ConquestExplorer

A Genomic and Phenotypic Data Repository and Processing Tool towards Omics Data

Rico Basekow[1], Jost Neigenfind[1], Axel Nagel[1], Christiane Gebhardt[2], Birgit Kersten[1]
Contact: basekow@mpimp-golm.mpg.de

[1]GabiPD Team, Bioinformatics, Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14424 Potsdam-Golm, Germany
[2]Max Planck Institute for Plant Breeding Research, Carl von Linné Weg 10, 50829 Cologne, Germany

Introduction

The GABI-Papatomics project is aiming at the validation of the diagnostic power of specific marker allele combinations for resistance to late blight and tuber quality traits in marker-assisted selection (MAS) experiments in populations of tetraploid potato genotypes. The used markers are derived from the former Gabi projects Gabi-Conquest and Gabi-Chips. Furthermore transcriptome, proteome and metabolome differences between MAS sub-populations contrasting genotypically and phenotypically for resistance to late blight will be analysed.

Efficient managing and processing of data are important tasks for any project which handles huge amount of data. ConquestExplorer is a data repository for genomic and phenotypic data. The tool provides an easy and well arranged access for biologists and breeders to their data.

ConquestExplorer

ConquestExplorer stores the data in a backend database, provides a graphical user interface (GUI) to maintain the data and has an interface to the statistic software R to analyze data. Currently, we manage data of the Conquest2 project and initial data of the Papatomics project with this tool. Conquest2 data consist of more than 4,000 PCR products with sequence information and tracefiles, about 45,000 SNPs, 52,000 PCR markers and 17,000 phenotypic data points. The current Papatomics data contain more than 13,000 marker data points of about 1,600 genotypes.

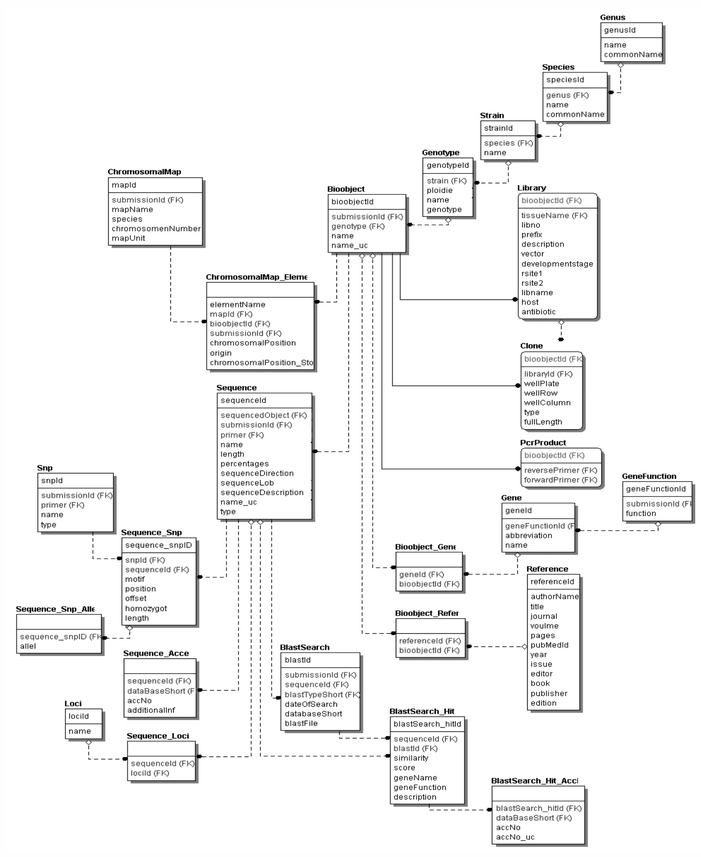

Figure 1: Extract from the PoMaMo[1] database schema. This schema was modified for the ConquestExplorer database schema.

Data Repository

The ConquestExplorer manages data in a relational database (Fig.1) which is accessed by an object-relational data layer. Thereby the underlying data model is flexible, easy expandable and adaptable to new data types and requirements. The tool allows to import data from files as well as to insert and to manipulate data on forms. To simplify data input a new input form similar to a spreadsheet was developed. Data can be explored by navigation and retrieved by a full-text search. Furthermore, a selection tool for marker profiles (Fig.2) was developed to support MAS.

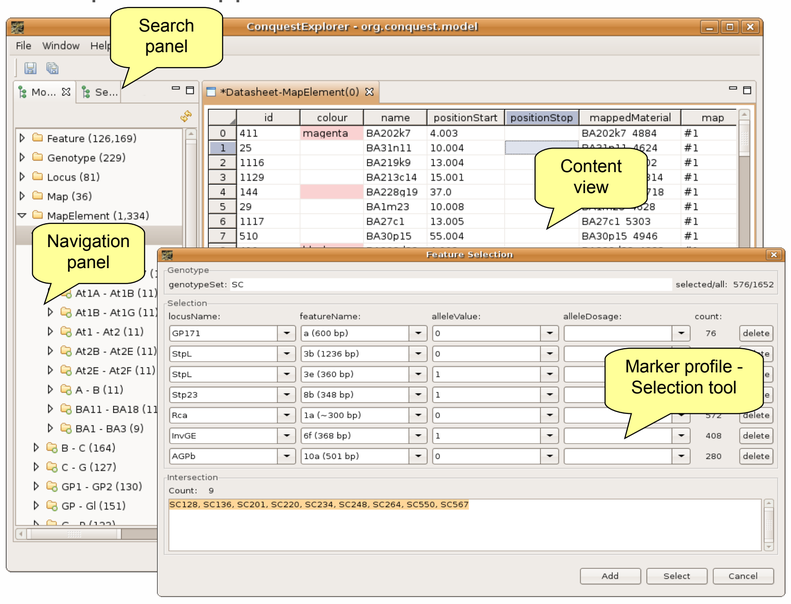

Figure 2: ConquestExplorer – Data management

Data Visualization

To get a fast and transparent overview about the different types of project data, the tool supports miscellaneous visualizations. There are text, map and image views. Those can show “raw” data like sequences or “correlated” data like SNPs that will be integrated in the tracefile of the original sequence (Fig.3).

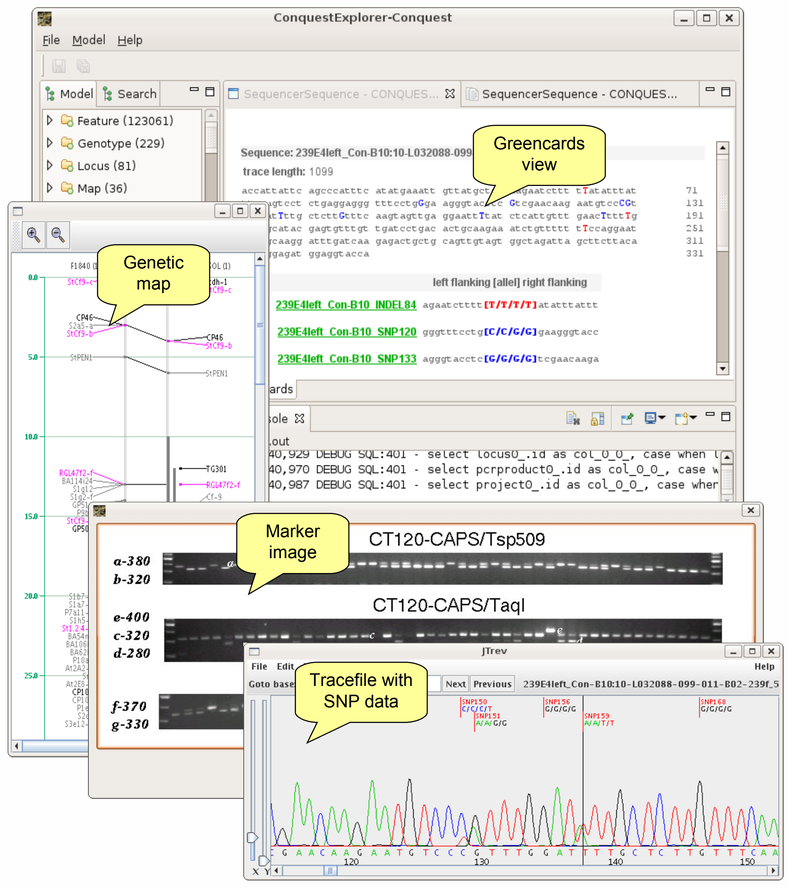

Figure 3: ConquestExplorer – Data visualization

Data Analysis

The integration of data management and data analysis makes data processing easier, more transparent and reproducible compared to a “manual” data analysis with a lot of cumbersome intermediate steps like data reformatting. Currently, standard statistical tests are integrated like Anova and Kruskal-Wallis (Fig.4). That enables correlation analyses between genotypic and phenotypic traits. The ConquestExplorer does also have an interface to the SATlotyper, a tool to predict haplotypes from unphased SNP data from heterzygous polyploids [2,3].

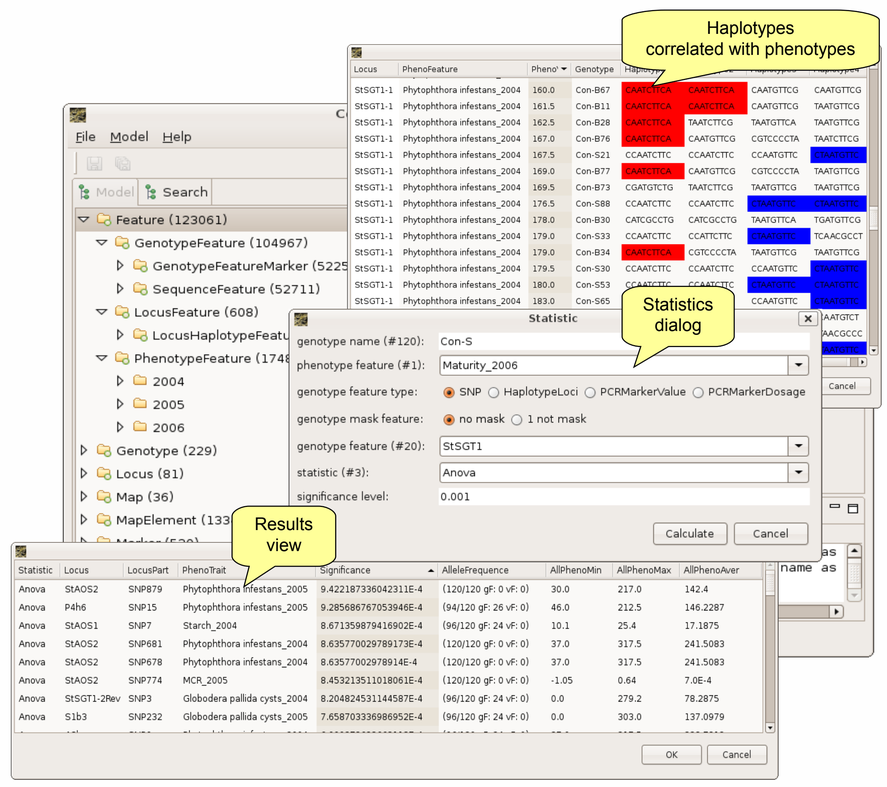

Figure 4: ConquestExplorer – Data analysis

Technologies

For the implementation of the ConquestExplorer we used mature and cutting-edge information technologies like

- an application platform (Eclipse Rich Client Platform),
- a database management system (HSQLDB),
- an object-oriented database interface (Hibernate),
- a fulltext search engine (Lucene) as well as
- analyzing software (GNU R).

Future Perspectives

ConquestExplorer will be extended to handle the new types of data generated within the Papatomics project like transcriptomic, proteomic and metabolomic data. We will also develop new or integrate available visualization tools for proteomic data. An interface to MapMan, a tool for visualization and pathway mapping of transcriptomic and metabolomic data, will be created. Furthermore, analysis tools for omics data will be developed and integrated.

gabi Genomanalyse im biologischen System Pflanze

References

[1] S. Meyer, A. Nagel and C. Gebhardt (2005) PoMaMo- a comprehensive database for potato genome data. Nucleic Acids Research 33 (Database Issue): D666-D670

[2] J. Neigenfind, G. Gyetvai, R. Basekow, S. Diehl, U. Achenbach, C. Gebhardt, J. Selbig and B. Kersten (submitted 2008) Haplotype inference from unphased SNP data in heterozygous polyploids based on the SAT. BMC Genomics, 9:356

[3] J. Neigenfind (2007) A generalized approach for calculating haplotypes in polyploid species based on the SAT-algorithm. Master Thesis, Free University Berlin