Package 'LOGICOIL'

February 19, 2015

Type Package

Version 0.99.0

Date 2014-04-12

Title LOGICOIL: multi-state prediction of coiled-coil oligomeric state.

Author Thomas L. Vincent <tlfvincent@gmail.com>, Peter J. Green and Derek N. Woolfson <D.N.Woolfson@bristol.ac.uk>

Maintainer Thomas L. Vincent <tlfvincent@gmail.com>

Depends R (>= 2.12), nnet

LazyData true

ZipData no

License GPL (>= 2)

Description This package contains the functions necessary to run the LOGICOIL algorithm. LOGICOIL can be used to differentiate between antiparallel dimers, parallel dimers, trimers and higher-order coiled-coil sequence. By covering >90 percent of the known coiled-coil structures, LOGICOIL is a net improvement compared with other existing methods, which achieve a predictive coverage of around 31 percent of this population. As such, LOGICOIL is particularly useful for researchers looking to characterize novel coiled-coil sequences or studying coiled-coil containing protein assemblies. It may also be used to assist in the structural characterization of synthetic coiled-coil sequences.

Repository CRAN

URL <http://coiledcoils.chm.bris.ac.uk/LOGICOIL>

NeedsCompilation no

Date/Publication 2014-04-14 00:04:00

R topics documented:

1

LOGICOIL–package   *Predict oligomerization state of coiled-coil sequences.*

Description

Functions for predicting the oligomeric state of user-defined coiled-coil sequencse. Those functions allow the user to separate between antiparallel dimer, parallel dimers, trimers and higher-order coiled coils.

Details

Package:	LOGICOIL
Type:	Package
Version:	1.0
Date:	2014-04-12
License:	What license is it under?

Functions in the LOGICOIL package allow users to apply the LOGICOIL prediction algorithm to predict the oligomeric state of coiled-coil sequences.

Author(s)

Thomas L. Vincent <tlfvincent@gmail.com>

References

Thomas L. Vincent, Peter J. Green and Dek N. Woolfson. "LOGICOIL-multi-state prediction of coiled-coil oligomeric state", 29(1), pp69-76, Bioinformatics, (2013). DOI: 10.1093/bioinformatics/bts648

Examples

```
library(nnet)
data(pssm)
data(Model_Parameters)
data(LOGICOILfit)

score <- LOGICOIL("GCN4wt", "MKQLEDKVEELLSKNYHLENEVARLKKLV",
    "abcdefgabcdefgabcdefgabcdefga", plot.result=TRUE)

# print output
print(score)
```

| EstimateProbability | *Estimate oligomeric state score at each position of the input coiled-coil sequence* |

Description

Sub-function used in LOGICOIL to compute the oligomeric state score of input coiled-coil sequences.

Usage

```
EstimateProbability(id, seq, reg, pssm, LOGICOILfit, Model_Parameters)
```

Arguments

id	A string that represents the id name of the input sequence
seq	A character string of the amino-acid sequence to be predicted. Valid characters are all uppercase letters except 'B', 'J', 'O', 'U', 'X', and 'Z';
reg	A character string of register assignements. Valid characters are the lowercase letters 'a' to 'g'. Register characters are not required to be in proper order and may start with any of the seven letters. It must always have the same length as the matching amino-acid sequence.
pssm	A profile scoring matrix generated from the LOGICOIL training data.
LOGICOILfit	The fitted model for LOGICOIL obtained using the original training data.
Model_Parameters	A list containing five elements, which are required to run the LOGICOIL model:

1. amino: a character vector containing allowed amino-acids.
2. register: a character vector containing allowed register assignments.
3. amino_interactions: a 87 x 2 matrix containing all amino-amino interactions considered in the LOGICOIL model.
4. register_interactions: a 87 x 2 matrix containing the matching register positions of the amino-amino interactions.
5. lag_matrix: a 7 x 7 matrix that gives the positional lag between register positions.

Value

Returns a matrix of dimension n x 4, where n is the length of the input coiled-coil sequence, and each column represents the LOGICOIL scores for antiparallel dimer, parallel, trimer and higher-order coiled coiled. For examples, the first row gives the predicted LOGICOIL scores of the first position of the input coiled-coil sequence. If a given sequence scores $f(x) == 1$, then there is no evidence for a particular oligomeric state. If $f(x) > 1$ then there is strong evidence for a particular oligimeric state, and if $f(x) < 1$ then there is less evidence against a particular oligimeric state.

Author(s)

Thomas L. Vincent <tlfvincent@gmail.com>

References

Thomas L. Vincent, Peter J. Green and Dek N. Woolfson. "LOGICOIL-multi-state prediction of coiled-coil oligomeric state", 29(1), pp69-76, Bioinformatics, (2013). DOI: 10.1093/bioinformatics/bts648

Examples

```
# load pssm data
library(nnet)
data(pssm)
data(Model_Parameters)
data(LOGICOILfit)

# priors obtained from LOGICOIL training set
prior <- c(0.6331,0.237,0.053,0.0769)

# compute test scores
cat("Estimating oligomeric state of coiled-coil sequences")
prob.oligo <- EstimateProbability("GCN4wt",
    "MKQLEDKVEELLSKNYHLENEVARLKKLV",
"abcdefgabcdefgabcdefgabcdefga",
      pssm,
      LOGICOILfit,
      Model_Parameters)

# compute final summary probability for whole sequence
score <- apply(prob.oligo, 2, mean) / prior

# print output
print(score)
```

LOGICOIL   *Predict oligomerization state of coiled-coil sequences.*

Description

Function for predicting the oligomeric state of a coiled-coil sequence. This function allows the user to separate between antiparallel dimer, parallel dimers, trimers and higher-order coiled coils.

Usage

```
LOGICOIL(id, seq, reg, plot.result = FALSE)
```

Arguments

`id`	A string identifier for the input coiled-coil sequence.
`seq`	The input coiled-coil sequence, which might a string or vector of amino-acids. The input must contain at least 14 characters. Valid characters should be upper-case letters and illegal terms include 'B', 'J', 'O', 'U', 'Z'. Invalid characters will not be tolerated and their use will result in a failure of the program. If the amino-acid in a coiled-coil sequence is unknown, it is permitted to use the identifier "X" instead
`reg`	The register assignments for the amino-acids in the coiled-coil sequence. The input may be either a string or a vector and must always have the same length as the `seq` argument. If this is not the case then the program will fail. Valid characters are the lowercase letters 'a' to 'g'. Register characters are not required to be in proper order
`plot.result`	A logical value indicating whether the results of the predictions should be plotted on a PNG file (default=FALSE).

Value

The function `LOGICOIL` is the parent function in the **LOGICOIL** package. It is used to call the LOGICOIL prediction algorithm in order to predict the oligomeric state of a new coiled-coil sequence. The LOGICOIL function returns a vector with 4 items, which gives the odd ratio that the input sequence belongs to an antiparallel dimer, parallel, trimer or higher-order coiled coiled. If a given sequence scores $f(x) == 1$, then there is no evidence for a particular oligomeric state. If $f(x) > 1$ then there is strong evidence for a particular oligimeric state, and if $f(x) < 1$ then there is against a particular oligimeric state.

Author(s)

Thomas L. Vincent <tlfvincent@gmail.com>

References

Thomas L. Vincent, Peter J. Green and Dek N. Woolfson. "LOGICOIL-multi-state prediction of coiled-coil oligomeric state", 29(1), pp69-76, Bioinformatics, (2013). DOI: 10.1093/bioinformatics/bts648

Examples

```
library(nnet)
data(pssm)
data(Model_Parameters)
data(LOGICOILfit)

score <- LOGICOIL("GCN4wt", "MKQLEDKVEELLSKNYHLENEVARLKKLV",
    "abcdefgabcdefgabcdefgabcdefga", plot.result=FALSE)

# print output
print(score)
```

LOGICOILfit   *Fit of the multinomial log-linear model obtained from the LOGICOIL training dataset.*

Description

A nnet object containing the results of fitting a muiltinomial log-linear model via neural network to the LOGICOIL training dataset. It is used to generate the LOGICOIL predictions.

Usage

```
data(LOGICOILfit)
```

Format

A nnet object

Source

DOI: 10.1093/bioinformatics/bts648

References

Thomas L. Vincent, Peter J. Green and Dek N. Woolfson. "LOGICOIL-multi-state prediction of coiled-coil oligomeric state", 29(1), pp69-76, Bioinformatics, (2013). DOI: 10.1093/bioinformatics/bts648 Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth edition. Springer.

Examples

```
data(LOGICOILfit)
names(LOGICOILfit)
LOGICOILfit$coefnames
```

Model_Parameters   *Model parameters for the LOGICOIL prediction algorithm.*

Description

A list containing five elements, each of which are required to run the LOGICOIL prediction algorithm.

Usage

```
data(Model_Parameters)
```

Format

A list containing five elements, which are required to run the LOGICOIL model:

1. amino: a character vector containing allowed amino-acids.

2. register: a character vector containing allowed register assignments.

3. amino_interactions: a 87 x 2 matrix containing all amino-amino interactions considered in the LOGICOIL model.

4. register_interactions: a 87 x 2 matrix containing the matching register positions of the amino-amino interactions.

5. lag_matrix: a 7 x 7 matrix that gives the positional lag between register positions.

Source

DOI: 10.1093/bioinformatics/bts648

References

Thomas L. Vincent, Peter J. Green and Dek N. Woolfson. "LOGICOIL-multi-state prediction of coiled-coil oligomeric state", 29(1), pp69-76, Bioinformatics, (2013). DOI: 10.1093/bioinformatics/bts648

Examples

```
data(Model_Parameters)
names(Model_Parameters)

# Optional to run
# summary(Model_Parameters$amino)
# summary(Model_Parameters$register)
# head(Model_Parameters$amino_interactions, 5)
# head(Model_Parameters$register_interactions, 5)
# Model_Parameters$lag_matrix
```

---

plot_LOGICOIL *Plot the LOGICOIL scores*

Description

This function can be used to plot the results of applying the LOGICOIL prediction algorithm to user-defined coiled-coil sequences. The plot helps visualize the LOGICOIL scores at each position of the coiled-coil sequence.

Usage

```
plot_LOGICOIL(prob.oligo, id)
```

Arguments

prob.oligo   A matrix of dimension n x 4, where n is the length of the input coiled-coil
              sequence, and each column represents the LOGICOIL scores for antiparallel
              dimer, parallel, trimer and higher-order coiled coiled.

id   A string that represents the id name of the input sequence.

Details

The required input for this function can be obtained by running the `EstimateProbability` function
(see example).

Value

A PNG file that shows the LOGICOIL prediction scores at each of the position of the input coiled-
coil sequence.

Author(s)

Thomas L. Vincent `<tlfvincent@gmail.com>`

References

Thomas L. Vincent, Peter J. Green and Dek N. Woolfson. "LOGICOIL-multi-state prediction of
coiled-coil oligomeric state", 29(1), pp69-76, Bioinformatics, (2013).

Examples

```
library(nnet)
data(pssm)
data(Model_Parameters)
data(LOGICOILfit)

cat("Estimating oligomeric state of coiled-coil sequences\n")
prob.oligo <- EstimateProbability("GCN4wt",
  "MKQLEDKVEELLSKNYHLENEVARLKKLV",
  "abcdefgabcdefgabcdefgabcdefga",
  pssm,
  LOGICOILfit,
  Model_Parameters)

# plot LOGICOIL results obtained from user-defined coiled-coil sequences
plot_LOGICOIL(prob.oligo, "GCN4wt")
```

---

pssm   *Profile scoring matrix derived from the LOGICOIL training set.*

Description

This data set contains the Position Specific Scoring Matrix (PSSM) used as the LOGICOIL training set. This dataset was compiled from the CC+ coiled-coil database. This PSSM is used by LOGICOIL to predict the oligomeric state of coiled-coil sequences. More details on the training set can be found in the reference below.

Usage

```
data(pssm)
```

Format

A data frame with 20295 observations on the following 18 variables.

amino a factor with levels A C D E F G H I K L M N P Q R S T V W X Y

register a factor with levels a b c d e f g

ab a factor with levels 0 AE IR NL

bc a factor with levels 0 ES

cd a factor with levels 0 LT QN

de a factor with levels 0 EL HD ID IK IL KD LK LQ NT SE TD TN TT

ef a factor with levels 0 FG IT QG TT YK

fg a factor with levels 0 KE KN LM QI RH RS SK TN TV

ga a factor with levels 0 EN KV NK RV VA VI YL

ad a factor with levels 0 AF AT IL LI LK LS NL RL VL

be a factor with levels 0 AE KE QN RQ

cf a factor with levels 0 SA

dg a factor with levels 0 EH LE

ea a factor with levels 0 IV YA

ae a factor with levels 0 KI KQ LE LT NR YF

bf a factor with levels 0 AK ED IH VT

cg a factor with levels 0 DR EH EI HE MA QE VT

da a factor with levels 0 II IL IS KI LK LN LR LV NI VI YA YT

Source

References

Thomas L. Vincent, Peter J. Green and Dek N. Woolfson. "LOGICOIL-multi-state prediction of coiled-coil oligomeric state", 29(1), pp69-76, Bioinformatics, (2013).

Examples

```
data(pssm)
head(pssm, 10)
```

Index