BIO594 – Summary

Topic: Landscape Genomics

February 12, 2019

Kevin Wong

The studies of Brauer et al. 2016 (BE) and Hancock et al. 2011 (HE) seek to determine if species inhabiting large geographical and environmental ranges have genomic variations that enable them to adapt to differing conditions. Even though both studies have similar objectives; the study system, approaches, and results were drastically different. BE focused on *Nannoperca australis*, an endangered freshwater perch with low dispersal potential with low genetic diversity and strong population structure among demes. Contrastingly, HE used a high-dispersal system, *Arabidopsis thaliana*, over a large distribution range and determined variant changes on loci that strongly correlate with climate. As both studies try to correlate genomic variation with climate or environmental differences, I will compare and contrast the following concepts: (1) the experimental design, (2) isolation of environmental variables, (3) determination of genetic variation, and (4) correlating the environmental variables to the genomic variations.

*1. Experimental Design*

HE performed a genome-wide scan of *A. thaliana across geography?,* in addition to a common-garden experiment to determine loci under selection that may allowed adaptation to different climates. 948 accessions of *A. thaliana* across the native Eurasian range and 40 climate variables were analysed to understand the genetic variation across environmental ranges. BE used 263 *N. australis* individuals from 25 locations across the distribution range in the Murray-Darling Basin (MDB) in Australia. 40 climate variables were also chosen in this study, as they were important environmental predictors for *N. australis*. *A. thaliana* has a high dispersal potential, indicating high gene flow between populations and a large distribution range across Eurasia. *N. australis,* however, is a low-dispersal species with isolated population across the MBD. Therefore, it is expected for genetic drift to be the main driver behind adaptive changes in *N. australis*, and mechanisms like hard-sweeps for *A. thaliana*.

*2. Isolation of Environmental Variables*

HE used a correlation analysis (pairwise Pearson Correlation) to eliminate extraneous variables and only retain ecologically relevant variables. BE binned all environmental variables into 5 categories and used a Variance Inflation Factor (VIF) analysis to eliminate correlated variables within categories. Using Principal Component Analysis (PCA), the variables with eigenvalues greater than 1 (i.e. no correlation with other variables) were kept for the genotype-environment association analysis. The BE study had a more rigorous and less biased approach in eliminating correlated variables than HE, which could influence downstream analyses and results.

*3. Detecting Genetic Variation*

The BE authors ddRAD in conjunction with a Bayesian approach (BAYESCAN) to detect neutral and outlier loci. Fst values between populations were determined with HIERFSTAT, which includes variations due genetic drift, which is expected in this case, when estimating local population divergence. To justify the loci were under selection, a genotype-environment association (GEA) analysis was performed. HE did not do any Fst comparisons between populations, as they decided to analyses all samples without grouping into populations to determine if their model would categorize them to their observed population. They provided ~215,000 SNPs to correlate against environmental variables and tested true signals of adaptation through enrichment analyses.

*4.Correlating the Environmental Variables with Genetic Variation*

When comparing how each study correlated environmental variables with the potential selective loci, HE used an approach that would incorporate the bias related to population history. Using a partial Mantel test, HE non-parametrically correlated all of their variants (SNPs) with a climate variable. Contrastingly, BE utilized a spatially explicit GEA analysis to explore the potential for signatures of selection at local and regional scales against five candidate environmental variables. The GEA consisted of two approaches: (1) a generalized linear mixed-model approach and (2) a partial redundancy analysis (RDA). The generalized linear mixed-model approach has two basic models, where the variance of allele frequency is dependent on a fixed environmental variable, and where there is no effect of an environmental variable. Both approaches by HE and BE control for genetic structure. However, to confirm these environmental-allele variation correlations, HE performed a common garden experiment to determine if they could predict the “fittest” phenotype for a known environmental condition generated by their model.

In conclusion, both studies took different approaches to answer similar hypotheses. However given the known life history and situation of their study systems, they both produced the theoretically anticipated results. The variation in *N. australis* populations is likely due to a combination of genetic drift and selection driven by the environment (i.e. hydrocline). The variation in *A. thaliana* is likely due to a “hard sweep” in each environment, creating local adaption between populations, which is evident in the transplant experiment.