Exercise 10. ANCOVA
Due: 11/12/2013
In collaboration with Will Crampton and Joe Waddell

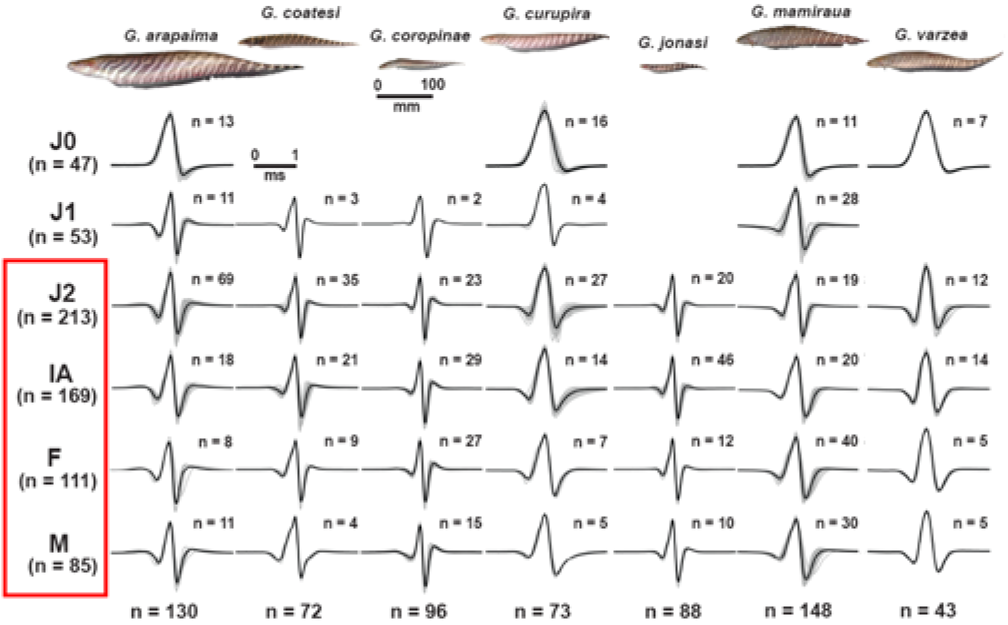

Fig. 1. Ontogeny of electric signals in seven sympatric species of the electric knife fish genus *Gymnotus* from a study site in the Central Amazon basin. J0 = post-larval; J1 = small juvenile; J2 = large juvenile; IA = Immature adult (immature but above minimum size of sexual maturity); F = Female; M = Male. All signals are recorded as head-to-tail waveforms in the far field. In this case study we will explore changes in signal structure during transitions from J2 to IA, and from IA to F or M.

Note: E-mail a single Word document with your results to both Instructors. All analytical work_needs to be done in R (unless otherwise noted). Scripts and output from R should be included in the Word document for full credit. *Please make sure you read all texts to understand the ecological context of the exercise, and be careful to answer all questions you find within the text!*

A study by Will Crampton and Joe Waddell investigated ontogenetic changes in signal structure in seven species of South American electric fish (above). They broke down signals into time-frequency coefficients and used these to construct a multi-dimensional ‘signal space’ within which distances between groups of individuals could be measured. Unlike some animal signals, electric fish signals serve both a reproductive function (only in mature specimens) and also non-reproductive signaling functions (e.g. territorial contests) in immature specimens. The signals in mature males (and sometimes females) are often more ‘showy’ than in immature specimens (e.g. more energetically demanding, or more conspicuous to eavesdropping predators). Because of this change in signal function during maturation, Crampton asked whether evidence for a previous history of Reproductive Character Displacement (RCD) could be gathered from changes in the spacing of a pair of species, in signal spacing during maturation.

The theoretical expectations are that:
1. If adults have signals that are confusingly similar (close by in signal space), the costs induced by heterospecific matings will drive selection for RCD.
2. If two species have similar signals at the immature stage, their signals will diverge a lot during maturation so that they don’t run a high risk of heterospecific mating as adults.
3. If two species have very different signals at the immature stage, their signals will not diverge a lot during maturation.

To begin, we measure the multivariate Mahalanobis distance D^2 from immature adult – to immature adult (of a given pair of species) (D^2 IA-IA) and then we measure D^2 from mature adult – to mature adult (of the same pair of species). For males this will be D^2 M-M, for females this will be D^2 F-F.

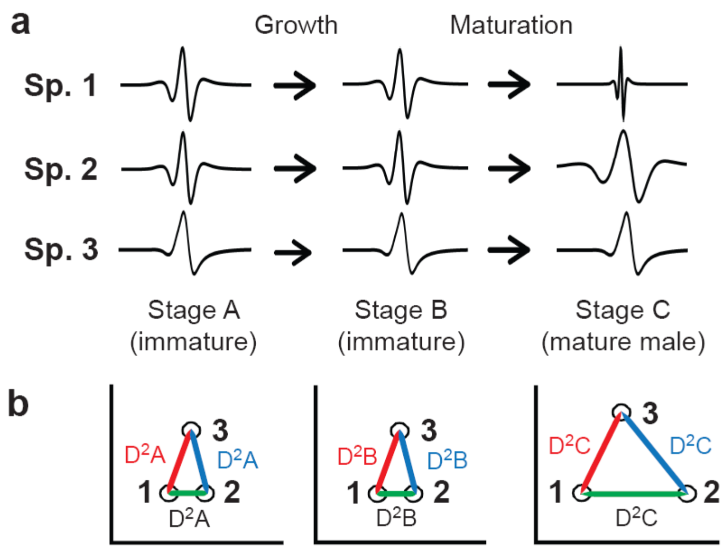

Fig. 2. Model of RCD in three hypothetical species, with hypothetical signals. a = waveforms. b = two-dimensional representation of signal space, with circles = centroids of each species (e.g. in Discriminant Function Analysis).

For any pair of heterospecifics (different species) we can then quantify the strength of RCD as the ***proportional increase in the distance in signal space*** from i. the immature stage, to ii. the mature stage (i.e. during maturation).

The **predictor** variable (on x axis) will therefore be: D^2 IA-IA
The **response** variable (on y axis) will be D^2 M-M/D^2 IA-IA (for transition from immature adult to mature males)
D^2 F-F/D^2 IA-IA (for transition from immature adults to mature females)

We have 7 species that coexist in the same ecological community, and breed synchronously.

Q1. How many paired combinations of species are there? [0.5 points]

Q2. Theory predicts that, for all paired combinations of species, the response variable will exhibit an overall NEGATIVE correlation to the predictor variable. Briefly explain why. [0.5 points]

In particular, we might predict that the actual relationship between the response (y) and predictor (x) variable might not be a simple linear relationship, but instead a non-linear relationship that resembles a POWER FUNCTION for lower values of x. Crampton modeled this prediction as the following graph.

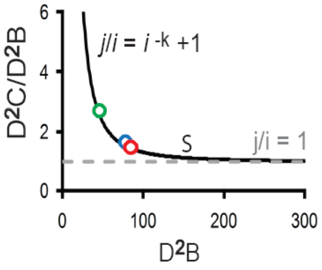

Fig. 3. Predicted response for maturation, under conditions of RCD.

Q3. Why might such a relationship emerge? [Hint: S here is a 'safe distance' beyond which two species are unlikely to accidentally breed and thus incur the costs of a heterospecific mating. Remember the y-axis is our measure of RCD. The x-axis is the distance in signal space of two species, prior to maturation. This graph refers back to Fig. 2 (see colors for inter-specific distances; for ontogenetic transition B to C)]. [0.5 points]

In EARLIER ontogenetic transitions – e.g. *GROWTH* from large juveniles to immature adults, we do not expect RCD, because there are no costs associated with either of the two stages being close by in signal space. Hence during a transition from large juveniles to immature adults we might expect the following kind of relationship. Here the transition is from Stage A to Stage B, and colors represent interspecific distances, as in Fig. 2.

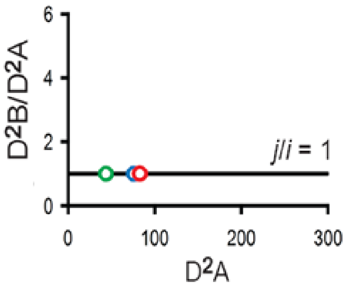

Fig. 4. Predicted response for growth, where there is no RCD (but instead stabilizing selection on juvenile signals).

Q4. What kind of a mathematical function is this? How does it differ from the previous graph? [0.5 points]

Q5. What statistical techniques can we use to: [0.5 points]
a) Distinguish between these two kinds of patterns?
b) Test for the significance of a slope for the regression of the predictor and response variables?

The dataset is contained in the file `ancova_fish.txt` **with the following conventions:**
***Age/Sex group:**
Juveniles (J) where the covariate is J2-J2 and the response is IA-IA/J2-J2
Males (M) where the covariate is IA-IA and the response is M-M/IA-IA
Female (F) where the covariate is IA-IA and the response is F-F/IA-IA
***Species:** ara = *Gymnotus arapaima*; coa = *G. coatesi*; cor = *G. coropinae*; cur = *G. curupira*; jon = *G. jonasi*; mam = *G. mamiraua*; var = *G. varzea*.

Q6. Plot the following three regressions: [1 point]
a) Growth from large juveniles to immature adults: i.e. x = J2-J2, y = IA-IA/J2-J2
b) Maturation from immature adults to males: i.e. x = IA-IA, y = M-M/IA-IA
c) Maturation from immature adults to females: i.e. x = IA-IA, y = F-F/IA-IA

Q7. Use AIC to work out which model (linear, Power or Power + 1) best fits each of regressions A-C. [1 point]

Q8. Work out if the slopes of regressions a, b, and c, are statistically significant. Use an appropriate transformation if necessary and choose whether to use a Frequentist or a Bayesian approach. What is the biological interpretation of these results? How well do they match theoretical predictions for growth versus maturation? [2.5 points]

Q9. Use **ANCOVA** to compare the slopes and intercepts between immature adults (a), males (b), and females (c). Choose whether to use a Frequentist or a Bayesian approach. Show a plot of your results. Explain your results, and why they are biologically significant. [2.5 points]

Q10. Search the literature to find about Bonferroni corrections in the frequentist framework. Think about the three ANCOVA results. How should the p-values be considered when assessing statistical significance? How does the application of Bonferroni corrections (or not) affect the possibility of Type I and Type II errors? [0.5 points]