26. VITAMIN: An Ontology-Driven Information System for Mutant Annotation

Óscar D. Sánchez1, Ana Conesa2, José Alberto Maldonado1, Montserrat Robles1, Manuel Talón2
1 Medical Informatics Group, Technical University of Valencia, Spain
2 Genomics Center, Institute for Agricultural Research of Valencia, Spain
Contact: osanchez@fis.upv.es

Characterization of mutant individuals generates a wealth of qualitative and quantitative, heterogeneous data. We present an approach based on the combination of ontologies to describe and structure these data, which has been implemented in the VITAMIN information system. In order to develop a more "biologist-friendly" application, user-centered design techniques have been taken into account.

INTRODUCTION

The generation of mutant genotypes by random mutagenesis techniques is a widely used approach in Plant Breeding Research to generate potential new varieties. In this research, characterization of mutant individuals plays a central role. Such characterization means the screening of large collection of putative mutants on many different physical, biochemical and developmental traits At the Genomics Center of the Institute for Agricultural Research of Valencia a great effort to improve several Citrus varieties is being carried out. At present, more than 15.000 putative mutant genotypes are being characterized.

The VITAMIN (Valencia Information Tracking and Annotation of Mutant Individuals) information system has been developed, in collaboration with the Medical Informatics Group of the Technical University of Valencia, in order to provide a suitable solution for the management of the data generated by the Citrus Mutagenesis Breeding Program of the IVIA. This information system aims to serve as a tool for Plant Breeding by offering a structured framework for the easy introduction, visualization, analysis and rapid spotting of potential interesting new phenotypes.

VITAMIN: AN ONTOLOGY-DRIVEN INFORMATION SYSTEM FOR MUTANT ANNOTATION

The development of the information system has involved two phases: the definition of a data model able to envisage all the possible data and the latter software implementation.

Data Model

We consider that the data gathered from mutant experimentation is composed of three elements: organ, attribute and value, as shown in Table 1. This model has been proposed by the Phenotype and Trait Ontology, PATO [1,2].

ORGAN	ATTRIBUTE	VALUE
leaf	colour	yellow
leaf	length	5 cm.

Table 1. Example of data, structured in three elements

Both organs and attributes are taken from ontologies. As the basis for the organ ontology the Plant Ontology [3] is used. The attribute ontology can be adapted from an existing biological ontology or created from scratch based on specific measurements and assays applied to the biological material. The

pair organ-attribute defines a trait. Through the definition of traits –i.e. the definition of the allowed organ-attribute pairs- the curator biologist controls the biological consistency of the data.

In the other hand, values can be physical quantities obtained from quantitative assays or qualitative values: terms of a controlled vocabulary, free comments, graphs, photographs, external documents, etc.

Then, we consider that any mutant has associated a set of annotations. An annotation is mainly composed of the reference to the mutant that it describes, the trait that is being observed, and a set of values, as described before. This approach gives the flexibility that the biologists need to describe the individuals. For example, for a mutant with an unusual fruit colour, it would possible to associate the measure of the colour index with a photograph and a comment of the researcher.

DEVELOPMENT OF A "BIOLOGIST-FRIENDLY" APPLICATION

The development of bioinformatics tools frequently fails to take into account the analysis and understanding of the biologist experience [4]. Through several interviews with the agricultural researchers we have tried to identify their needs and to figure out how the computer tool could support the biological research.

In our particular case, the data gathered is firstly written in notebooks and then transcribed to a computer tool. The interviews with biologists showed that it is one of the most intensive and least productive tasks of the research, as showed in [5]. The data was usually saved in spreadsheets, what was very inefficient.

In order to solve these problems we have designed a user interface –currently under test- that aims to speed up the data transcription process. We also intend to take advantage of this process offering statistical information about the recently entered and previously stored data. This functionality aims to facilitate the identification of errors and the at glance discovery of phenotypes of interest.

CONCLUSIONS AND FUTURE WORK

We have developed a data model based on the combination of ontologies to describe and structure the phenotypic information of mutant individuals. We have implemented this model into the VITAMIN information system.

VITAMIN is currently installed at the Institute for Agricultural Research of Valencia and is being applied to the characterization of Citrus mutants. However, it intends to be extensible to any species, given that the description of the data is done with ontologies based on the OBO format [1].

Also, we have taken into account user-centered design techniques to develop a bioinformatics tool close to the real needs of the biologists. The tool and its user interface are currently under test.

In the future, we plan to refine the data model and to continue the development of the application adding new functions and providing it with extensible plug-in architecture. It is also planned to release an open-source stable version for the bioinformatics community.

REFERENCES

1.   Open Biomedical Ontologies, [http://obo.sourceforge.net]
2.   Georgios V. Gkoutos, Eain CJ Green, Ann-Marie Mallon, John M Hancock and Duncan Davidson. Using ontologies to describe mouse phenotypes, Genome Biology 2004, 6.
3.   The Plant OntologyTM Consortium and Plant Ontologies, Comparative and Functional Genomics, 2002, Vol 3(2), 137-142.
4.   Homa Javahery, Ahmed Seffah, Thiruvengadam Radhakrishna. Making Bioinformatics Tools User-Centered. Communications of the ACM, Volume 47, Issue 11 (November 2004)
5.   Ron B. Yeh and Scott Klemmer. Field Notes on Field Notes: Informing Technology Support for Biologists.