Package 'predhy'

March 20, 2020

Type Package

Title Genomic Prediction of Hybrid Performance

Version 0.2.0

Author Yang Xu, Yanru Cui, Shizhong Xu, Chenwu Xu

Maintainer Yang Xu <xuyang_89@126.com>

Description Performs genomic prediction of hybrid performance with genomic best linear unbiased prediction (Xu S et al (2014) <doi:10.1073/pnas.1413750111>). The package also provides fast cross-validation and mating design for training population (Xu S et al (2016) <doi:10.1111/tpj.13242>; Xu S (2017) <doi:10.1534/g3.116.038059>).

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Depends R (>= 2.10)

NeedsCompilation no

Repository CRAN

Date/Publication 2020-03-20 11:30:02 UTC

R topics documented:

convertgen *Convert Genotype*

Description

Convert genotypes in HapMap format or in numeric format for hypred package.

Usage

```
convertgen(input_geno, type = c("hmp1", "hmp2", "num"),
  missingrate = 0.05, impute = TRUE)
```

Arguments

input_geno	genotype in HapMap format or in numeric format. The names of individuals should be provided. Missing (NA) values are allowed.
type	the type of genotype. There are three options: "hmp1" for genotypes in HapMap format with single bit, "hmp2" for genotypes in HapMap format with double bit, and "num" for genotypes in numeric format.
missingrate	max missing percentage for each SNP, default is 0.05.
impute	logical. If TRUE, imputation. Default is TRUE.

Value

A matrix of genotypes in numeric format, coded as 1, 0, -1 for AA, Aa, aa. Each row represents an individual and each column represents a marker. The rownames of the matrix are the names of individuals.

Examples

```
## load genotype in HapMap format with double bit
data(input_geno)

## convert genotype for hypred package
inbred_gen <- convertgen(input_geno, type = "hmp2")


## load genotype in numeric format
data(input_geno1)
head(input_geno1)

## convert genotype for hypred package
inbred_gen1 <- convertgen(input_geno1, type = "num")
```

`crodesign` *Generate Mating Design*

Description

Generate a mating design for a subset of crosses based on a balanced random partial rectangle cross-design (BRPRCD) (Xu et al. 2016).

Usage

```
crodesign(d, male_name, female_name, seed = 123)
```

Arguments

`d`	an integer denoting 1/d percentage of crosses to be evaluated in the field.
`male_name`	a character string for the names of male parents.
`female_name`	a character string for the names of male parents.
`seed`	the random number, default is 123.

Value

A data frame of mating design result with three columns. The first column is "crossID", the second column is the "male_Name" and the third column is the "female_Name".

References

Xu S, Xu Y, Gong L and Zhang Q. (2016) Metabolomic prediction of yield in hybrid rice. Plant J. 88, 219-227.

Examples

```
## generate a mating design with 100 male parents and 150 female parents
## for 1/d = 1/50 percentage of crosses to be evaluated in the field.
## the total number of potential crosses is 100 × 150 = 15000.
## The number of crosses to be field evaluated is 15000 × (1/50) = 300.

male_name <- paste("m", 1:100, sep = "")
female_name <- paste("f", 1:150, sep = "")
design <- crodesign(d = 50, male_name, female_name)
```

cv *Evaluate Trait Predictability via Cross Validation*

Description

Evaluate trait predictability of the GBLUP method via k-fold cross validation. For k-fold cross validation, the sample is randomly divided into k equal sized parts and each part is predicted once using parameters estimated based on the other k – 1 parts. The trait predictability is defined as the squared Pearson correlation coefficient between the observed and the predicted trait values.

Usage

```
cv(fix = NULL, y, kk, nfold = 5, seed = 123)
```

Arguments

fix	a design matrix of the fixed effects. If not passed, a vector of ones is added for the intercept.
y	a vector of the phenotypic values.
kk	a list of one or multiple kinship matrices.
nfold	the number of folds. Default is 5.
seed	the random number. Default is 123.

Value

Trait predictability

Examples

```
## load example data from hypred package
data(hybrid_phe)
data(input_geno)

## convert original genotype
inbred_gen <- convertgen(input_geno, type = "hmp2")

## infer the additive and dominance genotypes of hybrids
gena <- infergen(inbred_gen, hybrid_phe)$add
gend <- infergen(inbred_gen, hybrid_phe)$dom

## calculate the additive and dominance kinship matrix
ka <- kin(gena)
kd <- kin(gend)

## for the additive model
predictability <- cv(y = hybrid_phe[,3], kk = list(ka))
```

```
## for the additive-dominance model
predictability <- cv(y = hybrid_phe[,3], kk = list(ka,kd))
```

---

cv_fast *Evaluate Trait Predictability via the HAT Method*

Description

The HAT method is a fast algorithm for the ordinary cross validation. It is highly recommended for large dataset (Xu et al. 2017).

Usage

```
cv_fast(fix = NULL, y, kk, nfold = 5, seed = 123)
```

Arguments

fix	a design matrix of the fixed effects. If not passed, a vector of ones is added for the intercept.
y	a vector of the phenotypic values.
kk	a list of one or multiple kinship matrices.
nfold	the number of folds, default is 5. For the HAT Method, nfold can be set as the sample size (leave-one-out CV) to avoid variation caused by random partitioning of the samples, but it is not recommended for cv.
seed	the random number, default is 123.

Value

Trait predictability

References

Xu S. (2017) Predicted residual error sum of squares of mixed models: an application for genomic prediction. G3 (Bethesda) 7, 895-909.

Examples

```
## load example data from hypred package
data(hybrid_phe)
data(input_geno)

## convert original genotype
inbred_gen <- convertgen(input_geno, type = "hmp2")

## infer the additive and dominance genotypes of hybrids
gena <- infergen(inbred_gen, hybrid_phe)$add
```

```
gend <- infergen(inbred_gen, hybrid_phe)$dom

## calculate the additive and dominance kinship matrix
ka <- kin(gena)
kd <- kin(gend)

##for the additive model
predictability <- cv_fast(y = hybrid_phe[,3], kk = list(ka))

##for the additive-dominance model
predictability <- cv_fast(y = hybrid_phe[,3], kk = list(ka,kd))
```

`hybrid_phe` *Phenotypic data of hybrids*

Description

This dataset contains phenotypic data of 410 hybrids for grain yield in maize.

Usage

```
hybrid_phe
```

Format

A data frame with 410 rows and 3 variables:

`M` The names of male parents.

`F` The names of female parents.

`GY` The grain yield of hybrids.

`infergen` *Infer Genotype of Hybrids*

Description

Infer additive and dominance genotypes of hybrids based on their parental genotypes.

Usage

```
infergen(inbred_gen, hybrid_phe)
```

Arguments

inbred_gen	a matrix for genotypes of parental lines in numeric format, coded as 1, 0 and -1. The row.names of inbred_gen must be provied. It can be obtained from the original genotype using convertgen function.
hybrid_phe	a data frame with three columns. The first column and the second column are the names of male and female parents of the corresponding hybrids, respectively; the third column is the phenotypic values of hybrids. The names of male and female parents must match the rownames of inbred_gen. Missing (NA) values are not allowed.

Value

A list with following information is returned:

$add additive genotypes of hybrids

$dom dominance genotypes of hybrids

Examples

```
## load example data from hypred package
data(hybrid_phe)
head(hybrid_phe)
data(input_geno)

## convert original genotype
inbred_gen <- convertgen(input_geno, type = "hmp2")

gena <- infergen(inbred_gen, hybrid_phe)$add
gend <- infergen(inbred_gen, hybrid_phe)$dom
```

---

input_geno *Genotype in Hapmap Format*

Description

Genotypic data of 348 maize inbred lines in Hapmap format with double bit.

Usage

```
input_geno
```

Format

A data frame with 4979 rows and 359 columns.

input_geno1 *Genotype in Numeric Format*

Description

Genotypic data of 50 rice inbred lines with 1000 SNPs.

Usage

```
input_geno1
```

Format

A data frame with 1000 rows and 50 variables.

kin *Calculate Kinship Matrix*

Description

Calculate the additive and dominance kinship matrix.

Usage

```
kin(gen)
```

Arguments

gen	a matrix for genotypes, coded as 1, 0, -1 for AA, Aa, aa. Each row represents an individual and each column represents a marker.

Value

a kinship matrix

Examples

```
## random population with 100 lines and 1000 markers
gen <- matrix(rep(0,100*1000),100,1000)
gen <- apply(gen,2,function(x){x <- sample(c(-1,0,1), 100, replace = TRUE)})

## generate 100 × 100 kinship matrix
k <- kin(gen)
```

mixed *Solve Mixed Model*

Description

Solve linear mixed model using restricted maximum likelihood (REML). Multiple variance components can be estimated.

Usage

```
mixed(fix = NULL, y, kk)
```

Arguments

fix	a design matrix of the fixed effects. If not passed, a vector of ones is added for the intercept.
y	a vector of the phenotypic values.
kk	a list of one or multiple kinship matrices.

Value

A list with following information is returned:

$v_i the inverse of the phenotypic variance-covariance matrix

$var estimated variance components of genetic effects

$ve estimated residual variance

$beta estimated fixed effects

References

Xu S, Zhu D and Zhang Q. (2014) Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proc. Natl. Acad. Sci. USA 111, 12456-12461.

Examples

```
## load example data from hypred package
data(hybrid_phe)
data(input_geno)

## convert original genotype
inbred_gen <- convertgen(input_geno, type = "hmp2")

## infer the additive and dominance genotypes of hybrids
gena <- infergen(inbred_gen, hybrid_phe)$add
gend <- infergen(inbred_gen, hybrid_phe)$dom

## calculate the additive and dominance kinship matrix
```

```
ka <- kin(gena)
kd <- kin(gend)

## for the additive model
parm <- mixed(y = hybrid_phe[,3], kk = list(ka))

## for the additive-dominance model
parm <- mixed(y = hybrid_phe[,3], kk = list(ka, kd))
```

---

predhybrid *Predict the Performance of Hybrids*

Description

Predict all potential crosses of a given set of parents using a subset of crosses as the training sample.

Usage

```
predhybrid(inbred_gen, hybrid_phe, predparent_gen = inbred_gen,
  fix = NULL, fixnew = NULL, model = "AD", select = "all",
  number = NULL)
```

Arguments

inbred_gen	a matrix for genotypes of parental lines in numeric format, coded as 1, 0 and -1. The row.names of inbred_gen must be provied. It can be obtained from the original genotype using convertgen function.
hybrid_phe	a data frame with three columns. The first column and the second column are the names of male and female parents of the corresponding hybrids, respectively; the third column is the phenotypic values of hybrids. The names of male and female parents must match the rownames of inbred_gen. Missing (NA) values are not allowed.
predparent_gen	a matrix for genotypes of a given parental lines. All potential crosses derived from these parental lines will be predicted. Default is the same as inbred_gen.
fix	a design matrix of the fixed effects for the parental lines. If not passed, a vector of ones is added for the intercept.
fixnew	a design matrix of the fixed effects for all potential crosses. If not passed, a vector of ones is added for the intercept.
model	the model of prediction. There are two options: model = "AD" for the additive-dominance model, model = "A" for the additive model. Default is model = "AD".
select	the selection of hybrids based on the prediction results. There are three options: select = "all", which selects all potential crosses. select = "top", which selects the top n crosses. select = "top", which selects the bottom n crosses. The n is determined by the param number.

number — the number of selected top or bottom hybrids, only when select = "top" or select = "bottom".

Value

A data frame of prediction result with two columns. The first column denotes the names of male and female parents of the predicted hybrids, and the second column denotes the phenotypic values of the predicted hybrids.

Examples

```
## load example data from hypred package
data(hybrid_phe)
data(input_geno)
inbred_gen <- convertgen(input_geno, type = "hmp2")

## to save time, only predict 45 crosses derived from the first 10 lines of parental lines
## select all hybrids with additive-dominance model
pred1 <- predhybrid(inbred_gen, hybrid_phe, predparent_gen = inbred_gen[c(1:10),], model = "AD")

## select top 20 hybrids with additive model
pred2 <- predhybrid(inbred_gen, hybrid_phe, predparent_gen = inbred_gen[c(1:10),],
select = "top", number = 10, model = "A")
```

Index