

This manuscript is a simple, intuitive explanation of when parallel adaptation can be more likely than colonization via migration on patchy landscapes. I think the question is interesting and that the authors successfully capture the important points of the answer. I very much like the goal of writing a theoretical paper that is useful and accessible to empirical researchers. As much as I like this manuscript, I am going to recommend changing it quite a bit. Briefly, I think that the derivation of the main results (12), (21), and (25) can be made much shorter and easier to understand (for both empiricists and theorists), and that there need to be simulations that quantitatively confirm these results. Overly-detailed comments are below. (All page numbers, etc, refer to the corrected version of the manuscript.)

1. I think that the main result (12) can be derived much more simply, and also extended. By writing that the probability that a family that arrives at the patch is successful is approximately  $2s_p\mathbb{E}[K]/\xi^2$ , the authors have implicitly assumed that this quantity is  $\ll 1$ . But this implies that the fluctuations in the number of  $B$  individuals arriving at the patch can be ignored, since even a large fluctuation does not saturate the probability of success. This is valid for  $s_m \gg s_p$ . Given that the fluctuations can be ignored, (12) can be derived immediately from  $q(y)$  – simply apply Fick’s Law to  $q(y)\rho$  to get that the flux of new  $B$  individuals is  $\sigma^2|q'(y)|\rho/2$ , and then multiply by the success probability  $2s_p/\xi^2$ . (Actually, the result is slightly different from (12) – see X below.)
2. However, I think we should also care about the opposite limit,  $s_m \ll s_p$ . In this case, the fluctuations do matter – arrivals typically occur in clumps (families) large enough so that success is almost certain conditioned on arrival. Since on average each family consists of  $\mathbb{E}[K]$  individuals, the family flux is just the individual flux divided by  $\mathbb{E}[K]$ .  $\mathbb{E}[K] = \xi^2/(2s_m)$ , so the result is the same as above, except with  $s_p$  replaced by an additional factor of  $s_m$ . (This could be made more formal, but I don’t think that’s what the authors are going for.) In the intermediate case  $s_m \approx s_p$  which might be the most relevant for the mice<sup>1</sup>, the two limiting regimes cross over smoothly. So for the first limit ( $s_m \gg s_p$ ), all you need is  $q$ , and for the second, you only need to add basic facts about branching random walks (ie, limiting number distribution and localization at long times).
3. As it stands,  $\lambda_{\text{mig}}(y)$  is the rate of successful migrants at an infinitesimal area at distance  $y$ , so it needs to be integrated over an area to get the rate for a whole patch. The authors don’t do this, and as a result, expressions like (14) are dimensionally inconsistent. It’s a little annoying to define the right area of integration, but it’s straightforward to argue that up to a constant the answer should be given by integrating  $\lambda_{\text{mig}}(y)$  over a region of thickness  $\sigma/\sqrt{2s_m}$  around the boundary of the patch. I think it’s easier

---

<sup>1</sup>Although probably  $s_p > s_m$ , since these are effective parameters including density-dependent competition, which I guess is weaker in maladapted patches.

to work with fluxes and just integrate over the boundary of the patch, with no worries about thickness, so that's why I used them above. (Of course, you get the same answer either way, except for small or weirdly-shaped patches, where either or both can break down.)

4. I think that the authors need to show simulation results confirming the analytic approximations. Spatial simulations are a pain, but it's important to have a (somewhat) independent check on the math and assumptions. At a minimum, I think that there needs to be a figure showing that they have the right expression for the rate of successful migrants (12) and migration sojourn times (17) over a range of parameter values. If it's not too much work, I think it would be really nice to have a figure that doubles as a kind of graphical summary, showing a scattering of patches in two dimensions, with one starting as the homeland of the  $B$  allele and the rest colored according to their probability in simulations of being colonized by that original allele vs new mutations, and a curve marking the region predicted as being colonized by the analytical approximations, or something like that. But I understand if the authors don't want to spend their time on that.
5. It would be nice to have a table of symbol definitions.
6. And speaking of symbols, I think it's a bit nicer to write things in terms of the characteristic patch length  $a \sim A^{1/d}$ , rather than the area. This avoids having to talk about one-dimensional areas, and, since the ratio of the boundary volume to the total volume of a patch scales like  $\sigma/\sqrt{2s_m}/a$  in any dimension, it's the quantity that naturally shows up in formulas.
7. It would be nice to have a clear, concise summary of the quantitative assumptions needed for the results (all the stuff like  $a \gg \sigma/\sqrt{2s_m}$  or whatever).
8. In "Length of the hitchhiking haplotype", I would focus more on the asymptotic expressions for  $R \gg w \gg \sigma/\sqrt{2s_m}$ , since that's the regime where the analysis makes the most sense. Then the expressions become simpler and you can treat  $d = 2$  as well. If you want, you could even lead with the intuitive explanation at the end of the section, and save the Bessel functions for an appendix.
9. I think (24) follows directly from Barton, Heredity 1979. (If you look at the unnumbered equation above Fig. 1 in that paper and plug in  $x \sim a$ , you find the time scale  $r_{\text{eff}}$ .) I think that the displayed equation should just be  $r_{\text{eff}} \approx r\sigma/\sqrt{s}/a$ . (I feel like the extra correction terms and geometric factor go against the spirit of the paper, and in any case are probably small compared to the inaccuracies introduced just by applying the model to any real system.) I think you can cut essentially everything before the simple explanation just above (24); if people want the mathematical details, they can read Barton. If I am misunderstanding things and there

is a new result here, rather than just a good new exposition of an existing result, there should be a figure showing simulations confirming it.

10. In “Applications”, the authors spend a lot of time introducing the mice and calculating the values for Fig. 6, but then they don’t show where the actual data lie on those plots. Is there enough information to do some kind of comparison? How close are the nearest samples to each other? Have no examples of alleles spreading by migration been found? If not, how much of a bound does this put on the parameter values? Is there any data on the genetic map length of the region of reduced diversity around the selected locus?
11. A few sections or passages seem out of place:
  - A. I think “Simulation methods” should go in the “Materials and methods” section, or at least not right between “Establishment of a locally adaptive allele due to mutational influx” and “Establishment of a locally adaptive allele due to mutational influx”.
  - B. The sentence on p7 beginning “By our definition of fitness...”.
  - C. The paragraph starting “Also note that...” on p11.
12. Fig. 4:  $R/\sigma$  should have dimensions of  $\sqrt{\text{time}}$ . In the caption, I think it’s misleading to say that  $\lambda_{\text{mig}}$  is “relatively insensitive to  $s_m$ ”. It’s exponentially sensitive! I don’t understand the blue numbers in the right-hand panel. It might be easier to interpret these if they were contour plots of the total time to adaptation, colored according to the value of (14), although I’m not sure. Also, they aren’t really “phase diagrams”.
13. Top of p11 (just above “Multiple patches”): should be  $\sqrt{2s_m}/\sigma$ .
14. Between (49) and (51): should be “Green’s”.
15. (51)-(55) are missing factors of  $\sigma$ .
16. In (56) and following, it would be clearer to use different symbols for scaled and unscaled distances.