

1. Find the smallest value in the distance matrix (equivalent to the pair of taxonomic units closest to each other).
2. Combine the relevant taxonomic units into one group and calculate the distance of this new group to all other taxonomic units. The distance of the taxonomic unit T to this new group S is calculated as the arithmetic mean of the distances between the unit T and all the elements of group S. Furthermore, Group S can be considered as a hypothetical taxonomic unit.
3. If we have more than one taxonomic unit available, repeat the procedure from step 1.

By graphically representing the clustering process over the course of the algorithm described, we obtain the desired phylogenetic tree. The hypothetical taxonomic unit that was created last is its root.

Least Squares Method

In this case, we construct all sorts of phylogenetic trees and evaluate which one is the best. We can make the assessment according to the following prescription:

$$Q = \sum_{i=1}^N \sum_{j=1}^N (D_{i,j} - d_{i,j})^2,$$

where $d_{i,j}$ is the distance between the i and j nodes in the rated phylogenetic tree and $D_{i,j}$ is the distance between the corresponding taxonomic units in the distance matrix.

This procedure requires the design and evaluation of all possible phylogenetic trees, which, like aligning, is a NP-complete problem.

Minimal Evolution Method

The procedure is the same as for the least squares method, but we compare the individual trees by the sum of the lengths of all the branches.

Neighbor-joining

At the beginning, one star tree is created, where there is one internode, and all the solved taxonomic units are represented by leaves. This tree is gradually broken down by clustering the nearest taxonomic units so that the total length of the tree is reduced as much as possible in each step.

Maximum parsimony

The method of maximum parsimony seeks to find such a phylogenetic tree, which requires as few evolutionary events as possible, which would have to occur if this tree matched the course of evolution. In some cases, different weights are assigned to individual evolutionary events when assessing trees, such as when certain nucleotides or amino acids are known to mutate more easily or worse than others.

In the basic variant, this method again requires the design of all possible phylogenetic trees and their subsequent evaluation. Branch and bound method, for example, can be used to streamline searches through trees by selecting only "hopeful" trees.

Method of Maximum Likelihood

Here it is based on statistical methods and posterior probability. We are trying to estimate what the probability is that the statistical hypothesis presented by a particular phylogenetic tree is valid for the data we have available. For the H hypothesis and the D data, this probability can be calculated as follows:

$$P(H|D) = P(H) \cdot \frac{P(D|H)}{P(D)},$$

where $P(D|H)$ is the probability that we observe actual D data, assuming the H hypothesis is true.

The method requires a substitution model, on the basis of which we determine the probability of individual evolutionary changes (mutations). A tree that needs more of these changes to explain available phylogenetic data will have less credibility than a tree that makes do with fewer changes. Beyond that, we also notice the lengths of the individual branches.

Links

Related articles

- Phylogenetic taxonomy
- Evolution

External links

- Phylogenetic tree on Wikipedia

Sources

- Vladimír Hampl: Přednáška z molekulární taxonomie (<http://web.natur.cuni.cz/~vlada/moltax/>)
- Computational phylogenetics on Wikipedia