Package 'BCRANK'

November 12, 2024

Title Predicting binding site consensus from ranked DNA sequences

Version 1.68.0

Date 2010-08-25

Depends methods

Author Adam Ameur <Adam.Ameur@genpat.uu.se>

Description Functions and classes for de novo prediction of transcription factor binding consensus by heuristic search

Maintainer Adam Ameur <Adam.Ameur@genpat.uu.se>

License GPL-2

LazyLoad true

Collate AllClasses.R AllGenerics.R methods-BCRANKmatch.R methods-BCRANKsearch.R methods-BCRANKresult.R BCRANK.R

Imports Biostrings

Suggests seqLogo

biocViews MotifDiscovery, GeneRegulation

git_url https://git.bioconductor.org/packages/BCRANK

git_branch RELEASE_3_20

git_last_commit 5dc193f

git_last_commit_date 2024-10-29

Repository Bioconductor 3.20

Date/Publication 2024-11-12

Contents

bcrank *BCRANK: predicting binding site consensus from ranked DNA sequences*

Description

This function implements an algorithm for detection of short DNA sequences that are overrepresented in some part of the list. Starting from some initial consensus DNA sequence coded in IUPAC symbols, the method uses a heuristic search to improve the consensus until a local optimum is found. Individual predicted binding sites can be reported by the function `matchingSites`.

Usage

```
bcrank(fafile, startguesses=c(), restarts=10, length=10,
       reorderings=500, silent=FALSE, plot.progress=FALSE,
       do.search=TRUE, use.P1=FALSE, use.P2=TRUE, strip.desc=TRUE)
```

Arguments

`fafile`	a ranked fasta file containing DNA sequences.
`startguesses`	a character vector with consensus sequences in IUPAC coding to be used as starting sequences in the search. If empty, random start guesses will be generated.
`restarts`	number restarts of the algorithm when using random start guesses.
`length`	legth of random start guess.
`reorderings`	number of random reorderings of the DNA sequences performed when calculating score.
`silent`	reports progress status if FALSE.
`plot.progress`	if TRUE, the progress is displayed in a plot.
`do.search`	if FALSE, no search is performed. In that case the start guesses are assigned with scores and reported as results.
`use.P1`	Use penalty for bases other than A,C,G,T.
`use.P2`	Use penalty for motifs matching repetitive sequences.
`strip.desc`	Ignored (always treated as TRUE).

Value

The method returns an objcet of class `BCRANKresult-class`.

Author(s)

Adam Ameur, <adam.ameur@genpat.uu.se>

References

Ameur, A., Rada-Iglesias, A., Komorowski, J., Wadelius, C. Identification of candidate regulatory SNPs by combination of transcription factor binding site prediction, SNP genotyping and haploChIP. Nucleic Acids Res, 2009, 37(12):e85.

See Also

matchingSites, BCRANKresult-class

Examples

```
## Load example fasta file
fastaFile <- system.file("Exfiles/USF1_small.fa", package = "BCRANK")
## Run BCRANK
## Not run: BCRANKout <- bcrank(fastaFile, restarts=20)

## Show BCRANK results
toptable(BCRANKout)
## The top scoring result
topMotif <- toptable(BCRANKout,1)
## Plot BCRANK search path
plot(topMotif)
## Position Weight Matrix
pwm(topMotif, normalize=FALSE)
```

BCRANK-internal *BCRANK-internal*

Description

Internal methods for the **BCRANK** package.

Author(s)

Adam Ameur, <adam.ameur@genpat.uu.se>

See Also

bcrank

BCRANKmatch-class *Class "BCRANKmatch"*

Description

Holds the bcrank score for one IUPAC consensus sequence. Several objects of this class are collected in a BCRANKsearch-class object

Objects from the Class

Objects are not intended to be created directly but as a result from running bcrank.

Slots

consensus: consensus sequence in IUPAC coding

bcrankScore: bcrank score for the consensus

matchVec: vector with 0's (no match) and 1's (match) of same length as the ranked DNA sequences

Methods

consensus `signature(object = "BCRANKmatch")`: Returns the consensus sequence.

bcrankScore `signature(object = "BCRANKmatch")`: Returns the bcrank score.

matchVector `signature(object = "BCRANKmatch")`: Returns a vector with 0's (no match) and 1's (match) of same length and order as the ranked DNA sequences.

Author(s)

Adam Ameur, <adam.ameur@genpat.uu.se>

See Also

bcrank, BCRANKsearch-class

`BCRANKout` *BCRANK results for USF1 ChIP-chip data*

Description

Results from running `bcrank` on USF1 whole genome ChIP-chip data for the human liver cell line HepG2.

Usage

```
data(BCRANKout)
```

Source

Data from whole genome ChIP-chip experiments on human liver cell line HepG2. (Rada-Iglesias, A., et *al*. 2007)

References

Rada-Iglesias, A., et *al*. (2007) Whole-genome maps of USF1 and USF2 binding and histone H3 acetylation reveal new aspects of promoter structure and candidate genes for common human disorders.*Genome Research*, Accepted

See Also

bcrank

BCRANKresult-class *Class "BCRANKresult"*

Description

Holds the results from running bcrank. Contains a number of BCRANKsearch-class object, one for each restart of the bcrank search.

Slots

fname: the name of the fasta file used for running bcrank.

toplist: a list of BCRANKsearch-class objects, ranked by their scores.

funCall: the function call that was made to bcrank.

nrSeqs: number of sequences in the fasta input file.

restarts: number of restarts used in the bcrank search.

Methods

fname `signature(object = "BCRANKmatch")`: Returns the fasta file name.

toplist `signature(object = "BCRANKmatch", i=NULL)`: If `i` is NULL, returns a data frame containing consensus and score for the results for each restart of the bcrank search. Otherwise, the i'th BCRANKsearch-class object in the toplist is returned.

Author(s)

Adam Ameur, <adam.ameur@genpat.uu.se>

See Also

bcrank, BCRANKsearch-class,

BCRANKsearch-class *Class "BCRANKsearch"*

Description

Holds the whole search path from a single bcrank run. Each individual search step is stored in a BCRANKmatch-class object. Several objects of this class are collected in a BCRANKresult-class object

Objects from the Class

Objects are not intended to be created directly but as a result from running bcrank.

Slots

`searchPath`: a collection of `BCRANKmatch-class` objects, ontaining all bcrank search steps from a start guess to a locally optimal solution.

`final`: a `BCRANKmatch-class` object for the highest scoring consensus sequence (locally optimal solution) in this bcrank run.

`finalPWM`: position weight matrix for the highest scoring consensus sequence.

`finalNrMatch`: number of occurrences of the final consensus sequence in the fasta input file.

`nrIterations`: number of iterations required to move from the start guess to the final soloution in this bcrank run.

Methods

searchPath `signature(object = "BCRANKsearch", i=NULL)`: If `i` is NULL, returns a data frame containing consensus and score for the whole search path. Otherwise, the i'th `BCRANKmatch-class` object in the search path is returned.

pwm `signature(object = "BCRANKsearch", normalize=TRUE)`: Returns the position weight matrix (pwm) for the highest scoring consensus in this bcrank run. Matrix positions are between between 0 and 1 when `normalize` is TRUE. When FALSE, the number of matching sequences is reported.

plot `signature(x = "BCRANKsearch", y = "missing")`: A plot method for the searchPath.

Author(s)

Adam Ameur, <adam.ameur@genpat.uu.se>

See Also

`bcrank`, `BCRANKmatch-class`, `BCRANKresult-class`

---

matchingSites *Report IUPAC consensus occurrences in a fasta file*

Description

This function reports all occurrences of a consensus sequence in a fasta file. It can be used to extract transcription factor binding sites predicted by BCRANK or other motif search methods.

Usage

```
matchingSites(fafile, motifSequence, revComp=TRUE, strip.desc=TRUE)
```

Arguments

`fafile`	a ranked fasta file containing DNA sequences.
`motifSequence`	a character vector in IUPAC coding representing a DNA sequence.
`revComp`	set to TRUE if the reverse complement also be matched.
`strip.desc`	Ignored (always treated as TRUE).

Value

Returns a data frame with positions, strand and DNA sequence for the matching sites.

Author(s)

Adam Ameur, <adam.ameur@genpat.uu.se>

References

Ameur, A., Rada-Iglesias, A., Komorowski, J., Wadelius, C. Identification of candidate regulatory SNPs by combination of transcription factor binding site prediction, SNP genotyping and haploChIP. Nucleic Acids Res, 2009, 37(12):e85.

See Also

bcrank

Index