

# Package ‘ConCat’

November 24, 2014

**Title** Alignment concatenation and analysis package

**Version** 1.0

**Author** Ambuj Kumar

**Maintainer** Ambuj Kumar <ambuj@ufl.edu>

**Description** Perform alignment concatenation and analysis

**Depends** seqinr

**License** GPL (>= 2)

**URL** <https://github.com/Ambuj-UF/ConCat-R>

## R topics documented:

concat . . . . .	1
entcal . . . . .	2
fevol . . . . .	2
rcv . . . . .	3
read.nexus.data . . . . .	4
rycode . . . . .	5
<b>Index</b>	<b>7</b>

---

concat	<i>Alignment concatenation function</i>
--------	---

---

## Description

Performs sequence alignment concatenation. It adds "?" for missing taxon sequence data.

## Usage

```
concat(ext, form, outform, writeData = TRUE)
```

**Arguments**

ext	Input file extension. example: ".nex"
form	Input sequence alignment format. It can be nexus, phylip, fasta and clustal
outform	Output sequence alignment format. It can be nexus, phylip-relaxed, phylip-interleived, phylip-sequential, fasta and clustal
writeData	If TRUE (default = TRUE) the program creates a concatenated alignment file. Otherwise it return the concatenated dataframe element

**Author(s)**

Ambuj Kumar <ambuj@ufl.edu> Mrinal Mishra <mrnlmishra@gmail.com>

---

entcal	<i>Class "entropy"</i>
--------	------------------------

---

**Description**

Calculates entropy value for the given input alignment file.

**Usage**

```
entcal(file, type = "DNA")
```

**Arguments**

file	Sequence alignment file name specified by either a variable of mode character, or a double-quoted string; if file = "" (the default) then the program terminates with message
type	Alignment type. Can be DNA or Protein

**Author(s)**

Ambuj Kumar <ambuj@ufl.edu>

---

fevol	<i>Detects fast evolving sites</i>
-------	------------------------------------

---

**Description**

To calculate observed variability (OV), all sequences for a given position are compared in a pair-wise fashion. Mismatches are scored as 1 and matches as 0; the mean value amongst all the comparisons for a given position is used as the measure of character variability in the subsequent data sorting. Check reference for details.

**Usage**

```
fevol(file, cutoff = 0.9)
```

**Arguments**

file	Sequence alignment file name specified by either a variable of mode character, or a double-quoted string; if file = "" (the default) then the program terminates with message
cutoff	OV value cutoff for fast evolving site detection. cutoff = 0.9 is set as default.

**Author(s)**

Ambuj Kumar <ambuj@ufl.edu>

**References**

Goremykin VV1, Nikiforova SV, Bininda-Emonds OR. 2010. Automated removal of noisy data in phylogenomic analyses. J Mol Evol. 71:319-31. <http://www.ncbi.nlm.nih.gov/pubmed/20976444>

---

rcv	<i>Relative confidence value</i>
-----	----------------------------------

---

**Description**

Calculates RCV scores for the input alignment file

**Usage**

```
rcv(file, type = "DNA")
```

**Arguments**

file	Sequence alignment file name specified by either a variable of mode character, or a double-quoted string; if file = "" (the default) then the program terminates with message
type	Alignment type. Can be DNA or Protein

**Author(s)**

Ambuj Kumar <ambuj@ufl.edu>

read.nexus.data

*Read Character Data In NEXUS Format***Description**

This function reads a file with sequences in the NEXUS format.

**Usage**

```
read.nexus.data(file)
```

**Arguments**

**file** a file name specified by either a variable of mode character, or a double-quoted string.

**Details**

This parser tries to read data from a file written in a *restricted* NEXUS format (see examples below).

Please see files 'data.nex' and 'taxacharacters.nex' for examples of formats that will work.

Some noticeable exceptions from the NEXUS standard (non-exhaustive list):

- **I**Comments must be either on separate lines or at the end of lines. Examples:  
 [Comment] — **OK**  
 Taxon ACGTACG [Comment] — **OK**  
 [Comment line 1  
 Comment line 2] — **NOT OK!**  
 Tax[Comment]on ACG[Comment]T — **NOT OK!**
- **II**No spaces (or comments) are allowed in the sequences. Examples:  
 name ACGT — **OK**  
 name AC GT — **NOT OK!**
- **III**No spaces are allowed in taxon names, not even if names are in single quotes. That is, single-quoted names are not treated as such by the parser. Examples:  
 Genus\_species — **OK**  
 'Genus\_species' — **OK**  
 'Genus species' — **NOT OK!**
- **IV**The trailing end that closes the matrix must be on a separate line. Examples:  
 taxon AACCGGT  
 end; — **OK**  
 taxon AACCGGT;  
 end; — **OK**  
 taxon AACCCGT; end; — **NOT OK!**
- **V**Multistate characters are not allowed. That is, NEXUS allows you to specify multiple character states at a character position either as an uncertainty, (XY), or as an actual appearance of multiple states, {XY}. This information is not handled by the parser. Examples:  
 taxon 0011?110 — **OK**  
 taxon 0011{01}110 — **NOT OK!**  
 taxon 0011(01)110 — **NOT OK!**

- **VI**The number of taxa must be on the same line asntax. The same applies to nchar. Examples:  
ntax = 12 — **OK**  
ntax =  
12 — **NOT OK!**
- **VII**The word “matrix” can not occur anywhere in the file before the actual matrix command, unless it is in a comment. Examples:  
BEGIN CHARACTERS;  
TITLE 'Data in file "03a-cytochromeB.nex"';  
DIMENSIONS NCHAR=382;  
FORMAT DATATYPE=Protein GAP=- MISSING=?;  
["This is The Matrix"] — **OK**  
MATRIX  
  
BEGIN CHARACTERS;  
TITLE 'Matrix in file "03a-cytochromeB.nex"'; — **NOT OK!**  
DIMENSIONS NCHAR=382;  
FORMAT DATATYPE=Protein GAP=- MISSING=?;  
MATRIX

### Value

A list of sequences each made of a single vector of mode character where each element is a (phylogenetic) character state.

### Author(s)

Johan Nylander <nylander@scs.fsu.edu>

### References

Maddison, D. R., Swofford, D. L. and Maddison, W. P. (1997) NEXUS: an extensible file format for systematic information. *Systematic Biology*, **46**, 590–621.

### Examples

```
## Use read.nexus.data to read a file in NEXUS format into object x
## Not run: x <- read.nexus.data("file.nex")
```

---

rycode

"RY-coding"

---

### Description

Class "RY-coding"

### Usage

```
rycode(file, type = 3)
```

**Arguments**

file	Sequence alignment file name specified by either a variable of mode character, or a double-quoted string; if file = "" (the default) then the program terminates with message
type	RY coding position. If type=3 (Default=3), program performs RY coding at the third codon position whereas if type="all" then program performs RY coding for all the positions

**Author(s)**

Ambuj Kumar <ambuj@ufl.edu>

# Index

\*Topic **file**

read.nexus.data, [4](#)

concat, [1](#)

entcal, [2](#)

fevol, [2](#)

rcv, [3](#)

read.nexus.data, [4](#)

rycode, [5](#)