**ALA 5.1 Calculation of genetic relationships based on DNA marker information**

**Prerequisite**

Understanding of:

Detection and visualization of population structure

* 1. Measures of genetic similarity and distance
	2. Principle Component Analysis
	3. Cluster Analysis

**Purpose**

Provide understanding of the steps needed to convert DNA fingerprint information into a cladogram.

**Background**

Understanding the relationships among lines or genotypes of a breeding population or other germplasm are important in various contexts: establishing heterotic groups for hybrid breeding, selection of parents to develop breeding populations, variety protection, identification of duplicates in germplasm collections, among others. DNA markers are a useful tool to determine and quantify genetic relationships.

**Tasks**

**This is from ALA2.1 (should have been addressed earlier, not part of ALA5.1)**

****

Figure 1. AFLP electropherogram. Columns 1-11 show AFLP fingerprints of 11 different genotypes. Column 12 is a size standard and can be ignored for further calculations. AFLP fingerprints were generated using the same restriction enzymes and primer combination, and can thus be directly compared across the 11 genotypes.

1. Convert AFLP fingerprints into a 1-0 data matrix (for each genotype – band combination: a “1” signifies “band present”, a “0” “band missing”)
2. Determine “genetic similarity” (fraction of number of bands in common / all bands) for each pair of genotypes. Display genetic similarities in a matrix.
3. Provide a brief explanation of what you are doing in tasks 1 and 2.
4. Repeat 1-2 for two scenarios (you already completed one of the two above); include and exclude monomorphic DNA marker bands. How do these two scenarios differ, and why ?
5. Assume the 11 genotypes are either highly heterozygou or inbred lines – how does it affect interpretation of your cluster analysis ?

**Tasks for ALA5.1 that are directly connecting to ALA2.1, task 1:**

1. Calculate genetic similarities, while deciding for one of the similarity coefficients presented in the supplementary ppt file (slides 7-9). Explain why you decided for a particular similarity coefficient and create a similarity matrix.
2. Create a cladogram, by using the following “Pair Group Method Algorithm” (PGMA): (**1**) Identify the minimum distance between any two taxa,

(**2**) Combine these two taxa as a single pair,

(**3**) Re-calculate the average distance between this pair and all other taxa to form a new matrix, (**4**) Identify the closest pair in the new matrix, and so on, until the last two clusters are joined.

Note: this algorithm is valid for “dissimilarities” between pairs of genotypes, which is equal to 1-similarities. In other words, a dissimilarity of 0% is equal to a 100% similarity. An example on how to perform PGMA is given on slides 10-12 in the supplementary ppt file.

1. Draw a cladogram / phenogram, and repeat this process with at least one additional and different similarity coefficient (return to task 1). Compare cladograms obtained with different similarity coefficients.

**Tentative answers** (can differ, based on context / assumptions made)