Package 'ngsAssocPoly'

March 12, 2020

Type Package

Title Association analysis using NGS-genotyping data in autopolyploids

Version 1.0.0

Author Eiji Yamamoto

Maintainer Eiji Yamamoto <yame.repos@gmail.com>

Description

Tools for genetic mapping in autopolyploids using low-coverage NGS genotyping data.
Allele dosage of genetic mapping populations are estimated using read count information in NGS genotype data.
Association between the estimated allele dosage and phenotypes are analyzed using generalized linear regression models (GLM).

Imports utils, graphics, grDevices, vcfR, qqman, RColorBrewer

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.0.2

R topics documented:

`alleleDosageEstimation`

Allele dosage estimation using NGS-based genotyping data

Description

Performs allele dosage estimation from NGS-based genotyping data.

Usage

```
alleleDosageEstimation(vcf.file.name,
                       ploidy = 6,
                       min.dp = 10,
                       max.dp = 1000,
                       max.miss = 0.5,
                       max.freq = 0.95,
                       round.up = 1,
                       cut.off = 0,
                       read.err.prob = 0.001)
```

Arguments

vcf.file.name	Character string of VCF file name.
ploidy	Ploidy level of species analyzed. Default is 6.
min.dp	Minimum read depth. If a genotype has a read depth less than min.dp, it is assigned NA. Default is 10.
max.dp	Maximum read depth. If a genotype has a read depth greater than max.dp, it is assigned NA. Default is 1000.
max.miss	Maximum proportion of missing genotype data. 0 allows sites that are completely missing. 1 indicates no missing data allowed. Default is 0.5.
max.freq	Maximum proportion of major genotype (dosage) frequency. 1 indicates monomorphic markers allowed. Default is 0.95.
round.up	In allele dosage estimation, a dosage greater than this value is assigned 1, and the others are assigned 0. Default is 1.
cut.off	In allele dosage estimation, dosages less than this value is assigned 0. Default is 0.
read.err.prob	This value is set not to handle unknown experimental errors that breaks calculations within this function. Default is 0.001.

Details

Read depth (DP field in VCF) and reference allele depth (RD field) are used as input values for simple binomial probability function to calculate prior probabilities of allele dosage. Then, relative probabilities for all possible allele dosage (i.e. 0, 1/N, 2/N, ..., N/N, where N is the ploidy level) are returned as the result.

Value

Returns R binary files for estimated allele dosage information (XXX_Geno.Rda) and the chromosomal location information (XXX_Map.Rda). The estimated allele dosage information is a list vector whose each element is n x p matrix of probabilities for allele dosages. n is number of individuals in the population and p is ploidy + 1. The chromosomal location information is a matrix whose row element is SNP marker and the column elements are SNP marker name, chromosome and position.

Author(s)

Eiji Yamamoto

alleleDosageGLM *Association analysis based on estimated allele dosage*

Description

Performs association analysis based on estimated allele dosage with generalized linear model (GLM).

Usage

```
alleleDosageGLM(Geno.name,
                Map.name,
                pheno.file.name,
                method = "dogmat")
```

Arguments

Geno.name	Character string of file name of estimated allele dosage that is created by function alleleDosageEstimation.
Map.name	Character string of file name of map information for Geno.name that is created by function alleleDosageEstimation.
pheno.file.name	Character string of CSV file name that include phenotype data. The first colmn must be the line (or individual) name that corresponds to Geno.name.
method	Specifies mode of input genotype. "dogmat" is for matricies of estimated allele dosage probabilities. "continuous" is for vectors of continuous allele dosage values. "diplodized" is for genotypes in GT field of VCF files.

Details

Association analysis based on estimated allele dosage. If the trait is binary, binomial function is used. Otherwise, gaussian is used.

Value

Returns a data csv file where the first three columns are the marker name, chromosome, and position, and subsequent columns are the marker scores (-log_10p) for the traits, and qq and Manhattan plots are generated in newly created directly.

Author(s)

Eiji Yamamoto

Index