Package 'BREADR'

September 9, 2024

Title Estimates Degrees of Relatedness (Up to the Second Degree) for Extreme Low-Coverage Data

Version 1.0.2

Description The goal of the package is to provide an easy-to-use method for estimating degrees of relatedness (up to the second degree) for extreme low-coverage data. The package also allows users to quantify and visualise the level of confidence in the estimated degrees of relatedness.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

URL https://github.com/jonotuke/BREADR, https://jonotuke.github.io/BREADR/

BugReports https://github.com/jonotuke/BREADR/issues

Suggests spelling, knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

Imports data.table, dplyr, forcats, ggplot2, ggpubr, grDevices, magrittr, MASS, matrixStats, purrr, readr, stringr, tibble

Depends R (>= 2.10)

LazyData true

Language en-US

NeedsCompilation no

Author Jono Tuke [aut, cre] (<https://orcid.org/0000-0002-1688-8951>), Adam B. Rohrlach [aut] (<https://orcid.org/0000-0002-4204-5018>), Wolfgang Haak [aut] (<https://orcid.org/0000-0003-2475-2007>), Divyaratan Popli [aut] (<https://orcid.org/0000-0002-0305-6427>)

Maintainer Jono Tuke <simon.tuke@adelaide.edu.au>

Repository CRAN

Date/Publication 2024-09-09 11:00:02 UTC

Contents

`callRelatedness` *callRelatedness*

Description

A function that takes PMR observations, and (given a prior distribution for degrees of relatedness) returns the posterior probabilities of all pairs of individuals being (a) the same individual/twins, (b) first-degree related, (c) second-degree related or (d) "unrelated" (third-degree or higher). The highest posterior probability degree of relatedness is also returned as a hard classification. Options include setting the background relatedness (or using the sample median), a minimum number of overlapping SNPs if one uses the sample median for background relatedness, and a minimum number of overlapping SNPs for including pairs in the analysis.

Usage

```
callRelatedness(
  pmr_tibble,
  class_prior = rep(0.25, 4),
  average_relatedness = NULL,
  median_co = 500,
  filter_n = 1
)
```

Arguments

`pmr_tibble`	a tibble that is the output of the processEigenstrat function.
`class_prior`	the prior probabilities for same/twin, 1st-degree, 2nd-degree, unrelated, respectively.

`average_relatedness`
: a single numeric value, or a vector of numeric values, to use as the average background relatedness. If NULL, the sample median is used.

`median_co`
: if average_relatedness is left NULL, then the minimum cutoff for the number of overlapping snps to be included in the median calculation is 500.

`filter_n`
: the minimum number of overlapping SNPs for which pairs are removed from the entire analysis. If NULL, default is 1.

Value

results_tibble: A tibble containing 13 columns:

- row: The row number
- pair: the pair of individuals that are compared.
- relationship: the highest posterior probability estimate of the degree of relatedness.
- pmr: the pairwise mismatch rate (mismatch/nsnps).
- sd: the estimated standard deviation of the pmr.
- mismatch: the number of sites which did not match for each pair.
- nsnps: the number of overlapping snps that were compared for each pair.
- ave_re;: the value for the background relatedness used for normalisation.
- Same_Twins: the posterior probability associated with a same individual/twins classification.
- First_Degree: the posterior probability associated with a first-degree classification.
- Second_Degree: the posterior probability associated with a second-degree classification.
- Unrelated: the posterior probability associated with an unrelated classification.
- BF: A strength of confidence in the Bayes Factor associated with the highest posterior probability classification compared to the 2nd highest. (No longer included)

Examples

```
callRelatedness(counts_example,
  class_prior=rep(0.25,4),
  average_relatedness=NULL,
  median_co=5e2,filter_n=1
)
```

`counts_example` *counts_example*

Description

this is an example of the tibble made by processEigenstrat().

Usage

```
counts_example
```

Format

`counts_example`**:**

A data frame with 15 rows and 4 columns:

pair the pair of individuals that are compared

nsnps the number of overlapping snps that were compared for each pair.

mismatch the number of sites which did not match for each pair.

pmr the pairwise mismatch rate (mismatch/nsnps).

`get_column_new` *get column*

Description

get column

Usage

```
get_column_new(genofile, col = 1)
```

Arguments

`genofile`	genofile
`col`	column to return

Value

column of numbers

`plotLOAF` *plotLOAF*

Description

Plots all (sorted by increasing value) observed PMR values with maximum posterior probability classifications represented by colour and shape. Options include a cut off for the minimum number of overlapping SNPs, the max number of pairs to plot and x-axis font size.

Usage

```
plotLOAF(in_tibble, nsnps_cutoff = NULL, N = NULL, fntsize = 7, verbose = TRUE)
```

Arguments

in_tibble	a tibble that is the output of the callRelatedness() function.
nsnps_cutoff	the minimum number of overlapping SNPs for which pairs are removed from the plot. If NULL, default is 500.
N	the number of (sorted by increasing PMR) pairs to plot. Avoids plotting all pairs (many of which are unrelated).
fntsize	the fontsize for the x-axis names.
verbose	if TRUE, then information about the plotting process is sent to the console

Value

a ggplot object

Examples

```
relatedness_example
plotLOAF(relatedness_example)
```

---

plotSLICE *plotSLICE*

Description

A function for plotting the diagnostic information when classifying a specific pair (defined by the row number or pair name) of individuals. Output includes the PDFs for each degree of relatedness (given the number of overlapping SNPs) in panel A, and the normalised posterior probabilities for each possible degree of relatedness.

Usage

```
plotSLICE(
  in_tibble,
  row,
  title = NULL,
  class_prior = rep(1/4, 4),
  showPlot = TRUE,
  which_plot = 0,
  labels = NULL
)
```

Arguments

in_tibble	a tibble that is the output of the callRelatedness() function.
row	either the row number or pair name for which the posterior distribution is to be plotted.

title	an optional title for the plot. If NULL, the pair from the user-defined row is used.
class_prior	the prior probabilities for same/twin, 1st-degree, 2nd-degree, unrelated, respectively.
showPlot	If TRUE, display plot. If FALSE, just pass plot as a variable.
which_plot	if 1, returns just the plot of the posterior distributions, if 2 returns just the normalised posterior values. Anything else returns both plots.
labels	a length two character vector of labels for plots. Default is no labels.

Value

a two-panel diagnostic ggplot object

Examples

```
plotSLICE(relatedness_example, row = 1)
```

---

processEigenstrat *process Eigenstrat data - alternative version*

Description

A function that takes paths to an eigenstrat trio (ind, snp and geno file) and returns the pairwise mismatch rate for all pairs on a thinned set of SNPs. Options include choosing thinning parameter, subsetting by population names, and filtering out SNPs for which deamination is possible.

Usage

```
processEigenstrat(
  indfile,
  genofile,
  snpfile,
  filter_length = NULL,
  pop_pattern = NULL,
  filter_deam = FALSE,
  outfile = NULL,
  chromosomes = NULL,
  verbose = TRUE
)
```

Arguments

indfile	path to eigenstrat ind file
genofile	path to eigenstrat geno file.
snpfile	path to eigenstrat snp file.

filter_length	the minimum distance between sites to be compared (to reduce the effect of LD).
pop_pattern	a character vector of population names to filter the ind file if only some populations are to compared.
filter_deam	a TRUE/FALSE for if C->T and G->A sites should be ignored.
outfile	(OPTIONAL) a path and filename to which we can save the output of the function as a TSV, if NULL, no back up saved. If no outfile, then a tibble is returned.
chromosomes	the chromosome to filter the data on.
verbose	controls printing of messages to console

Value

out_tibble: A tibble containing four columns:

Examples

```
# Use internal files to the package as an example
indfile <- system.file("extdata", "example.ind.txt", package = "BREADR")
genofile <- system.file("extdata", "example.geno.txt", package = "BREADR")
snpfile <- system.file("extdata", "example.snp.txt", package = "BREADR")
processEigenstrat(
indfile, genofile, snpfile,
filter_length=1e5,
pop_pattern=NULL,
filter_deam=FALSE
)
```

---

processEigenstrat_old *process Eigenstrat data*

Description

A function that takes paths to an eigenstrat trio (ind, snp and geno file) and returns the pairwise mismatch rate for all pairs on a thinned set of SNPs. Options include choosing thinning parameter, subsetting by population names, and filtering out SNPs for which deamination is possible.

Usage

```
processEigenstrat_old(
  indfile,
  genofile,
  snpfile,
  filter_length = NULL,
  pop_pattern = NULL,
  filter_deam = FALSE,
  outfile = NULL,
  chromosomes = NULL,
  verbose = TRUE
)
```

Arguments

indfile	path to eigenstrat ind file
genofile	path to eigenstrat geno file.
snpfile	path to eigenstrat snp file.
filter_length	the minimum distance between sites to be compared (to reduce the effect of LD).
pop_pattern	a character vector of population names to filter the ind file if only some populations are to compared.
filter_deam	a TRUE/FALSE for if C->T and G->A sites should be ignored.
outfile	(OPTIONAL) a path and filename to which we can save the output of the function as a TSV, if NULL, no back up saved. If no outfile, then a tibble is returned.
chromosomes	the chromosome to filter the data on.
verbose	controls printing of messages to console

Value

out_tibble: A tibble containing four columns:

Examples

```
# Use internal files to the package as an example
indfile <- system.file("extdata", "example.ind.txt", package = "BREADR")
genofile <- system.file("extdata", "example.geno.txt", package = "BREADR")
snpfile <- system.file("extdata", "example.snp.txt", package = "BREADR")
processEigenstrat_old(
indfile, genofile, snpfile,
filter_length=1e5,
pop_pattern=NULL,
filter_deam=FALSE
)
```

---

read_ind *read_ind*

Description

read_ind

Usage

```
read_ind(filename)
```

Arguments

filename	a IND text file.

Value

tibble with column headings: ind (CHR), sex (CHR), pop (CHR)

Examples

```
ind_snpfile <- system.file("extdata", "example.ind.txt", package = "BREADR")
read_ind(ind_snpfile)
```

`read_snp` *read_snp*

Description

read_snp

Usage

```
read_snp(filename)
```

Arguments

`filename`	a SNP text file.

Value

tibble with column headings: snp (CHR), chr (DBL), pos (DBL), site (DBL), anc (CHR), and der (CHR).

Examples

```
std_snpfile <- system.file("extdata", "example.snp.txt", package = "BREADR")
broken_snpfile <- system.file("extdata", "broken.snp.txt", package = "BREADR")
read_snp(std_snpfile)
read_snp(broken_snpfile)
```

relatedness_example *relatedness_example*

Description

this is an example of the tibble made by callRelatedness()

Usage

```
relatedness_example
```

Format

`relatedness_example`**:**

A data frame with 15 rows and 13 columns:

row The row number

pair the pair of individuals that are compared.

relationship the highest posterior probability estimate of the degree of relatedness.

pmr the pairwise mismatch rate (mismatch/nsnps).

sd the estimated standard deviation of the pmr.

mismatch the number of sites which did not match for each pair.

nsnps the number of overlapping snps that were compared for each pair.

ave_re the value for the background relatedness used for normalisation.

Same_Twins the posterior probability associated with a same individual/twins classification.

First_Degree the posterior probability associated with a first-degree classification.

Second_Degree the posterior probability associated with a second-degree classification.

Unrelated the posterior probability associated with an unrelated classification.

BF A strength of confidence in the Bayes Factor associated with the highest posterior probability classification compared to the 2nd highest.

saveSLICES *saveSLICES*

Description

Plots all pairwise diagnostic plots (in a tibble as output by callRelatedness), as produced by plotSLICE, to a folder. Options include the width and height of the output files, and the units in which these dimensions are measured.

Usage

```
saveSLICES(
  in_tibble,
  outFolder = NULL,
  width = 297,
  height = 210,
  units = "mm",
  verbose = TRUE
)
```

Arguments

in_tibble	a tibble that is the output of the callRelatedness() function.
outFolder	the folder into which all diagnostic plots will be saved
width	the width of the output PDFs.
height	the height of the output PDFs.
units	the units for the height and width of the output PDFs.
verbose	Controls the printing of progress to console.

Value

nothing

Examples

```
saveSLICES(relatedness_example[1:3, ], outFolder = tempdir())
```

---

sim_geno *sim_geno*

Description

Simulated geno file of eigenstrat format

Usage

```
sim_geno(n_ind, n_snp, filename)
```

Arguments

n_ind	number of individuals
n_snp	number of SNPs
filename	filename of export

Value

NULL exports a file

Examples

```
## Not run:
sim_geno(10, 5, "geno.txt")

## End(Not run)
```

`split_line` *split line*

Description

takes a line for a SNP file and splits into parts.

Usage

```
split_line(x)
```

Arguments

x	line from SNP file

Value

tibble with 6 columns.

Examples

```
split_line("1_14.570829090394763     1        0.000000             14 A X")
split_line("rs3094315 1 0.0 752566 G A")
```

`test_degree` *test_degree*

Description

Test if a degree of relatedness is consistent with an observed PMR

Usage

```
test_degree(in_tibble, row, degree, verbose = TRUE)
```

Arguments

in_tibble	a tibble that is the output of the callRelatedness() function.
row	either the row number or pair name for which the posterior distribution is to be plotted.
degree	the degree of relatedness to be tested.
verbose	a logical (boolean) for whether all test output should be printed to screen.

Value

the associated p-value for the test

Examples

```
test_degree(relatedness_example, 1, 1)
```

Index