DNAread

Read DNA Sequences

Description:

This function reads the DNA sequences.

Usage:

DNAread("file")

Arguments:

file: DNA data having the same format used in PHYLIP.

Details:

This function reads DNA seqences from a file.

Note: The file name should be between double quotes " ".

Value:

The result of this function is $K \times n$ matrix containing the DNA data.

See Also:

Testpairs, Testoverall.

Remark:

The data should be in this format:-

The first line contains first the number of species, empty space then the number of nucleotides. The second line contains the first species name, empty space then the nucleotides all in one line without any spaces between them, and so on to the end of the species. Sites which contain characters other than A, C, G and T are omitted from the data.

Example for the Format:

3 12

species1 AAAGGGTCTGGT

species2 ATTGGAACCTAA

species3 CGTAAAGGGTCT

Testpairs

Test for Symmetry of Matched DNA Sequences

Description:
This function tests for symmetry between all the pairs of K matched DNA sequences.

Usage:
Testpairs(DNAdata)

Arguments:
DNAdata: a $K \times n$ matrix containing K matched sequences each of size n.

Details:
This function calculates Bowker's test for symmetry, Stuart's test for marginal symmetry and the test for internal symmetry. It depends on the $K \times n$ matrix.

Value:
A list of three lower triangle matrices:

1. the lower triangle of the matrix contains $(K - 1) \times (K - 1)$ values shows Bowker's test between all the possible pairs of the K sequences.

2. the lower triangle of the matrix contains $(K - 1) \times (K - 1)$ values shows Stuart's test between all the possible pairs of the K sequences.

3. the lower triangle of the matrix contains $(K - 1) \times (K - 1)$ values shows the internal test between all the possible pairs of the K sequences.

See Also:
DNAread, Testoverall.

Testoverall

Overall Test for Marginal Symmetry

Description:
This function tests for marginal symmetry between K matched DNA sequences.

Usage:
Testoverall(DNAdata)

Arguments:
DNAdata: a $K \times n$ matrix containing K matched sequences each of size n.

Details:
This function calculates overall test for marginal symmetry. It depends on the $K \times n$ matrix containing the nucleotides of K species. The test is related to that of Rzhetsky and Nei (1995).

Value:
A list of:

1. chisqstat gives the value of the test statistic.

2. df gives the degrees of freedom.

3. pvalue gives the P-value of the test with 3(K-1) degrees of freedom.

See Also:
DNAread, Testpairs.

Example.

We will give in this example the commands for how to read data and apply the statistical tests, distance function and plot the tree for a set of data.

`The bacterial data`.

Gaultier and Guoy (1995) inferred a phylogeny among five eubacterial species using the small-subunit ribosomal RNA sequences from *Aquafix pyrophilus, Thermotoga maritima, Thermus thermophilus, Deinococcus radiodurans,* and a fifth species chosen from the following genera: *Chlamydia, Spirochaeta, Bacterides, Agrobacterium, Escherichia, Fusobacterium, Clostridium, Bacillus, Micrococcus,* and *Anabaena.*

data=DNAread("bacdata")

Testoverall(data)

Testpairs(data)

```
> data=DNAread("bacdata")


> Testoverall(data)
$chisqstat [1] 103.4499
$df [1] 12
$pvalue [1] 1.110223e-16



> Testpairs(data)
$Bowker.test
          Aquifex Thermotoga Bacillus Deinococcu
Thermotoga  0.1317
Bacillus    0.0000     0.0000
Deinococcu  0.0000     0.0000   0.9948
Thermus     0.5087     0.4146   0.0000     0.0000
```

```
$Stuart.test
           Aquifex Thermotoga Bacillus Deinococcu
Thermotoga  0.3250
Bacillus    0.0000      0.0000
Deinococcu  0.0000      0.0000   0.9457
Thermus     0.7313      0.2673   0.0000      0.0000

$Internal.test
           Aquifex Thermotoga Bacillus Deinococcu
Thermotoga  0.0949
Bacillus    0.2953      0.4174
Deinococcu  0.7539      0.2975   0.9575
Thermus     0.2628      0.5457   0.5440      0.8632
```