



DIVERSIFOOD

Embedding crop diversity and networking for local high quality food systems

Grant agreement n°: 633571

H2020 - Research and Innovation Action

D3.2: User-friendly tools incorporating the relevant methods for decentralized on farm breeding

Due date: M32

Actual submission date: M48

Project start date: March 1st, 2015 **Duration:** 48 months

Workpackage concerned: WP3

Concerned workpackage leader: Isabelle Goldringer

Lead Beneficiary: INRA / RSP

Dissemination level:

- PU:** Public (must be available on the website)
- CO:** Confidential, only for members of the consortium (including the Commission Services)
- CI:** Classified, as referred to in Commission Decision 2001/844/EC

Abstract

This deliverable provides a list of software that allow implementing the methods relevant for PPB experiments and on-farm decentralized trials described in deliverable **D3.1: Smart methods specifically suited for decentralized on farm breeding**. Statistical analyses have been identified, adapted or developed by the partners involved in this activity (RSR, IPC, ITQB, RSP, ITAB, INRA). A first set of these methods has been implemented in a common framework in a new software: the R package PPBstats; this package constitutes the heart of D3.2.

Content

PPBstats package is based on the R software that is open source and widely used in the breeding and agronomy community. PPBstats aims at performing the analyses relevant for PPB programmes at four levels:

- agronomic trials
- nutritional analyses and organoleptic tests
- molecular data
- network of seeds circulation.

PPBstats is still under development and the code is hosted on Github to facilitate collaboration: <https://github.com/priviere/PPBstats>.

In the following, numbers (Dx and Mx) for the experimental designs and for the methods, refer to those presented in the decision tree (see **D3.1**) and organised according to the objectives of the experiments.

The following experimental designs can be used:

- **D1**: fully-replicated block design
- **D2**: incomplete block design
- **D3**: row-column design
- **D4**: satellite-farms & regional-farms

The following methods have been implemented:

- **M2**: Multivariate analyses (PCA)
- **M4a**: Anova
- **M4b**: Spatial analysis
- **M6**: AMMI and GGE
- **M7a**: Bayesian hierarchical model intra-location
- **M7b**: Bayesian hierarchical model GxE
- **M8**: Network analysis

The following methods are not yet implemented and can be performed through other software:

- **M1**: Non parametric; multivariate regression; classification & regression trees; random forest:

- Classification And Regression Trees (CART): rpart, the recursive partitioning algorithm, is the function used to train Classification And Regression Trees (CART). The function rpart is available in the R package rpart.
- Multivariate Linear Regression (MLR): lm, the linear model function is available in the base package of R. I.e., you don't need to open any specific R package.
- Multivariate Adaptive Regression Splines (MARS): function earth from the R-project.
- Random Forest: the function randomForest from the R package randomForest.
- **M2:** Multivariate analyses (clustering, discriminant analysis):
 - R package FactoMineR, <http://factominer.free.fr/index.html>
- **M3:** Analyses of molecular data such as genetic distances, trees, clustering. They can be done through :
 - R package adegenet
 - PowerMarker, <http://statgen.ncsu.edu/powermarker/downloads.htm> - V3.23 (Liu, 2002)
 - GENEPOP, <http://kimura.univ-montp2.fr/~rousset/Genepop.htm> - 4.0 (Raymond and Rousset 1995)
 - FSTAT, <http://www2.unil.ch/popgen/softwares/fstat.htm> - FSTAT v. 2.9.3.2, program package (Goudet 2002)
 - ARLEQUIN, <http://cmpg.unibe.ch/software/arlequin35/Arl35Downloads.html>) - ARLEQUIN ver. 3.0 (Excoffier et al., 2005)
 - PHYLIP, <http://evolution.genetics.washington.edu/phylip/getme.html> - PHYLIP ver. 3.6b software package (Felsenstein 1993)
 - STRUCTURE, <http://pritchardlab.stanford.edu/structure.html> - STRUCTURE ver. 2.3.3 (Pritchard et al., 2000)
 - STRUCTURE HARVESTER, <http://taylor0.biology.ucla.edu/structureHarvester/> - STRUCTURE HARVESTER v0.6.92 (Earl and van Holdt, 2012)
- **M5:** Mixed models for incomplete block designs: Genstats module.
- **M9a:** Multiple Factors Analyses; Projection Word Frequency:
 - R packages FactoMineR, <http://factominer.free.fr/index.html>
 - R package SensoMineR, <http://sensominer.free.fr/>
- **M9b:** ANOVA; Hierarchical Cluster Analysis; Correspondance Analysis on additional sensory descriptors:
 - R packages FactoMineR, <http://factominer.free.fr/index.html>
 - R package SensoMineR, <http://sensominer.free.fr/>
- **M9c:** Non parametric Test on Rank Sums; Friedman's Test: basic R functions.

Methods **M2**, **M5**, **M9a**, **M9b** and **M9c** will be implemented in PPBstats by the end of DIVERSIFOOD project.

A website dedicated to PPBstats and an exhaustive tutorial to collaborate and use the package can be found here: https://priviere.github.io/PPBstats_web_site.

-