Package 'PhyInformR'

November 13, 2016

Type Package

Title Rapid Calculation of Phylogenetic Information Content

Version 1.0

Date 2016-10-29

Maintainer Alex Dornburg <alex.dornburg@naturalsciences.org>

Description
Enables rapid calculation of phylogenetic information content using the latest advances in phylogenetic informativeness based theory. These advances include modifications that incorporate uneven branch lengths and any model of nucleotide substitution to provide assessments of the phylogenetic utility of any given dataset or dataset partition. Also provides new tools for data visualization and routines optimized for rapid statistical calculations, including approaches making use of Bayesian posterior distributions and parallel processing. Users can apply these approaches toward screening datasets for phylogenetic/genomic information content.

License GPL-3

Depends R (>= 3.3.0)

Encoding UTF-8

LazyData true

Imports phytools, splines, gplots, RColorBrewer, foreach, iterators, geiger, doParallel, gridExtra, hexbin, PBSmodelling, ape, ggplot2

NeedsCompilation no

Author Alex Dornburg [aut, cre, cph],
J. Nick Fisk [aut, cph],
Jules Tamagnan [aut, cph],
Jeffrey Townsend [aut, cph]

Repository CRAN

Date/Publication 2016-11-13 23:13:37

R topics documented:

---

allmodel.signal.noise *A function to compute results of QIRP, QIHP, and QIPP accomodating uneven branch lengths and user defined substitution models or base frequency distributions*

Description

This function calculates QIRP, QIHP, and QIPP across an uneven quartet while allowing for a user to define a substitution model and base frequencies.

Usage

```
allmodel.signal.noise(a, b, c, d, e, f, internode, Pi_T, Pi_C, Pi_A, Pi_G, rate_vector)
```

Arguments

a	substitution model parameter. a=rTC
b	substitution model parameter. b=rAT
c	substitution model parameter. c=rGT
d	substitution model parameter
e	substitution model parameter. e=rGC
f	substitution model parameter. f=rGA
internode	Four quartet branch lengths and internode c(62.49,62.49,65,65,3.2)->internode
Pi_T	Base Frequency of T
Pi_C	Base Frequency of C

Pi_A	Base Frequency of A
Pi_G	Base Frequency of G
rate_vector	site rate vector, converted to class 'matrix'

Details

Please note that base frequencies must sum to 1.

Value

Function returns a matrix of QIRP, QIHP. and QIPP values

Author(s)

A. Dornburg and J.N. Fisk

References

Townsend, J. P., Su, Z., and Tekle, Y. I. "Phylogenetic Signal and Noise: Predicting the Power of a Data Set to Resolve Phylogeny" Systematic biology 61, no. 5 (2012): 835–849. Su, Z., Zhuo, S., Zheng, W., Francesc, L.-G., and Townsend, J. P. "The Impact of Incorporating Molecular Evolutionary Model into Predictions of Phylogenetic Signal and Noise" Frontiers in Ecology and Evolution 2, (2014): doi:10.3389/fevo.2014.00011, Available at http://dx.doi.org/10.3389/fevo.2014.00011 Su, Z. and Townsend, J. P. "Utility of Characters Evolving at Diverse Rates of Evolution to Resolve Quartet Trees with Unequal Branch Lengths: Analytical Predictions of Long-Branch Effects" BMC evolutionary biology 15, (2015): 86.

Examples

```
as.matrix(rag1)->rates
a<-1
b<-1
c<-1
d<-1
e<-1
f<-1
Pi_T<-.25
Pi_C<-.25
Pi_A<-.25
Pi_G<-.25
internode<-c(12,12,15,15,.37)
allmodel.signal.noise(a, b, c, d, e, f, internode, Pi_T, Pi_C, Pi_A, Pi_G, rates)
```

Approximator   *Quantify Quartet Resolution Probabilities Using 2012 Formulation*

Description

Quantify QIRP, QIHP, or QIPP using the equations from Townsend et al. 2012.

Usage

```
Approximator(t, t0, rateVector, s)
```

Arguments

t	Time from tip of tree to focal internode
t0	Focal internode length
rateVector	An object containing a vector of site rates transformed to class "matrix"
s	A number representing the character state space that generated the site rates (e.g., s=2 for binary data)

Author(s)

A. Dornburg

References

Townsend, J. P., Su, Z., and Tekle, Y. I. "Phylogenetic Signal and Noise: Predicting the Power of a Data Set to Resolve Phylogeny" Systematic biology 61, no. 5 (2012): 835–849.

Examples

```
as.matrix(rag1)->rr
Approximator(100,0.5,rr,3)
```

---

cluster.signal.noise   *Quantify Quartet Resolution Probabilities Using 2012 Formulation and MCMC approach*

Description

Quantify Quartet Resolution Probabilities (QIRP, QIHP, and QIPP) Using 2012 Formulation and MCMC approach, outputs results to file

Usage

```
cluster.signal.noise(t, t0, rateVector, nsims, s, filename, imagename, image = "FALSE")
```

Arguments

t	Time from tip of tree to focal internode
t0	Focal internode length
rateVector	An object containing a vector of site rates transformed to class "matrix"
nsims	number of simulations
s	A number representing the character state space that generated the site rates (e.g., s=2 for binary data)
filename	Output file name for results of calculation
imagename	Output file name for results of visualization
image	True or False if image should visualized

Details

Please note that the two files must have different names

Value

returns QIRP, QIHP, and QIPP values

Author(s)

A. Dornburg

References

Townsend, J. P., Su, Z., and Tekle, Y. I. "Phylogenetic Signal and Noise: Predicting the Power of a Data Set to Resolve Phylogeny" Systematic biology 61, no. 5 (2012): 835–849.

Examples

```
as.matrix(rag1)->rr
Approximator(100,0.5,rr,3)

cluster.signal.noise(100, 2, rr, 10,
3, filename="output_values",
imagename="output_graphic", image = "TRUE")
```

---

defined.multi.profile *A function to plot phylogenetic informativeness profiles based on site rate cutoffs by partition*

Description

A function to plot phylogenetic informativeness profiles based on site rate cutoffs (e.g., loci)

Usage

```
defined.multi.profile(rate.vector, tree, breaks, values = "display")
```

Arguments

rate.vector	site rate vector, converted to class 'matrix'
tree	A phylogenetic tree
breaks	bounds for each profile
values	'display' or 'off': whether to return PI values for each time

Value

Graphical display of phylogenetic informativeness

Author(s)

A. Dornburg

References

Townsend, J. P. "Profiling Phylogenetic Informativeness" Systematic biology 56, no. 2 (2007): 222–231. Dornburg, A., Townsend, J. P., Friedman, M., and Near, T. J. "Phylogenetic Informativeness Reconciles Ray-Finned Fish Molecular Divergence Times" BMC evolutionary biology 14, (2014): 169. Prum, R. O., Berv, J. S., Alex, D., Field, D. J., Townsend, J. P., Lemmon, E. M., and Lemmon, A. R. "A Comprehensive Phylogeny of Birds (Aves) Using Targeted next-Generation DNA Sequencing" Nature 526, no. 7574 (2015): 569–573.

Examples

```
library("ape")
library("splines")
read.tree(system.file("extdata","Prumetal_timetree.phy",package="PhyInformR"))->tree
prumetalrates->Prates
as.matrix(Prates)->rates
lower<-c(0, 1594)
upper<-c(1595,2787)
breaks<-cbind(lower,upper)
defined.multi.profile(rates, tree, breaks, values = "off")
```

```
informativeness.profile
```
A function to plot phylogenetic informativeness profiles

Description

Plot phylogenetic informativeness profiles

Usage

```
informativeness.profile(rate.vector, tree, codon = "FALSE", values = "display")
```

Arguments

rate.vector	site rate vector, converted to class 'matrix'
tree	A phylogenetic tree
codon	True or false: whether to partition a given site rate matrix by codon
values	'display' or 'off': whether to return PI values for each time

Value

Graphical display of phylogenetic informativeness

Author(s)

A. Dornburg

References

Townsend, J. P. "Profiling Phylogenetic Informativeness" Systematic biology 56, no. 2 (2007): 222–231.

Examples

```
library("ape")
library("splines")
read.tree(system.file("extdata","polypterus_trees.phy",package="PhyInformR"))->trees
trees[[1]]->tree
as.matrix(rag1)->rates
informativeness.profile(rates, tree, codon = "FALSE", values = "display")
```

---

multi.profile	*A function to plot phylogenetic informativeness profiles based on site rate cutoffs by rate*

Description

A function to plot phylogenetic informativeness profiles based on site rate breaks

Usage

```
multi.profile(rate.vector, tree, breaks, values = "display")
```

Arguments

`rate.vector`	site rate vector, converted to class 'matrix'
`tree`	A phylogenetic tree
`breaks`	bounds for each profile
`values`	'display' or 'off': whether to return PI values for each time

Author(s)

Alex Dornburg

References

Townsend, J. P. "Profiling Phylogenetic Informativeness" Systematic biology 56, no. 2 (2007): 222–231.

Examples

```
library("ape")
library("splines")
read.tree(system.file("extdata","polypterus_trees.phy",package="PhyInformR"))->trees
trees[[1]]->tree
as.matrix(rag1)->rates
lower<-c(0.,0.003)
upper<-c(0.003000000001,10)
breaks<-cbind(lower,upper)
multi.profile(rates, tree, breaks, values = "off")
```

---

parallel.cluster.signal.noise

Quantify Quartet Resolution Probabilities Using 2012 Formulation and MCMC approach in parallel

---

Description

Quantify Quartet Resolution Probabilities (QIRP, QIHP, and QIPP) Using 2012 Formulation and MCMC approach, outputs results to file

Usage

```
parallel.cluster.signal.noise(t, t0,
rateVector, nsims, s, filename, imagename,
image = "TRUE")
```

Arguments

t	Time from tip of tree to focal internode
t0	Focal internode length
rateVector	
nsims	
s	A number representing the character state space that generated the site rates (e.g., s=2 for binary data)
filename	Output file name for results of calculation
imagename	Output file name for results of visualization
image	True or False if image should visualized

Details

Please note that the two files must have different names

Value

returns QIRP, QIHP, and QIPP values

Author(s)

A. Dornburg

References

Townsend, J. P., Su, Z., and Tekle, Y. I. "Phylogenetic Signal and Noise: Predicting the Power of a Data Set to Resolve Phylogeny" Systematic biology 61, no. 5 (2012): 835–849.

Examples

```
library("doParallel")
#Cores must be less than or equal to those had on your machine.
#registerDoParallel(cores=2)
as.matrix(rag1)->rr
parallel.cluster.signal.noise(100, 2, rr, 10, 3,
filename="output_values", imagename="output_graphic", image = "TRUE")
```

Plot.Another.TreeSI   *A function to visualize QIRP of additional data across a tree topology*

Description

A function to visualize QIRP of additional data across a tree topology

Usage

```
Plot.Another.TreeSI(tree, ratevector, s, color, type)
```

Arguments

tree	A phylogenetic tree
ratevector	site rate vector, converted to class 'matrix'
s	A number representing the character state space that generated the site rates e.g., s=2 for binary data
color	Color of line e.g., color="blue"
type	plotting function for lines e.g., type=3

Value

A graphical representation of QIRP at every node

Author(s)

A. Dornburg

References

Hwang, J., Jonathan, H., Qi, Z., Yang, Z. L., Zheng, W., and Townsend, J. P. "Solving the Ecological Puzzle of Mycorrhizal Associations Using Data from Annotated Collections and Environmental Samples - an Example of Saddle Fungi" Environmental microbiology reports 7, no. 4 (2015): 658–667.

Examples

```
library("ape")
##First plot a locus
as.matrix(rag1)->rr
read.tree(system.file("extdata","polypterus_trees.phy",package="PhyInformR"))->trees
trees[[1]]->tree
PlotTreeSI(tree,rr,3)
##Now plot another locus to compare
as.matrix(prumetalrates[1:1594])->candidate_locus
Plot.Another.TreeSI(tree,candidate_locus,3, color="red",type=3)
```

---

plotPosterior	*A function to visualize outputs of QIRP, QIHP, and QIPP computed across a distribution of trees*

Description

A function to visualize outputs of QIRP, QIHP, and QIPP computed across a distribution of trees

Usage

```
plotPosterior(final, plotType = "QIPs")
```

Arguments

final	output from su.bayes
plotType	type of plot, can be "QIPs" or "violin"

Value

Returns a graphical visulatization of values of either calculation densitys (plot='QIPS') or kernel denstities and quartiles (plot='violin')

Author(s)

A. Dornburg

References

Townsend, J. P., Su, Z., and Tekle, Y. I. "Phylogenetic Signal and Noise: Predicting the Power of a Data Set to Resolve Phylogeny" Systematic biology 61, no. 5 (2012): 835–849. Su, Z., Zhuo, S., Zheng, W., Francesc, L.-G., and Townsend, J. P. "The Impact of Incorporating Molecular Evolutionary Model into Predictions of Phylogenetic Signal and Noise" Frontiers in Ecology and Evolution 2, (2014): doi:10.3389/fevo.2014.00011, Available at http://dx.doi.org/10.3389/fevo.2014.00011 Su, Z. and Townsend, J. P. "Utility of Characters Evolving at Diverse Rates of Evolution to Resolve Quartet Trees with Unequal Branch Lengths: Analytical Predictions of Long-Branch Effects" BMC evolutionary biology 15, (2015): 86.

Examples

```
library("ape")
read.tree(system.file("extdata","polypterus_trees.phy",package="PhyInformR"))->trees
trees<-trees[1:10]
as.matrix(rag1)->rates
quart<-c("Polypterus_congicus", "Polypterus_bichir",
"Polypterus_ansorgii", "Polypterus_endlicheri")
a<-1
b<-1
c<-1
d<-1
e<-1
f<-1
Pi_T<-.25
Pi_C<-.25
Pi_A<-.25
Pi_G<-.25
##Not run
#su.bayes(a,b,c,d,e,f, Pi_T, Pi_C, Pi_A, Pi_G, rates, quart, trees)->final
#plotPosterior(final, plotType='violin')
##
```

PlotTreeSI   *A funtion to visualize QIRP across a tree topology*

Description

A funtion to visualize QIRP across a tree topology

Usage

```
PlotTreeSI(tree, ratevector, s)
```

Arguments

tree	A phylogenetic tree
ratevector	site rate vector, converted to class 'matrix'
s	A number representing the character state space that generated the site rates (e.g., s=2 for binary data)

Value

A graphical representation of QIRP at every node

Author(s)

A. Dornburg

References

Hwang, J., Jonathan, H., Qi, Z., Yang, Z. L., Zheng, W., and Townsend, J. P. "Solving the Ecological Puzzle of Mycorrhizal Associations Using Data from Annotated Collections and Environmental Samples - an Example of Saddle Fungi" Environmental microbiology reports 7, no. 4 (2015): 658–667.

Examples

```
library("ape")
as.matrix(rag1)->rr
read.tree(system.file("extdata","polypterus_trees.phy",package="PhyInformR"))->trees
trees[[1]]->tree
PlotTreeSI(tree,rr,3)
```

---

prumetalrates	*Rates for prumetal dataset*

Description

This data provides an example vector of rates to the user.

Usage

```
prumetalrates
```

Format

a vector of rates

Author(s)

Alex Dornburg

---

rag1	*Rates rag1 dataset*

Description

This data provides an example vector of rates to the user.

Usage

```
rag1
```

Format

a vector of rates

Author(s)

Alex Dornburg

`sample.rates`	*Simulated rates*

Description

This data provides an example vector of simulated rates to the user.

Usage

`sample.rates`

Format

a vector of rates

Author(s)

Alex Dornburg

`sample.tree`	*Sample tree*

Description

This data provides an example tree to the user.

Usage

`sample.tree`

Format

a phy file that can be read in by ape

Author(s)

Alex Dornburg

---

sample.tree2 *Another Sample tree*

Description

This data provides a second example tree to the user.

Usage

```
sample.tree2
```

Format

a phy file that can be read in by ape

Author(s)

Alex Dornburg

---

space.maker *Core Function required to assemble heatmaps of QIRP as in Prum et al. 2015*

Description

Core function required to create heatmaps as in Prum et al. 2015.

Usage

```
space.maker(rateVector, t, s)
```

Arguments

rateVector	An object containing a vector of site rates transformed to class "matrix"
t	Time of quartet (e.g, bigT from Townsend et al. 2012)
s	A number representing the character state space that generated the site rates (e.g., s=2 for binary data)

Value

Computes a range of QIRP values

Author(s)

A. Dornburg

References

Townsend, J. P., Su, Z., and Tekle, Y. I. "Phylogenetic Signal and Noise: Predicting the Power of a Data Set to Resolve Phylogeny" Systematic biology 61, no. 5 (2012): 835–849. Prum, R. O., Berv, J. S., Alex, D., Field, D. J., Townsend, J. P., Lemmon, E. M., and Lemmon, A. R. "A Comprehensive Phylogeny of Birds (Aves) Using Targeted next-Generation DNA Sequencing" Nature 526, no. 7574 (2015): 569–573.

Examples

```
as.matrix(rag1)->rr
space.maker(rr,15,3)
```

---

space.maker.narrow	*Core Function required to assemble more fine-scale heatmaps of QIRP as in Prum et al. 2015*

Description

Core function required to create heatmaps as in Prum et al. 2015.

Usage

```
space.maker.narrow(rateVector, t, s)
```

Arguments

rateVector	An object containing a vector of site rates transformed to class "matrix"
t	Time of quartet (e.g, bigT from Townsend et al. 2012)
s	A number representing the character state space that generated the site rates (e.g., s=2 for binary data)

Value

Computes a range of QIRP values

Author(s)

A. Dornburg

References

Townsend, J. P., Su, Z., and Tekle, Y. I. "Phylogenetic Signal and Noise: Predicting the Power of a Data Set to Resolve Phylogeny" Systematic biology 61, no. 5 (2012): 835–849. Prum, R. O., Berv, J. S., Alex, D., Field, D. J., Townsend, J. P., Lemmon, E. M., and Lemmon, A. R. "A Comprehensive Phylogeny of Birds (Aves) Using Targeted next-Generation DNA Sequencing" Nature 526, no. 7574 (2015): 569–573.

Examples

```
as.matrix(rag1)->rr
space.maker.narrow(rr,15,3)
```

---

su.bayes	*A function to compute results of QIRP, QIHP, and QIPP across a posterior distribution of trees*

Description

This function tracks a four-taxon tree over a user defined distribution of topologies to computer QIRP, QIHP, and QIPP while allowing for a user defined substitution model and base frequencies.

Usage

```
su.bayes(a, b, c, d, e, f, Pi_T, Pi_C, Pi_A, Pi_G, rate_vector, quart, tree)
```

Arguments

a	substitution model parameter. a=rTC
b	substitution model parameter. b=rAT
c	substitution model parameter. c=rGT
d	substitution model parameter. d=rCA
e	substitution model parameter. e=rGC
f	substitution model parameter. e=rGA
Pi_T	Base Frequency of T
Pi_C	Base Frequency of C
Pi_A	Base Frequency of A
Pi_G	Base Frequency of G
rate_vector	site rate vector, converted to class 'matrix'
quart	Four-taxa that comprise the phylogenetic tree quartet of interest quart<-c('taxonA','taxonB','taxonC','taxc
tree	A distribution of phylogenetic trees

Details

Please note that base frequencies must sum to 1.

Value

Function returns a matrix of branch lenghts, internode lengths, QIRP, QIHP. and QIPP values

Author(s)

A. Dornburg

References

Townsend, J. P., Su, Z., and Tekle, Y. I. "Phylogenetic Signal and Noise: Predicting the Power of a
Data Set to Resolve Phylogeny" Systematic biology 61, no. 5 (2012): 835–849. Su, Z., Zhuo, S.,
Zheng, W., Francesc, L.-G., and Townsend, J. P. "The Impact of Incorporating Molecular Evolution-
ary Model into Predictions of Phylogenetic Signal and Noise" Frontiers in Ecology and Evolution
2, (2014): doi:10.3389/fevo.2014.00011, Available at http://dx.doi.org/10.3389/fevo.2014.00011
Su, Z. and Townsend, J. P. "Utility of Characters Evolving at Diverse Rates of Evolution to Resolve
Quartet Trees with Unequal Branch Lengths: Analytical Predictions of Long-Branch Effects" BMC
evolutionary biology 15, (2015): 86.

Examples

```
library("ape")
read.tree(system.file("extdata","polypterus_trees.phy",package="PhyInformR"))->trees
trees<-trees[1:10]
as.matrix(rag1)->rates
quart<-c("Polypterus_congicus", "Polypterus_bichir",
"Polypterus_ansorgii", "Polypterus_endlicheri")
a<-1
b<-1
c<-1
d<-1
e<-1
f<-1
Pi_T<-.25
Pi_C<-.25
Pi_A<-.25
Pi_G<-.25
##Not run
#su.bayes(a,b,c,d,e,f, Pi_T, Pi_C, Pi_A, Pi_G, rates, quart, trees)->final
##
```

Index