

Package: Fluidigm (via r-universe)

September 9, 2024

Title Handling Fluidigm Data

Version 0.2

Description Designed to streamline the process of analyzing genotyping data from Fluidigm machines, this package offers a suite of tools for data handling and analysis. It includes functions for converting Fluidigm data to format used by 'PLINK', estimating errors, calculating pairwise similarities, determining pairwise similarity loci, and generating a similarity matrix.

License GPL-3

Encoding UTF-8

SystemRequirements perl (>= 5.10.0), plink (>=1.07)

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

Imports data.table (>= 1.9.4), graphics, grDevices, reshape (>= 0.8.8), stats, utils

Suggests knitr, rmarkdown, testthat, digest

VignetteBuilder knitr

Repository <https://fischuu.r-universe.dev>

RemoteUrl <https://github.com/fischuu/fluidigm>

RemoteRef HEAD

RemoteSha e315c14ea57a391d32bb4b32783ff0c1dd1a8dde

Contents

calculatePairwiseSimilarities	2
estimateErrors	3
Fluidigm	5
fluidigm2PLINK	6
fluidigmAnalysisWrapper	8
getPairwiseSimilarityLoci	11
similarityMatrix	12

Index	14
--------------	-----------

`calculatePairwiseSimilarities`*Run plink to Calculate Pairwise Similarities*

Description

This function serves as a wrapper to the 'PLINK' software, which is a free, open-source whole genome association analysis toolset. It specifically uses 'PLINK' to calculate pairwise similarities between genotypes.

Usage

```
calculatePairwiseSimilarities(  
  file,  
  db = NA,  
  map = NA,  
  out = NA,  
  sexing = TRUE,  
  verbose = TRUE,  
  verbosity = 1  
)
```

Arguments

<code>file</code>	A string representing the path to the filtered ped/map file pair (without ped/map file extension).
<code>db</code>	A string representing the path to an existing genotype database. If not provided, the function will proceed with the existing data.
<code>map</code>	A string representing the filepath to PlateDnoY.map file. If not provided, the function will use the map file with the same name as the ped file.
<code>out</code>	A string representing the path to the output. If not provided, the output will be written to a file with the same name as the input file, appended with "_oDB".
<code>sexing</code>	A logical value indicating whether the function should try to perform sexing. Default is FALSE.
<code>verbose</code>	A logical value indicating whether the output should be verbose. Default is TRUE.
<code>verbosity</code>	An integer representing the level of verbosity. Set to a higher number for more detailed output. Default is 1.

Details

The function first checks the input parameters and sets default values if necessary. It then constructs and executes a PLINK command to merge the genotype output with the existing genotype database if one is provided. Finally, it calculates pairwise similarities for all samples (and database individuals) using another PLINK command. If the 'sexing' parameter is set to TRUE, the function will also attempt to determine the sex of the individuals.

Value

A list containing the following elements: gensim, a matrix indicating if genotypes are called correctly for replicates and/or if genotypes are missing sums, a matrix with summary statistics

References

- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) *PLINK: a toolset for whole-genome association and population-based linkage analysis*. American Journal of Human Genetics, 81.
- Purcell, Shaun. PLINK. <https://zzz.bwh.harvard.edu/plink/>

Examples

```
## Not run:  
  
outdir <- tempdir()  
  
calculatePairwiseSimilarities(file=file.path(outdir, "example_data.csv.GOOD"))  
  
## End(Not run)
```

estimateErrors

Estimate Errors in 'PLINK' ped files

Description

This function processes the 'PLINK' ped files and estimates errors. It also performs sex assignment and species marker analysis if required.

Usage

```
estimateErrors(  
  file,  
  outdir = NA,  
  db = NA,  
  appendSamplesToDB = FALSE,  
  keep.rep = 1,  
  y.marker = NA,  
  x.marker = NA,  
  sp.marker = NA,  
  plots = TRUE,  
  neg_controls = NA,  
  allele_error = 5,  
  marker_dropout = 15,  
  no_marker = 50,  
  male.y = 3,
```

```

male.hetX = 0,
female.y = 0,
female.Xtot = 8,
female.hetXtot = 3,
warning.noYtot = 2,
warning.noHetXtot = 3,
sexing = FALSE,
verbose = TRUE,
verbosity = 1
)

```

Arguments

file	A string. Path to the ped input file.
outdir	A string specifying the output folder. If left empty the original folder path of the input file will be used.
db	A string. Name of the used database. Default is NA.
appendSamplesToDB	A logical. Should new samples be added to database? Default is FALSE.
keep.rep	A numeric. Keep only this n-fold replicates, default n=1.
y.marker	A vector. Y markers for sexing. Default is NA.
x.marker	A vector. X markers for sexing. Default is NA.
sp.marker	A vector. Markers used for species-identification. Default is NA.
plots	A logical. Should plots be created? Default is TRUE.
neg_controls	A vector. Names of negative controls. Default is NA.
allele_error	A numeric. Threshold for RERUN on Allele errors. Default is 5.
marker_dropout	A numeric. Threshold for RERUN on Marker dropout. Default is 15.
no_marker	A numeric. Number of markers. Default is 50.
male.y	A numeric. Threshold for sexing, male y-chromosome markers. Default is 3.
male.hetX	A numeric. Threshold for sexing, heterozygote x-chr markers. Default is 0.
female.y	A numeric. Threshold for sexing, female y-chromosome markers. Default is 0.
female.Xtot	A numeric. Threshold for sexing, total female x-chr markers. Default is 8.
female.hetXtot	A numeric. Threshold for sexing, heterozygote x-chr markers. Default is 3.
warning.noYtot	A numeric. Threshold for sexing, when should warning be triggered. Default is 2.
warning.noHetXtot	A numeric. Threshold for sexing, when should warning be triggered. Default is 3.
sexing	A logical. Should sexing be performed? Default is FALSE.
verbose	A logical or numeric. Should the output be verbose? Default is TRUE.
verbosity	A numeric. Level of verbosity, set to higher number for more details. Default is 1.

Details

This function processes the PLINK ped files and estimates errors. It checks if the first and second run of each sample have the same genotype and if both replicates are identical. It also performs sex assignment based on the provided Y and X markers. If species marker is provided, it performs species marker analysis. The function creates a consensus PED for all "GOOD" samples and exports it along with a .map file without the Y-markers. It also creates a database file, if provided.

Value

A list containing the following elements: gensim, a matrix indicating if genotypes are called correctly for replicates and/or if genotypes are missing sums, a matrix with summary statistics

Examples

```
## Not run:
# outdir is here the output directory from the fluidigm2PLINK function

# Estimate the errors with sexing applied
estimateErrors(file=file.path(outdir, "example_data.csv.ped"),
              keep.rep = 2)

# Estimate the errors and apply sexing with y and x markers defined
estimateErrors(file=file.path(outdir, "example_data.csv.ped"),
              keep.rep = 2,
              sexing=TRUE,
              y.marker = c("Y_scaffoldY158711_762",
                           "Y_scaffoldY42647_3017",
                           "Y_scaffoldY42656_3986"),
              x.marker = c("X_scaffold11905_7659",
                           "X_scaffold17088_4621",
                           "X_scaffold1915_14108",
                           "X_scaffold4825_648",
                           "X_scaffold5374_1437",
                           "X_scaffold10171:3154"))

## End(Not run)
```

 Fluidigm

Fluidigm

Description

Designed to streamline the process of analyzing genotyping data from Fluidigm machines, this package offers a suite of tools for data handling and analysis. It includes functions for converting Fluidigm data to format used by 'PLINK', estimating errors, calculating pairwise similarities, determining pairwise similarity loci, and generating a similarity matrix.

Title

Comprehensive Analysis of Fluidigm Genotyping Data

Description

A suite of tools designed to streamline the process of analyzing genotyping data from Fluidigm machines. It includes functions for converting Fluidigm data to 'PLINK' format, estimating errors, calculating pairwise similarities, determining pairwise similarity loci, and generating a similarity matrix.

Details

The package provides a comprehensive analysis pipeline for Fluidigm genotyping data. It starts by converting the raw data from the Fluidigm machine into a format that can be used with the 'PLINK' software. It then estimates errors in the data, calculates pairwise similarities between genotypes, determines pairwise similarity loci, and generates a similarity matrix. The package is designed to make it easier and more efficient for researchers to extract meaningful insights from their genotyping studies.

References

- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) *PLINK: a toolset for whole-genome association and population-based linkage analysis*. American Journal of Human Genetics, 81.
- Purcell, Shaun. PLINK. <https://zzz.bwh.harvard.edu/plink/>

Author(s)

Maintainer: Daniel Fischer <Daniel.Fischer@luke.fi> ([ORCID](#))

Authors:

- Robert Ekblom ([ORCID](#))

Other contributors:

- Helena Johansson [contributor, funder]

fluidigm2PLINK

Convert Fluidigm Output to 'PLINK' Format

Description

This function takes a Fluidigm output file and converts it into a 'PLINK' format. The 'PLINK' ped/map format is a widely used genetic variation data format. This function is useful for researchers who want to analyze Fluidigm data using tools that accept 'PLINK' format.

Usage

```
fluidigm2PLINK(  
  file = NA,  
  map = NA,  
  out = NA,  
  outdir = NA,  
  plots = TRUE,  
  rearrange = TRUE,  
  missing.geno = "0 0",  
  fixNames = TRUE,  
  overwrite = FALSE,  
  verbose = TRUE,  
  verbosity = 1  
)
```

Arguments

file	A string specifying the path to the input file in CSV format.
map	A string specifying the filepath to the map-file that should be used.
out	A string specifying the output file name. If left empty, the original basename of the input file will be used.
outdir	A string specifying the output folder. If left empty the original folder path of the input file will be used.
plots	A logical indicating whether additional figures for conversion should be plotted. Default is TRUE.
rearrange	A logical indicating whether the ped/map output should be rearranged in order of provided map file. Default is TRUE.
missing.geno	A character string specifying how missing values should be coded. Default is "0 0".
fixNames	A logical indicating whether whitespaces from sample names should be automatically removed. Default is TRUE.
overwrite	A logical indicating wheter the original map file should be overwritten or not. Default FALSE
verbose	A logical or numerical value indicating whether the output should be verbose. Default is TRUE.
verbosity	A numerical value indicating the level of verbosity. Set to a higher number for more details. Default is 1.

Details

The function first checks the input parameters and then imports the Fluidigm data from the CSV file. It creates a new MAP file based on the information provided in the given map file. The function then creates a PED file and exports both files. If requested, the function also generates plots for genotyping success and additional summary statistics.

This function uses the PLINK software. For more information about PLINK, please refer to the official documentation.

Value

A ped/map file pair and optional diagnostic plots.

References

PLINK: Whole genome data analysis toolset - Harvard University

Examples

```
file_path_csv <- system.file("extdata", "example_data.csv", package = "Fluidigm")
file_path_map <- system.file("extdata", "example_data_withY.map", package = "Fluidigm")
outdir <- tempdir()
```

```
fluidigm2PLINK(file=file_path_csv, map=file_path_map, outdir=outdir)
```

fluidigmAnalysisWrapper

Run the Fluidigm Analysis Script Together

Description

This function serves as a wrapper for the entire analysis pipeline. It takes a Fluidigm input file and performs several operations including conversion to 'PLINK' format, error estimation, calculation of pairwise similarities, determination of pairwise similarity loci, and calculation of the similarity matrix.

Usage

```
fluidigmAnalysisWrapper(  
  file,  
  out = NA,  
  outdir = NA,  
  db = NA,  
  appendSamplesToDB = FALSE,  
  map = NA,  
  keep.rep = 1,  
  neg_controls = NA,  
  y.marker = NA,  
  x.marker = NA,  
  sp.marker = NA,  
  plots = TRUE,  
  allele_error = 5,  
  marker_dropout = 15,  
  no_marker = 50,  
  male.y = 3,  
  male.hetX = 0,
```



```

female.y = 0,
female.Xtot = 8,
female.hetXtot = 3,
warning.noYtot = 2,
warning.noHetXtot = 3,
rearrange = TRUE,
group = NA,
fixNames = TRUE,
sexing = TRUE,
similarity = 0.85,
verbose = TRUE,
verbosity = 1,
missing.geno = "0 0",
overwrite = FALSE
)

```

Arguments

file	A string specifying the path to the Fluidigm input file.
out	A string specifying the output file name. If left empty, the original basename of the input file will be used.
outdir	A string specifying the output folder. If left empty the original folder path of the input file will be used.
db	A string specifying the filepath to the database file. If not provided, the function will proceed with the existing data.
appendSamplesToDB	A logical indicating whether new samples should be added to the database. Default is FALSE.
map	A string specifying the filepath to the PlateDnoY.map file. If not provided, the function will use the map file with the same name as the ped file.
keep.rep	A numeric value indicating the number of replicates to keep. Default is 1.
neg_controls	A vector specifying the names of negative controls. Default is NA.
y.marker	A vector specifying the Y markers for sexing. Default is NA.
x.marker	A vector specifying the X markers for sexing. Default is NA.
sp.marker	A vector specifying the markers used for species identification. Default is NA.
plots	A logical indicating whether plots should be created. Default is TRUE.
allele_error	A numeric value specifying the threshold for RERUN on Allele errors. Default is 5.
marker_dropout	A numeric value specifying the threshold for RERUN on Marker dropout. Default is 15.
no_marker	A numeric value specifying the number of markers. Default is 50.
male.y	A numeric value specifying the threshold for sexing, male y-chromosome markers. Default is 3.

<code>male.hetX</code>	A numeric value specifying the threshold for sexing, heterozygote x-chr markers. Default is 0.
<code>female.y</code>	A numeric value specifying the threshold for sexing, female y-chromosome markers. Default is 0.
<code>female.Xtot</code>	A numeric value specifying the threshold for sexing, total female x-chr markers. Default is 8.
<code>female.hetXtot</code>	A numeric value specifying the threshold for sexing, heterozygote x-chr markers. Default is 3.
<code>warning.noYtot</code>	A numeric value specifying the threshold for sexing, when should warning be triggered. Default is 2.
<code>warning.noHetXtot</code>	A numeric value specifying the threshold for sexing, when should warning be triggered. Default is 3.
<code>rearrange</code>	A logical indicating whether the ped/map output should be rearranged in order of provided map file. Default is TRUE.
<code>group</code>	A string specifying the sample identifier for statistics. Default is NA.
<code>fixNames</code>	A logical indicating whether whitespaces from sample names should be automatically removed. Default is TRUE.
<code>sexing</code>	A logical indicating whether sexing should be performed. Default is FALSE.
<code>similarity</code>	Similarity threshold. Default: 0.85.
<code>verbose</code>	A logical or numerical value indicating whether the output should be verbose. Default is TRUE.
<code>verbosity</code>	A numerical value indicating the level of verbosity. Set to a higher number for more details. Default is 1.
<code>missing.geno</code>	A character string specifying how missing values should be coded. Default is "0 0".
<code>overwrite</code>	A logical indicating wheter the original map file should be overwritten or not. Default FALSE

Details

The function first checks the input parameters and sets default values if necessary. It then runs the following functions in order:

- `fluidigm2PLINK`: Converts the Fluidigm data to PLINK format.
- `estimateErrors`: Estimates errors in the PLINK ped files.
- `calculatePairwiseSimilarities`: Calculates pairwise similarities between genotypes.
- `getPairwiseSimilarityLoci`: Determines the loci of pairwise similarities.
- `similarityMatrix`: Calculates the similarity matrix. The function prints a completion message when all operations are done.

Value

A list containing the following elements: `gensim`, a matrix indicating if genotypes are called correctly for replicates and/or if genotypes are missing sums, a matrix with summary statistics

References

- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) *PLINK: a toolset for whole-genome association and population-based linkage analysis*. American Journal of Human Genetics, 81.
- Purcell, Shaun. PLINK. <https://zzz.bwh.harvard.edu/plink/>

See Also

- [fluidigm2PLINK](#): Converts the Fluidigm data to PLINK format.
- [estimateErrors](#): Estimates errors in the PLINK ped files.
- [calculatePairwiseSimilarities](#): Calculates pairwise similarities between genotypes.
- [getPairwiseSimilarityLoci](#): Determines the loci of pairwise similarities.
- [similarityMatrix](#): Calculates the similarity matrix.

Examples

```
## Not run:  
fluidigmAnalysisWrapper(file="path/to/your/file.csv", map="path/to/your/mapfile.map")  
  
## End(Not run)
```

```
getPairwiseSimilarityLoci  
Get Pairwise Similarity Loci
```

Description

This function is a wrapper to a perl script that determines the loci of pairwise similarities. It is designed to work with data from the 'PLINK' software, which is commonly used in bioinformatics for whole genome association analysis.

Usage

```
getPairwiseSimilarityLoci(file, verbose = TRUE, verbosity = 1)
```

Arguments

file	A string. This is the path to the previously created PLINK ped file (with .ped extension). The ped file contains genotype information in a format that can be used for further analysis.
verbose	A logical. If TRUE, the function will print detailed messages during its execution to help you understand what it's doing at each step. Default is TRUE.
verbosity	An integer. This parameter controls the level of verbosity. The higher the number, the more detailed the messages. Default is 1.

Details

This function first checks the input parameters. It then constructs the output file names and the command to run the perl script. The command is executed using the system function. If the 'verbose' parameter is set to TRUE, the function will print a message when it has finished running.

The perl script was written by Doug Scofield, see references.

Value

This function does not return a value in the R environment. Instead, it creates an output file with the '.pairs' extension in the same directory as the input file. This output file contains the results of the pairwise similarity loci analysis.

References

The original code this function is based on can be found at: GitHub <https://github.com/douglasgscfield/bioinfo/blob/main/sc/pairwise-loci.pl>

Examples

```
## Not run:
  outdir <- tempdir()
  getPairwiseSimilarityLoci(file = file.path(outdir, "example_data.csv.GOOD"))

## End(Not run)
```

similarityMatrix	<i>Calculate the similarity matrix</i>
------------------	--

Description

Performs pairwise similarity analysis on genotypic data.

Usage

```
similarityMatrix(
  file = NA,
  mibs.file = NA,
  pairs.file = NA,
  ped.file = NA,
  group = NA,
  plots = TRUE,
  similarity = 0.85,
  verbose = TRUE,
  verbosity = 1
)
```

Arguments

<code>file</code>	Input file path. Default: NA.
<code>mibs.file</code>	MIBS input file path. Default: NA.
<code>pairs.file</code>	PAIRS input file path. Default: NA.
<code>ped.file</code>	PED input file path. Default: NA.
<code>group</code>	Sample identifier for statistics. Default: NA.
<code>plots</code>	Should plots be created? Default: TRUE.
<code>similarity</code>	Similarity threshold. Default: 0.85.
<code>verbose</code>	Should output be verbose? Default: TRUE.
<code>verbosity</code>	Verbosity level. Default: 1.

Details

Reads genotype data, performs pairwise similarity calculations, generates plots, and outputs data for further analysis.

Value

Does not return a value. Creates output files in the same directory as the input files.

Examples

```
## Not run:  
  similarityMatrix(file = file.path(outdir, "example_data.csv.GOOD"))  
  
## End(Not run)
```

Index

* **fluidigm**

Fluidigm, [5](#)

* **genotyping**

Fluidigm, [5](#)

calculatePairwiseSimilarities, [2](#), [11](#)

estimateErrors, [3](#), [11](#)

Fluidigm, [5](#)

Fluidigm-package (Fluidigm), [5](#)

fluidigm2PLINK, [6](#), [11](#)

fluidigmAnalysisWrapper, [8](#)

getPairwiseSimilarityLoci, [11](#), [11](#)

similarityMatrix, [11](#), [12](#)