Statistical Analysis of Genetic Data Math 155, Spring 2008 Jo Hardin HW #9

Due Friday April 25

1. Edwards and Cavalli-Sforza (1965) considered the following dissimilarity matrix between six families of bacteria:

	А	В	$\mathbf{C}$	D	Е	F
А	0					
В	5	0				
С	11	10	0			
D	11	6	6	0		
Е	14	13	17	13	0	
F	14	15	21	15	6	0

- (a) Partition the data into two clusters with PAM. (Do the algorithm by hand; don't use R. Though you should feel free to use R to check your answer.)
- (b) Did your objective function go down from the build to the swap stage? Justify your answer numerically.
- (c) Find the silhouette width for each observation.
- (d) Find the average silhouette width of the total data set.
- (e) Find the diameter for each cluster.
- (f) Find the separation for each cluster.
- (g) Is either cluster L or  $L^*$ ? Explain.