nature portfolio

Peer Review File

Adaptive tracking with antagonistic pleiotropy results in seemingly neutral molecular evolution

Corresponding Author: Professor Jianzhi Zhang

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

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this study, the authors analyse previously generated deep mutational scan data to estimate the fitness effects of amino acid substitutions. They conclude that beneficial mutations occur far more frequently than previously thought, on the order of 1% of all non-synonymous substitutions. The authors point out that even though beneficial mutations are still vastly outnumbered by neutral or deleterious mutations, their fixation probability is high enough that beneficial mutations represent approximately 99% of all substitutions.

This study has exciting ideas, tackles ambitious questions and covers substantial ground, ranging from analyses of previous data, simulations, to experimental evolution with extensive sequencing. The paper is well written. The amount of work undertaken is impressive. I appreciate the space in the methods that they authors take to explain the caveats and history of these data sets.

The dataset – especially the 21 gene data set is impressive, however, the claims made by the authors are even bigger – “a key premise is violated” (abstract) and “the foundations of theory are shaken” - the level of evidence required to substantiate these claims needs to be very high. While the quality of the study is high and data is voluminous, my major concern is whether we can be sure enough about the abundance of beneficial mutations, that these strong claims are supported.

Comments:

1. The authors point out caveats of the choice of 21 genes, but don’t mention (at least in my reading) that by omitting essential genes, the experiment avoids a huge class of mutations that would be lethal or profoundly deleterious (for example, any nonsense mutation in an essential gene would be non-viable). Excluding them raises the proportion of surviving mutations that could appear “beneficial.” In other words, there are biases built into the study that lead to beneficial mutations being over-represented. This might be unavoidable, but needs to be clearly stated, and due to the strength of the claims of this study, the authors need to explain why this limitation does not lead to the study overestimating beneficial mutations.

2. A really important part of estimating the beneficial potion of the DFE is the cut-off between neutral and beneficial. The vast majority of the DFE is made up of neutral and deleterious mutations, with very few with a substantial benefit and majority of beneficial mutations have tiny positive effects that hover around the detection limit. Slight changes in threshold or statistical stringency can flip those from beneficial to neutral. If the authors had been slightly more stringent (say, FDR 1% or a bigger effect size), the fraction of beneficial mutations could have been compatible with neutral theory. Sometimes a lenient threshold is used in variant classification to avoid missing interesting variants. In this study the danger is not missing interesting beneficial mutations, but overestimating the absolute number of beneficial mutations, so I would like the authors to explore using a more stringent cutoff.

3. The authors report the median beneficial effect, but it appears in Figure 1A that most beneficial effects in the distribution have less than 1% effect. Given the highly skewed nature of the beneficial mutation distribution, the authors should report addition information about the beneficial mutations, for example a zoomed in figure of the actual measurements that includes an indication of the spread for individual measurements for beneficial mutations. The Pearsons R of 0.9 is great, but this amount of variance would be enough to make the estimated fitness effect of very small effect mutations overlap with zero.

When this is published, the reader needs to have a sense of the individual measurements for each beneficial mutation.

4. As the number of beneficial mutations increases, it becomes more likely that lineages will carry multiple mutations—beneficial, neutral, and mildly deleterious hitchhikers—which collectively increase fitness variance in the population and subsequently decrease Ne. Have the authors considered the impact of variance on Ne on calculated omega values?

5. Under the adaptive tracking hypothesis, a switch in environment will result in a sudden large number of deleterious mutations segregating at non-trivial frequencies in the population – this sudden increase in genetic load could be a large burden on the population, and I would like to hear the authors have considered the implications of frequent environmental changes on genetic load.

6. In exploring the AdapTrack model, the authors varied parameters and observed that reductions in Ne and weakened selection shift the allele frequency spectrum, increasing the fractions of deleterious polymorphisms and substitutions, consistent with the nearly neutral theory. Have the authors manipulated effective population size in the adaptive model to observe changes in the beneficial substitution rate?

7. The sequencing effort is impressive, with many clones sequenced and mutations identified. However, calling single-clone-identified SNPs "fixed" is not sufficiently robust, and based on the sequencing data from experiments of similar length in the Sherlock and Desai labs, there are likely to be many low frequency neutral and deleterious “passenger” mutations counted among the fixed. I would like the authors to explain why they didn’t sequence two clones per population, or whole populations, so that they get a better idea of the actual fixed mutations, and/or how their analysis accounts for this.

8. The method for estimating beneficial mutations from yeast evolution experiments is unconventional. Typically, a conservative null model estimates the likelihood of genes being hit multiple times by chance, and then the observed results are compared to the expected. The authors should explain their null model calculation more explicitly. Line 300 refers readers to the methods for details, suggesting standard practice, but additional clarity is needed. Specifically, does a mutation count as a multi-hit if it recurs anywhere in the experiment or only within the same treatment? Given around 10,000 nonsynonymous mutations, many genes would naturally be expected to be hit twice by chance. I would specifically like the authors to show model support for the idea that two hits constitutes parallel evolution – this seems low.

9. The authors observe lower fb values in randomly changing environments compared to constant ones. While interesting, this result aligns with theoretical expectations—natural selection is less efficient in fluctuating environments. Although interesting and useful, this finding is somewhat orthogonal to the study's primary assertion against the neutral theory. My point is, the slower adaptation to fluctuating environments can be true without the violation of neutral theory. Would the authors agree? How does this evolution experiment explicitly test the key idea that there are more beneficial mutations around than we thought, and that this violates neutral theory?

10. On Line 349, the authors state: "The key prediction of our model that frequent environmental changes amplify the proportion of neutral substitutions by hindering the fixation of beneficial mutations is strongly supported by our yeast experimental evolution." However, “neutrality” is not directly addressed in the results of the evolution experiment ie the data are not able to show whether beneficial mutations become neutral or deleterious in changing environments. It seems more accurate to say frequent environmental changes prevent beneficial mutations from fixing, or that frequent environmental change limits natural selection’s effectiveness in increasing beneficial mutation frequencies.

11. The authors write on line 356 “While our model is a special form of fluctuating selection, prior studies of fluctuating selection did not notice nor reconcile the drastic difference between patterns of short-term and long-term molecular evolution.” Can the authors be more specific about their innovation – the Desai group paper (among the citations) was certainly different, but did mention long and short-term evolution.

Minor

Line 1046 includes a potential typo or unclear phrasing: "Original, originally..." This needs clarification or correction.

Can the authors please check whether the table s2 is in agreement with Johnson et al? Specifically the counts for each “strain” with improved fitness compared to the ancestor. Also, I think it should be populations, not strains. Also datasets S3 - S5 need an explanation in the form of a caption, and some kind of reference to the main text.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
In this manuscript, Song and colleagues bring together an extensive array of data (both reanalyzed and newly generated), theory, and simulations to support a model of adaptive tracking with antagonistic pleiotropy. Despite being mechanistically very different from standard ‘neutral’ models of evolution, it produces similar patterns (notably, a substantial proportion of neutral or nearly neutral substitutions) over relatively long time scales, concealing rapid adaptive evolution which fluctuates over shorter time scales, with changing environments.

While this model is compelling, the claims of novelty are overstated and often hyperbolic. The first Discussion sentence claims that it 'shakes the very foundation of the theory' but the following sentence more modestly states that this is not necessarily surprising. The neutral theory is portrayed as a bit of a straw man, when in reality the current understanding of drift-selection balance is more nuanced. The idea of adaptive tracking is not novel in itself, having been developed in reference 42, along with previous work by two of the authors (Chen & Zhang, NEE 2020), and supported by empirical work by Rudman (Science, 2022) among others. I do think that the key contribution in this work is the recognition that adaptive tracking creates patterns of neutral substitution that are more similar to neutral models, explaining the molecular clock better than a standard directional selection model. Even this realization is related to recent work by Torrillo & Lieberman (eLife 2024) describing an 'adaptive reversion' model to explain the time-dependence of estimated dN/dS values in microbial genomes. All in all, the work by Song et al. provides a useful synthesis of the available data, but the key novel contributions are sometimes hard to discern amid the abundant (sometimes excessive) analyses.

In many ways, the manuscript reads like a review or synthesis paper, alternating between background, discussion, and data. This is, at times, difficult to follow, with extended discussion/synthesis paragraphs in the Results (e.g. lines 113-133) and new results brought into the Discussion. Aside from these structural difficulties, the relevance of the data and experiments is not always clear.

First, the deep mutational scanning analyses (Fig 1) serve to refute the neutral theory, but it is not clear why the neutral theory needs refuting in the context of this manuscript. Moreover, this analysis raises many questions. The estimates of omega are extremely high (>1e4) – much higher than typically observed in comparative genomic studies, and certainly higher than the values obtained in SLiM simulations (Fig 2) and from experimental evolution (Fig 5). Is this because DMS provides a near-instantaneous fitness measurement? Or is it biased toward large-effect mutations? The alpha values are also very high, perhaps for these reasons. Moreover, although I appreciate the effort taken to bring together 4 different datasets, it is not at all clear if these few genes are representative of genome-wide patterns.

Second, the relevance of the evolution experiment is not clear. Although it is impressive work, it is not clear why the data from the 2020 experiment with similar design (Chen & Zhang, NEE 2020) could not have been used. The original 2020 experiment supports most predictions of the adaptive tracking hypothesis: the main result of figure 5 – that dN/dS (small omega) tends to be smaller (~0.8-1.8) in changing than constant (~1.5-2) environments is very similar to the result already reported in Chen & Zhang (Nature Ecology & Evolution, 2020). Certainly the new experiments add some subtlety and of course it's always good to replicate experiments, but they do not seem to address a major unsolved question.

Overall, this is a very interesting paper, but it would be more palatable with some major structural changes. Consider structuring it more like a review/synthesis paper, and perhaps removing some of the data.

Minor:

1. In the section 'Estimating omega and alpha from empirical DEs' (lines 496 onward), do the calculations setting a value for Ne? How is Ne set here?

2. In the section 'Estimation of the fraction of beneficial substitutions' (lines 749 onward), the authors state that at least two of the 12 replicates must have a substitution in the same gene for it to be counted. Presumably they mean substitutions relative to the ancestor, not to the reference genomes – because the reference genome used was not the same as the ancestor (if I understood earlier sections correctly). Please clarify.

3. Fig 5 refers to 'corresponding' constant environments but it is not clear how each changing environment is matched to a corresponding constant. It also seems from the method that there was only one changing environment, replicated 12 times, which was compared to 12 different constant environments. I think this explains why there is a boxplot (blue points in Fig 5b) for the constant environments, but a single red line for the changing environment. Was the t-test really done by comparing these 12 blue points to a single 'red' value? If so, this seems somewhat underpowered and maybe not the ideal statistical test. In any case, the general direction of the effect is clear from panels b, c and d combined.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
This is an interesting, thought-provoking, and well-written paper that investigates a model of fluctuating selection that could explain patterns of molecular evolution. I recommend the paper for publication, and of only a few comments. I think that's paper can be published more or less as it is.

At line 52, the neutral theory is no longer generally accepted, as there has been a great deal of evidence for adaptive molecular evolution.
See for example Smith and Eyre-Walker Nature 415(6875): 1022-4.

From line 71 onwards, estimates of the DFE are for very artificial environments. We don't really know what the DFE will look like in the natural environment. This requires more discussion.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
Dear Authors, i was asked to review this manuscript that I had already seen. I mentioned that to the editor that told me it was not an issue. So a lot of my comments will match my previous review as they were mostly commentaries on the tone and background of the paper, and some still hold true in the present manuscript, I omitted the ones that were answered in the current version (thanks).

In the manuscript entitled "Adaptive Tracking results in seemingly neutral molecular evolution" Song et al present a framework in which large fluctuation in the environment affect the effects of beneficial mutations that are much more frequent that a neutral model assumes. The models builds on observations that deep mutational scans as well as some mutation accumulation have revealed much larger than expected fraction of beneficial mutation which integrated over long evolutionary time scale would result in extremely high fraction of non-synonymous mutation compared the accumulation of synonymous mutations. This rapid adaptation in a given environment is also comforted by experimental evolution in which fast adaptation is observed but often recruits mutations that are at odds with the level of conservation observed in the targeted genes. The model proposed suggests that antagonistic pleiotropy may be at play and that mutations that are beneficial are so only locally and transiently and that they therefore do not reach fixation. To support that model, they build on simple simulations in which neutrals mutations keep being neutral but a fraction of mutations may switch randomly between beneficial and deleterious at pace that range from a few generations to few hundreds of generations. These simulations confirm obviously the fact that fixation of beneficial mutations in that context is marginal and the long-term pattern is overwhelmingly dominated by neutral mutations. An experimental validation is made through a large-scale experimental evolution involving stable of alternating environment over 800 generations of yeast evolution in the laboratory. Here, sampling individual clones after that evolution they confirm that there are less beneficial mutations and more neutrals in the fluctuating environment.

Besides a few technical comments I will present later, the article is clear and quite straight forward. On the positive side, it is timely , as data showing the intensity of adaptation and the extent of beneficial mutations are accumulating and contrast with macro evolutionary patterns. Yet, digging into the literature, the whole idea is also quite known in the field of evolution for a long time: antagonistic pleiotropy, G by E, standing genetic variation, fluctuation environment largely dealt with that concept so one may question the novelty of the approach. To support the novelty, I would say that it is simple here and all previous concepts have been quite widely used over the short term but very few from what I have seen have integrated the process to genomic patterns, so taht perspective is new.

I think however that a lot of statement about novelty should be downplayed a little to incorporate more the previous work:

In essence some models have been used to study environmental fluctuation, which usually is based on an additive effect of mutations on a trait connected to fitness, but a change in the optimal value of that trait. Here, the model is mutation centered and assumes shifts in effects but these are indeed equivalent, so connection with these models would be welcome. For instance, Graham Bell in “Fluctuating selection: the perpetual renewal of adaptation in variable environments » acknowledged that mild effect mutations could not easily rich fixation but that large effects ones could, somehow going beyond the present model in which large effect mutation would be mutations with fixation time faster than the environmental switching time. Similarly, Grossman et al in “Fluctuating Selection Models and Mcdonald-Kreitman Type Analyses” (PlosOne 2014) discussed how selection fluctuating impacts fixation, and identified the type of mutations that will in fine be fixed. Another case in, which the consequence of fluctuating s in the environment can be found in A. Desbiez-Piat et al “Interplay between extreme drift and selection intensities favors the fixation of beneficial mutations in selfing maize populations” (Genetics 2021), in which the high fluctuation in selection through time in plant is discussed for its consequences on fixation. Another related work is “Random fluctuation of selection coefficients and the extent of nucleotide variation in human populations” by Sayaka Miura, Zhenguo Zhang, and Masatoshi Nei (PNAS 2014) which starts with “It is well known that the selection coefficient of a mutant allele varies from generation to generation, and the effect of this factor on genetic variation has been studied by many theoreticians.” And finishes with “This result implies that there is no need of distinguishing between the diversity-enhancing and diversity-reducing models of fluctuating selection and the nucleotide polymorphism in human populations can be explained largely by neutral mutations when long-term evolution is considered.” So I think a full paragraph stating simply that the present work builds on these past work but focusses on the molecular evolution dimension would really be worth.

Finally, experimentally, fluctuating selection has been shown to limit the fixation of mutation, as seen in yeast recently by F. Abdul-Rahman et al “Fluctuating Environments Maintain Genetic Diversity through Neutral Fitness Effects and Balancing Selection ». So even if each of these works tackle the question differently the idea proposed here could be integrated in these framework. cannot be said to be new.

There is also a small papers of ours (Vigue Tenaillon PNAS 2023) that may match this various time scale of selection hypothesis and may bring some ideas: using 60000 genomes of E. coli and some inferred effects of mutation, we detected that some genes showed excess deleterious mutations even though there are core genes, suggesting that at local scale selection may favor lower functioning of these genes, that are conserved otherwise. This also raise the question of

heterogeneity across space and time, can the patterns of environmental variation observed here may be relevant at the species level to drive signals that will impact fixation rate…

Concerning the model: many layers of the potential complexity have been taken into account. I am globally in line with the modelling performed and appreciate its simplicity. The question though is over which time scale is the measure of s appropriate. For very short switching time, s over one generation is not appropriate but its average over time. Time should be counted as time needed to substantially change in frequency. Then the proposed model converge to a neutral model as the mean s of mutations is negative. Indeed, for both microbes and even more for higher organisms, the environment may fluctuate through a life time, so s is already an integration over some time, to what extent integrating over longer time scale is that different. Here the time scale of changes of the environment is somehow set to exactly match 1/s. Which allow substantial changes in allele frequency. The question is how likely is that to happen: if changes are slower, the adaptive model would predominate, if changes are faster, beneficial alleles will barely increase in frequency. The importance of this balance should be mentioned.

I think the consequence of the model on polymorphisms could be exploited further and more than qualitatively as done in figure 3. For instance, contrasting the site frequency spectrum with one including just neutral or background selection would be valuable and offer potential ways to test if these scenarios can be inferred from data. One key component of classical models is the probability of surviving drift, if burst of selection lead to rapid increase of selection this may affect this probability. Moreover, rapid decrease of frequency upon environment shifts will lead to profound reduction of effective population size that may also affect the probability of surviving drift.

Finally, for Figure 5, it is hard to know from the data how many mutations are used to infer omega, in other words, how robust are the estimates of omega. With low numbers noise may be high.

Overall, it is a simple and timely story, that should help many to value the different scales at which evolution acts, but to me it could still be more anchored in the existing literature, to avoid harsh criticism from people in the field of evolution.

Olivier Tenaillon (I always sign my reviews)

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I have reviewed the authors response. I appreciate the time they have taken to address my comments, and those of the other reveiwers, and especially to introduce parts of the response into the manuscript text and figures. All of my comments have been addressed, and I have no further questions. I congratulate the authors on a excellent piece of work.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
Thank you for the revisions and the detailed responses to my comments. All my substantial comments have been adequately addressed. Congratulations on this exciting work!

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I appreciate the fact that the author now cite a broader part of the literature from the field and have down toned a bit the novelty aspect fo the work. The focus on the molecular evolution consequence of fluctuation environment is a worth perspective.

The last sentence of for sure an opinion: "Hence, the adaptive tracking model may be even more important for multicellular than unicellular organisms". It could also be the exact opposite. As multicellular organisms have longer life span the fitness effect of mutation is already an average across a multitude of environment that an organism face during a life cycle and across a multitude of tissues, and as such is less likely to be beneficial contrary to microbial life and assays that can detect benefit in a very particular local environment. Si in my opinion the adaptive tracking may by less likely to happen for

multicellular. The question remains open.

(Remarks on code availability)

Response to reviewers

We thank the four reviewers for their valuable comments, which have helped improve the clarity and rigor of our manuscript. Below please find our point-to-point response in blue.

Reviewer #1

In this study, the authors analyse previously generated deep mutational scan data to estimate the fitness effects of amino acid substitutions. They conclude that beneficial mutations occur far more frequently than previously thought, on the order of 1% of all non-synonymous substitutions. The authors point out that even though beneficial mutations are still vastly outnumbered by neutral or deleterious mutations, their fixation probability is high enough that beneficial mutations represent approximately 99% of all substitutions.

This study has exciting ideas, tackles ambitious questions and covers substantial ground, ranging from analyses of previous data, simulations, to experimental evolution with extensive sequencing. The paper is well written. The amount of work undertaken is impressive. I appreciate the space in the methods that the authors take to explain the caveats and history of these data sets.

The dataset – especially the 21 gene data set is impressive, however, the claims made by the authors are even bigger – "a key premise is violated" (abstract) and "the foundations of theory are shaken" - the level of evidence required to substantiate these claims needs to be very high. While the quality of the study is high and data is voluminous, my major concern is whether we can be sure enough about the abundance of beneficial mutations, that these strong claims are supported.

We thank the reviewer's overall positive evaluation of our manuscript and address the reviewer's comments below.

Comments:

1. The authors point out caveats of the choice of 21 genes, but don't mention (at least in my reading) that by omitting essential genes, the experiment avoids a huge class of mutations that would be lethal or profoundly deleterious (for example, any nonsense mutation in an essential gene would be non-viable). Excluding them raises the proportion of surviving mutations that could appear "beneficial." In other words, there are biases built into the study that lead to beneficial mutations being over-represented. This might be unavoidable, but needs to be clearly stated, and due to the strength of the claims of this study, the authors need to explain why this limitation does not lead to the study overestimating beneficial mutations.

We agree that a potential bias exists because only nonessential genes were included in the genes analyzed. However, because <20% of genes are essential in both *E. coli* and yeast (Baba et al. 2006 *Mol Syst Biol*; Winzeler et al. 1999 *Science*), even if all nonsynonymous mutations in essential genes are selectively purged due to strong deleterious effects, the overall Ω of all genes would still be >0.8 times the Ω estimated from nonessential genes, which would still exceed 1

substantially. If a fraction of nonsynonymous mutations in essential genes can be fixed, Ω of all genes would be even greater.

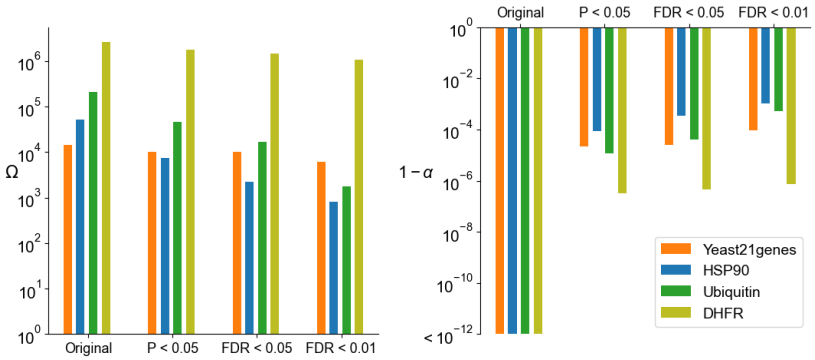
As for α, if we assume that there are no beneficial mutations in essential genes and that the deleterious portion of essential genes' DFE is the same as that of nonessential genes, one can derive α of all genes $\alpha_{all} > 4\alpha_{est}/(5-\alpha_{est})$, where α_{est} is the α estimated from nonessential genes. Because α_{est} is very close to 1, let us write $\alpha_{est} = 1-d$, where $0 < d << 1$. It can be shown that $\alpha_{all} > 1-5d/(4+d) > \alpha_{est} - 0.25d$. Hence, α_{all} is only slightly smaller than α_{est}. If some beneficial nonsynonymous mutations exist in essential genes and/or deleterious nonsynonymous mutations are more deleterious when occurring in essential than nonessential genes, α_{all} would be larger than the above estimate.

In other words, the estimation biases of Ω and α due to the exclusion of essential genes in the DFE data is minimal. We have added the above results to the manuscript (lines 507-520).

Importantly, in our manuscript, we applied even stricter criteria in the analysis of the DFEs by (1) considering only significantly beneficial mutations as truly beneficial, (2) assuming a 10-fold lower beneficial mutation rate than that observed, or (1) and (2) simultaneously (see Fig. 1bc: Significant, Original/10, and Significant/10). Under all of these conditions, Ω remains well above 1 and α remains very close to 1. Therefore, we believe that our conclusions regarding Ω and α are robust.

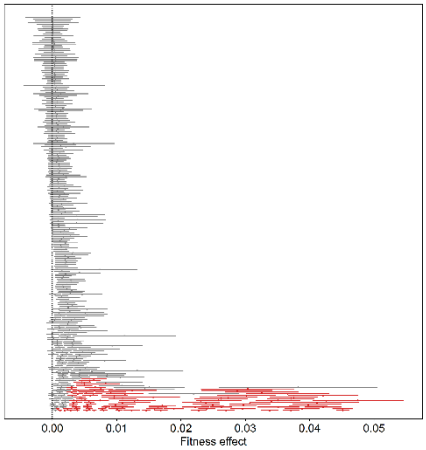
2. A really important part of estimating the beneficial potion of the DFE is the cut-off between neutral and beneficial. The vast majority of the DFE is made up of neutral and deleterious mutations, with very few with a substantial benefit and majority of beneficial mutations have tiny positive effects that hover around the detection limit. Slight changes in threshold or statistical stringency can flip those from beneficial to neutral. If the authors had been slightly more stringent (say, FDR 1% or a bigger effect size), the fraction of beneficial mutations could have been compatible with neutral theory. Sometimes a lenient threshold is used in variant classification to avoid missing interesting variants. In this study the danger is not missing interesting beneficial mutations, but overestimating the absolute number of beneficial mutations, so I would like the authors to explore using a more stringent cutoff.

This is an insightful comment. We have now applied three statistical significance cutoffs in calling beneficial mutations: nominal $P < 0.05$, FDR < 0.05, and FDR < 0.01. They yielded consistent results (see the newly added Fig. S2 shown below). We have added this finding to the manuscript (lines 76-78, 91-92, and 552-554).



3. The authors report the median beneficial effect, but it appears in Figure 1A that most beneficial effects in the distribution have less than 1% effect. Given the highly skewed nature of the beneficial mutation distribution, the authors should report addition information about the beneficial mutations, for example a zoomed in figure of the actual measurements that includes an indication of the spread for individual measurements for beneficial mutations. The Pearsons R of 0.9 is great, but this amount of variance would be enough to make the estimated fitness effect of very small effect mutations overlap with zero. When this is published, the reader needs to have a sense of the individual measurements for each beneficial mutation.

In Fig. 1a, many beneficial mutations in the DFE are not statistically significant. Following the reviewer's suggestion, we have added a zoomed-in graph in Fig. S1 (shown on the right) to present individual fitness effects of beneficial nonsynonymous mutations. Each dot shows the point estimate of the fitness effect of a mutation, with its standard error indicated by the error bar. Red indicates a significant fitness effect (nominal $P < 0.05$), whereas grey indicates a non-significant fitness effect.



4. As the number of beneficial mutations increases, it becomes more likely that lineages will carry multiple mutations—beneficial, neutral, and mildly deleterious hitchhikers—which collectively increase fitness variance in the population and subsequently decrease Ne. Have the authors considered the impact of variance on Ne on calculated omega values?

In the empirical DFE analysis, we primarily used $N_e = 10^7$, which is based on the empirically estimated long-term N_e from yeast. Using this empirically estimated long-term N_e means that the effect mentioned by the reviewer has been taken into account. In addition, we investigated a variety of N_e, and the results were similar qualitatively even when N_e was as small as 10^4 (Fig. S3cd). We also considered clonal interference and obtained qualitatively similar results (Fig. S3ab).

In the SLiM simulation, N_e was used to define the number of individuals in the Wright-Fisher model, so the decrease of N_e caused by a high fitness variance was inherently accounted for in the simulation. Additionally, we evaluated the impact of N_e on Ω in Fig. 2e-h, which showed that even when N_e is as small as 500, Ω is still much lower than 1 under the adaptive tracking model. Therefore, the decrease in N_e caused by high fitness variance should not affect our general conclusion.

5. Under the adaptive tracking hypothesis, a switch in environment will result in a sudden large number of deleterious mutations segregating at non-trivial frequencies in the population – this sudden increase in genetic load could be a large burden on the population, and I would like to hear the authors have considered the implications of frequent environmental changes on genetic load.

The reviewer is right that environmental changes can create a high genetic load. The exact size of the genetic load and its consequences await future studies. We have briefly discussed this point in lines 420-423.

6. In exploring the AdapTrack model, the authors varied parameters and observed that reductions in Ne and weakened selection shift the allele frequency spectrum, increasing the fractions of deleterious polymorphisms and substitutions, consistent with the nearly neutral theory. Have the authors manipulated effective population size in the adaptive model to observe changes in the beneficial substitution rate?

Yes. We investigated the influence of N_e on the substitution rate under the *Adaptive* model in Fig. 3c, showing that a larger N_e corresponds to a higher substitution rate.

7. The sequencing effort is impressive, with many clones sequenced and mutations identified. However, calling single-clone-identified SNPs "fixed" is not sufficiently robust, and based on the sequencing data from experiments of similar length in the Sherlock and Desai labs, there are likely to be many low frequency neutral and deleterious “passenger” mutations counted among the fixed. I would like the authors to explain why they didn’t sequence two clones per population, or whole populations, so that they get a better idea of the actual fixed mutations, and/or how their analysis accounts for this.

The reviewer is right that we used only a single clone to identify “substitutions”, which may contain low-frequency neutral or deleterious mutations. The sequencing data from constant-environment evolution experiments were obtained from a recent study (Chen and Zhang 2024 *Nat Commun*) where a single clone per population was sequenced due to the prohibitively high cost of sequencing over 3000 populations. To allow comparing with these data, we decided to similarly sequence one clone per population from changing-environment evolution experiments.

Because some of the identified “substitutions” are actually unfixed and because “beneficial substitutions” should on average have lower frequencies in changing environments (due to antagonistic selections) than in constant environments, treating all identified beneficial alleles as beneficial substitutions likely overestimates beneficial substitutions more in changing than in constant environments. Hence, our finding of a lower fraction of beneficial substitutions in changing environments than in constant environments (**Fig. 5bc**) is conservative; that is, the actual difference is expected to be even bigger than shown in Fig. 5bc. Similarly, treating unfixed mutations as “substitutions” will overestimate ω in changing environments more than in constant environments. Hence, our finding of a lower ω in changing environments than in constant environments (**Fig. 5d**) is conservative. In addition, in an earlier, much smaller experiment evolution study of constant and changing environments, we sequenced populations and obtained qualitatively similar results (Chen and Zhang 2020 *Nat Ecol Evol*).

We have now discussed the above points in Methods (lines 820-838) and added a note in lines 292-293.

8. The method for estimating beneficial mutations from yeast evolution experiments is unconventional. Typically, a conservative null model estimates the likelihood of genes being hit

multiple times by chance, and then the observed results are compared to the expected. The authors should explain their null model calculation more explicitly. Line 300 refers readers to the methods for details, suggesting standard practice, but additional clarity is needed. Specifically, does a mutation count as a multi-hit if it recurs anywhere in the experiment or only within the same treatment? Given around 10,000 nonsynonymous mutations, many genes would naturally be expected to be hit twice by chance. I would specifically like the authors to show model support for the idea that two hits constitutes parallel evolution – this seems low.

A mutation was considered beneficial if the mutated gene was hit at least twice among the 12 replicates of the treatment. This is mentioned in lines 786-797. We have now clarified it in the main text (lines 300-304).

Although there are around 10,000 "substitutions" in total in the data we generated, the number of substitutions per replicate population is on the order of 10, two to three orders of magnitude smaller than the number of genes in the yeast genome. Hence, it is unlikely that a gene is hit by a substitution in even two of the 12 replicate populations without positive selection.

We further quantitatively validated the approach of treating substitutions in multi-hit genes as beneficial substitutions by simulating a null model in which the same number of substitutions occurred randomly. The simulation showed that only 10% of identified beneficial substitutions are false positives, confirming that the majority are indeed beneficial. The detailed method was described in lines 793-797. Additionally, we previously designed a probabilistic method and analyzed the data from the constant-environment evolution experiment using the probabilistic method (Chen and Zhang 2024 *Nat Commun*). The result obtained by the conventional method is similar to that based on the probabilistic method.

9. The authors observe lower fb values in randomly changing environments compared to constant ones. While interesting, this result aligns with theoretical expectations—natural selection is less efficient in fluctuating environments. Although interesting and useful, this finding is somewhat orthogonal to the study's primary assertion against the neutral theory. My point is, the slower adaptation to fluctuating environments can be true without the violation of neutral theory. Would the authors agree? How does this evolution experiment explicitly test the key idea that there are more beneficial mutations around than we thought, and that this violates neutral theory?

The subheading of the experimental evolution section is "Testing the adaptive tracking model using experimental evolution". The purpose of the evolution experiment was not to test the neutral theory, because it had already been rejected by the abundance of beneficial mutations observed in empirical DFE data (Fig. 1). Rather, the experiment was designed to test the prediction of the adaptive tracking model that evolution in changing environments leads to a lower fraction of beneficial substitutions and a lower ω relative to evolution in corresponding constant environments.

10. On Line 349, the authors state: "The key prediction of our model that frequent environmental changes amplify the proportion of neutral substitutions by hindering the fixation of beneficial mutations is strongly supported by our yeast experimental evolution." However, "neutrality" is

not directly addressed in the results of the evolution experiment ie the data are not able to show whether beneficial mutations become neutral or deleterious in changing environments. It seems more accurate to say frequent environmental changes prevent beneficial mutations from fixing, or that frequent environmental change limits natural selection's effectiveness in increasing beneficial mutation frequencies.

We agree and have modified the sentence accordingly. It now reads "The key prediction of our model that frequent environmental changes prevent the fixation of beneficial mutations is strongly supported by our yeast experimental evolution." (lines 353-355)

11. The authors write on line 356 "While our model is a special form of fluctuating selection, prior studies of fluctuating selection did not notice nor reconcile the drastic difference between patterns of short-term and long-term molecular evolution." Can the authors be more specific about their innovation – the Desai group paper (among the citations) was certainly different, but did mention long and short-term evolution.

Our manuscript not only mentioned differences between long and short-term evolution, but more importantly, it reconciled the abundance of beneficial mutations in short-term evolution with the low ω observed in long-term evolution. We have modified the sentence to clarify the innovation of our model. It now reads, "While our model is a special form of fluctuating selection, prior studies of fluctuating selection did not reconcile nor recognize the contradiction between the abundance of beneficial mutations in short-term evolution and the low ω observed in long-term evolution." (lines 359-362)

Minor

Line 1046 includes a potential typo or unclear phrasing: "Original, originally..." This needs clarification or correction.

This is not a typo but presents an explanation for the definition of "Original" in Fig. 1bc.

Can the authors please check whether the table s2 is in agreement with Johnson et al? Specifically the counts for each "strain" with improved fitness compared to the ancestor. Also, I think it should be populations, not strains. Also datasets S3 - S5 need an explanation in the form of a caption, and some kind of reference to the main text.

We have now used "population" instead of "strain" in Data S2. We double checked the counts in Data S2 and have corrected some minor errors. Note that the values presented are numbers of populations that had significantly higher fitness at the end of the evolution experiment (generation 10,150) than at the previous time point. We only included non-contaminated populations in our analysis and performed *t*-test using selection coefficients estimated from the two replicates of each time point.

We have now added captions to Data S3-S5. The captions of all datasets are also shown at the end of Supplementary Materials.

Reviewer #2

In this manuscript, Song and colleagues bring together an extensive array of data (both reanalyzed and newly generated), theory, and simulations to support a model of adaptive tracking with antagonistic pleiotropy. Despite being mechanistically very different from standard 'neutral' models of evolution, it produces similar patterns (notably, a substantial proportion of neutral or nearly neutral substitutions) over relatively long time scales, concealing rapid adaptive evolution which fluctuates over shorter time scales, with changing environments.

While this model is compelling, the claims of novelty are overstated and often hyperbolic. The first Discussion sentence claims that it 'shakes the very foundation of the theory' but the following sentence more modestly states that this is not necessarily surprising. The neutral theory is portrayed as a bit of a straw man, when in reality the current understanding of drift-selection balance is more nuanced. The idea of adaptive tracking is not novel in itself, having been developed in reference 42, along with previous work by two of the authors (Chen & Zhang, NEE 2020), and supported by empirical work by Rudman (Science, 2022) among others. I do think that the key contribution in this work is the recognition that adaptive tracking creates patterns of neutral substitution that are more similar to neutral models, explaining the molecular clock better than a standard directional selection model. Even this realization is related to recent work by Torrillo & Lieberman (eLife 2024) describing an 'adaptive reversion' model to explain the time-dependence of estimated dN/dS values in microbial genomes. All in all, the work by Song et al. provides a useful synthesis of the available data, but the key novel contributions are sometimes hard to discern amid the abundant (sometimes excessive) analyses.

We thank the reviewer for the overall positive assessment. We have now toned down the statement about the novelty of our work. For example, the first sentence in Discussion now reads "Our analysis of the DFEs of nonsynonymous mutations showed that the key premise of the neutral theory is violated, questioning the general validity of the theory" (lines 341-342). We have now discussed Torrillo & Lieberman (2024 *eLife*) in lines 369-373.

The novelty of our work lies in three areas. **First**, we discovered that beneficial nonsynonymous mutations are too numerous to be compatible with the neutral theory of molecular evolution, thus rejecting the neutral theory for functional genes. **Second**, we proposed and demonstrated that adaptive tracking with antagonistic pleiotropy can explain both high beneficial mutation rates and seemingly neutral patterns of molecular evolution, suggesting that this model represents the actual process of evolution. **Third**, our model has profound implications. For example, it suggests that natural populations are always adapting to their (changing) environments yet are far from fully adapted. This is distinct from both the neutral and neo-Darwinian views of evolution.

In many ways, the manuscript reads like a review or synthesis paper, alternating between background, discussion, and data. This is, at times, difficult to follow, with extended discussion/synthesis paragraphs in the Results (e.g. lines 113-133) and new results brought into the Discussion. Aside from these structural difficulties, the relevance of the data and experiments is not always clear.

While the structure of our manuscript is somewhat unusual, the other three reviewers all commended the general clarity of our writing—"well-written" (reviewers 1&3) and "clear and quite straightforward" (reviewer 4). We thus did not alter the structure of our manuscript in the revision, but added some explanations in the manuscript that should help the reviewer and readers better understand this work.

First, the deep mutational scanning analyses (Fig 1) serve to refute the neutral theory, but it is not clear why the neutral theory needs refuting in the context of this manuscript. Moreover, this analysis raises many questions. The estimates of omega are extremely high (>1e4) – much higher than typically observed in comparative genomic studies, and certainly higher than the values obtained in SLiM simulations (Fig 2) and from experimental evolution (Fig 5). Is this because DMS provides a near-instantaneous fitness measurement? Or is it biased toward large-effect mutations? The alpha values are also very high, perhaps for these reasons. Moreover, although I appreciate the effort taken to bring together 4 different datasets, it is not at all clear if these few genes are representative of genome-wide patterns.

Rejection of the neutral theory motivates the search for an alternative model that explains patterns of molecular evolution so commonly explained by the neutral theory. Therefore, it is necessary to first show the invalidity of the neutral theory in our manuscript.

The reviewer is right that DMS provides a near-instantaneous fitness measurement because the measurement is taken in a constant environment. This is likely the reason why Ω and α estimates from DMS measurements are much higher than those observed from comparative genomics, as explained by our model of adaptive tracking with antagonistic pleiotropy (lines 137-158).

The DMS-based DFE data are not without caveats, but we believe that they provide qualitatively accurate information (see lines 483-520 for discussions of caveats and representativeness of the DFE datasets and genes).

Second, the relevance of the evolution experiment is not clear. Although it is impressive work, it is not clear why the data from the 2020 experiment with similar design (Chen & Zhang, NEE 2020) could not have been used. The original 2020 experiment supports most predictions of the adaptive tracking hypothesis: the main result of figure 5 – that dN/dS (small omega) tends to be smaller (~0.8-1.8) in changing than constant (~1.5-2) environments is very similar to the result already reported in Chen & Zhang (Nature Ecology & Evolution, 2020). Certainly the new experiments add some subtlety and of course it's always good to replicate experiments, but they do not seem to address a major unsolved question.

Chen and Zhang (2020 *Nat Ecol Evol*) performed experimental evolution in two changing environments, each composed of 5 different media. Only in one of the changing environments did the authors detect a signal of antagonistic pleiotropy. Hence, it was unclear how common antagonistic pleiotropy occurs in changing environments. The present study tested 10 changing environments, at least 9 of which showed signals of antagonistic pleiotropy, demonstrating that this is a very common phenomenon. We have explained the above motivation of the yeast evolution experiment in lines 268-277.

Overall, this is a very interesting paper, but it would be more palatable with some major structural changes. Consider structuring it more like a review/synthesis paper, and perhaps removing some of the data.

We thank the reviewer for the overall positive comment. As mentioned, because all other reviewers liked the structure of the manuscript, we decided not to alter it. We hope that the added explanations help the reviewer and readers better understand this work.

Minor:

1. In the section 'Estimating omega and alpha from empirical DEs' (lines 496 onward), do the calculations setting a value for Ne? How is Ne set here?

Yes. We first used the empirically estimated long-term N_e of yeast (10^7). We then used several different N_e values in the calculation to ensure that our conclusion is robust. We have now clarified it in lines 546-549.

2. In the section 'Estimation of the fraction of beneficial substitutions' (lines 749 onward), the authors state that at least two of the 12 replicates must have a substitution in the same gene for it to be counted. Presumably they mean substitutions relative to the ancestor, not to the reference genomes – because the reference genome used was not the same as the ancestor (if I understood earlier sections correctly). Please clarify.

The reviewer is correct that we called the substitutions relative to the ancestor. We have now clarified that in lines 771-782.

3. Fig 5 refers to 'corresponding' constant environments but it is not clear how each changing environment is matched to a corresponding constant. It also seems from the method that there was only one changing environment, replicated 12 times, which was compared to 12 different constant environments. I think this explains why there is a boxplot (blue points in Fig 5b) for the constant environments, but a single red line for the changing environment. Was the t-test really done by comparing these 12 blue points to a single 'red' value? If so, this seems somewhat underpowered and maybe not the ideal statistical test. In any case, the general direction of the effect is clear from panels b, c and d combined.

One changing environment is composed of 10 different media, so it corresponds to 10 constant environments. We indeed performed a *t*-test to test whether the red value significantly deviates from the distribution of the 10 blue points in each diagram of Fig. 5b. We agree that the test is not very powerful, which may explain why only 5 of the 10 sets of experiments in Fig. 5b showed a significant effect. This low power, however, is compensated by the increased power offered by the combined analysis of the 10 sets of experiments in Fig. 5c.

Reviewer #3

This is an interesting, thought-provoking, and well-written paper that investigates a model of fluctuating selection that could explain patterns of molecular evolution. I recommend the paper for publication, and of only a few comments. I think that's paper can be published more or less as it is.

We thank the reviewer for the very positive evaluation.

At line 52, the neutral theory is no longer generally accepted, as there has been a great deal of evidence for adaptive molecular evolution. See for example Smith and Eyre-Walker Nature 415(6875): 1022-4.

We have now added the above reference in line 55. Although we agree with the reviewer that there is now a great deal of evidence for adaptive molecular evolution, such evidence tends to be concentrated in a few species such as *Drosophila* or in a small subset of genes such as immunity genes. Considering the reviewer's comment, we have replaced the word "widely" with "commonly" in the sentence, which states that the neutral theory is "now commonly (albeit not universally) considered an accurate description of the evolutionary process for *most* genes and proteins". Whie some may disagree, we believe that this is a fair assessment of the field.

From line 71 onwards, estimates of the DFE are for very artificial environments. We don't really know what the DFE will look like in the natural environment. This requires more discussion.

We agree and have discussed this extensively in the section "Relevance to natural evolution" (starting line 110).

Reviewer #4

Dear Authors, i was asked to review this manuscript that I had already seen. I mentioned that to the editor that told me it was not an issue. So a lot of my comments will match my previous review as they were mostly commentaries on the tone and background of the paper, and some still hold true in the present manuscript, I omitted the ones that were answered in the current version (thanks).

In the manuscript entitled "Adaptive Tracking results in seemingly neutral molecular evolution" Song et al present a framework in which large fluctuation in the environment affect the effects of beneficial mutations that are much more frequent that a neutral model assumes. The models builds on observations that deep mutational scans as well as some mutation accumulation have revealed much larger than expected fraction of beneficial mutation which integrated over long evolutionary time scale would result in extremely high fraction of non-synonymous mutation compared the accumulation of synonymous mutations. This rapid adaptation in a given environment is also comforted by experimental evolution in which fast adaptation is observed but often recruits mutations that are at odds with the level of conservation observed in the targeted genes. The model proposed suggests that antagonistic pleiotropy may be at play and that mutations that are beneficial are so only locally and transiently and that they therefore do not reach fixation. To support that model, they build on simple simulations in which neutrals

mutations keep being neutral but a fraction of mutations may switch randomly between beneficial and deleterious at pace that range from a few generations to few hundreds of generations. These simulations confirm obviously the fact that fixation of beneficial mutations in that context is marginal and the long-term pattern is overwhelmingly dominated by neutral mutations. An experimental validation is made through a large-scale experimental evolution involving stable of alternating environment over 800 generations of yeast evolution in the laboratory. Here, sampling individual clones after that evolution they confirm that there are less beneficial mutations and more neutrals in the fluctuating environment.

Besides a few technical comments I will present later, the article is clear and quite straight forward. On the positive side, it is timely, as data showing the intensity of adaptation and the extent of beneficial mutations are accumulating and contrast with macro evolutionary patterns. Yet, digging into the literature, the whole idea is also quite known in the field of evolution for a long time: antagonistic pleiotropy, G by E, standing genetic variation, fluctuation environment largely dealt with that concept so one may question the novelty of the approach.

We thank the reviewer for the overall positive evaluation and for spending time re-reviewing our manuscript. We agree that the ideas of antagonistic pleiotropy, G by E, standing genetic variation, and fluctuating environments are not new. However, our analysis focused on the conflict between the abundance of beneficial mutations and apparently neutral molecular evolutionary patterns, which had not been studied previously. Associating fluctuating natural environments with antagonistic pleiotropy for the purpose of resolving this conflict is also new.

To support the novelty, I would say that it is simple here and all previous concepts have been quite widely used over the short term but very few from what I have seen have integrated the process to genomic patterns, so that perspective is new.

We thank the reviewer for recognizing the novelty of our work.

I think however that a lot of statement about novelty should be downplayed a little to incorporate more the previous work:

In essence some models have been used to study environmental fluctuation, which usually is based on an additive effect of mutations on a trait connected to fitness, but a change in the optimal value of that trait. Here, the model is mutation centered and assumes shifts in effects but these are indeed equivalent, so connection with these models would be welcome. For instance, Graham Bell in “Fluctuating selection: the perpetual renewal of adaptation in variable environments » acknowledged that mild effect mutations could not easily rich fixation but that large effects ones could, somehow going beyond the present model in which large effect mutation would be mutations with fixation time faster than the environmental switching time. Similarly, Grossman et al in “Fluctuating Selection Models and Mcdonald-Kreitman Type Analyses” (PlosOne 2014) discussed how selection fluctuating impacts fixation, and identified the type of mutations that will in fine be fixed. Another case in, which the consequence of fluctuating s in the environment can be found in A. Desbiez-Piat et al “Interplay between extreme drift and selection intensities favors the fixation of beneficial mutations in selfing maize populations” (Genetics 2021), in which the high fluctuation in selection through time in plant is

discussed for its consequences on fixation. Another related work is “Random fluctuation of selection coefficients and the extent of nucleotide variation in human populations” by Sayaka Miura, Zhenguo Zhang, and Masatoshi Nei (PNAS 2014) which starts with “It is well known that the selection coefficient of a mutant allele varies from generation to generation, and the effect of this factor on genetic variation has been studied by many theoreticians.” And finishes with “This result implies that there is no need of distinguishing between the diversity-enhancing and diversity-reducing models of fluctuating selection and the nucleotide polymorphism in human populations can be explained largely by neutral mutations when long-term evolution is considered.” So I think a full paragraph stating simply that the present work builds on these past work but focusses on the molecular evolution dimension would really be worth.

Finally, experimentally, fluctuating selection has been shown to limit the fixation of mutation, as seen in yeast recently by F. Abdul-Rahman et al “Fluctuating Environments Maintain Genetic Diversity through Neutral Fitness Effects and Balancing Selection ». So even if each of these works tackle the question differently the idea proposed here could be integrated in these framework. cannot be said to be new.

We thank the reviewer for the suggested references. We have now cited 14 papers about fluctuating selection (ref. 53-66 in our manuscript) and acknowledged their relevance to our adaptive tracking model in lines 359-373. We have also explicitly compared some fluctuating selection models with our model (lines 359-373). This said, we believe that our study is distinct from these earlier studies.

For example, Bell (2010) mentioned, “selection is generally rather strong and fluctuates on all time scales such that abrupt changes can occur over short periods of time and gradual directional change occurs over long periods of time”. While this is consistent with predictions of our adaptive tracking model, the focus of Bell’s discussion was on phenotypic evolution. Bell did not discuss molecular evolution and never pointed out the incompatibility between the abundance of beneficial mutations with the neutral theory. We cited this paper.

Grossman et al. (2014) showed that fluctuating selection, even with a zero expected fitness effect, tends to yield positive selection signals due to ascertainment bias, but this phenomenon is not related to what we study, which is based on direct measures of mutational fitness effects. We did not cite this paper.

Desbiez-Piat et al. (2021) used experimental evolution and simulation to investigate how the magnitude of drift affects molecular evolutionary dynamics under strong directional selection. However, because their study did not involve fluctuating selection or antagonistic pleiotropy of mutational effects, its findings are not directly relevant to our study. We did not cite this paper.

Miura et al. (2013) investigated the impact of fluctuating selection on genetic variation, which is marginally relevant to our focal subject. Miura et al. assumed an overall neutral fluctuating selection with the mean selection coefficient being zero, while our adaptive tracking model assumes that most non-neutral mutations are overall deleterious under fluctuating selection. We cited this paper.

Abdul-Rahman et al. (2021) reported that fluctuating environments can help maintain genetic polymorphisms, but it did not address the key question of our study, which is why the nonsynonymous substitution rate is so low despite the abundance of beneficial nonsynonymous mutations. Furthermore, fluctuating selection alone cannot explain this observation. We cited this paper.

While fluctuating selection has been studied multiple times, explicitly combining it with antagonistic pleiotropy is rare; the only exception may be our own earlier study (Chen and Zhang 2020 *Nat Ecol Evol*). To the best of our knowledge, fluctuating selection had never been combined with antagonistic pleiotropy to explain the contradiction between the abundance of beneficial nonsynonymous mutations and the seemingly neutral patterns of molecular evolution. In fact, this contradiction is a new discovery of our study.

There is also a small papers of ours (Vigue Tenaillon PNAS 2023) that may match this various time scale of selection hypothesis and may bring some ideas: using 60000 genomes of E. coli and some inferred effects of mutation, we detected that some genes showed excess deleterious mutations even though there are core genes, suggesting that at local scale selection may favor lower functioning of these genes, that are conserved otherwise. This also raise the question of heterogeneity across space and time, can the patterns of environmental variation observed here may be relevant at the species level to drive signals that will impact fixation rate…

If our understanding is correct, this paper suggests temporal variations in the functional constraint of a gene but is not about adaptive tracking or antagonistic pleiotropy. Hence, it appears minimally relevant to our manuscript.

Concerning the model: many layers of the potential complexity have been taken into account. I am globally in line with the modelling performed and appreciate its simplicity. The question though is over which time scale is the measure of s appropriate. For very short switching time, s over one generation is not appropriate but its average over time. Time should be counted as time needed to substantially change in frequency. Then the proposed model converge to a neutral model as the mean s of mutations is negative. Indeed, for both microbes and even more for higher organisms, the environment may fluctuate through a life time, so s is already an integration over some time, to what extent integrating over longer time scale is that different. Here the time scale of changes of the environment is somehow set to exactly match 1/s. Which allow substantial changes in allele frequency. The question is how likely is that to happen: if changes are slower, the adaptive model would predominate, if changes are faster, beneficial alleles will barely increase in frequency. The importance of this balance should be mentioned.

The reviewer is correct that environmental changes that occur too quickly relative to the generation time of a species may not matter, because there is no chance for allele frequencies to respond to such brief environmental changes. In nature, environmental changes occur at all time scales, so the adaptive tracking model is in principle applicable to almost all species. We have now added discussions about the time scale of environmental changes in lines 415-420. Note that our model will not converge to a neutral model, because under our model there are almost always many beneficial mutations available given that the population is almost never close to any fitness optimum.

In our SLiM simulation, we did not intentionally choose the timescale of the environmental change to match 1/*s*. We varied the frequency of environmental changes from once per 2 to once per 200 generations, and indeed, greater allele frequency fluctuations were observed when each environment lasts longer.

I think the consequence of the model on polymorphisms could be exploited further and more than qualitatively as done in figure 3. For instance, contrasting the site frequency spectrum with one including just neutral or background selection would be valuable and offer potential ways to test if these scenarios can be inferred from data. One key component of classical models is the probability of surviving drift, if burst of selection lead to rapid increase of selection this may affect this probability. Moreover, rapid decrease of frequency upon environment shifts will lead to profound reduction of effective population size that may also affect the probability of surviving drift.

We have compared the site frequency spectrum under various models in Fig. 2aeim. Additionally, we attempted to differentiate the adaptive tracking model from other models using population genetic patterns, as described in lines 393-395 and in Fig. S14. The change of fixation probability of neutral mutations due to environmental changes is automatically accounted for in the SLiM simulation. Furthermore, as the population size likely drops when a new environment is stressful, we considered a model in our SLiM simulation where the population size varies depending on the harshness of the environment. The results obtained (Fig. S5a-d) are not qualitatively different from our main results acquired under the constant population size.

Finally, for Figure 5, it is hard to know from the data how many mutations are used to infer omega, in other words, how robust are the estimates of omega. With low numbers noise may be high.

The data are now provided in Data S6.

Overall, it is a simple and timely story, that should help many to value the different scales at which evolution acts, but to me it could still be more anchored in the existing literature, to avoid harsh criticism from people in the field of evolution.

Olivier Tenaillon (I always sign my reviews)

We thank the reviewer for the positive evaluation and valuable comments. We have now referenced additional, relevant, past studies when discussing our adaptive tracking model.

Response to reviewers

Below please find our point-to-point response in blue.

Reviewer #1:
I have reviewed the authors response. I appreciate the time they have taken to address my comments, and those of the other reviewers, and especially to introduce parts of the response into the manuscript text and figures. All of my comments have been addressed, and I have no further questions. I congratulate the authors on an excellent piece of work.

We thank the reviewer for re-reviewing our manuscript and for approving our revision.

Reviewer #2:
Thank you for the revisions and the detailed responses to my comments. All my substantial comments have been adequately addressed. Congratulations on this exciting work!

We thank the reviewer for re-reviewing our manuscript and for approving our revision.

Reviewer #4:
I appreciate the fact that the author now cite a broader part of the literature from the field and have down toned a bit the novelty aspect fo the work. The focus on the molecular evolution consequence of fluctuation environment is a worth perspective.

The last sentence of for sure an opinion: "Hence, the adaptive tracking model may be even more important for multicellular than unicellular organisms". It could also be the exact opposite. As multicellular organisms have longer life span the fitness effect of mutation is already an average across a multitude of environment that an organism face during a life cycle and across a multitude of tissues, and as such is less likely to be beneficial contrary to microbial life and assays that can detect benefit in a very particular local environment. Si in my opinion the adaptive tracking may by less likely to happen for multicellular. The question remains open.

We thank the reviewer for re-reviewing our manuscript and for the additional comment. We agree with the reviewer that the relative importance of adaptive tracking in unicellular and multicellular organisms is unclear and requires future studies. We now write in the final paragraph of Discussion the following, with additional information in Methods and Supplementary Discussion.

"Compared with unicellular organisms, multicellular organisms generally have longer generations and smaller N_e, but their combined effect on the importance of adaptive tracking is unclear when the time scale of environmental changes is considered (see Methods and Supplementary Discussion)."