

Package ‘alleHap’

August 19, 2017

Type Package

Title Allele Imputation and Haplotype Reconstruction from Pedigree Databases

Version 0.9.9

Date 2017-08-19

Author Nathan Medina-Rodriguez and Angelo Santana

Maintainer Nathan Medina-Rodriguez <nathan.medina@ulpgc.es>

Description Tools to simulate alphanumeric alleles, impute genetic missing data and reconstruct non-recombinant haplotypes from pedigree databases in a deterministic way. Allelic simulations can be implemented taking into account many factors (such as number of families, markers, alleles per marker, probability and proportion of missing genotypes, recombination rate, etc). Genotype imputation can be used with simulated datasets or real databases (previously loaded in .ped format). Haplotype reconstruction can be carried out even with missing data, since the program firstly imputes each family genotype (without a reference panel), to later reconstruct the corresponding haplotypes for each family member. All this considering that each individual (due to meiosis) should unequivocally have two alleles per marker (one inherited from each parent) and thus imputation and reconstruction results can be deterministically calculated.

Imports abind, tools, stats, utils

License GPL (>= 2)

Suggests knitr

VignetteBuilder knitr

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2017-08-19 10:33:43 UTC

R topics documented:

alleHap-package	2
alleHaplotyper	3
alleImputer	5
alleLoader	6
alleSimulator	7

Index	9
--------------	----------

alleHap-package	<i>Allele Imputation and Haplotype Reconstruction from Pedigree Databases</i>
-----------------	---

Description

Tools to simulate alphanumeric alleles, impute genetic missing data and reconstruct non-recombinant haplotypes from pedigree databases in a deterministic way. Allelic simulations can be implemented taking into account many factors (such as number of families, markers, alleles per marker, probability and proportion of missing genotypes, recombination rate, etc). Genotype imputation can be used with simulated datasets or real databases (previously loaded in .ped format). Haplotype reconstruction can be carried out even with missing data, since the program firstly imputes each family genotype (without a reference panel), to later reconstruct the corresponding haplotypes for each family member. All this considering that each individual (due to meiosis) should unequivocally have two alleles per marker (one inherited from each parent), and thus imputation and reconstruction results can be deterministically calculated.

Details

Package:	alleHap
Type:	Package
Version:	0.9.9
Date:	2017-08-19
Depends:	abind, stats, tools, utils
License:	GPL (>=2)

Author(s)

Nathan Medina-Rodriguez and Angelo Santana

Maintainer: Nathan Medina-Rodriguez <nathan.medina@ulpgc.es>

References

Medina-Rodriguez, N. Santana A. et al. (2014) alleHap: an efficient algorithm to reconstruct zero-recombinant haplotypes from parent-offspring pedigrees. BMC Bioinformatics, 15, A6 (S-3).

Examples

```

## Generation of 10 simulated families with 2 children per family and 20 markers
dataset <- alleSimulator(10,2,20) # List with simulated alleles and haplotypes
datasetAlls <- dataset[[1]]      # Dataset containing alleles
datasetHaps <- dataset[[2]]     # Dataset containing haplotypes

## Loading of a dataset in .ped format with alphabetical alleles (A,C,G,T)
example1 <- file.path(find.package("alleHap"), "examples", "example1.ped")
datasetAlls1 <- alleLoader(example1)

## Loading of a dataset in .ped format with numerical alleles
example2 <- file.path(find.package("alleHap"), "examples", "example2.ped")
datasetAlls2 <- alleLoader(example2)

## Allele imputation of families with parental missing data
datasetAlls <- alleSimulator(10,4,6,missParProb=0.2)[[1]]
famsImputed <- alleImputer(datasetAlls)

## Allele imputation of families with offspring missing data
datasetAlls <- alleSimulator(10,4,6,missOffProb=0.2)[[1]]
famsImputed <- alleImputer(datasetAlls)

## Haplotype reconstruction for 3 families without missing data.
simulatedFams <- alleSimulator(3,3,6)
(famsAlls <- simulatedFams[[1]]) # Original data
famsList <- alleHaplotyper(famsAlls) # List containing families' alleles and haplotypes
famsList$reImputedAlls           # Re-imputed alleles
famsList$haplotypes              # Reconstructed haplotypes

## Haplotype reconstruction from a PED file
pedFamPath <- file.path(find.package("alleHap"), "examples", "example3.ped") # PED file path
pedFamAlls <- alleLoader(pedFamPath,dataSummary=FALSE)
pedFamList <- alleHaplotyper(pedFamAlls)
pedFamAlls           # Original data
pedFamList$reImputedAlls # Re-imputed alleles
pedFamList$haplotypes # Reconstructed haplotypes

```

alleHaplotyper

Haplotyping of a dataset composed by several families.

Description

By analyzing all possible combinations of a parent-offspring pedigree in which parents may be missing ($\text{missParProb} > 0$), as long as one child was genotyped, it is possible an unequivocal reconstruction of many parental haplotypes. When neither parent was genotyped ($\text{missParProb} == 1$), also it is possible to reconstruct at least two parental haplotypes in certain cases. Regarding offspring haplotypes, if both parents are completely genotyped ($\text{missParProb} == 0$), in majority of cases partial offspring haplotypes may be successfully obtained ($\text{missOffProb} > 0$).

Usage

```
alleHaplotyper(data, NASymbol = "?", alleSep = "", invisibleOutput = TRUE,
               dataSummary = TRUE)
```

Arguments

data	Data containing non-genetic and genetic information of families (or PED file path).
NASymbol	Icon which will be placed in the NA values of the haplotypes.
alleSep	Icon which will be used as separator of the haplotype alleles.
invisibleOutput	Data are not shown by default.
dataSummary	A summary of the data is shown by default.

Value

Re-imputed alleles and haplotypes for each loaded family.

References

Medina-Rodriguez, N. Santana A. et al. (2014) alleHap: an efficient algorithm to reconstruct zero-recombinant haplotypes from parent-offspring pedigrees. BMC Bioinformatics, 15, A6 (S-3).

Examples

```
## Haplotype reconstruction for 3 families without missing data.
simulatedFams <- alleSimulator(3,3,6)
(famsAlls <- simulatedFams[[1]]) # Original data
famsList <- alleHaplotyper(famsAlls) # List containing families' alleles and haplotypes
famsList$reImputedAlls # Re-imputed alleles
famsList$haplotypes # Reconstructed haplotypes

## Haplotype reconstruction of a family containing missing data in a parent.
infoFam <- data.frame(famID="FAM002", indID=1:6, patID=c(0,0,1,1,1,1),
                     matID=c(0,0,2,2,2,2), sex=c(1,2,1,2,1,2), phenot=c(2,1,1,2,1,2))
Mkrs <- rbind(c(1,4,2,5,3,6), rep(NA,6), c(1,7,2,3,3,2),
             c(4,7,5,3,6,2), c(1,1,2,2,3,3), c(1,4,2,5,3,6))
colnames(Mkrs) <- c("Mk1_1", "Mk1_2", "Mk2_1", "Mk2_2", "Mk3_1", "Mk3_2")
(family <- cbind(infoFam, Mkrs)) # Original data
famList <- alleHaplotyper(family) # List containing family's alleles and haplotypes
famList$reImputedAlls # Re-imputed alleles
famList$haplotypes # Reconstructed haplotypes

## Haplotype reconstruction from a PED file
pedFamPath <- file.path(find.package("alleHap"), "examples", "example3.ped") # PED file path
pedFamAlls <- alleLoader(pedFamPath, dataSummary=FALSE)
pedFamList <- alleHaplotyper(pedFamAlls)
pedFamAlls # Original data
pedFamList$reImputedAlls # Re-imputed alleles
```

```
pedFamList$haplotypes      # Reconstructed haplotypes
```

```
alleImputer                Imputation of missing alleles from a dataset composed by families.
```

Description

By analyzing all possible combinations of a parent-offspring pedigree in which parental and/or offspring genotypes may be missing; as long as one child was genotyped, in certain cases it is possible an unequivocal imputation of the missing genotypes both in parents and children.

Usage

```
alleImputer(data, invisibleOutput = TRUE, dataSummary = TRUE)
```

Arguments

data	Data containing the families' identifiers and the corresponding genetic data (or the path of the PED file).
invisibleOutput	Data are not shown by default.
dataSummary	A summary of the data is shown by default.

Value

Imputed markers, Homozygosity (HMZ) matrix, marker messages and number of unique alleles per marker.

References

Medina-Rodriguez, N. Santana A. et al. (2014) alleHap: an efficient algorithm to reconstruct zero-recombinant haplotypes from parent-offspring pedigrees. *BMC Bioinformatics*, 15, A6 (S-3).

Examples

```
## Imputation of families containing parental missing data
simulatedFams <- alleSimulator(10,4,6,missParProb=0.2)
famsAlls <- simulatedFams[[1]]      # Original data
alleImputer(famsAlls)              # Imputed alleles (genotypes)

## Imputation of families containing offspring missing data
datasetAlls <- alleSimulator(10,4,6,missOffProb=0.2)
famsAlls <- simulatedFams[[1]]     # Original data
alleImputer(famsAlls)              # Imputed alleles (genotypes)

## Imputation of a family marker containing missing values in one parent and one child
```

```

infoFam <- data.frame(famID="FAM03",indID=1:5,patID=c(0,0,1,1,1),
                    matID=c(0,0,2,2,2),sex=c(1,2,1,2,1),phenot=0)
mkr <- rbind(father=c(NA,NA),mother=c(1,3),child1=c(1,1),child2=c(2,3),child3=c(NA,NA))
colnames(mkr) <- c("Mkr1_1","Mkr1_2")
famMkr <- cbind(infoFam,mkr)          # Original data
alleImputer(famMkr)                 # Imputed alleles (genotypes)

```

alleLoader	<i>Data loading of nuclear families (in .ped format)</i>
------------	--

Description

The data to be loaded must be structured in .ped format and families must comprise by parent-offspring pedigrees.

Usage

```

alleLoader(data, invisibleOutput = TRUE, dataSummary = TRUE,
           missingValues = c(-9, -99))

```

Arguments

data	Data to be loaded.
invisibleOutput	Data are not shown by default.
dataSummary	A summary of the data is shown by default.
missingValues	Specification of the character/numerical values which may be missing.

Value

Loaded dataset.

References

Medina-Rodriguez, N. Santana A. et al. (2014) alleHap: an efficient algorithm to reconstruct zero-recombinant haplotypes from parent-offspring pedigrees. BMC Bioinformatics, 15, A6 (S-3).

Examples

```

## Loading of a dataset in .ped format with alphabetical alleles (A,C,G,T)
example1 <- file.path(find.package("alleHap"), "examples", "example1.ped")
example1Alls <- alleLoader(example1)
head(example1Alls)

## Loading of a dataset in .ped format with numerical alleles
example2 <- file.path(find.package("alleHap"), "examples", "example2.ped")

```

```
example2Alls <- alleLoader(example2)
head(example2Alls)
```

alleSimulator	<i>Simulation of genetic data (alleles) and non-genetic data (family identifiers)</i>
---------------	---

Description

Data simulation can be performed taking into account many different factors such as number of families to generate, number of markers (allele pairs), number of different alleles per marker, type of alleles (numeric or character), number of different haplotypes in the population, probability of parent/offspring missing genotypes, proportion of missing genotypes per individual, probability of being affected by disease and recombination rate.

Usage

```
alleSimulator(nFams = 2, nChildren = NULL, nMarkers = 3,
  numAllperMrk = NULL, chrAlleles = TRUE, nHaplos = 1200,
  missParProb = 0, missOffProb = 0, ungenotPars = 0, ungenotOffs = 0,
  phenProb = 0.2, recombRate = 0, invisibleOutput = TRUE)
```

Arguments

nFams	Number of families to generate (integer: 1..1000+)
nChildren	Number of children of each family (integer: 1..7 or NULL)
nMarkers	Number of markers or allele pairs to generate (integer: 1..1000+)
numAllperMrk	Number of different alleles per marker (vector or NULL)
chrAlleles	Should alleles be expressed as characters A,C,G,T ? (boolean: FALSE, TRUE)
nHaplos	Number of different haplotypes in the population (numeric)
missParProb	Probability of parents' missing genotype (numeric: 0..1)
missOffProb	Probability of offspring' missing genotype (numeric: 0..1)
ungenotPars	Proportion of ungenotyped parents (numeric: 0..1)
ungenotOffs	Proportion of ungenotyped offspring (numeric: 0..1)
phenProb	Phenotype probability, e.g. being affected by disease (numeric: 0..1)
recombRate	Recombination rate (numeric: 0..1)
invisibleOutput	Data are not shown by default.

Value

Families' genotypes and haplotypes.

References

Medina-Rodriguez, N. Santana A. et al. (2014) alleHap: an efficient algorithm to reconstruct zero-recombinant haplotypes from parent-offspring pedigrees. BMC Bioinformatics, 15, A6 (S-3).

Examples

```
## Generation of 5 simulated families with 2 children per family and 10 markers
simulatedFams <- alleSimulator(5,2,10) # List with simulated alleles and haplotypes
simulatedFams[[1]] # Alleles (genotypes) of the simulated families
simulatedFams[[2]] # Haplotypes of the simulated families
```


Index

alleHap (alleHap-package), [2](#)
alleHap-package, [2](#)
alleHaplotyper, [3](#)
alleImputer, [5](#)
alleLoader, [6](#)
alleSimulator, [7](#)