Solman (2020) Microbes: Is everything everywhere?

To do list:

1. Recover parameters from single archipelago simulation.

* Why does ms = m0/sqrt(area)? If m0 = rate of immigration per one unit area, shouldn’t the overall island immigration rate (ms) increase with island area? The likelihood of any one individual being replace by an immigrant should increase. Why does immigration rate decrease with island area?
* This is assuming that one unit area = one individual – that’s why it’s important to know how many individuals you have per unit area!
* If ms = 0.05 on an island of 100 units m0 = 0.05\*sqrt(100) = 0.5
* If ms = 0.05 on an island of 1000 units m0 = 0.05\*sqrt(1000) = 1.58
* So the 0.05 that I’ve been putting into the simulation actually corresponds with the likelihood that ANY individual on the WHOLE island will be replaced with an immigrant. 0.05/sqrt(100) = 0.005 = m0 = ms/sqrt(area) < shouldn’t this be the right way round? When I changed the equation, I came closer to retrieving my parameters. Email Ryan to clear this up.
* How to get best initial parameter estimates? Tried to implement the equations in the paper but when using the lambert W function keep getting inf- and NA results.
* Should I log the data before or after the fitting?

2. Recover parameters from multiple archipelago simulations. 100 archipelagos with these parameters:

|  |  |
| --- | --- |
| Migration Rate | Number of Niches |
| 0.005 | 5 |
| 0.01 | 10 |
| 0.015 | 15 |
| 0.02 | 20 |
| 0.025 | 25 |
| 0.03 | 30 |
| 0.035 | 35 |
| 0.04 | 40 |
| 0.45 | 45 |
| 0.5 | 50 |

* Run model fitting on each of the 100 archipelagos
* Make plots for m\_rate, niches and theta, with true value on x axis and recovered value on the y axis.

3. Show graph results to James/Ryan/Tom. Is the fitting process systematically over/under estimating any of our three variables?

4. For each dataset I have collected, does it meet the appropriate conditions of this model? Does it have area data for “islands” that can be immigrated to from source populations?

5. For each relevant dataset get estimations of per unit area population density. This will be the value of rho. Either collect these from the studies themselves, or find relevant information from the literature. Modify the model function, so that it can take a changing value of rho for each fitting.

6. For each relevant dataset, find initial parameter estimations – ask Ryan the best way to do this. This may be specific to microbial communities – ask Tom about this.

7. For each relevant dataset, run the model fitting procedure for each value of K. This probably won’t work for datasets that use diversity index values instead of OTUs or observed species. Diversity indexes may have to be back transformed, or more data acquired from the authors or the datasets excluded.

8. Use the best-fit parameter estimates to generate plotting points and estimate the critical area of each dataset. Am I estimating critical area appropriately? Plot these.

9. How can we quantify how isolated our “island archipelagos” are? Classify them into archipelago types and try and quantify isolation. Also, quantify the dispersal ability of the taxa.

10. Find a way to standardise the area variable, so that we can compare critical area across data sets. Perhaps we can only compare among certain datasets.

11. Run a multiple regression analysis on the successful fits with log(critical area) as the dependent variable, and archipelago type and taxonomic group as categorical explanatory variables. Does the multiple regression satisfy normality and homoscedasticity assumptions?)

12. What is the mean R^2 score of my fits? What is the mean adjusted R^2 score? What is the standard deviation of the R^2 scores? What is the range of the R^2 scores? Put the full data in supplementary materials and a summary in results.

13. What is the median value of the fitted fundamental biodiversity number, immigration parameter, number of niches? What is the most variable parameter across datasets?

14. Is there a correlation between the four model parameters (m, theta, K and the independently estimated parameter rho)? Try using Spearman’s rank to do this. Which parameters were most strongly correlated?

15. Are there differences in the critical area between archipelago types and taxonomic groups? Show these differences on a chart. Did the multiple regression show that critical area was smaller for taxa that have higher dispersal rates and archipelagos that are less isolated? What is the average difference between taxonomic groups and archipelago types? How much do taxonomic group and archipelago type explain the variance in log critical area?

16. Write up results.

17. Write up discussion.

18. Write up conclusion.

19. Go back and check introduction.

20. Go back and check methods.

21. Send the supervisors for comment.

22. Make everything as perfect as it can be and submit!