

## QUANTITATIVE GENETICS

Title Cluster AnalysisPurpose:

Have students begin with datasets from molecular marker core facility and go through all the steps to produce measures of similarity.

Goal:

Apply the concepts of identity in state to determine coefficients of realized relationship (parentage) among breeding lines.

ALA: How to perform cluster analysis with R

In ALA 4.5 you determined which distance metric would most accurately represent realized genomic relationships among lines. Consider the genotypic scores found in "ALA4.6\_ds.csv". These are from a genotyping facility and are only a small subset of the total data set. This subset consists of genotypic scores at 14 marker loci for 720 RILs representing nine families consisting of a sample of 80 RILs each.

The dataset received from the core facility is not coded in a way that R will recognize if you want to calculate genetic distances.

1. Translate the core facility genotypic scores into scores that R "dist" functions will recognize.
2. We note that the scores needed for the R dist functions are -1, 0, 1. These are SNP scores. How do we translate SNP scores into -1,0,1? What is the reference SNP score, because the -1,0,1 scores are relative to the reference?
3. Lastly, there are missing data in the data set. They are coded NA. NA is not the same as 0 (zero). For our current purposes 0 represents a heterozygote. R "dist" functions will not work for any values outside of the range of three values are included in the data set. The R code (and SAS code) will crash if there are values other than the three categorical metrics that we decided to use. What should be done about data that are not translatable into -1,0,1?
4. Utilize R code in "ALA4.6\_cluster anal.R" to visualize the genetic distance among the RILs using a cluster analysis.

5. Make a decision about whether the RIL's cluster as families.
6. Utilize R code in "ALA4.6\_cluster anal.R" to visualize the distance among the Markers.
7. Make a decision about whether any of the marker loci are linked.