Package 'MCIBD'

October 9, 2009

Type Package

Title Monte Carlo Identity-By-Descent Matrix Estimation

Version 0.1

Date 2009-10-01

Author Xia Shen

Maintainer Xia Shen <xia.shen@lcb.uu.se>

Description This is a package implemented for flexible identity-by-descent (IBD) matrix estimation in F2 pedigrees. The IBD matrix estimate is approached by a Monte Carlo strategy, where the segregation of founder alleles is easy to set up. Epistatic IBD matrices can also be calculated for two arbitrary loci. Such IBD matrix estimates are typically used in variance component quantitative trait loci (QTL) analysis. Parallelization is implemented to enhance the performance of Monte Carlo sampling.

License GPL

Depends sfsmisc

Suggests snow, snowfall

R topics documented:

MCIBD-package *Monte Carlo Identity-By-Descent Matrix Estimation*

Description

MCIBD is a package implemented for flexible identity-by-descent (IBD) matrix estimation in F2 pedigrees. The IBD matrix estimate is approached by a Monte Carlo strategy, where the segregation of founder alleles is easy to set up. Epistatic IBD matrices can also be calculated for two arbitrary loci. Such IBD matrix estimates are typically used in variance component quantitative trait loci (QTL) analysis. Parallelization is implemented to enhance the performance of Monte Carlo sampling.

Details

Package:	MCIBD
Type:	Package
Version:	0.1
Date:	2009-10-01
License:	GPL
Depends:	sfsmisc
Suggests:	snow, snowfall

Author(s)

Xia Shen

Maintainer: Xia Shen <xia.shen@lcb.uu.se>

References

Andersson, L., Haley, C.S., Ellegren, H., Knott, S.A., Johansson, M., Andersson, K., Andersson-Eklund L., Edfors-Lilja, I., Fredholm, M., Hansson, I., Hakansson, J. and Lundstrom, K. (1994). *Genetic Mapping of Quantitative Trait Loci for Growth and Fatness in Pigs*. Science, 263: 1771-1774.

Nettelblad, C., Holmgren, S., Crooks, L. and Carlborg, O. (2009). *cnF2freq: Efficient Determination of Genotype and Haplotype Probabilities in Outbred Populations Using Markov Models*. Lecture Notes in Bioinformatics (LNBI), 5462: 307-319, Springer-Verlag Berlin Heidelberg.

Shen, X., Ronnegard, L. and Carlborg, O. (2009). *A Monte Carlo Full Likelihood Approach in Variance Component Quantitative Trait Loci Analysis*. (Submitted.)

cnF2freq *Calculating Inheritance Probabilities using cnF2freq*

Description

`cnF2freq` is used to run the `cnF2freq` program or even compile `cnF2freq` in different operating systems. Usually on a shared cluster, compiling is required for the first time usage.

Usage

```
cnF2freq(ped.file, chr.file, os = "unix", compile = FALSE)
```

Arguments

`ped.file`	a file containing pedigree information. (See `pedi.ric` in `other.zip` or `other.tar` for example.)
`chr.file`	a file containing chromosome marker information. (See `chro.ric` in `other.zip` or `other.tar` for example.)
`os`	a string specifying the operating system, typically "unix" or "linux" or "windows".
`compile`	logical. If `TRUE`, `cnF2freq` is compiled before running.

Details

For computers with the user authorities, `cnF2freq` may not need to be compiled.

Value

An output file `cnF2freq.out` is produced in the working directory.

Author(s)

Xia Shen

References

Nettelblad, C., Holmgren, S., Crooks, L. and Carlborg, O. (2009). *cnF2freq: Efficient Determination of Genotype and Haplotype Probabilities in Outbred Populations Using Markov Models*. Lecture Notes in Bioinformatics (LNBI), 5462: 307-319, Springer-Verlag Berlin Heidelberg.

See Also

`MCIBD-package`

Examples

```
## Unzip other.zip or other.tar in the package library folder.
## Make sure you are using an operating system with g++ compiler installed.
## Copy the file cnF2freq.cpp in the working directory.
## Execute this function, setting compile = TRUE.

## You may also copy the file cnF2freq_unix or cnF2freq_linux
## or cnF2freq_windows.exe, then with user authority,
## this function can be executed without compiling.
```

MCIBD *Estimating Identity-By-Descent (IBD) Matrices using Monte Carlo Sampling*

Description

MCIBD estimates the IBD matrix at a given locus or the epistatic IBD matrix for two linked loci. Monte Carlo sampling is used to approach to the matrix estimator, where the output file from cnF2freq is required. Segregation of the founder alleles can be set up in the estimation. Parallelzation is available using snowfall package if executing on a multi-core computer or cluster.

Usage

```
MCIBD(loci, n.F2, pedigree = NULL, cnF2freq.out = NULL,
      output.Z = FALSE, read.file = FALSE, segregation = NULL,
      mc.size = 99, hpc = FALSE, n.cpus = 2)
```

Arguments

loci	typically an integer specifying a test locus, or a vector containing two linked loci.
n.F2	an integer telling the number of F2 individuals in the pedigree.
pedigree	a matrix or data.frame or file containing pedigree information. For file (when read.file is TRUE), see pedi.ric in other.zip or other.tar for example.
cnF2freq.out	a matrix or data.frame containing the output probabilities from cnF2freq. For file (when read.file is TRUE), see chro.ric in other.zip or other.tar for example.
output.Z	logical. If TRUE, average incidence matrix Z is estimated instead of the corresponding IBD matrix.
read.file	logical. If TRUE, data is loaded from files, where a pedigree information file is specified by pedigree and a probabilities output file cnF2freq.out in the working directory.
segregation	a vector specifying the segregation of founder alleles. (See Details.)
mc.size	an integer setting the sample size of the Monte Carlo sampling for one CPU.
hpc	logical. If TRUE, high performance computing is carried out by parallelization.
n.cpus	an integer telling the number of cores that parallelization is executed on. Only useful when hpc is TRUE. Not recommended when mc.size is small.

Details

To set up the segregation of the founder alleles, suppose that we have *m* founder(s) from one line and *n* from another line. Then in total there are *2 * (m + n)* alleles needed to be set up. According to the order in the pedigree information file, these alleles for the founders should be filled in segregation as a vector with length *2 * (m + n)*. For instance, if there is one male in one line and three females in another line, the vector might be created like c(1,1,2,2,3,3,2,3), which means that the male has two identical alleles whereas the genotypes of the females are (2,2), (3,3) and (2,3), respectively. If NULL, all the founder alleles are assumed to be different from each other.

Value

An output IBD matrix is saved in the working directory as a file named by the loci with extension `.ibd`.

Author(s)

Xia Shen

References

Shen, X., Ronnegard, L. and Carlborg, O. (2009). *A Monte Carlo Full Likelihood Approach in Variance Component Quantitative Trait Loci Analysis*. (Submitted.)

See Also

`cnF2freq`, `MCIBD.chro`, `MCIBD.epi2chro`, `MCIBD-package`

Examples

```
data(pedigree)
data(probabilities)

## Calculation on one CPU
MCIBD(loci = 80, n.F2 = 191, pedigree = pedigree,
      cnF2freq.out = probabilities, mc.size = 5)
## IBD matrix of dimension 191 x 191 at locus 80 is accomplished,
## where 5 imputes were sampled.

## Calculation on 2 CPUs with segregation of the founder alleles
MCIBD(loci = 90, n.F2 = 191, pedigree = pedigree,
      segregation = c(rep(1,4),rep(2,16)),
      cnF2freq.out = probabilities, mc.size = 5, hpc = TRUE, n.cpus = 2)
## IBD matrix of dimension 191 x 191 at locus 90 is accomplished,
## where 10 imputes were sampled.

## Calculation of epistatic IBD matrix
MCIBD(loci = c(88, 99), n.F2 = 191, pedigree = pedigree,
      segregation = c(rep(1,4),rep(2,16)),
      cnF2freq.out = probabilities, mc.size = 5, hpc = TRUE, n.cpus = 2)
## IBD matrix of dimension 191 x 191 for linked lici 88 and 99 is accomplished,
## where 10 imputes were sampled.
```

MCIBD.chro *Estimating Identity-By-Descent (IBD) Matrices along A Whole Chromosome*

Description

`MCIBD.chro` estimates the IBD matrices along a given chromosome. Monte Carlo sampling is used to approach to the matrix estimator, where the output file from `cnF2freq` is required. Segregation of the founder alleles can be set up in the estimation. Parallelzation is available using `snowfall` package if executing on a multi-core computer or cluster.

Usage

```
MCIBD.chro(dis = 5, n.F2, pedigree = NULL, cnF2freq.out = NULL,
           output.Z = FALSE, read.file = FALSE, segregation = NULL,
           mc.size = 99, hpc = FALSE, n.cpus = 2)
```

Arguments

dis	the distance between two nearest test loci in centimorgan.
n.F2	an integer telling the number of F2 individuals in the pedigree.
pedigree	a matrix or data.frame or file containing pedigree information. For file (when read.file is TRUE), see pedi.ric in other.zip or other.tar for example.
cnF2freq.out	a matrix or data.frame containing the output probabilities from cnF2freq. For file (when read.file is TRUE), see chro.ric in other.zip or other.tar for example.
output.Z	logical. If TRUE, average incidence matrix Z is estimated instead of the corresponding IBD matrix.
read.file	logical. If TRUE, data is loaded from files, where a pedigree information file is specified by pedigree and a probabilities output file cnF2freq.out in the working directory.
segregation	a vector specifying the segregation of founder alleles. (See Details.)
mc.size	an integer setting the sample size of the Monte Carlo sampling for one CPU.
hpc	logical. If TRUE, high performance computing is carried out by parallelization.
n.cpus	an integer telling the number of cores that parallelization is executed on. Only useful when hpc is TRUE. Not recommended when mc.size is small.

Details

To set up the segregation of the founder alleles, suppose that we have *m* founder(s) from one line and *n* from another line. Then in total there are *2 * (m + n)* alleles needed to be set up. According to the order in the pedigree information file, these alleles for the founders should be filled in segregation as a vector with length *2 * (m + n)*. For instance, if there is one male in one line and three females in another line, the vector might be created like c(1,1,2,2,3,3,2,3), which means that the male has two identical alleles whereas the genotypes of the females are (2,2), (3,3) and (2,3), respectively. If NULL, all the founder alleles are assumed to be different from each other.

Value

Output IBD matrices along the whole chromosome are saved in the working directory as files named by the loci with extension .ibd.

Author(s)

Xia Shen

References

Shen, X., Ronnegard, L. and Carlborg, O. (2009). *A Monte Carlo Full Likelihood Approach in Variance Component Quantitative Trait Loci Analysis*. (Submitted.)

See Also

MCIBD, MCIBD.epi2chro, MCIBD-package

MCIBD.epi2chro *Estimating Epistatic Identity-By-Descent (IBD) Matrices for Two Chromosomes*

Description

MCIBD.epi2chro estimates the epistatic IBD matrices for two given chromosomes. Monte Carlo sampling is used to approach to the matrix estimator, where the output file from cnF2freq is required. Segregation of the founder alleles can be set up in the estimation. Parallelzation is available using snowfall package if executing on a multi-core computer or cluster.

Usage

```
MCIBD.epi2chro(dis = 5, n.F2, pedigree = NULL, cnF2freq.out1 = NULL,
               cnF2freq.out2 = NULL, output.Z = FALSE, read.file = FALSE,
               segregation = NULL, mc.size = 99, hpc = FALSE, n.cpus = 2)
```

Arguments

dis	the distance between two nearest test loci in centimorgan.
n.F2	an integer telling the number of F2 individuals in the pedigree.
pedigree	a matrix or data.frame or file containing pedigree information. For file (when read.file is TRUE), see pedi.ric in other.zip or other.tar for example.
cnF2freq.out1	a matrix or data.frame containing the output probabilities of chromosome 1 from cnF2freq. For file (when read.file is TRUE), see chro.ric in other.zip or other.tar for example.
cnF2freq.out2	a matrix or data.frame containing the output probabilities of chromosome 2 from cnF2freq. For file (when read.file is TRUE), see chro.ric in other.zip or other.tar for example.
output.Z	logical. If TRUE, average incidence matrix Z is estimated instead of the corresponding IBD matrix.
read.file	logical. If TRUE, data is loaded from files, where a pedigree information file is specified by pedigree and a probabilities output file cnF2freq.out in the working directory.
segregation	a vector specifying the segregation of founder alleles. (See Details.)
mc.size	an integer setting the sample size of the Monte Carlo sampling for one CPU.
hpc	logical. If TRUE, high performance computing is carried out by parallelization.
n.cpus	an integer telling the number of cores that parallelization is executed on. Only useful when hpc is TRUE. Not recommended when mc.size is small.

Details

To set up the `segregation` of the founder alleles, suppose that we have *m* founder(s) from one line and *n* from another line. Then in total there are *2 * (m + n)* alleles needed to be set up. According to the order in the pedigree information file, these alleles for the founders should be filled in `segregation` as a vector with length *2 * (m + n)*. For instance, if there is one male in one line and three females in another line, the vector might be created like `c(1,1,2,2,3,3,2,3)`, which means that the male has two identical alleles whereas the genotypes of the females are (2,2), (3,3) and (2,3), respectively. If `NULL`, all the founder alleles are assumed to be different from each other.

Value

Output IBD matrices are saved in the working directory as files named by the loci with extension `.ibd`.

Author(s)

Xia Shen

References

Shen, X., Ronnegard, L. and Carlborg, O. (2009). *A Monte Carlo Full Likelihood Approach in Variance Component Quantitative Trait Loci Analysis*. (Submitted.)

See Also

`MCIBD`, `MCIBD.chro`, `MCIBD-package`

---

pedigree *A Pig Pedigree Information Dataset*

Description

(Andersson et al. 1994) An F2 cross has been bred from 2 European wild boars mated to 8 large white sows. Four F1 boars were mated to 22 F1 sows to produce 191 recorded F2 offspring in 26 families.

Usage

```
data(pedigree)
```

Format

A data frame with 227 observations on the following 2 variables.

`father` a numeric vector indicating fathers' ID

`mother` a numeric vector indicating mothers' ID

References

Andersson, L., Haley, C.S., Ellegren, H., Knott, S.A., Johansson, M., Andersson, K., Andersson-Eklund L., Edfors-Lilja, I., Fredholm, M., Hansson, I., Hakansson, J. and Lundstrom, K. (1994). *Genetic Mapping of Quantitative Trait Loci for Growth and Fatness in Pigs*. Science, 263: 1771-1774.

See Also

probabilities

Examples

```
data(pedigree)
```

`probabilities` *A Probabilities Dataset of Pig Chromosome 6 from cnF2freq*

Description

(Andersson et al. 1994) An F2 cross has been bred from 2 European wild boars mated to 8 large white sows. Four F1 boars were mated to 22 F1 sows to produce 191 recorded F2 offspring in 26 families. The genetic information on chromosome 6 came from 22 genotyped microsatellite markers at: 0.0, 8.6, 36.6, 49.7, 50.5, 62.9, 79.2, 80.4, 83.7, 84.1, 84.8, 90.6, 95.4, 100.7, 101.9, 115.9, 116.7, 119.0, 120.2, 124.0, 127.0 and 170.9 cM.

Usage

```
data(probabilities)
```

Format

A data frame with 38817 observations on 64 variables, where 38817 is the size of the pedigree 227 times the number of test loci 171 (every 1 cM per locus), and the 64 variables are the probabilities calculated from `cnF2freq`.

References

Andersson, L., Haley, C.S., Ellegren, H., Knott, S.A., Johansson, M., Andersson, K., Andersson-Eklund L., Edfors-Lilja, I., Fredholm, M., Hansson, I., Hakansson, J. and Lundstrom, K. (1994). *Genetic Mapping of Quantitative Trait Loci for Growth and Fatness in Pigs*. Science, 263: 1771-1774.

See Also

pedigree

Examples

```
data(probabilities)
```

Index