We thank the editor and reviewers for their thoughtful comments, which we think have significantly helped to improve the clarity of the paper. We have tried to address all of them as explained below.

**Editor's comments:**

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.

This has been done as suggested (lines 202-212).

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.

Done

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.

These approximations have been moved to Supplementary File S2; we have also simplified Figure 1 and the associated discussion (lines 303-336).

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).

Done

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

We have added a figure (Figure S6, mentioned lines 477-479) showing the different components of the genetic variance

Other comments

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

Done as suggested

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

We changed this reference to Charlesworth & Charlesworth 1987

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.

A panel showing fitness as a function of the scaled distance from the optimum for different values of Q has been added as suggested.

Ln 450: I think this should be Di,j = F Dij (your "D-terms" are reversed, no?)

Yes the terms were reversed, this has been corrected.

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 rho. What happened to just using rho = ½. Please clarify this.

We have added extra explanations on lines 412-419, with a new supplementary figure (Figure S5) showing the average identity disequilibrium for different map lengths

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).

Yes, this is now said on lines 435-436

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

Indeed, yes, the overall effect of Di,i is to increase the genetic variance. However, the indirect effect of Gij terms will be to decrease Di,i, thus resulting in a lower genetic variance than without identity disequilibria. This paragraph has been re-written so as to clarify this point

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.

Done

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

This has been corrected

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.

Added as suggested (lines 608-609)

**Reviewer 1's comments**

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.

Done (lines 130-132)

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.

We agree that it is desirable to find general analytical approximations for high selfing, but unfortunately we didn't find a way to do so. The main challenge is to deal with higher order genetic associations, which probably have important effects in this regime.

I also have a few very minor suggestions:

A table of symbols and definitions would help.

Done

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

Done

Two typos that I noticed:

- Line 271: "par" should be "per".
- Line 1004: Should be "average number of allele**s**".

Corrected

**Reviewer 2's comments**

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 \sigma 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

The new Figure S6 (mentioned lines 477-479) shows how the different components of the genetic variance change with the selfing rate

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.

We have added a sentence on genetic correlations in the discussion (lines 692-695), and tried to better explain the effects of n and m (lines 608-612, and 684-689). It is true that the strength of selection s acting on a mutation depends on m rather than n, but the load doesn't depend on s in the deterministic regime, while the variance of epistatic effects (and the probability that two mutations have compensatory effects) depends on n

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

As mentioned in the 3rd paragraph of the introduction, an advantage of the quantitative trait approach is that it provides a natural way of generating distributions of selection and epistasis coefficients, that seem compatible with the available data. We now mention in the discussion that we recover results from previous population genetics models (without epistasis) when n tends to infinity, as the variance of epistatic effects (on fitness) then tends to zero (lines 609-612).

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

Done (line 183, Table 1)

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

a^2 is introduced on line 179, and now appears in Table 1

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

This has been modified