Summary remaining analyses Sept 26/2018

Initial analysis under question:

Contribution of parallel natural selection (ecology) and geographic proximity to repeated genomic differentiation (FST). Ecology and geography quantified using binomial and continuous variables.

Initial problem:

Original analyses were run by transforming FST matrices into vectors and as a result did not account for the fact that FST was estimated between all pairs of populations so each one is included in multiple comparisons.

Initial solution:

Multiple regressions of distance matrices. They are an extension of Mantel tests that allow for the inclusion of multiple matrices and control for pseudorep noted above. They’re better than straight Mantels for a few reasons, one being that you can quantify the relative strength of associations while accounting for any correlations between our matrices. MRMs also provide you with information on the overall fit of your model, rather than just correlation coefficients for each matrix that is included (https://onlinelibrary.wiley.com/doi/epdf/10.1111/evo.12134). Could also run redundancy analyses here. They’re better than MRMs because they avoid Type I error that is a problem with Mantel tests. I do not have time to explain this. These runs can be found here: mrm\_stick\* (binomial) and mrm\_stick\_qual\* (continuous).

Second problem:

We realized there was more pseudorep in there as we have lakes in the same watershed that will undoubtedly have made use of the same variation.

Second solution:

Subsample one population from each watershed. Do this 1000 times so it doesn’t look like we cherry picked. You can see the results for the initial runs here (binomial) (quantitative). This approach was agreed on upon by all co-authors so I’m not sure why it’s under question now. If someone has a better idea that’s totally fine but I do not have time to run anything else. These runs can be found here: mrm\_stick\_mult\* (binomial) and mrm\_stick\_mult\_qual\* (continuous).

New analysis under question:

Repeated genomic differentiation with and without shared divergent selection. I personally didn’t realize we were also redoing this one but that’s probably my own failing. In any case I think the solution would be to run a Mantel test as there are only two matrices here. The problem is that Mantel tests (and their derivatives) require complete matrices. As Diana pointed out the file has been filtered: “In the comparisons where the same population are on both sides of the fst calculation e.g. when Boot\_Lake\_Roberts\_Lake is compared to Boot\_Lake\_Misty\_Lake there will automatically be a higher correlation coefficient because more of the same site will be divergent due a higher fraction of the sites being compared being the same - but this pattern is completely an artefact of having the reference population be the exact same on both sides”. The filtering is much more drastic than in the comparisons run above and we would lose all the non-divergent comparisons I think. Could someone with more knowledge of the system please confirm this? You can check this out yourself: mrm\_stick\_fst\_comp\_v2.Rmd