Phylogenetic Trees

Phylogenetic Trees

https://highered.mheducation.com/sites/9834092339/student_view0/chapter23/animation_-_phylogenetic_trees.html

Distance-based Methods for Tree Reconstruction

- Distance-based methods are the most popular
  - Neighbour Joining (NJ)
  - UPGMA

Distance-based Methods for Tree Reconstruction ...

- Distance-based methods involve 2 steps:
  - Measure the distances between pairs of sequences in the MSA
  - Transform the distance matrix into a tree

Neighbour Joining (NJ)

- [Naruya Saitou](#) and [Masatoshi Nei](#) in 1987.

- This method attempts to correct the UPGMA method for its strong assumption that the same rate of evolution applies to each branch.

- This method yields an un-rooted tree.

- Input is a distance matrix

- Output is phylogenetic tree

Neighbour Joining (NJ)

Step 1:

The method begins by finding the modified matrix. To do this the net difference of species I from all other taxa is calculated as follows.

$$r_i = \sum_{j=1}^{n} D_{ij}$$

Where $D_{ii}=0$

Neighbour Joining (NJ)

Step 2:

- The rate corrected matrix is the measure between i and j we use in the algorithm is

$$X_D(i,j) = D_{i,j} - \frac{r_i + r_j}{n-2}$$

- Where n represents number of taxas

Neighbour Joining (NJ)

Step 3:

- Join the two nodes/taxa with the smallest $X_D(i,j)$ and define the new branch lengths to this node as

$$l_{iu} = (D_{ij} / 2) + (r_i - r_j) / (2n - 4)$$

$$l_{ju} = D_{ij} - l_{iu}$$

Neighbour Joining (NJ)

Step 4:

- The new distances from node u to all others are defined as

$$D_{ku} = (D_{ik} + D_{jk} - D_{ij})/2$$

Neighbour Joining (NJ)

Step 5:

- Node i and j are removed, n is decreased by one and r_i

is recalculated.

This continuous until only two nodes remain and these two are linked with a branch length $l_{ij} = D_{ij}$

Neighbour Joining (NJ)

	A	B	C	D	E
B	5				
C	4	7			
D	7	10	7		
E	6	9	6	5	
F	8	11	8	9	8

We have in total 6 OTUs (n = 6).

Neighbour Joining (NJ)

	A	B	C	D	E
B	5				
C	4	7			
D	7	10	7		
E	6	9	6	5	
F	8	11	8	9	8

1. We calculate the net divergence r_i for each OTU from all other OTUs

$$r_A = d_{AB} + d_{AC} + d_{AD} + d_{AE} + d_{AF} = 5 + 4 + 7 + 6 + 8 = 30$$

Neighbour Joining (NJ)

	A	B	C	D	E
B	5				
C	4	7			
D	7	10	7		
E	6	9	6	5	
F	8	11	8	9	8

1. We calculate the net divergence r_i for each OTU from all other OTUs
$$r_A = d_{AB} + d_{AC} + d_{AD} + d_{AE} + d_{AF} = 5 + 4 + 7 + 6 + 8 = 30$$
Similarly, we calculate
$$r_B = 42$$
$$r_C = 32$$
$$r_D = 38$$
$$r_E = 34$$
$$r_F = 44$$

Neighbour Joining (NJ)

	A	B	C	D	E
B	5				
C	4	7			
D	7	10	7		
E	6	9	6	5	
F	8	11	8	9	8

1. We calculate the net divergence r_i for each OTU from all other OTUs

$$r_A = d_{AB} + d_{AC} + d_{AD} + d_{AE} + d_{AF} = 5 + 4 + 7 + 6 + 8 = 30$$

Similarly, we calculate

$r_B = 42$

$r_C = 32$

$r_D = 38$

$r_E = 34$

$r_F = 44$

2. Now we calculate a new distance matrix using for each pair of OTUs $M_{ij} = d_{ij} - [r_i + r_j]/(n - 2)$ or in the case of the pair A,B:

$$M_{AB} = d_{AB} - [(r_A + r_B)]/(n - 2) = -13$$

Neighbour Joining (NJ)

	A	B	C	D	E
B	5				
C	4	7			
D	7	10	7		
E	6	9	6	5	
F	8	11	8	9	8

2. Now we calculate a new distance matrix using for each pair of OTUs
$M_{ij} = d_{ij} - [r_i + r_j]/(n-2)$ or in the case of the pair A,B:
$M_{AB} = d_{AB} - [(r_A + r_B)]/(n-2) = -13$

	A	B	C	D	E
B	−13				
C	−11.5	−11.5			
D	−10	−10	−10.5		
E	−10	−10	−10.5	−13	
F	−10.5	−10.5	−11	−11.5	−11.5

Neighbour Joining (NJ)

Now we start with a star tree:

	A	B	C	D	E
B	5				
C	4	7			
D	7	10	7		
E	6	9	6	5	
F	8	11	8	9	8

	A	B	C	D	E
B	−13				
C	−11.5	−11.5			
D	−10	−10	−10.5		
E	−10	−10	−10.5	−13	
F	−10.5	−10.5	−11	−11.5	−11.5

Neighbour Joining (NJ)

3. We choose as neighbours those two OTUs for which M_{ij} is the smallest. These are A and B, and D and E. Let's take A and B as neighbours and we form a new node called U. Now we calculate the branch length from the internal node U to the external OTUs A and B.

$$l_{AU} = d_{AB}/2 + [r_A - r_B]/2(n-2) = 1$$
$$l_{BU} = d_{AB} - l_{AU} = 4$$

	A	B	C	D	E
B	−13				
C	−11.5	−11.5			
D	−10	−10	−10.5		
E	−10	−10	−10.5	−13	
F	−10.5	−10.5	−11	−11.5	−11.5

Neighbour Joining (NJ)

4. Now we define new distances from U to every other terminal node:

$d_{CU} = (d_{AC} + d_{BC} - d_{AB})/2 = 3$
$d_{DU} = (d_{AD} + d_{BD} - d_{AB})/2 = 6$
$d_{EU} = (d_{AE} + d_{BE} - d_{AB})/2 = 5$
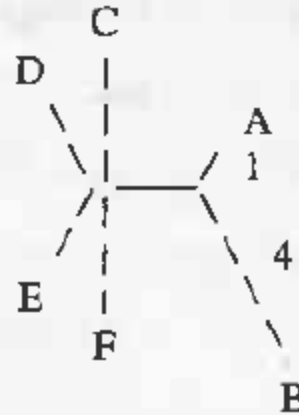$d_{FU} = (d_{AF} + d_{BF} - d_{AB})/2 = 7$

and we create a new matrix:

	U	C	D	E
C	3			
D	6	7		
E	5	6	5	
F	7	8	9	8

Neighbour Joining (NJ)

The resulting tree will be the following:

$$n = n - 1 = 5$$

Then the entire procedure is repeated starting from step 1.

Neighbour Joining (NJ)