Appendix 5 to:

Measures of Genetic Diversity

Transforming data from quantitative variables: An example

We have three characters in four individuals:

- Plant height (m)
- Seed weight (g)
- Diameter of the pollen grain (μ)

Before calculating distances, we first need to standardize the data by using the following formula:

$$\mathbf{m_{stand} = m - \bar{m}/\sigma}$$

After standardization, the measurement units are lost.

	m	m_{stand}	g	g_{stand}	μ	μ_{stand}
Indiv. 1	1.50	0.35	0.02	0.00	80.00	-0.15
Indiv. 2	1.20	-1.41	0.03	1.00	70.00	-1.32
Indiv. 3	1.45	0.06	0.01	-1.00	90.00	1.02
Indiv. 4	1.60	0.94	0.02	0.00	85.00	0.44
Average (X_i)	1.44	—	0.02	—	81.25	—
Deviation (s_i)	0.17	—	0.01	—	8.54	—

Distances may now be calculated for each pair of individuals by applying the formula we already know:

$d_{ij} = [\Sigma(X_{i} - X_{kj})^2]^{1/2}$	
$d12 = [(0.35 - (-1.41))^2 + (0.0 - 1.0)^2 + (-0.15 - (-1.32))^2]^{1/2} = 2.34$	d11 = 0
$d13 = [(0.35 - 0.06)^2 + (0.0 - (-1.0))^2 + (-0.15 - 1.02)^2]^{1/2} = 1.57$	d22 = 0
$d14 = [(0.35 - 0.94)^2 + (0.0 - 0.0)^2 + (-0.15 - 0.44)^2]^{1/2} = 0.83$	d33 = 0
$d23 = [(-1.41 - 0.06)^2 + (1.0 - (-1.0))^2 + (-1.32 - 1.02)^2]^{1/2} = 3.41$	d44 = 0
$d24 = [(-1.41 - 0.94)^2 + (1.0 - 0.0)^2 + (-1.32 - 0.44)^2]^{1/2} = 3.10$	
$d34 = [(0.06 - 0.94)^2 + (-1.0 - 0.0)^2 + (1.02 - 0.44)^2]^{1/2} = 1.45$	

Once we have the pairwise distances, we proceed to find the groups by using the UPGMA method (for details, see slides 58 and 59 in the module).

First, we organize our calculated distance values in a symmetrical table:

	I_1	I_2	I_3	I_4
I_1	0			
I_2	2.34	0		
I_3	1.57	3.41	0	
I_4	0.83	3.10	1.45	0

In the first cycle, we choose the shortest distance, which, in our case, is $\boldsymbol{d}_{1,4}$ = 0.83. A new matrix may therefore be formed by grouping Individual_1 with Individual_4 and calculating the combined distances:

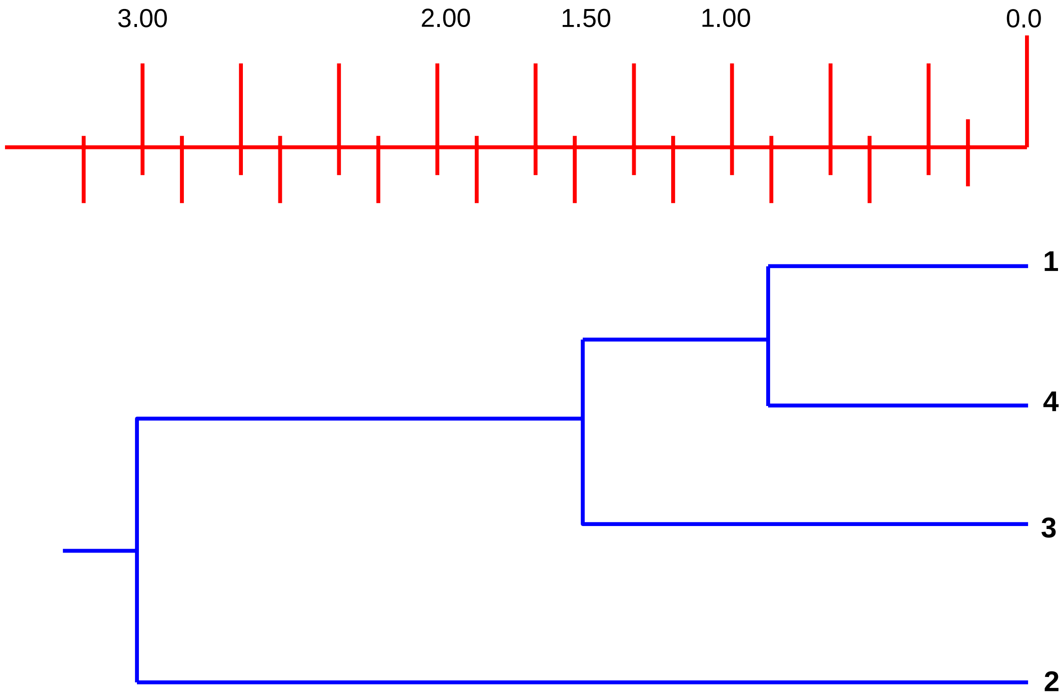
$$\boldsymbol{d}2(1,4) = (\boldsymbol{d}1,2 + \boldsymbol{d}2,4)/2 = (2.34 + 3.10)/2 = 2.72$$
$$\boldsymbol{d}3(1,4) = (\boldsymbol{d}1,3 + \boldsymbol{d}3,4)/2 = (1.57 + 1.45)/2 = 1.51$$

	$I_{1,4}$	I_2	I_3
$I_{1,4}$	0		
I_2	2.72	0	
I_3	1.51	3.41	0

We see that the shortest distance now is between $I_{1,4}$ and I_3. Then, in a new cycle, a new matrix is formed, grouping Individual_2 with group $I_{(1,4)3}$ and calculating the combined distance $\boldsymbol{d}_{((1,4)3)2}$ = 3.07

	$I_{1,4(3)}$	I_2
$I_{1,4(3)}$	0	
I_2	3.07	0

Based on the results above, we can proceed to draw the dendrogram, relating the four individuals of the example:


3.00
2.00
1.50
1.00
0.0
1
4
3
2