Package ‘ccChooser’

February 19, 2015

Type Package

Title Developing a core collections

Version 0.2.6

Date 2012-09-25

Author Marcin Studnicki and Konrad Debski

Maintainer Marcin Studnicki <`marcin_studnicki@sggw.pl`>

Depends R(>= 2.14.1), cluster (>= 1.13.1)

Description ccChooser can be used to developing and evaluation of core collections for germplasm collections (entire collection). This package used to develop a core collection for biological resources like genbanks. A core collection is defined as a sample of accessions that represent, with the lowest possible level of redundancy, the genetic diversity (the richness of gene or genotype categories) of the entire collection. The establishing a core collection that represents genetic diversity of the entire collection with minimum loss of its original diversity and minimum redundancies is an important problem for gene-banks curators and crop breeders. ccChooser establish core collection base on phenotypic data (agronomic, morphological, phenological).

License GPL (>= 2)

Repository CRAN

Date/Publication 2012-09-26 03:01:39

NeedsCompilation no

R topics documented:

`ccChooser-package` *Developing a core collections*

Description

ccChooser can be used to developing and evaluation of core collections for germplasm collections (entire collection). This package used to develop a core collection for biological resources like genbanks. A core collection is defined as a sample of accessions that represent, with the lowest possible level of redundancy, the genetic diversity (the richness of gene or genotype categories) of the entire collection. The establishing a core collection that represents genetic diversity of the entire collection with minimum loss of its original diversity and minimum redundancies is an important problem for gene-banks curators and crop breeders. ccChooser establish core collection base on phenotypic data (agronomic, morphological, phenological).

Details

Package:	ccChooser
Type:	Package
Version:	0.2.6
Date:	2012-09-25
License:	GPL (>= 2)

Author(s)

Marcin Studnicki and Konrad Debski Maintainer: Marcin Studnicki <marcin_studnicki@sggw.pl>

`allocc` *Allocation accesions*

Description

This function determining the numbers of entries from each cluster (from entire collection) included in a core collection.

Usage

```
allocc(x, groups, fraction = 0.1, method = ”Pro”)
```

Arguments

`x`	data frame including name of acessions, number of groups and value for quantitaive traits (phenotypic data)
`groups`	defining column where present number of groups (clusters)
`fraction`	defining size of core collections, as fraction of accesions from entire collection
`method`	string defining the allocation method. The four methods implemented are "Pro" (propotional) , "Log" (logarytmic), "D2" (D2 alocation method) and "D3" (D3 alocation method)

Details

Four allocation methods were used in this function. Two of these methods based on the group size, are commonly known as the proportional (Pro) and logarithmic (Log) methods. Franco et al. (2005) proposed allocation methods for determining the number of accessions taken from a group based on the mean of the Gower's distance between accessions within the group. The respective allocation methods suggested by Franco et al. (2005) are called as D2 and D3. The first, D2, allocation method determines that the size of the sample to be drawn from each group should be proportional to the mean squared Euclidean distance between the accessions within that group and size group. The D3 allocation method based on proportional to the mean squared Euclidean distance between the accessions within that group and logarithm from size of the group.

Value

Return a matrix of groups and numbers accesions including in core collection.

Author(s)

Marcin Studnicki and Konrad Debski

References

For information about the allocation method for developing core collection, see:

van Hintum Th., Brown A., Spillane C., Hodgkin T. 2000. Core collections of plant genetic resources. IPGRI Technical Bulletin No. 3. IPGRI, Rome

Franco J., Crossa J., Taba S., Shands H. 2005. A sampling strategy for conserving genetic diversity when forming core subsets. Crop Sci. 45:1035-1044

Studnicki, M., Madry, W., Kociuba, W. 2010. The efficiency and effectiveness of sampling strategies used to develop a core collection for the Polish spring triticale (Triticosecale Wittm.) germplasm resources. Communications in Biometry and Crop Science 5 (2), 127-137

Examples

```
data(dactylis_EC)
allocc(dactylis_EC, dactylis_EC$UPGMA)
allocc(dactylis_EC, dactylis_EC$UPGMA, fraction = 0.15, method = 'D2')
```

`dactylis_CC` *Core collection of orchadgrass*

Description

Core collection of orchadgrass (Dactylis glomerata L.) obtain from enitre collection, held at Botanical Garden of Plant Breeding and Acclimatization Institute in Bydgoszcz, Poland. Core collection establish used stratified method with Pro allocation method and random sampling.

Usage

```
data(dactylis_CC)
```

Format

A data frame with 197 accessions on the following 8 variables.

`X1`	plant height (cm)
`X2`	inflorescence length
`X3`	leaf width (cm)
`X4`	leaf length (cm)
`X5`	total seasonal yield (kg)
`X6`	1000 kernel weight (g)
`X7`	number of days since the 1st April to inflorescence emergence
`X8`	number of days since the 1st April to inflorescence

Examples

```
data(dactylis_CC)
data(dactylis_CC)
```

`dactylis_EC` *The orchadgrass germplasm collection*

Description

The orchadgrass (Dactylis glomerata L.) germplasm collection, held at Botanical Garden of Plant Breeding and Acclimatization Institute in Bydgoszcz, Poland. The number of accessions was limited by the availability of the data on the 8 important agro-morphological traits.). The accessions were assessed in the field trials located at Botanical Garden of Plant Breeding and Acclimatization Institute in Bydgoszcz, Poland (53.174 N, 18.046 E)

Usage

```
data(dactylis_EC)
```

Format

A data frame with 1971 accessions on the following 8 variables and cllassification by UPGMA methods.

`UPGMA`	classification accessions by UPGMA cluster analysis methods
`X1`	plant height (cm)
`X2`	inflorescence length
`X3`	leaf width (cm)
`X4`	leaf length (cm)
`X5`	total seasonal yield (kg)
`X6`	1000 kernel weight (g)
`X7`	number of days since the 1st April to inflorescence emergence
`X8`	number of days since the 1st April to inflorescence

Examples

```
data(dactylis_EC)
summary(dactylis_EC)
```

`evalucc` *Evaluation of core collection.*

Description

This function evaluation efficiency of the new core collection, using the four parameters.

Usage

```
evalucc(CC, EC)
```

Arguments

`CC`	data frame including value of quantitaive traits (phenotypic data) for core collection
`EC`	data frame including value of quantitaive traits (phenotypic data) for entire collection

Details

The DD The MD The VD The RR

Value

Return a matrix present a value of four parameters- DD

Author(s)

Marcin Studnicki and Konrad Debski

References

For information about the evaluation of core collection, see:

Franco J., Crossa J., Taba S., Shands H. 2005. A sampling strategy for conserving genetic diversity when forming core subsets. Crop Sci. 45:1035-1044

Hu J., Zhu J., Xu H.M. 2000. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. Theoretical and Applied Genetics 101:264-268

Kim K.W., Chung H.K., Cho G.T., Ma K.H., Chandrabalan D., Gwag J.G., Kim T.S., Cho E.G., Park Y.J. 2007. PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. Bioinformatics 23:2155-2162

Studnicki, M., Madry, W., Kociuba, W. 2010. The efficiency and effectiveness of sampling strategies used to develop a core collection for the Polish spring triticale (Triticosecale Wittm.) germplasm resources. Communications in Biometry and Crop Science 5:127-137

Wang J., Hu J., Zhang C.F., Zhang S. 2007. A strategy on constructing core collections by least distance stepwise sampling. Theoretical and Applied Genetics 15:1-8

Examples

```
data(dactylis_CC)
data(dactylis_EC)
dactylis_EC<-subset(dactylis_EC, select= -UPGMA)
evalucc(dactylis_CC, dactylis_EC)
```

`randomcc` *Random selection of core collections*

Description

This function developing core collection use random sampling methods.

Usage

```
randomcc(x, fraction = 0.1)
```

Arguments

`x`	data frame including value of quantitaive traits (phenotypic data)
`fraction`	defining size of core collections, as fraction of accesions from entire collection

Value

Return a data frame including selected to core collection acessions

Author(s)

Marcin Studnicki and Konrad Debski

References

For information about the random method used to developing core collection, see:

van Hintum Th., Brown A., Spillane C., Hodgkin T. 2000. Core collections of plant genetic resources. IPGRI Technical Bulletin No. 3. IPGRI, Rome

Examples

```
data(dactylis_EC)
randomcc(dactylis_EC)
randomcc(dactylis_EC, fraction = 0.2)
```

`stratcc` *Stratified selection of core collections*

Description

This function developing core collection use stratified sampling methods.

Usage

```
stratcc(x, groups, alloc = "Pro", fraction = 0.1, clustering = FALSE, cluster_method = "ward")
```

Arguments

`x`	data frame including classifications of accessions (number of groups) and value quantitaive traits (phenotypic data)
`groups`	defining column where present classifications of accessions (groups)
`alloc`	string defining the allocation method. The four methods implemented are "Pro" (propotional) , "Log" (logarytmic), "D2" (D2 alocation method) and "D3" (D3 alocation method)
`fraction`	defining size of core collections, as fraction of accesions from entire collection
`clustering`	if TRUE to selecting acessions in group used clustered sampling method. If FALSE to sample of accessions in group used random sampling method.
`cluster_method`	when clustering=TRUE defining cluster analysis methods.

Details

The first step of stratified methods in the development of core collections is stratification (grouping, classification). The next step was allocation, i.e. determining the number of entries from each cluster (group) included in a core collection. The third, final, step of stratified methods is selection of accesions. Two sampling methods were used. These are random and clustered sampling methods. The random sampling method was based on a simple random sample of accessions in each group (van Hintum et al. 2000). The clustered sampling method is a non-random (restricted-random) sampling method which involves dividing accessions in each group into homogeneous subgroups by cluster analysis. The number of subgroups for a group was specified in an allocation method and was equal to the nt number of accessions from this group being selected to a core. From each subgroup containing more accessions only one of them was randomly selected to a core subset and from each one-entry subgroup just this accession was selected.

Value

Return a data frame including selected to core collection acessions

Author(s)

Marcin Studnicki and Konrad Debski

References

For information about the stratified method used to developing core collection, see:

van Hintum Th., Brown A., Spillane C., Hodgkin T. 2000. Core collections of plant genetic resources. IPGRI Technical Bulletin No. 3. IPGRI, Rome

Franco J., Crossa J., Taba S., Shands H. 2005. A sampling strategy for conserving genetic diversity when forming core subsets. Crop Sci. 45:1035-1044

Studnicki, M., Madry, W., Kociuba, W. (2010). The efficiency and effectiveness of sampling strategies used to develop a core collection for the Polish spring triticale (Triticosecale Wittm.) germplasm resources. Communications in Biometry and Crop Science 5 (2), 127-137

See Also

`allocc`

Examples

```
data(dactylis_EC)
stratcc(dactylis_EC, dactylis_EC$UPGMA)
stratcc(dactylis_EC, dactylis_EC$UPGMA, clustering = TRUE, cluster_method = "average")
```

Index