




Peer Community In Evolutionary Biology

Understanding genetic variance, load, and inbreeding depression with selfing

Aneil F. Agrawal based on peer reviews by **Frédéric Guillaume**  and 1 anonymous reviewer

Diala Abu Awad and Denis Roze (2017) Effects of partial selfing on the equilibrium genetic variance, mutation load and inbreeding depression under stabilizing selection. Missing preprint_server, ver. Missing article_version, peer-reviewed and recommended by Peer Community in Evolutionary Biology. [10.1101/180000](https://doi.org/10.1101/180000)

Submitted: 26 August 2017, Recommended: 20 November 2017

Cite this recommendation as:

Agrawal, A. (2017) Understanding genetic variance, load, and inbreeding depression with selfing. *Peer Community in Evolutionary Biology*, 100041. <https://doi.org/10.24072/pci.evolbiol.100041>

Published: 20 November 2017

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

A classic problem in evolutionary biology is to understand the genetic variance in fitness. The simplest hypothesis is that variation exists, even in well-adapted populations, as a result of the balance between mutational input and selective elimination. This variation causes a reduction in mean fitness, known as the mutation load. Though mutation load is difficult to quantify empirically, indirect evidence of segregating genetic variation in fitness is often readily obtained by comparing the fitness of inbred and outbred offspring, i.e., by measuring inbreeding depression. Mutation-selection balance models have been studied as a means of understanding the genetic variance in fitness, mutation load, and inbreeding depression. Since their inception, such models have increased in sophistication, allowing us to ask these questions under more realistic and varied scenarios. The new theoretical work by Abu Awad and Roze [1] is a substantial step forward in understanding how arbitrary levels of self-fertilization affect variation, load and inbreeding depression under mutation-selection balance. It has never been entirely clear how selfing should affect these population genetic properties in a multi-locus model. From the single-locus perspective, selfing increases homozygosity, which allows for more efficient purging leading to a prediction of less variance and lower load. On the other hand, selfing directly and indirectly affects several types of multilocus associations, which tend to make selection less efficient. Though this is certainly not the first study to consider mutation-selection balance in species with selfing (e.g., [2-5]), it is perhaps the most biologically realistic. The authors consider a model where n traits are under stabilizing selection and where each locus affects an arbitrary subset of these traits. As others have argued [6-7], this type of fitness landscape model “naturally” gives rise to dominance and epistatic effects. Abu Awad and Roze [1] thoroughly investigate this model both with analytical approximations and stochastic simulations (incorporating the effects of drift). Their analysis reveals three major parameter regimes. The first regime

occurs under low mutation rates, when segregating deleterious alleles are sufficiently rare across the genome that multi-locus genetic associations (disequilibria) can be ignored. As expected, in this regime, increased selfing facilitates purging, thereby leading to less standing genetic variation, lower load and less inbreeding depression. In the second regime, mutation rates are higher and segregating deleterious alleles are more common. Though the effects of multilocus genetic associations cannot be ignored, Abu Awad and Roze [1] show that a good approximation can be obtained by considering only two-locus associations (ignoring the multitude of higher order associations). This is where the sophistication of their analysis yields the greatest insights. Their analysis shows that two different types of interlocus associations are important. First, selfing directly generates identity disequilibrium (correlation in homozygosity between two loci) that occurs because individuals produced through outbreeding tend to be heterozygous across multiple loci whereas individuals produced by selfing tend to be homozygous across multiple loci. These correlations reduce the efficiency of selection when deleterious effects are partially recessive [5]. Second, selfing indirectly affects traditional linkage disequilibrium. Epistatic selection resulting from the fitness landscape generates negative linkage disequilibrium between alleles at different loci that cause the same direction of deviation in a trait from its optimum. Because selfing reduces the effective rate of recombination, linkage disequilibrium reaches higher levels. Because selection tends to generate compensatory combinations of alleles, partially masking their deleterious effects, these associations also make purging less efficient. Their analysis shows the strength of the effect from identity disequilibrium scales with U , the genome-wide rate of deleterious mutations, but the effect of linkage disequilibrium scales with U/n because with more traits (higher n) two randomly chosen alleles are less likely to affect the same trait and so be subject to epistatic selection. Together, the effects of multilocus associations increase the load and can, in some cases, cause the load to increase as selfing increase from moderate to high levels. However, their analytical approximations become inaccurate under conditions when the number of epistatically interacting segregating mutations (proportional to U/n) becomes large relative to the effective recombination rate (dependent on outcrossing and recombination rates). In this third regime, higher order genetic associations become important. In the limit of no recombination, model behaves as if the whole genome is a single locus with a very large number of alleles, becoming equivalent to previous studies [2-3]. The study by Abu Awad and Roze [1] helps us better understand the “simplest” explanation for genetic variance in fitness—mutation-selection balance—in a model of considerable complexity involving multiple traits under stabilizing selection, which ‘naturally’ allows for pleiotropy and epistasis. Their model tends to confirm the classic prediction of lower variation in fitness, less load, and inbreeding depression in species with higher levels of selfing. However, their careful analysis provides a clearer picture of how (and by how much) epistasis and selfing affect key population genetic properties.

References:

- [1] Abu Awad D and Roze D. 2017. Effects of partial selfing on the equilibrium genetic variance, mutation load and inbreeding depression under stabilizing selection. bioRxiv, 180000, ver. 4 of 17th November 2017. doi: [10.1101/180000](<https://doi.org/10.1101/180000>)
- [2] Lande R. 1977. The influence of the mating system on the maintenance of genetic variability in polygenic characters. *Genetics* 86: 485–498.
- [3] Charlesworth D and Charlesworth B. 1987. Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*. 18: 237–268. doi: [10.1111/10.1146/annurev.es.18.110187.001321](<https://doi.org/10.1146/annurev.es.18.110187.001321>)
- [4] Lande R and Porcher E. 2015. Maintenance of quantitative genetic variance under partial self-fertilization, with implications for the evolution of selfing. *Genetics* 200: 891–906. doi: [10.1534/genetics.115.176693](<https://doi.org/10.1534/genetics.115.176693>)

[5] Roze D. 2015. Effects of interference between selected loci on the mutation load, inbreeding depression, and heterosis. *Genetics* 201: 745–757. doi: [10.1534/genetics.115.178533](https://doi.org/10.1534/genetics.115.178533)

[6] Martin G and Lenormand T. 2006. A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* 60: 893–907. doi: [10.1111/j.0014-3820.2006.tb01169.x](https://doi.org/10.1111/j.0014-3820.2006.tb01169.x)

[7] Martin G, Elena SF and Lenormand T. 2007. Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nature Genetics* 39: 555–560. doi: [10.1038/ng1998](http://dx.doi.org/10.1038/ng1998)

Reviews

Evaluation round #1

DOI or URL of the preprint: [10.1101/180000](https://doi.org/10.1101/180000)

Version of the preprint: 1

Authors' reply, 05 November 2017

We thank the reviewers and recommender for their thoughtful comments. We have tried to address all of them as explained in the attached pdf file, and hope that our paper can be recommended on *PCI Evol Biol*. If this is the case, please note that the last name of the first author is "Abu Awad" (not "Awad"). Thank you, sincerely, Denis Roze

[Download author's reply](#)

Decision by [Aneil F. Agrawal](#), posted 21 October 2017

Revise

This is a careful and thorough analysis of mutation load and inbreeding depression in a model with stabilizing selection for species with arbitrary levels of selfing. Undoubtedly, I will recommend this manuscript but I would like to see some revisions to improve the presentation.

Overall, I thought the Discussion was especially good both with respect to summarizing and explaining the results derived here and also in relating them to previous work (especially Lande and Porcher 2015) as well as to data.

This is a very thorough piece of work. The subject matter is dense and both reviewers and I had difficulty following all of it. It is difficult material and there is probably no way to make it all easy and clear. However, perhaps it is possible to do a better job presenting this for the "average" reader of this type of paper, leaving some of the more subtle points to the supplement (i.e., for very serious readers—primarily those that are hoping to extend this work in some way). I would like the authors to think about how best to accomplish this.

In addition to those from the reviewers, I have the following suggestions. The authors needn't follow all these suggestions but I would like them to seriously consider them.

Although I believe Roze has done it elsewhere, it would be helpful to show the relationship between the genetic associations used here ("D"-terms) and the classic association measures that also feature in this analysis (F, and Gij). I suggest doing this following line 201.

The simulation procedure is a straight forward simulation of the system described in the analytical part so I don't think needs to be described in the main text. Put the description in the supplement. The average reader doesn't need to be bogged down by this.

I would relegate the "near neutrality" part (ln 330-345) to the supplement. I would also move the "strong mutation" part (ln 374-384 and associated figures) to the supplement. Both these sections distract the average reader from the more interesting (but difficult) parts of the paper.

Ln 609-624 should be moved to a supplement and replaced with a single line (perhaps in the Methods) saying that allowing for multiple alleles per locus has a negligible effect on the major results (see supplement).

I would echo point 1 from the reviewer who provide the longer review.

Other comments

Going from 20 to 21 confused me more than it should have! Perhaps you could add in the phrase "using the relationship $s_i = \sum(\text{blah blah})$ " [ln 300].

Because Charlesworth and Charlesworth 2010 is a book, please provide reference to specific equations in it.

Figure 2. You might consider adding a panel (since you have an odd number as it is), that shows the fitness function for each value of Q.

Ln 450: I think this should be $D_{i,j} = F D_{i,j}$ (your "D-terms" are reversed, no?)

Ln 458: It is confusing to me that you seem to be saying, we can just use $\rho = 1/2$ because most loci are freely recombining (and this gives us eq. 38). Yet, in eq. 39 it is clear we need to know harmonic mean ρ . What happened to just using $\rho = 1/2$. Please clarify this.

Eq. 40 and Line 478: Does the "2" in that equation come from dominance being 1/4? (Please say so, otherwise the explanation provided doesn't seem to match the equation unless there is a link to recessivity).

Ln 483. I'm a bit confused about this line. Overall, genetic variance is increased by $D_{i,i}$ terms, right? But the $D_{i,i}$ terms are themselves reduced by the $G_{i,j}$ terms, right? But the $D_{i,i}$ terms are still positive (right) so, overall, genetic variance is increased by $D_{i,i}$. Am I correct in assuming that the "reduction" being discussed is the reduction due to $G_{i,j}$ term, not a reduction in genetic variance relative to genic variance? Please clarify.

Ln 488. The "increase [in] the genic variance" is DUE to reduced purging, right? Please say that here rather than waiting several lines to do so.

Ln 501 should it be "where the LAST TWO TERMS IN the brackets..."

Ln 654 (or somewhere) you should probably say that the effects of epistasis depend on U/n because this determines the number of "interacting" segregating mutations.

Reviewed by anonymous reviewer 1, 20 October 2017

This is an impressively thorough analysis of the effect of selfing on equilibria of Fisher's geometric model. I have to admit that I haven't gone through all the calculations or even carefully thought about all of the quantitative results, but everything that I've checked makes sense. I have just two suggestions for the overall framing:

1. I would make clear in the introduction that the model assumes no dominance for phenotype. As mentioned in the Discussion, dominance for phenotype produces some effects that cannot be replicated with just the effective dominance for fitness produced by Fisher's model, and I think it would be good to let the reader know up front that the authors have thought about this.
2. Any thoughts on how one might approach the rare-outcrossing regime analytically? Maybe there's some way to perturb away from the complete selfing case? I'm not suggesting that you attempt it here, but if you have some speculations I would enjoy reading them. I think a lot of populations may be in this regime.

I also have a few very minor suggestions:

1. A table of symbols and definitions would help.

2. In Figure 3, it might be nice to use different shapes for the top and bottom points.

Two typos that I noticed:

- Line 271: "par" should be "per".
- Line 1004: Should be "average number of alleles".

Reviewed by **Frédéric Guillaume** , 03 October 2017

Review of Awad & Roze : "Effects of partial selfing on the equilibrium genetic variance, mutation load and inbreeding depression under stabilizing selection" BioRxiv 2017

In this manuscript, Awad and Roze are asking how much self-fertilization affects the equilibrium genetic variance, mutation load, and inbreeding depression in a population under stabilizing selection for a set of quantitative traits. The model builds up from classical quantitative genetics theory with Gaussian selection on the traits and pleiotropic deleterious mutations. Mutations have additive effects on the traits, and because the mean population trait values are at their optima, all mutations are deleterious. It is worth mentioning that pleiotropic mutational effects are uncorrelated (independent). Finally, the derivations take account of disequilibria caused by epistasis, which is an 'emerging' property of the genotypes that depends on the curvature of the fitness function.

I found the paper very well written, and generally clear. The authors focus on the different kinds of disequilibria brought by selfing: identity and linkage disequilibria. The derivation of the equilibrium genetic variance ensues from the variation of those disequilibria caused by genetic associations. The authors present very elegant derivations for the change in disequilibria and equilibrium variance, load, and inbreeding depression.

The strength of the paper is to show how associations affect the change in homozygosity, and linkage and identity disequilibria in presence of self-fertilization, for which I am not aware of a systematic treatment of the sort. These effects are central to our understanding of how selfing affects the purging, or not, of the deleterious mutations. I will not summarize the results here but they have far reaching consequence on our understanding of evolution of selfed organisms and evolution of the mating system as well. The accounting of these effects is however often hard to maintain, and somewhat blurs our understanding of the overlay of their multifarious effects.

I will propose of few general comments to improve the discussion of the results and some corrections.

1 - as said above, I often lost track of the overall effect of associations on increase or decrease of the two components of the genetic variance and of the homozygosity. For instance about $D(i,i)$ (eq 40), *it may be useful to more clearly state that $D(i,i) < 0$ (or not), for instance on line 470 "This decrease in homozygosity is caused by *negative(?)* identity disequilibria".* also, it would be super useful to actually show, with a figure, how the three disequilibria vary with σ in the different cases, i'd love to see such a figure, it would act as a good summary of the treatment of effects of associations

2 - I found the discussion/treatment of the effects of pleiotropy rather poor. It is worth mentioning that allelic effects are here uncorrelated on the traits, and that (per Turelli 1985 and Bürger 2000, p294) stabilizing selection on the traits is then equivalent to the univariate case (i.e. selection acting independently on each trait, apparent selection is equivalent to actual selection). Then a discussion about what mutational correlations might change would be welcome. Genetic correlations among traits, also due to linkage disequilibrium, are pervasive in nature, this should be discussed. In general, I don't have a good sense of why parameter 'n' is more prevalent than the 'm', it seems to me that the average pleiotropic degree should have more importance than 'n' in the model since the strength of the selection acting on a mutation depends on 'm' and not 'n'. Apparently my intuition was wrong but I can't tell why from the model or the discussion.

3 - I am guessing that the general audience is more used to pure population-genetics treatment of the question of the evolution of the mutation load and inbreeding depression. The reasons why the authors chose a quantitative genetics approach may not seem obvious to all, so is the correspondance between the

two approaches. I'd hope to see a better justification and discussion of the pros and cons of the quant gen approach relative to the pop gen one

a few corrections: p9, line 175: do you mean $U=ul$ or $U=2ul$? clarify if it is the haploid or diploid genomic mutation rate

p12, equation 15: parameter a^2 not introduced yet, only comes on p14

p15, line 297: expression for F should use σ instead of α in $F = \alpha/(2 - \alpha)$