

2008 Bioinformatics Subcommittee Report

Prepared by Heiko Schoof (Chair, schoof@mpiz-koeln.mpg.de)

The main problem limiting a transforming impact of the genomics data available for Arabidopsis is still seen as being data integration and the availability of facile tools for comprehensive analysis across data sets. The last year saw the final activities of the Web Services project that was initiated early in 2006 with funding from the DFG and NSF, details of which can be found at the project web site (<http://bioinfo.mpiz-koeln.mpg.de/araws>). A third developers' workshop was held in May 2007 at JCVI (John Craig Venter Institute), and results were presented at the ICAR (International Conference on Arabidopsis Research) in Beijing. This project sparked the creation of dozens of web services provided from several sites, as well as client tools for the utilization of these services. These can be web pages that include data from remote sites (MAtdB gene reports <http://mips.gsf.de/proj/plant/jsf/athal/index.jsp>), standalone tools like the workflow design and enactment tool Taverna (<http://taverna.sf.net>) or web tools such as the JABBA aggregator (<http://bioinfo.mpiz-koeln.mpg.de/jabba>) that automatically collects all data on an AGI locus code from all web service providers registered with BioMoby (<http://biomoby.org>). An important outcome of the project is that the effort to implement web services is a matter of days for most data providers. The most successful format is implementation workshops where about three days are spent to teach the basics and set up a working environment, then two days can be used to implement the first web services on the data providers' own systems. For the first production services, the workshop environment with experienced web service developers at hand is very useful to overcome initial problems arising from the heterogeneity of server and development environments as well as the initial complexity of debugging web services. Detailed project reports are available on request.

An RCN proposal to continue these efforts was not funded. However, at the next ICAR in Montreal, July 2008, a workshop will again be organized, this time under the leadership of Rodrigo Gutierrez together with the MASC Systems Biology subcommittee. Topics will be: Data Generation for Systems Biology, Data Integration, Quantitative and Qualitative Modeling, and Bioinformatics tools for Systems Biology. This workshop will also involve members of the iPlant Collaborative (www.iplantcollaborative.org). This project has received major funding from the NSF to create cyberinfrastructure that will support biologist working groups on selected grand challenge questions. Proposals for these grand challenge working groups are being collected from the community and will be discussed in workshops to take place between August and December 2008. The bioinformatics subcommittee aims to participate in this process in order to ensure alignment of integration efforts between the iPlant Collaborative and Arabidopsis functional genomics.