Package 'acm4r'

February 19, 2015

Maintainer Andrea Benedetti <andrea.benedetti@mcgill.ca>

Depends MASS

License GPL

Title Align-and-Count Method comparisons of RFLP data

Type Package

Author Andrea Benedetti, Sahir Rai Bhatnagar, Xiaofei Zhao

Description Fragment lengths or molecular weights from pairs of lanes are compared, and a number of matching bands are calculated using the Align-and-Count Method.

Version 1.0

Date 2013-12-20

NeedsCompilation no

Repository CRAN

Date/Publication 2013-12-28 06:59:21

R topics documented:

acm.file.checkdelete *acm.file.checkdelete*

Description

acm.file.checkdelete abort the program if the file with the given filename already exists

Usage

```
acm.file.checkdelete(filename)
```

Arguments

filename	the name of the file which is tested for its existence

Value

none

Author(s)

XiaoFei Zhao <xiaofei.zhao@mail.mcgill.ca>

acm.file.copy *acm.file.copy (the modified file.copy function in acm)*

Description

acm.file.copy behaves similar to file.copy except that the program terminates and shows the already existing file which cannot be overwritten by "to" if "overwrite = FALSE"

Usage

```
acm.file.copy(from, to, overwrite = FALSE, recursive = FALSE,
  copy.mode = TRUE)
```

Arguments

from	see file.copy
to	see file.copy
overwrite	see file.copy
recursive	see file.copy
copy.mode	see file.copy

Value

none

Author(s)

XiaoFei Zhao <xiaofei.zhao@mail.mcgill.ca>

`acm.file.exists` *acm.file.exists (the modified file.exists function in acm)*

Description

acm.file.exists behaves similar to file.exists except that the program terminates and shows the already existing file if the file already exists

Usage

```
acm.file.exists(filename)
```

Arguments

`filename`	the name of the file which is tested for its existence

Value

none

Author(s)

XiaoFei Zhao <xiaofei.zhao@mail.mcgill.ca>

`acmwrapall` *acmwrapall (acm warp all)*

Description

This function takes only take two required parameters. This function calls the function call_erra, call_acm, call_bablbs, call_gd1, call_gd2, and clusters with their default arguments.

Usage

```
acmwrapall(replic, patient, work_dir = ".", dnum = 1, delete = FALSE)
```

Arguments

replic	is the full name of the file containing the rflp result of replicate strains
patient	is the full name of the file containing the rflp result of strains in patients
work_dir	is where the datasets should be stored
dnum	is the file number
delete	logical value indicating if you want to delete any pre-existing files. Default is FALSE

Value

none

Author(s)

XiaoFei Zhao <xiaofei.zhao@mail.mcgill.ca>

References

Salamon et al. (1998) Accommodating Error Analysis in Comparison and Clustering of Molecular Fingerprints. Emerging Infectious Diseases Vol. 4, No. 2, April-June 1998

---

call_acm *Align-and-Count Method comparisons of RFLP data*

Description

Fragment lengths or molecular weights from pairs of lanes are compared, and a number of matching bands are calculated using the Align-and-Count Method. Band intensities are not used in this method.This version of acm will perform all distinct pairwise comparisons between lanes. The first lane will be compared to all following lanes. The second lane will be compared to the third, fourth, etc..

Usage

```
call_acm(input, work_dir = ".", dnum = 1)
```

Arguments

input	The input file format for fragment and lane information are as the file example.in shows. Each line (record) should define a band. The file must be sorted so that the bands from a lane are together; the file must be sorted by unique lane identifier (ULI). There must be three fields (separated by white space – that is, seperated by any combination of spaces and tabs). The first field must be the ULI. The second field is the gel image identifier (GII). This is only critical for the first fragment read in for the lane; for subsequent fragments, the second field must simply be present. ULI and GII names must not contain any

spaces; I strongly suggest that these fields contain only letters, numbers, and underscore characters. The third field should contain the fragment length or molecular weight data for the fragment. Lines that cannot be parsed into these three fields will be recored in acm.log. Check acm.log after each run to see the number of lanes read in and to check for lines from the input file that could not be parsed. Lines must be no longer than 199 characters.

`work_dir` This is where the work is being done

`dnum` is the file number

Value

The output from acm or acmone gives a comparison of two lanes on each line of output in the form: ULI ULI numbands numbands num_matches The first field is identifier for the first lane in the comparison. The third field is the number of bands read in for this first lane. The second field is the identifier for the second lane in the comparison. The fourth field is the number of band read in for this second lane. The fifth field is the number of fragments found to match between the two lanes. The number of matches found may be different depending on whether the lanes came from the same gel image or from different gel images (see the file acm.par).

Note

Requires a parameter file generated by `call_erra`

Author(s)

Andrea Benedetti <andrea.benedetti@mcgill.ca>

Sahir Rai Bhatnagar

XiaoFei Zhao

References

Salamon et. al (1998) Accommodating Error Analysis in Comparison and Clustering of Molecular Fingerprints. Emerging Infectious Diseases Vol. 4, No. 2, April-June 1998

Abasci LLC. JAMES v1.0 User Documentation. 2002.

Examples

```
#perform the basic pairwise comparison via the align and count algorithm
res1<-call_acm("experiments.in")
```

call_bablbs *BABLBS (Band Added Band Lost Band Substituted)*

Description

BABLBS clustering means that two fingerprints match if they differ only through a lost band, a gained band, or a substituted band

Usage

```
call_bablbs(intable, minmatch = 2)
```

Arguments

intable	is the name of the acm table.
minmatch	is the threshold of band number such that two lanes are matching if and only if the number of matching bands between them is more than minmatch.

Value

a matrix of 2 columns where each row represents a pair of matching lanes.

Note

call_bablbs doesn't work for minmatch<5

Pass the result of this function to and pass this result to clusters to synthesize the results

Author(s)

Andrea Benedetti <andrea.benedetti@mcgill.ca>

Sahir Rai Bhatnagar

XiaoFei Zhao

References

Salamon et. al (1998) Accommodating Error Analysis in Comparison and Clustering of Molecular Fingerprints. Emerging Infectious Diseases Vol. 4, No. 2, April-June 1998

Abasci LLC. JAMES v1.0 User Documentation. 2002.

Examples

```
#bablbs, GD1, and GD2 all work with the results from call_acm
#now check matching by bablbs
res_bab<-call_bablbs(res1)
```

call_erra *Error Analysis for 1D Electrophoresis Gels*

Description

This function analyzes fingerprint replicate data. Replicates are repeated measurements of the same sample in one or more gels, which allows the reproducibility of the measurement to be ascertained.

Usage

```
call_erra(rep.file, work_dir = ".", dnum = 1, sd = 4, delete = FALSE)
```

Arguments

rep.file	is the full name and path of the replicate file
work_dir	is where the datasets should be stored
dnum	is the file number
sd	is the standard deviation needed to calculate all things, Salamon et al. (1998) suggested a value of 4 for sd
delete	logical value indicating if you want to delete any pre-existing acm.par files. Default is FALSE

Value

the parameter filed called acm.par, needed by call_acm. This file contains information on about the measurement error present in the data.

Note

call_erra doesn't work for unequal number of trials. This function needs to be run before call_acm.

Author(s)

Andrea Benedetti <andrea.benedetti@mcgill.ca>

Sahir Rai Bhatnagar

XiaoFei Zhao

References

Salamon et al. (1998) Accommodating Error Analysis in Comparison and Clustering of Molecular Fingerprints. Emerging Infectious Diseases Vol. 4, No. 2, April-June 1998

Abasci LLC. JAMES v1.0 User Documentation. 2002.

Examples

```
#generate the error parameter file needed by call_acm
call_erra("replicates.in",  dnum=1, sd=1, delete=TRUE)
```

`call_gd1` *Genetic Distance 1*

Description

Genetic Distance 1 which includes a lambda parameter

Usage

```
call_gd1(intable, gcrit = 5, lambda = 0.05)
```

Arguments

`intable`	is the name of the acm table.
`gcrit`	is the threshold of genetic distance such that two lanes are matching if and only if the genetic distance between them is less than gcrit. Is used to get a list of ids for subjects who have gd distance less than gcrit
`lambda`	is the minimum number of bands that need to match

Value

A matrix with two colums of fingerprint IDs. Each line represents a match according to `gcrit`.

Note

Pass the result of this function to and pass this result to `clusters` to synthesize the results

Author(s)

Andrea Benedetti <andrea.benedetti@mcgill.ca>

Sahir Rai Bhatnagar

XiaoFei Zhao

References

Salamon et. al (1998) Accommodating Error Analysis in Comparison and Clustering of Molecular Fingerprints. Emerging Infectious Diseases Vol. 4, No. 2, April-June 1998

Abasci LLC. JAMES v1.0 User Documentation. 2002.

Examples

```
#matching by GD1
res_gd1<-call_gd1(res1)
```

`call_gd2` *Genetic Distance 2*

Description

Genetic Distance 2 which includes a lambda parameter

Usage

```
call_gd2(intable, gcrit = 2)
```

Arguments

`intable`	is the name of the acm table
`gcrit`	is the threshold of genetic distance such that two lanes are matching if and only if the genetic distance between them is less than gcrit. Is used to get a list of ids for subjects who have gd distance less than gcrit

Value

A matrix with two colums of fingerprint IDs. Each line represents a match according to `gcrit`.

Note

`call_gd2` doesn't work for lambda=0

Pass the result of this function to and pass this result to `clusters` to synthesize the results

Author(s)

Andrea Benedetti `<andrea.benedetti@mcgill.ca>`

Sahir Rai Bhatnagar

XiaoFei Zhao

References

Salamon et. al (1998) Accommodating Error Analysis in Comparison and Clustering of Molecular Fingerprints. Emerging Infectious Diseases Vol. 4, No. 2, April-June 1998

Abasci LLC. JAMES v1.0 User Documentation. 2002.

See Also

`call_acm` for example of use

Examples

```
#matching by GD2
res_gd2<-call_gd2(res1)
```

`clusters` *Clustering*

Description

Clustering for exact matching and BABLBS matching

Usage

```
clusters(input, bablbs, type = "exact", work_dir = "")
```

Arguments

`input`	is the result of the call_acm function in the format of ULI ULI numbands numbands num_matches
`bablbs`	is the result of the call_bablbs, call_gd1 or call_gd2 function in the format of ULI ULI
`work_dir`	is where the datasets should be stored
`type`	value ("exact", "bablbs", "gd1", "gd2") indicating if you want exact matching or BABLBS/Genetic Distance. Default is "exact" i.e. exact matching.

Value

A list with 7 components. The first (SINGLE) is a 3 column matrix of all fingerprint IDs that do not belong to a cluster. The columns correspond to the Cluster_Number, the Cluster_size, and the Fingerprint ID. The second (CLUSTERED) is a matrix of all fingerprint IDs that do belong to a cluster. The columns correspond to the Cluster_Number, the Cluster_size, and the Fingerprint IDs that belong to that cluster. The third (BOTH) combines the others into one matrix. The fourth and fifth calculate RTIN and RTIn-1. The last two are used for the histograms that are produced by a call to this function.

Author(s)

Andrea Benedetti <andrea.benedetti@mcgill.ca>

Sahir Rai Bhatnagar

XiaoFei Zhao

References

Salamon et. al (1998) Accommodating Error Analysis in Comparison and Clustering of Molecular Fingerprints. Emerging Infectious Diseases Vol. 4, No. 2, April-June 1998

Abasci LLC. JAMES v1.0 User Documentation. 2002.

Examples

```
#synthesize the results
# Exact matching clusters
exact<-clusters(input=res1,type="exact")
names(exact)
exact$RTIN
exact$RTIN1
# Clustering based on BABLBS
bablbs<-clusters(input=res1, bablbs=res_bab,type="bablbs")
names(bablbs)
bablbs$RTIN
bablbs$RTIN1
# Clustering based on GD1
gd1<-clusters(input=res1, bablbs=res_gd1,type="gd1")
names(gd1)
gd1$RTIN
gd1$RTIN1
```

---

experiments.in *Sample fingerprint file*

Description

Sample fingerprint file

Author(s)

Andrea Benedetti <andrea.benedetti@mcgill.ca>

References

Salamon et. al (1998) Accommodating Error Analysis in Comparison and Clustering of Molecular Fingerprints. Emerging Infectious Diseases Vol. 4, No. 2, April-June 1998

Abasci LLC. JAMES v1.0 User Documentation. 2002.

---

replicates.in *Sample replicate file*

Description

Sample replicate file

Author(s)

Andrea Benedetti <andrea.benedetti@mcgill.ca>

References

Salamon et. al (1998) Accommodating Error Analysis in Comparison and Clustering of Molecular Fingerprints. Emerging Infectious Diseases Vol. 4, No. 2, April-June 1998

Abasci LLC. JAMES v1.0 User Documentation. 2002.

Index