



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

D7.1

**A registry of plant genomic
information**

Due date of deliverable: 31.5.2012 (Month 9)

Actual submission date: 31.5.2012 (Month 9)

Start date of project: 1.9.2011

Duration: 48 months

Organisation name of lead contractor for this deliverable: HMGU

Project co-funded by the European Commission within the Seventh Framework Programme (2011-2014)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Contributor

HMGU in cooperation with EBI and proof-reading and addition of data from all other transPLANT partners.

Introduction

Deliverable reference number: D7.1

A registry of plant genomic information: The implementation of a registry for plant genomic information; a publicly accessible webpage with information about computational and interactive resources produced by transplant partners and other researchers for the genomes present in the transplant hub.

The number of plant genome sequencing and (re-sequencing) projects is increasing; and for many of the more complex cereal genomes, progress towards the complete assembly of a reference genome is incremental, with a variety of projects producing genome, transcriptome and marker-based data from a range of technologies. Drawing on the broad expertise of the consortium, we will maintain a registry of important sequence-based resources for species of agricultural and economic importance as well as model systems. More specifically model genomes such as *Arabidopsis thaliana*, *A. lyrata*, *Medicago* and *Brachypodium* will be included as well as the genomes of tomato, potato, wine, cucumber, maize and apple as well as the extremely challenging and complex grass and Triticeae genomes of wheat, barley and rye.

Methods

We collected repository data for publicly available plant genome database systems making use of both existing compilations and a de-novo search of relevant websites and systems. For instance, we imported all relevant data from the plant genome website collection at http://www.phytozome.net/Phytozome_resources.php and performed extensive web searches for registering both species-specific and multi-species plant genome resources maintained by both transPLANT and non-transPLANT partners.

Results (if applicable, interactions with other workpackages)

A total of 187 distinct plant genome resources is registered at the transPLANT data registry at this time. We collected and structured the following information entities from the identified (see Methods) plant genome resources:

ID this will be the internal database ID, incremental

Provider_Shortname short name of the data providers institution (such as research center, university...)

Provider_Details details (such as full name) of the data providers institution

Resource_Shortname short name of the particular resource provided...this will be a system name such as *Phytozome*

Resource_Details details (such as full name, scope) of the data resource

Instance_Name this will describe a particular instance of the resource, such as the database for a single species within the system (such as the *Arab. thaliana* instance within *Phytozome*)

Instance_Description details (such as full name) of the instance

Species_Name scientific species name....this can contain more than one entry

Species_Commonname common species name

URL primary URL of the particular instance (should point to the entry page of the instance such as to the *Ara. thaliana* entry page in *Phytozome*)

Version_release_name this field can hold a release or version tag for the data within an instance...often this will be an annotation or assembly version, such as *TAIR10*

source_name_URL this field can hold a source URL for the data within an instance...often the data within an instance are derivative and were obtained from another primary resource...this URL can then point towards the original data resource

last_updated_or_release this field can hold a release version or last update data for the instance OR/AND resource...many resources are built in releases (such as *TAIR* or *Phytozome* in version 8.0) or updated regularly

data_type this field describes the type of data that is stored or provided from an instance...such as "genomic", "variation" or "expression"

tools this field can hold the names of useful tools provided within the instance or resources such as *GBrowse*

All registry data was stored in a relational database system and provided to the transPLANT partners in Excel format for proof-reading and adding data. The registry is hosted, maintained and updated at HMGU at the following URL:

<http://mips.helmholtz-muenchen.de/plant/transplant/index.jsp>

It is also fully accessible for search, query and linking (e.g. using cross-references from other data entities produced or integrated within transPLANT) from the official transPLANT website hosted by EBI:

<http://transplantdb.eu/>.

To ensure both registries being synchronous we will exchange updates and changes to the master registry monthly. In the near future, changes and updates to the registry can be performed by database providers themselves, lowering the maintaining cost and ensuring expert-curated and -driven information. For that, a web interface hosted by HMGU will be provided with secure access to the transPLANT genomic resources registry.