Package 'AncestryMapper'

September 28, 2016

Version 2.0

Date 2016-09-24

Title Assigning Ancestry Based on Population References

Author Tiago R Magalhaes, Eoghan T O'Halloran

Maintainer Eoghan T O'Halloran <eoghan.ohalloran2@mail.dcu.ie>

Description Assigns genetic ancestry to an individual and
studies relationships between local and global populations.

License GPL-2

LazyLoad yes

Suggests knitr

VignetteBuilder knitr

SystemRequirements None, but greater than 12 GB of RAM highly
recommended.

RoxygenNote 5.0.1

Depends R(>= 2.10)

Imports svd(>= 0.3.3-2), parallel(>= 3.2.1), grDevices, graphics,
stats, utils

NeedsCompilation no

Repository CRAN

Date/Publication 2016-09-28 19:28:04

R topics documented:

AncestryMapper	*Ancestry Mapper 2.0*

Description

Package computes the genetic distance, defined as the euclidean distance amongst a sample set of individuals and any number of human population references. The package allows for the visualisation of the relationship of sample individuals to the reference populations, thus permitting the user to assess the relationship of individuals to different geographic groupings.

Package comes with pre-loaded with toy data and toy references from various sources comprising 158 global populations.

Additional and full Population References can be downloaded from:

http://bit.ly/1OUstDP

Details

Package:	AncestryMapper
Type:	Package
Version:	2.0
Date:	2016-08-??
License:	None
LazyLoad:	yes

Author(s)

Eoghan T. O'Halloran ,Tiago R. Magalh\~aes, Darren J. Fitzpatrick

Maintainer: Eoghan T. O'Halloran <eoghan.ohalloran2@mail.dcu.ie>

References

Magalh\~aes et al, 2012 PLOS One accepted.

See Also

image dist

Examples

```
## Not run:
library(AncestryMapper)
Refs <- system.file ("data", package = "AncestryMapper")
```

```
tpeds <- system.file ("extdata", package = "AncestryMapper")

Corpheno <- system.file ("extdata", "CorPheno", package =
"AncestryMapper")

All00Frq <- system.file ("data", "MinMaxFreq.rda", package = "AncestryMapper")

genetic.distance <- calculateAMidsArith(pathTotpeds = tpeds,
                                        NameOut = "Example",
                                        pathToAriMedoids = Refs,
                                        pathAll00 = All00Frq)

plotAMids(AMids = genetic.distance, phenoFile = Corpheno, columnPlot = "I")
## End(Not run)
```

---

arithmeticRefsMedoids   *Object each reference file is assigned as it is loaded in or read.*

Description

arithmeticRefsMedoids is the name of the object each reference is assigned to after it is loaded or read.

Usage

```
arithmeticRefsMedoids
```

Format

A list of two character vectors, the first containing the name of each SNP and the second containing the arithmetic medoid value for each SNP.

Source

Ancestry Mapper from various datasets. See: http://bit.ly/1OUstDP

References

Various

---

calculateAMids   *Calculate genetic distances.*

Description

calculates and assigns Ancestry Mapper Ids (AMids) in a more crude, but faster manner than calculateAMids

Usage

```
calculateAMids(pedtxtFile, fileReferences)
```

Arguments

pedtxtFile   Character vector giving path to PED file to be used. The PED file should include all 51 HGDP references and the individuals for which the user wishes to calculate the genetic distance.

fileReferences   Character vector giving path to file File detailing the individuals in the ped file that correspond to the references, and the populations they refer to. A file that uses the 51 HGDP reference populations is provided with the package.

Examples

```
## Not run:
library(AncestryMapper)

HGDP_References <- system.file('extdata',
                               'HGDP_References.txt',
                                package = 'AncestryMapper')


HGDP_500SNPs <- system.file('extdata',
                            'HGDP_500SNPs.ped',
                             package = 'AncestryMapper')

Corpheno <- system.file('extdata',
                         'CorPheno',
                          package = 'AncestryMapper')

genetic.distance <- calculateAMids(pedtxtFile = HGDP_500SNPs,
                                   fileReferences = HGDP_References)

plotAMids(AMids = genetic.distance, phenoFile = Corpheno, columnPlot = "I")

## End(Not run)
```

---

calculateAMidsArith *Calculate genetic distances.*

Description

Calculates genetic distance between samples and population references.

Usage

```
calculateAMidsArith(pathTotpeds, pathToAriMedoids, AMmcapply = F, nrcores,
   seqchip = "", noseqdat = F, wd, NameOut = NULL, pathAll00)
```

Arguments

pathTotpeds   Character vector giving path to folder containing the plink tPED file(s) to be
                   used.

pathToAriMedoids

                   Character vector giving path to folder containing the arithmetic references to be
                   used.

AMmcapply   Logical value (TRUE or FALSE), specifying if the multicore funcion mcapply,
                   should be used. Inappropriate for most HPC cluster systems. Default = FALSE

nrcores   Numeric value detailing how many cores should be used if AMmcapply==TRUE.
                   If left unspecificed the number of cores will be detected and nrcores will be set
                   to that number -2.

seqchip   Character vector specifying if only references from one main SNP chip panel are
                   to be used. All references are marked with what chip panel they use at the end
                   of their file names, eg 'Yoruba.HGDP.20000.Illumina.ods' May be important if
                   your data has few SNPs in common with one panel. All toy references prepared
                   use 'Illumina' panels. Whole Genome sequence data is specified with 'WG'.
                   Supports custom designations, but will trigger a warning when used.

noseqdat   Logical value (TRUE or FALSE), specifying if sequence data is to be excluded,
                   will use only references that do not have names ending in '.WG.ods/rds/rda'.
                   Default = FALSE

wd   Character vector giving the desired working directory to house the outputs of
                   calculateAMidsArith. If left unspecified will use current working directory.

NameOut   Character vector giving the desired prefix name for the AMid file. Default is
                   NULL.

pathAll00   Character vector giving the path to a file containing the full data table of each
                   dbSNP and both alleles. An example version covering the SNPs used in the
                   example data is included. A full version can be found at: http://bit.ly/1OUstDP

Examples

```
## Not run:
Refs <- system.file('data', package = 'AncestryMapper')
tpeds <- system.file('extdata', package = 'AncestryMapper')
Corpheno <- system.file('extdata', 'CorPheno', package = 'AncestryMapper')
All00Frq <- system.file ('data', 'MinMaxFreq.rda', package = 'AncestryMapper')

genetic.distance <- calculateAMidsArith(pathTotpeds = tpeds,
                                        NameOut = 'Example',
                                        pathToAriMedoids = Refs,
                                        pathAll00 = All00Frq)

plotAMids(AMids = genetic.distance, phenoFile = Corpheno, columnPlot = "I")

## End(Not run)
```

---

createMedoid   *Creates Ancestry Mapper Population Reference.*

Description

Generates arithmetic population reference from PLINK tPED files.

Usage

```
createMedoid(pathTotpeds, AMmcapply = F, nrcores, wd, pathAll00,
  chipMan = "ChipMan", OutForm = "rda")
```

Arguments

pathTotpeds	Character vector giving path to folder containing tPED file(s) to be used.
AMmcapply	Logical value (TRUE or FALSE), specifying if the multicore funcion mcapply, should be used. Inappropriate for most HPC cluster systems. Default = FALSE
nrcores	Numeric value detailing how many cores should be used if AMmcapply==TRUE. If left unspecified the number of cores will be detected and mc.cores will be set to that number -1.
wd	Character vector giving the desired working directory to house the outputs of calculateAMids. If left unspecified will use current working directory.
pathAll00	Character vector giving the path to a file containing the full data table of each dbSNP and both alleles. A toy version covering the SNPs used in the toy data is included. A full version can be found at: http://bit.ly/1OUstDP
chipMan	Character vector giving name of company from which the SNP panel is derived. Eg 'Illumina', 'Affymetrix'. If no value is given will default to 'ChipMan'. If it is whole genome sequencing, please put 'WG'. The value will appear in the name of the arithmetic reference file. e.g. 'medoidArithmetic_Yoruba_HGDP_1000_Illumina.rda'.

OutForm   Character vector giving option for output format for arithmetic medoids. Can be
                       one of three options. 'ods' will generate a raw text file with the default extension
                       of '.ods'. This is the default option as is the most flexible format. 'rds' will save
                       the arithmetic medoid as a .rds file which can be loaded into R faster and is also
                       roughly a third the size of the raw text version. 'rda' will save the arithmetic
                       medoid as a .rda file which can be loaded into R faster and is also roughly a
                       third the size of the raw text version.

Examples

```
## Not run:
chipManExample <- 'Illumina'
tpeds <- system.file('extdata', package='AncestryMapper')

createMedoid(pathTotpeds = tpeds, chipMan = chipManExample)


## End(Not run)
```

MinMaxFreq   *Allele Variants for Demo SNPs According to dbSNP.*

Description

This data set contains the major and minor alleles for the 1000 SNPs used in the demo data accord-
ing to the strand orientation used on dbSNP.

Usage

```
MinMaxFreq
```

Format

A matrix containing 1000 observations of two variables.

Source

dbSNP.

References

"ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606/VCF/00-All.vcf.gz"

---

plotAMids   *Visualises genetic distances.*

Description

plotAMids is used to visualise the relationship amongst individuals and references.

Usage

```
plotAMids(AMids, phenoFile, columnPlot = "I", quantilePlot = TRUE,
  colorPlot = "BlBrewer", sepLinesPop = TRUE, plotIndNames = FALSE,
  legColor = TRUE, legRef = TRUE, legPheno = TRUE, legAxisPop = TRUE,
  legData = FALSE, bmar, lmar, tmar, rmar, cexref = 0.9, cexind = 0.8)
```

Arguments

AMids	Dataframe of genetic distances calculated by calculateAMids or calculateAMidsArith.
phenoFile	Optional file with phenotype, color and order information for individuals and populations. An example file, called CorPheno, is contained in the 'extdata' folder with the package.
columnPlot	Takes values 'I' or 'C'. 'I' is the default option. 'I' plots the normalised euclidean distances whereas 'C' plots the crude distances.
quantilePlot	Logical. Takes values TRUE or FALSE. TRUE is the default option. If columnPlot is 'C', TRUE will plot the quantiles, FALSE will plot the raw values.
colorPlot	Colors for the AMids. Possible choices are 'RedBl', 'RedBlGr' and 'BLBrewer'. The user can also provide a vector of colors.
sepLinesPop	Logical. Takes values TRUE or FALSE. The default is TRUE. If TRUE, a line demarcating populations is plotted.
plotIndNames	Logical. Takes values TRUE or FALSE. The default is FALSE. If TRUE, the individual ids are plotted on the left axis.
legColor	Logical. Takes values TRUE or FALSE. The default is TRUE. If TRUE, the legend for the colour gradient will be plotted in the top left.
legRef	Logical. Takes values TRUE or FALSE. The default is TRUE. If TRUE, text giving names of references will be plotted along the x axis.
legPheno	Logical. Takes values TRUE or 'no. The default is TRUE. If TRUE, will plot colour blocks relating to the population, dataset and regional origin of data if sample IDs have been given these in the Corpheno file, if not present, will plot them under 'Unspecified'.
legAxisPop	Logical. Takes values TRUE or FALSE. The default is TRUE. If TRUE text will be plotted giving name and number of populations for samples used on the right y axis of the plot.

legData	Logical. Takes values TRUE or FALSE. The default is FALSE. If TRUE, the reference to the dataset used to create the reference is appended to the reference population name on the bottom x axis.
bmar	Takes numeric value. Changes the size of the bottom outer margin of the plot. The default is empty. For more see ?par()
lmar	Takes numeric value. Changes the size of the left outer margin of the plot. The default is empty. For more see ?par()
tmar	Takes numeric value. Changes the size of the top outer margin of the plot. The default is empty. For more see ?par()
rmar	Takes numeric value. Changes the size of the right outer margin of the plot. The default is empty. For more see ?par()
cexref	Takes numeric value. Controls text size of reference names on y axis. Default is 0.9.
cexind	Takes numeric value. Controls text size of sample names on y axis. Default is 0.8. Individual sample IDs need plotIndNames = "yes" to display, this is set to FALSE by default.

Examples

```
## Not run:
Refs <- system.file('data', package = 'AncestryMapper')
tpeds <- system.file('extdata', package = 'AncestryMapper')
Corpheno <- system.file('extdata', 'CorPheno', package = 'AncestryMapper')
All00Frq <- system.file ('data', 'MinMaxFreq.rda', package = 'AncestryMapper')

genetic.distance <- calculateAMidsArith(pathTotpeds = tpeds,
                                        NameOut = 'Example',
                                        pathToAriMedoids = Refs,
                                        pathAll00 = All00Frq)

plotAMids(AMids = genetic.distance, phenoFile = Corpheno, columnPlot = "I")

## End(Not run)
```

---

refAdd   *Add individuals to the CorPheno file.*

Description

Add individuals to the CorPheno file in order to order and plot them with specific colours. Ids that aren't in the CorPheno will be plotted under 'Unspecified' and given a grey colour by plotAMids()

Usage

```
refAdd(phenoId, phenoValues, ignoreDupes = F, phenoFile, writeCor = T)
```

Arguments

phenoId	Path to file containing list of invididuals to be added to CorPheno with the three columns 'UNIQID','Fam','Pheno_Pop'. A full example called 'Example.phenoId' is present in the 'extdata' folder of the AncestryMapper package.
phenoValues	Path to file containing information on each population to be added, such as continental origin and colours as well as other information. A full example called 'Example.phenoValues' is present in the 'extdata' folder of the AncestryMapper package.
ignoreDupes	Logical value (TRUE or FALSE), specifying if the presence of individual IDs already in CorPheno should be ignored. Useful if the user knows this is the case and just wants the individuals not already included in the directed phenoFile. Default = FALSE
phenoFile	Main file with phenotype information for each individual. A sample file called CorPheno is included with the package in the ext folder. It contains values for the samples from the HGDP. This function augments this file with any novel individuals. If no value is given the sample file in the ext folder is used by default.
writeCor	Logical value (TRUE or FALSE), specifying if the new CorPheno should overwrite the file the given in 'phenoFile'. A backup of the previous CorPheno file with the same path as given in 'phenoFile' with '_Original' appended to the name will also be produced. You could alternatively write out your new file to your preferred location with write.table, making sure to keep the columns tab spaced. Default = TRUE

Examples

```
## Not run:
phenoIdPth <- system.file ("extdata", "Example.phenoId", package = "AncestryMapper")
PhenoValPth <- system.file("extdata", "Example.phenoValues", package = "AncestryMapper")
Corpheno <- system.file("extdata", "CorPheno", package = "AncestryMapper")

refAdd(phenoId = phenoIdPth, phenoValues = PhenoValPth, phenoFile = Corpheno)


## End(Not run)
```

Index