# dartR support for Admixture Analysis

Support for admixture analysis covers a number of approaches. Perhaps the best place to start is to examine structure among the sampled individuals with principal coordinates analysis (PCoA)(Gower, 1966). Ordination with PCA, PCoA and MDS essentially treat each locus as a variable, construct linear combinations of these original variables in order of the percentage of variation among individuals that they explain. This enables the visualization of the structure in the dataset in one, two, three or rarely more dimensions. Noise variation is pushed down to deeper dimensions which are disregarded.

When two populations diverge, they wander apart in ordinated space and their separation is represented in one linear dimension of the final ordination. Independent divergence of three populations will be represented in two ordinated dimensions, and so on. In this context, contemporary admixture will be evident as individuals falling between the two parental populations. An example is shown where individuals from the Murray-Darling Basin, the Paroo, Barcoo and Lake Eyre Basin are plotted in an ordinated space of two dimensions.

So a first step is to plot individuals in an ordinated space.

```
load(file="MDB_LEB.Rdata")
pcoa <- gl.pcoa(MDB_LEB)
gl.pcoa.plot(pcoa, MDB_LEB, xaxis=1, yaxis=2, labels="pop")
```
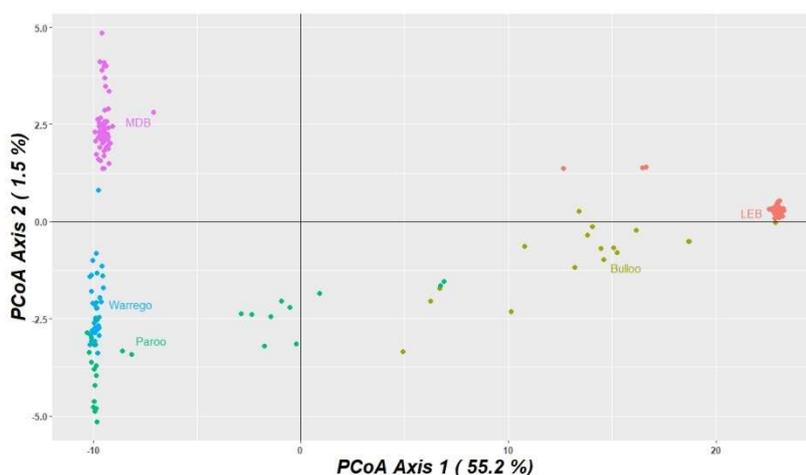

You can interrogate the plot using

```
gl.pcoa.plot(pcoa, MDB_LEB, xaxis=1, yaxis=2, labels="interactive")
ggplotyly()
```

For a quantitative analysis, we recommend NewHybrids (Anderson & Thompson, 2002). NewHybrids employs a statistical method for identifying species hybrids using data on multiple, unlinked markers which does not require that allele frequencies be known in the parental species. However, advance knowledge of the parental species, uncontaminated by introgressed individuals, increases the performance of the software.

The probability model used is one in which parentals and various classes of hybrids (F1s, F2s, and various backcrosses) form a mixture from which the sample is drawn. Using the framework of Bayesian model-based clustering, NewHybrids computes, using Markov chain Monte Carlo, the relative likelihood (posterior probability) that each individual belongs to each of the distinct hybrid classes.

NewHybrids is limited to ca 200 loci because of memory constraints, but this is not as serious a constraint as one might expect. You need relatively few loci to identify hybridization. We choose the best 200 available loci on the basis of their being different and fixed in each of the two nominated parental populations, and then on avgPIC (a DArT statistic measuring information content, akin to an even-ness score).

To run a NewHybrids analysis you need to have installed NewHybrids
(http://ib.berkeley.edu/labs/slatkin/eriq/software/software.htm#NewHybs)[http://ib.berk
eley.edu/labs/slatkin/eriq/software/software.htm#NewHybs].

If you have installed it, then by specifying the installation location, gl.nhybrids
will run NewHybrids and pass the results back to R. If you do not specify an
installation location, gl.nhbyrids() will simply produce an input file for NewHybrids.

```
gl.nhybrids(MDB_LEB,p0=parent0,p1=parent2,outpath=getwd(),outfile="MBD_LEB.nhy",
   nhyb.directory="C:/workspace/R/NewHybsPC",BurnIn=10000, sweeps=10000)
```

The output is a csv file in which all individuals are listed, and their assignment to
one of the parental populations is indicated, if the individuals are not admixed.
Admixed individuals will be assigned a posterior probability of membership of one of
the classes, F1, F2, backcross of F1 to parental 1, backcross of F1 to parental 2.

There is no value in going further than this, because the likelihood bins for more
distant crosses overlap to the point that the results are not that useful.

The following table shows the results for hybridization and introgression between the
introduced Emydura macquarii and the critically endangered Myuchelys georgesi in the
Bellinger River.

pop	Myuchelys	Emydura	F1	F2	F1xMy	F1xEm	N
p0	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	5
Myuchelys (Bellinger)	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	110
Myuchelys (Bellinger)	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	2
Myuchelys (Kalang)	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1
Myuchelys (Kalang)	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	3
Myuchelys (Taronga)	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	16
Myuchelys (Taronga)	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1
Emydura (Bellinger)	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	51
Emydura (Kalang)	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	14
p1	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	63

Here are the combined results of the PCoA and NewHybrids designations.
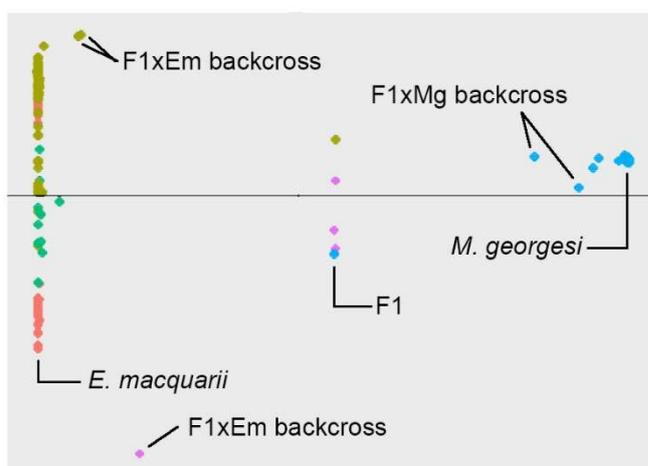
PCoA-1 (98.1%)

References

Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus
    genetic data. *Genetics*, *160*, 1217–1229.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis.
    Biometrika, *53*, 325–338.