

# Package ‘StageWise’

August 3, 2021

**Title** Two-stage analysis of multi-environment trials for genomic selection

**Version** 0.04

**Author** Jeffrey B. Endelman

**Maintainer** Jeffrey Endelman <endelman@wisc.edu>

**Description** Two-stage analysis of multi-environment trials for genomic selection

**Depends** R (>= 4.0)

**License** GPL-3

**LazyData** true

**RoxygenNote** 7.1.1

**Encoding** UTF-8

**Imports** Matrix, ggplot2, methods, asreml, ggrepel, rlang

**Suggests** knitr, rmarkdown

**Collate** 'Stage1.R'

'Stage2.R'

'blup.R'

'blup\_prep.R'

'class\_geno.R'

'class\_prep.R'

'class\_var.R'

'gwas\_threshold.R'

'manhattan\_plot.R'

'private\_functions.R'

'quantile.geno.R'

'read\_geno.R'

'summary.var.R'

## R topics documented:

blup	2
blup_prep	3
class_geno-class	3
class_prep-class	4
class_var-class	4
gwas_threshold	5
manhattan_plot	5
quantile	6

read_genom . . . . .	6
Stage1 . . . . .	7
Stage2 . . . . .	8
summary . . . . .	9
<b>Index</b>	<b>11</b>

---

blup	<i>BLUP</i>
------	-------------

---

**Description**

BLUP

**Usage**

```
blup(data, geno = NULL, what, index.coeff = NULL, gwas.ncore = 0L)
```

**Arguments**

data	object of <code>class_prep</code> from <code>blup_prep</code>
geno	object of <code>class_genom</code> from <code>read_genom</code>
what	"id" or "marker"
index.coeff	index coefficients for the locations or traits
gwas.ncore	Integer indicating number of cores to use for GWAS (default is 0 for no GWAS). Requires <code>what="markers"</code> .

**Details**

Argument `what="id"` leads to prediction of breeding values (BV) and genotypic values (GV), including the average fixed effect of the environments and any fixed effect markers. For argument `what="marker"`, fixed effects are not included in the BLUP. Argument `index.coeff` should be a named vector, matching the names of the locations or traits. Index coefficients are assumed to imply relative weights for the different locations or traits; as such, they are divided by the square root of the genetic variance estimate for that location/trait and then rescaled to have unit sum.

**Value**

Data frames of BLUPs

---

blup_prep	<i>Prepare data for BLUP</i>
-----------	------------------------------

---

**Description**

Prepare data for BLUP

**Usage**

```
blup_prep(data, vcov = NULL, geno = NULL, vars, mask = NULL)
```

**Arguments**

data	data frame of BLUEs from Stage 1
vcov	list of variance-covariance matrices for the BLUEs
geno	object of <code>class_geno</code> from <code>read_geno</code>
vars	object of <code>class_var</code> from <code>Stage2</code>
mask	(optional) data frame with column "id" and optional columns "env", "trait"

**Details**

The argument mask can be used to mask all observations for a particular individual, or by including a column named 'env', only the observations in particular environments can be masked. This is useful for cross-validation to test the accuracy of predicting into new environments.

**Value**

Object of `class_prep`

---

class_geno-class	<i>S4 class for marker genotype data</i>
------------------	--

---

**Description**

S4 class for marker genotype data

**Slots**

map	Marker map positions
coeff	Coefficients of the marker effects (dim: indiv x marker)
scale	Scaling factor between markers and indiv
G	Additive (genomic) relationship matrix
eigen.G	list of eigenvalues and eigenvectors

---

class_prep-class	<i>S4 class to prepare for blup</i>
------------------	-------------------------------------

---

### Description

S4 class to prepare for blup

### Slots

y Stage 1 BLUEs  
 Z Incidence matrix for random effects  
 id genotype identifiers  
 var.u variance of random effects  
 Vinv inverted covariance matrix of the Stage 1 BLUEs  
 Pmat P matrix from Searle  
 fixed fixed effect estimates  
 random random effect estimates  
 add logical whether additive effects predicted  
 loc.env data frame with loc, env  
 index.scale sqrt of genetic variances for the locations/traits  
 fixed.marker names of fixed effect markers

---

class_var-class	<i>S4 class for variances</i>
-----------------	-------------------------------

---

### Description

S4 class for variances

### Slots

add additive  
 g.resid genetic residual  
 resid residual  
 meanG mean of diagonal of G  
 meanOmega mean of diagonal of Omega  
 fixed.marker.var variance of marker fixed effects  
 fixed.marker.cov contribution of marker fixed effects to additive covariance between locations

---

gwas_threshold	<i>Compute GWAS discovery threshold</i>
----------------	---

---

**Description**

Compute GWAS discovery threshold

**Usage**

```
gwas_threshold(geno, n.core = 1, alpha = 0.05)
```

**Arguments**

geno	object of <code>class_geno</code>
n.core	number of cores to use
alpha	genome-wide significance level

**Details**

Uses a Bonferroni-type correction based on an effective number of markers that accounts for LD (Moskvina and Schmidt, 2008).

**Value**

$-\log_{10}(p)$  threshold

**References**

Moskvina V, Schmidt KM (2008) On multiple-testing correction in genome-wide association studies. Genetic Epidemiology 32:567-573. doi:10.1002/gepi.20331

---

manhattan_plot	<i>Create Manhattan plot</i>
----------------	------------------------------

---

**Description**

Create Manhattan plot

**Usage**

```
manhattan_plot(data, chrom = NULL, thresh = NULL, rotate = FALSE)
```

**Arguments**

data	data frame with columns for marker, chrom, position, and gwas.score
chrom	optional, to plot only one chromosome
thresh	optional, to include horizontal line at discovery threshold
rotate	TRUE/FALSE whether to rotate x-axis labels to be perpendicular

**Details**

Assumes position in bp

**Value**

ggplot2 object

---

quantile	<i>G matrix quantile</i>
----------	--------------------------

---

**Description**

G matrix quantile

**Arguments**

x	object of <code>class_genotype</code>
prob	probability

**Details**

Unlike the S3 method, prob must have length = 1

**Value**

data frame with the quantile of the G matrix coefficients for each id

---

read_genotype	<i>Read marker genotype data</i>
---------------	----------------------------------

---

**Description**

Read marker genotype data

**Usage**

```
read_genotype(filename, ploidy, map, eigen.tol = 1e-09)
```

**Arguments**

filename	Name of CSV file
ploidy	2,4,6,etc. (even numbers)
map	TRUE/FALSE
eigen.tol	1e-9

## Details

When `map=TRUE`, first three columns of the file are marker, chrom, position. When `map=FALSE`, the first column is marker. Subsequent columns contain the allele dosage for individuals/clones, coded 0,1,2,...ploidy (fractional values are allowed). Additive coefficients are computed by subtracting the population mean from each marker, and the additive (genomic) relationship matrix is computed as  $G = \text{tcrossprod}(\text{coeff})/\text{scale}$ . The scale parameter ensures the mean of the diagonal elements of  $G$  equals 1 under panmictic equilibrium. Missing genotype data is replaced with the population mean. For numerical conditioning, eigenvalues of  $G$  smaller than `eigen.tol` are replaced by `eigen.tol`.

## Value

Variable of class `class_geno`.

---

Stage1

*Stage 1 analysis of multi-environment trials*

---

## Description

Computes genotype BLUEs within each environment using ASReml-R

## Usage

```
Stage1(
  filename,
  traits,
  effects = NULL,
  silent = TRUE,
  workspace = "500mb",
  pworkspace = "500mb"
)
```

## Arguments

<code>filename</code>	Name of CSV file
<code>traits</code>	trait names (see Details)
<code>effects</code>	data frame specifying other effects in the model (see Details)
<code>silent</code>	TRUE/FALSE, whether to suppress ASReml-R output
<code>workspace</code>	memory limit for ASReml-R variance estimation
<code>pworkspace</code>	memory limit for ASReml-R BLUE computation

## Details

The input file must have one column labeled "id" for the individuals and one labeled "env" for the environments. The data for each environment are analyzed independently with a linear mixed model. Although not used in Stage1, to include a genotype x location effect in [Stage2](#), a column labeled "loc" should be present in the input file.

Including multiple traits in `trait` triggers a multivariate analysis, but for computational reasons, only 2 traits are analyzed at a time. With more than 2 traits, the software analyzes all pairs of traits. For single trait analysis, broad-sense  $H^2$  on a plot basis is computed from the variance components

for each env, with genotype as a random effect. The residuals from this analysis are also returned as a table and plots.

Argument `effects` is used to specify other i.i.d. effects besides genotype and has three columns: name, fixed, factor. The "name" column is a string that must match a column in the input file. The fixed column is a logical variable to indicate whether the effect is fixed (TRUE) or random (FALSE). The factor column is a logical variable to indicate whether the effect is a factor (TRUE) or numeric (FALSE).

Missing response values are omitted for single-trait analysis but retained for multi-trait analysis (unless both traits are missing), to allow for prediction in Stage 2. By default, the workspace and pworkspace limits for ASReml-R are set at 500mb. If you get an error about insufficient memory, try increasing the appropriate value (workspace for variance estimation and pworkspace for BLUE computation).

### Value

List containing

**blue** data frame of BLUEs for all environments

**vcov** list of variance-covariance matrices for the BLUEs, one per env

**H2** broad-sense H2 on a plot basis (only for single trait)

**resid** list containing boxplot, qqplot, and table of residuals (only for single trait)

---

Stage2

*Stage 2 analysis of multi-environment trials*

---

### Description

Stage 2 analysis of multi-environment trials

### Usage

```
Stage2(
  data,
  vcov = NULL,
  geno = NULL,
  fix.eff.marker = NULL,
  silent = TRUE,
  workspace = "500mb"
)
```

### Arguments

<code>data</code>	data frame of BLUEs from Stage 1 (see Details)
<code>vcov</code>	list of variance-covariance matrices for the BLUEs
<code>geno</code>	output from <a href="#">read_geno</a>
<code>fix.eff.marker</code>	markers in geno to include as additive fixed effect covariates
<code>silent</code>	TRUE/FALSE, whether to suppress ASReml-R output
<code>workspace</code>	Memory limit for ASReml-R variance estimation



## Details

Stage 2 of the two-stage approach described by Damesa et al. 2017, using ASReml-R for variance component estimation. The variable data has three mandatory column: id, env, BLUE. Optionally, data can have a column labeled "loc", which changes the main effect for genotype into a separable genotype-within-location effect, using a FA2 covariance model for the locations. Optionally, data can have a column labeled "trait", which introduces an unstructured covariance model for the traits. The multi-location and multi-trait analyses cannot be combined. Missing data are allowed in the multi-trait but not the single-trait analysis. The argument geno is used to partition genetic values into additive and non-additive (g.resid) components. Any individuals in data that are not present in geno are discarded.

The argument vcov is used to partition the macro- and micro-environmental variation, which are called GxE and residual in the output. vcov is a named list of variance-covariance matrices for the BLUES within each environment, with id on the rownames. The order in vcov and data should match. Both data and vcov can be created using the function [Stage1](#).

Because ASReml-R can only use relationship matrices defined in the global environment, this function creates and then removes global variables when either vcov or geno is used. By default, the workspace memory for ASReml-R is set at 500mb. If you get an error about insufficient memory, try increasing it. ASReml-R version 4.1.0.148 or later is required.

## Value

List containing

**aic** AIC

**vars** variances, as variable of class [class\\_var](#)

**fixed** Fixed effect estimates for env and markers

**random** Random effect predictions

**uniplot** uniplot of the genetic correlation between locations

## References

Damesa et al. 2017. Agronomy Journal 109: 845-857. doi:10.2134/agronj2016.07.0395

---

summary	<i>Summarize variances and correlations</i>
---------	---

---

## Description

Summarize variances and correlations

## Arguments

object                      object of [class\\_var](#)

## Details

When reporting the partitioning of variance, the variance component for the additive effect is multiplied by the mean diagonal of the G matrix.

**Value**

For univariate analysis, a matrix of variances. For multi-location or multi-trait analysis, a list containing the variance matrix and the correlation matrix.

# Index

blup, [2](#)  
blup\_prep, [2](#), [3](#)  
  
class\_geno, [2](#), [3](#), [5–7](#)  
class\_geno (class\_geno-class), [3](#)  
class\_geno-class, [3](#)  
class\_prep, [2](#), [3](#)  
class\_prep (class\_prep-class), [4](#)  
class\_prep-class, [4](#)  
class\_var, [3](#), [9](#)  
class\_var (class\_var-class), [4](#)  
class\_var-class, [4](#)  
  
gwas\_threshold, [5](#)  
  
manhattan\_plot, [5](#)  
  
quantile, [6](#)  
  
read\_geno, [2](#), [3](#), [6](#), [8](#)  
  
Stage1, [7](#), [9](#)  
Stage2, [3](#), [7](#), [8](#)  
summary, [9](#)