ZIGIA-DB: A versatile database for handling complex functional genomic data of *Arabidopsis thaliana*

Wingen, Luzie U.[a], Kim, Jan T.[b], Saedler, Heinz[a] and Dekker, Koen[a]

[a] Max Planck Institute for Breeding Research, ZIGIA,
Carl-von-Linné-Weg 10, D-50829 Köln, Germany

[b] Institut für Neuro- und Bioinformatik
Seelandstraße 1a, 23569 Lübeck, Germany

email: `wingen@mpiz-koeln.mpg.de`

The sequencing of the genome of the model plant *Arabidopsis thaliana* by the *Arabidopsis* Genome Initiative was finished in December 2000[1]. Sequence data revealed that the nuclear genome contains around 25,000 genes encoding proteins. This is a vast gain in information, but for most of these proteins, their function is largely unclear. Sequence comparisons can reveal homologies to other proteins which hopefully have known functions but still for a majority of proteins the function has to be determined experimetally *in vivo*.

Clarifying such gene–function relationships on a large scale in the model plant *Arabidopsis thaliana* is the goal of the ZIGIA project (Center for Functional Genomics in *Arabidopsis*[2]). To fulfill this task, an *Arabidopsis* knockout population was generated by mutagenising plants with the autonomous maize transposon *En-1/Spm*. The population currently consists of 11,000 individuals, each carrying 1–20 copies of the transposon. This population is used to study gene–function relationships using forward genetics (screening for plants with predicted differences in phenotype and identifying genes with transposon insertions responsible for the phenotype) or reverse genetics (screening for knockouts in specific genes and analysing the loss-of-function phenotypes of the identified plants). Several groups at the Max Planck Institute for Breeding Research (MPIZ) and furthermore industrial partners participate in the gene-phenotype discovery. Moreover, ZIGIA also offers to perform screens as a service to the scientific community.

The large amounts of different kinds of experimental data including plant identifiers, population genealogy, mutant descriptions, images of phenotypes, identified sequences, gene annotations, results of screens etc. which are entered or recalled by the members of the different groups demand a complex data management infrastructure. This infrastructure is required to support elimination of experimental redundancy, storing results in an easily accessible way and development of powerful tools for querying, analysing and integrating the data obtained in the project. This structure should make the ZIGIA data a useful resource for future users as the resources generated by ZIGIA will be integrated in the German plant genomics initiative, GABI

---

[1] `http://www.arabidopsis.org/genome_release.html`, *Nature* **408**: 796-815)
[2] `http://www.mpiz-koeln.mpg.de/~zigia/`

[3], in 2003.

For managing the complex data flow, a relational database was designed containing highly linked tables with the different data classes. The database, named ZIGIA-DB, was implemented as an Oracle database with a web based interface, so that easy data entry and access is possible from different locations and different computer platforms. As experimental data now start flowing into the database, the focus of ZIGIA-DB software development shifts towards making extensions and adjustments of functionality according to the need of the users. Thus, the database will soon manage the data obtained by large scale reverse genetics or other functional genomic studies.

---

[3]`http://www.mips.biochem.mpg.de/proj/gabi/`