Package 'MEET'

February 19, 2015

Type Package

Title MEET: Motif Elements Estimation Toolkit

Version 5.1.1

Date 2012-12-12

Author Joan Maynou and Erola Pairo.

Maintainer Joan Maynou <joan.maynou@upc.edu>

Description MEET (Motif Elements Estimation Toolkit) is a R-package that integrates a set of computational algorithms for the detection of Transcription Factor Binding Sites (TFBS).

License GPL (>= 2)

LazyLoad yes

LazyDataCompression bzip2

Depends R (>= 2.15.0), seqinr, pcaMethods, Matrix, ROCR,Hmisc,KernSmooth, methods, seqLogo

NeedsCompilation yes

Repository CRAN

Date/Publication 2013-02-22 16:07:16

R topics documented:

align.clustalw *Multiple sequence alignment by means of ClustalW*

Description

DNA sequences are aligned by means of ClustalW (Multiple Sequence Alignment).

Usage

```
align.clustalw(filein, fileout = "Sq.fa", call)
```

Arguments

filein	A set of nucleotide sequences in FASTA format.
fileout	Output file in FASTA format.
call	string of characters needed to call Clustalw from the working directory

Details

This function needs aaMI-package

Value

Fileout is a file in FASTA format with aligned nucleotide sequences.

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>.

See Also

align.MEME, align.MUSCLE

align.MEME *Multiple sequence alignment by means of MEME.*

Description

DNA sequences are aligned by means of MEME Version 4.4.0. (Multiple Expectation-Maximization for Motif Elicitation)

Usage

```
align.MEME(filein, fileout = "Sq.fa", iicc)
```

Arguments

filein	A set of nucleotide sequences in FASTA format.
fileout	Output file in FASTA format
iicc	A list of argument input.

Details

This function needs aaMI-package. This funtions works with meme<=4.3.0

Value

Output is a file in FASTA format with aligned nucleotide sequences.

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

align.clustalw, align.MUSCLE

---

align.muscle *Multiple sequence alignment by means of Muscle (MUltiple Sequence Comparison by Log-Expectation)*

Description

DNA sequences are aligned by means of Muscle Version 3.8. (Multiple Sequences Alignment)

Usage

```
align.muscle(filein , fileout = "Sq.fa", gapopen = gapopen, maxiters = maxiters, gapextend = gapextenc
```

Arguments

filein	A set of nucleotide sequences in FASTA format.
fileout	Output file in FASTA format.
gapopen	Gap open score.
maxiters	Maximum number of iterations.
gapextend	Gap extend score.
call	string of characters needed to call Muscle from the working directory

Details

Gapopen and gapextend must be negative. Output is a file in FASTA format with aligned nucleotide sequences. This function needs aaMI-package.

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

align.clustalw, align.MEME

`Alignment` *To line up Transcription Factor Binding sites through Multiple Sequence Alignment (MSA)*

Description

This function, Alignment, lines up Transcription Factor Binding Sites, TFBS, through MSA. There are different kinds of MSA: MEME, Clustalw and Muscle

Usage

```
Alignment(TF,iicc)
```

Arguments

`TF`	Transcription Factor
`iicc`	A set of initial conditions for the MEET-package

Author(s)

Joan Maynou <joan.maynouatupc.edu> and Erola Pairo <epeiroatibec.pcb.ub.es>

Examples

```
require("MEET")
data(iicc)
pathMEET <- system.file("sequences", package = "MEET")
Alignment(TF=paste(pathMEET, "AP1.fa", sep = "/"),iicc)
```

BackgroundOrganism — *Probabilities of each nucleotide in the* Homo sapiens *organism according to Thakurta et al.*

Description

Probabilities of each nucleotide in the Homo sapiens organism according to Thakurta et al.

Usage

```
data(BackgroundOrganism)
```

Format

Prob Probability of each nucleotide

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>.

Source

G. Guha Thakurta,Computational identification of transcriptional regulatory elements in dna sequence, Nucleic Acids Res., vol. 34, pp. 3585-3598, 2006.

CalculInformation — *Information content in each position of a set of aligned DNA sequences*

Description

Using as an input the set of aligned DNA sequences, information content in each position is calculated, taking into account background probability for each nucleotide. The first row of the returned sequence is the information in each postion, and the others are the loggods matrix.

Usage

```
CalculInformation(matriu, Prob)
```

Arguments

matriu	Aligned DNA sequences
Prob	Background Probability for the studied individual

Value

matriu is a matrix with a first row that is the information content in each position and the next ones the logodds matrix using information content.

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

See Also

kfold.MATCH, CalculScores

Examples

```
data(TranscriptionFactor)
data(BackgroundOrganism)
CalculInformation(matriu=TranscriptionFactor, Prob=Prob)
```

CalculPSSM *Position Specific Scoring Matrices from a set of aligned sequences*

Description

Calculate a PSSM of a set of DNA aligned sequences, taking into account background probabilities. The output is the logodds matrix.

Usage

```
CalculPSSM(matriu, Prob)
```

Arguments

matriu	Set of DNA aligned sequences
Prob	Probability distribution of the nucleotides in the background model

Value

logodds: logodds matrix calculated as a log2 of the frequency matrix

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

References

Gary D. Stormo. DNA binding sites: representation and discovery. Bioinformatics (2000) 16(1): 16-23 doi:10.1093/bioinformatics/16.1.16

See Also

CalculInformation

Examples

```
data(TranscriptionFactor)
data(BackgroundOrganism)
CalculPSSM(matriu=TranscriptionFactor, Prob=Prob)
```

CalculPWM *CalculPWM: To calculate Position Weight Matrix*

Description

This function calculates the PWM (Position Weigth Matrix) from set of aligned nucleotide sequences.

Usage

```
CalculPWM(matriu)
```

Arguments

matriu	A set of aligned nucleotide sequences

Value

A matrix of characters

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

Examples

```
data(TranscriptionFactor)
matriu<-TranscriptionFactor
resul<-CalculPWM(matriu)
```

`CalculRedundancy` *CalculRedundancy: To calculate the redundancy*

Description

This function calculates the redundancy (order) of a random variable X with N possible states.

Usage

```
CalculRedundancy(Factortrans, q,iicc)
```

Arguments

`Factortrans`	A set of aligned nucleotide sequences
`q`	Renyi Order
`iicc`	A list options

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

redundancy, entropy.Shannon,entropy.Renyi, correction.redundancy, correction.entropy, correction-aprox

`CalculScores` *Calcul Score of a Sequence, using a loggods matrix*

Description

Giving a logodds matrix, the Score of a sequence is calculated as the sum of the logodds of each nucleotide at each position of the sequence (equivalent to calculating the probability of each nucleotide at each position)

Usage

```
CalculScores(sequencia, logodds)
```

Arguments

`sequencia`	Sequence to analyze
`logodds`	logodds matrix for a given motif.

Value

score: score of the studied sequence

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

References

Gary D. Stormo. DNA binding sites: representation and discovery. Bioinformatics (2000) 16(1): 16-23 doi:10.1093/bioinformatics/16.1.16

See Also

CalculPSSM, CalculInformation

Examples

```
require("MEET")
#First example, calculating logodds of a Sequence
data(TFlogodds)
data(Sequence)
CalculScores(sequencia=Sequence, logodds=TFlogodds)
#given a Transfac matrix, calculate first loggods and then scores
data(TranscriptionFactor)
data(BackgroundOrganism)
data(Sequence)
Factortrans<-TranscriptionFactor
suma<-apply(Factortrans,2,function(y){sum(y=="-")})
Factortrans<-Factortrans[, suma==0]
logodds<-CalculInformation(matriu=Factortrans, Prob=Prob)
logodds <- logodds[,2:dim(logodds)[2]]
CalculScores(sequencia=Sequence,logodds=logodds)
```

CalculSimilarity *Similarity Score between a Sequence and a PSSM model*

Description

The Similarity between a Sequence and a PSSM model is calculated using the score of the Sequence, and the minimum and maximum scores that can be obtained with the model. As the score approaches to 1, the sequence is more likely to belong to the modeled motif

Usage

```
CalculSimilarity(current, minim, maxim)
```

Arguments

current	Score of the studied sequence
minim	minimum score obtained using the PSSM model
maxim	maximum score obtained using the PSSM model

Value

similarity: between 0 and 1, similarity of the given sequence to the model

Author(s)

Erola Pairo <epeiro@ibec.pcb.ub.es>

References

A.E. Kel , E. Gossling , I. Reuter , E. Cheremushkin , O.V. Kel-Margoulis , and E. Wingender MATCHTM: a tool for searching transcription factor binding sites in DNA sequences Nucl. Acids Res. 31: 3576-3579.

Examples

```
require("MEET")
data(TFlogodds)
data(BackgroundOrganism)
data(Sequence)
current<-CalculScores(Sequence, TFlogodds)

maxim<-0
minim<-0

for(j in 1:dim(TFlogodds)[1]){
  minim <- min(TFlogodds[j,])+minim
  maxim <- max(TFlogodds[j,])+maxim
  }
Similarity<-CalculSimilarity(current, minim, maxim)
```

---

chooseModel *ChooseModel: Choose the best model*

Description

This function uses AUC and its variation to choose the best parameters.

Usage

```
chooseModel(AUC, iicc)
```

Arguments

AUC	List of Areas under ROC curve produced by the studied method. Each AUC element of the list is a vector.
iicc	options of the MEET program

Details

The needed iicc is the vector of parameters to study

Value

y: list with the best parameter and the index of the position of this parameter in the studied range

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET, validation

Examples

```
require("MEET")
data(iicc)
detection(iicc)
```

---

classMODEL *classMODEL: To choose the model*

Description

This function uses transcription factor name, organism and method to choose the model.

Usage

```
classMODEL(org,method,nameTF)
```

Arguments

org	Organism
method	Method
nameTF	Transcription Factor name

Details

The needed iicc is the vector of parameters to study

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET,EntropyDROSOPHILA,EntropyHOMO,EntropyMUS,EntropyRATTUS,DivergenceHOMO,DivergenceMUS,Diverg

Examples

```
require("MEET")
data(iicc)
classMODEL(org=iicc$organism,method=iicc$method,nameTF=iicc$nameTF)
```

ConstructModel *A set of functions for training of motif discovery algorithms.*

Description

This function contains a set of functions for training of motif discovery algorithms. Specifically, the motif discovery algorithms are ITEME, MEME, MDscan, MATCH and Qresiduals.

Usage

```
ConstructModel(iicc, TF)
```

Arguments

iicc	Set of initial conditions for the MEET-package: mode, method, background,alignment,threshold,paramete
TF	A set of nucleotide sequences

Details

This function has two output: validation scores and ROC curve.

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET, detection

correction-class *Correction for finite sample effect*

Description

Correction for the finite sample effect

Format

Herrorl Mean Entropy

sderrorl Standard deviation Entropy

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>.

correction.entropy *Correction entropy from the Finite Sample Size Effect*

Description

Each training matrix is formed by a finite number of samples. The probability estimation error using the nucleotide frequency causes a bias on the uncertainty measurement

Usage

```
correction.entropy(q, p, long, iicc)
```

Arguments

q	Renyi Order
p	Sample size
long	Length of the binding site
iicc	A list of options

Details

Sample size has to be =< 50.

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

correctionaprox,entropy, redundancy, correction.redundancy

Examples

```
data(TranscriptionFactor)
data(iicc)
data(BackgroundOrganism)
data(RenyiOrder)
correction.entropy(q,p=nrow(TranscriptionFactor),long=1,iicc)
```

---

correction.redundancy *Correction redundancy from the Finite Sample Size Effect*

Description

Each training matrix is formed by a finite number of samples. The probability estimation error using the nucleotide frequency causes a bias on the uncertainty measurement

Usage

```
correction.redundancy(r, HXmax, Herror, finite)
```

Arguments

r	Length of the binding site
HXmax	Maximum entropy for N possible states
Herror	Error entropy for finit sample size effect
finite	Sample size

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

redundancy, entropy,CalculRedundancy, correction.entropy, correctionaprox

Examples

```
require("MEET")
data(TranscriptionFactor)
data(BackgroundOrganism)
data(RenyiOrder)
data(iicc)
Factortrans<-TranscriptionFactor
correction<-correction.entropy(q,p=nrow(Factortrans),long=1,iicc)
Herror<-slot(correction,"Herror")
```

```
HXmax<-iicc$HXmax
correction.redundancy(r=1,HXmax,Herror,finite=nrow(TranscriptionFactor))
```

`correctionaprox` *Correction Entropy Approximate from the Finite Sample Size Effect.*

Description

Each training matrix is formed by a finite number of samples. The probability estimation error using the nucleotide frequency causes a bias on the uncertainty measurement The approximation method is used for sample size is more than or equal to 50.

Usage

```
correctionaprox(x, matriu, s)
```

Arguments

`x`	Maximum entropy
`matriu`	A set of aligned nucleotide sequences.
`s`	Number of symbols

Details

Sample size has to be > 50.

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

entropy, redundancy, correction.redundancy, correction.entropy

Examples

```
require("MEET")
data(RenyiOrder)
data(iicc)
s<-4
correctionaprox(x=iicc$HXmax,matriu=iicc$TranscriptionFactor,s)
```

CreateConsensus *Consensus Sequence for a DNA motif*

Description

Using seqinr R package, constructs a consensus DNA sequences from a set of aligned Sequences

Usage

```
CreateConsensus(alignedSequences,iicc,filein)
```

Arguments

alignedSequences	Aligned DNA sequences
iicc	Initial options
filein	Fasta file with a set of aligned sequences with MEME

Details

package "seqinr" is required

Value

SeqCons: Consensus sequence (if gaps a "Na" is returned)

Author(s)

Erola Pairo <epeiroatibec.ub.pcb.es>

References

https://r-forge.r-project.org/projects/seqinr/

See Also

align.clustalw, align.MEME, read.aligned, align.muscle

Examples

```
require("MEET")
data(TranscriptionFactor)
data(iicc)
pathMEET <- system.file("sequences", package = "MEET")
SeqCons<-CreateConsensus(TranscriptionFactor,iicc,filein=NULL)
```

detection *Detection: A set of functions for detection of TFBS*

Description

This function contains a set of functions for detection of Transcription Factor Binding Sites. Specifically, motif discovery algorithms are ITEME, MEME, MDscan, MATCH and Qresiduals.

Usage

```
detection(iicc)
```

Arguments

iicc	Set of initial conditions for the MEET-package: mode, method, background,alignment,pvalor,parameters, nummotif,lenmotif,sentit,position,missing,vector,gapopen, maxiters,gapextend

Details

The main parameter is method (a set of motif discovery algorithms: ITEME(Entropy, divergence), Q-residuals, MEME, Match and MDscan)

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET, ConstructModel

Examples

```
require("MEET")
data(iicc)
detection(iicc)
```

detector_1rOrdre_diff *Detection of Transcription Factor Binding Sites Through Differential Renyi Entropy*

Description

This detection algorithm is based on Information Theory. Specifically, entropy algorithm uses a parametric uncertainty measurement called Renyi entropy. This computational method evaluates the variation on the total Renyi entropy of a set of sequences when a candidate sequence is assumed to be a a true binding site belonging to the set.The measurement of the variation of the total redundancy when the candidate sequence is added to the set has been computed by using the difference between the redundancy profile. This technique assumes independecy between positions in the binding sequence.

Usage

```
detector_1rOrdre_diff(training.set, val.set, iicc)
```

Arguments

`training.set`	A set of aligned nucleotide sequences
`val.set`	A candidate sequence
`iicc`	A set of inicial conditions for the MEET-package: mode, method,background,alignment,threshold,parame nummotif,lenmotif,sentit,position,missing,vector,gapopen,maxiters,gapextend

Details

Options parameter has to contain the next arguments: maximum entropy (HXmax), correction entropy and redundancy from the Finite Sample Size Effect(correction, Redundancia_corregida, Herror and ErrorHX)

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

detector_2Ordre, MEME, MDscan, Q-residuals and MATCH

Examples

```
require("MEET")
data(iicc)
test<-detector_1rOrdre_diff(training.set=iicc$Transcriptionfactor,val.set=NULL,iicc)
```

`detector_2nOrdre` *Detection of Transcription Factor Binding Sites Through Parametric PredictDivergence*

Description

This detection algorithm is based on Information Theory. Specifically, this method uses a parametric divergence. This algorithm evaluates the variation on the total Renyi entropy of a set of sequences assuming correlation between positions in the bindng sequence. When a candidate sequence is assumed to be a a true binding site belonging to the set.The measurement of the variation of the total redundancy when the candidate sequence is added to the set has been computed by using the difference between the redundancy profile.

Usage

```
detector_2nOrdre(training.set, val.set, iicc)
```

Arguments

`training.set`	A set of aligned nucleotide sequences
`val.set`	A candidate sequence
`iicc`	A set of inicial conditions for the MEET-package (mode, method, background, alignment, threshold, parameters, Transcriptionfactor, nummotif, lenmotif, sentit, position, missing, vector, gapopen, maxiters, gapextend) and the initial conditions for the divergence method (PredictDivergence, correction 1rOrdre, Exterior product entropic profile (Mperfil), maximun entropy, entropy and Renyi order)

Details

Options parameter has to contain the initial conditions for the divergence method: divergence matrix (D), maximum entropy (HXmax), correction entropy and redundancy from the Finite Sample Size Effect, exterior product entropic profile (Mperfil), entropy (HX) and the Renyi order.

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

detector_1rOrdre_diff

Examples

```
require("MEET")
data(BackgroundOrganism)
data(iicc)
training.set<-iicc$Transcriptionfactor
val.set<-sample(c('A','T','C','G'),ncol(training.set), replace=TRUE,Prob)
iicc<-detector_2nOrdre_init(training.set, val.set, iicc)
out<-detector_2nOrdre(training.set, val.set, iicc)
```

---

detector_2nOrdre_init *Detection of Transcription Factor Binding Sites Through Parametric PredictDivergence*

Description

This detection algorithm is based on Information Theory. Specifically, this method uses a parametric divergence. This algorithm evaluates the variation on the total Renyi entropy of a set of sequences assuming correlation between positions in the bindng sequence. When a candidate sequence is assumed to be a a true binding site belonging to the set.The measurement of the variation of the total redundancy when the candidate sequence is added to the set has been computed by using the difference between the redundancy profile.

Usage

```
detector_2nOrdre_init(training.set, val.set, iicc)
```

Arguments

training.set	A set of aligned nucleotide sequences
val.set	A candidate sequence
iicc	A set of initial conditions for the MEET-package

Details

This function calculates of initials conditions for divergence method.

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

detector_1rOrdre_diff

Examples

```
require("MEET")
data(iicc)
data(TranscriptionFactor)
data(BackgroundOrganism)
training.set<-TranscriptionFactor
val.set<-sample(c('A','T','C','G'),ncol(TranscriptionFactor), replace=TRUE,Prob)
detector_2nOrdre_init(training.set, val.set, iicc)
```

`diffInstructions` *The measurement of the variation of the total redundancy*

Description

This function measures the variation of the total redundancy.

Usage

```
diffInstructions(training.set, HX, HXmax, Herror, Redundancia_corregida)
```

Arguments

`training.set`	A set of aligned nucleotide sequences
`HX`	Entropy
`HXmax`	Maximum entropy
`Herror`	Entropy error. Correction of the Finite Sample Size Effect
`Redundancia_corregida`	Redundancy correction of the Finite Sample Size Effect

Details

This function depends on detector_1erOrdre_diff

Author(s)

Joan Maynou <joan.maynouatupc.edu>

Examples

```
require("MEET")
data(iicc)
data(BackgroundOrganism)
q<-iicc$q
training.set<-iicc$Transcriptionfactor
correction<-correction.entropy(q,p=nrow(training.set),long=1,iicc)
Herror<-slot(correction,"Herror")
HXmax<-iicc$HXmax
```

```
prob<-probability(training.set,Prob)
HX<-entropy.Shannon(prob)
Redundancia_corregida<-CalculRedundancy(training.set,q,iicc)
test<-diffInstructions (training.set,HX,HXmax,Herror,Redundancia_corregida)
```

`divergence.Renyi` *Renyi divergence*

Description

This function calculates parametric divergence (Renyi Order different 1)

Usage

```
divergence.Renyi(training.set, pmX, pmXY, q, correction)
```

Arguments

`training.set`	A set of aligned nucleotide sequences
`pmX`	Relative frequency of a nucleotide at a motif position (position independency model) as an estimation of the probability of this fact.
`pmXY`	To extend to pmX to include of correlated positions (position dependency model)
`q`	Renyi Order
`correction`	Correction of the Finite Sample Size Effect

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

divergence.Shannon, PredictDivergence

Examples

```
require("MEET")
data(iicc)
data(BackgroundOrganism)
training.set<-iicc$Transcriptionfactor
q<-iicc$q<-0.5
correction<-correction.entropy(q,p=nrow(training.set),long=1,iicc)
HXmax<-iicc$HXmax
pmX<-probability(training.set,Prob)
Probtrans<-probability.couple(Prob)
H<-entropy.Renyi(pmX,q)
pmXY<-joint.probability(training.set, Prob, Probtrans)
HXY<-entropy.joint(pmXY,q,iicc)
divergence.Renyi(training.set,pmX,pmXY,q,correction)
```

divergence.Shannon *Divergencia.Shannon: Mutual Information*

Description

This function calculates Mutual Information (Renyi Order equal 1) by means of Kullback-Leibler divergence

Usage

```
divergence.Shannon(training.set, H, HXY,correction)
```

Arguments

training.set	A set of aligned nucleotide sequences
H	Entropy
HXY	Joint Entropy
correction	Correction of the Finite Sample Size Effect

Details

Renyi Order has to be equal 1.

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

divergence.Renyi, PredictDivergence, kfold.divergence

Examples

```
require("MEET")
data(TranscriptionFactor)
data(BackgroundOrganism)
data(iicc)
q<-1
training.set<-TranscriptionFactor
correction<-correction.entropy(q,p=nrow(training.set),long=1,iicc)
HXmax<-entropy.Shannon(as.matrix(Prob))
pmX<-probability(training.set,Prob)
Probtrans<-probability.couple(Prob)
H<-entropy.Shannon(pmX)
pmXY<-joint.probability(training.set, Prob, Probtrans)
HXY<-entropy.joint(pmXY,q,iicc)
divergence.Shannon(training.set,H,HXY,correction)
```

DivergenceDROSOPHILA *DivergenceDROSOPHILA: Given a Transcription factor chooses the model for a specific organism and method.*

Description

This function uses transcription factor name to choose the model for the Drosophila organism and Divergence algorithm.

Usage

```
DivergenceDROSOPHILA(nameTF)
```

Arguments

nameTF	Transcription Factor name

Details

The needed iicc is the vector of parameters to study

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET,EntropyHOMO,EntropyDROSOPHILA,EntropyMUS,EntropyRATTUS,DivergenceHOMO,DivergenceMUS,Diverg

Examples

```
require("MEET")
data(iicc)
DivergenceDROSOPHILA(nameTF=iicc$nameTF)
```

DivergenceHOMO *DivergenceHOMO: Given a Transcription factor chooses the model for a specific organism and method.*

Description

This function uses transcription factor name to choose the model for the Homo organism and Divergence algorithm.

Usage

```
DivergenceHOMO(nameTF)
```

Arguments

nameTF	Transcription Factor name

Details

The needed iicc is the vector of parameters to study

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET,EntropyHOMO,EntropyDROSOPHILA,EntropyMUS,EntropyRATTUS,DivergenceHOMO,DivergenceMUS,Diverg

Examples

```
require("MEET")
data(iicc)
DivergenceHOMO(nameTF=iicc$nameTF)
```

DivergenceMUS — *DivergenceMUS: Given a Transcription factor chooses the model for a specific organism and method.*

Description

This function uses transcription factor name to choose the model for the Mus organism and Divergence algorithm.

Usage

```
DivergenceMUS(nameTF)
```

Arguments

nameTF	Transcription Factor name

Details

The needed iicc is the vector of parameters to study

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET,EntropyHOMO,EntropyDROSOPHILA,EntropyMUS,EntropyRATTUS,DivergenceHOMO,DivergenceMUS,Diverg

Examples

```
require("MEET")
data(iicc)
DivergenceMUS(nameTF=iicc$nameTF)
```

`DivergenceRATTUS` *DivergenceRATTUS: Given a Transcription factor chooses the model for a specific organism and method.*

Description

This function uses transcription factor name to choose the model for the Rattus organism and Divergence algorithm.

Usage

```
DivergenceRATTUS(nameTF)
```

Arguments

`nameTF`	Transcription Factor name

Details

The needed iicc is the vector of parameters to study

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET,EntropyHOMO,EntropyDROSOPHILA,EntropyMUS,EntropyRATTUS,DivergenceHOMO,DivergenceMUS,Diverg

Examples

```
require("MEET")
data(iicc)
DivergenceRATTUS(nameTF=iicc$nameTF)
```

`entropy.corrected` *Correction of the Finite Sample Size Effect*

Description

Each training matrix is formed by a finite number of samples. The probability estimation error using the nucleotide frequency causes a bias on the uncertainty measurement

Usage

```
entropy.corrected(H, ErrorHX, HXmax)
```

Arguments

`H`	Entropy
`ErrorHX`	Error Entropy
`HXmax`	Maximum Entropy

Details

This function uses the results of correction.entropy and correctionaprox functions.

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

correctionaprox, correction.entropy

Examples

```
require("MEET")
data(BackgroundOrganism)
data(iicc)
Factortrans<-iicc$Transcriptionfactor
correction<-correction.entropy(q=iicc$q,p=nrow(Factortrans),long=1,iicc)
ErrorHX<-slot(correction,"sderror")[nrow(Factortrans)]
HXmax<-iicc$HXmax
prob<-probability(Factortrans,Prob)
H<-entropy.Shannon(prob)
test<-entropy.corrected(H,ErrorHX,HXmax)
```

`entropy.joint` *To calculate joint entropy*

Description

This function calculates the joint entropy between two variables X and Y with N possible state.

Usage

```
entropy.joint(pmXY,q,iicc)
```

Arguments

`pmXY`	Joint probability
`q`	Renyi Order
`iicc`	A list of options

Details

If q is equal 1, Renyi joint entropy converges to Shannon joint entropy

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

entropy

Examples

```
require("MEET")
data(BackgroundOrganism)
data(iicc)
training.set<-iicc$Transcriptionfactor
Probtrans<-probability.couple(Prob)
pmXY<-joint.probability(training.set, Prob, Probtrans)
test<-entropy.joint(pmXY,q=iicc$q,iicc)
```

`entropy.Renyi` *Renyi Entropy*

Description

This function calculates the Renyi's entropy of the variable X with N possible states

Usage

```
entropy.Renyi(wind, q)
```

Arguments

`wind`	Probability matrix (4xm).
`q`	Renyi Order

Details

wind parameter is calculated by means of probability function

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

entropy.max, entropy.conjunta, probability, entropy.corrected

Examples

```
require("MEET")
data(TranscriptionFactor)
data(RenyiOrder)
data(BackgroundOrganism)
training.set<-TranscriptionFactor
wind<-probability(training.set,Prob)
test<-entropy.Renyi(wind,q)
```

`entropy.Shannon` *Shannon Entropy*

Description

This function calculates the Shannon's entropy of the variable X with N possible states

Usage

```
entropy.Shannon(wind)
```

Arguments

`wind`	Probability matrix (4xm).

Details

wind parameter is calculated by means of probability function

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

entropy.max, entropy. conjunta, probability, entropy.corrected

Examples

```
require("MEET")
data(TranscriptionFactor)
data(BackgroundOrganism)
training.set<-TranscriptionFactor
wind<-probability(training.set,Prob)
test<-entropy.Shannon(wind)
```

`EntropyDROSOPHILA` *EntropyDROSOPHILA: Given a Transcription factor chooses the model for a specific organism and method.*

Description

This function uses transcription factor name to choose the model for the Drosophila organism and Entropy algorithm.

Usage

```
EntropyDROSOPHILA(nameTF)
```

Arguments

nameTF	Transcription Factor name

Details

The needed iicc is the vector of parameters to study

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET,EntropyHOMO,EntropyDROSOPHILA,EntropyMUS,EntropyRATTUS,DivergenceHOMO,DivergenceMUS,Diverg

Examples

```
require("MEET")
data(iicc)
EntropyDROSOPHILA(nameTF=iicc$nameTF)
```

EntropyHOMO *EntropyHOMO: Given a Transcription factor chooses the model for a specific organism and method.*

Description

This function uses transcription factor name to choose the model for the Homo organism and Entropy algorithm.

Usage

```
EntropyHOMO(nameTF)
```

Arguments

nameTF	Transcription Factor name

Details

The needed iicc is the vector of parameters to study

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET,EntropyHOMO,EntropyDROSOPHILA,EntropyMUS,EntropyRATTUS,DivergenceHOMO,DivergenceMUS,Diverg

Examples

```
require("MEET")
data(iicc)
EntropyHOMO(nameTF=iicc$nameTF)
```

`EntropyMUS` *EntropyMUS: Given a Transcription factor chooses the model for a specific organism and method.*

Description

This function uses transcription factor name to choose the model for the Mus organism and Entropy algorithm.

Usage

```
EntropyMUS(nameTF)
```

Arguments

`nameTF`	Transcription Factor name

Details

The needed iicc is the vector of parameters to study

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET,EntropyHOMO,EntropyDROSOPHILA,EntropyMUS,EntropyRATTUS,DivergenceHOMO,DivergenceMUS,Diverg

Examples

```
require("MEET")
data(iicc)
EntropyMUS(nameTF=iicc$nameTF)
```

EntropyRATTUS *EntropyRATTUS: Given a Transcription factor chooses the model for a specific organism and method.*

Description

This function uses transcription factor name to choose the model for the Rattus organism and Entropy algorithm.

Usage

```
EntropyRATTUS(nameTF)
```

Arguments

nameTF	Transcription Factor name

Details

The needed iicc is the vector of parameters to study

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET,EntropyHOMO,EntropyDROSOPHILA,EntropyMUS,EntropyRATTUS,DivergenceHOMO,DivergenceMUS,Diverg

Examples

```
require("MEET")
data(iicc)
EntropyRATTUS(nameTF=iicc$nameTF)
```

Hmemory *Library of entropy values*

Description

This function calculates entropy values from all combinations possibles of nucleotides given a matrix of TFBS sequences.

Usage

```
Hmemory(iicc, training.set)
```

Arguments

iicc	A set of initial conditions for the MEET-package
training.set	A set of nucleotide sequences

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

Hread

Examples

```
require("MEET")
data(iicc)
factor<-iicc$Transcriptionfactor
Hmemory(iicc,factor)
```

Hread *To read Entropy values*

Description

This function reads entropy values saved in memory. From the entropy values, Hread calculates the variation of the total entropy when the candidate sequence is added to the set.

Usage

```
Hread(training.set.mes.rand,val.set,iicc)
```

Arguments

training.set.mes.rand	A set of nucleotide sequences
val.set	A candidate sequence
iicc	A set of inicial conditions for the MEET-package

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

Hmemory

Examples

```
require("MEET")
require("seqinr")
write.fasta <- get("write.fasta",pos="package:seqinr")
read.fasta <- get("read.fasta",pos="package:seqinr")
data(iicc)
training.set.mes.rand<-iicc$Transcriptionfactor
val.set<-c("A","T","C","C","G","T","A","T","T","A","C","G")
test<-Hread(training.set.mes.rand, val.set, iicc)
```

---

iicc *A set of initial conditions*

Description

A set of initial conditions for the MEET-package

Arguments

TF	A set of nucleotide sequences
nameTF	Transcription Factor name
seqin	Candidate sequence
alg	A set of Multiple Sequence Alignment
method	A set of motif discovery algorithms
mode	Training or detection
org	Background organism
vector	Characteristic parameters methodfor leave-one-out cross-validation.
num_motif	Number of motif
len_motif	Length of motif
direction	Forward, reverse or both
threshold	P-value
order	Renyi Order
model	model
position	Binding site position for training
mv	missing values
gapopen	The gap open score
maxiters	Maximum number of iterations
gapextend	The gap extend score
optionsFile	Path for MEME/MAST, clustalw and muscle

Details

Models (Entropy, Divergence and Qresiduals) for a set of TFBS. Homo sapiens:AP1,CREB1,E2F1,ELK1,ELK4,ESR1,ETS1,
Rattus norvegicus:AP1,CREB1,Ddit3_Cebpa,FEV,Foxd3,Foxq1,Mafb,NFATC2,NF_kappaB,NR3C1,SP1.
Mus musculus:AP1,CREB1,Ddit3_Cebpa,FEV,Foxd3,Foxq1,Mafb,NFATC2,NF_kappaB,NR3C1,SP1.
Drosophila melanogaster:Abd_B,Antp,Awh,B_H1,B_H2,C15,CG11085,CG11294,CG11617,CG13424,CG15696,CG32105

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>.

`JacksonParameters` *JacksonParameters: To calculates the parameters needed to transform a Q-residual to a confidence interval*

Description

Given PCA model, this function calculates the parameters to perform the Jackson statistics of the Q-residuals.

Usage

```
JacksonParameters(nPCs, TFBS)
```

Arguments

`nPCs`	number of principal components of the model
`TFBS`	numerical TFBS matrix

Value

output:list including the parameters h0 x1, x2 and x3 needed to calculate the Q-residuals statistics

Author(s)

Erola Pairo <epairoatibec.pcb.ub.es>

See Also

kfold.PCA, Predict.PCA, QtoJackson

Examples

```
require("MEET")
data(TranscriptionFactor)
data(iicc)
nPCs<-1
TFBS<-iicc$Transcriptionfactor
Prob<-iicc$background
missing<-50
NumericalMatrix<-numericalDNA(Prob)
suma<-apply(TFBS,2,function(y){sum(y=="-")})
threshold<-floor(nrow(TFBS)*missing/100)
TFBS<-TFBS[, suma<=threshold]
ncolTFBS<-ncol(TFBS)
TFBSnum<-apply(TFBS,1,function(x){as.vector(t(NumericalMatrix[x,]))})
TFBSnum<-t(TFBSnum)
model<-pca(TFBSnum, nPcs=nPCs, method="svd", center=TRUE)
JacksonPars<-JacksonParameters(nPCs,TFBSnum)
```

---

joint.probability *Joint Probability*

Description

This function calculates the joint probability of each base-couple among two positions of the training.set

Usage

```
joint.probability(training.set,Prob,Probtrans)
```

Arguments

training.set	A set of aligned nucleotide sequence
Prob	Background probability of each base within genome
Probtrans	Background probability of correlation among bases within genome

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

probability

Examples

```
require("MEET")
data(TranscriptionFactor)
data(BackgroundOrganism)
training.set<-TranscriptionFactor
Probtrans<-probability.couple(Prob)
pmXY<-joint.probability(training.set, Prob,Probtrans)
```

`kfold.Divergence` *Leave-one-out cross-validation for parametric divergence (ITEME).*

Description

Given a training sequence set, the optimal value for parametric divergence has been estimated by means of leave-one-out cross-validation from q-value set. For each q-value, the ROC curve has been calculated. From this results, the optimal q-value has been considered according to the area under convex surface maximum.

Usage

```
kfold.Divergence(iicc, TF)
```

Arguments

`iicc`	A set of inicial conditions for the MEET-package (mode, method, background, alignment, threshold, parameters, Transcriptionfactor, nummotif, lenmotif, sentit, position, missing, vector, gapopen, maxiters, gapextend)
`TF`	A set of nucleotide sequences

Details

This function integrates the Mutual information (Renyi Order equal 1) and parametric divergence (Renyi Order different 1). Moreover, it contains a set of function for the detection of transcription factor binding sites:correction.entropy.R, correction.redundancy.R, entropy.shannon.R, entropy.renyi.R, entropy.corrected.R, probability.R, CalculRedundancy.R, diff.instructions.R, redundancy.R, ROC-model.R, detector_2nOrdre.R, pvalue.R.

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

kfold.Entropy, kfold.MEME, kfold.MDscan, kfold.MATCH and kfold.PCA

kfold.Entropy *Leave-one-out cross-validation for Renyi entropy (ITEME)*

Description

Given a training sequence set, the optimal value for Renyi entropy has been estimated by means of leave-one-out cross-validation from q-value set. For each q-value, the ROC curve has been calculated. From this results, the optimal q-value has been considered according to the area under convex surface maximum.

Usage

```
kfold.Entropy(iicc, TF)
```

Arguments

iicc	A set of inicial conditions for the MEET-package (mode, method, background, alignment, threshold, parameters, Transcriptionfactor, nummotif, lenmotif, sentit, position, missing, vector, gapopen, maxiters, gapextend)
TF	A set of nucleotide sequences

Details

This function integrates the Shannon entropy for Renyi Order equal 1. Moreover, it contains a set of function for the detection of transcription factor binding sites:correction.entropy.R, correction.redundancy.R, entropy.Shannon.R,entropy.Renyi.R, entropy.corrected.R, probability.R, CalculRedundancy.R, diff.instructions.R, redundancy.R, ROCmodel.R, detector_1rOrdre_diff.R, pvalue.R.

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

kfold.Divergence, kfold.MEME, kfold.MDscan, kfold.MATCH and kfold.PCA

kfold.MATCH *MATCH validation process*

Description

For a vector of Core cut values this function calculates returns a matrix of Similarities and labels indicating if a sequence position is a binding site or not. To calculate the Similartiy a leave-one-out cross validation model is used. With the utput of this function a ROC curve can be calculated for each Core Similarity and the results can be compared.

Usage

```
kfold.MATCH(iicc, Seqin)
```

Arguments

iicc	List of options described in the MEET program
Seqin	DNA sequence

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

References

A.E. Kel , E. Gossling , I. Reuter , E. Cheremushkin , O.V. Kel-Margoulis , and E. Wingender MATCHTM: a tool for searching transcription factor binding sites in DNA sequences Nucl. Acids Res. 31: 3576-3579.

See Also

Match, MEET

Examples

```
require("MEET")
require("seqinr")
data(iicc)
data(TranscriptionFactor)
pathMEET <- system.file("sequences", package = "MEET")
iicc$method<-"MATCH"
iicc$vector<-c(0.5, 0.8)
kfold.MATCH(iicc, Seqin = paste(pathMEET, "AP1.fa", sep = "/"))
```

---

kfold.MDscan *Leave-one-out cross-validation for MDscan.*

Description

Given a training sequence set, the optimal length and number motif has been estimated by means of leave-one-out cross-validation from length and number motif set. From each value, the ROC curve has been calculated. From this results, the optimal value has been considered according to the area under conver surface maximum

Usage

```
kfold.MDscan(iicc, TF)
```

Arguments

iicc	A set of inicial conditions for the MEET-package (mode, method, background, alignment, threshold, parameters, Transcriptionfactor, nummotif, lenmotif, sentit, position, missing, vector, gapopen, maxiters, gapextend)
TF	A set of nucleotide sequences

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

kfold.Divergence, kfold.Entropy, kfold.MEME, kfold.MATCH and kfold.PCA

---

kfold.MEME *Leave-one-out cross-validation for MEME*

Description

Given a training sequence set, the optimal length and number motif has been estimated by means of leave-one-out cross-validation from length and number motif set. From each value, the ROC curve has been calculated. From this results, the optimal value has been considered according to the area under conver surface maximum

Usage

```
kfold.MEME(iicc, TF)
```

Arguments

iicc	A set of inicial conditions for the MEET-package (mode, method, background, alignment, threshold, parameters, Transcriptionfactor, nummotif, lenmotif, sentit, position, missing, vector, gapopen, maxiters, gapextend)
TF	A set of nucleotide sequences

Details

This function needs MEME/MAST software.

Author(s)

Joan Maynou <joan.maynouatupc.edu>

References

[5] T. Bailey and C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers, in Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology.AAAI Press, August 1994, pp. 28-36

See Also

kfold.PredictDivergence, kfold.PredictInformationTheory, MDscan, kfold.MATCH and kfold.PCA

`kfold.PCA` *PCA*

Description

For a vector of principal components the Q-residuals of a DNA symbolical sequence are calculated using a PCA model of the numerical DNA motif. First sequences are converted to numerical DNA sequences and then the model is applied. A matrix is returned for each number of PC with the value of the Q-residuals in each position and a label indicating if the sequence belong to a binding site or not

Usage

```
kfold.PCA(iicc, TF)
```

Arguments

`iicc`	Options described in the MEET function
`TF`	DNA sequences used to construct the motif model

Details

Alignment method has to be installed in your computer.

Value

As a list, for each nPCs

`matriuROC[[nPCs]]`
: matrix with two colums, in the first one the Q-residuals for each studied sequence, and the second one indicates if the sequence belong to a TFBS

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

References

Jolliffe I.T. Principal Component Analysis, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illus. ISBN 978-0-387-95442-4 Stacklies, Wolfram, Redestig, Henning, Scholz, Matthias, Walther, Dirk, and Selbig, Joachim: pcaMethods a bioconductor package providing PCA methods for incomplete data, Bioinformatics 23(9), volume 23, 1164-1167, 2007

See Also

MEET, kfold.Entropy, kfold.MATCH, kfold.Divergence, PCanalysis, llegir_DNA, convertDNA, numericalDNA.

Examples

```
require("MEET")
require("seqinr")
pathMEET <- system.file("sequences", package = "MEET")
data(iicc)
data(TranscriptionFactor)
iicc$method<-"PCA"
#Define the numner of principal components
iicc$vector<-c(1,3,5)
kfold.PCA(iicc, TF = paste(pathMEET, "AP1.fa", sep = "/"))
```

---

kfold.transMEME *Leave-one-out cross-validation for MEME/MAST through training.matrix aligned with MUSCLE or CLUSTALW.*

Description

This function does leave-one-out cross-validation for MEME/MAST. In this case, a set of nucleotide sequences is lined up MUSCLE and CLUSTALW. This is main difference between transMEME and MEME.

Usage

```
kfold.transMEME(iicc, TF)
```

Arguments

iicc	Set of initial conditions for the MEET-package: mode, method, background,alignment,threshold,paramete
TF	A set of nucleotide sequence

Details

This function needs MEME/MAST software.

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET, kfold.Entropy, kfold.transMEME,kfoldMEME, kfold.Divergence, kfold.PCA

MEET *MEET: Motif Elements Estimation Toolkit*

Description

MEET (Motif Elements Estimation Toolkit) is a R-package that integrates a set of computational algorithms for the detection of Transcription Factor Binding Sites (TFBS)

Usage

```
MEET(TF,nameTF, seqin, alg, method, mode, org, vector, num_motif, len_motif, direction, threshold, or
```

Arguments

TF	A set of nucleotide sequences
nameTF	Transcription Factor name
seqin	Candidate sequence
alg	A set of Multiple Sequence Alignment
method	A set of motif discovery algorithms
mode	Training or detection
org	Background organism
vector	Characteristic parameters methodfor leave-one-out cross-validation.
num_motif	Number of motif
len_motif	Length of motif
direction	Forward, reverse or both
threshold	P-value
order	Renyi Order
model	model
position	Binding site position for training
mv	missing values
gapopen	The gap open score
maxiters	Maximum number of iterations
gapextend	The gap extend score
optionsFile	Path for MEME/MAST, clustalx and muscle

Details

This function need the next packages: seqinr, fields, pcaMethods, Matrix, ROCR, Hmisc, KernSmooth. Moreover, it needs the next software: MEME/MAST Version 4.4.0, MATCH Version 1.0, MDscan, ClustalW and Muscle Version 3.8.

MImemory *Library of PredictDivergence values*

Description

This function calculates divergence values from all combinations possibles of a couple of nucleotides given a matrix of TFBS sequences. The divergence values is calculated only by TFBS positions correlated.

Usage

```
MImemory(iicc, training.set)
```

Arguments

iicc	A set of inicial conditions for the MEET-package
training.set	A set of nucleotide sequences

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

MIread

Examples

```
require("MEET")
data(iicc)
iicc <- detector_2nOrdre_init(training.set=iicc$Transcriptionfactor, val.set=val.set, iicc)
MImemory(iicc,training.set=iicc$Transcriptionfactor)
```

MIread *To read PredictDivergence values*

Description

This function reads divergence values saved in memory. From the divergence values, MIread calculates the variation of the total divergence when the candidate sequence is added to the set.

Usage

```
MIread(training.set,val.set,iicc)
```

Arguments

training.set	A set of nucleotide sequences
val.set	A candidate sequence
iicc	A set of inicial conditions for the MEET-package

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

MImemory

Examples

```
require("MEET")
data(DrosophilaDivergence)
model<-list()
model$D<-iicc[["a1"]]$model$parameterModel$D
model$HXmax<-iicc[["a1"]]$model$parameterModel$HXmax
model$correctioc_1rOrdre<-iicc[["a1"]]$model$parameterModel$correction_1rOrdre
model$Entropy<-iicc[["a1"]]$model$parameterModel$HX
model$Mperfil<-iicc[["a1"]]$model$parameterModel$Mperfil
model$interA<-iicc[["a1"]]$model$parameterModel$interA
model$interB<-iicc[["a1"]]$model$parameterModel$interB
model$Divergence<-iicc[["a1"]]$model$model

test<-MIread(training.set=iicc[["a1"]]$Transcriptionfactor, val.set=iicc[["a1"]]$Transcriptionfactor[1,],iicc=
```

Model-class *A set of Models for the detection*

Description

A set of models for the detection

Format

Model A List of the models

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>.

ModelDivergence *To create Model Divergence*

Description

From the best parameter, this function, ModelDivergence, calculates the divergence matrix of the set of aligned nucleotide sequences.

Usage

```
ModelDivergence(iicc)
```

Arguments

iicc	A set of inicial conditions for the MEET-package

Author(s)

Joan Maynou <joan.maynouatupc.edu>

Examples

```
require("MEET")
data(iicc)
iicc$parametersIdeal<-0.5
ModelDivergence(iicc)
```

ModelEntropy *To create Model Entropy*

Description

From the best parameter, this function, ModelEntropy, calculates the redundancy profile of the set of aligned nucleotide sequences.

Usage

```
ModelEntropy(iicc)
```

Arguments

iicc	A set of inicial conditions for the MEET-package

Author(s)

Joan Maynou <joan.maynouatupc.edu>

Examples

```
require("MEET")
data(iicc)
iicc$parametersIdeal<-1
ModelEntropy(iicc)
```

`ModelMATCH` *Match algorithm to detect TFBS in a sequence*

Description

Match algorithm is used to construct a PWM model of the intput TFBS sequences. The returned values are the PWM model and the parameters needed to predict using MATCH and the PWM, as the position of the Core, the minimum and maximum scores for a query sequence and its core, the Corecut parameter and the dimensions of teh TFBS matrix.

Usage

```
ModelMATCH(iicc)
```

Arguments

`iicc`	options of the MEET program

Details

The specific parameters for this detection are: the transcription factor to model and the background probabilities for each nucleotide

Value

output: a list with the model and the parameters

`model`	logodds matrix using information per site
`parametersModel`	posCore=position of the core, minim_core=minimum score for the core, maxim_core= maximum score for the core, minim=minimu score, maxim=maximum score, Corecut= percentage of core score that we want to consider, core=logodds for the core, ncolTFBS=dimension of the TFBS matrix

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

References

E. Kel , E. Gossling , I. Reuter , E. Cheremushkin , O.V. Kel-Margoulis , and E. Wingender MATCHTM: a tool for searching transcription factor binding sites in DNA sequences Nucl. Acids Res. 31: 3576-3579.

See Also

MEET, kfold.MATCH, ModelPCA

Examples

```
require("MEET")
data(iicc)
data(TranscriptionFactor)
iicc$method<-"MATCH"

iicc$pvalor<-0.8
iicc$parameter<-60
ModelMATCH(iicc)
```

`ModelMDscan` *MDscan algortihm to detect TFBS within a sequence*

Description

MDscan algorithm is used to construct a PWM model of the intput TFBS sequences. The returned values is the file with the motif

Usage

```
ModelMDscan(iicc)
```

Arguments

`iicc`	options of the MEET program

Details

The output is the name of the file where the results are writen

Value

output:

`nameMDscan`	name of the MDscan file

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

See Also

MEET, kfold.MDscan

ModelMEME *MEME algortihm to detect TFBS within a sequence*

Description

MEME algorithm is used to construct a PWM model of the intput TFBS sequences. The returned values is the file with the motif

Usage

```
ModelMEME(iicc)
```

Arguments

iicc	options of the MEET program

Details

The output is the name of the file where the results are writen

Value

output:

nameMEME	name of the MEME file

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

See Also

MEET, kfold.MEME

ModelPCA *PCA model for a set of TFBS*

Description

Performs a principal components analysis of the input DNA aligned sequences. It can be used to construct a model with the number of components chosen using the validation method, or the number of components entered by the user. The PCA model for the aligned sequences, and the parameters to calculate the Q-residuals statistics are returned.

Usage

```
ModelPCA(iicc)
```

Arguments

iicc	options described in the MEET program

Details

The specific options for this program are: Order (Number of Principal Components used), and Missing (Percentage threshold of unknown nucleotides in a given position to take into account this position. Default 50)

Value

output:list with the PCA model (same output than in pcaMEthods package), and the parameters of the model: the parameters needed to calculate the Jackon statistics, the numerical TFBS matrix and the dimensions of the TFBS matrix used once the missing values are estimated.

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

See Also

PredictPCA, kfold.PCA, MEET, detection

Examples

```
require("MEET")
data(iicc)
data(TranscriptionFactor)
iicc$parameterIdeal<-2
iicc$pvalor<-0.1
iicc$parameter<-2
TFBSmodel<-ModelPCA(iicc)
```

Models *To create Detection Model*

Description

From the best parameter, this function,ConstructModel, calculates the detection model of the set of aligned nucleotide sequences. There are different kinds of models: Entropy, Divergence, PCA, MEME, MDscan, MATCH and transMEME

Usage

```
Models(iicc)
```

Arguments

iicc	A set of inicial conditions for the MEET-package

Author(s)

Joan Maynou <joan.maynouatupc.edu> and Erola Pairo <epeiroatibec.pcb.ub.es>

Examples

```
require("MEET")
data(iicc)
iicc$parametersIdeal<-1
Models(iicc)
```

`ModeltransMEME` *To create Model transMEME*

Description

From the best parameter, this function, ModeltransMEME, calculates the best MEME model of the set of aligned nucleotide sequences.

Usage

```
ModeltransMEME(iicc)
```

Arguments

`iicc`	A set of inicial conditions for the MEET-package

Author(s)

Joan Maynou <joan.maynouatupc.edu> and Erola Pairo <epeiroatibec.pcb.ub.es>

`motif.mast` *MEME format to training matrix*

Description

Given a set of DNA sequences in fasta format, this function converts this set of sequences in a training matrix

Usage

```
motif.mast(input,Factortrans,k,m)
```

Arguments

input	a set of DNA seqeunces in fasta format
Factortrans	Transcription Factor Binding Sites
k	Length of the candidate sequence
m	Binnding Sites sequences that is extracted from training matrix

Author(s)

Joan Maynou <joan.maynouatupc.edu> and Erola Pairo <epeiroatibec.pcb.ub.es>

---

numericalDNA *Conversion of nucleotides to numerical vectors*

Description

Convert a nucleotide to a numerical vector, using a representation where each nucleotide is placed at the vertex of a regular tetrahedron.

Usage

```
numericalDNA(background)
```

Arguments

background	Corresponding point in the tetrahedron of the input nucleotide

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

References

Silverman, B.D., and Linske, R.: A measure of DNA periodicity, Journal of Theoretical Biology 118, volume 118, 295-300, 1986

See Also

convertDNA

Examples

```
require("MEET")
background<-c(0.25, 0.25, 0.25, 0.25)
numericA<-numericalDNA(background)
```

organism *Probability for each nucleotide according to different organism*

Description

Probability for each nucleotide according to different organism: *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Homo sapiens*, *Escherichia coli*, *Caenorhabditis elegans*, *Plasmodium falciparum*. *Streptomyces coelicolor*, *Gallus gallys*, *Mus musculus* and *Rattus norvegicus*

Usage

```
data(organism)
```

Format

organism Table of probability for each nucleotide according to different organism

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>.

PCanalysis *PC analysis on numerical DNA sequences*

Description

Performs a PC analysis of numerical DNA sequences, using the pcaMethods package and projects the studied DNA sequence to the subspace. The Q-residuals (euclidean distance between the sequences and the modeled subspace) of the DNA studied sequence are returned as output.

Usage

```
PCanalysis(TFBS, nPCs, Sequences)
```

Arguments

TFBS	DNA numerical matrix of known TFBS
nPCs	Number of principal components used to build the model
Sequences	numerical matrix of DNA sequences to study (converted as a numerical matrix)

Value

residus: Euclidean distance from the studied sequences to the modeled motif

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

See Also

PCA, kfold.PCA

`PredictDivergence` *A set of functions for detection of Transcription Factor Binding Sites by means of Divergence*

Description

PredictDivergence contains a set of functions for detection of Transcription Factor Binding sites through parametric divergence. This algorithm evaluates the variation on the total Renyi entropy of a set of sequences assuming correlation between positions in the bindng sequence.

Usage

```
PredictDivergence(iicc)
```

Arguments

`iicc`	A set of inicial conditions for the MEET-package (mode, method, background, alignment, threshold, parameters, Transcriptionfactor, nummotif, lenmotif, sentit, position, missing, vector, gapopen, maxiters, gapextend)

Details

This function integrates the Mutual Information (Renyi order equal 1) and parametric divergence (Renyi order different 1)

Author(s)

Joan Maynou <joan.maynouatupc.edu>

References

J. Maynou, M. Vallverdu, F. Claria, J.J. Gallardo-Chacon, P. Caminal and A. Perera," Transcription Factor Binding Site Detection through Position Cross-Mutual Information variability analysis". 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

See Also

PredictEntropy, PredictMEME, PredictMDscan, PredictPCA and PredictMATCH

Examples

```
require("MEET")
data(HomoDivergence)
iicc[["AP1"]]$threshold<-0.01
PredictDivergence(iicc[["AP1"]])
```

`PredictEntropy` *PredictEntropy: Detection of Transcription Factor Binding Sites by means of Renyi entropy*

Description

Detection of transcription factor binding sites through parametric uncertainty measurement (Renyi entropy). This detection algorithm evaluates the variation on the total Renyi entropy of a set of sequences when a candidate sequence is assumed to be a true binding site belonging to the set.

Usage

```
PredictEntropy(iicc)
```

Arguments

`iicc`	A set of inicial conditions for the MEET-package (mode, method, background, alignment, threshold, parameters, Transcriptionfactor, nummotif, lenmotif, sentit, position, missing, vector, gapopen, maxiters, gapextend)

Details

This function contains a set of function for the detection of transcription factor binding sites:correction.entropy.R, correction.redundancy.R, entropy.R entropy.max.R, entropy.corrected.R, probability.R, CalculRedundancy.R, diff.instructions.R, redundancy.R, ROC.R, detector_1rOrdre_diff.R, pvalue.R

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

PredictDivergence, PredictMEME, PredictMDscan, PredictPCA, PredictMATCH

Examples

```
require("MEET")
data(iicc)
PredictEntropy(iicc)
```

Prediction *To detect Transcription Factor Binding sites by means of a model*

Description

This function, Prediction, detects Transcription Factor Binding Sites, TFBS, from the model. There are different kinds of model: Entropy, Divergence, PCA, MEME, MDscan, MATCH and trans-MEME. This model is obtained from ConstructModel function.

Usage

```
Prediction(iicc)
```

Arguments

iicc	A set of inicial conditions for the MEET-package

Author(s)

Joan Maynou <joan.maynouatupc.edu> and Erola Pairo <epeiroatibec.pcb.ub.es>

Examples

```
require("MEET")
data(iicc)
Prediction(iicc)
```

PredictMATCH *MATCH algorithm to detect TFBS in a sequence*

Description

The model constructed using the MATCH algorithm and the specific parameters for this method are used to detect TFBS within a sequence

Usage

```
PredictMATCH(iicc)
```

Arguments

iicc	options of the MEET program

Details

Uses the MATCH constructed motif, and as a parameter the chosen corecut

Value

output: the detected sequences and its position within the sequence.

`Detected Factors` factors detected

`P-value` Similarity Score

`Position` Position of the detected BS within the large DNA sequence

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

See Also

MEET, kfold.MATCH, ModelMATCH

Examples

```
require("MEET")
data(iicc)
data(TranscriptionFactor)
iicc$method<-"MATCH"
iicc$parametersIdeal<-iicc$parameters
iicc$model<-ModelMATCH(iicc)
DetectedSequences<-PredictMATCH(iicc)
```

PredictMDscan *MDscan algorithm to detect TFBS in a sequence*

Description

MDscan constructed motif is used to find TFBS into the query sequence.

Usage

```
PredictMDscan(iicc)
```

Arguments

`iicc` options of the MEET program

Details

Uses the MEME constructed motif

Value

output: the detected sequences and its position within the sequence.

`Detected Factors`	factors detected
`P-value`	Similarity Score
`Position`	Position of the detected BS within the large DNA sequence

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

See Also

MEET, kfold.MDscan, ModelMDscan

`PredictMEME` *MEME algorithm to detect TFBS in a sequence*

Description

MEME constructed motif is used to run MAST to find TFBS into the query sequence.

Usage

```
PredictMEME(iicc)
```

Arguments

`iicc`	options of the MEET program

Details

Uses the MEME constructed motif

Value

output: the detected sequences and its position within the sequence.

`Detected Factors`	factors detected
`P-value`	Similarity Score
`Position`	Position of the detected BS within the large DNA sequence

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

See Also

MEET, kfold.MEME, ModelMEME

`PredictPCA` *Q-residuals detection of TFBS, using a principal components model*

Description

Performs the detection of a modeled motif in a chormosomics sequence

Usage

```
PredictPCA(iicc)
```

Arguments

`iicc`	options described in the MEET program

Details

The specific options for this program are the model constructed for the motif, the background probability for the studied organism and the DNA query sequence

Value

output:list of the TFBS found in the studied DNA sequence: Sequence (BS found), pvalue (Pvalue of the given BS) and position (Position within the studied DNA sequence).

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

See Also

ModelPCA, biol.PCA, MEET, detection

Examples

```
require("MEET")
data(iicc)
data(TranscriptionFactor)
iicc$method<-"PCA"
iicc$pvalor<-0.1
iicc$parameterIdeal<-2
iicc$model<-ModelPCA(iicc)
DetectedSequences<-PredictPCA(iicc)
```

PredicttransMEME *MAST algorithm to detect TFBS in a sequence*

Description

PWM in transfac format is used to find TFBS into a query sequence.

Usage

```
PredicttransMEME(iicc,TF)
```

Arguments

iicc	options of the MEET program
TF	Transcription Factor

Details

Uses the transMEME constructed motif

Value

output: the detected sequences and its position within the sequence.

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

See Also

MEET, kfold.transMEME, ModeltransMEME

Prob *Probabilities of each nucleotide in the* Homo sapiens *organism according to Thakurta et al.*

Description

Probabilities of each nucleotide in the Homo sapiens organism according to Thakurta et al.

Usage

```
data(BackgroundOrganism)
```

Format

Prob Probability of each nucleotide

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>.

Source

G. Guha Thakurta,Computational identification of transcriptional regulatory elements in dna sequence, Nucleic Acids Res., vol. 34, pp. 3585-3598, 2006.

`probability` *Probability*

Description

This function calculates the probability of random varible with N states considering missing values.

Usage

```
probability(wind, Prob)
```

Arguments

`wind`	A set of nucleotide aligned sequences
`Prob`	Probability of each base within genome

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

joint.probability, probIdem, probtransicio

Examples

```
require("MEET")
data(TranscriptionFactor)
data(BackgroundOrganism)
wind<-TranscriptionFactor
probability(wind,Prob)
```

probability.couple *Background joint probability*

Description

Background joint probability

Usage

```
probability.couple(Prob)
```

Arguments

Prob	Probability of each base within genome

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

probability, joint.probability, probIdem, probtransicio

Examples

```
require("MEET")
data(BackgroundOrganism)
probability.couple(Prob)
```

pvalue *P value*

Description

This function calculates p-value under numerical distribution

Usage

```
pvalue(a, dist)
```

Arguments

a	Numerical value
dist	Numerical distribution

Details

This function needs the next packages: KernSmooth and Hmisc

Author(s)

Helena Brunel <helena.brunelatupc.edu>

Examples

```
require("MEET")
dist<-rnorm(10000,0,1)
pvalue(a=3.298,dist)
```

---

q *Renyi Order*

Description

A value for the Renyi order

Usage

```
data(RenyiOrder)
```

Format

q q value

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>.

QresidualsDROSOPHILA *QresidualsDROSOPHILA: Given a Transcription factor chooses the model for a specific organism and method.*

Description

This function uses transcription factor name to choose the model for the Drosophila organism and Qresiduals algorithm.

Usage

```
QresidualsDROSOPHILA(nameTF)
```

Arguments

nameTF	Transcription Factor name

Details

The needed iicc is the vector of parameters to study

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET,EntropyHOMO,EntropyDROSOPHILA,EntropyMUS,EntropyRATTUS,DivergenceHOMO,DivergenceMUS,Diverg

Examples

```
require("MEET")
data(iicc)
QresidualsDROSOPHILA(nameTF=iicc$nameTF)
```

QresidualsHOMO *QresidualsHOMO: Given a Transcription factor chooses the model for a specific organism and method.*

Description

This function uses transcription factor name to choose the model for the Homo organism and Qresiduals algorithm.

Usage

```
QresidualsHOMO(nameTF)
```

Arguments

nameTF	Transcription Factor name

Details

The needed iicc is the vector of parameters to study

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET,EntropyHOMO,EntropyDROSOPHILA,EntropyMUS,EntropyRATTUS,DivergenceHOMO,DivergenceMUS,Diverg

Examples

```
require("MEET")
data(iicc)
QresidualsHOMO(nameTF=iicc$nameTF)
```

`QresidualsMUS` *QresidualsMUS: Given a Transcription factor chooses the model for a specific organism and method.*

Description

This function uses transcription factor name to choose the model for the Mus organism and Qresiduals algorithm.

Usage

```
QresidualsMUS(nameTF)
```

Arguments

`nameTF`	Transcription Factor name

Details

The needed iicc is the vector of parameters to study

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET,EntropyHOMO,EntropyDROSOPHILA,EntropyMUS,EntropyRATTUS,DivergenceHOMO,DivergenceMUS,Diverg

Examples

```
require("MEET")
data(iicc)
QresidualsMUS(nameTF=iicc$nameTF)
```

QresidualsRATTUS *QresidualsRATTUS: Given a Transcription factor chooses the model for a specific organism and method.*

Description

This function uses transcription factor name to choose the model for the Rattus organism and Qresiduals algorithm.

Usage

```
QresidualsRATTUS(nameTF)
```

Arguments

nameTF	Transcription Factor name

Details

The needed iicc is the vector of parameters to study

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET,EntropyHOMO,EntropyDROSOPHILA,EntropyMUS,EntropyRATTUS,DivergenceHOMO,DivergenceMUS,Diverg

Examples

```
require("MEET")
data(iicc)
QresidualsRATTUS(nameTF=iicc$nameTF)
```

QtoJackson *Q to Jackson: transform a Q-residual into a confidence interval*

Description

Given Q-residual, this function uses the parameters calculated in JAcksonParameters to convert a Q-residual into a confidence interval

Usage

```
QtoJackson(Q,h0, x1, x2, x3)
```

Arguments

Q	Q-residual
h0	h0 parameter
x1	x1 parameter
x2	x2 parameter
x3	x3 parameter

Value

output:confidence interval of the given residual.

Author(s)

Erola Pairo <epairoatibec.pcb.ub.es>

See Also

kfold.PCA, Predict.PCA, JacksonParameters

Examples

```
require("MEET")
data(TranscriptionFactor)
data(iicc)
nPCs<-1
TFBS<-iicc$Transcriptionfactor
Prob<-iicc$background
missing<-iicc$missing
NumericalMatrix<-numericalDNA(Prob)
suma<-apply(TFBS,2,function(y){sum(y=="-")})
threshold<-floor(nrow(TFBS)*missing/100)
TFBS<-TFBS[, suma<=threshold]
ncolTFBS<-ncol(TFBS)
TFBSnum<-apply(TFBS,1,function(x){as.vector(t(NumericalMatrix[x,]))})
TFBSnum<-t(TFBSnum)
model<-pca(TFBSnum, nPcs=2, method="svd", center=TRUE)
JacksonPars<-JacksonParameters(nPCs,TFBSnum)
Qres<-3.45
confidence<-QtoJackson(Qres,h0=1,x1=1,x2=1,x3=1)
```

Read.aligned *Read nucleotide sequences*

Description

This function reads nucleotide sequences in format .fasta.

Usage

```
Read.aligned(file)
```

Arguments

file	A set of nucleotide sequences

Details

The file has to be in format .fasta

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

ReadSequence, ReadTF

Examples

```
require("MEET")
pathMEET <- system.file("sequences", package = "MEET")
Read.aligned(file=paste(pathMEET, "AP1.fa", sep = "/"))
```

---

read.mast *Read output mast*

Description

This function reads the output of MAST

Usage

```
read.mast(input, Factortrans, k, m)
```

Arguments

input	Output MAST
Factortrans	A set of aligned nucleotide sequences
k	Length DNA candidate sequence
m	Number of TFBS sequence used on leave-one-out cross-validation

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

readMEME *Read MEME motifs and consensus sequences*

Description

DNA discovered motifs and consensus sequences were read from the MEME results file

Usage

```
readMEME(resultat, num_motif)
```

Arguments

resultat	MEME results file
num_motif	number of MEME motifs

Details

This function works with MEME<=4.3.0

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

align.MEME

Examples

```
resultat<-"/memeout/meme.txt"
num_motif=2
readMEME(resultat, num_motif)
```

ReadSequence *Convert a DNA sequence in a numerical DNA matrix*

Description

Convert a DNA symbolical sequence to a DNA numerical matrix of length m, using a conversion where each nucleotide is placed at the vertex of a numerical tetrahedron

Usage

```
ReadSequence(Seq, m,background,convertDNA)
```

Arguments

Seq	Vector containing a DNA symbolical sequence
m	length of the matrix
background	Probability background
convertDNA	Numerical representation of each nucleotide

Value

matriuSeq:numerical matrix representing de DNA sequence.

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es>

See Also

convertDNA, numericalDNA, readTF

---

redundancy *To calculate redundancy*

Description

This function calculates the redundancy of random variable X with N possibles states

Usage

```
redundancy(HX, HXmax, Herror, finite, correction = TRUE)
```

Arguments

HX	A entropy vector
HXmax	Maximum entropy
Herror	Error entropy
finite	Finite Sample Size
correction	If TRUE, considers correction of the finite Sample size

Details

This function depends the correction.redundancy

Author(s)

Joan Maynou <joan.maynouatupc.edu>

See Also

correction.redundancy, correction.entropy, entropy.Shannon,entropy.Renyi

Examples

```
require("MEET")
data(BackgroundOrganism)
data(iicc)
q<-iicc$q
Factortrans<-iicc$Transcriptionfactor
HXmax<-iicc$HXmax
correction<-correction.entropy(q,p=nrow(Factortrans),long=1,iicc)
Herror<-slot(correction,"Herror")
pmX<-probability(Factortrans,Prob)
HX<-switch(iicc$classentropy,"Shannon"=entropy.Shannon(pmX),"Renyi"=entropy.Renyi(pmX,q))
test<-redundancy(HX,HXmax,Herror,finite=nrow(Factortrans),correction=TRUE)
```

ROCmodel *To choose the best paramater for a model*

Description

This function selects the best paramater for a model. Output's ROCdetector are: Area under curve Roc, Roc curve performance, the best parameters and the model.

Usage

```
ROCmodel(Scores, labels, iicc)
```

Arguments

Scores	Output Scores validation
labels	Output Labels validation
iicc	Set of initial conditions for the MEET-package: mode, method, background,alignment,threshold,paramete

Details

This functions needs the ROCR package.

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

run.read.MDscan *Run and read MDscan on validation*

Description

This function runs and reads MDscan on validation mode.

Usage

```
run.read.MDscan(input, k, len_motif, num_motif,call.MDscan)
```

Arguments

input	DNA sequence
k	Length DNA candidate seqeunce
len_motif	Length motif
num_motif	Number motif
call.MDscan	Path MDscan

Details

This function needs seqinr-Package and MDscan software.

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEET, kfold.PredictInformationTheory, MEME, kfold.PredictDivergence, kfold.PCA

scoreMDscan *Output MDscan method*

Description

This function writes the output MDscan method. The output contains the next fields:sequence, direction, score MDscan.

Usage

```
scoreMDscan(input, k, matriu, direction)
```

Arguments

input	Score MDscan
k	Length DNA sequence
matriu	A set of aligned nucleotide sequence
direction	Direction of DNA sequence

Details

Output's run.read.MDscan is input's scoreMDscan

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

run.read.MDscan

Sequence *A sequence with binding evidence.*

Description

A sequence with binding evidence

Usage

```
data(Sequence)
```

Format

Sequence Binding sites sequence

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>.

standardout *Standard output detector*

Description

This function writes output detector (ITEME, MEME, transMEME and MDscan) in standard format.

Usage

```
standardout(Rdetector, iicc)
```

Arguments

Rdetector	Output detector for ITEME, MEME, transMEME and MDscan
iicc	Set of initial conditions for the MEET-package: mode, method, background,alignment,threshold,paramete

Details

Rdetector is a list of numercial vectors.

Author(s)

Joan Maynou <joan.maynouatupc.edu>

TFlogodds *Logodds matrix*

Description

Logodds matrix

Usage

```
data(TFlogodds)
```

Format

TFlogodds Log odds for a transcription factor binding sites

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>.

TranscriptionFactor *A set of aligned binding sites sequences*

Description

Binding sites sequences

Usage

```
data(TranscriptionFactor)
```

Format

TranscriptionFactor A set of aligned binding sites sequences

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>.

writeMEME *Write a training matrix in a MEME/MAST format*

Description

This function rewrites a training matrix (a set of aligned nucleotide sequences) in a MEME/MAST format.

Usage

```
writeMEME(matriu, m)
```

Arguments

matriu	A set of aligned nucleotide sequences
m	Number of TFBS sequence used on leave-one-out cross-validation

Author(s)

Erola Pairo <epeiroatibec.pcb.ub.es> and Joan Maynou <joan.maynouatupc.edu>

See Also

MEME,transMEME,DMEME,DtransMEMEME

Examples

```
require("MEET")
data(TranscriptionFactor)
matriu<-TranscriptionFactor
writeMEME(matriu, m=1)
```

`writeResultsHTML` *Writes the results of a MEET detection to HTML.*

Description

This function writes the results of a MEET detection step into an HTML file.

Usage

```
writeResultsHTML(resultsMEET, fileName='index.html')
```

Arguments

`resultsMEET`	The results previously got from a detection performed using MEET.
`fileName`	The name of the file to where write the HTML content. By default it is initialized to index.html

Details

This function creates the HTML file in the working directory.

Author(s)

Marco Paulo Seco <marco.paulo.secoatestudiant.upc.edu>

See Also

MEET, detection

Index