

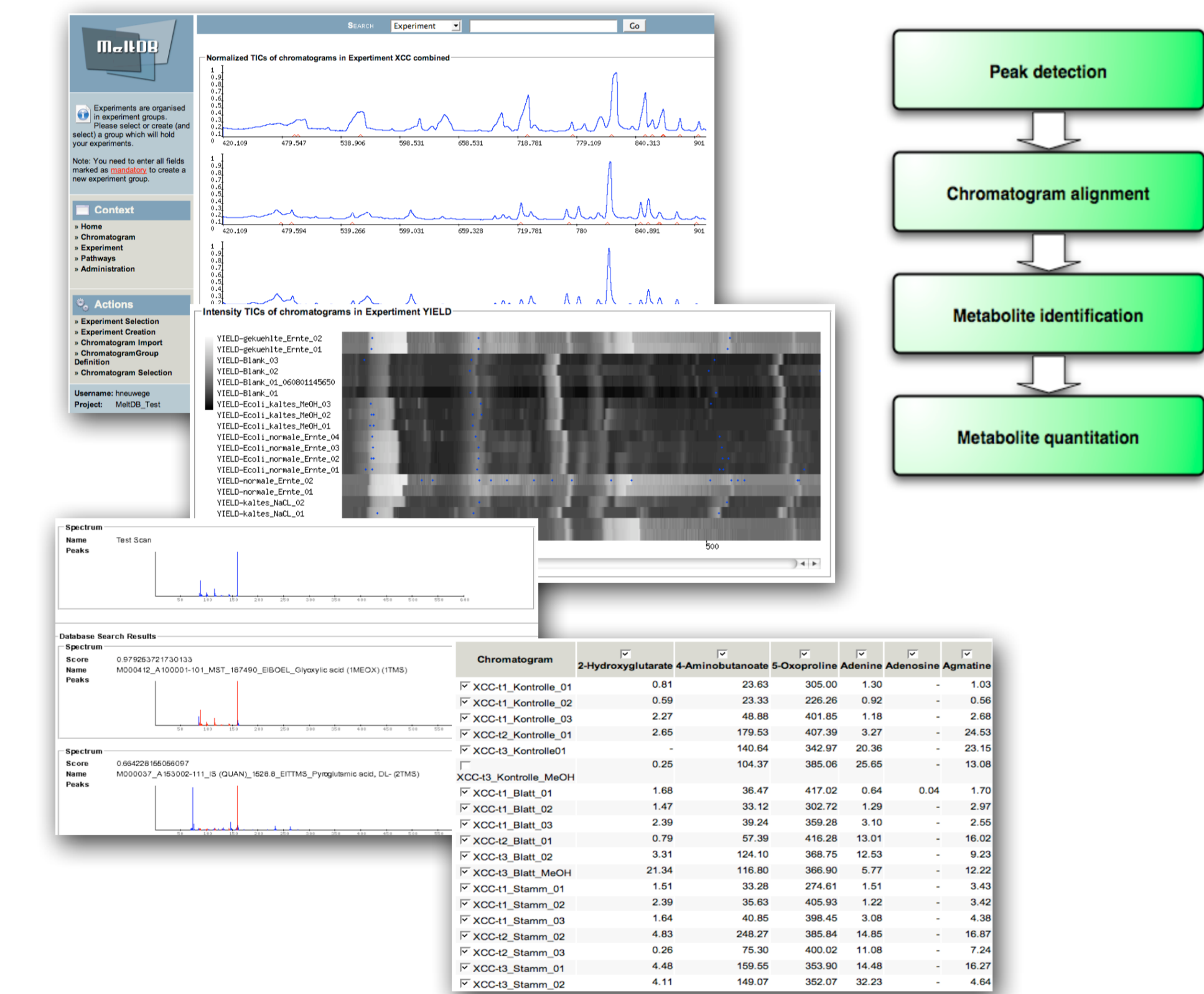
Introduction

Metabolomics is a rapidly maturing field and expanding the scope of functional genomics beyond microarray analysis and proteomics to the level of metabolism. In order to allow a global understanding of metabolic processes, an integration of information generated on all these levels has to be accomplished:

MeltDB is a web-based system for the management, analysis and integration of metabolomic datasets. The system allows organization and preprocessing of raw chromatographic data.

Currently, MeltDB supports the netCDF input format. Automated metabolite identification, quantitation and normalization methods are provided that generate metabolite pool datasets. These can be analyzed using R functionality integrated in MeltDB.

Extensible visualizations such as the interactive mapping of metabolite concentrations to metabolic pathways together with other 'omics' data are provided



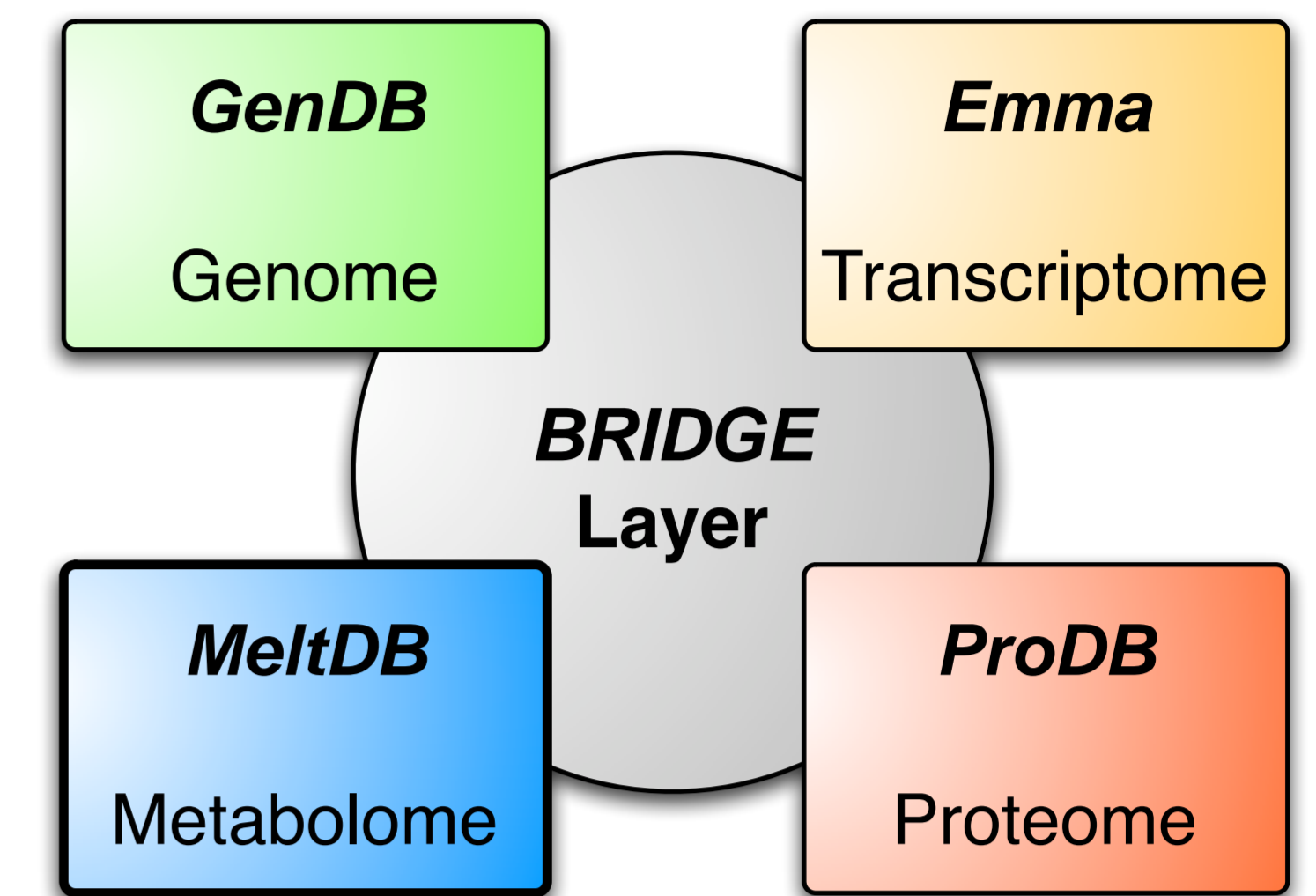
Analysis Pipeline

A pipeline concept has been defined that allows the integration of existing tools such as XCMS (Smith et al. 2006) and importers for results generated by proprietary software from the vendors of GC/LC-MS machinery.

For mass spectral database searches the algorithm proposed by Stein and Scott (1994) has been re-implemented and extended. Queries to the freely available GMD (Kopka et al. 2005) or user defined spectral databases in NIST or AMDIS format are supported.

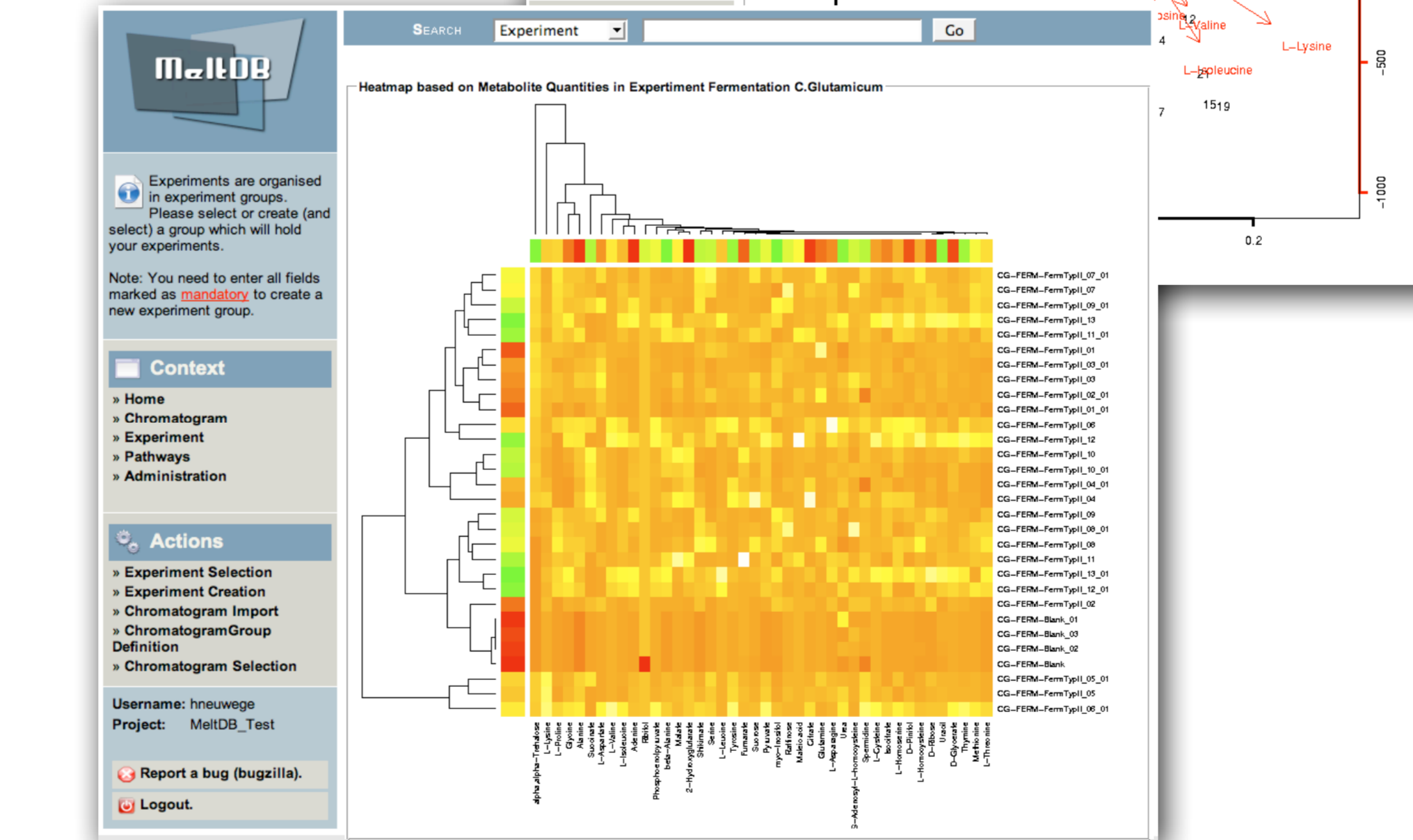
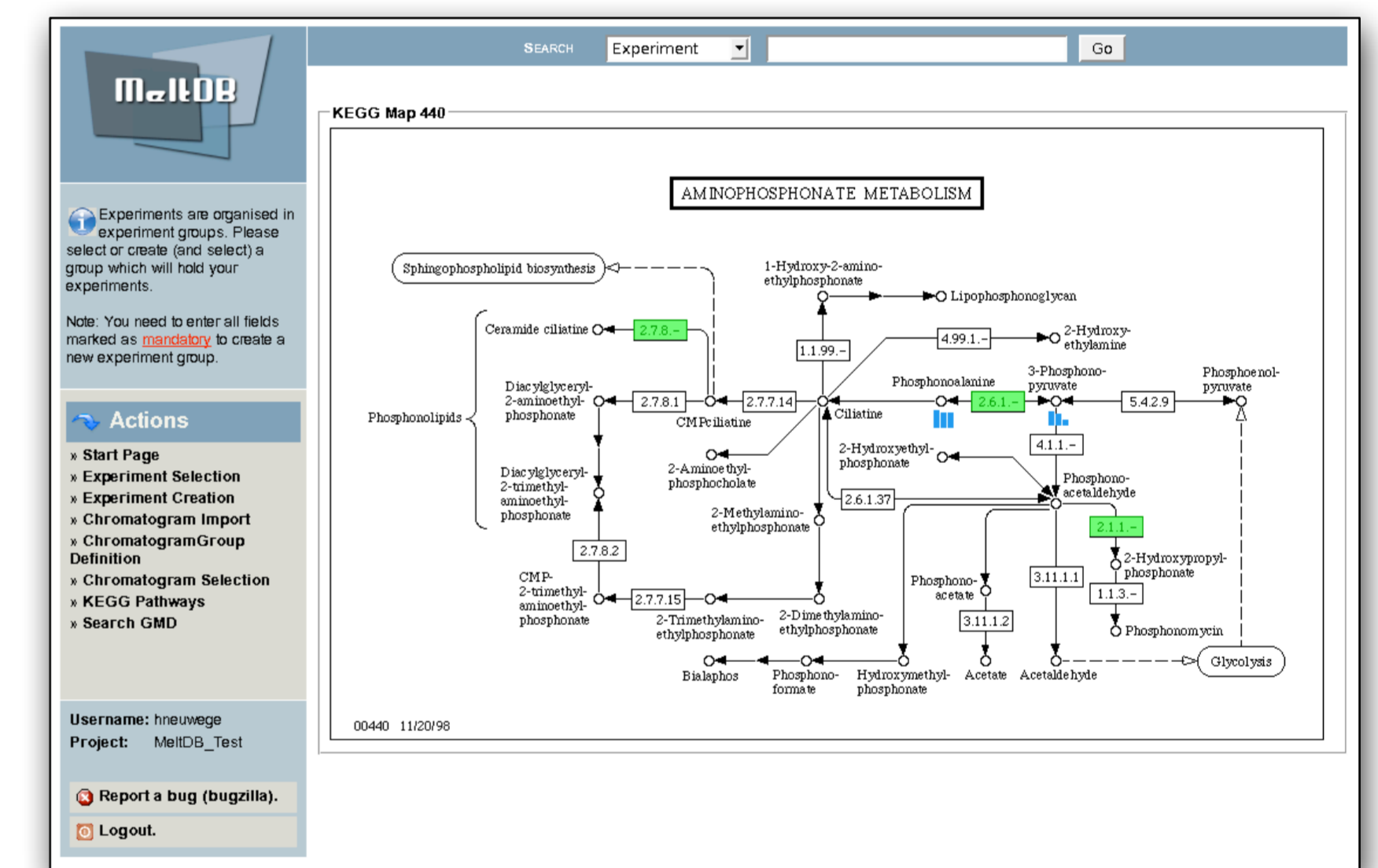
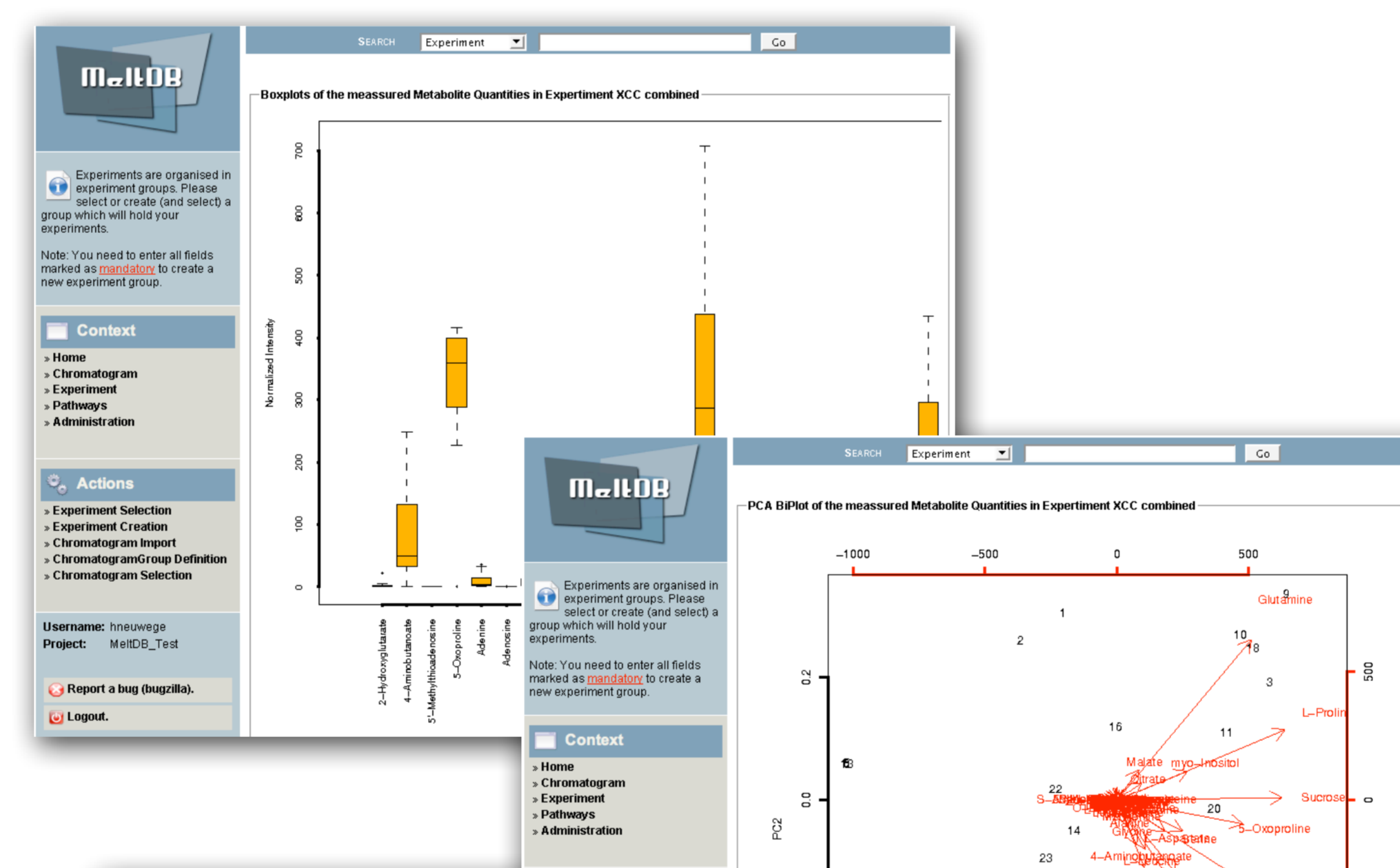
Additionally, peak detection and quantification routines have been implemented. Retention Index calculation together with a Dynamic Time Warping approach for chromatogram alignment is provided and contributes to the improved identification of metabolites.

These methods form the basis of the preprocessing pipeline of MeltDB that can automatically generate normalized metabolite concentration matrices from raw chromatographic datasets.



Integration with existing systems

Through the integration with other BRIDGE (Goesmann et al. 2005) compliant functional genomics software packages like GenDB or EMMA it is possible to generate comprehensive visualizations of metabolic pathways (e.g. KEGG). The level of transcript and metabolite concentrations as well as the gene content of the examined organism can be visualized. Direct access to the underlying datasets and their representations is achieved via the BRIDGE layer.



Current status and further development

The web based MeltDB system already provides functionality for the automated processing and analysis of GC/LC-MS chromatographic data. In the near future we would like to include additional input formats such as mzData/mzXML.

For improved statistical analysis, we would like to provide profiling and classification methods from machine learning such as Random Forest and SVMs.

Acknowledgements

Heiko Neuweiger would like to acknowledge the financial support by the International NRW Graduate School in Bioinformatics and Genome Research.



Statistical and Explorative Analysis

Statistical tests (T-Test, Manova, ...), hierarchical clustering and PCA can be performed on the obtained metabolite concentration matrices. The resulting visualizations are accessible via the MeltDB web interface and facilitate biological interpretation of the analyzed experiments.

References

Smith et al., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem*, 2006, 78, 779-787

Goesmann et al., Building a BRIDGE for the integration of heterogeneous data from functional genomics into a platform for systems biology. *J Biotechnol*, 2003, 106, 157-167

Kopka et al. GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics*, 2005, 21, 1635-1638

Stein and Scott, Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification, *American Society for Mass Spectrometry*, 1994, 859-866

Truong et al., Learning a Complex Metabolomic Dataset using Random Forests and Support Vector Machines, *Proc. of the tenth ACM SIGKDD int. conference on Knowledge discovery and data mining*, 2004, 835 - 840