In this study, Deshpande and Fronhofer investigate the evolution of the density-dependent dispersal reaction norm under two models that assume different genetic architectures for it. In the first model (the "RN" model), the reaction norm is assumed to be a threshold function, the shape of which has been derived from first principles by previous authors. There is no dispersal below a threshold density, and a saturating increase in the probability of dispersal above it. Here, only the threshold value evolves, and the overall shape of the reaction norm is fixed. The second model considers a more complicated, mechanistic genetic architecture for the norm, that makes no assumption regarding its shape (the "GRN" model). Instead, the reaction norm emerges from a gene-regulatory network inspired by the Wagner model. The network behaves analogously to a recurrent neural network, that takes either the local population density or local population density and the sex of its bearer as input, processes those inputs, and outputs a dispersal probability. In this second model, the weights, slopes and thresholds of all the genes in the network evolve. The authors first consider equilibrium metapopulation dynamics, and compare the reaction norm evolving under the GRN model to the one obtained under the RN model (which is known to be the optimal one). They find that the GRN is able to generate the optimal reaction norm with reasonable accuracy over the population density values that frequently occur in the simulations, even with a small number of genes in the intermediate layer (n=4). They then let the population undergo range expansion, and show that the GRN leads to more efficient colonisation than the classical RN model because the GRN retains large amounts of standing variation for densities that do not occur under equilibrium conditions, allowing it to react more efficiently to selection. Overall, I found the study to be quite interesting but I think it has a few shortcomings, and could be strengthened in a few ways. Please find my comments below.

Introduction

The Introduction is very long and difficult to follow. I think it could be much shorter and to the point. Here is a few suggestions.

- The introduction contains a lot of information that is not directly relevant to the study being presented, which causes the reader to be confused as to what the study is going to be about. For example, the sentence starting on line 25, about dispersal not being random and being associated with other traits does not seem useful to me in the context

of the study. Similarly, I did not get the point of the discussion of the Harman (2020) on lines 34 to 38. The discussion of Li and Kokko (2019) on lines 38 to 41 is not helpful either, because the authors are not addressing any of the problems identified in this previous study.

- The text is repetitive on a few occasions. For instance, the first paragraph finishes with the statement that the mechanisms underlying dispersal plasticity have not been modelled, and the second paragraph begins with 4 lines (46 to 50) essentially repeating the same thing. The paragraph on lines 64 to 72 says the same thing again: a lot of work has been done on density-dependent reaction norms, using different approaches, but none of the previous studies have considered the underlying mechanisms.

- The Wagner model should either not be explained in detail in the introduction (leaving explanations for the methods section), or the description given on lines 74-76 should be expanded and made less abstract.

Model description

- l. 129: Mention at this point that each cell of the grid corresponds to a patch where the population follows a Beverton-Holt model. In fact, it might be better to introduce the landscape in one sentence, and then rather than describing the approach taken to analyse the model (which can be done in the "Analysis" part at the end), immediately explain the life cycle.
- l. 131: typo "plasticity"
- l. 137: explain what "periodic" boundary conditions are.

- The authors need to give the readers more explanations on the Wagner model, especially since it is the heart of the paper. The authors should walk us through eq. (2) and how an iteration of the network happens.

- l. 194-195: What does discarding the GRNs that do not reach fixation mean? Are the corresponding individuals killed? How should we interpret this biologically?

Results

- Based on the visual inspection of Figures 2 and 3, the authors claim that the GRN is able to accurately produce the optimal reaction norm. While it does seem to perform reasonably well for some parameter values, comparison of the fit of a cloud of points with a curve is usually not a very good way of representing goodness-of-fit. Why not average the norms produced by the 1000 GRNs shown, and plot this averaged norm as a second curve? This would give a much more convincing depiction of how well the network actually performs. It could be interesting to do this over the whole range of densities and not just the ones that most frequently occur, as it could help to prove the next result on the network performing well only at the densities that actually occur in simulations.

- This latter point is intuitive, but could be made in a more convincing way. For instance, the histograms showing the occurrence of the densities could be moved to the main text, accompanied by a plot showing (for instance) the variance in the dispersal probability given by the 1000 networks as a function of the frequency of occurrence of the considered density. I expect lower variance to be found where the network was able to optimise.

- The network seems to be performing much better for some combinations of ϵ and μ than others, but the authors do not discuss this. If it is possible, I would be curious to get some explanation as to why that is.

- Have the authors tried to use networks with different numbers of genes involved, to see how it affects accuracy for instance? On that note, why did the authors decide to stop iterating the network at I=20?

- The part of the results on range expansion could be improved: the authors have not really proven what they claim to be the mechanism (high standing variation for some population densities) producing faster expansion in the GRN model. Perhaps a plot showing the phenotypic variance over the range of population densities, as suggested above, could be useful.