Package ‘LDAandLDAS’

October 14, 2022

Type Package

Title Linkage Disequilibrium of Ancestry (LDA) and LDA Score (LDAS)

Version 1.0.2

Maintainer Yaoling Yang <yaoling.yang@bristol.ac.uk>

Description Computation of linkage disequilibrium of ancestry (LDA) and linkage disequilibrium of ancestry score (LDAS). LDA calculates the pairwise linkage disequilibrium of ancestry between single nucleotide polymorphisms (SNPs). LDAS calculates the LDA score of SNPs. The methods are described in Barrie W, Yang Y, Attfield K E, et al (2022) <doi:10.1101/2022.09.23.509097>.

License GPL-3

Encoding UTF-8

LazyData true

Imports Rcpp (>= 1.0.8)

LinkingTo Rcpp

RoxygenNote 7.2.1

Depends R (>= 3.5.0)

NeedsCompilation yes

Author Yaoling Yang [aut, cre] (<https://orcid.org/0000-0003-4905-8097>), Daniel Lawson [aut] (<https://orcid.org/0000-0002-5311-6213>)

Repository CRAN

Date/Publication 2022-10-14 08:45:23 UTC

R topics documented:

`LDAandLDAS-package` *Linkage Disequilibrium of Ancestry (LDA) and LDA Score (LDAS)*

Description

This is the software for Linkage disequilibrium of ancestry (LDA) and LDA score (LDAS) which is proposed by the papar Genetic risk for Multiple Sclerosis originated in Pastoralist Steppe populations, Barrie W, Yang Y, Attfield K E, et al (2022).

LDA quantifies the correlations between the ancestry of two SNPs, measuring the proportion of individuals who have experienced a recombination leading to a change in ancestry, relative to the genome-wide baseline.

LDA score is the total amount of genome in LDA with each SNP (measured in recombination map distance), which is useful for detecting the signal of "recombinant favouring selection".

The codes for LDA and LDAS are hosted at `https://github.com/YaolingYang/LDAandLDAscore`.

Author(s)

Maintainer: Yaoling Yang <`yaoling.yang@bristol.ac.uk`> (ORCID)

Authors:

- Daniel Lawson <`Dan.Lawson@bristol.ac.uk`> (ORCID)

References

Barrie W, Yang Y, Attfield K E, et al. Genetic risk for Multiple Sclerosis originated in Pastoralist Steppe populations. bioRxiv (2022).

`cal_lda` *LDA of a pair of SNPs*

Description

Computation of the pairwise Linkage Disequilibrium of Ancestry (LDA) between a pair of single nucleotide polymorphisms (SNPs).

Usage

```
cal_lda(data_resample, data_base, data_experiment, n_ancestry)
```

Arguments

`data_resample`	a data frame of the first SNP's ancestry probabilities after resampling. Different ancestry probabilities are in different columns.
`data_base`	a data frame of the first SNP's ancestry probabilities. Different ancestry probabilities are in different columns.
`data_experiment`	a data frame of the second SNP's ancestry probabilities. Different ancestry probabilities are in different columns.
`n_ancestry`	a positive integer representing the number of different ancestries.

Details

This function computes the LDA between two SNPs. Resampling of one of the SNPs' painting data is required prior to implementing this function. To compute pairwise LDA between multiple pairs of SNPs, please use `LDA`.

Value

a numeric number representing the pairwise LDA of the two SNPs.

References

Barrie W, Yang Y, Attfield K E, et al. Genetic risk for Multiple Sclerosis originated in Pastoralist Steppe populations. bioRxiv (2022).

Examples

```
# compute the LDA between the 50th SNP and the 55th SNP
# painting data for the 50th SNP (2 ancestries)
data_base <- cbind(LDAandLDAS::example_painting_p1[,50],
                   LDAandLDAS::example_painting_p2[,50])

# painting data for the 55th SNP (2 ancestries)
data_experiment <- cbind(LDAandLDAS::example_painting_p1[,55],
                         LDAandLDAS::example_painting_p2[,55])

# resample painting data for the 50th SNP
data_resample <- data_base[sample(1:nrow(data_base)),]

#compute their pairwise LDA
LDA_value <- cal_lda(data_resample,data_base,data_experiment,2)
```

`example_map` *Example genetic maps*

Description

Example genetic maps including the physical position and the genetic distance of single nucleotide polymorphisms (SNPs).

Usage

```
data("example_map")
```

Format

A data frame with 2,000 haploid genomes of the physical position, named `SNP`, and the genetic distance of these SNP, named gd.

Examples

```
data(example_map)
```

`example_painting_p1` *Example painting data for population 1.*

Description

Example painting data (average ancestry probabilities) of a chromosome for population 1, including 2000 genomes (observations) and 200 single nucleotide polymorphisms (SNPs).

Usage

```
data("example_painting_p1")
```

Format

A data frame with 1500 haploid genomes (observations) on 200 SNPs (variables).

Examples

```
data(example_painting_p1)
```

`example_painting_p2` *Example painting data for population 2.*

Description

Example painting data (average ancestry probabilities) of a chromosome for population 2, including 2000 genomes (observations) and 200 single nucleotide polymorphisms (SNPs).

Usage

```
data("example_painting_p2")
```

Format

A data frame with 1500 haploid genomes (observations) on 200 SNPs (variables).

Examples

```
data(example_painting_p2)
```

`LDA` *LDA of all pairs of SNPs*

Description

Computation of the pairwise Linkage Disequilibrium of Ancestry (LDA) between all pairs of single nucleotide polymorphisms (SNPs).

Usage

```
LDA(paintings, SNPlimit = NULL, verbose = FALSE)
```

Arguments

`paintings`	a list of data frames of the N*k painting data (average ancestry probabilities) from different populations (N is the number of genomes, k is the number of SNPs).
`SNPlimit`	a positive integer representing the maximum number of SNPs at each side of a SNP that is used to calculate the pairwise LDA for the SNP. The value shouldn't be larger than the total number of SNPs. We may set a limit if the LDAs between SNPs far in distance are not to be investigated.
`verbose`	logical. If verbose=TRUE, print the process of calculating the pairwise LDA for the i-th SNP. By default, verbose=FALSE

Details

Linkage Disequilibrium of Ancestry (LDA) quantifies the correlations between the ancestry of two SNPs, measuring the proportion of individuals who have experienced a recombination leading to a change in ancestry, relative to the genome-wide baseline.

Value

a data frame of the pairwise LDA, with SNPs in the decreasing order of physical position on a chromosome.

References

Barrie W, Yang Y, Attfield K E, et al. Genetic risk for Multiple Sclerosis originated in Pastoralist Steppe populations. bioRxiv (2022).

Examples

```
# visualize the painting data
# Painting data are the average probabilities of different populations
head(LDAandLDAS::example_painting_p1[1:5,],10)

# combine the painting data for two populations as a list
# to make to input data for function 'LDA'.
paintings=list(LDAandLDAS::example_painting_p1,
          LDAandLDAS::example_painting_p2)

# calculate the pairwise LDA of SNPs
LDA_result <- LDA(paintings)
```

`LDAS` *LDA Score*

Description

Computation of the Linkage Disequilibrium of Ancestry Score (LDAS) of each single nucleotide polymorphism (SNP).

Usage

```
LDAS(LDA_data, map, window = 5, verbose = TRUE)
```

Arguments

`LDA_data`	a data frame of LDA between all pairs of SNPs that are within the 'window'. SNPs should be in the decreasing order of physical position on a chromosome. This is the output from `LDA`.
`map`	a data frame of the physical position and genetic distance of all the SNPs contained in 'LDA_data'. 'map' contains two columns. The first column is the physical distance (unit: b) of SNPs in the decreasing order. The second column is the genetic distance (unit: cM) of SNPs.
`window`	a positive number specifying the genetic distance that the LDA score of each SNP is computed within. By default, window=5.
`verbose`	logical. Print the process of calculating the LDA score for the i-th SNP.

Details

LDA score is the total amount of genome in LDA with each SNP (measured in recombination map distance). A low LDA score is the signal of "recombinant favouring selection".

Value

a data frame of the LDA score and its upper and lower bound at the physical position of each SNP.

References

Barrie W, Yang Y, Attfield K E, et al. Genetic risk for Multiple Sclerosis originated in Pastoralist Steppe populations. bioRxiv (2022).

Examples

```
# visualize the painting data
# Painting data are the average probabilities of different populations
head(LDAandLDAS::example_painting_p1[1:5,],10)

# combine the painting data for two ancestries as a list
# to make to input data for function 'LDA'.
paintings=list(LDAandLDAS::example_painting_p1,
          LDAandLDAS::example_painting_p2)

# calculate the pairwise LDA of SNPs
LDA_result <- LDA(paintings)

# map is the data containing two columns
# The first column is the physical position (unit: b) (decreasing order)
# The second column is the recombination distance (unit: cM) of the SNPs
head(LDAandLDAS::example_map,10)

# calculate the LDA score for the SNPs
LDA_score <- LDAS(LDA_result,LDAandLDAS::example_map,window=10)

#visualize the LDA scores
```

```
plot(x=LDA_score$SNP,y=LDA_score$LDAS)
```

Index