



www.diversifood.eu

# SMART METHODS FOR DECENTRALIZED ON-FARM BREEDING

## A decision tree to identify smart methods

Participatory Plant Breeding (PPB) is based on decentralized on-farm breeding, which requires appropriate experimental methods.

A decision tree (*cf. page 2*) has been developed within DIVERSIFOOD, to match **experimental design** and **statistical methods** to a particular PPB project, according to its objectives and experimental constraints.

#### **AT FIRST GLANCE**

Decentralized on-farm breeding requires appropriate methods due to:
 - on farms specific conditions;
 -collaborative approach with farmers.
A decision tree has been developed to select the most appropriate experimental designs and methods, according to the objectives and the experimental constraints.



Embedding crop diversity and networking for local high quality food systems

Most of the methods have been implemented in an R Package: **PPBStats** (developed in DIVERSIFOOD deliverable **D3.2** and described in **D3.1**), whose code is hosted on: <a href="https://github.com/priviere/PPBstats">https://github.com/priviere/PPBstats</a>

The decision tree is organised according to the objectives of the experiments. For each, there are several methods based on different experimental designs that require specific conditions (e.g. number of plots per location; of replicated germplasms within and between locations).

#### Identifying the objectives of the experiment

Data analyses from PPB programmes may have one or multiple objectives. The first step is to identify **the objective(s)** of the experiment (*in green on the decision tree*), among the following:

Four types of information can be considered: agronomic/nutritional data, sensory data, network topology

- To improve the prediction of a target variable for selection by analysing agronomic /nutritional traits.
- To compare different varieties or populations (hereafter called germplasms) evaluated for selection in different locations by analysing agronomic/nutritional traits and by sensory analysis.
- To study the response of germplasms under selection over several environments by analysing agronomic traits.
- To study diversity structure and identify parents for crossing based on either good complementarity or similarity for some traits by analysing agronomic traits and molecular data.
- To study networks of seed circulation by analysing network topology.

of the seed circulation and molecular data.

### Experimental designs and statistical methods available

Four **experimental designs** can be used: **D1**: fully-replicated block design; **D2**: incomplete block design; **D3**: row-column design; **D4**: satellite-farms & regional-farms.

Nine **statistical methods** can be used, depending on the design, the type of data and the objectives: **M1**: Non parametric methods; **M2**: Multivariate analyses (PCA); **M3**: Genetic distances & trees; **M4a**: Anova; **M4b**: Spatial analysis; **M5**: Mixed models for incomplete block designs; **M6**: AMMI and GGE; **M7a**: Bayesian hierarchical model intra-location; **M7b**: Bayesian hierarchical model GxE; **M8**: Network analysis; **M9a**: Napping tests; **M9b**: Hedonic tests; **M9c**: Ranking tests.

#### Suggested readings

Rivière, P., J.C. Dawson, I. Goldringer, and O. David. 2015. "Hierarchical Bayesian Modeling for Flexible Experiments in Decentralized Participatory Plant Breeding." Crop Science 55 (3).

Gauch, H.G. 2006. "Statistical Analysis of Yield Trials by AMMI and GGE." Crop Sci 46 (4): 1488–1500.

M. Singh, R. S. Malhotra, S. Ceccarelli, A. Sarker, S. Grando And W. Erskine (2003) Spatial variability models to improve dryland field trials. Experimental Agriculture 39(2): 151-160.





														AB	II : ngisəU	
	thods for PPB	Decision tree	M6a - AWMI M6b - GCE	M7b - Bayesian hierarchical model G&E	M1 - Non parametric; multivariate regression; classification & regression trees; random forest					M4a - Anova	M4b - Spatial analysis	ted M7a - Bayesian hierarchical t ocations model intra-location	M5 - Mixed models for incomplete block design		mz - munyanare analysis (PCA, clustering, discriminant analysis)	Method
	statisticals methods for		Same entries in all locations, all entries are replicated at least twice in each location Dr - fully replicated	All locations share one replicated control or more; entries are not replicated within and among locations D4 - stallite and regional farms	Same entries in all locations, all entries are replicated at least twice in each location D1 - fully replicated			9		All entries are replicated at least twice Di - fully-replicated Eurl or incomplete replications	control is replicated in rows and columns <b>D3 - row-column</b>	All locations share one replicated control or more; entries are not replicated within and among locations D4 - stallite and regional fams	Entries are replicated at least twice and distributed among environments D2 - incomplete block design	Same entries in all locations	all entries are replicated at least twice in each location DI - fully replicated	Experimental design
	and		At least two locations and one year or more	At least 25 environments (i.e. number location x number of year ≥ 25)	At least one environment (i.e. number location x number of year ≥ 1)	W9a - Multiple factors	analyses; Projection word frequency	M9b - ANOVA; Herarchical cluster analysis; Correspondance analysis on additionnal sensory descriptors	M9c - Non parametric test on rank sums; Friedman's Test	One or several locations	aid old of several years	At least 25 environments (i.e. number location x number year ≥ 25)	At least one environment (i.e. number location x number year ≥ 1)		At least one environment (i.e. number location x number year ≥ 1)	Experimental constraints
M8 - Network analysis	M8 - Network analysis	M8 - Network analysis	Number of plots per location: large	Number of plots per location: low	Number of plots per location: large		Number of product < 12 Number of tasters > 10	Number of product < 7 Number of tasters > 60	Number of product < 6 Number of tasters > 12	Number of plots per	ocanon: raige	Number of plots per location: low		M3 - Genetic distances; trees	Number of plots per location: large	Experimental constraints
Bipart network for germplasmand location	Unipart network for location	Unipart network for seed lots	Quantitative traits, single- trait approach		Quantitative or qualitative traits, multi-trait approach		Napping test : characterize product sensory properties	Hedonic test : understand and check consumers preferences	Ranking test: rank product according to few sensory descriptors		Quantitative traits, single- trait approach			Individual genetic data	Quantitative or qualitative traits, multi-trait approach	Type of data
	Network topology		Agronomic /	Nutritional traits traits	Agronomic / Nutritional traits traits			Sensory		Agronomic / Nutritional traits				Molecular data	Agronomic / Nutritionnal traits traits	Information type
	Study network of see circulation		Study the response of populations under	selection over several environments	Improve the prediction of a target variable for selection				Compare different populations evaluated for selection in different	locations				Study diversity structure and identify parents to	good complementarity or similarity for some traits	<b>Cbjectives</b>