

The authors present new results from a previously published experiment on range expansions in microcosms using a parasitic wasp to examine the effects of connectivity. In their previous work, the authors showed that the different connectivity treatments resulted in different expansion dynamics, as quantified by the degree to which the expansions were pushed vs. pulled. In the new data and analyses, the authors attempt to link these observations to patterns in trait evolution by measuring a suite of traits in wasps collected at various points throughout the experiment.

I found this manuscript well-written and very intriguing. The question put forward is highly relevant and important to current research on eco-evolutionary dynamics in range expansions. For the most part the authors describe their experiments and statistical analyses well (see a few minor comments below) and the analyses are appropriate and well justified. My only concern with this manuscript is one that the authors appear to be well aware of, which is the difficulty of drawing meaningful inference from the low sample sizes used in the experiments quantifying the density dependence of fecundity and dispersal.

My concern regarding this arises for two reasons. First and most obvious, it is extremely difficult to draw meaningful interpretations on trait evolution from a single replicate per connectivity X strain treatment. I realize multiple wasps were taken from core and edge locations of these landscapes, but they only represent a single replicate of the evolutionary processes underpinning these range expansions. The authors are aware of this limitation and mention it several times (with a particularly nice discussion of it on lines 389-398). However, there are also several instances where the authors take these results and interpret them more broadly in the context of the behavior of edge or core populations broadly (e.g. lines 376-379, 384-389, and 429-431), which is a difficult assertion to justify given the small sample size. My second concern regarding the small sample sizes is that evolutionary dynamics in range expansions are known to be strongly affected by neutral, stochastic forces (i.e. gene surfing). This means that alleles can rise to high frequencies (or be lost) in edge populations purely by chance in range expansions, making it even more difficult to draw inferences on trait evolution. Gene surfing increases the importance of replication in these studies as it makes distinguishing evolutionary changes due to adaptation vs. neutral dynamics very difficult in a single range expansion.

This is not to say that the authors should not report on their results, as they do provide intriguing directions for future research. However, I would suggest the authors use more careful language throughout their manuscript as this experiment's ability to make broad inferences on trait evolution in edge populations at varying connectivity levels is a bit low. Below, I have included a few minor comments with line numbers.

Line 83: The authors switch to "spatial selection" here from "spatial sorting." This is in line with existing definitions of spatial selection vs. spatial sorting (e.g. Perkins et al. 2013 which the authors cite), but the authors should perhaps include an explicit definition if they are to use this term so that casual readers do not conflate the two.

Line 101-105: This is the first time the authors introduce the idea of ecological context affecting trait evolution during range expansions. As ecological context could refer to a wide variety of things (e.g. biotic interactions), I suggest the authors define what they mean by ecological context and/or use more precise language here.

Line 169-172: Were the 8 landscapes truly chosen randomly (i.e. was there no attempt to balance among strains or connectivity levels)? In general, I have this same question for most of the traits measured.

Line 219-222: The authors should state whether it was an edge or core population that was excluded here and why (presumably low numbers).

Line 227-228: How did the 47 populations divide into 99 groups? Also, which four populations were excluded and why? I realized this is asking for a lot of details, but given the low replication of this experiment (4 replicates per strain x connectivity combination) these numbers of excluded populations could make a difference depending on how they are distributed.

Line 252-255: I'm again having trouble figuring out how these numbers add up to 72.

Line 286-291: Is the choice to not break out results for p and λ separately driven by a parameter identifiability issue or simply because it does not add any new information? I am curious as it seems possible that experimental factors could impact these two variables differently, so it would be interesting to know if the authors are able to show how they each respond to connectivity, location, strain, etc.