Package 'EvaluateCore'

June 20, 2022

Title Quality Evaluation of Core Collections

Version 0.1.2.9000

Description Implements various quality evaluation statistics to assess the
value of plant germplasm core collections using qualitative and
quantitative phenotypic trait data according to Odong et al. (2015)
<doi:10.1007/s00122-012-1971-y>.

Copyright 2018-2022, ICAR-NBPGR

License GPL-2 | GPL-3

Encoding UTF-8

RoxygenNote 7.2.0

URL https://github.com/aravind-j/EvaluateCore

https://CRAN.R-project.org/package=EvaluateCore

https://aravind-j.github.io/EvaluateCore/

https://doi.org/10.5281/zenodo.3875930

BugReports https://github.com/aravind-j/EvaluateCore/issues

RdMacros mathjaxr,
Rdpack

Depends R (>= 3.5.0)

Imports agricolae,
car,
cluster,
dplyr,
entropy,
ggcorrplot,
ggplot2,
grDevices,
gridExtra,
kSamples,
mathjaxr,
psych,
reshape2,
Rdpack,
stats,
vegan

1

Suggests corehunter,
      pander,
      rJava (>= 0.9-8)

LazyData true

R topics documented:

---

bar.evaluate.core   *Bar Plots*

Description

Plot Bar plots to graphically compare the frequency distributions of qualitative traits between entire collection (EC) and core set (CS).

Usage

```
bar.evaluate.core(data, names, qualitative, selected)
```

Arguments

data   The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.

names   Name of column with the individual names as a character string

qualitative   Name of columns with the qualitative traits as a character vector.

selected   Character vector with the names of individuals selected in core collection and present in the names column.

Value

A list with the ggplot objects of relative frequency bar plots of CS and EC for each trait specified as qualitative.

See Also

barplot, geom_bar

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

bar.evaluate.core(data = ec, names = "genotypes",
                  qualitative = qual, selected = core)
```

---

box.evaluate.core   *Box Plots*

Description

Plot Box-and-Whisker plots (Tukey 1970; McGill et al. 1978) to graphically compare the probability distributions of quantitative traits between entire collection (EC) and core set (CS).

Usage

```
box.evaluate.core(data, names, quantitative, selected)
```

Arguments

data   The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.

`names`	Name of column with the individual names as a character string
`quantitative`	Name of columns with the quantitative traits as a character vector.
`selected`	Character vector with the names of individuals selected in core collection and present in the `names` column.

Value

A list with the `ggplot` objects of box plots of CS and EC for each trait specified as `quantitative`.

References

McGill R, Tukey JW, Larsen WA (1978). "Variations of box plots." *The American Statistician*, **32**(1), 12.

Tukey JW (1970). *Exploratory Data Analysis. Preliminary edition.* Addison-Wesley.

See Also

boxplot, geom_boxplot

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

box.evaluate.core(data = ec, names = "genotypes",
                  quantitative = quant, selected = core)
```

---

cassava_CC   *IITA Cassava Germplasm Data - Core Collection*

Description

An example germplasm characterisation data of a core collection generated from 1591 accessions of IITA Cassava collection (International Institute of Tropical Agriculture et al. 2019) using 10 quantitative and 48 qualitative trait data with CoreHunter3 (corehunter). The core set was generated using distance based measures giving equal weightage to Average entry-to-nearest-entry distance (EN) and Average accession-to-nearest-entry distance (AN). Includes data on 26 descriptors for 168 (10 % of cassava_EC) accessions. It is used to demonstrate the various functions of EvaluateCore package.

Usage

```
cassava_CC
```

Format

A data frame with 58 columns:

CUAL Colour of unexpanded apical leaves

LNGS Length of stipules

PTLC Petiole colour

DSTA Distribution of anthocyanin

LFRT Leaf retention

LBTEF Level of branching at the end of flowering

CBTR Colour of boiled tuberous root

NMLB Number of levels of branching

ANGB Angle of branching

CUAL9M Colours of unexpanded apical leaves at 9 months

LVC9M Leaf vein colour at 9 months

TNPR9M Total number of plants remaining per accession at 9 months

PL9M Petiole length at 9 months

STRP Storage root peduncle

STRC Storage root constrictions

PSTR Position of root

NMSR Number of storage root per plant

TTRN Total root number per plant

TFWSR Total fresh weight of storage root per plant

TTRW Total root weight per plant

TFWSS Total fresh weight of storage shoot per plant

TTSW Total shoot weight per plant

TTPW Total plant weight

AVPW Average plant weight

ARSR Amount of rotted storage root per plant

SRDM Storage root dry matter

Details

Further details on how the example dataset was built from the original data is available online.

References

International Institute of Tropical Agriculture, Benjamin F, Marimagne T (2019). "Cassava mor-
phological characterization. Version 2018.1." *www.genesys-pgr.org*.

Examples

```
data(cassava_CC)
summary(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

lapply(seq_along(cassava_CC[, qual]),
       function(i) barplot(table(cassava_CC[, qual][, i]),
                           xlab = names(cassava_CC[, qual])[i]))

lapply(seq_along(cassava_CC[, quant]),
       function(i) hist(table(cassava_CC[, quant][, i]),
                        xlab = names(cassava_CC[, quant])[i],
                        main = ""))
```

---

cassava_EC   *IITA Cassava Germplasm Data - Entire Collection*

Description

An example germplasm characterisation data of a subset of IITA Cassava collection (International
Institute of Tropical Agriculture et al. 2019). Includes data on 26 (out of 62) descriptors for 1684
(out of 2170) accessions. It is used to demonstrate the various functions of EvaluateCore package.

Usage

```
cassava_EC
```

Format

A data frame with 58 columns:

CUAL Colour of unexpanded apical leaves

LNGS Length of stipules

PTLC Petiole colour

DSTA Distribution of anthocyanin

LFRT Leaf retention

LBTEF Level of branching at the end of flowering

CBTR Colour of boiled tuberous root

NMLB Number of levels of branching

ANGB Angle of branching

CUAL9M Colours of unexpanded apical leaves at 9 months

LVC9M Leaf vein colour at 9 months

TNPR9M Total number of plants remaining per accession at 9 months

PL9M Petiole length at 9 months

STRP Storage root peduncle

STRC Storage root constrictions

PSTR Position of root

NMSR Number of storage root per plant

TTRN Total root number per plant

TFWSR Total fresh weight of storage root per plant

TTRW Total root weight per plant

TFWSS Total fresh weight of storage shoot per plant

TTSW Total shoot weight per plant

TTPW Total plant weight

AVPW Average plant weight

ARSR Amount of rotted storage root per plant

SRDM Storage root dry matter

Details

Further details on how the example dataset was built from the original data is available online.

References

International Institute of Tropical Agriculture, Benjamin F, Marimagne T (2019). "Cassava morphological characterization. Version 2018.1." *www.genesys-pgr.org*.

Examples

```
data(cassava_EC)
summary(cassava_EC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

lapply(seq_along(cassava_EC[, qual]),
       function(i) barplot(table(cassava_EC[, qual][, i]),
                           xlab = names(cassava_EC[, qual])[i]))

lapply(seq_along(cassava_EC[, quant]),
       function(i) hist(table(cassava_EC[, quant][, i]),
                        xlab = names(cassava_EC[, quant])[i],
                        main = ""))
```

---

chisquare.evaluate.core

Chi-squared Test for Homogeneity

Description

Compare the distribution frequencies of qualitative traits between entire collection (EC) and core set (CS) by Chi-squared test for homogeneity (Pearson 1900; Snedecor and Irwin 1933).

Usage

```
chisquare.evaluate.core(data, names, qualitative, selected)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string
qualitative	Name of columns with the qualitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the names column.

Value

A a data frame with the following columns.

Trait	The qualitative trait.
EC_No.Classes	The number of classes in the trait for EC.
EC_Classes	The frequency of the classes in the trait for EC.
CS_No.Classes	The number of classes in the trait for CS.
CS_Classes	The frequency of the classes in the trait for CS.
chisq_statistic	
	The χ^2 test statistic.
chisq_pvalue	The p value for the test statistic.
chisq_significance	
	The significance of the test statistic (*: $p \leq 0.01$; **: $p \leq 0.05$; ns: $p > 0.05$).

References

Pearson K (1900). "X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **50**(302), 157–175.

Snedecor G, Irwin MR (1933). "On the chi-square test for homogeneity." *Iowa State College Journal of Science*, **8**, 75–81.

See Also

chisq.test

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

chisquare.evaluate.core(data = ec, names = "genotypes",
                        qualitative = qual, selected = core)
```

---

corr.evaluate.core *Phenotypic Correlations*

Description

Compute phenotypic correlations (Pearson 1895) between traits, plot correlation matrices as correlograms (Friendly 2002) and calculate mantel correlation (Legendre and Legendre 2012) between them to compare entire collection (EC) and core set (CS).

Usage

```
corr.evaluate.core(data, names, quantitative, qualitative, selected)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string
quantitative	Name of columns with the quantitative traits as a character vector.
qualitative	Name of columns with the qualitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the names column.

Value

A list with the following components.

Correlation Matrix

> The matrix with phenotypic correlations between traits in EC (below diagonal) and CS (above diagonal).

Correlogram   A correlogram of phenotypic correlations between traits in EC (below diagonal) and CS (above diagonal) as a `ggplot` object.

Mantel Correlation

> A data frame with Mantel correlation coefficient (r) between EC and CS phenotypic correlation matrices, it's p value and significance (*: p $\leq$ 0.01; **: p $\leq$ 0.05; ns: p $>$ 0.05).

References

Friendly M (2002). "Corrgrams." *The American Statistician*, **56**(4), 316–324.

Legendre P, Legendre L (2012). "Interpretation of ecological structures." In *Developments in Environmental Modelling*, volume 24, 521–624. Elsevier.

Pearson K (1895). "Note on regression and inheritance in the case of two parents." *Proceedings of the Royal Society of London*, **58**, 240–242.

See Also

cor, cor_pmat ggcorrplot, mantel

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

corr.evaluate.core(data = ec, names = "genotypes", quantitative = quant,
                   qualitative = qual, selected = core)
```

---

coverage.evaluate.core

Class Coverage

Description

Compute the Class Coverage (Kim et al. 2007) to compare the distribution frequencies of qualitative traits between entire collection (EC) and core set (CS).

Usage

```
coverage.evaluate.core(data, names, qualitative, selected)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string
qualitative	Name of columns with the qualitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the names column.

Details

Class Coverage (Kim et al. 2007) is computed as follows.

$$Class\,Coverage = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{A_{CS_i}}{A_{EC_i}} \right) \times 100$$

Where, A_{CS_i} is the sets of categories in the CS for the ith trait, A_{EC_i} is the sets of categories in the EC for the ith trait and n is the total number of traits.

Value

The Class Coverage value.

References

Kim K, Chung H, Cho G, Ma K, Chandrabalan D, Gwag J, Kim T, Cho E, Park Y (2007). "PowerCore: A program applying the advanced M strategy with a heuristic search for establishing core sets." *Bioinformatics*, **23**(16), 2155–2162.

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL
```

```
core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

coverage.evaluate.core(data = ec, names = "genotypes",
                       qualitative = qual, selected = core)
```

---

cr.evaluate.core   *Coincidence Rate of Range*

Description

Compute the Coincidence Rate of Range (CR) (Hu et al. 2000) (originally described by (Diwan et al. 1995) as Mean range ratio) to compare quantitative traits of the entire collection (EC) and core set (CS).

Usage

```
cr.evaluate.core(data, names, quantitative, selected)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string
quantitative	Name of columns with the quantitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the `names` column.

Details

The Coincidence Rate of Range (CR) is computed as follows.

$$CR = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{R_{CS_i}}{R_{EC_i}} \right) \times 100$$

Where, R_{CS_i} is the range of the ith trait in the CS, R_{EC_i} is the range of the ith trait in the EC and n is the total number of traits.

A representative CS should have a CR value no less than 70% (Diwan et al. 1995) or 80% (Hu et al. 2000).

Value

The CR value.

References

Diwan N, McIntosh MS, Bauchan GR (1995). "Methods of developing a core collection of annual *Medicago* species." *Theoretical and Applied Genetics*, **90**(6), 755–761.

Hu J, Zhu J, Xu HM (2000). "Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops." *Theoretical and Applied Genetics*, **101**(1), 264–268.

See Also

wilcox.test

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

cr.evaluate.core(data = ec, names = "genotypes",
                 quantitative = quant, selected = core)
```

---

dist.evaluate.core   *Distance Measures*

Description

Compute average Entry-to-nearest-entry distance ($E\text{-}EN$), Accession-to-nearest-entry distance ($A\text{-}EN$) and Entry-to-entry distance ($E\text{-}E$) (Odong et al. 2013) to evaluate a core set (CS) selected from an entire collection (EC).

Usage

```
dist.evaluate.core(data, names, quantitative, qualitative, selected, d = NULL)
```

Arguments

`data`	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
`names`	Name of column with the individual names as a character string
`quantitative`	Name of columns with the quantitative traits as a character vector.
`qualitative`	Name of columns with the qualitative traits as a character vector.
`selected`	Character vector with the names of individuals selected in core collection and present in the `names` column.
`d`	A distance matrix of class `"dist"` with individual names in the `names` column in data as labels. If `NULL` (default), then a distance matrix is computed using Gower's metric. (Gower 1971).

Value

A data frame with the average values of *E-EN*, *A-EN* and *E-E*.

References

Gower JC (1971). "A general coefficient of similarity and some of its properties." *Biometrics*, **27**(4), 857–871.

Odong TL, Jansen J, van Eeuwijk FA, van Hintum TJL (2013). "Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation." *Theoretical and Applied Genetics*, **126**(2), 289–305.

See Also

[evaluateCore](#)

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

dist.evaluate.core(data = ec, names = "genotypes", quantitative = quant,
                   qualitative = qual, selected = core)
```

```
###################################
# Compare with corehunter
###################################

library(corehunter)
# Prepare phenotype dataset
dtype <- c(rep("RD", length(quant)),
           rep("NS", length(qual)))
rownames(ec) <- ec[, "genotypes"]
ecdata <- corehunter::phenotypes(data = ec[, c(quant, qual)],
                                 types = dtype)

# Compute average distances
EN <- evaluateCore(core = rownames(cassava_CC), data = ecdata,
                   objective = objective("EN", "GD"))
AN <- evaluateCore(core = rownames(cassava_CC), data = ecdata,
                   objective = objective("AN", "GD"))
EE <- evaluateCore(core = rownames(cassava_CC), data = ecdata,
                   objective = objective("EE", "GD"))
EN
AN
EE
```

---

freqdist.evaluate.core

Frequency Distribution Histogram

Description

Plot stacked frequency distribution histogram to graphically compare the probability distributions of traits between entire collection (EC) and core set (CS).

Usage

```
freqdist.evaluate.core(
  data,
  names,
  quantitative,
  qualitative,
  selected,
  highlight = NULL,
  include.highlight = TRUE,
  highlight.se = NULL,
  highlight.col = "red"
)
```

Arguments

data   The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.

names	Name of column with the individual names as a character string
quantitative	Name of columns with the quantitative traits as a character vector.
qualitative	Name of columns with the qualitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the names column.
highlight	Individual names to be highlighted as a character vector.
include.highlight	
	If TRUE, the highlighted individuals are included in the frequency distribution histogram. Default is TRUE.
highlight.se	Optional data frame of standard errors for the individuals specified in highlight. It should have the same column names as in data.
highlight.col	The colour(s) to be used to highlighting individuals in the plot as a character vector of the same length as highlight. Must be valid colour values in R (named colours, hexadecimal representation, index of colours [1:8] in default R palette() etc.).

Value

A list with the ggplot objects of stacked frequency distribution histograms plots for each trait specified as quantitative and qualitative.

See Also

hist, geom_histogram

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))


freqdist.evaluate.core(data = ec, names = "genotypes",
                       quantitative = quant, qualitative = qual,
                       selected = core)

checks <- c("TMe-1199", "TMe-1957", "TMe-3596", "TMe-3392")

freqdist.evaluate.core(data = ec, names = "genotypes",
```

```
                                  quantitative = quant, qualitative = qual,
                                  selected = core,
                                  highlight = checks, highlight.col = "red")

quant.se <- data.frame(genotypes = checks,
                       NMSR = c(0.107, 0.099, 0.106, 0.062),
                       TTRN = c(0.081, 0.072, 0.057, 0.049),
                       TFWSR = c(0.089, 0.031, 0.092, 0.097),
                       TTRW = c(0.064, 0.031, 0.071, 0.071),
                       TFWSS = c(0.106, 0.071, 0.121, 0.066),
                       TTSW = c(0.084, 0.045, 0.066, 0.054),
                       TTPW = c(0.098, 0.052, 0.111, 0.082),
                       AVPW = c(0.074, 0.038, 0.054, 0.061),
                       ARSR = c(0.104, 0.019, 0.204, 0.044),
                       SRDM = c(0.078, 0.138, 0.076, 0.079))

freqdist.evaluate.core(data = ec, names = "genotypes",
                       quantitative = quant,
                       selected = core,
                       highlight = checks, highlight.col = "red",
                       highlight.se = quant.se)
```

---

iqr.evaluate.core	*Interquartile Range*

Description

Compute the Interquartile Range (IQR) (Upton and Cook 1996) to compare quantitative traits of the entire collection (EC) and core set (CS).

Usage

```
iqr.evaluate.core(data, names, quantitative, selected)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string
quantitative	Name of columns with the quantitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the names column.

Value

A data frame with the IQR values of the EC and CS for the traits specified as quantitative.

References

Upton G, Cook I (1996). "General summary statistics." In *Understanding statistics*. Oxford University Press.

See Also

> [IQR](#)

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

iqr.evaluate.core(data = ec, names = "genotypes",
                  quantitative = quant, selected = core)
```

---

levene.evaluate.core   *Levene's Test*

Description

Test for of variances of the entire collection (EC) and core set (CS) for quantitative traits by Levene's test (Levene 1960).

Usage

```
levene.evaluate.core(data, names, quantitative, selected)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string
quantitative	Name of columns with the quantitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the names column.

Value

A data frame with the following columns

`Trait`	The quantitative trait.
`EC_V`	The variance of the EC.
`CS_V`	The variance of the CS.
`EC_CV`	The coefficient of variance of the EC.
`CS_CV`	The coefficient of variance of the CS.
`Levene_Fvalue`	The test statistic.
`Levene_pvalue`	The p value for the test statistic.
`Levene_significance`	
	The significance of the test statistic (*: $p \leq 0.01$; **: $p \leq 0.05$; ns: $p > 0.05$).

References

Levene H (1960). "Robust tests for equality of variances." In Olkin I, Ghurye SG, Hoeffding W, Madow WG, Mann HB (eds.), *Contribution to Probability and Statistics: Essays in Honor of Harold Hotelling*, 278–292. Stanford University Press, Palo Alto, CA.

See Also

leveneTest

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

levene.evaluate.core(data = ec, names = "genotypes",
                     quantitative = quant, selected = core)
```

---

pca.evaluate.core   *Principal Component Analysis*

Description

Compute Principal Component Analysis Statistics (Mardia et al. 1979) to compare the probability distributions of quantitative traits between entire collection (EC) and core set (CS).

Usage

```
pca.evaluate.core(
  data,
  names,
  quantitative,
  selected,
  center = TRUE,
  scale = TRUE,
  npc.plot = 6
)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string
quantitative	Name of columns with the quantitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the names column.
center	either a logical value or numeric-alike vector of length equal to the number of columns of x, where 'numeric-alike' means that as.numeric(.) will be applied successfully if is.numeric(.) is not true.
scale	either a logical value or a numeric-alike vector of length equal to the number of columns of x.
npc.plot	The number of principal components for which eigen values are to be plotted. The default value is 6.

Value

A list with the following components.

EC PC Importance	
	A data frame of importance of principal components for EC
EC PC Loadings	A data frame with eigen vectors of principal components for EC
CS PC Importance	
	A data frame of importance of principal components for CS
CS PC Loadings	A data frame with eigen vectors of principal components for CS
Scree Plot	The scree plot of principal components for EC and CS as a ggplot object.
PC Loadings Plot	
	A plot of the eigen vector values of principal components for EC and CS as specified by npc.plot as a ggplot2 object.

References

Mardia KV, Kent JT, Bibby JM (1979). *Multivariate analysis.* Academic Press, London; New York. ISBN 0-12-471250-9 978-0-12-471250-8 0-12-471252-5 978-0-12-471252-2.

See Also

prcomp

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

pca.evaluate.core(data = ec, names = "genotypes",
                  quantitative = quant, selected = core,
                  center = TRUE, scale = TRUE, npc.plot = 4)
```

---

pdfdist.evaluate.core   *Distance Between Probability Distributions*

Description

Compute Kullback-Leibler (Kullback and Leibler 1951), Kolmogorov-Smirnov (Kolmogorov 1933; Smirnov 1948) and Anderson-Darling distances (Anderson and Darling 1952) between the probability distributions of collection (EC) and core set (CS) for quantitative traits.

Usage

```
pdfdist.evaluate.core(data, names, quantitative, selected)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string

quantitative	Name of columns with the quantitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the `names` column.

Value

A data frame with the following columns.

Trait	The quantitative trait.
KL_Distance	The Kullback-Leibler distance (Kullback and Leibler 1951) between EC and CS.
KS_Distance	The Kolmogorov-Smirnov distance (Kolmogorov 1933; Smirnov 1948) between EC and CS.
KS_pvalue	The p value of the Kolmogorov-Smirnov distance.
AD_Distance	Anderson-Darling distance (Anderson and Darling 1952) between EC and CS.
AD_pvalue	The p value of the Anderson-Darling distance.
KS_significance	
	The significance of the Kolmogorov-Smirnov distance (*: $p \leq 0.01$; **: $p \leq 0.05$; ns: $p > 0.05$).
AD_pvalue	The significance of the Anderson-Darling distance (*: $p \leq 0.01$; **: $p \leq 0.05$; ns: $p > 0.05$).

See Also

[KL.plugin](), [ks.test](), [ad.test]()

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

pdfdist.evaluate.core(data = ec, names = "genotypes",
                      quantitative = quant, selected = core)
```

---

percentdiff.evaluate.core
Percentage Difference of Means and Variances

Description

Compute the following differences between the entire collection (EC) and core set (CS).

- Percentage of significant differences of mean ($MD\%_{Hu}$) (Hu et al. 2000)
- Percentage of significant differences of variance ($VD\%_{Hu}$) (Hu et al. 2000)
- Average of absolute differences between means ($MD\%_{Kim}$) (Kim et al. 2007)
- Average of absolute differences between variances ($VD\%_{Kim}$) (Kim et al. 2007)
- Percentage difference between the mean squared Euclidean distance among accessions ($\bar{d}D\%$) (Studnicki et al. 2013)

Usage

```
percentdiff.evaluate.core(data, names, quantitative, selected, alpha = 0.05)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string
quantitative	Name of columns with the quantitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the names column.
alpha	Type I error probability (Significance level) of difference.

Details

The differences are computed as follows.

$$MD\%_{Hu} = \left(\frac{S_t}{n} \right) \times 100$$

Where, S_t is the number of traits with a significant difference between the means of the EC and the CS and n is the total number of traits. A representative core should have $MD\%_{Hu} < 20$ % and $CR > 80$ % (Hu et al. 2000).

$$VD\%_{Hu} = \left(\frac{S_t}{n} \right) \times 100$$

Where, S_t is the number of traits with a significant difference between the variances of the EC and the CS and n is the total number of traits. Larger $VD\%_{Hu}$ value indicates a more diverse core set.

$$MD\%_{Kim} = \frac{1}{n} \sum_{i=1}^{n} \frac{|M_{EC_i} - M_{CS_i}|}{M_{CS_i}}$$

Where, M_{EC_i} is the mean of the EC for the ith trait, M_{CS_i} is the mean of the CS for the ith trait and n is the total number of traits.

$$VD\%_{Kim} = \frac{1}{n} \sum_{i=1}^{n} \frac{|V_{EC_i} - V_{CS_i}|}{V_{CS_i}}$$

Where, V_{EC_i} is the variance of the EC for the ith trait, V_{CS_i} is the variance of the CS for the ith trait and n is the total number of traits.

$$\overline{d}D\% = \frac{\overline{d}_{CS} - \overline{d}_{EC}}{\overline{d}_{EC}} \times 100$$

Where, $\overline{d}_{CS}$ is the mean squared Euclidean distance among accessions in the CS and $\overline{d}_{EC}$ is the mean squared Euclidean distance among accessions in the EC.

Value

A data frame with the values of $MD\%_{Hu}$, $VD\%_{Hu}$, $MD\%_{Kim}$, $VD\%_{Kim}$ and $\overline{d}D\%$.

References

Hu J, Zhu J, Xu HM (2000). "Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops." *Theoretical and Applied Genetics*, **101**(1), 264–268.

Kim K, Chung H, Cho G, Ma K, Chandrabalan D, Gwag J, Kim T, Cho E, Park Y (2007). "PowerCore: A program applying the advanced M strategy with a heuristic search for establishing core sets." *Bioinformatics*, **23**(16), 2155–2162.

Studnicki M, Madry W, Schmidt J (2013). "Comparing the efficiency of sampling strategies to establish a representative in the phenotypic-based genetic diversity core collection of orchardgrass (*Dactylis glomerata* L.)." *Czech Journal of Genetics and Plant Breeding*, **49**(1), 36–47.

See Also

snk.evaluate.core, snk.evaluate.core

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")
```

```
ec[, qual] <- lapply(ec[, qual],
                       function(x) factor(as.factor(x)))

percentdiff.evaluate.core(data = ec, names = "genotypes",
                          quantitative = quant, selected = core)
```

---

qq.evaluate.core   *Quantile-Quantile Plots*

Description

Plot Quantile-Quantile (QQ) plots (Wilk and Gnanadesikan 1968) to graphically compare the probability distributions of quantitative traits between entire collection (EC) and core set (CS).

Usage

```
qq.evaluate.core(data, names, quantitative, selected)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string
quantitative	Name of columns with the quantitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the names column.

Value

A list with the ggplot objects of QQ plots of CS vs EC for each trait specified as quantitative.

References

Wilk MB, Gnanadesikan R (1968). "Probability plotting methods for the analysis for the analysis of data." *Biometrika*, **55**(1), 1–17.

See Also

qqplot

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)
```

```
quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
          "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
         "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
         "PSTR")

ec[, qual] <- lapply(ec[, qual],
                    function(x) factor(as.factor(x)))

qq.evaluate.core(data = ec, names = "genotypes",
               quantitative = quant, selected = core)
```

---

shannon.evaluate.core   *Shannon-Weaver Diversity Index*

Description

Compute the Shannon-Weaver Diversity Index (H'), Maximum diversity (H'_{max}) and Shannon Equitability Index (E_H) (Shannon and Weaver 1949) to compare the phenotypic diversity for qualitative traits between entire collection (EC) and core set (CS).

Usage

```
shannon.evaluate.core(data, names, qualitative, selected)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string
qualitative	Name of columns with the qualitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the names column.

Details

Shannon-Weaver Diversity Index (H') is computed as follows.

$$H' = -\sum_{i=1}^{k} p_i \ln(p_i)$$

Where p_i denotes the proportion in the group k.

The maximum value of the index (H'_{max}) is $\ln(k)$. This value occurs when each group has the same frequency.

The Shannon equitability index (E_H) is the Shannon diversity index divided by the maximum diversity.

$$E_H = \frac{H'}{\ln(k)}$$

Value

A data frame with the following columns.

Trait	The qualitative trait.
EC_H	The Shannon-Weaver Diversity Index (H') for EC.
EC_H	The Shannon-Weaver Diversity Index (H') for CS.
EC_Hmax	The Maximum diversity value (H'_{max}) for EC.
CS_Hmax	The Maximum diversity value (H'_{max}) for CS.
EC_EH	The Shannon Equitability Index (E_H) for EC.
CS_EH	The Shannon Equitability Index (E_H) for CS.

References

Shannon CE, Weaver W (1949). *The Mathematical Theory of Communication*, number v. 2 in The Mathematical Theory of Communication. University of Illinois Press.

See Also

[shannon](#)

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

shannon.evaluate.core(data = ec, names = "genotypes",
                      qualitative = qual, selected = core)
```

---

signtest.evaluate.core

Sign Test

Description

Test difference between means and variances of entire collection (EC) and core set (CS) for quantitative traits by Sign test ($+$ versus $-$) (Basigalup et al. 1995; Tai and Miller 2001).

Usage

```
signtest.evaluate.core(data, names, quantitative, selected)
```

Arguments

`data`	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
`names`	Name of column with the individual names as a character string
`quantitative`	Name of columns with the quantitative traits as a character vector.
`selected`	Character vector with the names of individuals selected in core collection and present in the `names` column.

Details

The test statistic for Sign test (χ^2) is computed as follows.

$$\chi^2 = \frac{(N_1 - N_2)^2}{N_1 + N_2}$$

Where, where N_1 is the number of variables for which the mean or variance of the CS is greater than the mean or variance of the EC (number of $+$ signs); N_2 is the number of variables for which the mean or variance of the CS is less than the mean or variance of the EC (number of $-$ signs). The value of χ^2 is compared with a Chi-square distribution with 1 degree of freedom.

Value

A data frame with the following components.

`Comparison`	The comparison measure.
`ChiSq`	The test statistic (χ^2).
`p.value`	The p value for the test statistic.
`significance`	The significance of the test statistic (*: $p \leq 0.01$; **: $p \leq 0.05$; ns: $p > 0.05$).

References

Basigalup DH, Barnes DK, Stucker RE (1995). "Development of a core collection for perennial *Medicago* plant introductions." *Crop Science*, **35**(4), 1163–1168.

Tai PYP, Miller JD (2001). "A Core Collection for *Saccharum spontaneum* L. from the World Collection of Sugarcane." *Crop Science*, **41**(3), 879–885.

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)
```

```
quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

signtest.evaluate.core(data = ec, names = "genotypes",
                       quantitative = quant, selected = core)
```

---

snk.evaluate.core   *Student-Newman-Keuls Test*

Description

Test difference between means of entire collection (EC) and core set (CS) for quantitative traits by Newman-Keuls or Student-Newman-Keuls test (Newman 1939; Keuls 1952).

Usage

```
snk.evaluate.core(data, names, quantitative, selected)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string
quantitative	Name of columns with the quantitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the names column.

Value

A data frame with the following components.

Trait	The quantitative trait.
EC_Min	The minimum value of the trait in EC.
EC_Max	The maximum value of the trait in EC.
EC_Mean	The mean value of the trait in EC.
EC_SE	The standard error of the trait in EC.
CS_Min	The minimum value of the trait in CS.
CS_Max	The maximum value of the trait in CS.
CS_Mean	The mean value of the trait in CS.
CS_SE	The standard error of the trait in CS.

SNK_pvalue   The p value of the Student-Newman-Keuls test for equality of means of EC and CS.

SNK_significance
                  The significance of the Student-Newman-Keuls test for equality of means of EC and CS.

References

Keuls M (1952). "The use of the „studentized range" in connection with an analysis of variance." *Euphytica*, **1**(2), 112–122.

Newman D (1939). "The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation." *Biometrika*, **31**(1-2), 20–30.

See Also

SNK.test

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

snk.evaluate.core(data = ec, names = "genotypes",
                  quantitative = quant, selected = core)
```

---

ttest.evaluate.core   *Student's t Test*

Description

Test difference between means of entire collection (EC) and core set (CS) for quantitative traits by Student's t test (Student 1908).

Usage

```
ttest.evaluate.core(data, names, quantitative, selected)
```

Arguments

`data`	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
`names`	Name of column with the individual names as a character string
`quantitative`	Name of columns with the quantitative traits as a character vector.
`selected`	Character vector with the names of individuals selected in core collection and present in the `names` column.

Value

`Trait`	The quantitative trait.
`EC_Min`	The minimum value of the trait in EC.
`EC_Max`	The maximum value of the trait in EC.
`EC_Mean`	The mean value of the trait in EC.
`EC_SE`	The standard error of the trait in EC.
`CS_Min`	The minimum value of the trait in CS.
`CS_Max`	The maximum value of the trait in CS.
`CS_Mean`	The mean value of the trait in CS.
`CS_SE`	The standard error of the trait in CS.
`ttest_pvalue`	The p value of the Student's t test for equality of means of EC and CS.
`ttest_significance`	
	The significance of the Student's t test for equality of means of EC and CS.

References

Student (1908). "The probable error of a mean." *Biometrika*, **6**(1), 1–25.

See Also

[t.test](#)

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
```

```
                    function(x) factor(as.factor(x)))

ttest.evaluate.core(data = ec, names = "genotypes",
                    quantitative = quant, selected = core)
```

---

vr.evaluate.core   *Variable Rate of Coefficient of Variation*

Description

Compute the Variable Rate of Coefficient of Variation (VR) (Hu et al. 2000) to compare quantitative traits of the entire collection (EC) and core set (CS).

Usage

```
vr.evaluate.core(data, names, quantitative, selected)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string
quantitative	Name of columns with the quantitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the names column.

Details

The Variable Rate of Coefficient of Variation (VR) is computed as follows.

$$VR = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{CV_{CS_i}}{CV_{EC_i}} \right) \times 100$$

Where, CV_{CS_i} is the coefficients of variation for the ith trait in the CS, CV_{EC_i} is the coefficients of variation for the ith trait in the EC and n is the total number of traits

Value

The VR value.

References

Hu J, Zhu J, Xu HM (2000). "Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops." *Theoretical and Applied Genetics*, **101**(1), 264–268.

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

vr.evaluate.core(data = ec, names = "genotypes",
                 quantitative = quant, selected = core)
```

---

wilcox.evaluate.core   *Wilcoxon Rank Sum Test*

Description

Compare the medians of quantitative traits between entire collection (EC) and core set (CS) by Wilcoxon rank sum test or Mann-Whitney-Wilcoxon test or Mann-Whitney U test (Wilcoxon 1945; Mann and Whitney 1947).

Usage

```
wilcox.evaluate.core(data, names, quantitative, selected)
```

Arguments

data	The data as a data frame object. The data frame should possess one row per individual and columns with the individual names and multiple trait/character data.
names	Name of column with the individual names as a character string
quantitative	Name of columns with the quantitative traits as a character vector.
selected	Character vector with the names of individuals selected in core collection and present in the names column.

Value

Trait	The quantitative trait.
EC_Med	The median value of the trait in EC.
CS_Med	The median value of the trait in CS.

Wilcox_pvalue   The p value of the Wilcoxon test for equality of medians of EC and CS.

Wilcox_significance

                  The significance of the Wilcoxon test for equality of medians of EC and CS.

References

Mann HB, Whitney DR (1947). "On a test of whether one of two random variables is stochastically larger than the other." *The Annals of Mathematical Statistics*, **18**(1), 50–60.

Wilcoxon F (1945). "Individual comparisons by ranking methods." *Biometrics Bulletin*, **1**(6), 80.

See Also

wilcox.test

Examples

```
data("cassava_CC")
data("cassava_EC")

ec <- cbind(genotypes = rownames(cassava_EC), cassava_EC)
ec$genotypes <- as.character(ec$genotypes)
rownames(ec) <- NULL

core <- rownames(cassava_CC)

quant <- c("NMSR", "TTRN", "TFWSR", "TTRW", "TFWSS", "TTSW", "TTPW", "AVPW",
           "ARSR", "SRDM")
qual <- c("CUAL", "LNGS", "PTLC", "DSTA", "LFRT", "LBTEF", "CBTR", "NMLB",
          "ANGB", "CUAL9M", "LVC9M", "TNPR9M", "PL9M", "STRP", "STRC",
          "PSTR")

ec[, qual] <- lapply(ec[, qual],
                     function(x) factor(as.factor(x)))

wilcox.evaluate.core(data = ec, names = "genotypes",
                     quantitative = quant, selected = core)
```

Index