

Package ‘SoyNAM’

November 18, 2015

Type Package

Title Soybean Nested Association Mapping Dataset

Version 1.2

Date 2015-11-16

Author

Alencar Xavier, William Beavis, James Specht, Brian Diers, Reka Howard, William Muir, Katy Rainey.

Maintainer Alencar Xavier <xaviera@purdue.edu>

Description Genomic and multi-environmental soybean data. Soybean Nested Association Mapping (SoyNAM) project dataset funded by the United Soybean Board (USB), pre-formatted for general analysis and genome-wide association analysis using the NAM package.

License GPL-3

Imports NAM, lme4, reshape2

Depends R (>= 3.0.2)

NeedsCompilation yes

Repository CRAN

Date/Publication 2015-11-18 09:31:14

R topics documented:

SoyNAM-package	2
BLUP	2
Dataset	4

Index	5
--------------	----------

SoyNAM-package

Soybean Nested Association Mapping Dataset

Description

Genomic and multi-environmental soybean data. Soybean Nested Association Mapping (SoyNAM) project dataset funded by the United Soybean Board (USB), pre-formatted for general analysis and genome-wide association analysis using the NAM package.

Details

Package: NAM
Type: Package
Version: 1.2
Date: 2015-11-16
License: GPL-3

Author(s)

Alencar Xavier, William Beavis, James Specht, Brian Diers, William Muir, Katy Rainey.

Maintainer: Alencar Xavier <xaviera@purdue.edu>

See Also

Functions: BLUP, data(soybase) and data(soynam).

BLUP

Best Linear Unbias Predictor

Description

Genetic values for a given trait computed by REML.

Usage

```
BLUP(trait="yield", family="all", env="all",  
      MAF=0.05, use.check=TRUE, impute="FM", rm.rep=TRUE)
```

Arguments

trait	Character. Trait of interest. The options are: "yield" (grain yield in Kg/ha), "maturity" (days to maturity), "height" (plant height in cm), "lodging" (lodging score from 1 to 5), "protein" (protein percentage in the grain), "oil" (oil percentage in the grain), "size" (seed size = mass of 100 seeds in grams) and "fiber" (fiber percentage in the grain).
family	Numeric vector or "all". Which SoyNAM families to use.
env	Numeric vector or "all". Which environments to use. The environments are coded as follows: 1 (IA_2012), 2 (IA_2013), 3 (IL_2011), 4 (IL_2012), 5 (IL_2013), 6 (IN_2012), 7 (IN_2013), 8 (KS_2012), 9 (KS_2013), 10 (MI_2012), 11 (MO_2012), 12 (MO_2013), 13 (NE_2011), 14 (NE_2012), 15 (OHmc_2012), 16 (OHmc_2013), 17 (OHmi_2012) and 18 (OHmi_2013).
MAF	Numeric. Minor allele frequency threshold for the markers.
use.check	Logical. If TRUE, it includes a control term as fixed effect in the model.
impute	NULL, 'RF' or 'FM'. If 'RF', it imputes missing genotypes using random forest. If 'FM' is imputes missing genotypes using a forward algorithm, filling missing loci with the most likely genotype based on the previous marker, as a Markov model.
rm.rep	Logical. If TRUE, it removes replicated genotypes. Genotypes are treated as identical when the genotypes are more than 95 percent identical. This argument requires imputed genotypes.

Details

This function uses the raw dataset, allowing user-defined data quality control for genotypes and BLUPs of genetic values.

The algorithm start from selecting the chosen families and environment that will be used for the best linear unbiased predictor (BLUP). The BLUP values are calculates based on the following model: ($Trait = Control + Environment + Genotype$). Where control is a covariate set as fixed effect based on the checks of each set (microenvironment); Environment is a random effect that represents the combination of location and year; and Genotype is the random effect associated to the lines. The BLUP values are the regression coefficients corresponding to the Genotype effect. The BLUP is calculated using the R package lme4 (Bates 2010) using REML.

If checks are used as covariate (use.check=TRUE), then the best linear unbiased estimator (BLUE) of the check effects is assigned to each set as a micro-environmental control. Each set had between one and five controls, including the SoyNAM parents and five other cultivars. These genotypes are normalized by environment and the BLUE of each set is calculated. All genotypes in a same set will have the same check effect.

Value

This function returns a list with four objects. A numeric vector with the BLUP solution of the phenotypes ("Phen"); the corresponding genotypes ("Gen"); a vector with the respective family ("Fam"); and a numeric vector with the number of SNPs per chromosome ("Chrom"). The output of this function has the exact input format for the NAM package to perform genome-wide association analysis.

Author(s)

Alencar Xavier

References

Bates, D. M. (2010). lme4: Mixed-effects modeling with R. URL <http://lme4.r-forge.r-project.org/book>.

Examples

```
Test=BLUP(trait="yield", family=2:5, env=2:7)
```

Dataset

Datasets

Description

Genotypes and phenotypes from quality assured (soybase) or original (soynam) datasets.

Usage

```
data(soybase)  
data(soynam)
```

Details

Datasets of the SoyNAM project, original and quality assured (QA) versions. Data was downloaded on November 16th 2015 from soynam.org.

Genotypic matrices are named "gen.raw" and "gen.qa" for the raw and QA versions, respectively. In each dataset, phenotypes are allocated into two objects, one with the lines ("data.line") and one with checks and parents ("data.checks"). Information on data objects include year, location, environment (combination of year and location), strain, family, set (set in each environment), spot (combination of set and environment), height (in centimeters), R8 (number of days to maturity), planting date (501 represents may 1), flowering (701 represents july 1), maturity (901 represents september 1), lodging (score from 1 to 5), yield (in Kg/ha), moisture, protein (percentage in the seed), oil (percentage in the seed), fiber (percentage in the seed), seed size (in grams of 100 seeds).

Author(s)

Alencar Xavier

Examples

```
data(soybase)  
data(soynam)
```

Index

`.Random.seed` (Dataset), 4

BLUP, 2

`data.check` (Dataset), 4

`data.line` (Dataset), 4

Dataset, 4

`gen.qa` (Dataset), 4

`gen.raw` (Dataset), 4

`soybase` (Dataset), 4

`soynam` (Dataset), 4

SoyNAM-package, 2