Evaluation of Core Collections with `EvaluateCore`

Aravind, J.[1], Kaur, V.[2], Wankhede, D. P.[3], and Nanjundan, J.[4]

2025-03-04

1. Division of Germplasm Conservation, ICAR-National Bureau of Plant Genetic Resources, New Delhi. 2. Division of Germplasm Evaluation, ICAR-National Bureau of Plant Genetic Resources, New Delhi. 3. Division of Genomic Resources, ICAR-National Bureau of Plant Genetic Resources, New Delhi. 4. ICAR-Indian Agricultural Research Institute, Regional Station, Wellington, Tamil Nadu.

Contents

Overview

A core collection is a *"limited set of accessions representing, with minimum repetitiveness, the genetic diversity of a crop species and its wild relatives"*(Frankel, 1984).

In case of several large and unwieldy germplasm collections conserved in genebanks, development of several such core collections has facilitated increasing the efficiency of their characterisation and in turn the utilisation.

There are different criteria available for evaluating the core collections by comparing with the original collections from which they have been sampled. The choice of the suitable evaluation criteria should be dictated by the objectives of the core collection. Several such criteria including the distance based metrics described in Odong et al. (2013) have been implemented in `EvaluateCore` and this document shows how to use them. This document assumes a basic knowledge of `R` programming language.

Installation

The package can be installed using the following functions:

```
# Install from CRAN
install.packages('EvaluateCore', dependencies=TRUE)

# Install development version from Github
devtools::install_github("aravind-j/EvaluateCore")
```

Then the package can be loaded using the function

```
library(EvaluateCore)
```

Version History

The current version of the package is 0.1.3. The previous versions are as follows.

Table 1. Version history of `EvaluateCore` `R` package.

Version	Date
0.1.1	2020-06-03
0.1.2	2021-06-12

To know detailed history of changes use `news(package='EvaluateCore')`.

Table 2 : Core evaluation methods implemented in `EvaluateCore`.

Metric	Criteria/Method	Function	Type of Variable	Type of Core	Type of comparison	Reference
Genetic distance	Average entry-to-nearest-entry distance ($E - EN$)	`dist.evaluate.core`	Quantitative & Qualitative	CC–I	Multivariate	Odong et al. (2013)
Genetic distance	Average accession-to-nearest-entry distance ($A - EN$)	`dist.evaluate.core`	Quantitative & Qualitative	CC–X	Multivariate	Odong et al. (2013)
Genetic distance	Average entry-to-entry distance ($E - E$)	`dist.evaluate.core`	Quantitative & Qualitative	CC–X	Multivariate	Odong et al. (2013)
Mean	Student-Newman-Keuls test	`snk.evaluate.core`	Quantitative	CC–D	Univariate	Newman (1939); Keuls (1952)
Median	Wilcoxon rank sum test or Mann-Whitney-Wilcoxon test or Mann-Whitney U test	`wilcox.evaluate.core`	Quantitative	CC–D	Univariate	
Mean	Student's t test	`ttest.evaluate.core`	Quantitative	CC–D	Univariate	
Mean and Variance	Sign test	`signtest.evaluate.core`	Quantitative	CC–D	Multivariate	
Variance	Levene's test	`levene.evaluate.core`	Quantitative	CC–D	Univariate	
Variability	Interquartile range	`iqr.evaluate.core`	Quantitative	CC–D	Univariate	
Range	Coincidence rate of range	`cr.evaluate.core`	Quantitative	CC–D	Multivariate	
Variance	Variable rate of coefficient of variation	`vr.evaluate.core`	Quantitative	CC–D	Multivariate	
Probability distribution	Quantile-quantile plots	`qq.evaluate.core`	Quantitative	CC–D	Univariate	
Probability distribution	Kullback-Leibler distance	`pdfdist.evaluate.core`	Quantitative	CC–D	Univariate	
Probability distribution	Kolmogorov-Smirnov distance	`pdfdist.evaluate.core`	Quantitative	CC–D	Univariate	
Probability distribution	Anderson-Darling distance	`pdfdist.evaluate.core`	Quantitative	CC–D	Univariate	
Probability distribution	Box-and-whisker plot	`box.evaluate.core`	Quantitative	CC–D	Univariate	
Mean	Percentage of significant differences of mean	`percentdiff.evaluate.core`	Quantitative	CC–D	Multivariate	
Variance	Percentage of significant differences of variance	`percentdiff.evaluate.core`	Quantitative	CC–D	Multivariate	
Mean	Average of absolute differences between means	`percentdiff.evaluate.core`	Quantitative	CC–D	Multivariate	
Variance	Average of absolute differences between variances	`percentdiff.evaluate.core`	Quantitative	CC–D	Multivariate	
Euclidean distance	Percentage difference between the mean squared Euclidean distance	`percentdiff.evaluate.core`	Quantitative	CC–D	Multivariate	

Metric	Criteria/Method	Function	Type of Variable	Type of Core	Type of comparison	Reference
Probability distribution	Principal component analysis	`pca.evaluate.core`	Quantitative	CC–D	Multivariate	
Frequency distribution	Bar plot	`bar.evaluate.core`	Qualitative	CC–D	Univariate	
Frequency distribution	Chi-squared test for homogeneity	`chisquare.evaluate.core`	Qualitative	CC–D	Univariate	
Frequency distribution	Class coverage	`coverage.evaluate.core`	Qualitative	CC–I	Univariate	
Diversity	Shannon-Weaver diversity index	`shannon.evaluate.core`	Qualitative	CC–I	Univariate	
Diversity	Maximum Shannon-Weaver diversity index	`shannon.evaluate.core`	Qualitative	CC–I	Univariate	
Diversity	Shannon Equitability index	`shannon.evaluate.core`	Qualitative	CC–I	Univariate	
Phenotypic correlation	Pearson correlation coefficients	`corr.evaluate.core`	Quantitative & Qualitative	CC-D	Multivariate	
Correlation matrix	Mantel correlation coefficient	`corr.evaluate.core`	Quantitative & Qualitative	CC-D	Multivariate	
Frequency distribution	Stacked frequency distribution histogram	`freqdist.evaluate.core`	Quantitative & Qualitative	CC-D	Univariate	

References

Frankel, O. H. (1984). "Genetic perspective of germplasm conservation," in *Genetic Manipulations: Impact on Man and Society*, eds. W. Arber, K. Limensee, W. J. Peacock, and P. Stralinger (Cambridge University Press: Cambridge, UK), 161–170.

Keuls, M. (1952). The use of the „studentized range" in connection with an analysis of variance. *Euphytica* 1, 112–122. doi:10.1007/BF01908269.

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, 20–30. doi:10.1093/biomet/31.1-2.20.

Odong, T. L., Jansen, J., Eeuwijk, F. A. van, and Hintum, T. J. L. van (2013). Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. *Theoretical and Applied Genetics* 126, 289–305. doi:10.1007/s00122-012-1971-y.