Response to Reviews for: 23 July, 2022

"The role of pseudo-overdominance in maintaining inbreeding depression"
by D. Abu-Awad and D. Waller https://doi.org/10.1101/2021.12.16.473022
Submitted by Diala Abu Awad 04 Jan 2022 12:15

We thank **Sylvain Glémin** for handling our ms and for his response of 25 Feb 2022. He obtained two reviews and generously provided comments himself. We are gratified that all appreciated our work and found it to be of general interest. Below, we respond first to the general points made about our assumptions and approach and the range of genetic parameters explored. We then respond to all the specific points that follow. We indent and color our responses to make them clear here and below.

Sylvain Glémin:

This manuscript presents a theoretical model based on both analytical predictions and simulations that addresses the question of the maintenance of pseudo-overdominance and its possible role in explaining inbreeding depression, especially persistence of inbreeding depression in highly selfing species. As myself, the two reviewers appreciated this work, which could be of general interest for the community. However, reviewer 2 raised several **concerns about the modelling assumptions**, hence the biological conclusions that could be drawn from the results. Both reviewers also asked for **clarification in presentation and terminology**.

We fully address concerns about modelling assumptions by 'going back to the drawing board.' We completely rewrote the paper. It now includes a completely new Introduction (to provide better motivation and more background), new simulations and results (to cover a wider range of scenarios and parameter values), and a new Discussion (to provide better focus and integrate the fuller set of results). Throughout, we further clarify our approach, methods, and use of terms, e.g., substituting "POD zone" for "POD" to disambiguate the phenomenon of overdominance from the genomic region supporting overdominance effects. The paper is substantially improved.

In line with reviewer 2, I think the two main weaknesses of the manuscript are: i) the quite loose hypothesis about the origin of POD regions, and ii) the limited exploration of deleterious mutations parameters.

We are less interested here in how POD regions arise than in their fate once they exist. We explain this better in the current version. We now point out several ways in which POD regions could originate noting that our hybrid F1 model is but one of these. It still serves to demonstrate feasibility via that mechanism and set the stage for analyses of POD region persistence. Note, too, that the hypothesis that pseudo-overdominance arises following hybridization between sub-populations with independently fixed loads it not ours – it was enunciated by B. Charlesworth (2018). We do not consider this scenario either loose or weak.

All models involve assumptions and simplifications. We considered several alternative approaches when we conceived this paper. We chose the particular assumptions and parameter values we did in order to pursue a simple straightforward "proof of concept"

modeling exercise. It was never our intention to explore an exhaustive wider domain of parameter space. Given the dozens to hundreds of load loci, with each potentially expressing different levels of selection, dominance, linkage relations, etc., that is clearly impractical. We regret that our efforts to explain our rationale, model structure, and choice of parameter values fell short of what was expected.

Nevertheless, we have complied with requests to substantially expand the domain of parameter values that we explore. These additional extensive simulations took much time and effort, given that we are dealing with the effects of **many genetic and demographic parameters**: levels of dominance (h), the selective effects of individual loci (several different s parameters), population size (N), the number of POD load loci (n), the background mutation rate (U), the recombination rate (map distance), total genome size (R), and the selfing rate (σ). This is hardly a "limited" set of parameters. We also focused on what we considered to be a reasonable (often based on published) set of values for these parameters. We have little wish to try the reader's attention (and our own time and attention) to explore yet wider ranges of variation in all these parameters.

This additional work considerably strengthens the paper by demonstrating the generality of our conclusions. That is, while we obtain some interesting new results, our previous results are largely upheld and reinforced. They now apply to a wider range of situations that we simulated (as shown by the new Figures 3-5 which we added to the paper). These new and interesting results confirm the generality of our results, further enhancing the value of our paper.

About the plausibility of the scenario for the origin of POD, if I understood correctly, the initial heterosis experienced by the hybrid is very high: the hybrid as a fitness twice as high as the parental population 1/(1-sH). This seems rather unlikely.

We were clear about our specific assumptions and the structure of our models and now explain these more clearly. Crosses between populations often result in dramatic examples of 'genetic rescue' like those reviewed by Tallmon et al. (2004, TREE). The fact that the "hybrid vigor" these F1 clonal lineages "had an <u>average</u> fitness that was **over 35 times** that of resident lineages" clearly demonstrates that assuming a fitness twice as high as the parental population is quite reasonable. We also cite Kim et al. 2018 who note a similar effect, namely that "when ancestry from a larger population is added to a smaller population, the ancestry from the larger population dramatically increases in frequency because it carries fewer deleterious mutations."

Large segregating fitness effects have been found to persist in many species now including self-fertilizing ones. Examples include: *Eucalyptus* (Hedrick et al. 2016); *Drosophila* (Gilligan et al. 2005, Latter 1998); *Zea mays* (Kardos et al. 2016, Brandenburg et al. 2017; Lariepe et al. 2012); *Arabidopsis* (Seymour et al. 2016); *Mimulus* (Brown and Kelly 2020); and *C. elegans* (Chelo et al. 2019, Seidel et al. 2008, Bernstein et al. 2019). Waller 2021 (cited) discusses this evidence in detail, concluding "these studies support the existence and strength of pseudo-overdominant effects in many species." We now cite much more of this literature but assume that the Editor

does not wish us to duplicate the 2+ journal pages of material that Waller 2021 has already reviewed).

The numerical example used is based on some empirical data but is somewhat arbitrary. It shows that many weakly, partially recessive mutations are required, but two highly recessive sublethal mutations can also generate POD, as can be shown with equation 4.

Agreed – "all models are simplistic / arbitrary, but some are useful." We mention prominently the essential similarity in effects between strong / true overdominance at one or a few loci and the muti-locus POD models we present. However, our new simulations reveal an interesting and important effect here: having fewer, stronger load loci destabilizes the POD region (as there are fewer distinct segregants with bigger fitness differences). Note, too, that in addition to the results we present, we explored a far wider set of parameter values in our simulations.

It is not likely that a structure with lethal in repulsion emerge in a selfing species (contrary to blocks of linked weakly deleterious mutations) but it can emerge in an outcrossing species and persist after a transition from outcrossing to selfing. So it would be interested to explore more the parameter space. At least, exploring other values of h and s would help better understanding the conditions under which POD regions are more likely to be maintained (and also to relate the work to Zhao and Charlesworth 2016). Ideally, considering both highly recessive lethal or sublethal and weakly deleterious and less recessive mutations could be useful (or even a distribution of h and s).

Good point –pre-existing recessive lethals linked in repulsion may well persist even as lineages start to inbreed. As noted above, we have taken this suggestion to heart and now present a wider set of parameter values including variation in *h* and *s*.

A side note: one reason it took 2+ years for us to write this paper is that we continually felt the need to explore additional parameter values and scenarios. We carefully selected the particular ones to include in the paper – to represent key effects and conclusions and crucial zones of transitions among outcomes. We would prefer to publish this, admittedly limited and incomplete paper, sooner rather than a more perfect paper with further delays.

Another parameter which is not widely explored is recombination rate. I would guess that change in regime could depend on the ratio r/u where u is the mutation rate towards deleterious mutations.

We initially included simulation results for both tight and loose linkage along the genome (R = 1 and 10 Morgans) in an early version of our paper, but were not fully convinced by the qualitative difference in results. We have now changed how our results are presented, allowing us to include these results. However, we did not analyze these outcomes in terms of the R / U ratio. If there are specific parameter values that the reviewer feels are important to include, please specify those.

Overall I thus recommend to reject this manuscript in the current form I encourage the authors to properly address the different issues, and especially to justify more clearly whether the proposed scenario is biologically possible or not, and under which conditions.

I have also a few minor comments:

L 118 : “This much balancing selection” : strange wording

> We modified this to read: “This amount of balancing selection is required for the POD region to persist.”

Equation 6 : why not use the more accurate expression given by Roze 2015 ?

> The expression provided in Roze 2015 does not account for genetic drift. We had initially wanted to use Roze’s expression, then found that for N = 1000, the expression provided by Bataillon and Kirkpatrick (2000) gave much better results (in Roze 2015, N = 20 000).

L157 : here the term inbreeding load is used whereas it is for inbreeding depression. The term inbreeding load is usually used in a slightly different way as the slope of log-fitness with inbreeding coefficient (see Morton Crow Muller model)

> Inbreeding depression is indeed most accurately quantified as **-B**, the inbreeding load slope we refer to. Inbreeding depression, in contrast, depends on how much the individuals being compared differ in their levels of inbreeding. Commonly, delta compares the fitness of selfed and outcrossed progeny. In such cases, as Keller and Waller (2002) point out, inbreeding loads and inbreeding depressions can be interconverted using the formula: delta = $1 - e^{-B*F}$ (their Eqn IV). We seek to use these terms precisely.

L158 : “A2from” : a space is missing

> Fixed.

Eq 8 : the use of i as a parameter is a bit disturbing at first sight because we can think about a counting index (but this is just a matter of taste)

> We chose to keep the same notation as that used in Kimura and Ohta, allowing readers to easily refer back to the original text.

Review #1 - *Yaniv Brandvain*, *17 Feb 2022 20:18*

In this manuscript, the authors explore how a “pseudo over dominant (POD) block of the genome” evolve, and how their evolution both impacts and is impacted by ongoing mutations at other regions of the genomes. Pseudo-over dominant genomic regions are portions of the genome for which heterozygotes have higher fitness than homozygotes, due to the action of numerous partially recessive mutations in repulsion phase disequilibrium, rather than the action of truly over variants. The authors consider how the answers to these questions can change depending on mating system and population size.

The work has a few take-home messages – (1) POD regions are often unstable (2) POD regions are more stable when they are made of many tightly linked mutations, and/or other regions of the genome, and or heterozygosity is favored genome wide (3) POD regions are less stable when made of fewer more loosely linked mutations in populations are small (or effectively so because linked

selection decreases Ne) and/or more selfing. (4) When POD regions are maintained high inbreeding depression can be maintained – even in largely selfing populations.

On the whole, the work seems correct, and like a nice contribution to the ongoing interest / appreciation of pseudo overdominance. However, I have numerous questions / concerns about decisions with regards to model conception and parameters choice etc. I also recommend the authors consult our recent preprint [Sianta et al. first posted may 2021, reposted Dec 2021 https://www.biorxiv.org/content/10.1101/2021.05.20.445016v2] which modelled the origin of POD, whose results complement this work (In hope to integrate the findings from this paper into ours).

Thank you for the link to this interesting work, which we now cite.

Concerns: *Beginning with a POD region*

The authors start their study with a stylized POD region already existing. Their underlying model is that two isolated populations have themselves accumulated a "local drift load" composed of n equally spaced deleterious mutations per haplotype in some specific genomic region, and with no other deleterious mutations elsewhere in the genome (although they can arise after this initial starting point). The biological basis for such a model is unclear. Harkness, Brandvain, and Goldberg (2019, JEB) began with a similar premise – assuming that isolated populations each fixed their local drift load but did not assume such variants to be highly localized to a single genomic region, and it is unclear why this would occur biologically. My sense is that it would not, and that rather than being strongly biologically motivated, this model is a hack to get at the interesting questions the authors hope to explore. If this is the case, the authors should make that clear, if not, they should provide more evidence that this is a reasonable biological scenario (e.g. perhaps such regions have exceptionally low recombination rates or high mutation rates, such that the load can be purged from other genomic regions, but not these?).

We regret that we did not apparently fully explain the "biological basis" for F1 heterosis between two populations fixed for alternative deleterious mutations. We are now more explicit and clear n the revised manuscript and cite numerous examples to show that such heterosis can be quite large. As already noted, it was Charlesworth 2018 who presented the idea that hybrids between populations subject to drift load may often express appreciable pseudo-overdominance. We took his approach to provide a straightforward and biologically plausible mechanism for POD region origination (though there are others that we also now mention). We now explain our rationale explicitly and cite Charlesworth 2018 and other papers (demonstrating high heterosis) to better justify our approach. We have also insisted in the methods section on the existence a segregating load in the initial populations (before hybridization) for simulations in which $U > 0$, as the burn-in time was aimed at allowing the source populations to reach mutation-selection-drift-equilibrium along the genome. In our initial version, all figures in which we compare heterozygosity along the genome and inbreeding depression with and without POD were made using simulations in which background mutations had also been introduced in the parental population.

We never state that only one predictable circumscribed region of the genome will show overdominance effects. Clearly, many to most genomic regions in F1 hybrids could show such effects. Regions will also surely vary widely in the number of alternate

mutations fixed, their effect sizes (s), dominance levels (h), etc. Given such complexity and uncertainty, we chose to investigate the effects of creating one such region with a clearly defined number of mutations with known effect sizes and dominance, while incorporating background mutations. This presents a reasonable starting point for this simple "proof of concept" paper.

Composition of POD regions

In addition to being surprised by the pre-existence of POD regions, I was confused by the properties of mutations in these regions. (as well as the ongoing mutations experienced by the population). The authors assume a dominance coefficient of 0.2 and a selective coefficient of 0.01. These choices were justified by pointing to the mean values from Agrawal and Whitlock's efforts to estimating the distribution of these values from yeast knock outs. It is not obvious to me why this value is relevant for mutations that could make up a POD region – which should not be random draws from the 2D distribution of fitness and dominance effects – lets alone equal one value in the center of the distribution. A dominance value of h=0.2 is higher than what I imagine would make up PODs generated and maintained by a process other than the one considered in this model – so I worry about the applicability of these parameters.

The values for **h** and **s** we assume were chosen to be realistic in that they reflect values similar to those reported for *Drosophila* and yeast. In reality, we know that mutations span a wide range of s and h values (with s and h negatively correlated). We agree that this variation deserves to be investigated, but doing so in convincing detail would require far more space that we have available in this simple "proof of concept" paper. The models and results we present are already complex. Some consequences are clear. Lower h values, for example, would reduce selection against heterozygotes, increasing inbreeding depression (a well-known result), *enhancing* the emergence of POD's. Thus, we considered an h of 0.2 to be *conservative*. Stronger selection, in general, favors POD region formation by eliminating more homozygotes, allowing fewer load loci to create strong POD effects.

To address this criticism, we substantially extended the range of parameter values and now provide explicit results regarding *s* and *h*, with corresponding revisions to our Methods, Results, and Discussion. These include considering the effects of fewer, stronger, load loci – a situation that can shorten POD persistence by increasing selection differentials and thus the power of selection.

Explanation of results

Often the explanation of the results seems pretty shallow. For example the authors state that "Counter-intuitively, POD heterozygosity persists much longer (relative to neutral heterozygosity) in smaller populations than in larger populations (compare the red and blue lines in Fig. S3 where N = 100 to Fig. 2 and Fig. S2 where N = 1000 and 5000)." I could logic why the authors found this to be counterintuitive, and I can logic an explanation for this result, but neither were provided.

This was a good suggestion. We have substantially revised the paper to provide further explanations and explorations of the many results. This particular result was counter-intuitive because one might expect POD heterozygosity, reflecting selection, to be closer to that for neutral loci (as Ne s is close to 1). This, however, is not at all what we observe. The

> explanation for this result is that selection against individual loci is much more efficient in larger populations, which quickly destabilized the POD zone by increasing the efficacy of purging. The maintenance of POD in part depends on drift acting to reinforce linkage among loci within the POD zone, hence increasing the strength of POD selection. More simply put: In large populations with little drift, selection acts more often on single loci making up the POD zone. In smaller populations with stronger drift, selection acts more against the group of loci within the POD zone.

Limit discussion of the work to what was found

The abstract and discussion suggest that the work is related to the formation and persistence of POD regions, however the methods and results focus only on the persistence. My suggestion is to not discuss the formation of POD regions as it is pretty far afield from what was modelled. It is also somewhat strange to have statements like “it seems unlikely that POD’s would arise in isolated populations” which depend on modelling assumptions of h = 0.2, etc (see above)..

> We accepted and incorporated this suggestion. We did not explicitly model how mutations accumulate in low-recombining regions as our aim was to present a “proof of concept” approach. We now mention other mechanisms that could generate POD regions and explain that we use the F1 population fusion scenario simply to demonstrate one. We also now explore the consequences of variable spacing (linkage) among the load loci to address the criticism that our models only treat artificial ideal conditions for Pseudo-Overdominance to emerge.
>
> The phrase cited refers to the far higher likelihood that POD will emerge from crosses between sub-populations fixed for alternative load mutations than spontaneously within a single isolated population. Isolated populations are likely to undergo bottlenecks, inbreeding, and drift, robbing them of the heterozygosity necessary to allow POD to emerge. That is, POD regions can only emerge when the same individual hosts two distinct haplotypes fixed for different sets of mutations. Admixture thus creates many more opportunities for POD regions to arise. We changed wording to clarify these points.

Enhance clarity of figures, parameter names, etc etc.

I found all the figures and parameters to be difficult to follow. Most increased my cognitive load substantially. I am afraid this could turn readers off and decrease the impact of this work. For example, denoting the population size as N, and the number of loci as n is a bit confusing, and made my life hard. Similarly, the sometimes the “control” (i.e. no POD region) was labelled as “no POD” and other times as s = 0. Direct labelling of figures would make results easier to process, and little changes like have the x-axis as “selfing rate (\sigma)” rather than “\sigma” would make it all easier to follow. Finally, I may be being dense, but I am having trouble seeing if there is any different information in the top (a & b) and bottom (c & d) parts of figure 5

> We modified the figures accordingly. Concerning Fig. 5, in a and b, one could not see that there is a bimodal distribution in levels of inbreeding depression. This is now clearer in the new Figure 5.

*Reviewed by **Lei Zhao**, 21 Jan 2022 18:45*

Review on The role of pseudo-overdominance in maintaining inbreeding

Following several first-order approximations (mainly inspired by Kimura M, Ohta T 1971), this manuscript theoretically discussed the potential role POD blocks (Pseudo Overdominance blocks) could play in maintaining population diversity and inbreeding depression. I am excited about the manuscript and find its results valuable.

The Brief Idea of the Manuscript:

This manuscript uses model studies and simulations to investigate:

1. The mechanisms of maintenance and erosion of the POD block.

2. The interactions between POD and the “background selection” elsewhere.

- Recurrent deleterious mutations to POD block
- POD block to recurrent deleterious mutations

3. POD’s influences on inbreeding depression

The overdominance generated by a POD increases the amount of inbreeding coefficient, and it will reduce the decline in inbreeding depression (\delta) caused by the increase of \sigma.

The model the authors adopted a specific model where mild mutations evenly distributed along the incipient POD in repulsion to alternating mutations on the opposite chromosome. The authors argued this model maximize the effects of POD, and they also tested more general model settings with randomly distributed mutations within the POD.

Comments:

The followings are some thoughts I want to share with the authors and readers:

- In the original literature, Equation 1 is based on the assumption “without loss of generality, s_2<=s_1 and s_2 < 0.5”. And I suggest to write Equation 1 as

σ<2s(1−s)

s1+s2−2s1s2

where $s = min\{s1, s2\} < 0.5$

In the line above equation 1, the authors mentioned with both alleles at a frequency of 0.5, which is not right, as this is only true when s_1=s_2.

> Thank you for pointing this out. It was left over from a previous version. Both points have been modified as suggested.

I do not like the terminology “background selection” across the manuscript.

> We amended this to refer to “background mutations” to avoid any explicit reference to other processes.

1. Since the literature has carefully discussed the impacts of recurrent recessive deleterious mutations: When the (effective) population size/selection is sufficiently large, such mutations will

reduce the neutral diversity, we refer to such effects as background selection; while when the population size is small, the strongly recessive deleterious mutations will help to maintain neutral diversity, such effects are called associative overdominance.2. In several parts of this manuscript, the author observed the opposite effects of background selection, e.g., in lines 345-348, the author wrote “When fewer loci (n) contribute to a POD and linkage is loose, recurrent deleterious mutations speed the decay of heterozygosity.” “In fact, when selfing and mutation rates are high (U = 0.5), POD heterozygosity actually increases”; in Figure S9a, the author observed an unexpected increase of inbreeding depression as \sigma grows. since a larger proportion of selfing is assumed in both cases, the effective population size Ne will be relatively small, given the authors chose s_d = 0.01 and h_d = 0.2 for recurrent deleterious mutations, I highly suspect it is the AOD effects of the recessive mutations that dominate both cases (See Zhao and Charlesworth 2016 for details). Since Zhao and Charlesworth 2016 was based on a higher-order moment iteration, it should be able to give a better explanation than Eqn 6 did for figure S9a. If the authors want to reduce the unexpected increase of POD heterozygosity, I suggest to try h_d=0.5, as it always in the regime of background selection.

3. I believe the suprising results the authors mentioned in the paragraph starting at line 310 can also be phrased and explained using the similar idea as Zhao and Charlesworth 2016.

4. So briefly speaking, in my point of view, the interactions between POD and the recurrent deleterious mutations elsewhere can be summarised as AOD (associative overdominance) and POD collaborate to maintain the diversity, while BGS (background selection) reduces the POD heterzygosity. As the authors indicated, the current hypothesis of the origin of POD is less convincing. AOD might also be the source to build up POD blocks from smaller ones, but this might need to be tested.

> We see your points here and agree that associations, like AOD and POD, are both operating. For the reasons mentioned, we prefer using broader terms like “associations among loci” rather than referring specifically to AOD or POD in discussing complex selective interactions between BGS and POD selection as higher-order associations may also be important. These interactions among loci can either increase He, ID, and the inbred load (by sheltering more of the load or interfering with purifying selection) or reduce these (when selection is reinforcing among loci), making it difficult to generalize. We considerably expanded the Discussion to explore these ideas further. The effects of background mutations do indeed differ depending on the genetic context, as we present and discuss (see new Figures 4 and S7). When the load increases, via increases in the mutation rate or higher levels of sheltered load, protected within the POD zone, inbreeding depression also increases – as expected.
>
> Thank you, too, for suggesting that we simulate effects of co-dominant mutations. We did this and found that there seems to be agreement with the results presented in Zhao and Charlesworth (2016). There is a slight increase in Ne when the background mutations are recessive (h_d = 0.2), but only for selfing populations (Supp. Figs S6 and S7). This implies that identity disequilibria may be a main contributing force. There seems to be a balance between loss of neutral diversity due to BGS and that maintained by AOD, The heterzygosity maintained at the selected loci, both inside and outside the POD zone, likely results from a form of reinforced AOD, as our new results show.

Though “pure” BGS (with h_d = 0.5) had some effect on increasing heterozygosity, most of the increase probably reflects AOD (for h_d = 0.2).

We like the idea that AOD might act to favor the origin or growth of POD blocks and agree that it deserves exploration. Although we do not have enough room to explore this ideally thoroughly, we now go into more detail in the Introduction and Discussion about scenarios that might serve to initiate POD and favor POD zones. Making use of the heterosis model as a basis for forming POD zones, as we do, does not invalidate other mechanisms, as we now emphasize in both the Introduction and Discussion.