Homework 13. A GLMM for survival in a heat shock experiment.

The file "heat_shock_all.csv" includes data from the silverleaf whitefly heat shock experiment that you looked at in homework 2.

Each row of the dataset is an experimental replicate. In each replicate, 10 flies were subjected to a heat shock, and the number of survivors at the end is recorded in the column 'Survival', while the proportion of survivors is in 'proportion'. The experiment used flies from two regions in Colombia, the Caribbean region (sea level, uniformly hot) and the Southwest region (in the Andes, more variable temperature). The researchers hypothesized that the flies from these two regions might be locally adapted to thermal conditions, leading to different heat shock tolerances. Experimental replicates were performed on males and females separately, to quantify any effect of sex on heat shock tolerance.

Each replicate had 10 flies, of a single Sex. All of the flies in a replicate were derived from a single isofemale line, meaning they were all third-generation descendants from a single female. There are a total of 56 isofemale lines used in the experiment. This means these flies from the same isofemale are all highly related to each other, compared to the other flies, and the variation among isofemale lines is a measured of genetic variation in heat shock tolerance. Furthermore, the isofemales were derived from a set of 9 populations in two regions, as shown in this figure:

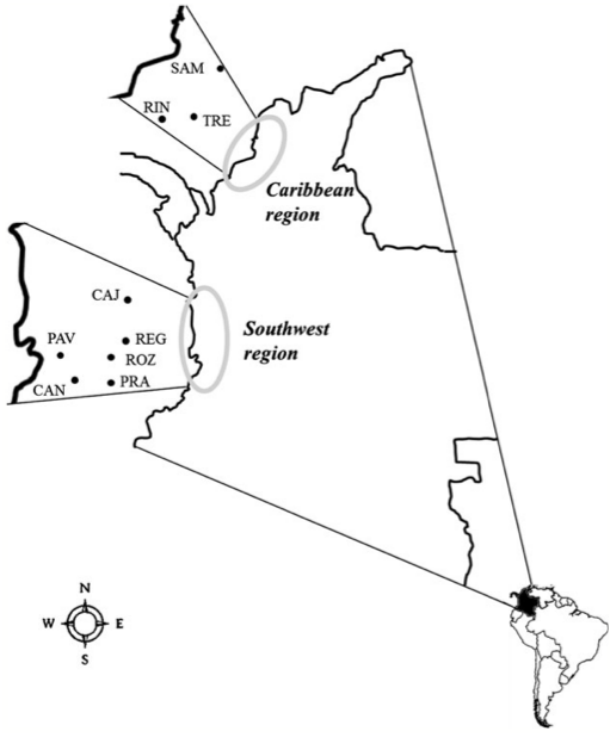

Fig. 1 Populations of the whitefly *B. tabaci* analysed in this study from Colombia. The codes for each sample are as in Table 1.

Each population contributed between 4 and 11 isofemale lines. Clearly the data have a hierarchical structure: replicates within isofemales within populations within regions.

The experiment used two different heat shock treatments. One is the 'hardening' treatment, where flies were acclimated to 40ºC for one hour, reduced to 25ºC for one hour, and then shocked at 45ºC for one hour. The authors call this the 'hardening' treatment because it allows the flies to potentially acclimate to higher temperatures before getting shocked. The authors also used a 'base' treatment, where flies where just shocked at 45ºC for one hour, without a pretreatment at 40ºC. The difference between the 'base' and 'hardening' treatments gives some info on the plasticity of the heat shock response, i.e. whether the preconditioning at 40ºC allows flies to acclimate to higher temperature and survive better at 45ºC.

We can use these data to test the following questions:

1. Do males and females have different heat shock tolerance?
2. Do flies from different regions have different heat shock tolerance?
3. Do the base and hardening treatments differ, i.e., is there acclimation?
4. Does the acclimation effect differ between the two regions?
5. Does the acclimation effect differ between the two sexes?
6. Does the effect of sex differ between the two regions?

Fit a GLMM that incorporates random effects for isofemale and population, and also includes fixed effects and interactions that test the 6 questions. We also need to account for potential overdispersion in the binomial response. The best way to do this in lme4 is with an 'individual level random effect'. Make a new factor where each row of the dataset gets a level, like this:

```
mydata$replicateID = factor(1:nrow(mydata))
```

Using this as a random effect allows for random variation at the level of the replicate that is not accounted for by the binomial distribution.

Fit this whole model, and then use likelihood ratio tests to see which of the interactions is significant (LRTs tend to be anti-conservative for fixed effects, but this model has a lot of data, and doing a parametric bootstrap would take a long time with this model). Drop the non-significant interactions, which will allow you to test and interpret the main effects more clearly. Now use LRTs to test the significance of the remaining fixed effects in the model.

Plot the fitted fixed effects. How do you interpret these results, in light of the 6 questions listed above?

Use the random effects estimates, and plots of those estimates, to discuss how much genetic variation there is among populations, and among isofemales within populations, and how much extra (overdispersed) variation there is among replicates. Plotting the individual-level random effect for replicate is not advised; there are too many levels for the function. Think about the relative magnitude of the fixed and random effects. How does the random variation across groups compare to the size of the fixed effect differences?

The authors were interested in another question as well: Does acclimation to high temperature (i.e., the treatment effect) exhibit genetic variation among isofemale lines? To test this, you need to allow treatment to vary by isofemale. In other words, you need to include a fixed*random interaction. Put this in the model, and use a likelihood ratio test to see if it is important. Use a plot to visualize how the treatment effect varies by isofemale. Does the genetic variation seem substantial? What is the correlation between the random Intercept and random treatment effect among isofemales? Can you explain what this correlation implies? Does including the treatment-by-isofemale interaction change the results for the fixed effects in the model? If so, how?