

[Subscribe](#)[Share ▼](#)[Past Issues](#)[Translate](#)

transPLANT Newsletter Autumn 2014 Issue 2

[View this email in your browser](#)

trans-National Infrastructure for Plant Genomic Science

INSIDE THIS ISSUE

- [The transPLANT website: new data, new tools, easier navigation](#)
- [Training: new resources and events](#)
- [Two new data warehouses: GrapeMine and Wheat3BMine](#)
- [e!DAL - Store, cite and share research data](#)
- [Filling the gap between sequence and function - a bioinformatics approach](#)

COMMENTS?

- We want to hear from you! Any comments or suggestions please contact us at transplant_help@ebi.ac.uk

The transPLANT website: new data, new tools, easier navigation

The screenshot shows the transPLANT website interface. At the top, there is a navigation bar with links for Home, News, Tools, Training, Standards, and About. A search bar is located in the top right corner. Below the navigation bar, the page title is "trans-national infrastructure for plant genomic science". The main content area is divided into several sections:

- CURRENT SEARCH:** Found 37 results.
- FILTER BY DATABASE:**
 - Ensembl Plants (16)
 - PlantsDB (13)
 - OPTIMAS-DW (8)
- FILTER BY DATA TYPE:**
 - Sequence feature (16)
 - transcript (13)
 - Sequence (8)
- FILTER BY SPECIES:**
 - Arabidopsis thaliana (11)
 - Zea mays (8)
 - Medicago truncatula (6)
 - Brassic... subsp. pekinensis

The search results for "PAD4" are displayed, showing a search bar with "PAD4" entered and a "Search" button. The results for "MTR_8g094370" are shown, including a description, species (Medicago truncatula), and location (8:27218509-27222910). Below the search results, there is an integrated genome browser showing tracks for Sequence, Genes, and Transcripts. The browser displays a 27.22 Mb region with a red bar indicating the location of the gene MTR_8g094370. Callouts highlight key features: "One search accessible on all pages", "Results are divided by category (data objects, resource registry and the website)", "Integrated genome browser", and "Filter results by database and object type".

The transPLANT [website](#) has been reworked, to provide clearer access to the growing range of data and information available. New features include access to [Ephesis](#) and [LAILAPS](#); a growing library of training [videos](#), plus other training [resources](#) and [news](#); and a new genome assembly decision support system, to guide scientists in their choice of technologies and tools based on a systematic analysis of previous datasets. In addition, a simplified search interface now allows users to simultaneously search 10 different plant databases, the transPLANT resource registry and the transPLANT website itself. Direct access to a genome browser is now available from the search results.

Training: new resources and events

New documents in the training resources registry

To assist and train potential users of transPLANT services and resources who are unable to attend one of our workshops, all the transPLANT training materials from previous workshops are now available on the transPLANT portal at <http://www.transplantdb.eu/training-resources>. The materials cover a broad range of available transPLANT services and resources, including many hands-on training exercises and real-life use cases. Contact persons are given in each presentation to ensure help in case of questions. All training material from upcoming workshops and presentations will be deposited to this repository, too.

The screenshot shows the transPLANT website interface. At the top, there is a navigation bar with links for Home, News, Tools, Training, Standards, and About. A search bar is located on the right. Below the navigation bar, the main content area is titled "trans-national infrastructure for plant genomic science". On the left, there is a section for "SEARCH TRAINING RESOURCES" with a note that a faceted search is coming soon. The main content area features a "Training resources" section, which includes a paragraph about upcoming training courses and a list of attachments for a workshop on Triticeae genomics and phenotyping standards. The attachments are listed in a table with columns for "Attachment" and "Size".

Attachment	Size
Announcement, Application form, and Programme.	94.72 KB
An overview of Ensembl Plants.	2.12 MB
The Triticeae genomes in Ensembl Plants.	1.2 MB
A walk-through tutorial of Ensembl Plants functionality.	6.42 MB
Sample solutions to the tutorial.	140 KB

A new video material to train on barley resources



Genome data from many crop species such as wheat and barley present great challenges for data analysis but also for data integration and representation due to complex genetics, high repeat content and large genome sizes. As a result, data and resources tend to be distributed and the integration of highly heterogeneous data and data types is laborious and challenging.

transPLANT partners have made significant efforts to integrate and synchronize data for the complex triticeae genomes. To provide users with an idea about how transPLANT genome resources for complex plant genomes are linked together and can be used for custom analyses we are producing a series of **transPLANT user training videos**. A first user training video has been produced for the transPLANT genome resources for **barley** (*Hordeum vulgare*) and is available from the [transPLANT webpage](http://www.transplants.eu). We are currently in the process of generating a similar user training video for the complex wheat genome.

A new training workshop on 'Exploiting and understanding Solanaceous genomes'

The third transPLANT user training workshop took place at the Wageningen University & Research centre (WUR), the Netherlands, from October 13-14, 2014. It focused on transPLANT resources for solanaceous genomes, notably for potato and tomato. A total of 29 participants attended, representing 17 different nationalities

(Europe - 7, South-America - 5, Asia - 4, Africa - 1) and five companies. 11 tutors from the transPLANT community gave 1-2 hour tutorials with hands-on exercises, introducing database resources and data analysis tools for everyday work:

- [GnplS system](#) for integrating and mining genomic data (Pommier, URGI INRA)
- [ENSEMBL plants](#) (Bolser, EBI), notably with respect to solanaceous genomes
- query variation in tomato genomes (Janssen, da Ponta, KeyGene)
- methods and resources to predict the biological processes plant proteins are involved in (van Dijk, Warris, WUR)
- The [LAILAPS](#) search engine (Lange, IPK)
- tools for breeding: [BreeDB](#) (Finkers, WUR).

All workshop material including exercises, tutorials and solutions are now available from the transPLANT [user training website](#).



Training in data standardization for Marie-Curie ITN fellows



Within its WP3, transPLANT works on recommendations concerning standardized collection and exchange of experimental data on plants. Currently the main effort is directed towards data commonly called as "phenotypic". However, the developed recommendations are compatible with standards worked out by others for sequence or molecular data.

All project partners actively promote solutions in training workshops or by implementing them in data resources offered to the plant research community. On 16-17 September 2014, at the Amsterdam University, the group from the Institute of Plant Genetics PAS in Poznań that coordinates WP3, organized a training session on "Standardization of experimental information" for young scientists, PhD students and post-docs, working in the Marie-Curie ITN project [EpiTRAITS](#). The mission of EpiTRAITS (which financed the workshop) is to train young researchers in epigenetic

gene regulation and flowering in the model plant *Arabidopsis thaliana* and the crop plants maize and barley.

The training started with a demonstration of the main deliverables of transPLANT. Then the need for standardization, the distinction between metadata and data, and the elements of datasets that require standardization (content, annotation, format) were explained. The collection of standards registered at the [Biosharing platform](#) was presented, including the MIAPPE recommendations developed by transPLANT. The most important ontology services (EBI Lookup Service, Bioportal) were presented, as well as publicly available tools for data annotation and formatting (ISA Tools, Ontomaton). The sessions included practical training with selected exemplary data sets.

Two new data warehouses: GrapeMine and Wheat3BMine

Wheat3BMine, a data warehouse dedicated to wheat chromosome 3B



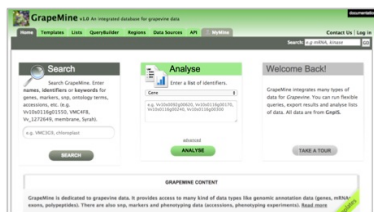
URGI develops a multi-species integrative information system: [GnpIS](#) (Steinbach *et al.* 2013 <http://dx.doi.org/10.1093/database/bat058>) which manages plants data from genomics to genetics and phenomics.

From this master database, we have generated a new data warehouse, [Wheat3BMine](#), dedicated to wheat 3B chromosome. It is developed using InterMine technology, providing a fast, flexible and user friendly access to integrated data by multiple ways: a browser, a query builder and a region search tool. Wheat3BMine users can filter their favorite features, save their own queries, and export results under many formats (GFF3, BED or XML). It contains heterogeneous data and is gene-centered. Our typical gene card centralizes relevant information like function, ontology terms and overlapping features.

Wheat3BMine provides access to genomic annotation data (genes, mRNA, polypeptides, transposable elements), polymorphisms data (markers), genetic mapping data (QTL, metaQTL) and phenotyping data. Indeed, useful links are available from a card toward [wheat 3B genome browser](#) (Choulet *et al.* 2014 doi: 10.1126/science.1249721) or toward additional details in GnpIS.

An online documentation and pre-computed queries are also available.

GrapeMine, a new data warehouse dedicated to grape genome



Following the same methodology, URGI developed a second datawarehouse, GrapeMine, dedicated to grape genome. It allows retrieving information of the v1 annotation of the grapevine genome (mRNAs, genes, exons and polypeptides), on ontologies (Plant Ontology, Gene Ontology, Prosite, Pfam), on polymorphisms data (Snp, indels or markers), on phenotyping data and genetic resources (germplasms) data.

Additional external links allow users to complete searches on:

- URGI GBrowse → https://urgi.versailles.inra.fr/gb2/gbrowse/vitis_12x_pub
- GnpIS → <https://urgi.versailles.inra.fr/gnpis/>
- BioMart → <https://urgi.versailles.inra.fr/biomart/martview/>
- Ontologies website → <http://www.ebi.ac.uk/QuickGO/>,
<http://www.plantontology.org/>

A tutorial with use cases is also available at the [GrapeMine tutorial pages](#).

e!DAL – store, cite and share research data



transPLANT is working on infrastructures on plant genomic science. One important aspect in the cooperation of trans-national collaborators is the sharing and publishing of plant genomic related data.

As joint effort of German Plan Phenotyping Network and transPLANT, IPK Gatersleben, Germany, published e!DAL as a general applicable framework for sharing and publication of research data (<http://dx.doi.org/10.1186/1471-2105-15-214>). The *e!DAL –API* is a very flexible, lightweight open source framework and condense all major accepted standards of long-term preservation for digital objects into a JAVA library. Its design enables an application as embeddable solution for data management in single tools as well as its operation as software stack for comprehensive research data archives. An already productive, e!DAL driven research data repository for plant research data is hosted at:

<http://edal.ipk-gatersleben.de/applications/repositories.html>

e!DALs major features are:

- DublinCore aware metadata management
- version management and data history tracking

- comprehensive search functionality
- interface integration into DataCite consortium
- data publication by assigned DOIs
- embedded HTTP(S) access and remote API

Further documentation, sample code and use cases are available on the project website <http://edal.ipk-gatersleben.de>. The software is available as Java libraries, source code or Maven artifacts.

Filling the gap between sequence and function - a bioinformatics approach

On October 28, 2014, Joachim W. Bargsten successfully defended his thesis at Wageningen University. The thesis was prepared at Plant Research International, Wageningen, the Netherlands in the context of transPLANT.

This thesis focuses on deriving function from sequence information, with the emphasis on plant sequence data. Unravelling the impact of genomic elements, in most cases genes, on the phenotype of an organism is a major challenge in biological research and modern plant breeding. An important part of this challenge is the (functional) annotation of such genomic elements. With the advent of next generation sequencing platforms, vast amounts of sequence data are generated. This data are used in connection with the available experimental data to derive function from a bioinformatics perspective. The connection between sequence information and function was approached on the level of chromosome structure and of gene families using combinations of existing bioinformatics tools.

The applicability of using interaction networks for function prediction was demonstrated by first markedly improving an existing method and by exploring the role of network topology in function prediction. The combination of methods and results presented indicate the potential as well as the current state-of-the-art of function prediction in (plant) bioinformatics. A new approach to improve the prediction of protein function in terms of biological processes is developed that is particularly attractive for sparsely annotated plant genomes. The combination of the network-based prediction method Bayesian Markov Random Field (BMRF) with the sequence-based prediction method Argot2 shows significantly improved performance. The approach was applied to predict biological processes for the proteomes of rice, barrel clover, poplar, soybean and tomato. The full thesis is available on the [transPLANT website](#).

[Back to top](#)

Copyright © transPLANT 2014



The transPLANT project is funded by the **European Commission** within its 7th Framework Programme, under the thematic area "Infrastructure", **contract number 283496**.

If you wish to unsubscribe from this mailing list, please email:
transplant_help@ebi.ac.uk