Package 'stepwise'

February 20, 2015

Title Stepwise detection of recombination breakpoints

Version 0.3

Author Jinko Graham <jgraham@stat.sfu.ca>, Brad McNeney <mcneney@sfu.ca>, Francoise Seillier-Moiseiwitsch <seillier@math.umbc.edu>, R interface by Sigal Blay <sblay@sfu.ca>

Description A stepwise approach to identifying recombination breakpoints in a sequence alignment.

Maintainer Brad McNeney <mcneney@stat.sfu.ca>

Depends R (>= 2.5.0)

License GPL (>= 2)

URL http://stat.sfu.ca/statgen/research/stepwise.html

Repository CRAN

Date/Publication 2012-11-24 08:11:17

NeedsCompilation yes

R topics documented:

`maxchi` *Stepwise detection of recombination breakpoints using MaxChi*

Description

Stepwise detection of recombination breakpoints using the maximum chi-square (MaxChi) method at each step

Usage

```
maxchi(input_file, breaks, winHalfWidth, permReps)
```

Arguments

`input_file`	character string indicating the name of a Phylip format data input file
`breaks`	an integer vector of ordered site(s) just before the previously declared breakpoints
`winHalfWidth`	the window half width to use
`permReps`	the number of Monte Carlo replicates to use for the permutation distribution

Details

The maxchi function implements the maximum chi-square (MaxChi) method for detecting recombination breakpoints (Maynard Smith 1992) using a moving window of fixed width. Breakpoints detected in previous steps of a stepwise search may be conditioned upon.

For a given position of the moving window on the sequence alignment, and for a given pair of sequences, a chi-square statistic is computed to compare two proportions: the proportion of sites at which the sequences agree in the left half-window and the proportion of sites in at which the sequences agree in the right half-window. Discordance between the two proportions may reflect a recombination event, located at the window centre, in the history of the two sequences. The maximum chi-square over all sequence pairs is regarded as a summary of the evidence for recombination at the window centre. The individual chi-square statistics may also be of interest for suggesting pairs of sequences segments that descend from historical recombination events. Significance of observed chi-square statistics is assessed by a Monte Carlo permutation test. When conditioning on breakpoints proposed at previous steps of a stepwise search, permutation is restricted to sites within blocks defined by the previously proposed breakpoints, as described by Graham et al. (2004).

Value

`polyposn`	The site numbers of all ungapped polymorphic sites in the alignment
`chisqs`	Observed chi-square statistics that exceed the 90th percentile of the permutation null distribution
`winlocs`	Window centres corresponding to the chi-square statistics in `chisqs`
`pairmem1`	First member of each pair that lead to a significant chi-square statistic in `chisqs`
`pairmem2`	Second member of each pair
`quants`	90th and 95th percentiles of the permutation distribution

Author(s)

Brad McNeney <mcneney@stat.sfu.ca>, Jinko Graham <jgraham@stat.sfu.ca>, Sigal Blay <sblay@sfu.ca>

References

Graham J, McNeney B and Seillier-Moiseiwitsch F (2004). Stepwise detection of recombination breakpoints in sequence alignments. *Bioinformatics* Sep 23; [Epub ahead of print]

Maynard Smith J (1992). Analyzing the mosaic structure of genes. *J Mol Evol*, **34**:126-129.

http://stat-db.stat.sfu.ca/stepwise

See Also

summary.maxchi, phylpro

Examples

```
demo(maxchi)
```

`phylpro` *Stepwise detection of recombination breakpoints using Phylpro*

Description

Stepwise detection of recombination breakpoints using phylogenetic profiling (Phylpro) at each step

Usage

```
phylpro(input_file, breaks, winHalfWidth, permReps)
```

Arguments

`input_file`	character string indicating the name of a Phylip format data input file
`breaks`	an integer vector of ordered site(s) just before the previously declared breakpoints
`winHalfWidth`	the window half width to use
`permReps`	the number of Monte Carlo replicates to use for the permutation distribution

Details

The phylpro function implements phylogenetic profiling (Phylpro) for detecting recombination breakpoints (Weiller 1998) using a moving window of fixed width. Breakpoints detected in previous steps of a stepwise search may be conditioned upon.

For a given position of the moving window on the sequence alignment, and for a given "target" sequence, a correlation is computed to compare two distance vectors: the distance between the target sequence and all other sequences in the left half-window and the distance between the target

sequence and all others in the right half-window. The pair-wise distance measure used is the proportion of sites at which the sequences differ. Discordance between the two distance vectors may reflect a recombination event, located at the window centre, in the history of the target sequence. The minimum correlation over all target sequences is regarded as a summary of the evidence for recombination at the window centre. The individual correlations for the target sequences may also be of interest for suggesting sequence segments that descend from historical recombination events. Significance of observed correlation statistics is assessed by a Monte Carlo permutation test. When conditioning on breakpoints proposed at previous steps of a stepwise search, permutation is restricted to sites within blocks defined by the previously proposed breakpoints, as described by Graham et al. (2004).

Value

`polyposn`	The site numbers of all ungapped polymorphic sites in the alignment
`corrs`	Observed correlations that exceed the 90th percentile of the permutation null distribution
`winlocs`	Window centres corresponding to the correlations in `corrs`
`target.seqs`	The target sequence that lead to a significant correlation in `corrs`
`quants`	90th and 95th percentiles of the permutation distribution

Author(s)

Brad McNeney <mcneney@stat.sfu.ca>, Jinko Graham <jgraham@stat.sfu.ca>, Sigal Blay <sblay@sfu.ca>

References

Graham J, McNeney B and Seillier-Moiseiwitsch F (2004). Stepwise detection of recombination breakpoints in sequence alignments. *Bioinformatics* Sep 23; [Epub ahead of print]

Weiller G (1998). Phylogenetic profiles: A graphical method for detecting genetic recombination in homologous sequences. *Mol Biol Evol*, **15**:326-335.

http://stat-db.stat.sfu.ca/stepwise

See Also

`summary.phylpro`, `maxchi`

Examples

```
demo(phylpro)
```

stepwise.seqs *Simulated sequence alignment from the stepwise package*

Description

Simulated sequence alignment comprised of 30 sequences, each of length 1000 bases.

Usage

```
data(stepwise.seqs)
```

Format

A 30-by-1000 matrix of mode character. Each row represents a genetic sequence. Each column is a nucleotide in that genetic sequence.

Details

A simulated data set analysed by Graham et al. (2004) to illustrate the stepwise recombination detection methods. The alignment was simulated using Treevolve version 1.3 (Grassly and Holmes 1997). For details on the parameter values used in the simulation, see Graham et al. (2004).

Source

Graham J, McNeney B and Seillier-Moiseiwitsch F (2004). Stepwise detection of recombination breakpoints in sequence alignments. *Bioinformatics* Sep 23; [Epub ahead of print]

References

Grassley NC and Holmes EC (1997). A likelihood method for the detection of selection and recombination using nucleotide sequences. *Mol Biol Evol*, **14(3)**: 239-247

Examples

```
data(stepwise.seqs)
dim(stepwise.seqs) # 30 x 1000
## Not run:
# write the 30 sequences to a Phylip-format input file
cat(paste(nrow(stepwise.seqs),ncol(stepwise.seqs),"\n"),
file="stepwise.phy")
write.table(stepwise.seqs,file="stepwise.phy",append=TRUE,quote=FALSE,
sep="", row.names=TRUE,col.names=FALSE)

## End(Not run)
```

summary.maxchi *Summary of a maxchi object*

Description

Print a summary of a maxchi object

Usage

```
 ## S3 method for class 'maxchi'
summary(object, ...)
```

Arguments

object	A list object of class maxchi output by the maxchi function
...	Additional arguments to the summary function (currently unused)

Details

The output of maxchi includes information on all site- and pair-specific chi-squares that exceed the 90th percentile of the permutation distribution. The summary.maxchi function computes maximum site-specific measures (i.e. maxima over all sequence-pair chi-squares from the same site) and reports all sequence pairs that tied for these maximum values.

Value

siteWinlocs	Window locations with significant recombination signal
siteChisqs	The maximum chi-square at these window locations
pairs	The pair(s) that give the maximum chi-square statistic at each window location

Author(s)

Brad McNeney <mcneney@stat.sfu.ca>, Jinko Graham <jgraham@stat.sfu.ca>, Sigal Blay <sblay@sfu.ca>

See Also

maxchi

Examples

```
demo(maxchi)
```

summary.phylpro *Summary of a phylpro object*

Description

Print a summary of a phylpro object

Usage

```
 ## S3 method for class 'phylpro'
summary(object, ...)
```

Arguments

object	A list object of class phylpro output by the phylpro function
...	Additional arguments to the summary function (currently unused)

Details

The output of phylpro includes information on all site- and target sequence-specific correlations that exceed the 90th percentile of the permutation distribution. The summary.phylpro function computes minimum site-specific measures (i.e. minima over all target-sequence correlations from the same site) and reports all target sequences that tied for these minimum values.

Value

siteWinlocs	Window locations with significant recombination signal
siteCorrs	The minimum correlation at these window locations
target.seqs	The target sequence(s) that give the minimum correlation at each window location

Author(s)

Brad McNeney <mcneney@stat.sfu.ca>, Jinko Graham <jgraham@stat.sfu.ca>, Sigal Blay <sblay@sfu.ca>

See Also

phylpro

Examples

```
demo(phylpro)
```

Index