EEB C234 - Final Project

Nicholas Ruffolo Michael Alfaro Marc Suchard

Introduction

In the natural world, color serves a number of purposes for organisms including but not limited to sexual selection, camoflauge, and mimicry (Cuthill *et al.* 2017). For the aforementioned reasons, color has always been a captivating trait to study for evolutionary biologists, dating as far back to both Charles Darwin and Alfred Russel Wallace (Darwin 1871; Caro 2017). To analyze color, visual ecologists have used different color spaces that fall into three classes: spectral spaces, psychological color spaces, and psycophysical color spaces (Renoult *et al.* 2017). For our work, we have chosen to work in a spectral space through the use of reflectance spectra. Reflectance spectra are the proportion of light reflected at a given wavelength over a set of wavelength values. The key advantage to using reflectance spectra is that it allows us to generate trait values that are independent of any viewer bias.

To put reflectance spectra in an evolutionary context, we have chosen to take a function-valued approach. By using a function-valued approach, it allows us to explore how the entire spectra has evolved over time, rather than tracking a singular data point and losing a significant portion of data. The primary drawback to a function-valued approach for a trait's evolution is that this methodology been an underutilized tool within the field of evolutionary biology (Gomulkiewicz *et al.* 2018). Previous work using function-valued traits have focused on ancestral state reconstructions using phylogenetic Gaussian process regression and phylogenetic generalized least squares techniques (Hadjipantelis *et al.* 2013; Goolsby 2015). In this project, I propose a new method for the comparative methods of function-valued traits, utilizing cubic spline functions and allowing for their coefficients to diffuse under a Brownian motion process in a correlated manner in a quick and convenient framework.

Methods

To start, we gathered reflectance data and a phylogeny on 51 tanager species (Burns & Shultz 2012; Burns *et al.* 2014).

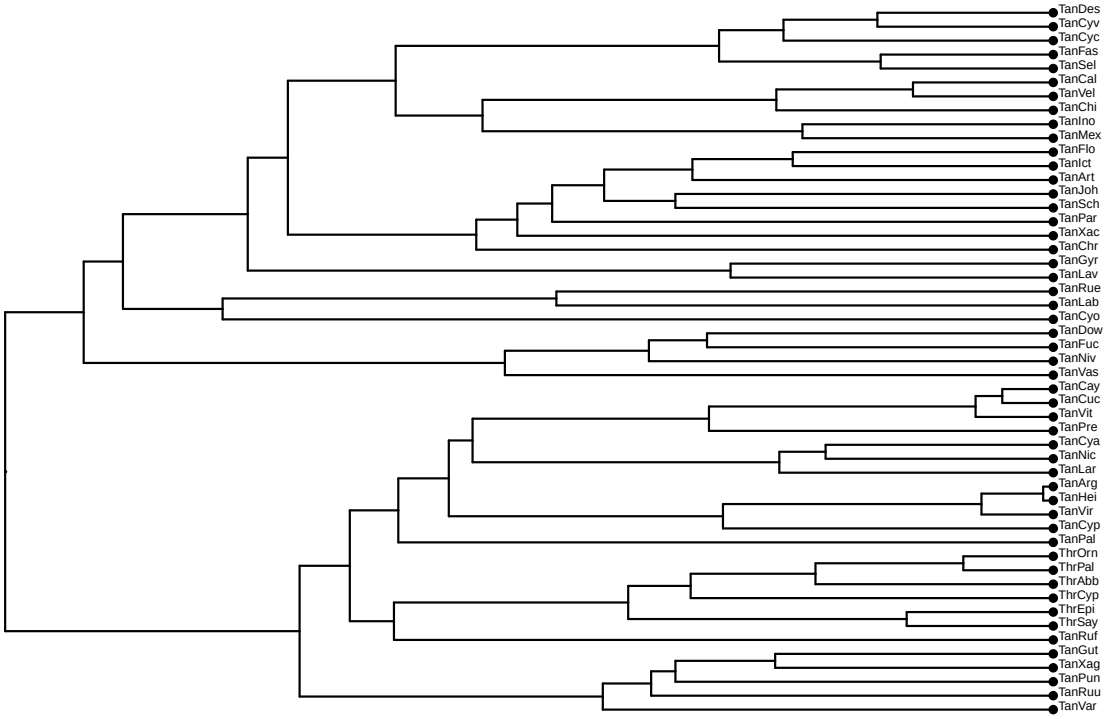


Figure 1: Pruned Tanager Phylogeny from Burns et al. (2014)

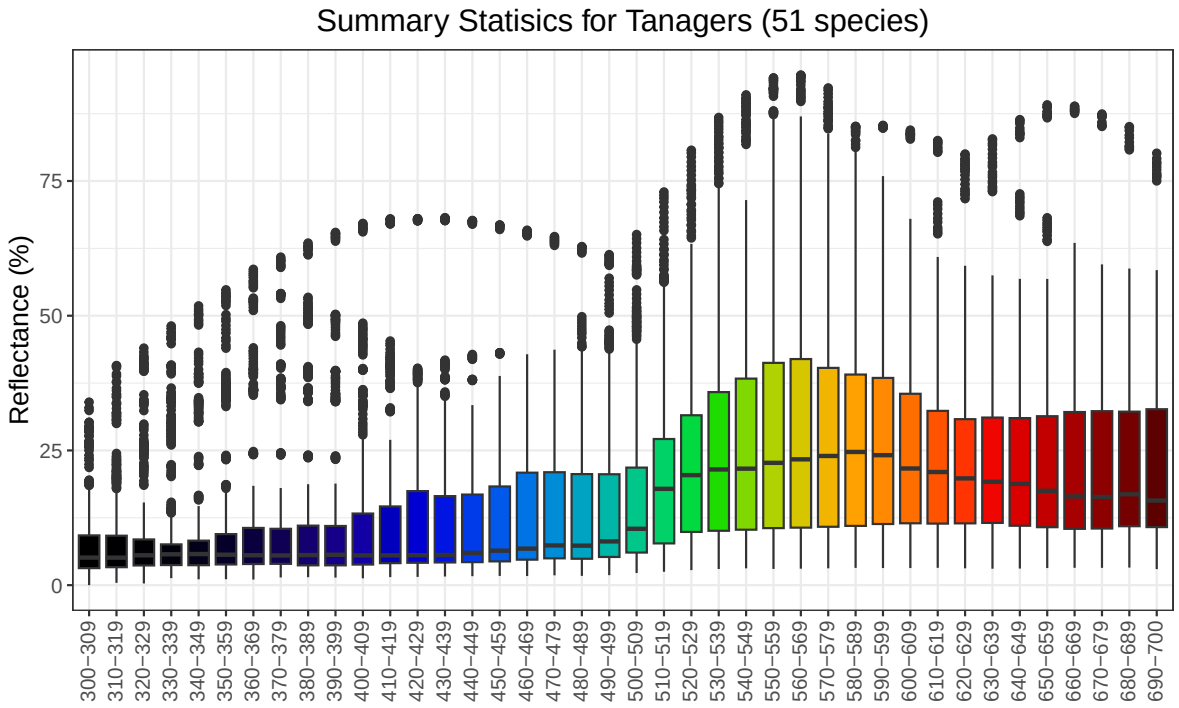


Figure 2: Boxplots for each 10nm interval of light in the visible spectrum. Each boxplots color corresponds to the color each subset of light represents in RGB color space.

Next, we will be fitting a cubic splines to our reflectance data. Now you may be asking what a cubic spline is. In its most rudimentary sense, a spline function is a piecewise polynomial fit to data. What happens is you segment your data into non-overlapping clusters at points called knots. Two knots will be placed at the endpoints of your data. Knots placed at the endpoints of data are called boundary knots. All knots located in between the boundary knots are called internal knots. In R, you have the luxury of specifying where you place the knots for your data, otherwise they will be placed at $\frac{1}{(n+1)}$ quantiles, where n is the number of knots that you are placing.

Please note that it is important to use the same domain for each species. For example, our data can range anywhere between 300-1000 nanometers depending on the instrument, and if you use a different range, then the parameterizations (e.g. knot placement) will not be uniform. Different parameterizations will provide you with different trait spaces and therefore innaccurate conclusions.

Another key option you can choose from is the degree of polynomials that you would like to fit to the data. This can range from 1 to any other natural number, although a degree of 3 (cubic polynomials) should suffice for a large quantity of datasets. Degree is important for determining the degrees of freedom of your spline fit. Degrees of freedom are calculated

as $n + d$, where n is the number of knots that you would like and d is the degree of your polynomials.

For our model, we have selected to use $d = 3$ and $n = 35$ for our spline parameters. This segments our wavelength range (300 to 700 nanometers) into bins of approximately 11nm bins. This provides enough coverage to each bin of data without worry of overfitting while also adequately capturing the variation found in any given reflectance spectra.

For our regression model, we used a generalized least squares model (`gls` from the `nlme` package). We use this regression technique as we can specify the error structure for our model. Changing the error structure is important as the errors in the regression model are not going to be independent from one another (a key assumption when using linear regression [`lm`]). The non-independence arises from the fact that the reflectance at any given wavelength will be highly dependent of the reflectance at surrounding wavelengths.

In the chunk below, we apply our gls model to our spline fit.

Here, I have a brief explanation of what is extracted and from what function:

- **Degree** -> `bs` function from the `splines2` package
- **Boundary knots** -> `bs` function from the `splines2` package
- **Internal knots** -> `bs` function from the `splines2` package
- **Model Coefficients** -> `coef` function from base R
- **AIC, BIC, and logLik scores** -> extract from model using the `summary` function from base R
- **`rspec` reflectance** -> `as.rspec` function from the `pavo` package
- **RGB hexcode** -> `spec2rgb` function from the `pavo` package.

After extracting our spline coefficients, we then allowed our coefficients to diffuse in two separate frameworks. Our first approach was using a maximum-likelihood method, using the R package `mvMORPH`. Our second approach, we used a Bayesian method. We used a fixed tree topology, Figure 1, and ran a MCMC for 20000 runs, sampling every 100 iterations, and used a burn-in of 2000 iterations. To view our Bayesian results, we then sampled 1000 trees and randomly drew 100 trees from our subset to view our results.

Results

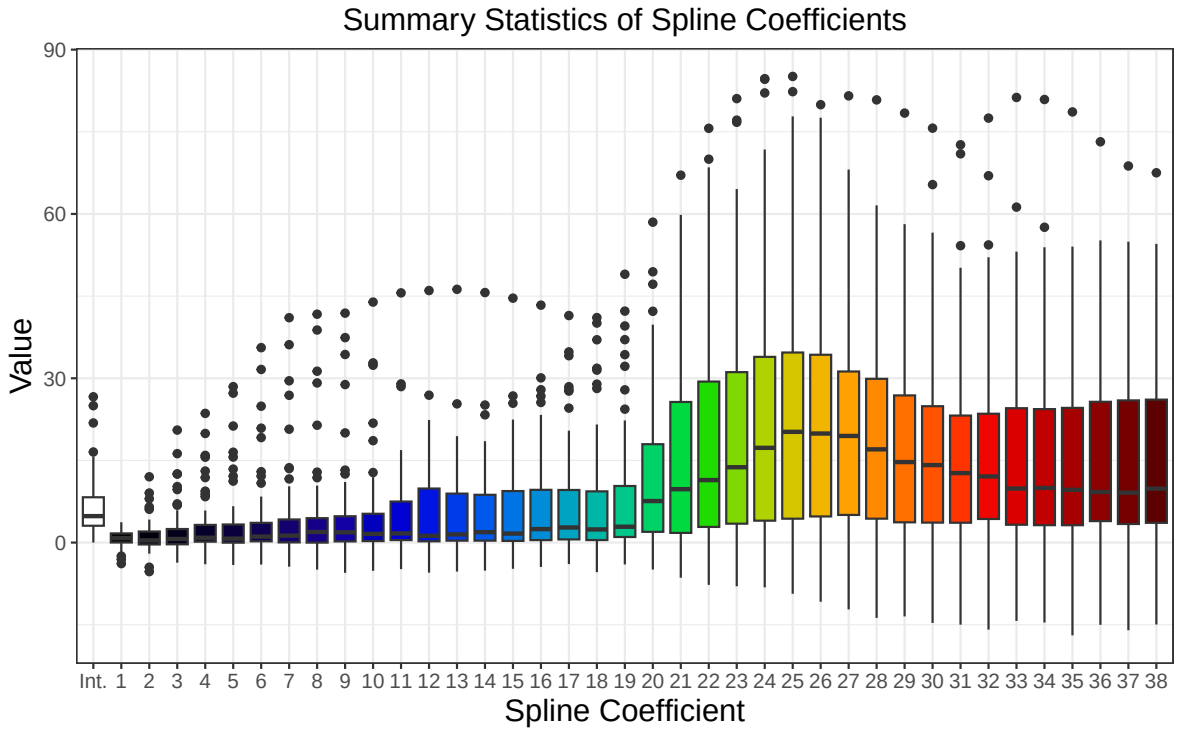
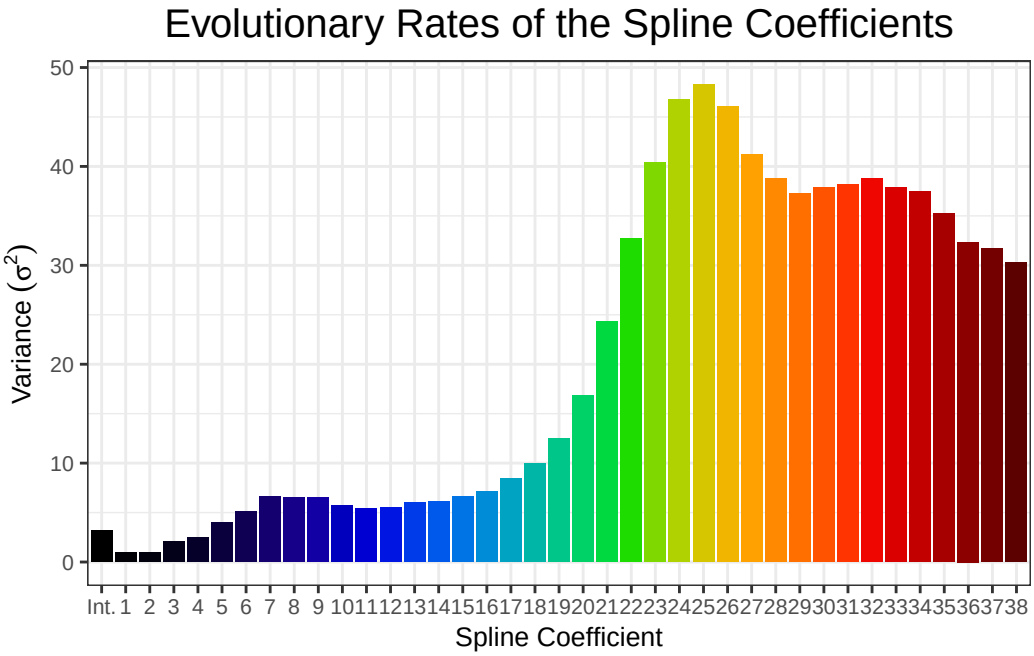


Figure 3: Boxplots for the each of the 39 spline coefficients used to represent the reflectance spectra for the 51 species of tanagers.

When allowing our spline coeficients to diffuse under a Brownian motion process in a correlated manner, we find that our spline coefficients evolve at rates that we would expect based on their mean and spread of our spline coefficients in extant taxa.

A possible hypothesis as to why we see this trend in our evolutionary rates could stem from the color production mechanisms used in bird coloration. Birds that use pigments to color themselves (associated with higher wavelengths of light) might experience less constraints when transitioning to other colors while colors produced by structural means might have stricter physiological constraints.



After seeing how fast each coefficient was evolving, we wanted to see if any parts of our reflectance spectra were coevolving. We can do this by transforming our variance-covariance matrix into a correlation matrix. What we see are approximately 4 islands of highly correlated coefficients, with each island being approximately 8 to 11 coefficients in size. Each island represents distinct color regions as follows: -

- Island 1 = UV to blue wavelengths
- Island 2 = Blue to light green wavelengths
- Island 3 = Green to light orange wavelengths
- Island 4 = Orange to NIR wavelengths.

Additionally, we notice that coefficients that are farther from a given coefficient experience little to no correlation in their coevolution, and as coefficients are located even further along on a reflectance spectrum, they turn out to be negatively correlated with one another.

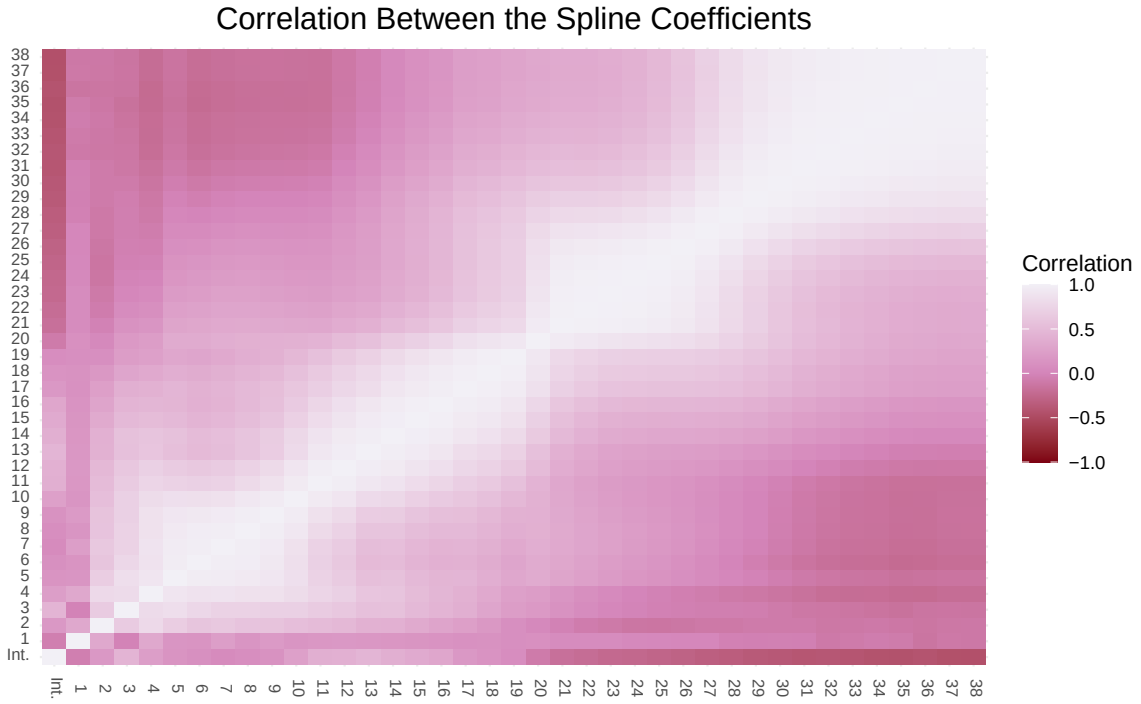


Figure 4: Heatmap showing the correlation between spline coefficients through a Brownian motion diffusion.

References

Burns, K.J. & Shultz, A.J. (2012). Widespread cryptic dichromatism and ultraviolet reflectance in the largest radiation of Neotropical songbirds: Implications of accounting for avian vision in the study of plumage evolution. *The Auk*, **129**, 211–221.

Burns, K.J., Shultz, A.J., Title, P.O., Mason, N.A., Barker, F.K., Klicka, J., Lanyon, S.M. & Lovette, I.J. (2014). Phylogenetics and diversification of tanagers (Passeriformes: Thraupidae), the largest radiation of Neotropical songbirds. *Molecular Phylogenetics and Evolution*, **75**, 41–77.

Caro, T. (2017). Wallace on Coloration: Contemporary Perspective and Unresolved Insights. *Trends in Ecology & Evolution*, **32**, 23–30.

Cuthill, I.C., Allen, W.L., Arbuckle, K., Caspers, B., Chaplin, G., Hauber, M.E., Hill, G.E., Jablonski, N.G., Jiggins, C.D., Kelber, A., Mappes, J., Marshall, J., Merrill, R., Osorio, D., Prum, R., Roberts, N.W., Roulin, A., Rowland, H.M., Sherratt, T.N., Skelhorn, J.,

Speed, M.P., Stevens, M., Stoddard, M.C., Stuart-Fox, D., Talas, L., Tibbetts, E. & Caro, T. (2017). The biology of color. *Science*, **357**, eaan0221.

Darwin, C. (1871). *The descent of man, and Selection in relation to sex, Vol 1.* John Murray, London.

Gomulkiewicz, R., Kingsolver, J.G., Carter, P.A. & Heckman, N. (2018). Variation and Evolution of Function-Valued Traits. *Annu. Rev. Ecol. Evol. Syst.*, **49**, 139–164.

Goolsby, E.W. (2015). Phylogenetic Comparative Methods for Evaluating the Evolutionary History of Function-Valued Traits. *Systematic Biology*, **64**, 568–578.

Hadjipantelis, P.Z., Jones, N.S., Moriarty, J., Springate, D.A. & Knight, C.G. (2013). Function-valued traits in evolution. *J. R. Soc. Interface.*, **10**, 20121032.

Renoult, J.P., Kelber, A. & Schaefer, H.M. (2017). Colour spaces in ecology and evolutionary biology: Colour spaces in ecology and evolutionary biology. *Biol Rev*, **92**, 292–315.