**Microcosm Evolution**

Some researchers design an experiment to investigate evolution in relatively small populations. They set up ten microcosms, each with a population of flies in it. The flies are initially randomly allocated to microcosms. The flies are then allowed to breed and the experiment is run for many generations. The microcosm equipment is designed so that some movement among microcosms can be allowed or closed off by the researchers. In this experiment, movement is initially closed off, but at some point during the experiment it is opened up and allowed. The equipment also allows a temperature gradient to be applied across the microcosms, so that microcosm 1 is a little hotter than microcosm 2, which is a little hotter than microcosm 3, which is a little hotter than microcosm 4, and so on, with microcosm 10 being the coolest. Alternatively, all microcosms can be set to the same intermediate temperature. In this experiment, temperature is first set to have a gradient across the microcosms, then at some point set to be equal across the microcosms, and then again set to have a gradient across the microcosms.

The researchers sample from the flies in each microcosm at each generation, and use some expensive new equipment to measure five functional traits, five morphological traits, and three semi-cryptic traits for each individual fly sampled. They also test 20 marker loci and measure the number of a specific test marker allele at each of the loci – giving either 0, 1 or 2 of the test alleles at each of the marker 20 loci. The markers were specifically chosen to be ‘neutral’ markers, ie the expectation is that they would not be under selective pressures. The trait and genetic data are recorded in a separate file for each generation. At the end of the experiment the researchers are faced with a large amount of data and not much idea how to deal with it. As a recent graduate with some skills and experience in ‘big biological data’, you are called in to help! 😊

Work through the “look at data.R” script, checking that you understand what the code is doing. If you are unsure of anything, make sure you ask for clarification.

When you are ready, extend the code to look at different traits and different generations, following the suggestions in the script.

You will want to consider some of the following general questions…

Do some of the genes become fixed or lost? Does this happen for only some genes or all genes? Only some populations or all populations? Some generations or all generations?

Are there trends of change over time in any of the traits and/or genetic frequencies? Are these consistent over time, or do they change at certain times? Do trait values and/or genetic frequencies in the populations diverge from each other, or remain close, or diverge and then become close again? Do any of these patterns indicate that any of the traits or genes are under selective pressures at any times? And/or that there is gene flow among populations at any time?

For the assignment, you will need to enter answers to some more specific questions about this data set through an LMS quiz.