

# **Supplementary Material for "slimr: An R package for integrating data and tailor-made population genomic simulations over space and time"**

Russell Dinnage<sup>1, 2</sup>

Stephen D. Sarre<sup>2</sup>

Richard P. Duncan<sup>2</sup>

Christopher R. Dickman<sup>3</sup>

Scott V. Edwards<sup>4</sup>

Aaron Greenville<sup>3</sup>

Glenda Wardle<sup>3</sup>

Bernd Gruber<sup>2</sup>

## **Author Affiliations**

1. Institute of Environment, Department of Biological Sciences, Florida International University, Miami, FL, USA
2. Centre for Conservation Ecology and Genomics, Institute for Applied Ecology, University of Canberra, Canberra, ACT, Australia
3. Desert Ecology Research Group, School of Life and Environmental Sciences, University of Sydney, NSW 2006, Australia
4. Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

## **Full Example**

### **Investigating population genomics of small mammals in a periodic environment**

The context for this example is a long-term ecological study in the Simpson Desert in central Australia. Several authors of this paper have studied the population dynamics of small mammals and reptiles in this desert for more than 30 years (C. Dickman, Wardle, Foulkes, & de Preu, 2014; Greenville, Dickman, & Wardle, 2017; Greenville, Wardle, Nguyen, & Dickman, 2016). Recently, we have begun sequencing tissue samples taken from animals captured during the

past 15 years, and obtained single nucleotide polymorphism (SNP) data using DArT (Diversity Arrays Technology Pty Ltd) technology. Here, we use SNP data from 167 individuals of a common native rodent species, the sandy inland mouse *Pseudomys hermannsburgensis*, sampled at 7 sites over three years (2006-2008), and subsequently aggregated to 3 subpopulations for analysis. The three sample years span periods before and after a major rainfall event at the end of 2006; big rains occur infrequently in the study region (every 8-12 years) (Greenville, Wardle, & Dickman, 2012) but drive major population eruptions. We used the SNP data to calculate pairwise  $F_{st}$  values among the three subpopulations in each year, revealing that pairwise  $F_{st}$  values dropped rapidly to nearly zero immediately after the rainfall event from a high recorded just prior to the event when the populations were more genetically differentiated. We interpreted this result to mean that the rainfall event, which caused the sandy inland mouse population to rapidly increase, also allowed animals to move out of spatially scattered refuge patches to which they had been confined during the preceding dry period (C. R. Dickman, Greenville, Tamayo, & Wardle, 2011). This movement allowed the subpopulations to mix, leading to a decrease in population genetic structure as measured by  $F_{st}$ . We go through a whole development process for a simulation based on this system, using `slimr`, in the vignette, available here:

[https://rdinnager.github.io/slimr/articles/Main\\_manuscript\\_example\\_v2.html](https://rdinnager.github.io/slimr/articles/Main_manuscript_example_v2.html)

To formalize our ideas based on that vignette a little more we ran an Approximate Bayesian Computation (ABC) analysis to derive an approximate posterior distribution of model parameters that produced a good fit to our short  $F_{st}$  time series (see ***figshare link*** for the R script and data used to run it, [This has been attached to the manuscript for reviewers convenience, and will be posted to figshare if accepted]). We were able to easily move from simulation exploration

to a more formal fitting exercise because the simulation was already in R (thanks to `slimr`), and so only a small amount of code was required to convert the input and output of our simulation to the format required by the `easyABC` package (Jabot et al. 2013), which we used for this analysis.

## ABC Results

After extracting the parameters of a sample of the approximate posterior distribution we reran simulations based on those parameters, calculated mean  $F_{st}$  and plotted them next to the observed  $F_{st}$  values (Figure 6a, main text). The simulated  $F_{st}$ s do cluster around the observed values though it does appear that the simulated values for 2018 (the year after the rainfall) do tend to be a bit lower than the observed values in the simulations.

The marginal posterior distribution based on samples from the ABC analysis confirms that our data do not constrain individual parameters much, with a fairly wide distribution for most parameters providing a good fit to our data (Figure 6b, main text, diagonal panels). The only exception was perhaps mutation rate, for which the lower values that we simulated tended to provide a better fit. The parameter of most interest to us was the abundance threshold (`abund_threshold` in Figure 6b, main text), which specified the population size above which a subpopulation would 'turn on' migration, that is, start exporting individuals to the other subpopulations (in the real system this population size change is driven by rainfall). In this simulation an abundance threshold of zero or less would be migration always happening, and one of 20 or over would be migration almost never happening. Some simulations produced well fitting  $F_{st}$  values for nearly all relevant values of the abundance threshold, with some falloff at either end. However, when we start looking at combinations of multiple parameters we see that the value of the abundance threshold parameter does constrain what values of other parameters

will make for a good fit to the data. For example, if the abundance threshold is low, and thus migration is always on, only simulations with very low migration rates and very low mutation rates can provide a good fit to the data (figure 6b, main text, panels in rows 3 and 4 in column 2). All in all this suggests that there are two approaches to improving our ability to distinguish how different processes lead to the patterns we see (besides just collecting more data): 1) try adding new summary statistics besides just pairwise  $F_{st}$ , which may capture some other aspect of the data, and 2) use some independent sources of data or information to estimate and constrain the parameter space of our simulations closer to that of the real system. In particular, approach 1 could be tested without having to collect more data by doing more simulations: we could simulate our model, then simulate data collection and calculate our new summary statistic on the simulated data. We can then see if we can recover the parameters of our simulation better than we could before after incorporating our new statistic.

The results from these preliminary simulations will thus be invaluable in guiding which individuals and time periods we should focus our sequencing on, and what summary statistics to use, to maximize the chances of distinguishing among competing hypotheses that might explain the combined population and genetic patterns in the data. Ultimately, we aim to use this approach to understand how future climate change could alter the population and genetic structure of desert animals, highlighting the value of `slimr` in a scientific workflow.

## References

- Dickman, C., Wardle, G., Foulkes, J., & de Preu, N. (2014). Desert complex environments. *Biodiversity and Environmental Change: Monitoring, Challenges and Direction*, 379–438.

- Dickman, C. R., Greenville, A. C., Tamayo, B., & Wardle, G. M. (2011). Spatial dynamics of small mammals in central Australian desert habitats: the role of drought refugia. *Journal of mammalogy*, 92(6), 1193–1209.
- Greenville, A. C., Dickman, C. R., & Wardle, G. M. (2017). 75 years of dryland science: Trends and gaps in arid ecology literature. *Plos One*, 12(4), e0175014.
- Greenville, A. C., Wardle, G. M., & Dickman, C. R. (2012). Extreme climatic events drive mammal irruptions: regression analysis of 100-year trends in desert rainfall and temperature. *Ecology and Evolution*, 2(11), 2645–2658.
- Greenville, A. C., Wardle, G. M., Nguyen, V., & Dickman, C. R. (2016). Population dynamics of desert mammals: similarities and contrasts within a multispecies assemblage. *Ecosphere*, 7(5), e01343.
- Jabot, F., Faure, T., & Dumoulin, N. (2013). Easy ABC: performing efficient approximate Bayesian computation sampling schemes using R. *Methods in Ecology and Evolution*, 4(7), 684-687.