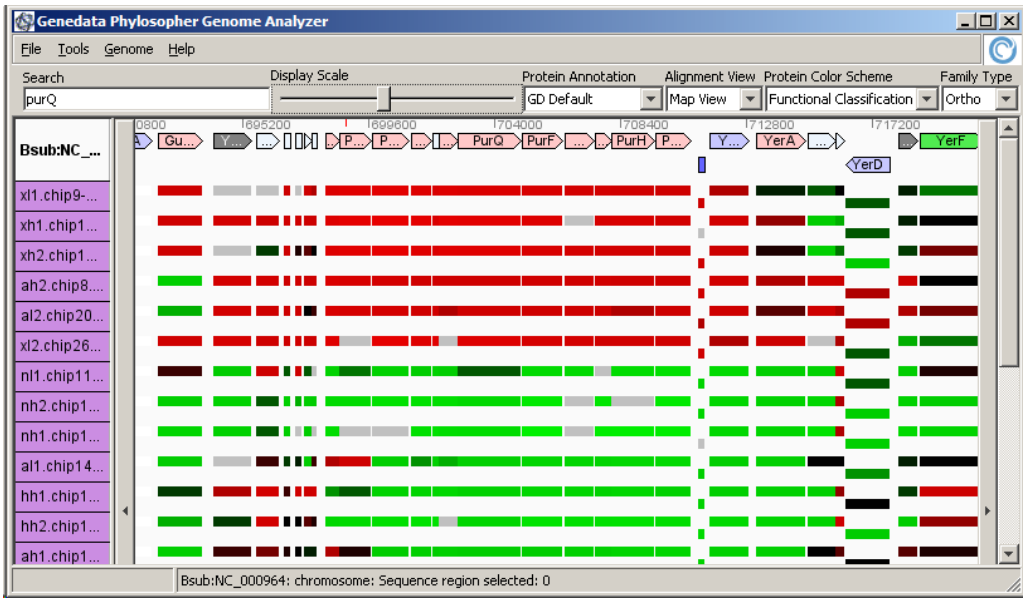
BaSysBio is a collaborative research programme (Integrated Project) funded by the European Commission under the 6th Framework Programme in the Life Sciences, Genomics and Biotechnology for Health

Development of databases for large-scale experiments

Hans Peter Fischer – Genedata, Switzerland

Genedata has developed a central database to support the storage of experimental data produced within the BaSysBio consortium, as well as tools for analyzing the biological context of the data derived from molecular profiling studies. This Central BaSysBio Data Management System (CBDMS) is largely based upon Genedata's established Phylosopher's technology. The CBDMS has been implemented on a central server that is hosted by Genedata's research IT services group in Basel. Within the first months of the project, the data models underlying the CBDMS have been successfully validated by integrating a variety of different *Bacillus subtilis* relevant data sources, partially from public sources, partially of proprietary origin provided by the BaSysBio consortium. Most noteworthy in this context is the integration of a whole-genome scaffold to be used as a common genetic reference system for the consortium. In addition, the first NimbleGen tiling array designs (DTU, Copenhagen) have been integrated and mapped onto the CBDMS reference genome. To facilitate the biological interpretation of the molecular profiling data being produced in the consortium, partner INRA has identified a variety of existing collections of Bacillus-relevant annotation sources which have been integrated into the CBDMS by Genedata. In addition, a first version of an enhanced metabolic reconstruction of the *Bacillus subtilis* metabolic network has been developed by Genedata scientists.



A genome-centric display of *Bacillus subtilis* data that has been integrated into the CBDMS. The figure shows an interactive, Java-based viewer that enables the querying and analysis of data representing different data types such as genome sequences, funct ional gene annotations, and whole-genome mRNA expression values. This particular example shows the *Bacillus subtilis* Pur locus, with the genes being color-coded according the cellular functions of the respective gene products. In addition, a number of different mRNA expression experiments representing different stress conditions are shown underneath the genome map with each row corresponding to color –coded expression signals derived from microarray signals (red rectangles = up-regulated genes, green rectangles = down-regulated genes). This example shows pronounced co-expression of neighbouring genes, suggesting that the genes of the Pur locus represent a so called operon structure that is subject to the same co-regulation mechanism.

From Work-Package 4.1 - Information system standardisation

Contributors: J. Retey, T. Hartsch, W. Vahrson, J. Traechslin, L. Macko, S. Ribrioux, H. P. Fischer (all Genedata), P. Bessieres (INRA), H. Jarmer (DTU)

January 2008

contact: Julia.Retey@genedata.com

EC contract reference LSHG-CT-2006-037469