

### EEB313: Mid-project Update

For this project I will explore the ecological factors that might explain the variation in *Apistogramma* phenotype, and how incorporation of these data may influence the topology of a phylogenetic tree. I will examine ecological and phenotypic traits for approximately 30 species in the genus (Römer, 2006). I believe that the visual conditions where the fish live are responsible for the general phenotype of the species that evolve, and that convergence within the genus is due to environment, not shared ancestry. I predict that analysis will cluster species by phenotype best when explained by water type rather than ancestry, as it encapsulates numerous ecological factors and determines the visual environment for this highly communicative species. I predict that a tree built using the combined ecological, morphological and molecular data will be a better representation of the genus than previously published trees built using either morphology or molecular data alone.

The data come from two tables published in a book by Dr. Uwe Römer, as well as phylogenetic data from a paper by Tougaard *et al* (2017). The first table is a compilation of categorical ecological data for 102 species of *Apistogramma*, including water body type, water type, habitat type, and river system. The second table is a compilation of 51 morphological characters used for phylogenetic cluster analysis of the same *Apistogramma* species. These characters include physical traits like colour pattern, body shape, and dentition, as well as reproductive behaviour. Tougaard *et al.* used molecular data from 30 species of *Apistogramma* to construct a phylogenetic tree, resulting in four lineages. I will use the species identified in the paper and in addition, based on personal experience and published data, I will assign lineage to species that are suspected to be within the same species complex. I will knit the morphological, ecological, and lineage data by species, then remove rows (species) with any missing values. The ecological data contains some 'occasionally' values, which will be converted to 'yes', as they are based on presence/absence and so an answer of occasionally likely reflects sampling effort. Behavioural traits such as broodcare and family structure will be omitted, as these are context-dependent and the author did not state how these were determined. The total data will be filtered so that only species with lineage data are included.

I will conduct a Multiple Correspondence Analysis (MCA) using the R packages 'FactoMineR' and 'factoextra'. This will cluster like species together and identify the variables that best explain the shape of the data. I will plot the individual species within the MCA and colour iterations by the ecological factors (most importantly whitewater and blackwater) as well as lineage, to visualize the difference in groupings. If plots appear similar, I will use ANOVA to compare the means of each grouping to see if they vary significantly. From the MCA I will extract results for the variable categories to determine which contribute most to the principal components, either ecological or lineage. The MCA will assign new dimensions for the data, from which I can create two columns of continuous data per species (Dimension1, Dimension2). I will add the dimension data to the total data and create a maximum parsimony tree using the R package 'ape'. The results will then be compared to trees in Tougaard *et al.*(2017) and Römer (2006) to see whether

the addition of ecological data creates a tree in concordance with one based on either molecular or morphological data alone.

Römer U. 2006. Cichlid Atlas Volume 2. 1<sup>st</sup> Ed. Melle, Germany: Mergus.

Tougard C, García Dávila CR, Römer U, Duponchelle F, Cerqueira F, Paradis E, Guinand B, Angulo Chávez C, Salas V, Quéroutil S, Sirvas S, Renno JF. 2017. Tempo and rates of diversification in the South American cichlid genus *Apistogramma* (Teleostei: Perciformes: Cichlidae). PLoS One. Sep 5;12(9):e0182618.