Elaine Shen

February 12, 2019

BIO 594

Discussion Paper 1:

Landscape Genomics

In order to measure the magnitude and speed with which populations are able to respond to environmental stress, it is necessary to understand the standing genetic variation present among subpopulations that is able to facilitate phenotypic plasticity, adaptive selection, and organisms’ abilities to successfully range shift and exchange genetic material with other subpopulations. With the rise of computationally-extensive genomic data and multivariate statistical and modeling approaches, it has become easier to infer exactly which environmental parameters and ecological data are contributing to the distribution of populations’ variation across diverse natural gradients. Brauer et al. (2016) and Hancock et al (2019) use single nucleotide polymorphisms (SNPs) across geographically diverse aquatic and terrestrial conditions in order to examine a species’ ability to respond to future environmental stressors. As expected, both papers found loci where soft/partial selective sweeps were predicted to be occurring and tested the significance of these loci against putatively neutral loci with a variety of experimental and statistical tests.

Brauer et al. (2016) hypothesized that the steep hydroclimatic gradient of the Murray-Darling Basin has contributed to adaptive divergence of an endangered specialist fish, the southern pygmy perch (*N. australis*). After filtering for outlier (non-neutral) and neutral loci, calculating effective population sizes, converting geographical distances to more biologically relevant distances using multidimensional scaling (MDS), and removing collinear environmental factors/creating synthetic environmental variables using principal component analyses (PCA), Brauer et al. (2016) used genotype-environment association approaches (GEA), which is able to separate geographic and environmental effects and account for the effect of shared population history. In particular, they tested spatially-explicit generalized linear mixed-models, where the spatial effect is a random effect in both models they tested where either the environmental variables of interest had a fixed effect on population allele frequencies or no effect at all. To check their PCA results, which indicated that temperature, flow, and topography were explaining a large majority of the variance, Brauer et al. (2016) also utilized partial redundancy analysis (RDA) to create the final model where it was confirmed that temperature, rainfall, and topography were major environmental factors in this system, while human disturbances only affected local variation). Overall, the authors were able to disentangle the effects of drift and hydroclimatic-driven selection, noting that transplanting individuals from the Lower Murray to other areas as a way to simulate natural gene flow across these now fragmented waterways may have positive conservation impacts.

On the other hand, Hancock et al. (2019) utilized genome-wide scans and validated their variants predicting the fitness of those variants on a common garden with the generalist plant and widely-distributed *Arabidopsis thaliana*. Instead of GEA, Hancock et al. (2019) used a partial Mantel test to determine the correlation between a given SNP and their pruned non-autocorrelated environmental data (based on Pearson r2 coefficients). While the partial Mantel test was also able to control for population structure, it was unable to disentangle spatial effects as explicitly as GEA. However, Hancock et al. (2019) justified using kinship matrices, where distance matrices of phenotypes or individual genetic variants were the response variables to a pairwise distance matrix of the climate variable, rather than linear modelling approaches because the latter tended to identify variants that were affected mainly or only by climatic outliers. The spatial component of the phenotypes and environmental data were overlaid onto a geographic map for a more qualitative picture of their distributions. The advantage that Hancock et al. (2019) had was that they were able to isolate 107 ecologically important phenotypes, which gave their results more interpretive power. In addition, rather than rely on multiple statistical and modeling tests like Brauer et al. (2016), they were able to raise a geographically diverse group of *Arapidopsis* individuals in a common garden and successfully predict their fitness given the conditions of the common garden relative to the individuals’ geographic origin.

Overall, Brauer et al. (2016) gave a more convincing argument for the relative effects of random drift versus adaptive selection in their study system because they utilized a variety of approaches to confirm which environmental variables were driving their outlier loci in comparison to their neutral loci, the main strength being their spatially-explicit multivariate approaches. Hancock et al. (2019) cleverly circumvented some of the more computationally intensive modelling tests by confirming their initial predictions with an experimental plot, which was similarly convincing but did not account for random spatial effects. Brauer et al. (2016) noted that *N. australis* lived under more complicated environmental regimes and therefore it was unlikely they could have experimentally confirmed their results, though it is plausible that future transplantations could provide the basis for a natural experiment where changes in adaptive genetic variants in the Upper Murray cachements are monitored over time.

Works Cited:

Brauer, Chris J., Michael P. Hammer, and Luciano B. Beheregaray. "Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin." *Molecular Ecology* 25.20 (2016): 5093-5113.

Hancock, Angela M., et al. "Adaptation to climate across the Arabidopsis thaliana genome." *Science* 334.6052 (2011): 83-86.