of the RNA-seq results, all codons within the same gene will have the same transcription level.



We have removed "tRNA depletion" when citing Frumkin et al., and also explicitly mentioned that we have not proved the existance of tRNA depletion in Discussion, which is pasted below. We hope these revision have better elucidate this point.

> … in terms of the molecular mechanism behind our observation, our working model is tRNA depletion triggered by excessive translation of genes with minimum *D*. Although the results from our experiments with GmR and AmpR, as well as those obtained from tRNA overexpression and Ribo-Seq, can parsimoniously be explained by this model, it has not been proven conclusively. Moreover, we did not rule out other molecular mechnisms that could contribute to GmR's translational cost, such as increased plasmid copy number, transcription level, or translational initiation rate. Future research will be necessary to fully elucidate the molecular mechanisms involved.

Comment 3

3. Their calculation of the D measure – the extent of deviation from the tRNA pool and codon usage is critical, especially since the whole argument is based on the difference between D that approach zero and D that is slightly above zero. My problem is with their estimation of the tRNA pool which is needed for equation in line 358. Particularly

they say “Xij represents the tRNA supply of the cell, approximated by the fraction of codon j among the synonymous codons of amino acid i for the highly expressed genes in the transcriptome of the cell” ref #47. I’m not sure if and how ref #47 provides us with a tRNA pool. So a few comments on this: 1. I’d like to see their results reproduced with other, more conventional definitions of the tRNA pool. Again, they could do tRNA seq. There’s also the classical tAI – based tRNA pool deduced from tRNA gene copy number. Is their D optimized to be slightly above zero also when using these more conventional estimations of the tRNA pool? The way in which ref #47 appears to be estimating the tRNA pool is affected by codon demand, this is potentially problematic.

Response:

Thank you very much for this important comment. We would like to address from two angles.

1. The assumption on CUB of highly expressed genes has been optimized to match the tRNA supply

The logic behind our estimation of tRNA supply is twofold. **First**, the actual tRNA supply is neither the copy number of the tRNA gene nor the abundance of tRNA transcripts. Instead, it is the abundance of the ternary complex of aminoacyl-tRNA+EF-Tu+GTP (or replace EF-Tu with EF1A in eukaryotes). Currently, the abundance of ternary complexes for different types of tRNAs cannot be systematically measured. In addition, the kinetic parameters between different codon-anticodon pairs will also influence the "effective" tRNA supply, further complicating the estimation process. **Second**, because of the first point, we resorted to the assumption that the cellular tRNA supply and demand (codon usage) of the whole transcriptome have been evolutionarily optimized by natural selection to ensure a close match between supply and demand. This assumption allows us to approximate the supply of tRNA by using the codon usage (the demand) of the top 10% expressed gene (which dominate the transcriptome) in the transcriptome. It is important to note that we are using the transcriptome, not the genome, which means the codon usage has been weighted by the expression of those genes before being summed up. Furthermore, this assumption of supply-demand match has been empirically supported (PMID: 22479199 and PMCID: PMC12262485), and has been used in our previous publications (PMID: 32123323 and 34977494). It is also important to note that a supply-demand match on a transcriptome level does not necessarily imply a supply-demand match for individual genes. Therefore, this assumption is neither circular nor illogical.

2. Using measurements of tRNA transcript abundance for tRNA supply does not change our conclusion.

Per your suggestion, we obtained tRNA expression levels based on high-throughput tRNA sequencing data from a previous publication (PMID: 31353208, GEO accession GSE128812), which were then scaled by codon-anticodon affinity and used as the tRNA supply. This metric is similar to the “adaptiveness value” (w) of a codon, as used in the popular tAI metric (PMID:15448185). It was found that the grouping

(minimum, small, or large *D*) in **Figure 3** remained unchanged, suggesting that our conclusion is not affected by the change in the method used to quantify tRNA supply. We have now included this result in **Supplementary Figure S2**, and it is pasted below.

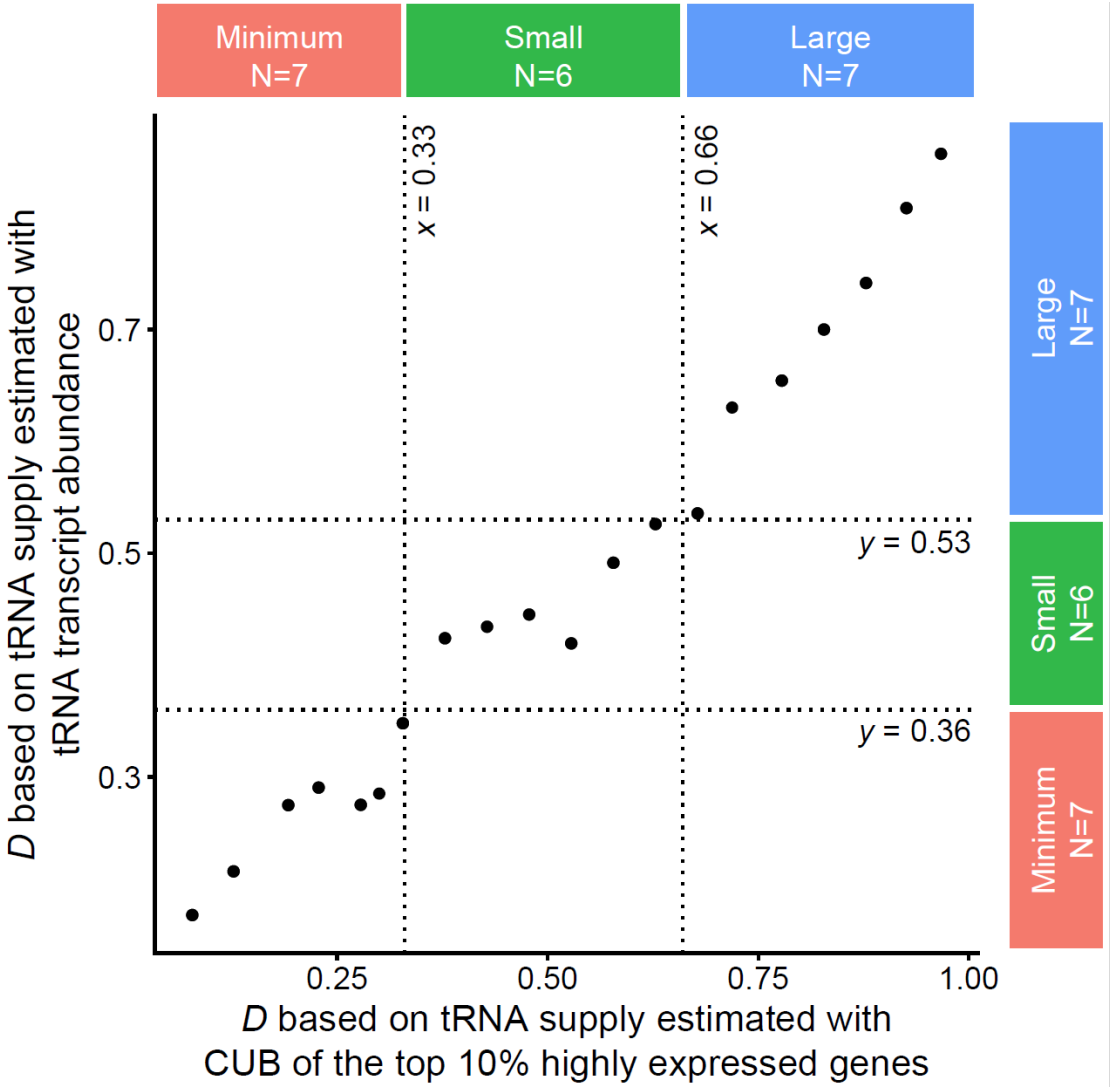


Figure S2. Grouping by D of GmR strains remains unchanged if high-throughput tRNA sequencing data are used to estimate tRNA supply

The D values of designed GmR synonymous variants were calculated with tRNA supply based on either transcriptomic codon usage of the top 10% highly expressed genes (x axis) or tRNA transcript abundance derived from high-throughput tRNA sequencing (y axis). The black lines indicate the grouping, which is identical for both types of D values. Please note that each dot has up to three unique sequences as biological replicates (see **Table S2**).

Comment 4

4. Their estimates for genes functional importance in case of natural genes is based on fitness upon replacement with Kan_R. I'm not sure if this measure is adequate because it represents a total deletion of a gene, while here we ask about reduction in its translation efficiency. I'm not sure how to solve this problem, but I'd be better convinced if they had additional measures of genes' functional importance. Perhaps by employing bioinformatics of evolutionary conservation of genes they could generate an estimation of functional importance but not open total deletion.

Response:

Thank you very much for your comment. We agree that using fitness after gene deletion is not ideal. Ideally, we would like to see fitness data for the same gene with slightly reduced expression. However, this type of data is not currently available at the genomic level. Therefore, we have resorted to the crude indirect approximation currently used in our study. We did not try evolutionary conservation here because it is determined by functional constraints instead of functional importance (PMID:26055156).

Point-by-point response to reviewers' comments

Reviewer #1:

Overall Comment:

The authors are commended on thoroughly addressing my comments.

Response:

We thank you for your thorough evaluation of our work and your very constructive suggestions.

Reviewer #2:

Overall Comment:

Following the first round of reviews, Chen and collaborators have considerably improved their manuscript entitled " A slight mismatch between a gene's codon usage and the cellular tRNA supply is beneficial", by addressing most of reviewers' comments. I will go through the main comments below.

Response:

Please accept our sincere thanks for acknowledging our improvement. Below, we will respond to your further comments one-by-one.

Comment 1

In my view, they have satisfactorily addressed Reviewer 1's main objection related to the use of gentamycin, which has potential confounding effects as it affects translation. They have performed additional experiments using ampicillin and an ampicillin-resistant gene in their genetic editing work, which confirm their original observations.

Response:

Thank you for your support.

Comment 2

The second objection, brought forward by all three reviewers, is that their measure of distance with respect to the tRNA supply (noted D in the manuscript) is an imperfect approximation. All three reviewers highlighted the fact that D is rather a measure of distance with respect to the synonymous codon frequencies of the 10% most highly expressed genes, and not a direct measure of the distance with respect to the actual tRNA supply. The authors argue that the two should match, under the hypothesis that translational selection has optimized the codon usage bias of the 10% most highly expressed genes to match the tRNA supply. That is indeed what is expected, under the "conventional" translational selection model (which is disputed in this article), for species where translational selection is strong - that is, species with high effective population size. However, in species with low Ne, codon usage bias does not match tRNA supply even for highly expressed genes - again, see Bénitière et al. Nevertheless, the authors have followed the reviewers' suggestion and included in the revised manuscript a measure of tRNA supply based on tRNA gene expression levels. They show that their classification of GmR versions into categories (low D, medium D and high D) does not change with this definition of tRNA supply. This confirms the validity of most of their results, but not all - indeed, the analyses presented in Figure 4, 5 and 6 (which go beyond the analysis of E coli strains with different GmR versions) have not been redone with this measure of D. As these analyses are important for the authors' model, they need to be validated with this new measure of D.

Response:

Thank you for your question. We have now redone all relevant results in Figure 4, 5 and 6 using *D* calculated based on the tRNA expression, whose results remained qualitatively unchanged. Please note that we have also slightly updated the results derived from the mutation accumulation lines in order to better reflect mutation biases. More specifically, when generating random mutations to assess the null distribution, not only is the number of mutations matched, but also the type of mutations matched. This renders the original positive control for MA data in Figure S7 (currently Figure S10) insignificant, indicating that the sample size is insufficient for a proper test. As a result, we were forced to include 15% most highly expressed genes. The tRNA expression-based results corresponding to Figure 4/5/6 are pasted below.

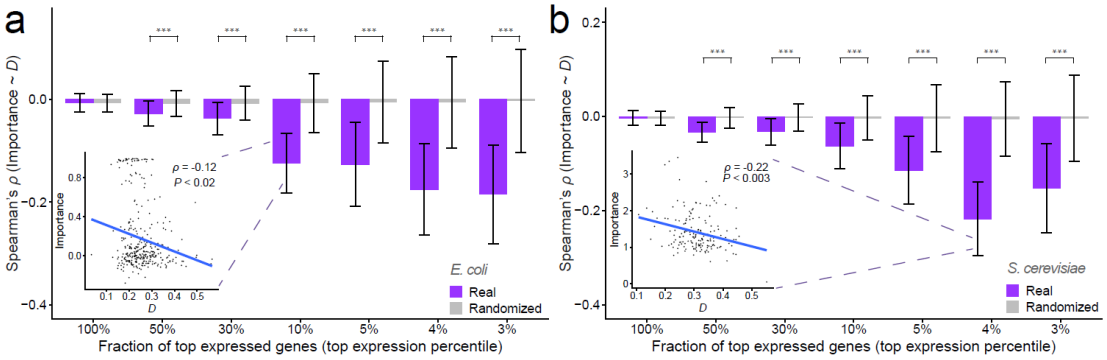


Figure S6. The effect of functional payoff on native codon usage of E. coli endogenous genes replicated by tRNA expression-based *D*

These panels are similar to **Figure 4a-b**, except that the tRNA supply (X_{ij} in D) is estimated using tRNA expression level instead of the codon usage of the 10% most highly expressed genes.

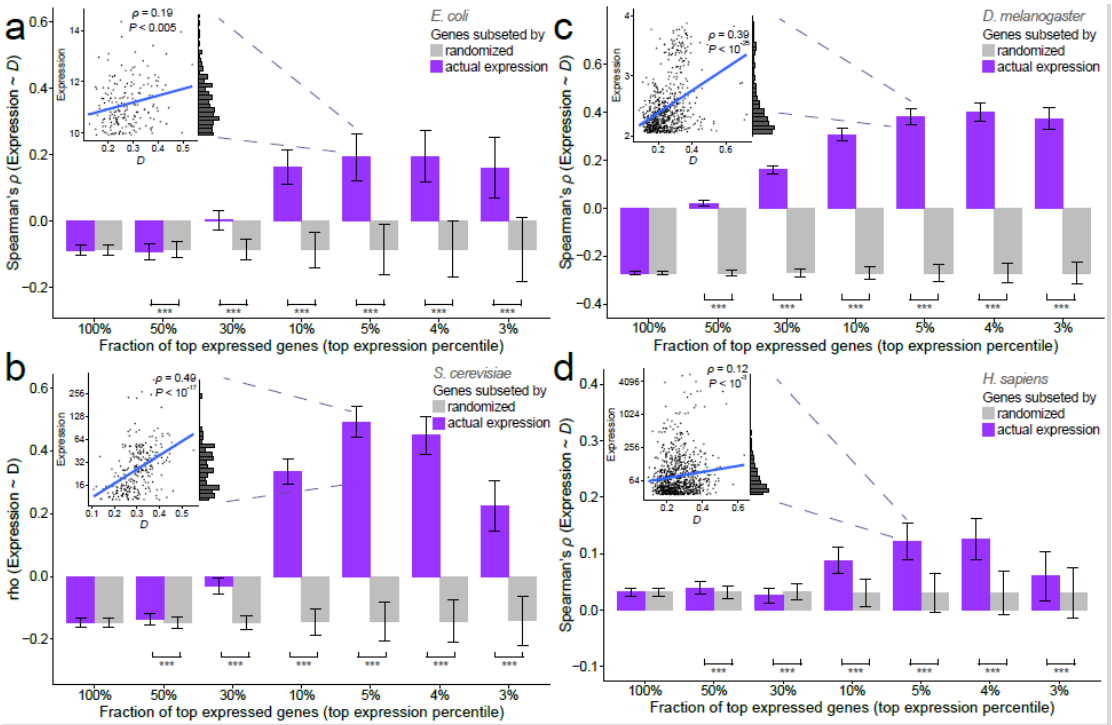


Figure S7. The correlation between *D* and expression for endogenous genes is recapitulated by tRNA expression-based *D*

These panels are similar to **Figure 5b-e**, except that the tRNA supply (X_{ij} in D) is estimated using tRNA expression level instead of the codon usage of the 10% most highly expressed genes.

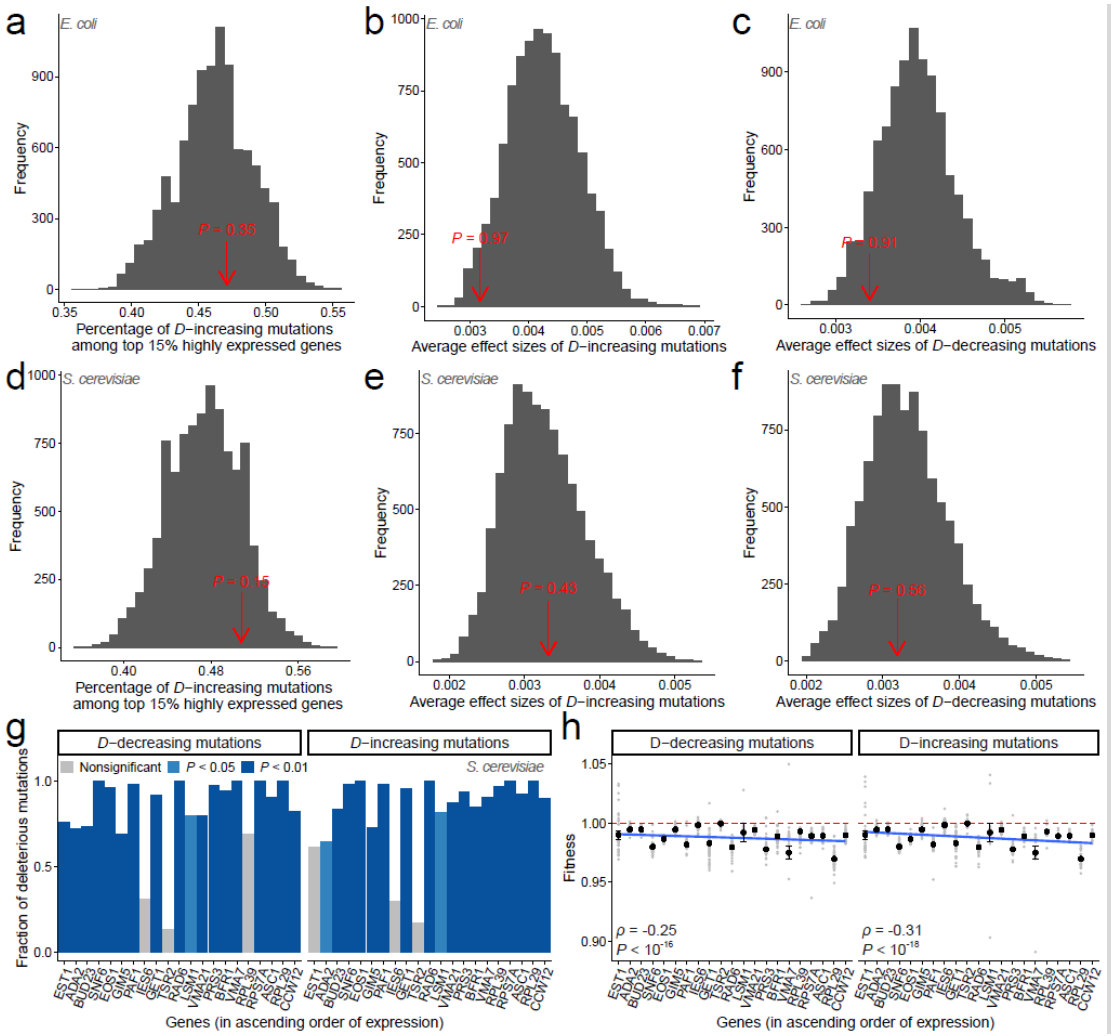


Figure S11. Selective maintenance of the mismatch between endogenous gene's codon usage and cellular tRNA supply is recapitulated by tRNA expression-based *D*

These panels are similar to **Figure 6c-h**, except that the tRNA supply (X_{ij} in D) is estimated using tRNA expression level instead of the codon usage of the 10% most highly expressed genes.

Comment 3

On the same point, I still have trouble with the circularity of the argument: the authors rely on a measure of D that only makes sense under the "conventional" translational selection theory to refute the "conventional" translational selection theory. Under their model, codon usage bias should match the tRNA supply only for highly expressed *and essential/important* genes, for which the functional payoff is more important than the translational cost. The authors should analyze whether their analyses are still valid when defining D based on the codon usage of the most highly expressed and essential/important genes. I am not suggesting that this new D measure should replace the one originally proposed by the authors throughout the manuscript, as that would be circular as well - but rather I am asking for additional analyses, to be presented in the supplementary material.

Response:

Thank you for your insightful question. Please allow us to clarify that our argument is not circular, and that there is a subtle difference between the conventional translational selection model and our assumption underlying the definition of *D*. Specifically, we assume that the 10% most highly expressed genes COLLECTIVELY match the tRNA supply, but individually they may not necessarily match it. By contrast, the conventional translational selection model assumes that selection optimizes the codon usage bias of INDIVIDUAL genes so as to match the tRNA supply. To illustrate this, we plotted the synonymous codon frequency of each amino acid (i.e., fraction of sites of an amino acid being encoded by one of its synonymous codon) for each of the 10% most highly expressed genes. In addition, we plotted the expression-weighted average of the synonymous codon frequency as a red triangle. As shown in **Figure RR1** below, the codon frequencies of individual genes all deviate from the collective codon frequencies in one way or another, demonstrating that the collective codon usage of top 10% highly expressed genes may differ from those of the individual genes. In other words, the assumption underlying *D* (about the collective codon usage) is different from the prediction of the conventional translational selection model (about codon usage of individual genes).

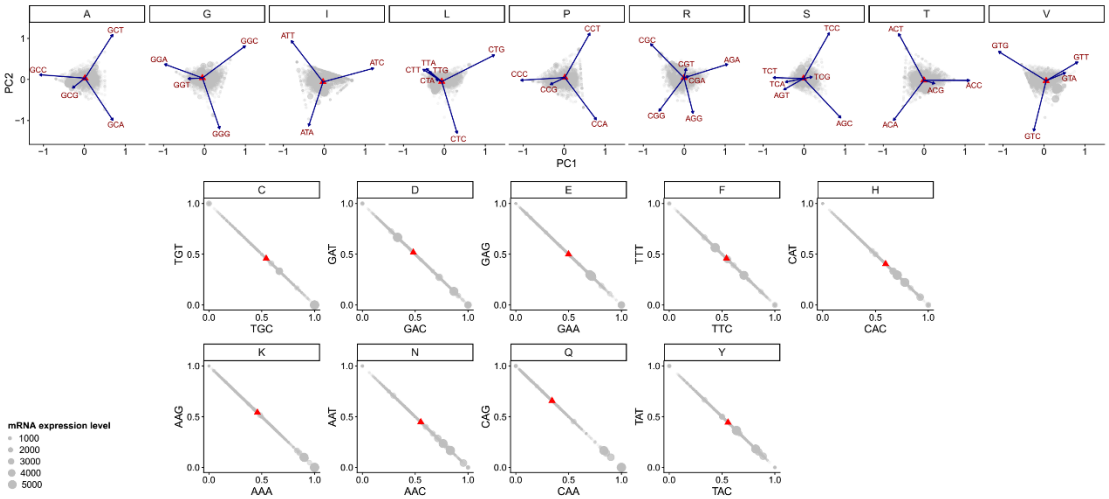


Figure RR1. Synonymous codon frequencies of 10% most highly expressed genes

The frequency of synonymous codons in individual genes (gray points) and their expression-weighted averages (red triangles) are plotted. The size of the gray point reflects the expression level of the focal gene, as indicated by the legend at the bottom left. For amino acids with two synonymous codons (bottom two rows of panels), the two codon frequencies are directly shown on both axes. For amino acids with three or more synonymous codons (top row of panels), the first two principal components of the codon frequencies are shown, while the loadings of frequencies of the individual synonymous codons are indicated by blue arrows.

In addition, we would like to emphasize that we are not advocating that the most important genes should perfectly match the supply of tRNA. Rather, we believe that even highly expressed

important genes should exhibit a slight mismatch with the tRNA supply. This has now been clarified in the Discussion.

> … In this work, we further demonstrated that this concept can be extended to endogenous genes native to the genome. In other words, to ensure optimal cellular growth, any gene expressed in a cell, including those that are functionally very important, should have a CUB that slightly deviates from the cellular tRNA supply, though the specific direction of deviation may differ for each gene. …

Comment 4

The authors have addressed my comment regarding the translational effects on endogenous genes by performing Ribo-seq and analyzing the decoding time of codons that are frequent in the GmR. They show that in E coli strains where the GmR has low D, there is a positive correlation between decoding time and the frequency of codons in the GmR sequence, in particularly for genes that are highly translated. I appreciate the effort that the authors have put into performing the Ribo-seq experiments and I am satisfied with the result.

Response:

Thanks again for your constructive suggestions that pushed us to improve our analysis.

Comment 5

Regarding the mutation accumulation experiments, the authors have followed my suggestion to only include synonymous mutations. However, I am puzzled as to why they argue that it is not useful to look at weakly expressed genes. In fact, weakly expressed genes should provide a "control" of sorts, precisely because their CUB is not expected to have been optimized by purifying selection. I would like to see a comparison between highly expressed and weakly expressed genes for the analyses presented in Figure 6.

Response:

Thank you for this clarification. We agree with you that weakly expressed genes can serve as a type of control. We therefore randomly sampled a group of weakly expressed genes (defined as 20% least expressed genes) that matched the highly expressed genes in terms of number of genes, number of mutations of each type (to control mutation bias) and GC content. Using the same analyses as for highly expressed genes, we found no evidence for bias for *D*-increasing synonymous mutations. Note that this result should be explained by a lack of strong translational selection in weakly expressed genes rather than by the selective maintenance of the slight mismatch between CUB and tRNA supply. The related results are now included as Supplemental Figure S9c-d, which is pasted below.

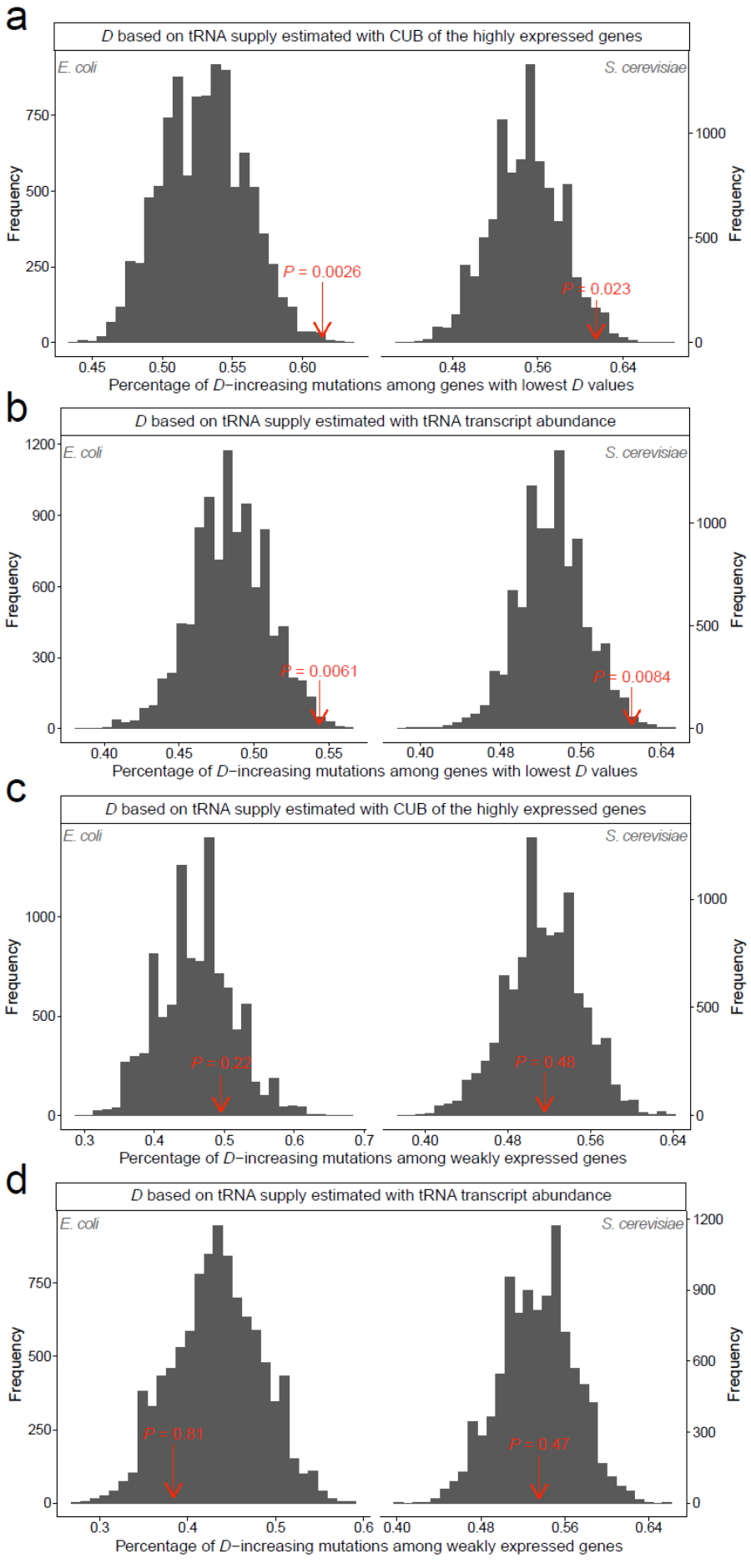

a
D based on tRNA supply estimated with CUB of the highly expressed genes
E. coli
S. cerevisiae
Frequency
P = 0.0026
P = 0.023
Percentage of D−increasing mutations among genes with lowest D values
b
D based on tRNA supply estimated with tRNA transcript abundance
P = 0.0061
P = 0.0084
Percentage of D−increasing mutations among genes with lowest D values
c
D based on tRNA supply estimated with CUB of the highly expressed genes
P = 0.22
P = 0.48
Percentage of D−increasing mutations among weakly expressed genes
d
D based on tRNA supply estimated with tRNA transcript abundance
P = 0.81
P = 0.47
Percentage of D−increasing mutations among weakly expressed genes

Figure S9. The fraction of *D*-increasing mutations is significantly higher than random expectation among genes with minimal *D*, but not among weakly expressed genes.

(a) This is similar to **Figure 6c** and **f**, except that mutations among genes with 10% lowest *D* values were analyzed. This is a positive control for Figure 6c and f, demonstrating that the observation therein is not due to insufficient statistical power. See **Methods** for details. **(b)** This is similar to panel a, except that tRNA supply (X_{ij} in *D*) is estimated using tRNA expression level instead of the codon usage of the 10% most highly expressed genes. **(c-d)** These are similar to panel a-b, except that a group of lowly expressed genes (defined as 20% least expressed genes) that matched the highly expressed genes in **Figure 6c** and **f** were analyzed as a control. The number of genes, number of mutations of each type in a gene and GC content of a gene (±15%) were matched.

Comment 6

Regarding the analysis presented in Figure 5 (correlation between D and expression level, among highly expressed genes), I am still not entirely convinced by the authors' interpretation, mainly because of the results observed for human. The corerlation between D and expression level is 0.15 for human and 0.13 for yeast for the top 10% most highly expressed genes, although in yeast translational selection (be it "conventional" or according to the authors' new model) should be much more efficient than in humans. Because D is defined as the distance with respect to the CUB of the 10% most highly expressed genes, and that the CUB is computed based on the transcriptome not on the genome, I wonder if part of this correlation may be artefactual (though I confess that I have trouble imagining what the confounding factor might be!). Again, here it is important to redo the analysis with the D measure that is based on tRNA abundance. Some protein-coding gene families (e.g. ribosomal proteins) are greatly over-represented among the most highly expressed genes. Could that induce a biased *amino-acid* usage, that could explain some of this correlation? Perhaps if one could compute a D value by amino-acid, and check if all amino-acid D values show the same pattern?

Response:

Thanks for your comment. We are in agreement with you that species with stronger translational selection should exhibit a stronger correlation between *D* and expression level. But the issue here is that gene expression distributions may vary between species, making it difficult to directly compare quantile thresholds (*x* axis in **Figure 5**). Accordingly, we believe that the maximum correlation achieved should be a better reflection of the strength of translational selection, which is >0.2 in *E. coli*, *S. cerevisiae* and *D. melanogaster*, but never reaches 0.2 in human.

For the re-analyses using *D* derived from tRNA expression, please refer to the above Comment 2. We also followed your suggestion to separately analyze *D* for each amino acids, i.e. D_i. The results have been added as Supplemental Figure S9 (too big to paste inline here). Essentially, all of these results show the same pattern as our main results based on *D*.

Comment 7

I am otherwise satisfied by authors' reply to my suggestions (including my questions related to YFP transcription level, plasmid copy number variations and GC content effects). Overall I believe that this is a valuable manuscript and that it should ultimately be published in Nature Communications, following an additional round of review that would address the remaining points listed above.

Response:

Thanks again for your very thorough evaluation of our work !

Point-by-point response to reviewers' comments

Reviewer #2:

Overall Comment:

I thank again the authors for their thorough reply. I remain puzzled by some of the results (in particular the apparent high translational selection levels in humans, according to the authors' observations), but overall I do not have any further objections to publication. This manuscript deserves to be published and further examined by the scientific community.

Response:

We thank you for your very constructive suggestions and your support of our work.