Project No. *283496*

transPLANT

Trans-national Infrastructure for Plant Genomic Science

Instrument: **Combination of Collaborative Project and Coordination and Support Action**

Thematic Priority: FP7-INFRASTRUCTURES-2011-2

D3.2
Format specifications for data exchange by flat file and web services

Due date of deliverable: 31.8.2013 (M24)
Actual submission date: 20.8.2013

Start date of project: 1.9.2011 Duration: 48 months

Organisation name of lead contractor for this deliverable: EMBL-EBI

Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission	
RE	Restricted to a group specified by the consortium (including the Commission	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Contributor
EMBL in collaboration with IPK, INRA, IPG PAS, Biogemma, DLO, Keygene

Introduction

Deliverable reference number:D3.2

Content

1. Introduction

In recent years, increasing efforts have been directed at characterizing the phenotypes of crop plants, such as morphological traits or adaptation to environment and diseases, measured in field or greenhouse experiments. Traditional "phenotypic" experiments have long been performed in field or greenhouse trials suited for specific conditions of plant material and experimental factors. However, due to the widely acknowledged need for a systems biology approach and the fast development of new high-throughput measurement protocols, new "-omics" data are now being collected in large experiments with a factorial structure, including repeated measurements in time. Many of the recent phenotypic studies involve also measuring results of gene action at the transcriptional, protein or metabolite level, or features of chromatin e.g. through the characterization of DNA-protein interaction events or chromatin states, or image data.

Although the laboratory protocols for different measurements require very different data collection schemes and device-specific preprocessing algorithms, in every case the data have to be collected for clearly identified "samples" or "experimental units" to allow their integration with existing knowledge about the system under study. To achieve this goal we are developing recommendations concerning the annotation of data sets containing phenotypic observations with meta data, to support integrative analysis on phenotypes and "-omics" data measured at different levels of plant organization. We do this with recognition of standardization achievements already undertaken in particular areas of "-omics" studies such as transcriptomics, proteomics, and metabolomics. However, these are domain-specific and generally do not allow for inclusion of morphological, yield-related, quality or resistance traits. In consequence, we propose the "Minimum Information about Plant Phenotypic Experiments" (MIAPPE) checklist, which is based on existing "Minimum Information" recommendations. Furthermore, we have developed an implementation of this standard in the form of a file format for exchange of phenotypic information between databases, web services and data analysis tools. The format is based on the ISA-TAB structure, which has already been applied successfully in other biological domains. Our ISA-TAB implementation is compatible with the structures used in other plant science applications. The presented standards and recommendations will be used by the transPLANT consortium to develop tools for data storage, exchange, information retrieval and integrated data analysis of phenotypic data.

The transPLANT consortium needs, for its own purposes, to define a set of internal data standards for the exchange of this data (which is at present not centrally archived and is not consistently recorded: developing the Ephesis repository, maintained at INRA, as a resource for such data is a deliverable in work package 6. But in addition to use within the consortium, we also intend to discuss these standards with a wider community, comprising our collaborators and other groups active in this rapidly evolving area (e.g. plant phenotyping centres, crop breeders involved in field trials, etc.) We are working in conjunction with other projects also interested in this area with the goal of establishing a common standard. For example, we are already in discussion with the developers of the Trait Ontology; with the European Plant Phenotyping Network; with Bioversity International; and with various European consortia interested in this domain. We will push these activities forward over the remaining period of the grant.

Methods
Literature and Internet studies Minimum Information approach Data formatting Implementation of standards

Results (if applicable, interactions with other workpackages)

2. Proposition of requirements for Minimum Information About a Plant Phenotypic Experiment (MIAPPE)

In order to describe standardized schema for keeping information about phenotypic experiments we consider the following elements:

a) The plant phenotypic experiment as an entity.

b) Minimum information requirements.

c) The set of ontologies necessary to annotate a phenotypic experiment.

d) A data format to store and exchange information.

In this report we concentrate on points a, b and d, as point c is the subject of report on deliverable D3.1.

By a "plant phenotypic experiment" we understand a set of experimental units, organized in an experimental design, in which a set of biological materials ("biosources") is studied through exposure to treatments at some level (Table 1). By "phenotype" we understand any quantitative or qualitative trait (transcript level, protein level, metabolite level, morphology trait, image phenotyping trait, yield component, etc.) measured on samples taken from the biosources.

Each row in the description of the experiment corresponding to a sample (measurement) contains a number of attributes. The list of attributes that must be given for the appropriate understanding of data is called the "checklist". The checklist should be consulted when preparing the data set for deposition or exchange. It is especially important that the description of the experiment includes not only the attributes that vary over the units, but also the ones that are constant, but are necessary for understanding of the experiment and future integration of data with another datasets.

When designing the checklists for plant phenotypic experiments, we took a set of requirements, as follows:

- the attributes must describe adequately:
    - study,
    - environment,
    - biosource,
    - treatments,
    - experimental design,
    - sample collection, processing, management,
    - phenotypic traits: type, measurement protocol, processing protocol, scale, units,
- the attributes must be able to describe both designed experiments and observational studies,
- existing standards must be re-used as much as possible; standards which have been successfully adopted (for ontology annotation and data formats) are preferred.

As a result of literature and Internet studies, it was decided to form the Minimum Information about Plant Phenotypic Experiment according to the structure shown in Table 2.

Existing minimum information standards, that were considered as possible sources of attributes are listed and characterized in Table 3. The final choice of standards is given in Table 4. In this table, a "raw data file"(possibly one per sample) contains raw observations in a device-specific format. A "derived data file" contains observations processed by application of the specific protocols. A "sufficient data file" contains functions of observations (statistics) necessary and sufficient to draw conclusions from the experiment. The minimum information requirements for phenotypes (phenotypic traits) consist of three attributes: Name, Method and Scale. This list is based on the "Trait/Method/Scale" triplet approach

applied by Generation Challenge Program Crop Ontology (http://www.cropontology.org, Shrestha et al., 2012).

3. Proposition of data exchange format

Selection of the data exchange format was based on the following requirements:

- the data format must allow for definition of broad spectrum of phenotypic traits,

- the data format must support annotation of all attributes and their values by ontologies, controlled vocabularies or databases,

- the data format should be sufficiently human-readable to allow for manual preparation by biologists (possibly with the help of some dedicated software),

- the data format should allow for separation of metadata valid for the whole experiment from attributes specific for its parts,

These requirements are fulfilled by ISA-TAB format (www.isa-tools.org, Sansone and Rocca-Serra 2012). This format utilizes a hierarchical Investigation/Study/Assay structure, which can be used for both simple experiments and large projects performing a number of assays on common biological materials. Two important advantages of the format are (i) ISA-TAB has already been accepted by some communities or databases in similar research areas (see e.g. the Genomespace platform, www.genomespace.org), (ii) it can be used for integration of different types of traditional phenotypic data with new "-omics" observations. It supports annotation: the values to be annotated can be stored in the form of a triplet comprising "a term, a reference for the source of the term, and a database accession number". The dataset in ISA-TAB format consists of a collection of text files, so can be prepared in any text or sheet-format editor; special software supporting file preparation, data input and annotation is available (ISACreator, ISAConfigurator).

The proposed implementation of ISA-TAB format for phenotypic data:

- is based - at the investigation and study level - on the configurations developed by metabolomics groups concentrated around Metabolomics Standard Initiative and Metabolights database (Haug et al. 2012),

- defines a new "assay" configuration specific for the needs of phenotyping experiments called "phenotyping assay"; the resulting ISA-TAB structure at assay level is shown in Fig. 1; the rows of ISA-TAB study, assay and derived data files are linked by either "source" or "sample" attributes,

- defines a "trait definition file" (tdf), which contains descriptions of the phenotypic traits.

4. Implementation process in 2011/13

The collective work in transPLANT WP3 on standards for phenotyping was carried according to the planned schedule:

1. A meeting was held in Hinxton, 8th-9th December 2011, bringing together members of the transPLANT project with breeders and ontology developers from Europe and America to discuss the potential for the extension of existing ontologies for the annotation of field trials and other agricultural applications. The participation of transPLANT consortium members in this discussion has informed subsequent developments within the project.
2. Partner IPG PAS distributed the document describing MIAPPE and its practical implementation in ISA-TAB format to all partners by 28 February 2013.
3. WP3 (and other transPLANT partners interested) tested the solutions (at least one data set) and sent comments on it to IPG PAS by 31 March 2013.
4. A corrected version of the document was sent to partners by 30 April 2013.
5. The Consortium agreed on the document on 31 May 2013.

The document will be promoted at two conferences: EPSO 2013 in Greece (1-5 September) and Phenodays 2013 in Holland (16-18 October). It will be distributed to selected units (people) outside the project and opinions will be collected by the end of 2013.

5. Testing phase results

During tests among partners, no fundamental problems emerged concerning the MIAPPE formulation. One can attribute this to the fact that MIAPPE consists mostly of standards accepted after thorough consultations by other application fields.

For data formatting, datasets were obtained from: IPG PAS, Keygene, WUR, IPK and INRA. Below we collected some comments made during testing and discussions with partners. An ISA "phenotyping" configuration file was developed after consultation with the Metabolights database developers (a resource for metabolomics data being developed around ISA-TAB standards), and has been sent for evaluation to ISA-TAB group.

Dataset 1: IPG PAS test example

Artificial data on 5 barley genotypes, block design, 3 phenotypic traits

This example was used to demonstrate the features of the data format at project meetings and teleconferences.

Dataset 2. Keygene – Arabidopsis image phenotyping

7273 measurement points, 93 biosources (NASC accessions), 2-4 replications, 192 pots, 39 days of measurement (incomplete or overcomplete for pots x days), 18 traits derived from raw dataset.

This example served, in particular, for discussion on the following topics:

- if a special ISA-TAB configuration for time phenotyping is needed; it seems that "time" or "repeated measurement" can be considered as a factor in the experiment and data from experiments in time can be formatted using the general configuration; what is important is its proper annotation so that the correct data modelling is applied.

- several attributes must be used to preserve annotation to experiment – pot label, analysis_Id, etc.,

- for image phenotyping there is a basic problem with annotation of traits – there is no ontology of low-level features obtained from the image recognition process (e.g. expressed in pixels).

Dataset 3. WUR (DLO) - Arabidopsis gene expression by RNA-Seq

19 biosources (NASC Arabidopsis accessions), 2 replications, 33602 traits (100 formatted) – expression of genes obtained by RNAseq as RPKM values from:

http://mus.well.ox.ac.uk/19genomes/gene_expression/expression_tables/MERGED_COMBINED_COUNTS_1_04_04_2011.tab

This example concerns gene expression data obtained by RNA-Seq. Note that for the gene expression data there is another profile (configuration) in ISA-TAB, for "transcription profiling assays". This profile serves as an updated version of MAGE-TAB, the format used for submission of microarray data. For this exercise we used the "phenotypic" profile to test it. The difference between the two profiles is in the metadata. Also, in "transcription profiling" configuration it is assumed that all traits are expression values, so there is no need for special description of traits (tdf file).

Dataset 4. IPK – evaluation gene bank data

2416 accessions, no replications, 42 traits

In this example, the challenge may be to annotate traits; if they are not available in the reference TO, gene bank definitions could be used as the source of trait names.

Dataset 5. INRA – exemplary data from EPHESIS database

8 accessions of wine grape studied for 2 phenotypic traits, a fragment of data from a replicated experiment in a block design, with replicated sampling within plots.

This example was used also to test ISACreator software. When using ISA Creator ver. 1-7 program, a problem was encountered: files created under "phenotyping" configuration became corrupted at some point during processing by the ISA program. The problem was not located, but it seems that it concerns also other configuration (default ISA-TAB); it also seems that the problem does not exist in ISA Creator version 1-6. Version 1-6 is adequate for use until this problem is fixed.

The general problems raised during discussion on MIAPPE and ISA-TAB were:

1. How to propose and discuss application of ISA-TAB format with FPPN, EPPN and PODD groups.
2. Conversion from Ephesis to ISA-TAB file format (and back).
3. Strengths and weaknesses of those formats:
    a. Strengths:
        i. the ability to explicitly integrate investigation>study>assay in a single archive rather than multiple files,
        ii. a community supporting the format in related areas of biology and plant sciences,
        iii. clear separation of data and metadata,
        iv. clear handling of time series.
    b. Weaknesses:
        i. unstable ISACreator tool – but files can be created independently of the tool,
        ii. metadata in tab format which is not easily parsed; XML or RDF approach like for PODD might be better suited – the tools for ISA-TAB to RDF conversion should be studied.

6. Interactions with other packages

Work in WP3, in particular on this deliverable D3.2, is interacting with:

- WP 2: the developed standards were proposed as solutions to some projects collaborating with transPLANT partners IPG PAS, INRA.

- WP 4: the standards were promoted at the training workshop organized by transPLANT in Poznań in June 2013 and by EU ITN project Epitraits in Amsterdam in March 2013.

- WP 6: the proposed data format will be used in the phenotypic data search system.

- WP 10: with respect to definition and content of the "sufficient data file".

7. Conclusions

We have concluded that the standardization of phenotypic data should be based on the "minimum information" approach. The "Minimum Information about a Plant Phenotypic Experiment "requirements have been proposed. In the metadata layer, they are mostly based on existing standards developed and accepted in other areas of experimental plant science. For data content, new requirements have been described which are suited to the special situation of phenotypic experiments, in which many traits of different nature are measured. As the data exchange format the existing ISA-TAB has been chosen. Its main advantage is the existence of a community promoting it in other experimental applications. This will permit not only exchange of phenotypic data, but also integration with other dataset prepared in "omics" studies. Fundamental discussion on applicability of MIAPPE and phenotypic ISA-TAB configuration within the transPLANT consortium has been completed. The solutions will be now promoted outside the project by presentations at selected meetings and discussions with invited scientists.

References

Shrestha R. et al. (2012). Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the CropOntology developed by the crop communities of practice. Frontiers in Physiology 3: 326.
Sansone SA, Rocca-Serra P. (2012). On the evolving portfolio of community-standards and data sharing policies: turning challenges into new opportunities. GigaSience 1: 10.
Haug K et al. (2012). MetaboLights - an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic Acid Research 41: 781-786.
Brazma A. et al. (2001). MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. Nature Genetics 29: 365-371.
Zimmerman P et al. (2006). MIAME/Plant – adding value to plant microarrray experiments. Plant Methods 2: 1.
Fiehn O et al. (2007a). The metabolomics standards initiative (MSI). Metabolomics 3: 175-178.
Fiehn O et al. (2007b). Minimum reporting standards for plant biology context information in metabolomic studies. Metabolomics 3: 195-201.
Morrison N et al. (2007). Standard reporting requirements for biological samples in metabolomics experiments: environmental context. Metabolomics 3: 203-210.
Rocca-Serra P et al. (2010). ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. Bioinformatics 26: 2354-2356.
Yilmaz P et al. (2011). The genomic standards consortium: bringing standards to life for microbial ecology. ISME Journal 5: 1565-1567.
Taylor CF et al. (2007). The minimum information about a proteomics experiment (MIAPE). Nat Biotechnol 25: 887-893.

Publications

In preparation

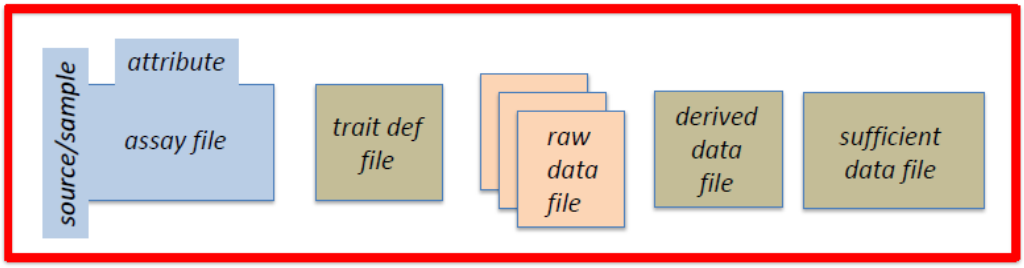

Fig. 1. The files constituting phenotyping assay data set

Table 1. The elements of plant phenotypic experiment

	Sample identifier	Description of block structure		Description of biosource	Description of treatments	Observed phenotypic traits		
Symbol:	**id**	**B**		**S**	**T**	**Y**		
Meaning:	Vector of unique sample (measurement) identifiers	Columns of matrix **B** describe **design of experiment**, contain codes (allocation of samples to experimental units, plots, pots, blocks, rows, columns etc.)		Columns of matrix **S** describe the genotypes used in the experiment (species, varieties, accessions, etc.)	Columns of matrix **T** describe allocation of experimental units to levels of **factors**: conditions, time points, treatment regimes, etc.	Columns of matrix **Y** contain **phenotypes** observed for samples taken from corresponding experimental units, entries are real numbers or category codes		
Example:	**id**	**Unit**	**Replication**	**Variety**	**Drought**	**Yield**	**Lodging**	**Proline cont**
	1	1	1	Maresi	I	1.34	a	12.56
	2	2	2	Maresi	I	2.45	a	12.67
	3	3	3	Maresi	I	1.65	a	13.78
	4	4	1	Morex	C	1.67	b	16.12
	5	5	2	Morex	C	2.56	b	18.34
	....	...	...	...	...	...	...	...
	n	n	3	CamB1	C	1.78	b	19.23

Table 2. Proposed structure of MIAPPE

Checklist part	**Proposition**
Study	Accept existing list(s)
Environment	Accept existing list(s)
Biosource	Accept existing list(s)
Treatments	Accept existing list(s)
Experimental design	New list (gap)
Sample: collection, processing, management,	Accept existing list(s) + additional attributes (gap)
Phenotypic traits: values, type, measurement protocol, processing protocol, scale, units	New list (gap)

Table 3. Minimum information standards relevant to plant phenotypic experiments

Minimum information document (initiative)	**Microarrays** MIAME (MGED)	**Metabolomics** CIMR (MSI)	**Sequence** MIxS (GSC)	**Proteomics** MIAPE (PSI)
Full name (initiative), publication, document	Minimum Information about a **Microarray** Experiment (Microarray Gene Expression Database Group), Brazma et al. (2001) http://www.mged.org/Workgroups/MIAME/miame_2.0.html	Core Information for **Metabolomics** Reporting (Metabolomics Standards Initiative), Fiehn et al. (2007a) http://msi-workgroups.sourceforge.net/	Minimum Information about **any sequence** specifications (Genomic Standards Initiative) Yilmaz et al. (2011) http://gensc.org/gc_wiki/index.php/MIxS	Minimum Information about a **Proteomics** Experiment (Proteomics Standards Initiative) Taylor et al. (2007) http://www.psidev.info/groups/miape
Relevant extensions, publication, checklists	**MIAME/Plant** Zimmerman et al. (2006) http://www.mged.org/Workgroups/MIAME/MIAME-plant_Dec2005.pdf	**CIMR: Plant Biology Context ,** Fiehn et al. (2007b) http://msi-workgroups.sourceforge.net/bio-metadata/reporting/pbc/doc.rtf **CIMR: Environmental Analysis Context** Morrison et al. (2007) http://msi-workgroups.sourceforge.net/bio-metadata/reporting/env/reporting-requirements/ECWSG_reporting_requirements_v1.rtf	MIxS **Plant-associated environmetal package** Yilmaz et al. (2011) http://gensc.org/gc_wiki/images/9/90/MIMARKS_26_01_11.xls	None (only assay-specific documents)
MIBBI project http://mibbi.sourceforge.net/portal.shtml	Checklists stored	Checklists stored	MIGS, MIMIS checklists partially stored, link to full MIMARKS list at GSC site	No own checklists, checklists of MIBBI used
Compliant data exchange format	MAGE-TAB Rayner et al. (2006)	ISA-TAB Rocca-Serra et al. (2010)	Several, depending on the database	PRIDE XML (?)
Databases able to store compliant information	**ArrayExpress** (EBI) – MAGE-TAB, spreadsheet submission, online tool submission **Gene Expression Omnibus** (NCBI) – various formats, and tools	**MetaboLights** (EBI) – ISA-TAB	**Genebank** (NCBI) – various formats and tools **Sequence Read Archive** (NCBI) – SRA XML, various tools	**PRIDE** (EBI), PRIDE XML, PRIDE Converter 2 (no quantitative data)

Table 4. Proposed checklists for MIAPPE elements

Checklist section	Checklist accepted	Attributes	No. of attributes (1, m - many)	Remarks	Alternative checklist
Study	-	identifier title description submission date public release date design type publications	1 1 1 1 1 1 m	List accepted from default ISA-TAB tools configuration	-
Environment	CIMR: Environmental Analysis Context	Table 4.1	m	As at MIBBI, some attributes removed	MIxS environment
Biosource	MIxS: Plant-associated environmental package	Table 4.2	m	As at GSC, some attributes removed	CIMR: Plant BiologyContext, or MIAME/Plant
Treatments	MIxS: Plant-associated environmental package	Table 4.2	m	As at GSC, some attributes removed	CIMR: Plant BiologyContext, or MIAME/Plant
Experimental design	-	Exp. design attributes	m	New list	-
Sample collection, processing management	-	Table 4.2	m	List exist for assays in CIMR, MIAPE, MIMARKS (ISA-TAB configurations) + new attributes for phenotypic assays	-
Phenotype		Raw data file Derived data file Sufficient data file Processing protocols	1 1 1 m	Approach taken after metabolomics standards	

Table 4.1. Attributes for environment, biosource and treatments (1 or many)

Description of any field environment Geographic location Altitude or depth Habitat Meteorological conditions Lunar or solar phase All other measured parameters Description of any laboratory environment Laboratory address and contact details Description of terrestrial environment Inclination and aspect Substrate type Substrate temperature All other measured parameters	Description of aquatic environment Sample(s) was submerged and emerged (how deep and for how long in this condition)Water temperature Tidal phase All other measured parameters Description of atmospheric environment Atmospheric temperature All other measured parameters Description of biotic environment Description of host organism Relationship of organism(s) to host All other measured parameters

Table 4.2. Attributes for biosource, treatments and sample description

Biosource			non-mineral nutrient regimen		m
host taxid		1	radiation regimen	m	
infra_specific_name		1	rainfall regimen		m
infra_specific_rank		1	salt regimen		m
host common name		1	watering regimen		m
genotype		m	water temperature regimen		m
Treatments			standing water regimen		m
climate environment		m	pesticide regimen	m	
seasonal environment		m	pH regimen		m
air temperature regimen		m	perturbation		m
antibiotic regimen		m			
chemical administration		m	**Sample collection, processing, management**		
chemical mutagen		m	plant body site (organ)		1
disease status		m	age		1
fertilizer regimen		m	life stage		1
fungicide regimen		m	plant product		1
gaseous environment		m	organism count		m
gravity		m	temperature		1
growth hormone regimen		m	oxygenation status of sample		1
herbicide regimen	m		sample salinity		1
mechanical damage		m	sample storage duration		1
mineral nutrient regimen		m	sample storage location		1
humidity regimen	m		sample storage temperature		1

Appendix 1. ISA-TAB files for Dataset 1 (partial views)

Investigation

ONTOLOGY SOURCE REFERENCE			
Term Source Name	OBI	EO	NCBITaxon
Term Source File		http://www.ebi.ac.uk/ontology-lookup/	http://bioportal.bioontology.org/ontolo
Term Source Version		Jun 2010	47845
Term Source Description	Ontology for Biomedical Investigations	Plant Environmental Conditions	NCBI organismal classification
INVESTIGATION			
Investigation Identifier			
Investigation Title	Investigation		
Investigation Description			
Investigation Submission Date			
Investigation Public Release Date			
Comment [Created with configuration]			
Comment [Last Opened With Configuration]	Phenotyping		
Comment [Created With Configuration]	C:Documents and SettingspkraMoje dokumentyISA-TABConfigurationsPhenotyping		
INVESTIGATION PUBLICATIONS			
Investigation PubMed ID			
Investigation Publication DOI			
Investigation Publication Author List			
Investigation Publication Title			
Investigation Publication Status			
Investigation Publication Status Term Accession Number			
Investigation Publication Status Term Source REF			
INVESTIGATION CONTACTS			
Investigation Person Last Name			
Investigation Person First Name			
Investigation Person Mid Initials			
Investigation Person Email			
Investigation Person Phone			
Investigation Person Fax			
Investigation Person Address			
Investigation Person Affiliation			
Investigation Person Roles			
Investigation Person Roles Term Accession Number			
Investigation Person Roles Term Source REF			
STUDY			
Study Identifier	Phenotypic1		
Study Title	TRANSPLANT WP3 Phenotypic standards test		
Study Description	The study is aimed at testing standards for phenotypic data collection, formatting and annotation in transPlant proj		
Study Submission Date			
Study Public Release Date			
Study File Name	s_Study1.txt		
STUDY DESIGN DESCRIPTORS			
Study Design Type	randomized complete block design		

Study

Source Name	Characteristics[Organ	Term Source REF	Term Accession Num	Characteristics[Infra-s	Term Source REF	Term Accession Numb	Characteristics[Organism part]	Term Source REF
source1	Hordeum vulgare sub	NCBITaxon	xon_112509	Sebastian	EURISCO	500076	stem	PO
source2	Hordeum vulgare sub	NCBITaxon	xon_112509	Sebastian	EURISCO	500076	stem	PO
source3	Hordeum vulgare sub	NCBITaxon	xon_112509	Amarena	EURISCO	500079	stem	PO
source4	Hordeum vulgare sub	NCBITaxon	xon_112509	Amarena	EURISCO	500079	stem	PO
source5	Hordeum vulgare sub	NCBITaxon	xon_112509	Nagradowicki	EURISCO	500067	stem	PO
source6	Hordeum vulgare sub	NCBITaxon	xon_112509	Nagradowicki	EURISCO	500067	stem	PO
source7	Hordeum vulgare sub	NCBITaxon	xon_112509	HOR 198	EURISCO		stem	PO
source8	Hordeum vulgare sub	NCBITaxon	xon_112509	HOR 198	EURISCO		stem	PO
source9	Hordeum vulgare sub	NCBITaxon	xon_112509	Basza	EURISCO	500072	stem	PO
source10	Hordeum vulgare sub	NCBITaxon	xon_112509	Basza	EURISCO	500072	stem	PO

Term Accession Num	Protocol REF	Protocol REF	Factor Value[Treatment]	Term Source REF	Term Accession Number	
9047	drought application	sample collection	Control			
9047	drought application	sample collection	Drought			
9047	drought application	sample collection	Control			
9047	drought application	sample collection	Drought			
9047	drought application	sample collection	Control			
9047	drought application	sample collection	Drought			
9047	drought application	sample collection	Control			
9047	drought application	sample collection	Drought			
9047	drought application	sample collection	Control			
9047	drought application	sample collection	Drought			

Assay

Source Name	Sample Name	Factor Value[Block]	Term Source	Term Acces	Raw Data Fil	Protocol REF	Derived Data File	Trait Definition File
source1	sample1	1				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source1	sample2	2				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source1	sample3	3				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source1	sample4	4				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source1	sample5	5				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source2	sample6	1				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source2	sample7	2				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source2	sample10	5				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source3	sample11	1				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source3	sample12	2				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source3	sample13	3				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source3	sample14	4				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source3	sample15	5				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source4	sample16	1				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source4	sample17	2				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source4	sample18	3				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source4	sample19	4				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source4	sample20	5				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source5	sample21	1				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source5	sample22	2				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source5	sample23	3				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source5	sample24	4				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source5	sample25	5				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source6	sample26	1				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source6	sample27	2				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source6	sample28	3				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source6	sample29	4				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source6	sample30	5				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx
source7	sample31	1				data transformation	a_study1_processed_data.xlsx	a_study1_tdf.xlsx

Processed data file

Sample Name	Trait Value[len]	Trait Value[Colour]	Term Source REF	Term Accession Number	Trait Value[Stem diameter]
sample1	0,7221	green	PATO	320	1,08
sample2	0,4288	yellow	PATO	324	1,15
sample3	0,5585	green	PATO	320	1,46
sample4	0,1394	yellow	PATO	324	1,23
sample5	0,5313	yellow	PATO	324	1,23
sample6	0,8398	green	PATO	320	1,31
sample7	0,8084	green	PATO	320	1,62
sample10	0,1138	green	PATO	320	1,46
sample11	0,9038	green	PATO	320	1,00
sample12	0,1341	yellow	PATO	324	1,00
sample13	0,2490	yellow	PATO	324	1,69
sample14	0,0910	green	PATO	320	1,92
sample15	0,4247	green	PATO	320	1,38
sample16	1,9308	yellow	PATO	324	1,69
sample17	1,1836	yellow	PATO	324	1,00
sample18	1,6157	green	PATO	320	1,31
sample19	1,3986	green	PATO	320	1,62
sample20	1,7506	yellow	PATO	324	1,46
sample21	0,8363	yellow	PATO	324	1,85
sample22	0,7670	yellow	PATO	324	1,00
sample23	0,2674	green	PATO	320	1,38
sample24	0,3717	green	PATO	320	1,31
sample25	0,9469	yellow	PATO	324	1,69
sample26	0,4396	green	PATO	320	1,15
sample27	1,4546	green	PATO	320	1,46
sample28	1,8874	yellow	PATO	324	1,38
sample29	1,2617	yellow	PATO	324	1,31
sample30	1,6363	green	PATO	320	1,62
sample31	0,8264	green	PATO	320	1,69
sample32	0,9446	yellow	PATO	324	1,38
sample33	0,3708	yellow	PATO	324	1,62

Trait definition file

Trait Name	Trait Source REF	Trait Term Accession Number	Method Name	Method Source REF
len	TO	576	Stem length measuring method	BO
Colour	PATO	14	Color assessed visually by 2 specialists	
Stem diameter			Diameter measured in the middle	

Method Term Accession Nun	Scale Name	Scale Source REF	Scale Term Accession Number
12	cm	UO	15
	colour scale	PATO	14
	mm	UO	16