D1.4 Methodologies for the combined analysis of measures in different phenotyping setups for whole plant architectures based on common experiments

Xavier Draye (UCL)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731013. This publication reflects only the view of the author, and the European Commission cannot be held responsible for any use which may be made of the information contained therein.

Document information

EU Project N°	731013	**Acronym**	EPPN2020
Full Title	European Plant Phenotyping Network 2020		
Project website	www.eppn2020.plant-phenotyping.eu		

Deliverable	**N°**	D1.4	**Title**	Methodologies for the combined analysis of measures in different phenotyping setups for whole plant architectures based on common experiments
Work Package	**N°**	JRA1/WP1	**Title**	Novel techniques and methods for environmental and plant measurements, including model assisted phenotyping

Date of delivery	**Contractual**	31/10/2021 (Month 54)	**Actual**	10/11/2021 (Month 55)
Dissemination level	X	**PU Public, fully open, e.g. web**		
		CO Confidential, restricted under conditions set out in Model Grant Agreement		
		CI Classified, information as referred to in Commission Decision 2001/844/EC.		

Authors (Partner)	UCL			
Responsible Author	**Name**	Xavier Draye	**Email**	xavier.draye@uclouvain.be

Version log			
Issue Date	**Revision N°**	**Author**	**Change**
25/10/2021	0	Xavier Draye	First version
08/11/2021	1	François Tardieu	Validated version

Executive Summary

An essential objective of the EPPN2020 Joint Research Activities (JRA) was to improve the quality of plant phenotyping experiments in such a way that a broad scientific community, not necessarily involved in conducting the experiments, would be able to find, understand and re-interpret the results in a new context, typically to address new questions, use or develop new methods, often combining the results of different experiments. During the project, JRAs have developed new concepts, methods and tools towards this objective and intense communication was maintained across JRAs to keep this alignment.

The specific role of JRA1.4 was to generate a case study where the three JRAs could collectively evaluate their recommendations and discuss their options. JRA1.4 was designed by the consortium as a common multi-platform experiment involving most EPPN2020 installations. The activity started in year 2 of EPPN2020 and fully reached its objective of collectively testing methods with a common case study. At every step of the experiment, from the design phase to the exploitation of the combined datasets, the methods and tools developed within the three JRAs have been systematically applied. The reciprocal interactions between JRA contributors and with the participating partners has led to the identification of areas of improvement and stimulated the continued improvement of JRA methods, until the very end of the project.

JRA1.4 has helped identifying the methods that can be readily disseminated in a broader community and those which still require further development. JRA1.4 has therefore been instrumental in demonstrating how the methodological developments of the JRAs have allowed setting a strong basis for the establishment of the EU plant phenotyping infrastructure EMPHASIS.

This Deliverable provides indicators of (i) the work that has been achieved within JRA1.4, (ii) explains how the activity has developed during the project and (iii) illustrates how the recommendations and tools of the three JRAs have been jointly implemented and modified. A side, but important, output of the cross-platform experiment was also to create a common dataset, based on sound scientific hypotheses, whose full analysis already began but will be continued over years after the project.

Table of contents

1. DESIGN AND COURSE

1.1. Design principles

According to the DoA, JRA1.4 was planned as a set of two multi-platform experiments, on root and shoot phenotyping, respectively, with little connections between the two sets of experiments and only a tentative reference to the three JRAs.

In line with the development of JRA1, 2 and 3 during the first year of the project, and based on a critical appraisal of the difficulties met during the previous EPPN project, the initial (naive) plan for JRA1.4 has been revisited. In the initial description, the focus was mainly placed on the scientific objectives. In the revisited plan, the focus has been moved to the design of the task in order to secure the possibility to evaluate practically the options implemented the three JRAs. The scientific objectives have also been made stronger (see 3.1).

The objectives of the three JRAs are reminded below:

- *JRA1* developed methods at the measurement side (environmental characterization and sensor calibration, see D1.2) enabling the comparison (and combination) of plant phenotypic datasets obtained in different setups (experiments, sensors, installation);
- *JRA2* developed statistical tools and software to design phenotyping experiments, to check data quality (outliers identification), to apply spatial corrections to individual measurements and to conduct multi-trait and multi-platform analyses of the combined data;
- *JRA3* developed concepts and software to organise and produce datasets that obey the FAIR principles, thereby enabling the exchange and combination of datasets.

The model of a multi-platform experiment was appropriate to evaluate these methods and tools. In particular, partners who joined the experiment had very diverse background in terms of (i) experimental conditions, (ii) sensor configuration and equipment, (iii) expertise in experimental design and data analysis and (iv) data management procedures.

The concepts of *levels* defined within each JRA have been instrumental to ensure that all partners contributing to the experiment would match minimum requirements.

The details of the interactions with the JRAs are explained in Section 2 of this document, including an appreciation of how easy or complicated the adoption and implementation of the different methods can be.

1.2. Adhesion of partners

According to the DoA, four partners were expected to contribute to the experiment. After two rounds of discussion during EPPN2020 annual meetings, 12 partners finally asked to be part of the experiment, without extra funding. All of them had previously shown to comply with the level 1 rules defined in the three JRAs. If the motivation of some partners was to contribute to the scientific hypotheses, the enthusiasm of most of them was explained by the training opportunities provided by the experiment. It was, indeed, clearly announced that contributors of the three JRAs would be along during the course of the experiment to help partners in the implementation of the methods and to take their feedback into account.

The following 12 partners joined the process: ABER, ALSIA, FZJ, INRAE (PHENO3C, 4PMI, M3P and DIAPHEN installations), IPK, MTA, NaPPI, PHENO3C, SPPU, UCLouvain, UCPH, UNOTT, which represented 15 installations. Two installations were not able to reach the end of the process, due to platform maintenance issues (UNOTT) or severe climatic conditions (PHENO3C).

The range of installations was therefore very broad in terms of environmental conditions (from growth chambers to fields), experimental setup and types of substrates (aeroponics, rhizotubes, rhizotrons and pots), phenotyping capabilities (type of sensor, time and spatial, resolution, number of dimensions) and acquaintance with statistical and data management tools and standards. For a multi-platform experiment, one could see this as a "worst case scenario", thereby very well suited to grasp the reality of many platforms.

1.3. Data management

Given the diversity of partners background in terms of data management standards, a common approach has been adopted to ensure that the 13 dataset would be available and in the appropriate format at the end of EPPN2020: (i) each partner would manage its dataset locally with its own procedures and (ii) provide a version of the dataset in a simple format that is accessible to all. Hence, the principles at the base of JRA3 were considered by all partners (level 1), but an installed and operational information system was not taken as a requirement to participate to the multi-platform experiment, because this would have limited the number of participants.

A group of scientists involved in all three JRA met at several occasion to analyse various options, after having assembled a list of expected variables and their nature (e.g. time-series, dimensions, units, method). The group selected the ISATAB format, which is a text files container archive complying with the MIAPPE specifications for experimental metadata. ISATAB also comes with a R package to easily extract data from the archive.

Storage-wise, it was agreed to store the ISATAB files of the 13 experiments on a Google drive shared between all JRA1.4 contributing partners. This temporary storage solution was selected for its simplicity and accessibility to all partners.

1.4. Course of the experiment

Once the (three) definitive scientific hypotheses of the experiments were established (see 3.1), each platform was assigned to one hypothesis depending on throughput and phenotyping specifications.

The interest for improving the environmental characterisation to level 2 (maps of light and temperature, JRA1.1) was discussed with each partner, where relevant. Three partners (INRAE-Dijon, NaPPI and UCLouvain) engaged in this optional venture. Another set of 8 partners (ABER, ALSIA, FZJ, HMGU, IPK, MTK and UHEL) engaged in the cross-platform camera calibration prototype developed in JRA1.3.

With the help of JRA2 scientists, partners developed an experimental design for their own experiment. The design was selected to be suitable to their platform specifications and to the analysis of the whole multi-platform dataset.

The logistics of the experiment was managed by INRAE staff who prepared and distributed the seed material to individual platforms as soon as they were ready to start the experiment. The experiments were conducted between 2018 and 2021.

In June 2021, the data template was finalised and presented to all partners during the annual meeting (see 1.3). It was then decided that all contributing partners would try the JRA2 tools (spatial correction and outlier detection) on at least two of their phenotypic variables and, for these variables, provide the raw and the corrected version of the data, for later inspection by JRA2 scientists.

A training session followed by two weeks of "hotline" support was organised in July 2021 to ensure a dynamic kick-off of the data collection. At the time of writing of this deliverable, the datasets of the 13 experiments are, at least partly, available in the shared drive and partners continue interacting with the JRAs contributors to finalise this part in the shortest possible timeframe.

2. INTERACTIONS WITH THE THREE JRAS

Retrospectively, JRA1.4 has offered many opportunities for reciprocal interactions between the JRAs and with the partners:

- partners have received close custom training on the tools developed in EPPN2020;
- while interacting individually with partners, JRA contributors have better realized the diversity of situations and have received useful feedback which was helpful to improve their tools;
- the three JRA have exchanged intensively on issues related to data management, JRA1 helping to understand the diversity of variables and JRA2 to set the requirements for smooth analysis and exploitation of data in a multi-platform context.

The following sections detail the different JRA methods and tools that were used in jRA1.4 and benefited from the JRA1.4 experience.

2.1. JRA1

JRA1 developed, tested and disseminated methods and practices for proper environmental characterization (see D1.2). Within JRA1.4, it had to be further ensured that these environmental data would be shareable in the context of a multi-platform experiment (e.g. units conversion, variable definition). With the help of partners, a wiki-like document containing a list of sensors used in EPPN2020 and recommendations for their usage was established. The tutorials for light and temperature mapping were assessed at three installations and avenues for improvement have been identified to ease the dissemination of these methods.

JRA1 also tested in real size its calibration strategy to allow measurements on similar objects, obtained in installations with different imaging specifications (e.g. camera position, focal length) to be converted consistently into absolute real-life units (e.g. meter). The previous EPPN project pointed to the need for suitable camera calibration procedures, but previous attempts (e.g. moving plants across platforms, fake plants) to evaluate the magnitude of resulting errors failed largely. The novel strategy proposed by JRA1, based on artificial images (see D1.3), has been tested thanks to the contribution of 8 JRA1.4 partners and has allowed the cross-calibration of their cameras, demonstrating the validity of the new method.

2.2. JRA2

JRA2 has developed the Design Generator, a web-based application to generate experimental design suitable for plant phenotyping experiments (see D2.5). JRA1.4 has convinced many partners to test the application. In return, the feedback provided to the web developers has been a major source for improvement and further developments of the application.

JRA2 has also developed an R package (StatgenHTP) for quality testing of phenotypic data (e.g. outlier detection) and for the correction of spatial trends in greenhouse and growth chambers. The different steps of the analysis pipeline had been handled one by one during the project, but JRA1.4 has given an opportunity to test the pipeline at partner sites. Setting up the whole pipeline did not reveal any major difficulty, thanks to the rich educational material developed during the project and tested in different courses (within and outside EPPN2020)

Yet to come is the test of novel multi-trait and multi-platform linear models developed at WUR, for which the JRA1.4 dataset will be instrumental.

2.3. JRA3

JRA3 established a set of minimum requirements to enable the safe combination of plant phenotyping datasets on the same biological material but from different platforms. It turned out during the project that the underlying concepts were not easy to understand, even inside the EPPN2020 community, presumably due to the very different levels of understanding of data management concepts. Accompanying JRA1.4, JRA3 scientists developed very practical tools (online variable and URI generators) that have considerably helped participants to take ownership of these concepts. This is a (major) first step towards the adoption of EPPN2020 information systems and MIAPPE specifications by more partners.

Through interacting with JRA1 and JRA2 in the context of JRA1.4, JRA3 scientists have proposed adjustments to the MIAPPE specifications that are included in the MIAPPE2 that was recently issued, as well as to the ISATAB archive format to include a larger set of variable types.

JRA1.4 also helped JRA3 contributors to realize the breadth of the challenge of deploying information systems at many sites. According to JRA1.4 return, one of the main limiting factors is the time needed upstream of the deployment to design local implementations of the EPPN2020 information systems.

The exchanges between all JRAs and partners during the data collection step indicated that the ISATAB format might be adopted by several partners as a preferred way for long-term storage of plant phenotyping datasets. However, the time that is required for manual re-formatting of datasets by each partner demonstrated the value of having an information system at each site.

3. EXPLOITATION OF RESULTS

3.1. Scientific hypotheses

Testing scientific hypotheses was not the primary objective of the multi-platform experiment in the context of EPPN2020. However, the perspective of a challenging scientific objective, of the availability of a common dataset allowing further work (beyond EPPN2020) and of publishing high level papers was a potent incentive for partners to participate in this task. In year 2 of EPPN2020, a two day seminar was organised with 8 partners to elaborate robust scientific hypotheses that would frame the multi-platform experiment, next to the methodological objectives detailed in sections 1 and 2 of this document. The partners agreed on three working hypotheses to be texted, presented below.

The first scientific objective addressed by the multi-platform experiment was to assess the number of independent phenotypic variables (e.g. dimensionality) required to describe root system architecture. The literature reports many representations (e.g. topological, geometric, model-based) but comparisons of the information content of different representations is lacking. This question is not as naïve as it may seem, because most root phenotyping platform are designed with a specific representation in mind. Two out of the 13 experiments addressed this question, using a large number of replicates (n = 30).

The second scientific objective of the multi-platform experiment was to assess the loss of information that is caused by the rhizotron observation method, where only a fraction of the root system is visible to the imaging equipment. Four out of the 13 experiments addressed this question.

The third scientific objective of the JRA1.4 experiment was to evaluate to what extent part of the shoot growth response to substrate-induced water limitation is mediated by root architecture. All experiments are involved in this objectives, with six platforms measuring root architecture (see objectives 1 and 2) and seven platforms measuring shoot growth responses to water limitation in pots or in the field.

A common biological material was selected for the three objectives, comprised on 15 inbred maize lines (*Z. Mays*) and their corresponding hybrids with a common tester line. This material was proposed based on previous experiments (more than 30 field and platform experiments) whose data is also available to JRA1.4.

3.2. Next steps

At the time of writing this document, the multi-platform experiment has generated an open dataset grounded on solid scientific aims. The hypotheses testing will start after the project with a subset of partners committed to exploit these results beyond the practical achievements discussed above. The experience of several partners from previous projects (e.g. the FP7 project DROPS) shows that carefully designed and conducted experiments lead to publications during many years after the end of the project. The case of the JRA1.4 experiment will be a demonstrating one, given the efforts made by all JRAs to ensure the feasibility of statistical analysis across platforms.

4. CONCLUSIONS

In the context of JRA1.4, all principles developed inside JRA1, 2 and 3 have been applied to a common experiment conducted in 13 phenotyping installations with very different backgrounds. The reciprocal interactions between JRA contributors and with the participating partners has led to the identification of areas of improvement and stimulated the continued improvement of JRA methods, until the very end of the project.

At this stage, JRA1.4 has helped identify the methods that can be readily disseminated in a broader community and those which still require further developments. JRA1.4 has therefore been instrumental in demonstrating how the methodological developments of the JRAs have allowed setting a strong basis for the establishment of the EU plant phenotyping infrastructure EMPHASIS.

Glossary

EPPN: European Plant Phenotyping Network (EU project, 2012-2015)
EPPN2020: European Plant Phenotyping Network – 2020 (EU project, 2015-2021)
MIAPPE: Minimum Information About a Plant Phenotyping Experiment (miappe.org)