Package: vhica (via r-universe)

September 3, 2024

Type Package

Title Vertical and Horizontal Inheritance Consistence Analysis

Version 0.2.8

Date 2023-03-07

Author Arnaud Le Rouzic

Suggests ape, plotrix, parallel, seqinr, gtools

Maintainer Arnaud Le Rouzic
`<arnaud.le-rouzic@universite-paris-saclay.fr>`

Description The ``Vertical and Horizontal Inheritance Consistence Analysis'' method is described in the following publication: ``VHICA: a new method to discriminate between vertical and horizontal transposon transfer: application to the mariner family within Drosophila'' by G. Wallau. et al. (2016) <DOI:10.1093/molbev/msv341>. The purpose of the method is to detect horizontal transfers of transposable elements, by contrasting the divergence of transposable element sequences with that of regular genes.

License GPL-2

NeedsCompilation no

Date/Publication 2023-03-08 08:10:02 UTC

Repository https://lerouzic.r-universe.dev

RemoteUrl https://github.com/cran/vhica

RemoteRef HEAD

RemoteSha 2928afa300d91c3a81ad1f3faac39f1850a10dcf

Contents

`vhica-package` *Vertical and Horizontal Inheritance Consistence Analysis*

Description

The package implements the VHICA method described in Wallau et al. (in prep). The purpose of the method is to detect horizontal transfers of transposable elements, by contrasting the divergence of transposable element sequences with that of regular genes. Two files should be provided, for both a set of reference genes and transposable element sequences: (i) pairwise divergence across species (e.g., dS), (ii) codon usage bias for all genes and elements in all species.

Details

Package:	vhica
Type:	Package
License:	GPL-v2

This package contains three main functions.

- `read.vhica`: reads sequence files and generates an object of class `vhica` that will be used for further analysis.
- `plot.vhica`: plots the VHICA regression between two species, and displays how far transposable elements (or any kind of other sequences) are from the reference genes.
- `image.vhica`: plots the consistency of a specific element across all species, which makes it possible to build evolutionary scenarios.

In addition, it provides tools to calculate divergence (`div`) and codon usage bias (`CUB`), which are necessary to apply the VHICA method.

Author(s)

Implementation: Arnaud Le Rouzic <arnaud.le-rouzic@universite-paris-saclay.fr>
Scientists who designed the method: Gabriel Wallau, Aurélie Hua-Van, Arnaud Le~Rouzic.

Maintainer: Arnaud Le Rouzic <arnaud.le-rouzic@universite-paris-saclay.fr>

Repository: https://github.com/lerouzic/vhica

References

Gabriel Luz Wallau, Arnaud Le Rouzic, Pierre Capy, Elgion Loreto, Aurélie Hua-Van. VHICA: A new method to discriminate between vertical and horizontal transposon transfer: application to the mariner family within Drosophila. Molecular biology and evolution 33 (4), 1094-1109.

Examples

```
file.cb <- system.file("extdata", "mini-cbias.txt", package="vhica")
file.div <- system.file("extdata", "mini-div.txt", package="vhica")
file.tree <- if(require("ape")) system.file("extdata", "phylo.nwk", package="vhica") else NULL
vc <- read.vhica(cb.filename=file.cb, div.filename=file.div)
plot(vc, "dere", "dana")
im <- image(vc, "mellifera:6", treefile=file.tree, skip.void=TRUE)
summary(im)
```

`CUB` *Computes the Codon Usage Bias of DNA sequences*

Description

The function reads aligned sequences in a fasta file and estimates the codon usage bias for each sequence. Several methods exist to estimate CUB; so far, only the "Effective Number of Codons" (ENC) calculation is available.

Usage

```
CUB(file = NULL, sequence = NULL, method = "ENC")
```

Arguments

`file`	FASTA file in which aligned sequences are stored.
`sequence`	Alternatively, the result of seqinr::read.fasta.
`method`	The method used to compute CUB. "ENC": Effective Number of Codons, as described in Wright (1990).

Value

A named vector of CUB scores. Names correspond to sequence names in the dataset.

Author(s)

Aurelie Hua-Van and Arnaud Le Rouzic.

References

Wright, F. (1990). The 'effective number of codons' used in a gene. Gene, 87(1), 23-29.

See Also

`div`

Examples

```
seq.file <- system.file("extdata/Genes", "Amd.fas", package="vhica")
CUB(seq.file)
```

`div` *Computation of the synonymous divergence between sequences*

Description

The divergence between DNA sequences can be synonymous (neutral) or non-synonymous. Synonymous differences are generally considered as a better proxy for evolutionary divergence, as it is not affected by selection. This function computes the synonymous divergence between sequences.

Usage

```
div(file = NULL, sequence = NULL, sqs = NULL, method = "LWL85",
 pairwise = TRUE, max.lim = 3)
```

Arguments

`file`	FASTA file in which aligned sequences are stored.
`sequence`	Alternatively, the result of seqinr::read.fasta.
`sqs`	Vector of sequence names to be compared. If not provided, all pairwise comparisons will be performed.
`method`	Method used to compute the divergence. So far, only the LWL85 method (from Li et al. 1985).
`pairwise`	Boolean: should the divergence be calculated for each pair of sequences or on the whole dataset? This is of particular importance when indels (gaps) are present in sequences, as codons with gaps are generally discarded by most methods. Setting this option to `TRUE` is thus more likely to give accurate results with multiple-gap sequences, but the calculation will also be slower.
`max.lim`	Maximum value for divergence. Depending on the algorithm, various corrections can bring the divergence value above 100%. Values larger than `max.lim` will be replaced by `NA`s, as they can be problematic for further statistical tests.

Details

The `LWL85` method is a wrapper around the `kaks` function from the `seqinr` package.

Value

A 3-column data frame with the following fields:

- div: The divergence score
- sq1: The first sequence in the comparison
- sq2: The second sequence in the comparison

Author(s)

Aurelie Hua-Van and Arnaud Le Rouzic

References

Li, W. H., Wu, C. I., & Luo, C. C. (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Molecular biology and evolution, 2(2), 150-174.

See Also

`CUB`

Examples

```
seq.file <- system.file("extdata/Genes", "Amd.fas", package="vhica")
div(seq.file)
```

`image.vhica` *Consistency matrix for a transposable element in the VHICA analysis.*

Description

This function plots a composite figure summarizing the evolutionary properties of a transposable element in a group of related species. Discrepancies may indicate horizontal transfers.

Usage

```
## S3 method for class 'vhica'
image(x, element = "", H1.test = "bilat", treefile = NULL,
skip.void = FALSE, species = NULL, p.threshold = 0.05,
p.adjust.method = "bonferroni",  ncolors = 1024,
main = element, threshcol=0.1, colsqueeze=1,
species.font.family="mono", species.font.cex=1,
    max.spname.length=10, ...)
```

Arguments

`x`	An object of class `vhica`, created by the function `read.vhica`.
`element`	The name of the transposable element, as specified in the data files. If the element is not present in the data, the program halts.
`H1.test`	A value among `"bilat"`, `"lower"`, or `"greater"`.
`treefile`	A Newick file containing a phylogenetic tree. Species names in the tree need to match the data. If absent, the figure will not display the phylogenetic relationship (which makes the interpretation impossible).
`skip.void`	Whether or not the figure should show species that do not contain the transposable element.
`species`	A named character vector to display pretty species names. The names of the vector are the real species names (as they will appear in the figure), the content of the vector are species codes as in the data files.

`p.threshold`	Threshold for the p-value (above which the color gradient increases).
`p.adjust.method`	As documented in `p.adjust`.
`ncolors`	Number of colors in the gradient.
`main`	Main title of the figure (default: the name of the transposable element).
`threshcol`	Part of the color spectrum devoted to non-significant values.
`colsqueeze`	Values larger than 1 shrink the color gradient around the threshold.
`species.font.family`	Font family for the species names.
`species.font.cex`	Font size of the species names.
`max.spname.length`	Maximum length of species names. Longer labels are truncated.
`...`	Further arguments to the generic function `image`.

Details

The figure displays in blue TE copies that are more divergent than expected between species, and in red copies that are less divergent than expected. If several lineages of copies are present in a species, the table will be split to display both lineages. Keys for the interpretation of the pattern and the reconstruction of an evolutionary scenario are provided in the original publication.

Value

The function returns (invisibly) a list of elements (object of class `vhicaimage`) which can be used for further analysis: `tree` contains the phylogenetic tree (object of class `"phylo"`), `species` is the vector of species, `stats` is a matrix of log10(P-values) (positive elements are minus log10(P) corresponding to positive residuals), codedS is a matrix reminding the divergence rates from the data. Calling the method `summary.vhicaimage` on this object returns a nicely formatted data frame.

Author(s)

Implementation: Arnaud Le Rouzic <lerouzic@legs.cnrs-gif.fr>
Scientists who designed the method: Gabriel Wallau, Aurélie Hua-Van, Arnaud Le~Rouzic.

References

Gabriel Luz Wallau, Arnaud Le Rouzic, Pierre Capy, Elgion Loreto, Aurélie Hua-Van. VHICA: A new method to discriminate between vertical and horizontal transposon transfer: application to the mariner family within Drosophila. Molecular biology and evolution 33 (4), 1094-1109.

See Also

`read.vhica`, `plot.vhica`, `summary.vhicaimage`.

Examples

```
file.cb <- system.file("extdata", "mini-cbias.txt", package="vhica")
file.div <- system.file("extdata", "mini-div.txt", package="vhica")
file.tree <- if(require("ape")) system.file("extdata", "phylo.nwk", package="vhica") else NULL
vc <- read.vhica(cb.filename=file.cb, div.filename=file.div)
plot(vc, "dere", "dana")
im <- image(vc, "mellifera:6", treefile=file.tree, skip.void=TRUE)
summary(im)
```

`plot.vhica` *Plots a VHICA regression between two species.*

Description

The VHICA method is based on a contrast between gene divergence and codon usage bias. A regression between divergence and codon usage provides a reference, and sequences of interest (typically, transposable elements) will be compared to the reference genes.

Usage

```
## S3 method for class 'vhica'
plot(x, sp1 = NULL, sp2 = NULL, ...)
```

Arguments

`x`	An object of class `vhica`, created by `read.vhica`.
`sp1`	Name of the first species, as in the data files.
`sp2`	Name of the second species, as in the data files.
`...`	Additional options for `plot`.

Details

The resulting figure displays genes as circles, and transposable elements as symbols.

Author(s)

Implementation: Arnaud Le Rouzic <lerouzic@legs.cnrs-gif.fr>
Scientists who designed the method: Gabriel Wallau, Aurélie Hua-Van, Arnaud Le~Rouzic.

References

Gabriel Luz Wallau, Arnaud Le Rouzic, Pierre Capy, Elgion Loreto, Aurélie Hua-Van. VHICA: A new method to discriminate between vertical and horizontal transposon transfer: application to the mariner family within Drosophila. Molecular biology and evolution 33 (4), 1094-1109.

See Also

`read.vhica`, `image.vhica`

Examples

```
file.cb <- system.file("extdata", "mini-cbias.txt", package="vhica")
file.div <- system.file("extdata", "mini-div.txt", package="vhica")
file.tree <- if(require("ape")) system.file("extdata", "phylo.nwk", package="vhica") else NULL
vc <- read.vhica(cb.filename=file.cb, div.filename=file.div)
plot(vc, "dere", "dana")
image(vc, "mellifera:6", treefile=file.tree, skip.void=TRUE)
```

`read.vhica` *Reads divergence and codon usage data files for the VHICA method.*

Description

The VHICA method relies on two sources of information: (i) the divergence between sequences, and (ii) the codon usage bias. This function reads two data files and creates an object of class `vhica` that can be further explored by `plot.vhica` and `image.vhica`. Input can be either (1) two vectors of fasta file names (one for the genes, one for the putatively transfered genes), or (2) already processed files containing codon usage bias and divergence data (see Details).

Usage

```
read.vhica(gene.fasta=NULL, target.fasta=NULL,
 cb.filename=NULL, div.filename=NULL,
 reference = "Gene", divergence = "dS",
 CUB.method="ENC", div.method="LWL85", div.pairwise=TRUE,
 div.max.lim=3, species.sep="_", gene.sep=".", family.sep=".", ...)
```

Arguments

`gene.fasta`	Sequence files (FASTA format) containing the aligned sequences (respecting the translation phase) for all species of the reference genes.
`target.fasta`	Sequence files (FASTA format) containing the aligned sequence of the putatively transfered genes.
`cb.filename`	File name for the codon usage bias data. If FASTA files are provided, this file will be created.
`div.filename`	File name for the divergence data. If FASTA files are provided, this file will be created.
`reference`	Name of the reference type in the codon usage file. Default is "Gene".
`divergence`	Name of the divergence column in the divergence file. Default is "dS".
`CUB.method`	Method to be used for Codon Usage Bias calculation (see `CUB`).
`div.method`	Method to be used for divergence calculation (see `div`).
`div.pairwise`	Whether divergence should be calculated from the whole alignment of between pairs of sequences (see `div`).
`div.max.lim`	Maximum divergence score. Estimated divergence much larger than 100% are likely to be problematic and should not be considered.

`species.sep`	Separator for species (or equivalent) labels in sequence names. Any character string following this separator will be disregarded – be careful about potential duplicates.
`gene.sep`	Separator for gene names from gene sequence files.
`family.sep`	Separator for target sequence sub-families.
`...`	Further parameters for the internal function `.reference.regression`.

Details

Details about CUB and divergence calculations can be found in `CUB` and `div`. If CUB and/or divergence need to be calculated by an external program, it is possible to provide them in the following format:

- Codon usage bias Example of data file:

```
        Type    sp1     sp2     sp3
CG4231  Gene    42.3    51.1    47.2
CG2214  Gene    47.2    44.9    53.2
Pelem1  TE      36.2    47.0    44.4
...
```

  - Row names (or first column)sequence index
  - Type whether the sequence is a reference (default: Gene) or a focal sequence (transposable element, ...)
  - Following columns a measurement of codon bias (ENC, CBI...) for every species
- Divergence Example of data file:

```
seq     dS      sp1     sp2
CG4231  0.84    Dmel    Dsim
CG4231  0.46    Dmel    Dana
CG4231  0.58    Dsim    Dana
CG2214  0.10    Dmel    Dsim
...
```

  - First column (or row names): sequence index
  - Second column: divergence measurement
  - Columns 3 and 4: the pair of species on which the divergence is calculated
  - Row names and Col names are allowed but disregarded

Value

The function returns an object of class `vhica`, a list containing:

- cbias: A codon bias array
- div: The divergence matrix
- reg: The result of all pairwise regressions
- reference: The `reference` option
- target: The sequence type that is not the reference
- divergence: The `divergence` option
- family.sep: The character used to indicate TE sub-families

Author(s)

Implementation: Arnaud Le Rouzic
Scientists who designed the method: Gabriel Wallau, Aurelie Hua-Van, Arnaud Le Rouzic.

References

Gabriel Luz Wallau, Arnaud Le Rouzic, Pierre Capy, Elgion Loreto, Aurelie Hua-Van. VHICA: A new method to discriminate between vertical and horizontal transposon transfer: application to the mariner family within Drosophila. Molecular biology and evolution 33 (4), 1094-1109.

See Also

`plot.vhica`, `image.vhica`, `CUB`, `div`

Examples

```
file.cb <- system.file("extdata", "mini-cbias.txt", package="vhica")
file.div <- system.file("extdata", "mini-div.txt", package="vhica")
file.tree <- if(require("ape")) system.file("extdata", "phylo.nwk", package="vhica") else NULL
vc <- read.vhica(cb.filename=file.cb, div.filename=file.div)
plot(vc, "dere", "dana")
image(vc, "mellifera:6", treefile=file.tree, skip.void=TRUE)
```

`summary.vhicaimage` *Provides a data.frame that nicely displays the information returned by* `image.vhica`.

Description

The `image.vhica` routine displays visually the statistical support for horizontal transfers, and can return an object of class `vhicaimage`. The current `summary` method reorganizes this object into a dataframe that can be displayed or reused in further analysis.

Usage

```
## S3 method for class 'vhicaimage'
summary(object, divrate=NA, p.thresh=1, ...)
```

Arguments

`object`	An object of class `vhicaimage`, created by `image.vhica`.
`divrate`	Optional divergence rate (in neutral substitutions per Myr).
`p.thresh`	Optional p-value threshold. By default, all data is returned.
`...`	Additional options for `summary` (unused).

Value

The resulting `data.frame` has 4 or 5 columns. The two first columns are `sp1` and `sp2`, the two species between which the horizontal transfer is tested, in an arbitrary order. The column `p.value` contains the p-value calculated as in `image.vhica` (including the possible correction for multiple testing). The `dS` column is a copy of the corresponding divergence from the original data. The last, optional column `Time(Mya)` is a molecular clock estimate of the time of divergence between the two sequences, based on the divergence rate (when provided).

Author(s)

Implementation: Arnaud Le Rouzic <lerouzic@legs.cnrs-gif.fr> and Gabriel Wallau
Scientists who designed the method: Gabriel Wallau, Aurélie Hua-Van, Arnaud Le~Rouzic.

References

Gabriel Luz Wallau, Arnaud Le Rouzic, Pierre Capy, Elgion Loreto, Aurélie Hua-Van. VHICA: A new method to discriminate between vertical and horizontal transposon transfer: application to the mariner family within Drosophila. Molecular biology and evolution 33 (4), 1094-1109.

See Also

`read.vhica`, `image.vhica`

Examples

```
file.cb <- system.file("extdata", "mini-cbias.txt", package="vhica")
file.div <- system.file("extdata", "mini-div.txt", package="vhica")
file.tree <- if(require("ape")) system.file("extdata", "phylo.nwk", package="vhica") else NULL
vc <- read.vhica(cb.filename=file.cb, div.filename=file.div)
plot(vc, "dere", "dana")
im <- image(vc, "mellifera:6", treefile=file.tree, skip.void=TRUE)
summary(im)
```

Index