# Response to editor and point-by-point response to reviewers' comments on "*Detecting diversifying selection for a trait from within and between-species genotypes and phenotypes*"

## Editor:

After reading this manuscript again, I could not help but relate this study to Lynch (1990, Am Nat 36: 727-741) that the skull of human is the only trait that exceed the minimum rates of evolution expected from neutral model of phenotypic evolution among traits that he had information for. Those included dental and skeletal measures. One interpretation of the empirical analyses shown here is that brain size, which is closely related to skull size in hominids (as well as all birds and mammals), might be an exceptional case (at least among morphological traits). This point should perhaps be mentioned in discussion in line 418-420 by citing Lynch (1990).

Along the same vein, one could see this study as a (much advanced) sequel to Lynch (1990), at least in its core spirit. Even though the method and the goal is more like Ovaskainen et al (2011), Equation (3) of Lynch (1990) is a ratio of within- and between-species variance (Lynch termed these var(w) and var(b), pretty much as the authors did!). The current narrative is already pretty clear so nothing much is necessary, but mentioning Lynch (1990) somewhere in intro, perhaps at the very end in line 93-98, would further solidify the context of this work.

Thanks a lot for this clarification, we now mention Lynch (1990) in the *Introduction*, changing from
"The ratio that we propose can be considered as a neutrality index for any quantitative trait articulating trait and nucleotide variation within and between species." to:
"The ratio that we propose can be considered as a neutrality index for any quantitative trait (Lynch, 1990), while articulating trait and nucleotide variation within and between species."

And in the *Discussion*, changing from:
"However, such datasets will become more and more accessible and we showed the applicability of our method by applying it to the illustrative example of mammals' brain and body mass, showing signals of diversifying selection. The consensus on macro-evolutionary studies, assuming constancy of $N_e$, generation time and mutation rates, is that empirical rates of evolution calculated on phylogenetic trees and the fossil record are far inferior to the expected under drift (Lynch and Crease, 1990; Uyeda et al., 2011). Our finding of diversifying selection on body and brain mass could be seen as an argument against that interpretation. In fact, rates of nucleotide evolution also show a tendency for slowing down on a longer timescale (Rolland et al., 2023). One possible interpretation is that normalization by nucleotide divergence could absorb this observed slowing rate of evolution. Altogether, further empirical and theoretical studies are required to disentangle this discrepancy between these different interpretations." to:

"However, such datasets will become more and more accessible and we showed the applicability of our method by applying it to the illustrative example of mammals' brain and body mass, both showing signals of diversifying selection. As such, this result corroborates studies relying solely on changes in mean trait values across mammals, showing strong statistical support for several distinct evolutionary regimes for body- and brain mass (Mitov et al, 2019). Interestingly, our strongest signal is for brain mass, corroborating studies in hominids where skull size is the only trait that exceeded the expected rate of phenotypic evolution under a neutral model (Lynch, 1990). Hence, one first interpretation here is that brain mass (related to brain mass) might be an exceptional case among many phenotypic traits (e.g. dental and skeletal measures). Second, from a macro-evolutionary perspective, the consensus is that empirical rates of evolution calculated on phylogenetic trees and the fossil record are far inferior to the expected under drift (Lynch and Crease, 1990; Uyeda et al., 2011), where such methods assume constancy of $N_e$, generation time and mutation rates. Our finding of diversifying selection on body and brain mass could be seen as an argument against that interpretation. In fact, rates of nucleotide evolution also show a tendency for slowing down on a longer timescale (Rolland et al., 2023). One possible interpretation is that normalization by nucleotide divergence could absorb this observed slowing rate of evolution. Altogether, further empirical and theoretical studies are required to disentangle this discrepancy between these different results and interpretations."

Line 72: two low -? Too low

Sorry for this mistake.

Line 82 - 84: Probably authors forgot to remove "Between to" at the beginning? Please check and correct accordingly.

Completely agreed it is more straightforward without the "Between to"

Line 90: remove "a"

Thanks for spotting it.

References:

Lynch (1990) The Rate of Morphological Evolution in Mammals from the Standpoint of the Neutral Expectation. The American Naturalist 136: 727-741

# Reviewer I:

The authors have thoroughly revised their manuscripts, building up constructively upon the suggestions from the previous round of review. I agree with them that the manuscript's clarity has improved in the process. In particular, I appreciate how they have modified their treatment and explanation of molecular and phenotypic divergence, I find the new version much more explicit and convincing. I therefore only have a couple comments, which relate to news additions that are not central to the ms.

Thank you for this comment.

First, I think the statements on saturation of phenotypic divergence (399-401 in Discussion, and section 4 in appendix) need to be toned down. There is no clear consensus that "the phenotype is […] ultimately bounded" because it "encoded by a genetic architecture", such that "phenotypic divergence should plateau at some point". This can happen or not, depending precisely on assumptions about the genetic architecture: number of alleles per locus, whether new mutations add something to the phenotype or replace the previous effect by a new one, etc. So I think this argument but should be left more open, using more conditional statements, such as: "The phenotype may be bounded by what the current genetic architecture can produce, and this could cause phenotypic divergence to plateau after all possible genetic change at underlying loci has been exhausted". And then describe this as a possible scenario, which you explore in the appendix. In my opinion, if mutation limits (rather than selection) were such a prevalent cause for the slowdown of evolutionary rates over time and for the paradox of stasis, then this would have already been shown several times in the literature, which doesn't seem to be the case. But I still think that this scenario is interesting to study, and that the appendix is a useful addition to the paper.

We agree that our phrasing in the *Discussion* needs to be toned down, and that our explanation can be more precise on the slowdown of evolutionary rates over time. The first argument is indeed that under a constant genetic architecture, the phenotype is constrained by the possible genetic changes leading to a phenotype that is ultimately bounded. But second, and probably more important in our case, the phenotypic divergence is reduced since any mutation that hits the same loci will revert the phenotype, resulting in a slowdown of evolutionary rates over time. In other words, the reduction in phenotypic variance occurs even before the genetic changes at underlying loci have been exhausted. We rephrased the *Discussion*:

"However, the phenotype is encoded by a genetic architecture, and is thus ultimately bounded, even more so if the trait is encoded by a few loci (see table 2). At the macro-evolutionary scale, phenotypic divergence should plateau at some point, resulting in a reduced between-species trait variation. We argue that this effect can result in a spurious signal of stabilizing selection ($\rho<1$), especially for deeper phylogeny (figure S2 and section S4)" to:

"However, the phenotype may be bounded by what the genetic architecture can produce, and this could cause a slowdown of phenotypic divergence over time due to the erosion of possible phenotypic changes at the underlying loci. Typically, such an effect depends on the number of alleles per locus, whether new mutations are generating new alleles or instead reverting to previous alleles. Altogether, in our simulation setting under a constant genetic architecture with a fixed number of loci, such a slowdown of phenotypic divergence can result in a spurious signal of stabilizing selection ($\rho<1$), especially for deeper phylogeny (figure S2 and section S4)"

And in the *Appendix (section S4)*, from:

"The phenotype is encoded by a genetic architecture and is thus ultimately bounded. At the macro-evolutionary scale, phenotypic divergence should plateau at some point, ultimately resulting in a decreased $\sigma^2_W$." to:

"In our simulation setting, the genetic architecture underlying the phenotype is not changing along the phylogeny. As a result, the phenotype may be bounded by what such genetic architecture can produce, and this could cause a slowdown of phenotypic divergence over time, ultimately resulting in a decreased $\sigma^2_W$."

Another minor point regards phenotypic plasticity. This happens to be a topic on which I work a lot, and I don't quite agree with what the authors wrote about in Table 2 (and their letter). Phenotypic plasticity is the response of individual phenotypes to their environment of development/expression. It can indeed modify the phenotypic variance within a population, if this population's habitat includes fine-grained, micro-environmental variation (ie individuals occupy different patches with different environments). However, the more common issue on this topic (e.g. in Qst–Fst analysis) is that different populations (or here species) are likely to experience different macro-environments, and thus have different mean phenotypes because of phenotypic plasticity; think for instance of populations/species spread along a latitudinal or elevation gradient, with different temperatures, precipitation, and so on. This is one of the main reasons why a common garden is needed to study phenotypic divergence. There are many references on this topic, but perhaps the most useful would be: Stamp, M. A., & Hadfield, J. D. (2020). The relative importance of plasticity versus genetic differentiation in explaining between population differences; a meta-analysis. Ecology letters, 23(10), 1432-1441. A more recent paper that is also somewhat relevant (although it fails to explicitly mention plasticity, unfortunately) is : Schraiber, J. G., & Edge, M. D. (2024). Heritability within groups is uninformative about differences among groups: Cases from behavioral, evolutionary, and statistical genetics. Proceedings of the National Academy of Sciences, 121(12), e2319496121. Of course I understand that this is a secondary point in the paper, but if the authors want to mention plasticity, I think they should also account for its effect on the mean trait, which is arguably more important than that on the variance.

Thanks a lot for this comment and the references. We agree that the effect of phenotypic plasticity on trait mean and variance is more nuanced than what we have stated. In the study most similar to our objective (i.e. including within species variation at the phylogenetic scale), phenotypic plasticity has so far been seen as affecting the trait variance and not the trait mean (Rohlfs & Nielsen, 2015). However, we agree the effect on the trait mean should be acknowledged and the consequences discussed.

Altogether, we have removed the line in table 2 about phenotypic plasticity, but we added this paragraph to the *Discussion*, also mentioning your examples and the consequences for our test:

"Additionally, phenotypic plasticity also affects the genotype-phenotype relationship with intricate consequences for our test of selection. First, at the level of within species variation, individuals might occupy different patches with different environments. Responding to these individual environmental conditions, phenotypic plasticity would then result in increased trait variation within species. In this scenario, as hypothesized in Rohlfs & Nielsen (2015), phenotypic plasticity then leads to a reduced ratio of between to within species variations, thus ultimately leading to our tests of diversifying selection being underpowered although not invalid. Alternatively, it is also possible that different species are experiencing different macro-environments, for example with species spread along a latitudinal or elevation gradient, with

different temperatures or precipitation. These species could thus have different mean phenotypes solely because of phenotypic plasticity, while such changes are not encoded in their genome (Stamp & Hadfield, 2020; Schraiber & 2024). Such an effect can lead to ρ>1 erroneously interpreting diversifying selection. The test of ρ>1 would however be correct that the changes in mean phenotypes across species is due to change in environment, albeit in such a case not encoded by the genotype of individuals but due to phenotypic plasticity.``

Minor points (line numbers are from the diff pdf):
82 : typo: “rate too low”
Completely agreed

310: I would write “flattening the fitness landscape”, as fixing has another meaning in evolution, and even regardless of that, “fixed” implies “constant”, but there could be selection with a constant fitness landscape.
Completely agreed, “flattening” conveys the proper meaning while “fixing” was misleading.

429: it would seem like Lande (1979) should be cited here, since it is the L in LGGD ?
Completely agreed

## Reviewer II:

I am mostly satisfied with the author's reviews of my previous comments, and I appreciate all the changes that the authors made. I have one potential concern: I appreciate that the authors disentangled the Multivariate Brownian process from the Bayesian estimate. However, I am left with many questions about this. The text could be clearer on what is actually being done. As far as I could understand, the test is still being performed on each trait individually, after the appropriate diagonal entries of Σ are estimated. If that is correct, the only multivariate part is the estimation of the model parameters, but the test itself is still univariate. There is no problem with that, but that has to be made clear. I do wonder, though, what are the benefits of doing this instead of independently estimating rate parameters for each trait? You are adding more parameters (such as covariances) that do not seem to be used by the method and that are harder to estimate. Overall, I am left with many questions of the true utility of a multivariate method that is not taking advantage of the multivariate estimates (covariances). I think the authors should try to explain a bit better what they are trying to achieve here or rethink if leaving this part is essential to the message they are trying to convey.

This is an excellent remark and completely correct. We tried to clarify the text to explain our rationale on when and why to use a single multi-dimensional process instead of many 1-dimensional processes. There is currently one advantage of including covariances in the estimation which is an improved estimation of evolutionary rates (Adams & Collyer, 2018). We changed the section *Multivariate Brownian process* from:

"In the previous section, ρ is estimated independently for each trait of interest. Here we generalize to K traits co-varying along the phylogenetic tree. Trait variation along the phylogenetic tree is modeled as a K-dimensional Brownian process $B$ (1×K) starting at the root and branching along the tree topology (Huelsenbeck and Rannala, 2003; Lartillot and Poujol, 2011; Lartillot and Delsuc, 2012; Latrille et al., 2021). The rate of change of the Brownian process is determined by the positive semi-definite and symmetric covariance matrix between traits Σ (K×K). The branch lengths of the tree used to model the Brownian process runs is measured in units of 4d (d being the nucleotide divergence). The off-diagonal elements of Σ are the covariance between traits, and the diagonal elements are the variance of each trait when measured in 4d units, and thus equate to $\sigma^2_B$ (see section S2.1)." to:

"In the previous section, ρ is estimated independently for each trait of interest. Here we generalize to K traits co-varying along the phylogenetic tree, since simultaneously estimating all $\sigma^2_B$ can improve their estimation (Adams & Collyer, 2018). More specifically, trait variation along the phylogenetic tree is modeled as a K-dimensional Brownian process $B$ (1×K) starting at the root and branching along the tree topology (Huelsenbeck and Rannala, 2003; Lartillot and Poujol, 2011; Lartillot and Delsuc, 2012; Latrille et al., 2021). The rate of change of the Brownian process is determined by the positive semi-definite and symmetric covariance matrix between traits Σ (K×K). The branch lengths of the tree used to model the Brownian process runs is measured in units of 4d (d being the nucleotide divergence). The off-diagonal elements of Σ are the covariance between traits, and the diagonal elements are the variance of each trait when measured in 4d units, and thus equate to $\sigma^2_B$ (see section S2.1). Of note, modeling trait evolution as a multi-dimensional process is reliable only if K<N, meaning that the number of species is largely superior to the number of traits (Adams & Collyer, 2018). Thus, relying on a K-dimensional process should be reserved for a handful of allometric traits (e.g. brain mass and body mass). If K is large, the traits are better tested independently each with a 1-dimensional Brownian process, which is a specific case of the multi-dimensional process."

And finally, to be more precise, we changed the text in the section *Bayesian estimate* from:
"From $\sigma^2_W$ and $\Sigma$, the posterior distribution of $\rho$ (as in eq. 28) is obtained for each trait." to:
"From $\sigma^2_W$ for each trait and diagonal elements of $\Sigma$ (i.e. the $\sigma^2_B$ for each trait), the posterior distribution of $\rho$ (as in eq. 28) is obtained for each trait."

Adams, D. C. and Collyer, M. L. 2018. Multivariate Phylogenetic Comparative Methods: Evaluations, Comparisons, and Recommendations. Systematic Biology, *67*(1): 14–31.

Despite this, I have no great issue with the manuscript, other than thinking that clarity could be improved on some points and the prose is still a bit difficult. Below, I list some of these minor concerns with suggestions that I think might help the authors prepare the final version of their manuscript. I will refer to the line numbering on the tracked change document.
We acknowledge that the manuscript is heavy and reading can be improved. In light of your comment and upon several readings of the manuscript, discussion with colleagues, and also from your comments from the previous round of reviews, we agree that the hat() notations for estimates is actually not doing the manuscript a favor, causing more friction than bringing clarity. Thus we have removed this notation altogether and adapted a few sentences to match this change. We believe this change makes the text and the figures more lightweights, and allow for a smoother reading.

L51- Felsenstein (1988. Annual Review of Ecology and Systematics, 19(1), 445-471.) could also be cited here.
Completely agreed, thanks a lot we added the reference to Felsenstein (1988).

L69- Felsenstein (1988) and Felsenstein, J. (1985. The American Naturalist, 125(1), 1-15.) could be cited as well.
Completely agreed, we added the reference to Felsenstein (1988) and Felsenstein (1985).

L75- Hansen (1997. Evolution, 51(5), 1341-1351) is also a good reference here.
Completely agreed.

L82-83- the added phrase is breaking the flow of the points being made here. Maybe add a "second" instead of also, and elevate the current "second" to a "third"?
Completely agreed, we implemented your suggested changes.

L90- Mitov et al (2019. Proceedings of the National Academy of Sciences, 116(34), 16921-16926) also fits here
Thanks a lot for pointing this out, we agree that Mitov et al (2019) is highly relevant to our manuscript. However, this sentence (l. 90-92) is about *"the importance of modeling within-species variation together with changes in mean trait value"*, which is not the message of Mitov et al (2019) which focused solely on changes in mean trait value across species.
Altogether, we added the reference to Mitov et al (2019) in the *Discussion*, changing from:

"However, such datasets will become more and more accessible and we showed the applicability of our method by applying it to the illustrative example of mammals' brain and body mass, showing signals of diversifying selection. " to:
"However, such datasets will become more and more accessible and we showed the applicability of our method by applying it to the illustrative example of mammals' brain and body mass, both showing signals of diversifying selection. As such, this result corroborates studies relying solely on changes in mean trait values across mammals, showing strong statistical support for several distinct evolutionary regimes for body- and brain mass (Mitov et al, 2019)."

L92-94- There are many other models and processes that do exactly the same thing as the "ratios" method discussed here. One of them is the eigenvalue regression method developed by Ackermann, R. R., & Cheverud, J. M. (2004. Detecting genetic drift versus selection in human evolution. Proceedings of the National Academy of Sciences, 101(52), 17946-17951), which basically do a "ratio test" to infer the presence of selection in multivariate data. This method has been used in many publications (e.g., Marroig, G., & Cheverud, J. M. 2004.The American Naturalist, 163(3), 417-428.; Machado, F. A. 2020. The American Naturalist, 196(2), 197-215, de Azevedo et al. (2017. Proceedings of the National Academy of Sciences, 114(47), 12442-12447; Schroeder, L., & Ackermann, R. R. 2017. Journal of human evolution, 111, 1-17). Some (like de Azevedo et al. 2017) have taken the difference of slopes in the eigenvalue regression to mean differences in evolutionary processes (stabilizing or directional selection). Other method that touches on this is the relative eigenanalysis proposed by Bookstein, F. L., & Mitteroecker, P. (2014. Evolutionary biology, 41, 336-350.) which identifies axes of variation that mostly differ among within and between populations. Simon et al. (2016. Proceedings of the Royal Society B: Biological Sciences, 283(1841), 20161783.) used this method (along with the eigenvalue regression) to argue for directions of most selection in a macroevolutionary context.
That said, all of these methods require strong assumptions about the trait quantitative genetics (like proportionality between Ps and Gs) and ignore completely divergence times, effective population sizes, and all these issues that the authors have already pointed out.
This is true, but these methods are at the within species level, while Rohlfs (2014, 2015) is about between species variations, to make this clear, we changed the text from:
"within-species variation has also been used to infer diversifying selection by estimating the ratio of between to within species variation of many traits and test for deviation from the average ratio across traits." to:
"Across many species, within-species variation has also been used to infer diversifying selection by estimating the ratio of between to within species variation of many traits and test for deviation from the average ratio across traits."

L106- is this method truly a generalization of Fst-Qst? If so, that would mean that standard Fst-Qst methods are special cases of this current method. I would suggest the authors to consider if this is actually the case. If not, use a more adequate term ("version", "analog" etc).
Thank you for this comment, we completely agree that "analog" is much better adapted than "generalization". Also in the discussion we changed from "extension" to "analog" when referring to $Q_{ST}$-$F_{ST}$ methods and their derivatives.

L257-258- the added phrase is a bit out of place here. Maybe move it to go with the other added phrase at L 260

We completely agree, we moved it.

L369- Did you mean “body size”?

Yes, thank you for pointing this out.

Table 1, lines 9-10. Aren’t the posterior probabilities supposed to be 1 (as per L379-380)?

Thank you very much for spotting this, there is absolutely an error in the text. We double checked and re-ran the analysis, the tables are the good output, but the reporting in results was wrong. We changed the text from:

“We also analyzed a similar dataset for body mass focusing this time only at Primates (Table 1). For primates body mass, we found posterior probabilities of diversifying selection of 1.0 for males and 0.914 for females when assuming a uniform distribution for the heritability of body mass between 20% and 40%. Assuming complete heritability of body mass did not change the posterior probability for males, but increased the one for female to 1.0. Evidence for diversifying selection on body mass was therefore more pronounced in Primates than in mammals.” to:

“We also analyzed a similar dataset for body and brain mass focusing this time only at Primates (Table 1). For primates body mass, assuming complete heritability, body mass was found to be under diversifying selection with posterior probabilities of 0.0 for both males and females, exactly as in the mammal dataset. However, we found posterior probabilities of diversifying selection of 1.0 for males and 0.914 for females when assuming a uniform distribution for the heritability of body mass between 20% and 40%. For brain mass, assuming complete heritability or not (between 20% and 40%) did not change the posterior probability of diversifying selection, which was 1.0. Evidence for diversifying selection on both brain and body mass was therefore more pronounced in Primates than in mammals.”