nature portfolio

Peer Review Information

Journal: Nature Ecology & Evolution
Manuscript Title: Low protein expression enhances phenotypic evolvability by intensifying selection on folding stability
Corresponding author name(s): Andreas Wagner

Editorial Notes:

Reviewer Comments & Decisions:

Decision Letter, initial version:

28th January 2022

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Shraddha,

Your Article, "Low expression enhances protein evolvability by mitigating folding defects" has now been seen by three reviewers. You will see from their comments copied below that while they find your work of considerable potential interest, they have raised quite substantial concerns that must be addressed. In light of these comments, we cannot accept the manuscript for publication, but would be very interested in considering a revised version that addresses these serious concerns.

We hope you will find the reviewers' comments useful as you decide how to proceed. If you wish to submit a substantially revised manuscript, please bear in mind that we will be reluctant to approach the reviewers again in the absence of major revisions.

If you choose to revise your manuscript taking into account all reviewer and editor comments, please highlight all changes in the manuscript text file in Microsoft Word format.

We are committed to providing a fair and constructive peer-review process. Please do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

If revising your manuscript:

* Include a "Response to reviewers" document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the referees along with the revised manuscript.

* If you have not done so already we suggest that you begin to revise your manuscript so that it conforms to our Article format instructions at http://www.nature.com/natecolevol/info/final-submission. Refer also to any guidelines provided in this letter.

* Include a revised version of any required reporting checklist. It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review. A revised checklist is essential for re-review of the paper.

Please use the link below to submit a revised paper:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

If you wish to submit a suitably revised manuscript we would hope to receive it within 6 months. If you cannot send it within this time, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Ecology & Evolution or published elsewhere.

Nature Ecology & Evolution is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. This applies to primary research papers only. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

Please do not hesitate to contact me if you have any questions or would like to discuss the required revisions further.

Thank you for the opportunity to review your work.

[REDACTED]

Reviewer expertise:

Reviewer #1: protein evolvability, theory and experiments

Reviewer #2: protein evolvability, microbial experimental evolution

Reviewer #3: computational evolutionary biology, biophysical mechanisms of protein evolution

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

In this manuscript, Wagner and colleagues approached the very important and interesting question about how gene expression level affects evolvability of protein phenotype. They conducted direct evolution experiments, in which GFP driven by two promoters whose expression levels differed by ~3-fold, and then investigated the evolvability of GFP to a new phenotype (Cyan fluorescence). They have some unexpected findings, most notably more nonsynonymous and neo-functionalizing variants in the low expression group. After modelling their experiment, they come to the conclusion of abundance-stability compensation, which helps lowly expressed proteins evolve novel phenotypes. They supplement their model with some additional in vitro experiments. Overall, the authors targeted a valid question, found some interesting observations, and explained it with a well-described theoretical model. Unfortunately, there are several serious flaws in their experimental design, which leaves the main conclusion of the study in doubt. I will detail my concerns below.

My first major concern was the biological relevance of the selective scheme in both experiments and theoretical modelling. There is an implicit unrealistic assumption crucially responsible for the experimental observation and theoretical expectation. Namely, natural selection does not select the top number or fraction of genotypes. Instead, whether a specific genotype can get through selection qualitatively depends on its relative fitness. As a simple example, imagine a population with 100 individuals, where the relative fitness (w) of only 5 individuals equals 1, and all others have w=0.01. In an artificially defined selective scheme such as the one used in this manuscript, say 10% of the genotypes make it to the next generation, five lucky genotypes with w=0.01 can survive. In reality, however, one can easily see that these w=0.01 genotypes have really small chance of surviving, only one hundredth of those w=1. Let us name this distribution the "TOY distribution", which I will refer to again below. The main "unrealistic" part here is that the chance of surviving is no longer consistent with the value of fitness (w) if one uses a criteria of top 10% fitness.

This unrealistically assumed selective scheme might be fine for some evolutionary investigations, but seems particularly problematic for this study, or at least cannot be generalized to natural selection (i.e., could be applicable to artificial selection such as the experiment here. Also, I am not questioning this selective scheme applied in other context, such as an earlier Science paper by the same senior author). In particular, as the authors modelled the selection in their simulation (similarly in their experiment), they chose the top 1% cells with highest fluorescence (and repeat this process 100 times to keep the population size constant. See line 200 in Supplementary material). In the TOY distribution scenario, if we use the top 10% fluorescence as the criteria of selection, the resulting selected

population after the first generation will have 50% cells with w=0.01. In the second generation, the author applied their log-normal distribution again on the abundance, a treatment they consider reflecting expression noise. In my demonstration here, this is the same as applying the TOY distribution again to the population, therefore the next generation will still have 50% cells with w=0.01. In other words, this type of modelled selection has really weak capacity of purging individuals with w=0.01.

To rephrase the problem, the 10% criteria combined with the log-normal distribution set the criteria of surviving at a really low level such that destabilizing mutations can sneak into the next generation. This phenomenon is stronger for H population compared to L population because high abundance proteins have a more skewed distribution of expression. This phenomenon could be even stronger when we used an even more exaggerating distribution such as the TOY distribution (the authors should agree with me on this as they've discussed similarly on a heavy-tailed distribution as one requirement of their theory on line 349). This should not be call an "abundance-stability compensation", it should be call "increased tolerance due to large variance".

Note that the authors have also failed to explain this ultimate cause ("increased tolerance due to large variance") of their observation in Introduction or Abstract. They instead just keep using "abundance-stability compensation". The "abundance-stability compensation" could be understand in two ways, one is the "increased tolerance due to large variance" as modelled by the authors, the other is "increase amount of proteins could lead to tolerance of destabilizing mutations as long as the total functional activity meets some requirement". The difference between this two understandings can be illustrated with a normal distributed expression. In the first understanding, as shown by the authors, normal distribution cannot create the "abundance-stability compensation". In the second understanding, however, normal distribution can still create the "abundance-stability compensation".

The correct way a simulation for selection should instead be carried out is that the probability of surviving is proportional to the fitness (fluorescence). In the case of this study, the authors should sample the ancestral population 10,000 times with replacement, and the probability of a cell being sampled equals its fluorescence intensity. As a result, this selective scheme will allow x% of cells in the H population to survive, but y% of cells in the L population to survive, where $y > x$ (in contrast, $x = y$ in the current selective scheme by the authors). I can only be persuaded on this major concern if the authors could repeat their conclusion using this type of more realistic selection.

My second major concern is highly related to the first one, but it is independent. They authors have mixed expression noise and among-individual variation of mean expression in their simulation. expression noise is largely log-normal, but among-individual variation of mean expression is not. The selective agent (trait being selected) should not be the instantaneous expression of a cell, which is heavily influenced by the expression noise. Instead, the selective agent should be mean expression (averaged across different timepoints), which is much less affected by expression noise.

The third major concern I have is about the experimental design. Green and cyan are two very similar colors. The excitation spectrum of GFP heavily overlapped with the cyan channel (the authors show it in Fig.3A). In this case, the H group GFP could be so strong that its signal on the cyan channel is already strong enough to mask any neo-functionalizing mutations. On the contrary, the L group GFP is

relatively weak such that the cyan channel can be used as an efficient selective agent as designed. As a result, the selection for cyan is expected to be weaker for the H group compared to that for the L group.

In a similar vein, this problem of experimental design extends to the problem of correlation between traits. I believe evolvability for a highly correlated trait can be a totally different story, compared to the evolvability for an orthogonal trait. The authors should also discuss this point.

There are also minor concerns below.
The results in Table S3 is rather confusing. How can a single mutation have such different effects (especially D113N) between H and L? I think increasing a certain fluorescence or not is a physical property, and has nothing to do with expression level.

The sequence difference among variants is quite small, which could lead to heteroduplexes during PCR before SMRT sequencing, and increase sequencing error therein. How is this problem corrected ?

Line 62, "compensated" in terms of what ? fitness? amount of functional molecules ? fluorescence intensity ?

Line 123 "T230S" should be "T203S"?

Line 150 "10% of GFP population survives, light gray", light gray should be black. Also, Fig3D and E is confusing, what is the difference between the gray-color contrast and the dashed vertical line ? It looks like there are two types of selection .

Line 223 "Figure 3D" should be a wrong mention

Line 262 "Figure 3B" should be a wrong mention.

Line 271 "variant" to "variants".

Line 288 "Figure 3D", is that a wrong mention too ?

Reference 11 proposed a model for mistranslation-induced misfolding. If the authors are mostly talking about misfolding only (regardless mistranslation), I believe PMID:20959819 is a more accurate reference.

Supplementary line 74 "the H population has a relative fluorescence intensity" should be "the H population has a low relative fluorescence intensity" ?

FigureS6 C, there no text in bracket.

Reviewer #2 (Remarks to the Author):

Karve et al. conducted a directed evolution experiment on GFP, in which they selected for cyan fluorescence. Their experiment had one key treatment variable: the strength of the promoter expressing GFP, causing one set of replicate populations to have high GFP abundance, and the other set to have low GFP abundance. Strikingly, the novel cyan phenotype evolved more readily in the low abundance treatment. Karve et al. use computational modeling to illustrate that with a fitness function in which the protein phenotype under selection is proportional to the product of stability and abundance, populations with higher GFP abundance can meet the selective threshold in their experiment with lower stability proteins. By contrast, the low abundance populations do not have this "luxury". The authors then demonstrate that the proportionally higher selection pressure on protein stability in the low abundance populations is responsible for their key finding, by testing two predictions of the model: they demonstrate that the fluorescent proteins in the low abundance treatment are more stable/foldable than those in the high abundance treatment (refolding kinetics), and they conducted a second evolution experiment in which they show that the difference in evolvability disappears when they use stabilizing selection to increase the (hypothesized) number of stabilizing mutations in both populations, before applying strong directional selection.

Overall, the basic finding (low abundance proteins in this system are more evolvable) is interesting, and the reason why (the higher abundance proteins evolved lower stability) is both surprising as well as well supported by the model and follow-up experiments. I very much enjoyed reading this work, and I do not have any critical comments to make about the logic or technique used to establish these results. I think that this work may be widely read by researchers working in molecular evolution.

I do have several minor comments to make about the presentation. I also think that the implications of this work should be more framed more precisely. In particular, many readers will read this work with prior expectations about reading about whether high abundance <=> high stability, or about the causes of the abundance/evolutionary rate correlation. So, it's possible that readers who skim this work by reading just the abstract or conclusions may fundamentally misunderstand the main findings and their importance.

First, minor comments about presentation:
1) often, box plots are used to show differences between the treatment. Based on my reading of the legend, it seems that each box plot represents 4 data points (the mean for each population in the treatment). I would much rather see the underlying four data points for each (maybe jittered to show the distribution). If the box plots are being used to summarize many more datapoints (say distribution of fluorescence for all 4 populations together), then I would appreciate a brief comment in the legend stating exactly what the underlying data points are (10,000 cell counts per pop, or something like that?)

2) Many of the figures have inconsistent fonts, both within and across panels in single figures, and across figures. It would be nice to make this consistent, unless this is something that the publisher is going to handle.

3) Figure 3 C and G have these straight lines, sometimes dashed, in the panels. I can't tell whether these are meaningful or if they are visual artifacts. If they are meaningful, please describe what they mean, if not, please remove.

4) I find the argument on lines 201-208 really hard to understand, starting with "The reason why ...". The arrow on Figure 3G is not much help either. I'm reading Fig. 3 C and G as hyperbolic isoclines based on the product of stability*abundance. I am really confused about how Fig 3G is showing something about a log-normal distribution. Based on Figure 3D and 3E, it just looks like the L populations have stronger selection on relative stability compared to H populations. The log-normal shape seems to be critical based on lines 206-208, but I don't understand this point, a good figure would make a big difference, but Fig 3G isn't cutting it. Help!

Now, minor points about framing:

1) Perhaps this is a major point: the title of the paper should be rewritten. It's misleading. The first part "low expression enhances protein evolvability" is great, but low expression isn't mitigating folding defects! A chaperone mitigates folding defects-- here, proteins with folding defects in the low expression treatment don't survive selection. Something like "High abundance reduces protein evolvability by reducing the strength of selection on protein stability" seems more accurate, or "Low expression enhances protein evolvability by intensifying selection on foldability", or really, whatever is most accurate.

2) In the abstract, the following is really confusing given my prior assumptions about unrelated questions in the field: "proteins with folding defects preferentially accumulate when a protein is highly expressed, because high protein abundance can compensate for such folding defects." This read to me like the per unit rate of misfolding increases with gene expression, because interactions between proteins somehow is helping the defective proteins to fold properly. I get it after reading the paper carefully, but if I had just read the abstract, I really wouldn't understand what this paper is about. What really helped me was the equation for fitness in the paper-- then I immediately understood what was going on (at least up to the log-normal stuff). Something like "Under strong directional selection for cyan fluorescence, high fitness can be achieved by either a smaller number of GFP with few folding defects, or a larger number of GFP with proportionally more folding defects. We call this principle XXXX".

I also find the "abundance-stability compensation principle" misleading for the following reason. Abundance and stability can compensate for each other to get pass the selective threshold in these experiments. But abundance-stability compensation is not a principle of protein evolvability-- it's not that evolvable proteins have to be either abundant or have to be stable. Rather, the low abundance proteins are forced to be more stable and thus evolvable. The fact that they __cannot__ compensate is what makes them evolvable!

Then the sentence on line 20: "Its consequence is that lowly expressed proteins are __under stronger selection__ for proper folding, which facilitates their evolvability".

These same criticisms hold for the language in the Introduction on lines 60-63, and the Discussion on lines 330-344 and lines 368-375. To reiterate, the lack of compensation for stability is what is causing increased evolvability in the L populations, so stating that "abundance-stability compensation facilitates the evolution of new phenotypes" gets it backwards (especially since abundance does not

compensate for stability in terms of evolvability; it's not symmetric in this regard), and will be really confusing for readers who do not read this work carefully.

And the last line of the abstract: "to help" -> "by which", since the lowly expressed proteins do not have agency in their evolution.

I really enjoyed reading this paper, and I hope these minor criticisms help this work reach a wider audience.

Rohan Maddamsetti

Reviewer #3 (Remarks to the Author):

This is a nice contribution, and I generally agree with the way the analysis is set up and with the conclusions that were drawn. I just feel that at times (and in particular in the abstract), readers could be confused by statements such as "proteins with folding defects preferentially accumulate when a protein is highly expressed", as they seem to contradict the widely accepted result that more highly expressed proteins are less likely to exhibit folding defects.

I think the authors understand exactly what is happening in their system and the Discussion is reasonably clear about how everything fits together, but there's still some room for improvement in the writing I think. To me, what resolves the paradox is the difference in time scales. Over long evolutionary times, genes that are highly expressed will slowly evolve into variants that are less likely to misfold. The selection pressure that causes this effect is quite weak, though, and likely doesn't have any effect in short-term experiments such as the ones presented here.

I would like to encourage the authors to read the entire paper carefully and see whether there are places where they can emphasize the time scale dependence. For example, throughout the paper, the authors write about what happens under strong directional selection, and many of these phrases could possibly improved by adding "over short time scales" or similar. Likewise, when the authors state that selection against toxic misfolding is not present in their system, they could add 1-2 explanatory sentences explaining why we think this selection pressure isn't there.

There's one other aspect to this work that I'd like the authors to consider and possibly discuss. If I understand correctly, the selection via FACS really measures two things at the same time: (i) Is there enough fluorescence to be detected; (ii) Is the fluorescence in the right wavelength. For low expression genes, they cannot afford losing any fluorescence signal at all, since it's low already, and hence they are selected for high quality proteins that don't misfold and fluoresce well. By contrast, for high expression genes, they can afford to lose signal intensity and hence they do so, because there is no actual selection pressure to maintain a strong signal. If this is correct, then the entire situation reminds me a bit of the idea of drift robustness discussed by LaBar and Adami (see e.g. https://pubmed.ncbi.nlm.nih.gov/29044114/). It's not exactly the same effect, but the low expression genes are essentially drift-robust because any drift away from their optimum likely causes them to be

removed from the population. By contrast the high expression genes can afford to drift because they have a fitness buffer.

Signed,
Claus Wilke

Author Rebuttal to Initial comments

Subject: Revision of the manuscript NATECOLEVOL-211215405

Response to Reviewer I

Reviewer 1: In this manuscript, Wagner and colleagues approached the very important and interesting question about how gene expression level affects evolvability of protein phenotype. They conducted direct evolution experiments, in which GFP driven by two promoters whose expression levels differed by ~3-fold, and then investigated the evolvability of GFP to a new phenotype (Cyan fluorescence). They have some unexpected findings, most notably more nonsynonymous and neo-functionalizing variants in the low expression group. After modelling their experiment, they come to the conclusion of abundance-stability compensation, which helps lowly expressed proteins evolve novel phenotypes. They supplement their model with some additional in vitro experiments. Overall, the authors targeted a valid question, found some interesting observations, and explained it with a well-described theoretical model. Unfortunately, there are several serious flaws in their experimental design, which leaves the main conclusion of the study in doubt. I will detail my concerns below.

Response: Before our point-by-point reply, we would like to thank you very much for your positive comments and constructive suggestions. They helped us to improve the quality of the manuscript substantially and to make the narrative clearer.

Reviewer 1: My first major concern was the biological relevance of the selective scheme in both experiments and theoretical modelling. There is an implicit unrealistic assumption crucially responsible for the experimental observation and theoretical expectation. Namely, natural selection does not select the top number or fraction of genotypes. Instead, whether a specific genotype can get through selection

qualitatively depends on its relative fitness. As a simple example, imagine a population with 100 individuals, where the relative fitness (w) of only 5 individuals equals 1, and all others have w=0.01. In an artificially defined selective scheme such as the one used in this manuscript, say 10% of the genotypes make it to the next generation, five lucky genotypes with w=0.01 can survive. In reality, however, one can easily see that these w=0.01 genotypes have really small chance of surviving, only one hundredth of those w=1. Let us name this distribution the "TOY distribution", which I will refer to again below. The main "unrealistic" part here is that the chance of surviving is no longer consistent with the value of fitness (w) if one uses a criteria of top 10% fitness.

This unrealistically assumed selective scheme might be fine for some evolutionary investigations, but seems particularly problematic for this study, or at least cannot be generalized to natural selection (i.e., could be applicable to artificial selection such as the experiment here. Also, I am not questioning this selective scheme applied in other context, such as an earlier Science paper by the same senior author). In particular, as the authors modelled the selection in their simulation (similarly in their experiment), they chose the top 1% cells with highest fluorescence (and repeat this process 100 times to keep the population size constant. See line 200 in Supplementary material). In the TOY distribution scenario, if we use the top 10% fluorescence as the criteria of selection, the resulting selected population after the first generation will have 50% cells with w=0.01. In the second generation, the author applied their log-normal distribution again on the abundance, a treatment they consider reflecting expression noise. In my demonstration here, this is the same as applying the TOY distribution again to the population, therefore the next generation will still have 50% cells with w=0.01. In other words, this type of modelled selection has really weak capacity of purging individuals with w=0.01.

To rephrase the problem, the 10% criteria combined with the log-normal distribution set the criteria of surviving at a really low level such that destabilizing mutations can sneak into the next generation. This phenomenon is stronger for H population compared to L population because high abundance proteins have a more skewed distribution of expression. This phenomenon could be even stronger when we used an even more exaggerating distribution such as the TOY distribution (the authors should agree with me on this as they've discussed similarly on a heavy-tailed distribution as one requirement of their theory on line 349).

Response: Thank you for raising these important issues. You have raised three concerns that we summarize again briefly before responding to them individually. First, our selection does not mimic natural selection which is expected to act on relative fitness in a probabilistic manner. Second, our selection criterion appears to be weak, so that it can allow the fixation of deleterious mutations. Third, we use a log-normal distribution which exaggerates the difference between the L and H populations. We discuss these concerns below.

First, we agree that our selection scheme may not seem to mimic natural selection because of the survival of a fixed fraction of cells in our population. However, we posit that it is quite similar to probability dependent selection. In our experiments we generate in every mutagenesis cycle ~10^5 mutants which are distributed over ~10^9 cells that we subject to sorting. In this case, every genotype is represented by ~10^4 cells and the expression level of GFPs in these cells is not the same but follows a distribution due to the presence of expression noise. In other words, our population comprises 10^5 isogenic subpopulations, each with ~10^4 cells. We select 50,000 cells from this population based on their fluorescence. We emphasize that we do not select a fixed number or percentage of variants but a fixed number of cells that fluoresce in the top 1% (or top 5%) of all cells.

The higher the mean fluorescence of an isogenic population with a probabilistic distribution of GFP numbers, the greater the number of cells that may fluoresce above the top 1% threshold at the time of selection. In consequence, the probability that a cell survives selection does scale with the mean fluorescence of the isogenic population that it is a part of. To rephrase, selection acting in the presence of expression noise mimics probability-dependent selection. In addition to this qualitative argument, we also performed new simulations which show that the mechanism that we propose also works under strong selection that is probabilistic: Such selection results in lowly abundant proteins being more evolvable (Figure S11).

This brings us to the second issue, which regards the strength of selection. We argue that our selection pressure is actually not weak but very strong, for the following reasons. First, we select the top 1% of cells. More precisely, in light of the argument in the preceding paras, we choose 5×10^4 cells from all ~10^9 cells (which harbour ~10^5 variants that are generated during the mutagenesis). This selection regime will allow the ~5 best genotypes to survive. We then perform another selection for the top 5% of the surviving variants (without mutagenesis). In combination, the two selection steps allow only the top 0.05% of cells

to survive. We emphasize again that selection in both steps mimics the probability-dependent selection due to the presence of expression noise.

In addition, our sequencing (Figure2) and phenotypic data (Table S2) support our assertion that selection is strong. Specifically, our populations of surviving cells show an increase in cyan fluorescence with a gain of just one or two mutations on average. For example, most variants in the populations H2 and H4 harbour 1 or 2 mutations (Figure R1 below) after the second round of directed evolution, and show a 4.6 and 3.3 fold increase in cyan fluorescence respectively. (Table S2). This shows that out of the ~10^5 variant populations only those that confer a large fitness benefit are selected.

Finally and most importantly, the same selection scheme when preceded by stabilizing selection, yields an almost 40 fold increase in cyan fluorescence within five rounds of directed evolution in both H and L populations (Table S2). If selection was effectively weak, we could not possibly have observed such a dramatic change in cyan fluorescence. All of the above shows that selection was indeed very strong in our experiments.

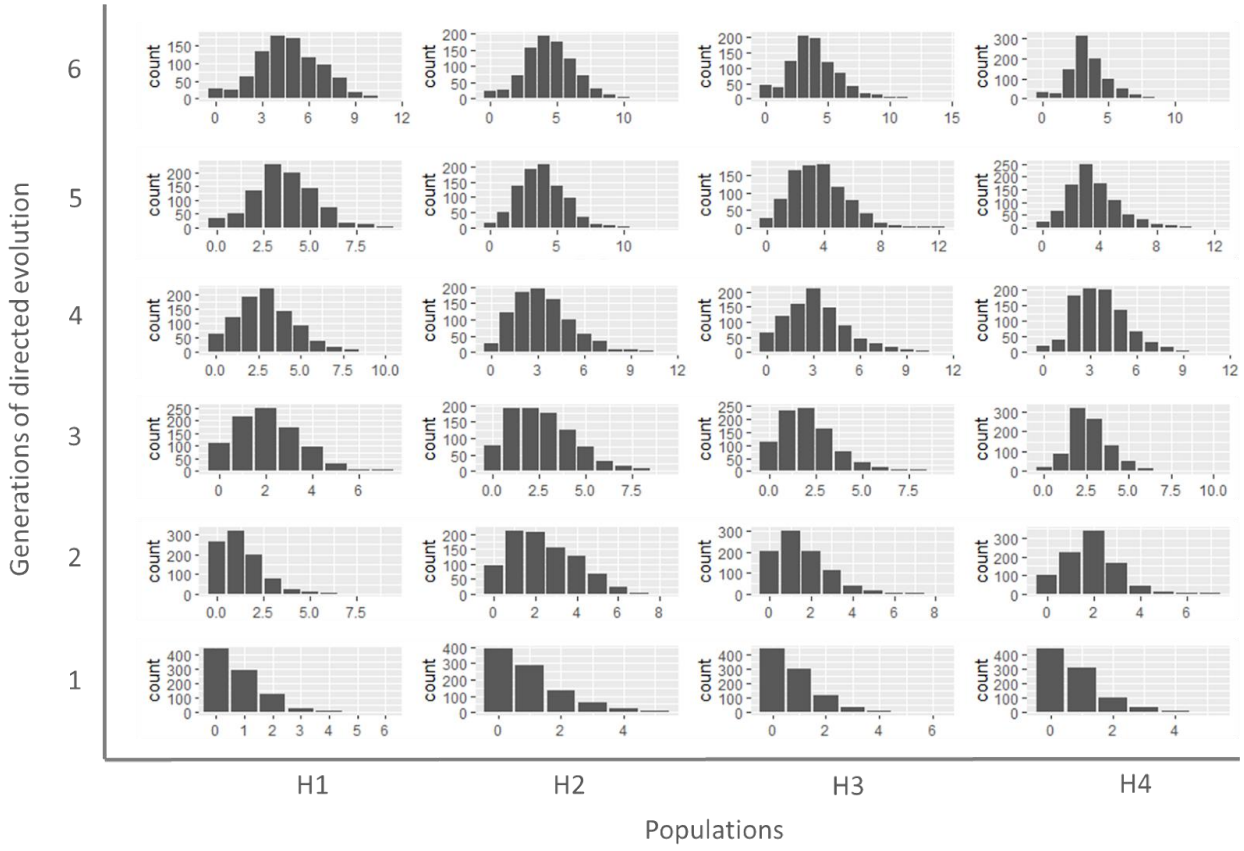

Figure R1: Distribution of the number of mutations in GFP molecules for the four replicate H populations during the six rounds of the first directed evolution experiment. Each row along the y-axis shows the distribution of the number of mutations in one generation of directed evolution, while each column along the x-axis shows the number of these mutations for a single replicate population during all six generations.

All that being said, we agree that your concern about selection regime and strength are important and need to be discussed. We have thus now added the following passages to the Results to explain how our selection criteria mimic probability-dependent selection. We have also underlined the necessity of strong selection for the higher evolvability of lowly expressed proteins to manifest, and the independence of this phenomenon from the particular kind of selection (the top 0.05% percent of a population) that we use in our experiments.

Lines 355-372 "*Scarcity-stability synergy requires two further conditions to be met. The first is that directional selection must be strong, otherwise high protein abundance will not be able to compensate effectively for reduced stability (Figure 3H). Although we use threshold-based (truncation) selection in our experiments, we emphasize that a synergy between scarcity and stability can also manifest itself during other kinds of selection, as long as selection is strong. For instance, SSS also causes low evolvability for highly expressed proteins under a probabilistic form of selection, i.e., when a cell's probability of surviving selection decreases with decreasing fluorescence (Supplementary information, Figure S11). In fact, the mere presence of gene expression noise effectively causes selection to be probabilistic. For example, in our experiments selection acted on subpopulations of ~10^4 genetically identical cells for each of ~10^5 different variants created by mutation. Each of the genetically identical cell is subject to gene expression noise, which causes different cells to express different amounts of the same protein, such that their fluorescence varies probabilistically. In other words, the population of individuals surviving selection is sampled from a meta-distribution made up of ~10^5 individual distributions. Cells in a subpopulation with higher mean fluorescence also have a higher probability of surviving selection, which renders selection probabilistic.*

The second requirement relates to gene expression noise [1,2,37]. In the absence of gene expression noise, scarcity-stability synergy does not lead to higher evolvability at low expression (Supplementary information, Figure S8)."

The third issue you have raised is the assumption of log-normality. Please also see the detailed justification we have given below in the response to the third major concern. Briefly, we show that low expression will enhance evolvability whenever expression noise follows a heavy-tailed distribution (Figure S10). It is important in this regard that heavy-tails are indeed a general property of gene expression noise. Expression patterns of genes follow a heavy-tailed distribution in organisms as different as bacteria, yeast, and humans (references 38 to 42 in the main text).

In sum, our method of selection exerts strong selection pressure, because in each round of directed evolution, we select cells whose fluorescence intensity is higher than 99.95% of the rest of the population. In addition, low expression leads to high evolvability also under other, probabilistic selection regimes

more akin to natural selection. Finally, as a result of gene expression noise, our selection procedure is, in fact, probabilistic.

Reviewer 1: This should not be call an “abundance-stability compensation”, it should be call “increased tolerance due to large variance”. Note that the authors have also failed to explain this ultimate cause (“increased tolerance due to large variance”) of their observation in Introduction or Abstract. They instead just keep using “abundance-stability compensation”. The “abundance-stability compensation” could be understand in two ways, one is the “increased tolerance due to large variance” as modelled by the authors, the other is “increase amount of proteins could lead to tolerance of destabilizing mutations as long as the total functional activity meets some requirement”. The difference between this two understandings can be illustrated with a normal distributed expression. In the first understanding, as shown by the authors, normal distribution cannot create the “abundance-stability compensation”. In the second understanding, however, normal distribution can still create the “abundance-stability compensation”.

Response: We completely agree with you on this point and we have dropped all instances of the ambiguous term “abundance-stability compensation” from the manuscript. We want to underline though that the large variance that endows proteins with increased tolerance for destabilizing mutations is highly correlated with large abundance, such that the two are not necessarily separable. Since our experiments controlled for abundance but not variance, we feel that it is more appropriate to use a term emphasizing abundance in describing the mechanism we discovered. We thus coined a new term which reflects the essence of our observations, namely that there is a synergy between protein scarcity and stability, in the sense that low protein expression (scarcity) promotes the evolution of stability. We call this phenomenon ‘scarcity-stability synergy’ (SSS). This term also precludes the other possible interpretation of ‘abundance-stability compensation’, namely that increased amounts of a protein can help tolerate destabilizing mutations as long as the total activity exceeds some threshold. Below are all the instances where we use this newly coined term:

Lines 24-26 *"Because many proteins meet the essential requirements for this scarcity-stability synergy (SSS), it may be a widespread mechanism by which low expression helps proteins evolve new phenotypes and functions."*

Lines 344-346 "*The mechanism we identified for the high evolvability of lowly expressed proteins relies on a synergy between protein scarcity and stability (SSS), in the sense that low protein abundance favors stable proteins (which in turn facilitate evolvability)*[21]*.*"

Line 355 "*Scarcity-stability synergy requires two further conditions to be met.*"

Lines 359-361 "*SSS also causes low evolvability for highly expressed proteins under a probabilistic form of selection, i.e., when a cell's probability of surviving selection decreases with decreasing fluorescence*"

Lines 370-371 "*In the absence of gene expression noise, scarcity-stability synergy does not lead to higher evolvability at low expression*"

Lines 391-392 "*Scarcity-stability synergy is likely to be important far beyond GFP, because it exists wherever abundance and stability contribute multiplicatively to a protein phenotype*[46-48]*.*"

Lines 397-398 *"In consequence, scarcity-stability synergy may be widespread in facilitating the evolution of new phenotypes and functions in lowly expressed proteins.*

Finally, we note that even though the phrase 'scarcity-stability synergy' does not explicitly include the variance component, we highlight the importance of this component in several places, including the following:

Lines 370-373 "*In the absence of gene expression noise, scarcity-stability synergy does not lead to higher evolvability at low expression (Supplementary information, Figure S8). Also, gene expression noise must cause a heavy-tailed abundance distribution, i.e., a distribution in which a greater proportion of proteins show high abundance than in a normal distribution (Figure S9)."*

Lines 384-388 "*. A complication in distinguishing the role of gene expression noise from that of average expression is that the two are correlated – low mean gene expression entails greater expression noise[43]. Our results suggest that whenever gene expression noise facilitates adaptive evolution, the lower average expression of evolving proteins may be part of the reason.*"

Reviewer 1: The correct way a simulation for selection should instead be carried out is that the probability of surviving is proportional to the fitness (fluorescence). In the case of this study, the authors should sample the ancestral population 10,000 times with replacement, and the probability of a cell being sampled equals its fluorescence intensity. As a result, this selective scheme will allow x% of cells in the H population to survive, but y% of cells in the L population to survive, where y > x (in contrast, x = y in the current selective scheme by the authors). I can only be persuaded on this major concern if the authors could repeat their conclusion using this type of more realistic selection.

Response: Thank you very much for this insightful suggestion, which we gladly followed by performing simulations to compare the surviving fractions of the H and L populations. In doing so, we implemented the selection scheme that you suggested. Briefly, we generated two ancestral populations with 50,000 cells each to represent H and L populations. The distribution of protein abundance was log-normal in these populations, the only difference being that the H population had a higher mean and standard deviation compared to the L population. We used a mean abundance of 100, and 300 copies, and a standard deviation of 30 and 90 copies for the L and the H populations, respectively. We also used a Gaussian distribution of protein stabilities (ΔG) with ΔG_{mean}=-12 kcal/mol (corresponding to that of WT GFP), and ΔG_{sd}=2 kcal/mol. We chose these values from previously measured estimates of protein folding stabilities (see: *PNAS* 101.46 (2004): 16192-16197; *Journal of molecular biology* 369.5 (2007): 1318-1332). We then sampled 10000 cells from these populations with replacement, where the sampling probability increased linearly from the first to the 100th percentile of the fluorescence intensity,

$$P(sampling\ from\ percentile\ i) = \frac{i}{\sum_{i=1}^{100} i} = \frac{i}{5050}, \qquad \text{(Equation R1)}$$

and repeated this sampling procedure 1000 times. Each time, we recorded the average stability of the GFPs that survived selection in the simulated H and L populations. We then compared their average stabilities. This analysis showed that the GFPs in the L population are more stable than in the H

population (see figure R2, $p \sim 10^{-16}$; Wilcoxon rank sum test). In sum, our results hold also if we implement selection in the way you suggest.

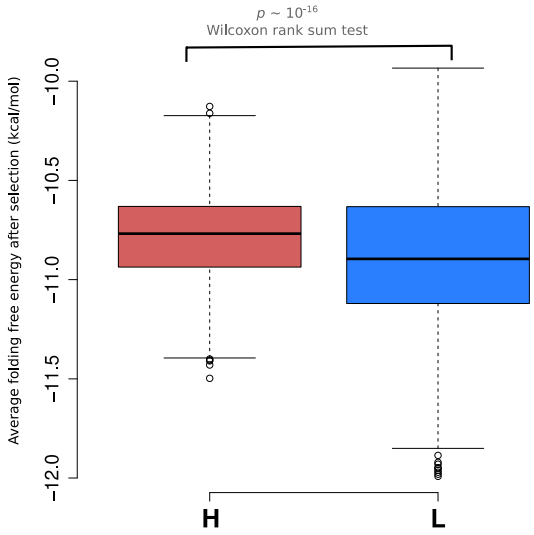

Figure R2. The average folding stability of selected GFPs in H and L populations in 1000 simulations. In each simulation, we sampled 10000 GFPs from an ancestral populations with 50000 cells. The probability that a cell survives selection in these simulations was proportional to its fluorescence intensity (Equation R1).

Reviewer 1: My second major concern is highly related to the first one, but it is independent. They authors have mixed expression noise and among-individual variation of mean expression in their simulation. Expression noise is largely log-normal, but among-individual variation of mean expression is not. The selective agent (trait being selected) should not be the instantaneous expression of a cell, which is heavily influenced by the expression noise. Instead, the selective agent should be mean expression (averaged across different time points), which is much less affected by expression noise.

Response: Thanks a lot for pointing out that we did not clearly distinguish between the instantaneous GFP expression and the mean GFP expression of cells in our writing. We agree that selection will often act on the time-average expression of the traits of interest, for example, in the case of resistance to beta-lactam antibiotics by β-lactamases. However, our simulations needed to model the experiments that we

performed. In these experiments we select for cells that pass through the flow-cell of a FACS's cell sorter. We select fifty thousand cells that fluoresce in the top 1% (or top 5%) bin *at the time of* sorting.

However, the following considerations indicate, perhaps surprisingly, that the mean fluorescence values relevant for selection in our experiment are not normally but also log-normally distributed at the time of selection. That is, they also show the kind of heavy-tailedness that is required for scarcity-stability synergy. We illustrate this principle with an example and with additional simulations. As mentioned before, our mutation libraries have ~10^5 variants distributed among ~10^9 cells, i.e. ~10^4 cells of each variant. We thus simulated 10^5 log-normal distributions, randomly sampled 10^4 cells from each distribution, pooled all these cells, and examined their distribution. This 'meta-distribution', which includes expression noise as well as among-individual variation, is also log-normal (Figure R3, p~0; likelihood ratio test). The reason is that our final distribution is not an average of different log-normal distributions, which would have been a normal distribution as you correctly pointed out. Instead, the final distribution is a convolution of different log-normal distributions, which is itself log-normal (see: Mitchell, R. L. (1968). Permanence of the log-normal distribution. *JOSA*, *58*(9), 1267-1272.)

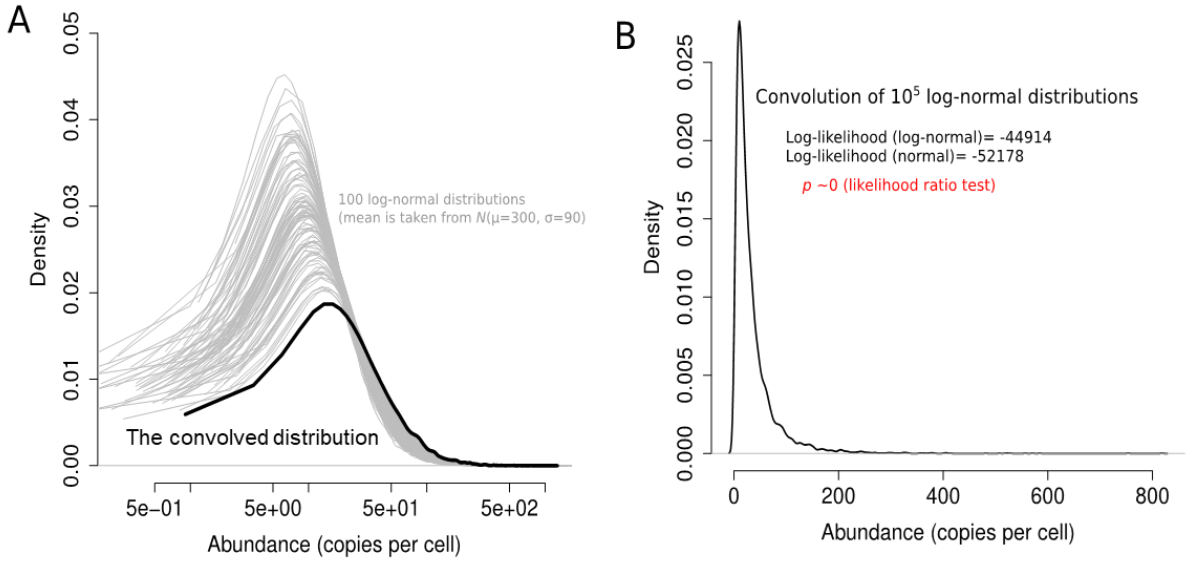

Figure R3: The convolution of log-normal distributions is a log-normal distribution itself. A) An example where gene expression noise is modelled by log-normal distributions. We generated 100 log-normal distributions (shown in grey) and show the convolution of all these 100 distributions in black. B) The convolution of 10^5 log-normal distributions is log-normally distributed. We used a likelihood ratio test to compare the likelihoods of a normal and a log-normal distribution fitting the data. In both panels A and B, the mean and standard deviation of each log-normal distribution were taken from a normal distribution with μ=300 and σ=90 copies, respectively.

To make the distinction between expression noise and among-individual variation clear, we added the following statement to the results:

Line 181-186 "*This expression level is subject to gene expression noise, which causes expression variation within a population, which we modeled through an empirically motivated log-normal distribution. Note that in our populations, ~10^5 GFP variants created by mutations are distributed among ~10^9 cells. This implies that at every instance, the fluorescence distribution of the 10^9 cells can be thought of as being composed of ~10^5 log-normal distributions. The 'meta-distribution' of these log-normal distributions is also log-normally distributed (Supplementary Information, Figure S13).*"

as well as in the discussion:

Lines 362-367 "*For example, in our experiments selection acted on subpopulations of ~10^4 genetically identical cells for each of ~10^5 different variants created by mutation. Each of the genetically identical cell is subject to gene expression noise, which causes different cells to express different amounts of the same protein, such that their fluorescence varies probabilistically. In other words, the population of individuals surviving selection is sampled from a meta-distribution made up of ~10^5 individual distributions.*"

In addition, we have added figure R3 to the supplementary methods (Figure S13).

Reviewer 1: The third major concern I have is about the experimental design. Green and cyan are two very similar colors. The excitation spectrum of GFP heavily overlapped with the cyan channel (the authors show it in Fig.3A). In this case, the H group GFP could be so strong that its signal on the cyan channel is already strong enough to mask any neo-functionalizing mutations. On the contrary, the L group GFP is relatively weak such that the cyan channel can be used as an efficient selective agent as designed. As a result, the selection for cyan is expected to be weaker for the H group compared to that for the L group.

Response: Point well taken. The spectra do indeed overlap, but we note that the excitation maxima are more than 80 nm apart from each other. More importantly, however, this overlap does not affect the detection of neo-functionalizing mutations in H populations, because the neo-functionalizing mutations clearly shift the fluorescence from green to cyan in *both* H and L populations (Figure S2). In fact we observe that the H populations harbour a higher average number of neo-functionalizing mutations per GFP molecule than the L populations, which suggests that the neo-functionalizing mutations might get selected more effectively (but certainly no less effectively) in H populations compared to L populations. We also observe that H populations can effectively increase cyan fluorescence after the stabilizing phase (Figure 4). If the stronger cyan signal in H populations was preventing selection to act effectively, it would have manifested in this second directed evolution experiment as well.

That being said, we agree that this matter should be discussed. We have thus now added the following explanation to the methods:

Lines 502-504 "*The green (ancestral) and cyan (new) fluorescence spectra overlap to some degree but the excitation maxima are more than 80 nm apart. More importantly, however, this overlap does not affect the selection of GFP variants with neo-functionalizing mutations in the evolving GFP populations (Figure 2B).*"

Reviewer 1: In a similar vein, this problem of experimental design extends to the problem of correlation between traits. I believe evolvability for a highly correlated trait can be a totally different story, compared to the evolvability for an orthogonal trait. The authors should also discuss this point.

Response: Thank you for the suggestion. We completely agree that the evolvability of two correlated traits may differ from that of orthogonal traits. We note, however, that orthogonal traits are quite rare when proteins evolve new functions. Most new protein functions are correlated to previous primary or secondary (moonlighting) functions (see: Aharoni et al. (2005). "The evolvability of promiscuous protein functions. *Nature genetics 37*(1): 73-76 as well as Khersonsky and Tawfik (2010). Enzyme Promiscuity: A

Mechanistic and Evolutionary Perspective. *Annual review of biochemistry 79*(1): 471-505). For this reason, we believe that the evolvability of correlated traits (like the traits we study) is more relevant to the evolution of novel traits than the evolvability of entirely orthogonal traits. We now address this point of correlated and orthogonal traits in the methods:

Lines 504-507 "*Correlations between two traits are the rule rather than the exception in evolving proteins, because ancestral and derived phenotypes are correlated in many proteins. That is, most novel protein functions are initially correlated with an ancestral function, and diverge only later through mechanisms such as gene duplication*[55,56]*."*

Reviewer 1: The results in Table S3 is rather confusing. How can a single mutation have such different effects (especially D113N) between H and L? I think increasing a certain fluorescence or not is a physical property, and has nothing to do with expression level.

Response: We thank you for raising this issue, and were just as intrigued by it even before submitting the original manuscript. As you point out, a change in fluorescence due to a mutation is a physical property. Although we express any one mutant protein in either the H or L background, the fold changes in fluorescence should be comparable across H and L populations. But even though we conducted additional experiments and analyses described below, we have not yet found a satisfactory explanation for this observation.

As a first step to better understand this observation, we repeated our sequencing and fluorescence measurements to confirm the observed pattern. It held true. We also checked whether the individual GFP variants are less stable in the H expression background than in the L background, which might lead to a reduced increase in fluorescence due to misfolding. This was not the case (Figure R4 below, p=0.5, paired t test). More precisely, while the estimated Tm values are not exactly identical for the same variant when expressed in the H and L background, the differences are small in magnitude and not systematic in nature.

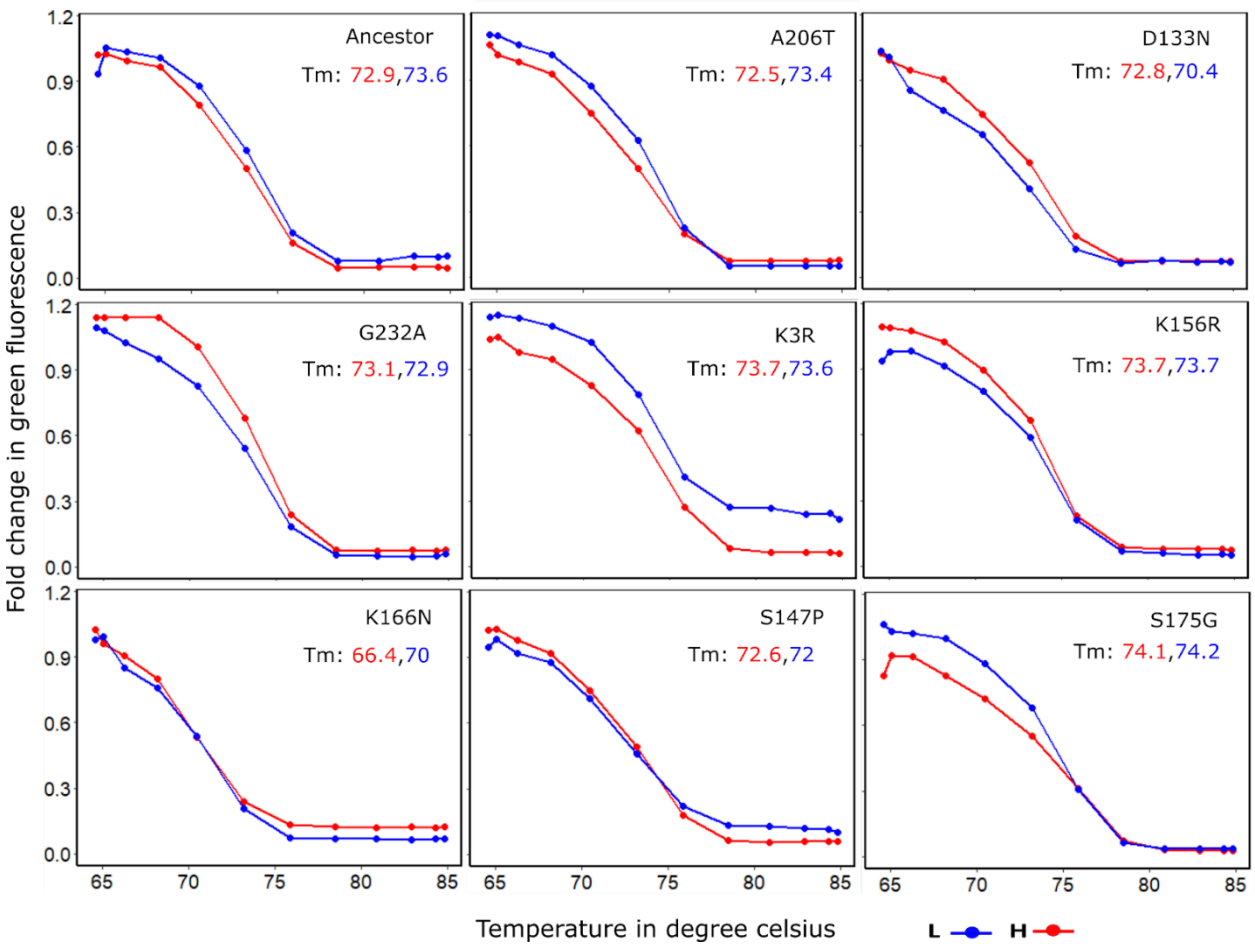

Figure R4: Each panel represents the loss of green fluorescence with increasing temperature for a single mutant generated in the H (red) or L (blue) genetic background. Each circle represents a mean over three replicate measurements. The engineered mutation and the estimated Tm for H and L populations are shown in each every panel.

In addition, we scoured the literature for a possible mechanism that might explain this puzzling observation but failed to locate any before we submitted the manuscript. Only a very recent paper has shown that the fitness effect of a single mutation can be expression dependent (see: Wu et. al. (2022). Expression level is a major modifier of the fitness landscape of a protein coding gene. *Nature ecology & evolution*, *6*(1), 103-115). The authors also show that high expression affects most mutations negatively,

i.e. most beneficial mutations become less beneficial, and most deleterious mutations become more deleterious when highly expressed. Although this work suggests a possible explanation for our observation, it also differs from ours in one important respect: The phenotype that the authors assess is reproductive fitness of the cell, and misfolding toxicity explains their observations. Although misfolding toxicity has no role in our experiments, because the growth rates of H and L populations are indistinguishable throughout the evolution experiment (Table S1), we suspect that expression level may affect the fluorescent output of some single mutants. We plan to investigate this hypothesis in future.

Reviewer 1: The sequence difference among variants is quite small, which could lead to heteroduplexes during PCR before SMRT sequencing, and increase sequencing error therein. How is this problem corrected ?

Response: Thank you for raising this valid concern. Based on previous directed evolution experiments with the fluorescent proteins in our lab, we anticipated the possibility that selected GFP populations may have very few differences in the nucleotide sequences. Thus we had already optimized our protocol to minimize the formation of heteroduplexes. First, we used the Phusion hot start enzyme to minimize the reaction time. This high-fidelity polymerase is expected to replicate one GFP amplicon in fewer than 5 seconds. Second, we used a very small amount of template (~40 ng in 25 µl reaction mix). Third, we used as few PCR cycles as possible (5 cycles for the first step and 25 for the second step). Even if these precautions may not prevent the formation of heteroduplexes altogether, we note that the incidence of such error will be comparable across H and L populations. They should thus not affect our conclusions, which are based on a comparison between H and L populations.

Reviewer 1: Line 62, “compensated” in terms of what ? fitness? amount of functional molecules ? fluorescence intensity ?

Response: Point well-taken. The sentence now reads as follows:

Lines 63-65 “*We demonstrate that under strong directional selection, the deleterious effects of a destabilizing mutation that reduces fluorescence intensity can be compensated for by increased protein abundance.*”

Reviewer 1: Line 123 “T230S” should be “T203S”?

Response: Thank you. Rectified.

Reviewer 1: Line 150 “10% of GFP population survives, light gray”, light gray should be black.

Response: Thanks. Corrected.

Reviewer 1: Also, Fig3D and E is confusing, what is the difference between the gray-color contrast and the dashed vertical line ? It looks like there are two types of selection .

Response: Thanks for raising this point, which requires clarification. The dashed vertical lines show the relative abundances of proteins that lie in the 90th percentile of absolute protein abundances in simulated H and L populations. Because of log-normality and the increased skewness of the abundance distribution of H population, the 90th percentile of protein abundance corresponds to smaller relative abundances in these populations. We changed the legend to clarify this point.

Reviewer 1: Line 223 “Figure 3D” should be a wrong mention

Response: Replaced by ‘Figure 3I’. Thank you for pointing this out.

Reviewer 1: Line 262 “Figure 3B” should be a wrong mention.

Response: Thank you. Replaced by ‘Figure 3I’.

Reviewer 1: Line 271 “variant” to “variants”.

Response: Rectified.

Reviewer 1: Line 288 “Figure 3D”, is that a wrong mention too ?

Response: It was incomplete. Thanks for bringing it to our attention. Replaced by ‘Figure 3D and 3E’.

Reviewer 1: Reference 11 proposed a model for mistranslation-induced misfolding. If the authors are mostly talking about misfolding only (regardless mistranslation), I believe PMID:20959819 is a more accurate reference.

Response: Thank you for suggesting this more appropriate reference. We now refer to the Yang et. al. paper you refer to in the context of misfolding.

Reviewer 1: Supplementary line 74 “the H population has a relative fluorescence intensity” should be “the H population has a low relative fluorescence intensity”?

Response: Corrected.

Reviewer 1: FigureS6 C, there no text in bracket.

Response: Thank you for bringing this to our attention. Rectified.

Response to Reviewer 2

Reviewer2: Karve et al. conducted a directed evolution experiment on GFP, in which they selected for cyan fluorescence. Their experiment had one key treatment variable: the strength of the promoter expressing GFP, causing one set of replicate populations to have high GFP abundance, and the other set to have low GFP abundance. Strikingly, the novel cyan phenotype evolved more readily in the low abundance treatment. Karve et al. use computational modeling to illustrate that with a fitness function in which the protein phenotype under selection is proportional to the product of stability and abundance, populations with higher GFP abundance can meet the selective threshold in their experiment with lower stability proteins. By contrast, the low abundance populations do not have this "luxury". The authors then demonstrate that the proportionally higher selection pressure on protein stability in the low abundance populations is responsible for their key finding, by testing two predictions of the model: they demonstrate that the fluorescent proteins in the low abundance treatment are more stable/foldable than those in the high abundance treatment (refolding kinetics), and they conducted a second evolution experiment in which they show that the difference in evolvability disappears when they use stabilizing selection to increase the (hypothesized) number of stabilizing mutations in both populations, before applying strong directional selection.

Overall, the basic finding (low abundance proteins in this system are more evolvable) is interesting, and the reason why (the higher abundance proteins evolved lower stability) is both surprising as well as well supported by the model and follow-up experiments. I very much enjoyed reading this work, and I do not have any critical comments to make about the logic or technique used to establish these results. I think that this work may be widely read by researchers working in molecular evolution.

Response: Thank you very much for your positive feedback. We are immensely pleased to know that you enjoyed reading our manuscript. We also thank you for your suggestions and comments. Please find below our point-by-point reply to the individual minor issues you raised.

Reviewer2: I do have several minor comments to make about the presentation. I also think that the implications of this work should be more framed more precisely. In particular, many readers will read this work with prior expectations about reading about whether high abundance <=> high stability, or about the causes of the abundance/evolutionary rate correlation. So, it's possible that readers who skim this work by reading just the abstract or conclusions may fundamentally misunderstand the main findings and their importance.

Response: Thanks a lot for pointing this out. In response to this comment and other issues raised by the first reviewer, we have now modified the abstract to state the implications of our study more clearly. We have also added the following lines to the discussions to highlight that our study reveals an effect of expression level on the evolvability of protein *phenotypes*, and doesn't address the causes of the expression-evolutionary rate anticorrelation, which is about protein *genotypes* (sequences).

Lines 323-326 "*Our results pertain to the role of protein abundance in the evolvability of protein phenotypes on short evolutionary time scales. We emphasize that our observations are independent from and do not contradict the misfolding avoidance hypothesis, which can help explain why the sequences of highly expressed proteins evolve slowly on long evolutionary time scales.*"

Reviewer2: often, box plots are used to show differences between the treatment. Based on my reading of the legend, it seems that each box plot represents 4 data points (the mean for each population in the treatment). I would much rather see the underlying four data points for each (maybe jittered to show the distribution). If the box plots are being used to summarize many more datapoints (say distribution of fluorescence for all 4 populations together), then I would appreciate a brief comment in the legend stating exactly what the underlying data points are (10,000 cell counts per pop, or something like that?)

Response: Thanks a lot for this suggestion. As you correctly point out, every box plot represents four data points. We have now added the individual data points (jittered) in each case.

Reviewer2: Many of the figures have inconsistent fonts, both within and across panels in single figures, and across figures. It would be nice to make this consistent, unless this is something that the publisher is going to handle.

Response: Thank you for bringing this to our notice. We have now made all the fonts consistent across panels and figures.

Reviewer2: Figure 3 C and G have these straight lines, sometimes dashed, in the panels. I can't tell whether these are meaningful or if they are visual artefacts. If they are meaningful, please describe what they mean, if not, please remove.

Response: The lines were artefacts. Corrected. Thank you for pointing this out.

Reviewer2: I find the argument on lines 201-208 really hard to understand, starting with "The reason why ...". The arrow on Figure 3G is not much help either. I'm reading Fig. 3 C and G as hyperbolic isoclines based on the product of stability*abundance. I am really confused about how Fig 3G is showing something about a log-normal distribution. Based on Figure 3D and 3E, it just looks like the L populations have stronger selection on relative stability compared to H populations. The log-normal shape seems to be critical based on lines 206-208, but I don't understand this point, a good figure would make a big difference, but Fig 3G isn't cutting it. Help!

Response: Thanks for this suggestion and for pointing out the confusion that Figure 3G causes. Although we select a fixed fraction of top fluorescent cells in both populations, this fixed fraction corresponds to a lower relative (to maximum) fluorescence intensity in the H populations. This is caused by the log-normal distribution of protein abundance. In log-normal distributions, skewness increases with the standard deviation such that the relative value of a fixed percentile gets smaller with increasing standard deviation. We revised Figure 3G substantially to better illustrate these two points. Now the panel shows two schematic distributions with blue and red colours that correspond to the distributions of the relative abundance of the L and H populations, respectively. The distribution of the relative abundance of cells in the H populations is left-skewed compared to the L populations. As a result, the relative abundance of a

fixed percentile of protein abundance in the H population is lower than that of the L populations. This is shown by dashed blue and red lines for the L and the H populations, respectively. The difference in distribution leads to the survival of cells in H population that are more likely to express GFPs with lower stabilities than those in L populations. Compare the red and the blue areas on the surface plot shown in the revised Figure 3G.

Reviewer2: Perhaps this is a major point: the title of the paper should be rewritten. It's misleading. The first part "low expression enhances protein evolvability" is great, but low expression isn't mitigating folding defects! A chaperone mitigates folding defects-- here, proteins with folding defects in the low expression treatment don't survive selection. Something like "High abundance reduces protein evolvability by reducing the strength of selection on protein stability" seems more accurate, or "Low expression enhances protein evolvability by intensifying selection on foldability", or really, whatever is most accurate.

Response: Thank you very much for this great suggestion. Following your suggestion we have now changed the title of the manuscript to ‘Low protein expression enhances phenotypic evolvability by intensifying selection on folding stability’.

Reviewer2: In the abstract, the following is really confusing given my prior assumptions about unrelated questions in the field: "proteins with folding defects preferentially accumulate when a protein is highly expressed, because high protein abundance can compensate for such folding defects." This read to me like the per unit rate of misfolding increases with gene expression, because interactions between proteins somehow is helping the defective proteins to fold properly. I get it after reading the paper carefully, but if I had just read the abstract, I really wouldn't understand what this paper is about. What really helped me was the equation for fitness in the paper-- then I immediately understood what was going on (at least up to the log-normal stuff). Something like "Under strong directional selection for cyan fluorescence, high fitness can be achieved by either a smaller number of GFP with few folding defects, or a larger number of GFP with proportionally more folding defects. We call this principle XXXX".

Response: Thank you very much for this suggestion. We have now revised the abstract completely to make it clearer and more accurate to match our claims. We hope that the current framing of the abstract has eliminated the confusion you mentioned.

Reviewer2: I also find the "abundance-stability compensation principle" misleading for the following reason. Abundance and stability can compensate for each other to get pass the selective threshold in these experiments. But abundance-stability compensation is not a principle of protein evolvability-- it's not that evolvable proteins have to be either abundant or have to be stable. Rather, the low abundance proteins are forced to be more stable and thus evolvable. The fact that they __cannot__ compensate is what makes them evolvable!

Response: Point well-taken. We have now changed the term to one that we believe is more accurate. Because the scarcity of a protein (its low expression) favours stability, and is in this sense synergistic with stability, we refer to the phenomenon as ‘scarcity-stability synergy’. We also do no longer refer to it as a principle of protein evolvability but as a phenomenon (‘something that happens or exists’).

Reviewer2: Then the sentence on line 20: "Its consequence is that lowly expressed proteins are __under stronger selection__ for proper folding, which facilitates their evolvability".

Response: Rectified. Thank you for the suggestion.

Reviewer2: These same criticisms hold for the language in the Introduction on lines 60-63, and the Discussion on lines 330-344 and lines 368-375. To reiterate, the lack of compensation for stability is what is causing increased evolvability in the L populations, so stating that "abundance-stability compensation facilitates the evolution of new phenotypes" gets it backwards (especially since abundance does not compensate for stability in terms of evolvability; it's not symmetric in this regard), and will be really confusing for readers who do not read this work carefully.

Response: We completely agree with your point that the phrasing is not accurate. In the light of this and the previous comments we have rephrased the inaccurate framings in the previous version. They now read as follows:

Lines 61-65 "*Our results show that high expression diminishes evolvability of a protein, and it does so for reasons that are unrelated to the misfolding toxicity of highly expressed proteins. We demonstrate that under strong directional selection, the deleterious effects of a destabilizing mutation that reduces fluorescence intensity can be compensated for by increased protein abundance."*

Lines 344-351 "*The mechanism we identified for the high evolvability of lowly expressed proteins relies on a synergy between protein scarcity and stability (SSS), in the sense that low protein abundance favors stable proteins (which in turn facilitate evolvability)[21]. The reason is that fluorescence intensity, our focal phenotype, depends on the product of protein abundance and stability. In consequence, mutations can reduce protein stability to some extent, as long as protein abundance can compensate for the reduced stability. This is why the stability of GFP evolving in H populations can decrease to a greater extent than in L populations. Because the ability to compensate for reduced folding stability is smaller in L populations, these populations evolve genotypes with higher folding stability.*"

Lines 391-398 "*Scarcity-stability synergy is likely to be important far beyond GFP, because it exists wherever abundance and stability contribute multiplicatively to a protein phenotype[46-48]. This multiplicative relationship has been shown to accurately predict the changes in bacterial growth rate with different orthologs of essential proteins, such as dihydrofolate reductase (DHFR)[8], as well as resistance conferring enzymes, such as β-lactamase[6].The additional requirements of a heavy-tailed protein abundance distribution and strong selection also apply to a broad range of proteins and organisms, both during experimental evolution and in the wild[21,38-42,49-51]. In consequence, scarcity-stability synergy may be widespread in facilitating the evolution of new phenotypes and functions in lowly expressed proteins.*"

Reviewer2: And the last line of the abstract: "to help" -> "by which", since the lowly expressed proteins do not have agency in their evolution.

Response: Rectified. Thank you.

Reviewer2: I really enjoyed reading this paper, and I hope these minor criticisms help this work reach a wider audience.

Response: We sincerely thank you for the positive remarks and the constructive criticism, which have helped us make the narrative much clearer.

Response to Reviewer 3

Reviewer3: This is a nice contribution, and I generally agree with the way the analysis is set up and with the conclusions that were drawn.

Response: Thank you very much for your positive feedback! Please find below the point-by-point reply to the concerns you raised.

Reviewer3: I just feel that at times (and in particular in the abstract), readers could be confused by statements such as "proteins with folding defects preferentially accumulate when a protein is highly expressed", as they seem to contradict the widely accepted result that more highly expressed proteins are less likely to exhibit folding defects.

Response: Thank you for pointing this out. In light of this comment and the comments from other reviewers, we have re-written the abstract. The aforementioned sentence now reads as:

Lines 22-23 "*The reason is that high fluorescence can be achieved by either few proteins that fold well, or by many proteins that fold less well.*"

We have also modified other, similar instances in the manuscript.

In the introduction, we now write:

Lines 65-66 "*As a result, proteins that are highly expressed allow more genetic variants that cause folding defects, which reduces their evolvability.*"

In the results:
Line 135 "*Figure 3. Modeled H populations retain more destabilizing mutations*."

Line 147-148 "*Figure 3G: Increased skewness of relative GFP abundance in H populations can explain the retention of more destabilizing mutations in these populations compared to L populations.*"

Lines 163-164 "*The most prominent class of such mutations impair protein folding[20,25], and we hypothesized that such variants are retained when GFP is highly expressed.*"

Lines 169-170 "*To identify a mechanism that may cause H populations to retain more destabilizing deleterious mutations than L populations, we developed a computational model for the directed evolution of GFP.*"

Lines 207-209 "*Importantly, although the same percentile of fluorescent cells survived selection in both populations, the H population harbored more destabilizing mutations in the model ($p < 10\text{-}16$, Wilcoxon's rank sum test; Figure 3F).*"

Lines 221-223 "*In sum, when GFP is highly expressed during evolution, it retains more destabilizing mutations than when it is lowly expressed (Figure 3H).*"

And in the discussion:
Lines 313-315 *"The reason was that highly expressed GFP populations retained more deleterious mutations that caused protein misfolding."*

Reviewer3: I think the authors understand exactly what is happening in their system and the Discussion is reasonably clear about how everything fits together, but there's still some room for improvement in the writing I think. To me, what resolves the paradox is the difference in time scales. Over long evolutionary times, genes that are highly expressed will slowly evolve into variants that are less likely to misfold. The selection pressure that causes this effect is quite weak, though, and likely doesn't have any effect in short-term experiments such as the ones presented here.

Response: Thank you for this valuable suggestion. We fully agree that purifying selection acting on the sequences of highly expressed proteins should be very weak on the short timescale of directed evolution (also see: Shibai et. al. bioRxiv 2022.03.02.482674; doi: https://doi.org/10.1101/2022.03.02.482674). We now explicitly write about this difference in the time scales:

Lines 323-336 "*Our results pertain to the role of protein abundance in the evolvability of protein phenotypes on short evolutionary time scales. We emphasize that our observations are independent from and do not contradict the misfolding avoidance hypothesis, which can help explain why the sequences of highly expressed proteins evolve slowly on long evolutionary time scales. This hypothesis posits that highly expressed proteins impose a fitness cost on a cell when misfolded, because protein misfolding can be toxic, and especially so for highly expressed proteins. Destabilizing mutations in such proteins are thus rarely tolerated, which leads to low rates of amino acid sequence evolution on longer evolutionary time scales. The selection pressure that causes this effect may be too weak to manifest itself on short evolutionary time scales. Misfolded proteins can be costly to a cell when present in large numbers and reduce the reproductive fitness of the cell. The resulting misfolding toxicity does not play a major role in our experiments, because first, the amount of GFP expressed in H and L populations is not very high (see Methods for details of expression levels). Second and more important, the growth rates of H and L populations are indistinguishable throughout our evolution experiment (Table S1). In other proteins or experiments where misfolding toxicity plays a role, it may cause an additional evolvability disadvantage for highly expressed proteins.*"

Reviewer3: I would like to encourage the authors to read the entire paper carefully and see whether there are places where they can emphasize the time scale dependence. For example, throughout the paper, the authors write about what happens under strong directional selection, and many of these phrases could possibly improved by adding "over short time scales" or similar.

Response: We fully agree with you that the time scale dependence was not emphasized sufficiently in the previous draft. We made the following changes to better emphasize the relevance of time scale:

In the abstract:

Lines 20-22 " *Lowly expressed proteins are under stronger selection for proper folding, which facilitates their evolvability on short evolutionary time scales.*"

In the discussion:

Lines 323-324 "*Our results pertain to the role of protein abundance in the evolvability of protein phenotypes on short evolutionary time scales.*"

Lines 328-330 "*Destabilizing mutations in such proteins are thus rarely tolerated, which leads to low rates of amino acid sequence evolution on longer evolutionary time scales. The selection pressure that causes this effect may be too weak to manifest itself on short evolutionary time scales.*"

Reviewer3: Likewise, when the authors state that selection against toxic misfolding is not present in their system, they could add 1-2 explanatory sentences explaining why we think this selection pressure isn't there.

Response: Thank you for this suggestion. We now explain this briefly in the discussion:

Lines 328-335 "*. Destabilizing mutations in such proteins are thus rarely tolerated, which leads to low rates of amino acid sequence evolution on longer evolutionary time scales. The selection pressure that causes this effect may be too weak to manifest itself on short evolutionary time scales. Misfolded proteins can be costly to a cell when present in large numbers and reduce the reproductive fitness of the cell. The resulting misfolding toxicity does not play a major role in our experiments, because first, the amount of GFP expressed in H and L populations is not very high (see Methods for details of expression levels). Second and more important, the growth rates of H and L populations are indistinguishable throughout our evolution experiment (Table S1).*"

Reviewer3: There's one other aspect to this work that I'd like the authors to consider and possibly discuss. If I understand correctly, the selection via FACS really measures two things at the same time: (i) Is there enough fluorescence to be detected; (ii) Is the fluorescence in the right wavelength. For low expression genes, they cannot afford losing any fluorescence signal at all, since it's low already, and hence they are selected for high quality proteins that don't misfold and fluoresce well. By contrast, for high expression genes, they can afford to lose signal intensity and hence they do so, because there is no actual selection pressure to maintain a strong signal. If this is correct, then the entire situation reminds me a bit of the idea of drift robustness discussed by LaBar and Adami (see e.g. https://pubmed.ncbi.nlm.nih.gov/29044114/). It's not exactly the same effect, but the low expression genes are essentially drift-robust because any drift away from their optimum likely causes them to be removed from the population. By contrast the high expression genes can afford to drift because they have a fitness buffer.

Response: Thank you very much for this insightful suggestion! We agree that the strong selection for the folding stability in L populations is somewhat analogous to the evolution of drift-robustness in small populations. Specifically, under strong selection, low protein abundance in L populations restricts the allowable variation in folding stability. This results in protein populations without small-effect deleterious mutations. We now mention this parallel in the discussion:

Lines 347-353 "*In consequence, mutations can reduce protein stability to some extent, as long as protein abundance can compensate for the reduced stability. This is why the stability of GFP evolving in H populations can decrease to a greater extent than in L populations. Because the ability to compensate for reduced folding stability is smaller in L populations, these populations evolve genotypes with higher folding stability. The phenomenon resembles the evolution of drift-robustness, in which genotypes from small populations can evolve reduced vulnerability to genetic drift, and become less likely to accumulate small-effect deleterious mutations*[36]*.*"

Decision Letter, first revision:

19th April 2022

Dear Shraddha,

Thank you for submitting your revised manuscript "Low protein expression enhances phenotypic evolvability by intensifying selection on folding stability" (NATECOLEVOL-211215405A). It has now been seen again by the original reviewers and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Ecology & Evolution, pending minor revisions to satisfy the reviewers' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Ecology & Evolution. Please do not hesitate to contact me if you have any questions.

[REDACTED]

Reviewer #1 (Remarks to the Author):

I want to thank the authors for their time and effort spent responding to my comments.

In general, I found their response to be satisfactory and the manuscript has been substantially improved. I only have a couple of remaining minor comments listed below.

1. I am surprised to learn that the convolution of log-normal distribution does not follow the Central Limit Theorem. The authors should cite the paper titled "Permanence of the log-normal distribution" at around line 186.

2. The suspected reason for the difference in effect of D113N between H and L populations should be discussed somewhere (Result or Discussion).

Reviewer #2 (Remarks to the Author):

Overall the quality of this work is very high and the authors did a good job addressing the reviewers' criticisms. However, I still find Figure 3G and the explanation in the text of how the log-normal distribution matters very unsatisfactory.

specific comments:

line 35: it should be clarified that misfolding toxicity -> ER-anticorrelation is a hypothesis, rather than an established fact. There are several other hypotheses for the ER-anticorrelation (as discussed in the Nat Rev Genetics paper cited here), and its specific causes are still debated. Since the misfolding hypothesis is brought up later, I would either move the citations to the previous sentence and strike out this statement that misfolding toxicity is THE cause, or I would qualify that this has been hypothesized as the cause for the ER-anticorrelation. I would probably take the first course, since ER-anticorrelation is not the point of this paper.

line 323: I think the key difference is that this work is about positive directional selection, and that the misfolding -avoidance hypothesis is about negative purifying selection. Timescales matter too, but I think this is secondary, and the recent Wu et al. (2022) paper "Expression level is a major modifier of the fitness landscape of a protein coding gene" in Nat Eco Evo, along with my 2021 GBE paper, suggest that the timescale difference is much less important than has been previously believed.

line 370: I think this should rather be something like "The second requirement is for a heavy-tailed distribution of protein abundance per cell, which is known to be caused by gene expression noise". My reasoning is that there may be other sources of stochasticity in the cell that could cause the heavy tailed abundance distribution, and Fig S9 shows that gene expression noise in itself is not sufficient; it has to be a heavy-tailed distribution. I initially misread this as expression noise is the key factor, when the reasoning here is a bit more subtle.

Figure 3G: At first glance this looks great. But when I tried to understand it, Figure 3G is super confusing. There are three pieces here: 1) a density plot of protein abundance per cell in the two populations 2) the isoclines for selection for fluorescence in each population (like panel C) 3) a density plot of fluorescence per cell, with thresholds for selection (like panel B). The 3rd piece does not even exist, but it looks like it does, given the confusing y-axis labels.

The main issue is that the axis labels for the density plots are missing. The x-,y-,z- axis refer to the 3D isocline. The y-axis for the density plot is "Density", and the x-axis is relative abundance. The vertical dashed lines by eye seem to be selective thresholds for the fluorescence, based on how they line up with the isoclines. But really, these are 10% relative abundance lines. But the 10% relative abundance line seems more or less irrelevant since it doesn't map nicely to the actual selection pressures. I would actually also cut the vertical dashed lines in D & E for that reason. There are selected cells below the this "threshold" in both panels D & E, and there are unselected cells above this threshold in panel E, which demonstrates how it's not directly relevant to the selection.

Finally the logic about the leftward-shift relative abundance in the high expression populations is super confusing, because selection in this system doesn't care about relative protein abundance per cell in itself. Lower relative stability does not necessarily evolve from having lower relative abundance, especially since the point of this paper is that the populations with lower absolute abundance evolve higher absolute stability! I think the simplest fix at this stage would be to better just to completely cut Figure 3G, and eliminate the dashed vertical lines in Fig 3D and Fig 3E. And I would just cut lines 209-216, and go straight from "... Figure 3F)." to "This observation, which rests..." That seems easier than doing more writing at this point.

Alternatively the authors could briefly describe the facts in Figure S8 and Figure S9, which that the log-normal distribution in abundance is critical, as shown in the simulations, without giving the conceptually convoluted explanation involving lower relative abundance, which raises more questions than answers.

The y-axis labels in panels A and B of Figures S8 and S9 should be absolutely consistent. Indeed, it looks like the log-normal distribution prevents the simulated H populations from achieving high stability, due to a long tail of cells with very high abundance. This long tail gets scaled away in the relative abundance calculation, and I think this is why the authors' explanation is so convoluted (focus on relative abundance). My feeling is there is probably a much simpler explanation involving absolute abundance.

Regardless of these criticisms in presentation, I think this paper is very fine work.

signed,
Rohan Maddamsetti

Reviewer #3 (Remarks to the Author):

Thank you for your careful revisions. I have no further comments.

Our ref: NATECOLEVOL-211215405A

22nd April 2022

Dear Dr. Karve,

Thank you for your patience as we've prepared the guidelines for final submission of your Nature Ecology & Evolution manuscript, "Low protein expression enhances phenotypic evolvability by intensifying selection on folding stability" (NATECOLEVOL-211215405A). Please carefully follow the step-by-step instructions provided in the attached file, and add a response in each row of the table to indicate the changes that you have made. Please also check and comment on any additional marked-up edits we have proposed within the text. Ensuring that each point is addressed will help to ensure that your revised manuscript can be swiftly handed over to our production team.

We would like to start working on your revised paper, with all of the requested files and forms, as soon as possible (preferably within two weeks). Please get in contact with us immediately if you anticipate it taking more than two weeks to submit these revised files.

When you upload your final materials, please include a point-by-point response to any remaining reviewer comments.

If you have not done so already, please alert us to any related manuscripts from your group that are under consideration or in press at other journals, or are being written up for submission to other journals (see: https://www.nature.com/nature-research/editorial-policies/plagiarism#policy-on-duplicate-publication for details).

In recognition of the time and expertise our reviewers provide to Nature Ecology & Evolution's editorial process, we would like to formally acknowledge their contribution to the external peer review of your manuscript entitled "Low protein expression enhances phenotypic evolvability by intensifying selection on folding stability". For those reviewers who give their assent, we will be publishing their names alongside the published article.

Nature Ecology & Evolution offers a Transparent Peer Review option for new original research manuscripts submitted after December 1st, 2019. As part of this initiative, we encourage our authors to support increased transparency into the peer review process by agreeing to have the reviewer comments, author rebuttal letters, and editorial decision letters published as a Supplementary item. When you submit your final files please clearly state in your cover letter whether or not you would like to participate in this initiative. Please note that failure to state your preference will result in delays in accepting your manuscript for publication.

Cover suggestions

As you prepare your final files we encourage you to consider whether you have any images or illustrations that may be appropriate for use on the cover of Nature Ecology & Evolution.

Covers should be both aesthetically appealing and scientifically relevant, and should be supplied at the best quality available. Due to the prominence of these images, we do not generally select images featuring faces, children, text, graphs, schematic drawings, or collages on our covers.

We accept TIFF, JPEG, PNG or PSD file formats (a layered PSD file would be ideal), and the image should be at least 300ppi resolution (preferably 600-1200 ppi), in CMYK colour mode.

If your image is selected, we may also use it on the journal website as a banner image, and may need to make artistic alterations to fit our journal style.

Please submit your suggestions, clearly labeled, along with your final files. We'll be in touch if more information is needed.

Nature Ecology & Evolution has now transitioned to a unified Rights Collection system which will allow our Author Services team to quickly and easily collect the rights and permissions required to publish your work. Approximately 10 days after your paper is formally accepted, you will receive an email in providing you with a link to complete the grant of rights. If your paper is eligible for Open Access, our Author Services team will also be in touch regarding any additional information that may be required to arrange payment for your article.

Please note that Nature Ecology & Evolution is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including <a href="https://www.nature.com/nature-portfolio/editorial-policies/self-archiving-and-license-to-publish. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

Please note that you will not receive your proofs until the publishing agreement has been received through our system.

For information regarding our different publishing models please see our Transformative Journals page. If you have any questions about costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com.

Please use the following link for uploading these materials:
[REDACTED]

If you have any further questions, please feel free to contact me.

[REDACTED]

Reviewer #1:
Remarks to the Author:

I want to thank the authors for their time and effort spent responding to my comments.

In general, I found their response to be satisfactory and the manuscript has been substantially improved. I only have a couple of remaining minor comments listed below.

1. I am surprised to learn that the convolution of log-normal distribution does not follow the Central Limit Theorem. The authors should cite the paper titled "Permanence of the log-normal distribution" at around line 186.

2. The suspected reason for the difference in effect of D113N between H and L populations should be discussed somewhere (Result or Discussion).

Reviewer #2:
Remarks to the Author:
Overall the quality of this work is very high and the authors did a good job addressing the reviewers' criticisms. However, I still find Figure 3G and the explanation in the text of how the log-normal distribution matters very unsatisfactory.

specific comments:

line 35: it should be clarified that misfolding toxicity -> ER-anticorrelation is a hypothesis, rather than an established fact. There are several other hypotheses for the ER-anticorrelation (as discussed in the Nat Rev Genetics paper cited here), and its specific causes are still debated. Since the misfolding hypothesis is brought up later, I would either move the citations to the previous sentence and strike out this statement that misfolding toxicity is THE cause, or I would qualify that this has been hypothesized as the cause for the ER-anticorrelation. I would probably take the first course, since ER-anticorrelation is not the point of this paper.

line 323: I think the key difference is that this work is about positive directional selection, and that the misfolding -avoidance hypothesis is about negative purifying selection. Timescales matter too, but I think this is secondary, and the recent Wu et al. (2022) paper "Expression level is a major modifier of the fitness landscape of a protein coding gene" in Nat Eco Evo, along with my 2021 GBE paper, suggest that the timescale difference is much less important than has been previously believed.

line 370: I think this should rather be something like "The second requirement is for a heavy-tailed distribution of protein abundance per cell, which is known to be caused by gene expression noise". My reasoning is that there may be other sources of stochasticity in the cell that could cause the heavy tailed abundance distribution, and Fig S9 shows that gene expression noise in itself is not sufficient; it has to be a heavy-tailed distribution. I initially misread this as expression noise is the key factor, when the reasoning here is a bit more subtle.

Figure 3G: At first glance this looks great. But when I tried to understand it, Figure 3G is super confusing. There are three pieces here: 1) a density plot of protein abundance per cell in the two

populations 2) the isoclines for selection for fluorescence in each population (like panel C) 3) a density plot of fluorescence per cell, with thresholds for selection (like panel B). The 3rd piece does not even exist, but it looks like it does, given the confusing y-axis labels.

The main issue is that the axis labels for the density plots are missing. The x-,y-,z- axis refer to the 3D isocline. The y-axis for the density plot is "Density", and the x-axis is relative abundance. The vertical dashed lines by eye seem to be selective thresholds for the fluorescence, based on how they line up with the isoclines. But really, these are 10% relative abundance lines. But the 10% relative abundance line seems more or less irrelevant since it doesn't map nicely to the actual selection pressures. I would actually also cut the vertical dashed lines in D & E for that reason. There are selected cells below the this "threshold" in both panels D & E, and there are unselected cells above this threshold in panel E, which demonstrates how it's not directly relevant to the selection.

Finally the logic about the leftward-shift relative abundance in the high expression populations is super confusing, because selection in this system doesn't care about relative protein abundance per cell in itself. Lower relative stability does not necessarily evolve from having lower relative abundance, especially since the point of this paper is that the populations with lower absolute abundance evolve higher absolute stability! I think the simplest fix at this stage would be to better just to completely cut Figure 3G, and eliminate the dashed vertical lines in Fig 3D and Fig 3E. And I would just cut lines 209-216, and go straight from "... Figure 3F)." to "This observation, which rests..." That seems easier than doing more writing at this point.

Alternatively the authors could briefly describe the facts in Figure S8 and Figure S9, which that the log-normal distribution in abundance is critical, as shown in the simulations, without giving the conceptually convoluted explanation involving lower relative abundance, which raises more questions than answers.

The y-axis labels in panels A and B of Figures S8 and S9 should be absolutely consistent. Indeed, it looks like the log-normal distribution prevents the simulated H populations from achieving high stability, due to a long tail of cells with very high abundance. This long tail gets scaled away in the relative abundance calculation, and I think this is why the authors' explanation is so convoluted (focus on relative abundance). My feeling is there is probably a much simpler explanation involving absolute abundance.

Regardless of these criticisms in presentation, I think this paper is very fine work.

signed,
Rohan Maddamsetti

Reviewer #3:
Remarks to the Author:
Thank you for your careful revisions. I have no further comments.

Author Rebuttal, first revision:

Response to reviewers

Reviewer #1

Reviewer #1: I want to thank the authors for their time and effort spent responding to my comments. In general, I found their response to be satisfactory and the manuscript has been substantially improved. I only have a couple of remaining minor comments listed below.

Response: Thank you very much for your comments. Addressing those comments helped made the narrative clear. We have dealt with both of your remaining concerns in this version of the manuscript.

Reviewer #1: I am surprised to learn that the convolution of log-normal distribution does not follow the Central Limit Theorem. The authors should cite the paper titled "Permanence of the log-normal distribution" at around line 186.

Response: Thank you for suggesting this citation. We have now added the citation to the results.

Lines 144-145 "*The 'meta-distribution' of these log-normal distributions is also log-normally distributed [27](Supplementary Information, Figure S13).*"

Reviewer #1: The suspected reason for the difference in effect of D113N between H and L populations should be discussed somewhere (Result or Discussion).

Response: Thank you for the suggestion. As discussed in our previous response to your comments, a mutant protein with a single mutation does not increase cyan fluorescence to similar extent when expressed in H and L genetic backgrounds. The D133N mutation is no exception. It increases cyan fluorescence ~5 folds in the L but not in the H genetic background. We now mention this intriguing observation along with possible reasons in the results.

Lines 105-110 "*The increase in cyan fluorescence was not the same when a given variant was expressed in the H or L genetic background (Table S3). We suspect that the effects of expression*

level are not restricted to the reproductive fitness of an individual carrying the variant, as recently demonstrated25, but also extend to the fluorescence output for some variants. Despite these differences in the extent of the fluorescent increase, all nine variants exceptD133N in the H background increased the intensity of cyan fluorescence (Table S3)."

Reviewer #2

Reviewer #2: Overall the quality of this work is very high and the authors did a good job addressing the reviewers' criticisms. However, I still find Figure 3G and the explanation in the text of how the log-normal distribution matters very unsatisfactory.

Response: Thank you very much for your concern. Please find below the point-by-point reply to your specific comments

Reviewer #2: line 35: it should be clarified that misfolding toxicity -> ER-anticorrelation is a hypothesis, rather than an established fact. There are several other hypotheses for the ER-anticorrelation (as discussed in the Nat Rev Genetics paper cited here), and its specific causes are still debated. Since the misfolding hypothesis is brought up later, I would either move the citations to the previous sentence and strike out this statement that misfolding toxicity is THE cause, or I would qualify that this has been hypothesized as the cause for the ER-anticorrelation. I would probably take the first course, since ER-anticorrelation is not the point of this paper.

Response: We completely agree that misfolding toxicity is only one of the hypotheses that may explain the ER-anticorrelation. As you point out, misfolding toxicity is not the point of our paper and hence we have removed the sentence pertaining to misfolding toxicity from the introduction altogether.

Lines 34-35 "*Specifically, sequences of highly expressed proteins evolve slowly*[4,10,11,12]*.*"

Reviewer #2: line 323: I think the key difference is that this work is about positive directional selection, and that the misfolding -avoidance hypothesis is about negative purifying selection.

Timescales matter too, but I think this is secondary, and the recent Wu et al. (2022) paper "Expression level is a major modifier of the fitness landscape of a protein coding gene" in Nat Eco Evo, along with my 2021 GBE paper, suggest that the timescale difference is much less important than has been previously believed.

Response: Thank you for pointing out the recent advance in the understanding of the role of timescales. We agree with you that the role of positive directional selection is likely to be dominant. However, as reviewer #3 mentions, the timescale is also crucial and we would like to retain it as per his recommendation.

Reviewer #2: line 370: I think this should rather be something like "The second requirement is for a heavy-tailed distribution of protein abundance per cell, which is known to be caused by gene expression noise". My reasoning is that there may be other sources of stochasticity in the cell that could cause the heavy tailed abundance distribution, and Fig S9 shows that gene expression noise in itself is not sufficient; it has to be a heavy-tailed distribution. I initially misread this as expression noise is the key factor, when the reasoning here is a bit more subtle.

Response: Thank you very much for this suggestion. We have now modified the lines as per your suggestion.

Limes 317-318 "*The second requirement is a heavy-tailed distribution of protein abundance per cell, which is known to be caused by gene expression noise*[1,2,39]."

Reviewer #2: Figure 3G: At first glance this looks great. But when I tried to understand it, Figure 3G is super confusing. There are three pieces here: 1) a density plot of protein abundance per cell in the two populations 2) the isoclines for selection for fluorescence in each population (like panel C) 3) a density plot of fluorescence per cell, with thresholds for selection (like panel B). The 3rd piece does not even exist, but it looks like it does, given the confusing y-axis labels.

The main issue is that the axis labels for the density plots are missing. The x-,y-,z- axis refer to the 3D isocline. The y-axis for the density plot is "Density", and the x-axis is relative abundance. The vertical

dashed lines by eye seem to be selective thresholds for the fluorescence, based on how they line up with the isoclines. But really, these are 10% relative abundance lines. But the 10% relative abundance line seems more or less irrelevant since it doesn't map nicely to the actual selection pressures. I would actually also cut the vertical dashed lines in D & E for that reason. There are selected cells below the this "threshold" in both panels D & E, and there are unselected cells above this threshold in panel E, which demonstrates how it's not directly relevant to the selection.

Finally the logic about the leftward-shift relative abundance in the high expression populations is super confusing, because selection in this system doesn't care about relative protein abundance per cell in itself. Lower relative stability does not necessarily evolve from having lower relative abundance, especially since the point of this paper is that the populations with lower absolute abundance evolve higher absolute stability! I think the simplest fix at this stage would be to better just to completely cut Figure 3G, and eliminate the dashed vertical lines in Fig 3D and Fig 3E. And I would just cut lines 209-216, and go straight from "... Figure 3F)." to "This observation, which rests..." That seems easier than doing more writing at this point.

Alternatively the authors could briefly describe the facts in Figure S8 and Figure S9, which that the log-normal distribution in abundance is critical, as shown in the simulations, without giving the conceptually convoluted explanation involving lower relative abundance, which raises more questions than answers.

Response: Thank you very much for your concern. We agree that figure 3G was difficult to digest. In response to your comment, we have thus removed figure 3G and added a supplementary figure (S14) which shows the main differences between the H and L populations. Importantly, in this figure we addressed the concerns related to the axis labels, and separated the figure into three panels with distinct labels (panel A= stability factor vs. abundance, panel B: the distribution of protein stabilities, panel C= distribution of protein abundances). We have also removed the vertical dashed lines from Figures 3D and 3E

In addition, we have also revised the paragraph in our main text where we explained the main difference between the H and L populations (lines 209-216 in the previous version of the manuscript) and particularly removed the term "relative fluorescence" which was confusing. In the new version, we have focused on the range of expression level in selected cells. We now emphasize that because of log-normality the same percentile of selected (surviving) cells corresponds to a lower fluorescence intensity in the H population than in the L population. Because higher fluorescence is more likely to be associated destabilizing mutations (destabilizing mutations are more enriched in cells harbouring GFPs with lower stabilities), the stability of GFPs in H populations span a wider range than in L populations. Specifically, we now write:

Lines 169-176 "*The reason is that although we selected the same percentile of fluorescent cells in both populations, the range of expression levels in surviving cells (from minimum to maximum) is broader in H populations than in L populations. This is mainly caused by the log-normality of the abundance distribution, whose skewness increases with its standard deviation. GFP abundance in H population has a higher standard deviation and is more left-skewed. Therefore, and because maximum fluorescence is more likely achieved by GFPs with high stability, the higher range of fluorescence in the H populations corresponds to a wider range of stabilities (Figure S14). Indeed, when we used a normal distribution to model protein abundance, GFP stability did not decrease with higher mean abundance (Supplementary information, Figure S9)*."

Reviewer #2: The y-axis labels in panels A and B of Figures S8 and S9 should be absolutely consistent. Indeed, it looks like the log-normal distribution prevents the simulated H populations from achieving high stability, due to a long tail of cells with very high abundance. This long tail gets scaled away in the relative abundance calculation, and I think this is why the authors' explanation is so convoluted (focus on relative abundance). My feeling is there is probably a much simpler explanation involving absolute abundance.

Response: Thank you very much for pointing this out. We now use the exact same range for the y-axis in Figures S8 and S9.

Reviewer #2: Regardless of these criticisms in presentation, I think this paper is very fine work.

Response: Thank you, we are grateful for your positive comments and constructive suggestions.

Reviewer #3

Reviewer #3: Thank you for your careful revisions. I have no further comments.

Response: Thank you very much.

Final Decision Letter:

19th May 2022

Dear Shraddha,

We are pleased to inform you that your Article entitled "Low protein expression enhances phenotypic evolvability by intensifying selection on folding stability", has now been accepted for publication in Nature Ecology & Evolution.

Over the next few weeks, your paper will be copyedited to ensure that it conforms to Nature Ecology and Evolution style. Once your paper is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required

After the grant of rights is completed, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

You will not receive your proofs until the publishing agreement has been received through our system

Due to the importance of these deadlines, we ask you please us know now whether you will be difficult to contact over the next month. If this is the case, we ask you provide us with the contact information (email, phone and fax) of someone who will be able to check the proofs on your behalf, and who will be available to address any last-minute problems . Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies

(see www.nature.com/authors/policies/index.html). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our web site).

Please note that Nature Ecology & Evolution is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including <a href="https://www.nature.com/nature-portfolio/editorial-policies/self-archiving-and-license-to-publish. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

In approximately 10 business days you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

You will not receive your proofs until the publishing agreement has been received through our system.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Ecology & Evolution as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Ecology & Evolution logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

You can generate the link yourself when you receive your article DOI by entering it here: <a href="http://authors.springernature.com/share">http://authors.springernature.com/share<a>.

[REDACTED]

P.S. Click on the following link if you would like to recommend Nature Ecology & Evolution to your librarian http://www.nature.com/subscriptions/recommend.html#forms

** Visit the Springer Nature Editorial and Publishing website at www.springernature.com/editorial-and-publishing-jobs for more information about our career opportunities. If you have any questions please click here.**