NC STATE UNIVERSITY

Department of Statistics

Inferring Home Features from Indoor and Outdoor Microbial Fungi

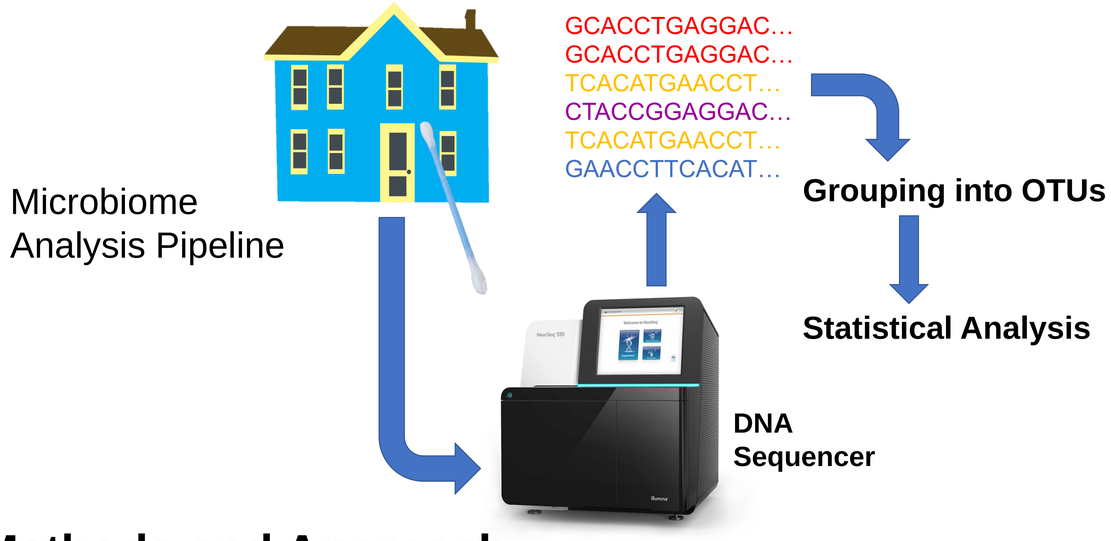
Keltin Grimes[1], Neal Grantham[2], Brian Reich[2]

[1]William G. Enloe Magnet High School, Raleigh, NC 27610*

[2]Department of Statistics, North Carolina State University, Raleigh, NC 27695

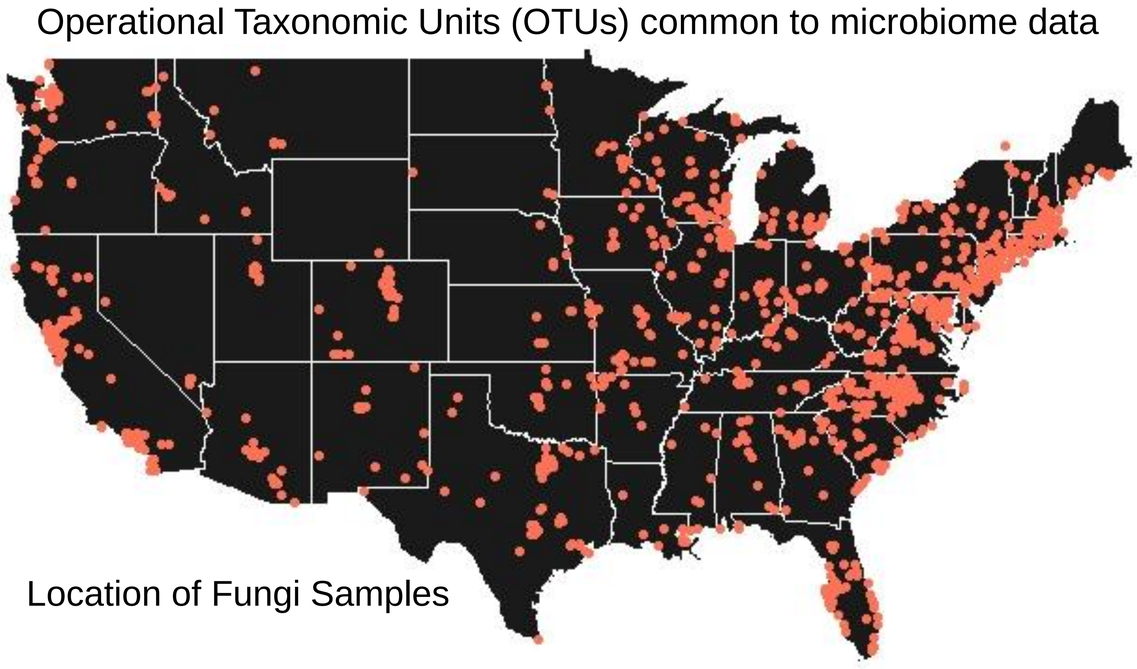
Objective

- To develop a model that quantifies the relationship between the microbial fungi inside and outside one's home and features of that home, such as the number of occupants, their allergies, type of pets they own, does a smoker live in the house, etc.



Methods and Approach

- Create a model that can effectively deal with the large number of Operational Taxonomic Units (OTUs) common to microbiome data

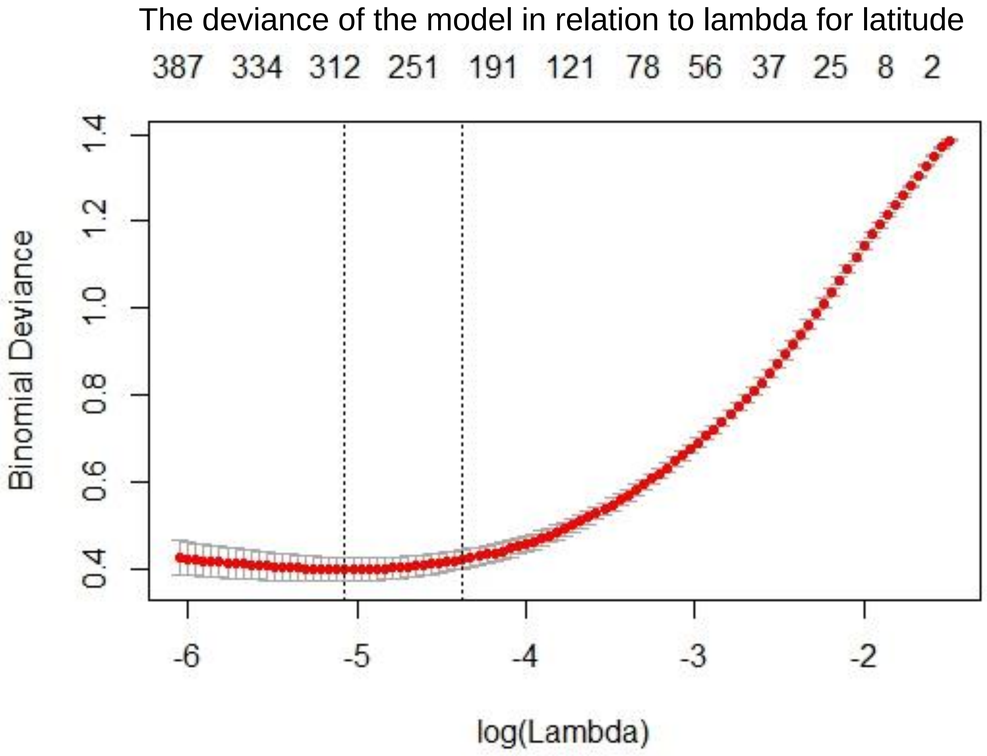


Location of Fungi Samples

glmnet solves this equation over a grid of λ covering the entire range

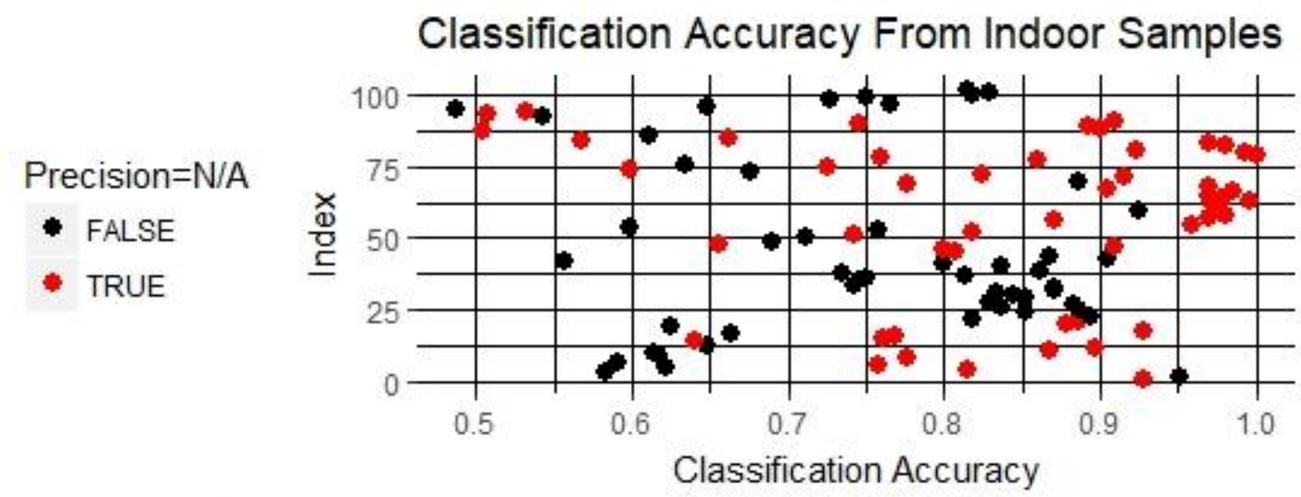
$$\min_{\beta_0,\beta} \frac{1}{N} \sum_{i=1}^{N} w_i l(y_i, \beta_0 + \beta^T x_i) + \lambda \left[(1-\alpha)||\beta||_2^2/2 + \alpha||\beta||_1 \right]$$

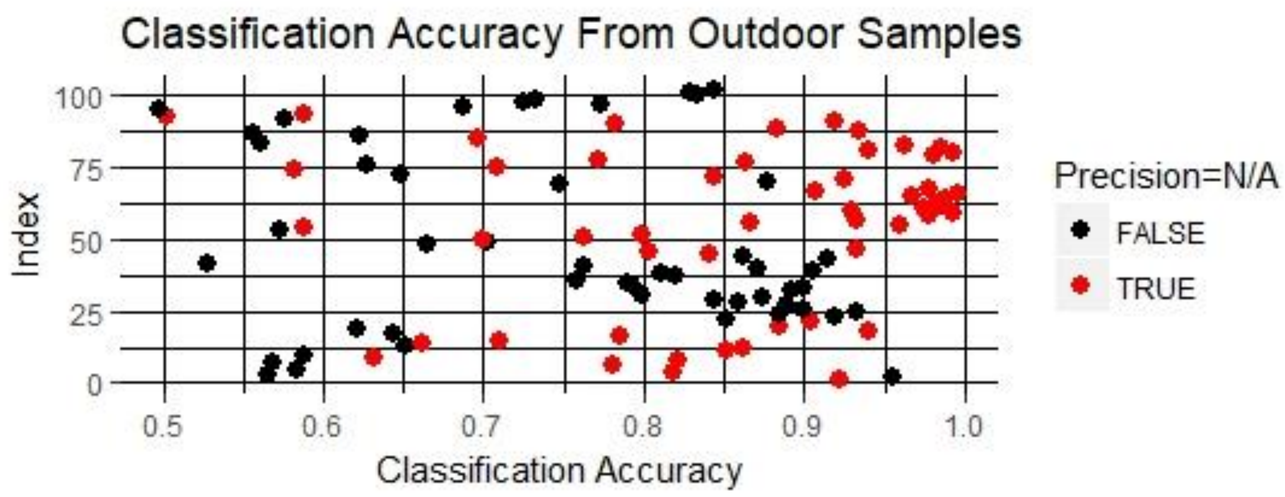
The deviance of the model in relation to lambda for latitude



Results

- Implemented with the glmnet R package, the logistic regression model includes a LASSO penalty that deals well with the large number of OTUs
- Analyzing fungi data, the model predicts the value of a specified metadata variable, and yields classification accuracy, an error matrix, and the most positive and negative coefficients along with which OTUs those coefficients represent
- **The model can predict east vs west, north vs south, most environmental variables (e.g. high mean temperature vs low mean temperature), and diversity statistics (e.g. high amphibian diversity vs low amphibian diversity) with 87-93% classification accuracy**
- With variables that the model cannot accurately predict, the model often will have a precision of N/A, meaning that it only predicts an absence of the variable (shown below in red)



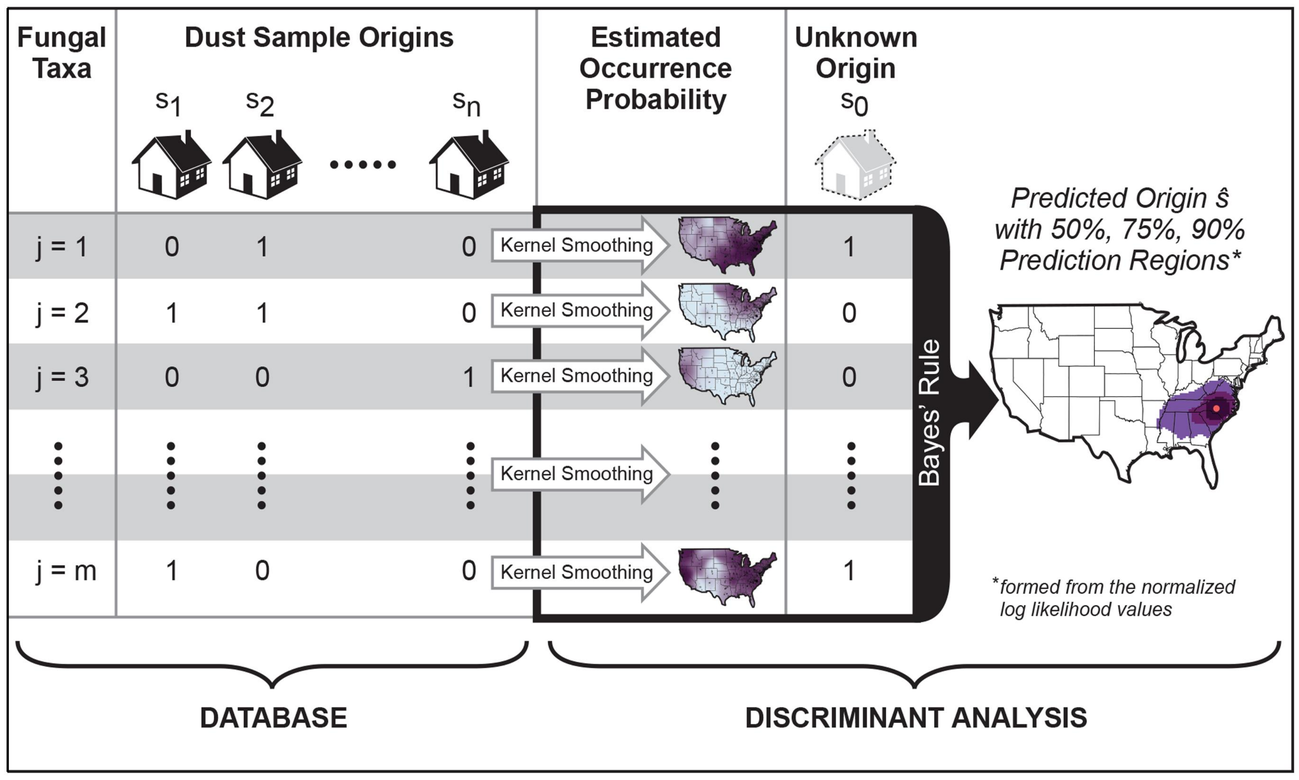


Conclusions

- Many covariates are simply not correlated enough to OTUs
- Most of the success comes from variables measured independently from the house, such as mean temperature and precipitation

Future Work

- Tune model to specific covariates
- Use success with environmental variables to advance work on identifying geographic origins of fungal samples



Fungi Identify the Geographic Origin of Dust Samples, Grantham et al., 2015

Acknowledgements

- Data provided by Rob Dunn and the Rob Dunn Lab
- Thanks to Neal Grantham, Brian Reich, and Benjamin Hu for guiding me through the research process

NC STATE UNIVERSITY