1. Page 3: “The possibility that neutral variants are affected more strongly by selection at linked sites than by genetic drift, even in organisms like Drosophila, is one possible explanation for the ‘paradox of variation’, the fact that genetic diversity and population size often do not scale linearly”.

How do the authors reconcile this suggestion with the fact that natural populations of *Drosophila* show very small levels of LD? (e.g. Mackay et al. 2012)

2. Pages 3-6: A long discussion of analytical math is provided in the introduction only to conclude that it is not applicable to the scenario investigated in this paper - which only uses simulations. The math section should be shortened significantly or moved to the supplement.

3. Page 6: It seems like only beneficial mutations are modeled in the background loci. The reality is that most background mutations are deleterious, so beneficial focal mutations are more likely to be “dragged” down by linked deleterious alleles than “drafted” up by beneficial ones. Is it realistic to completely ignore negative selection on linked sites?

4. Page 6: “the population experiences a constant influx of beneficial mutations at the background loci. These mutations’ fitness effects are drawn from an exponential distribution”. In page 4 (mathematical description of the system): “If x is the background fitness of a lineage, then the distribution of background fitnesses f (x) is assumed to be roughly Gaussian”. How can sampling from an exponential distribution produce a Gaussian background? Any part of the mathematical discussion that does not pertain to the assumptions of this study should either be removed or explicitly mentioned to be not relevant to simulations done here.

5. Page 6: “any time a locus becomes monomorphic, a mutation at that locus is injected into a random individual in the population. In this way, the population experiences a constant influx of beneficial mutations at the background loci.” This practice seems to me to mask one of the important effects of population size: the total number of mutations introduced every generation. So, it does not seem appropriate at least when the question is the effect of Ne on likelihood or time of valley crossing.

6. Pages 7-9: Do all simulations always successfully cross the valley? If not, the rate of success should be reported in addition to the required time.

7. Page 7-9: As the authors discuss, when σ is high, the effect of linked background loci might be more important in determining the fate of a lineage than the focal loci. A question then arises: what is the distribution of background fitness of the lineages that make the crossing? Do only focal mutations on favorable backgrounds cross the valley? If all of them cross, then a plot or summary statistics of dependence of τ on the lineages’ background fitness (averaged over loci and time to simplify) would be informative.

8. Page 8, right column: “In Figure 5, we set $s_{sweep} = 0.01$, $s_{sweep} = 0.1$, and $m = 10^{-5}$”. Should be “$s_{valley} = 0.1$”.

9. Page 8: “Figure 5b shows the expected pattern that deeper valleys in smaller populations take longer to cross”. This seems contrary to earlier arguments for easier crossing in smaller populations due to weaker selection on deleterious intermediates. Please clarify.

10. Page 10: The authors mention the contrast of their results to those of Ochs and Desai. Can they offer any explanations as to the reason?