Revision round 2

Reviewer #2: In this revised manuscript the authors have essentially addressed all of my concerns from the previous round of review. However, there are still a few relatively minor issues remaining:

(1) In the revised supplement, the authors have gotten a little tripped up deriving the neutral phenotypic distribution under their model (Section 1.3) by jumping too early to a continuous approximation which they have trouble solving. In fact, the neutral phenotypic distribution is just a binomial distribution with parameters n and (1-1/K) that has been rescaled so that its support ranges from 0 to 1 in increments of 1/n rather than from 0 to n in increments of 1. This stationary distribution is obvious once you consider that under neutrality each site has an independent 1/K chance of being the optimal allele. As a result, the curves on pg. 3 of the supplemental material giving the stationary phenotypic variance based on simulations have the analytical formula (1/n)(1/K)(1-1/K).

I suggest that the authors update the supplement to reflect these simpler results. This could potentially also help with the results in Section 1.6, especially via the normal approximation to the binomial, but it is not clear to me how it plays out.

Admittedly we tripped up deriving the equilibrium by jumping to a continuous approximation. The discrete distribution is indeed a rescaled binomial $B(n, 1 - 1/K) / n$ that can be approximated by a Normal distribution $N(1 - 1/K, (1/n)(1/K)(1-1/K))$. We updated section 1.3 and 1.6 to account for this more audible and simpler result. Simulations in section 1.6 also provides the insight that the variance at mutation-selection equilibrium is lower than the theoretical variance at mutational equilibrium $(1/n)(1/K)(1-1/K)$. The normal approximation is however not sufficient to derive an analytical result for the expectation and variance at mutation-selection equilibrium, due to the complexity brought by the fitness function.

(2) Although the authors have generally done a good job addressing my concerns about distinguishing clearly between formally proven results and approximations/simulations, there is still a "proven" in the abstract, in particular that omega is independent of N_e "a result proven under general conditions". I suggest simply leaving this clause out, particularly since the abstract goes on to describe how the current manuscript characterizes HOW omega changes as a function of N_e, which would not make sense if one could prove that omega is independent of N_e.

The clause "a result proven under general conditions" has been removed from the abstract.

(3) My computer was having difficulty printing the supplemental material, and I suspect that it is due to the large images (with transparency?) on page 16. These could perhaps be converted to bitmaps.

Indeed, the difficulty was due to a large of number of points displayed in the pdf as vectorized objects. We rasterized these clouds in order to reduce printing time.

Revision round 1

Dear Editor,

Thank you very much.

We are grateful for your comment and the ones raised by the reviewers, all of which greatly enriched our perspective on the questions addressed in our manuscript.

-Size your figures to 1-column or 2-column size. The font is quite small in some of them.

All figures are sized to span the text width, also we increased the font of figures.

-In captions, refer to symbols by their conceptual meaning in addition to the symbol (e.g. "effective population size N_e" rather than simply "N_e").

Absolutely, the symbols (N_e, ω, $\Delta\Delta G$) are now also referred by their conceptual meaning in captions.

-Use (author, year) citations according to the journal format.

The bibliographic format has been changed accordingly.
We also noticed that some references were not appropriately formatted (capitalization, pages, ...), hence we fixed these issues.

-The journal format does not have "Materials and methods" at the end.
I suggest changing this to "Appendix" and referring to it from appropriate points in the manuscript.

"Materials and methods" has been changed to "Appendix"

Reviewer: 1

Latrille and Lartillot outline a theoretical model to examine the response of ω to changes in N_e and expression level through the use of thermodynamics-based genotype-fitness mapping. Overall, I feel this manuscript is well written and introduces an interesting framework. I have included some minor suggestions that I feel would improve the quality of this manuscript.

Page 3 line 42: This paragraph seems disconnected to the previous one, a little rewording might help to resolve this issue. Consider spelling out why it is important to consider a genotype fitness map instead of a distribution of fitness effects is important when thinking about protein-coding sequences.

Indeed, in order to articulate more cohesively the two paragraphs, it has been rephrased to: However, the DFE may not be fixed a priori, instead it may naturally emerge from an underlying fitness landscape and thus vary depending on the position of the population in the fitness landscape (Lourenco, 2013). In an alternative approach, based on genotype-fitness maps modelling the underlying fitness landscape, the selective effect of a mutation depends on the fitness of both the source and the target amino acids involved in the mutation event (Halpern and Bruno, 1998; Rodrigue et al., 2010; Tamuri and Goldstein, 2012).

Figure 1: It was initially confusing why x* was introduced here until I read much further in the paper. It also does not seem to relate to anything you are showing in this figure. You could consider removing that portion of the figure and just explaining what x* is in the text of the manuscript.

Absolutely, the reference to x* is unnecessary and has been removed from the figure.

Equation 6: This feels unfinished. Maybe it is something about the bracket that makes it seems like this should be set equal to some matrix.

The bracket is meant to encompass all possible changes of phenotype after a mutation, this equation has thus been reformatted to make it clearer that it describes the set of all possible probabilities.

Figure 2: There is a lot going on in this figure and it takes awhile to unpack. Maybe adding some axis labels would help. Also add a description about the gray dash lined vs the gray solid line in the legend. Change the text to reflect the correct colors in the following paragraph.

Indeed the figure could be simplified, we made it clearer how the equilibrium phenotype (x*) is changing with N_e and added the description of grey lines in the legend.

Equation 17: I may have missed something but I'm not sure where the first "n" (number of sites) term comes from in the delta delta G n part. It seems like this should just be delta delta G. You may want to double check that equation. If I am mistaken please disregard this comment.

Absolutely, the first "n" (number of sites) was a mistake, and as been removed.

Section 2.5 time to relaxation: This section needs more set up. Its clear from the figure that a lot of simulation work went into coming up with the result, but it would be helpful if a brief overview were giving in the manuscript and leaving the more detailed description in the figure legend.

The section has been updated to give more background on the simulation setup and goal. We also first describe the non-equilibrium behaviours such as to better understand the underlying mechanisms driving the change in ω after a sudden change in N_e. Subsequently, we describe the relationship between parameters of the model and the relaxation time to reach the new equilibrium.

Discussion: It seems strange that a large portion of the discussion section is spent talking about analysis which are described in the supplemental material. Maybe add a short paragraph before the discussion that addresses these findings separately.

Admittedly there are some more developments in the supplemental material not described in the results, but these are not theoretical results, which is the essential of the paper. These developments are just complements, based on analysis of elementary data, in order to enrich the discussion regarding the empirical confrontation. Probably for this reason, after trying to reconfigure the text as suggested, we realized that the text was losing a lot of clarity. So, we prefer to keep all these developments in the discussion, where they find a more natural context.

In general, while reading this I kept thinking back to the assumption that sites contribute additively to delta G. Which is a perfectly fine simplifying assumption, but I was wondering if that might also contribute to why the model and the empirical results do not match. There isn't much discussion on how this assumption would impact findings.

We agree there is not so much discussion about the assumption that sites contribute additively to ΔG. On the other hand, under the model presented in figure 3 (panel D), free energy is computed based on the 3D conformation of the protein and using statistical potentials. As a result, the stabilizing or destabilizing effect of an amino acid at a particular site depends on amino acids present in the vicinity in 3D conformation, thus implementing specific epistasis (the sites contribution to ΔG depends on other sites). However even in this case the theoretical model and the empirical results do not match. As a result we believe

that although it could be a contributive factor, the additive contribution of sites to ΔG is not the sole responsible for the discrepancy between theoretical model and the empirical results.

Reviewer: 2

This manuscript describes a weak mutation approach to purifying selection on a trait determined by n additive loci with identical mutation rates and identical effects on the trait. The authors use these results to explore possible explanations for the relationship between dN/dS and protein expression level.

I apologize to the authors for the delay in submitting this review.

My overall feeling is that this piece is a nice elaboration of a set of influential models. Nonetheless it has some issues that should be addressed prior to publication.

1. The authors argue that the equilibrium value for the additive trait is the value where advantageous and deleterious substitutions are balanced, so that the expected change in fitness with time is zero. Although used previously by others in this literature (e.g. by Joshua Cherry, Richard Goldstein), this argument is not formally correct.

The flaw is that the model being examined is stochastic and not deterministic, and so there is a stationary distribution of trait values rather than a single equilibrium point.

The heuristic of looking for the fitness where deleterious and advantageous substitutions are balanced does produce a good approximation in this instance, but can result in qualitatively incorrect predictions under other circumstances (a good example is Kingman's House of Cards model with a Gaussian distribution of fitnesses introduced by mutation, where the apparent equilibrium point only produces half of the true expected fitness gain at stationarity).

The authors need to make additional arguments to support the validity of their approximation. For one treatment of these issues see e.g. Van Kampen, Stochastic Processes in Physics and Chemistry, Chapter V The Master Equation, Section 8 The Macroscopic Equation.

It is true that our model is stochastic and not deterministic, such that the equilibrium is a distribution of trait and not an equilibrium trait value. We indeed acknowledge that our developments are indeed approximations.

As recommended, we sought to derive the distribution of phenotype at equilibrium in our cases. This derivation is done in two steps, first by deriving the equilibrium at the mutational balance (without selection), second by subsequently deriving the mutation-selection equilibrium distribution. We could provide an approximation in the case of 2 states (K=2) for the mutational and mutation-selection equilibrium (shown in supplementary materials, section 1.3 and 1.6). However the general case ($K > 2$) was not analytically tractable. Moreover, we derive the phenotypic variance in the case of 2 states (K=2) at the mutational equilibrium, which equals $1/4n$ where n is the number of sites. Unfortunately we could not

derive the phenotypic variance in the general case, nor at the mutation-selection equilibrium. However, simulations done under various n shows that the variance at equilibrium is always lower than 1/4n. As a result, for large number of sites, variance is negligible such that the equilibrium distribution essentially reduces to a single point mass.

Altogether, we modified the manuscript to make it clearer that our formal results (and of Cherry and Goldstein) are approximates and not formally demonstrated since they assume that the fluctuations around the equilibrium points are negligible. We refer to the supplementary materials for the simulations and analytical approximations of the variance at equilibrium.

2. More broadly, the authors should better clarify throughout which results (both their own and in the literature) are derived via formal mathematical proof (i.e. have been "proved"), which are approximations, which have been demonstrated numerically, etc. An example of this is line 49 "Conversely, one striking theoretical result was the proof that omega is in fact predicted to be independent of N_e under relatively general circumstance [...] [Cherry 1998]". Cherry argues this (not completely correctly, see point 1), but does not actually provide a formal proof demonstrating that the result is logically entailed under a set of well-defined assumptions.

We absolutely agree with the distinction between results that have been formally demonstrated ("proved"), the results that are approximations, and finally the results that are numerically verified. We edited the manuscript to clearly emphasize the distinctions.

3. Although the authors in general do an excellent job in their treatment and synthesis of the prior molecular evolution literature, they appear to have missed closely related work by Brian Charlesworth considering the case where fitness is a quadratic function of the underlying additive trait:

Charlesworth, B., 2013. Stabilizing selection, purifying selection, and mutational bias in finite populations. Genetics, 194(4), pp.955-971.

While not identical, the Charlesworth paper ultimately produces very similar approximate expressions and conclusions to some presented here. The authors should clarify the relationship between these two pieces.

Thank you for pointing this literature. It is true that we had mistakenly missed this Charlesworth paper. We also agree that it produces very similar approximate expressions (equation 7b in Charlesworth reflects equation 9 in our manuscript) and conclusions (scaled intensity of selection is nearly independent of the effective population and is strongly determined by the logarithm of the mutational bias). However, they are notable differences between the two papers. First, the Charlesworth paper focuses on the equilibrium of traits values in the context of stabilizing selection, bi-allelic loci and a theoretical fitness function.

Instead our genotype to fitness mapping is derived from thermodynamics-based principles in the context of directional selection. Second, we derive the relationship between the phenotypic traits at equilibrium (x*) and the rate of evolution (ω, equation 11) which is an observable of evolutionary trajectory and can be empirically determined by phylogenetic codon models. Third, as noted in (1), we did not focus so much on the equilibrium but mostly on the change of equilibrium after a change in effective population size (equation 12-14) or protein expression level (equation 18). Finally, our work focuses more on the scaling relationship between observable thermodynamic parameters (n, $\Delta\Delta G$, β), population-genetics parameters (Ne) and phylogenetic observable (ω), and how they relate together. Altogether, the Charlesworth paper confirms that some of our findings are more general and are reproducible under a different set of assumptions.

The main manuscript has been updated (in introduction and discussion) to acknowledge the Charlesworth paper and outline the difference between our results and his.

Minor comments:

- Overall. There is a disconnect between the Introduction and Discussion, which are very focused on N_e and omega, vs. the results which are more general. The Abstract mentions this, but it would be helpful to clarify it again in the Introduction, and to maintain a consistent level of generality in the results (e.g. Equation 6 silently assumes amino acids)

Both introduction and discussion has been edited to articulate more cohesively around the results.

For equation 6, we edited the previous paragraph to:
To derive the probabilities of such events to occur, we assume that the state of each site is one of the 20 amino-acids, considered equivalent, and also make the simplifying assumption that all transitions between the amino acids are equiprobable.

- Line 38. Word usage. Allometry is a specific concept different from the more generic concept that omega scales as some power of N_e.

Allometric has been removed.

- Line 208. Marginal stability refers to empirical observation that the delta G of folding for a typical protein is around -10 kcal/mol (i.e. of the same order as a single hydrogen bond) rather than the general concept that there is an evolutionary equilibrium of stability.

The paragraph has been edited to clarify the difference:

Starting from an optimal sequence, mostly destabilizing mutations will occur, some of which may reach fixation and accumulate until selection coefficients against new deleterious mutations is too strong, at which point the protein will reach a point of equilibrium. This equilibrium corresponds to marginal stability (Taverna and Goldstein, 2002; Bloom et al., 2007)

- Line 166 Is this the correct citation? I don't recall these equations from Goldstein 2011.

This is true that the approximations are not from Goldstein (2011), hence we updated the paragraph to clarify the citation.
Analytically, the fitness function is given by the Fermi Dirac distribution (Goldstein, 2011) and is typically close to 1 (Serohijos and Shakhnovich, 2014), leading to a first-order approximation.

- Equation 4. Missing a minus sign.

Absolutely, sorry for this mistake, it has been corrected.

- Line 221. Worth clarifying here that this argument is specific to proteins and wouldn't work for bi-allelic loci.

It has been rephrased to clarify that this argument is specific to proteins and to mutations between the 20 amino acids:
In contrast, stabilizing mutations toward an optimal amino acid are rare, while on the other hand destabilizing mutations from an optimal to a non-optimal amino acid have a low probability of fixation.

- Figure 2, missing division bar, ln((1-x)/x).

The division bar was indeed not displayed in the manuscript but present in the original pdf figure, which is now correctly displayed.

- Line 377. Missing minus sign? (i.e. should the correlation be negative?).

The minus sign is undoubtedly missing in $\hat{\chi}$=0.02, it has been changed to $\hat{\chi}$=-0.02.

- Line 399 "Indeed equimutability . . . " Is this sentence strictly true?

The sentence has been rephrased to:
Indeed, a necessary condition to observe independence between ω and N_e is equimutability, or namely that the distribution of $\Delta\Delta G$ of mutations is independent of ΔG.

- Line 502. What is the tree being referred to here . . . isn't there only a single lineage per simulation?

Indeed there is only a single lineage per simulation, the simulator was originally meant to describe evolution along a tree and was specialized for this manuscript. The paragraph has been rephrased to:

Averaged over the simulation, ω is:

$$\omega = \int_t \omega(t) dt,$$

where the integral is taken over the complete simulation, while the integrand ω(t) is a piece-wise function changing after every point substitution event.

- Figures 3,4. The shaded area is supposed to be a "90% confidence interval". What does that mean precisely? Is it the 90% confidence interval for the mean value of omega (individual observations will be spread much more widely than the shaded band) or does the shaded region actually contain 90% of the simulated values (90% inter-quantile range)? In addition, what is the main source of variability, e.g. stochasticity between runs in the parameters, slow convergence to the true stationary omega across runs, etc.?.

Indeed the wording was imprecise, the shaded region actually contain 90% of the simulated values (90% inter-quantile range). This has been modified in the manuscript. The main of variability comes mainly from the the change in parameters between simulations. More precisely, each simulation has a randomly drawn optimal sequence, hence the optimal amino-acid for each site is different between two different simulations.