Data Expedition- Liana Burghardt and Colin Maxwell
Teaching Dates: 1:30-5:30pm Nov 20th, Nov 25th, and Dec 2nd
Class: Cellular and Molecular Biology—Professor Daniele Armaleo,

Duke graduate students Colin Maxwell and Liana Burghardt recently co-created and taught a 12 hour module for Duke undergraduates on analyzing yeast biofilm formation. The students worked with data to explore both the genetics and gene expression signatures associated with biofilms. The goals of the module were for the students to 1) learn the basics of programming with data 2) summarize, visualize, and interpret a complex dataset, and 3) synthesize the relationships between genotype, gene expression, and phenotype.

The students explored the genes responsible for biofilm formation in yeast and the mechanism by which those genes cause biofilms. Biofilms are communities of microorganisms that surround themselves with sticky secretions. Pathogenic yeast often form biofilms in their hosts, and understanding the genetic basis for biofilm formation may help design treatments to remove biofilms. Using an unpublished dataset generated by Paul Magwene's lab at Duke University, the students explored what genes are associated with biofilms and what genes are turned on in yeast that form biofilms.

This module was presented at the end of a course taught by Professor Daniele Armaleo course entitled "Cellular and Molecular Biology". In this course, the students actually perform historical and cutting edge experiments on the model fungi *Saccharomyces cerevisiae* including generating data similar to the data analyzed in the module. The Data Expeditions module will be incorporated into future classes taught by Dr. Armaleo.

This module paired active learning exercises that provided the statistical intuition behind commonly used methods to map genes responsible for traits, to find patterns in complex datasets and to find common functions in sets of genes with a workbook that led students through conducting these analyses in the R programming environment. The course also featured short lecture components and discussion background information derived from the primary literature. One student noted, "I liked the use of R and all the interactive learning activities because I find it easier to understand a topic by practicing the concept rather than talking about it".

The module was successful in its aims; after taking this module 60% of the students who had not had programming experience before the unit stated that they were more likely to take a programming class as a result of this course. One student commented in their evaluations, "I am still not entirely comfortable [with R] but would like to learn more because I realize now that it is a powerful tool!"

The learning was not a one-way street, Colin and Liana also learned a lot about teaching from this experience and are looking forward to teaching a more refined version of the module next spring.