Nature Precedings : doi:10.1038/npre.2009.3108.1 : Posted 20 Apr 2009

ChlamyCyc - a comprehensive database and web-portal centered on *Chlamydomonas reinhardtii*

Jan-Ole Christian[1,2], Patrick May[1], Stefan Kempa[2], Dirk Walther[1]

1 Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany
2 Institute of Biology and Biochemistry, University Potsdam, Germany

Overview

The unicellular alga *Chlamydomonas reinhardtii* is an important eukaryotic model organism for the study of photosynthesis and growth, as well as flagella development and other cellular processes. It has recently attracted substantial interest in the context of bio-fuel and hydrogen production. In the era of modern high-throughput technologies there is an imperative need to integrate large-scale data sets from high-throughput experimental techniques using computational methods and database resources.

ChlamyCyc provides a curated and integrated systems biology repository that will enable and assist in systematic studies of fundamental cellular processes. The ChlamyCyc database and web-portal is freely available: **http://chlamycyc.mpimp-golm.mpg.de.** It provides comprehensive information about the whole cellular system of an single organism on all levels of molecular organization.

It was assembled using an integrative approach combining the recently published genome sequence, bioinformatics methods such as orthology and paralogy assignments, and experimental data from metabolomics and proteomics experiments such as metabolic profiling. We also analyzed and integrated a combination of primary and secondary database resources, such as existing genome annotations from JGI, EST collections, orthology information, and MapMan [1] classification.

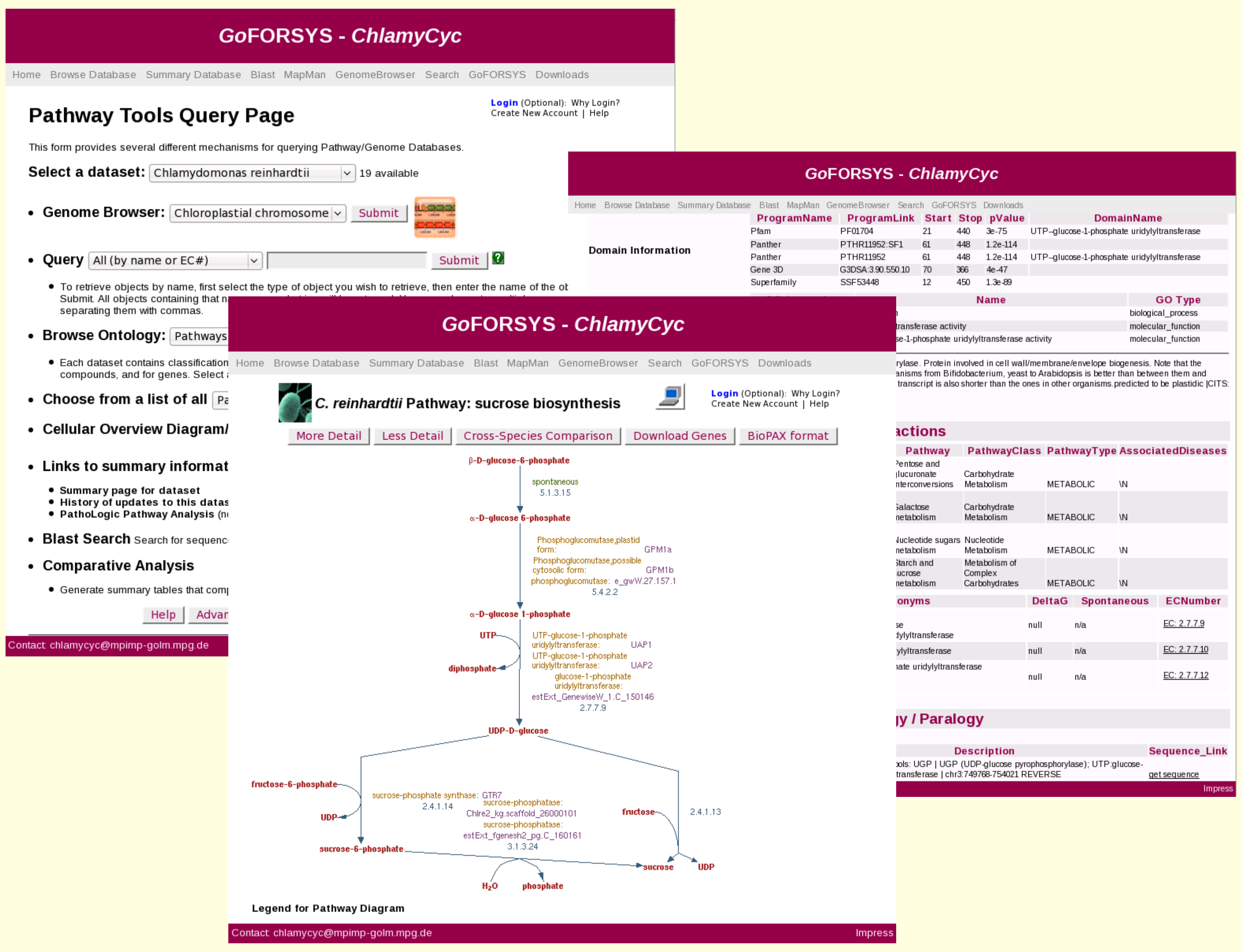
Metabolomic network analysis and visualization

Based on the adopted MapMan ontology and high-throughput " omics" data, we reconstructed the metabolic network and built a pathway/genome database for *Chlamydomonas reinhardtii* with Pathway-Tools Software [2] 12.5. It currently features:

- 270 Pathways
- 1395 Reactions, 6919 Enzymatic Reactions
- 14 Regulation
- 1024 Compounds
- 14657 Proteins
- 93 tRNAs, rRNAs, microRNAs

Comparative Analysis

We are mirror for 20 other pathway/genome databases from the BioCyc [3] family. This brings the possibility to compute comparisons across 18 organism, including 12 plants, yeast, mouse and human



Gene report pages

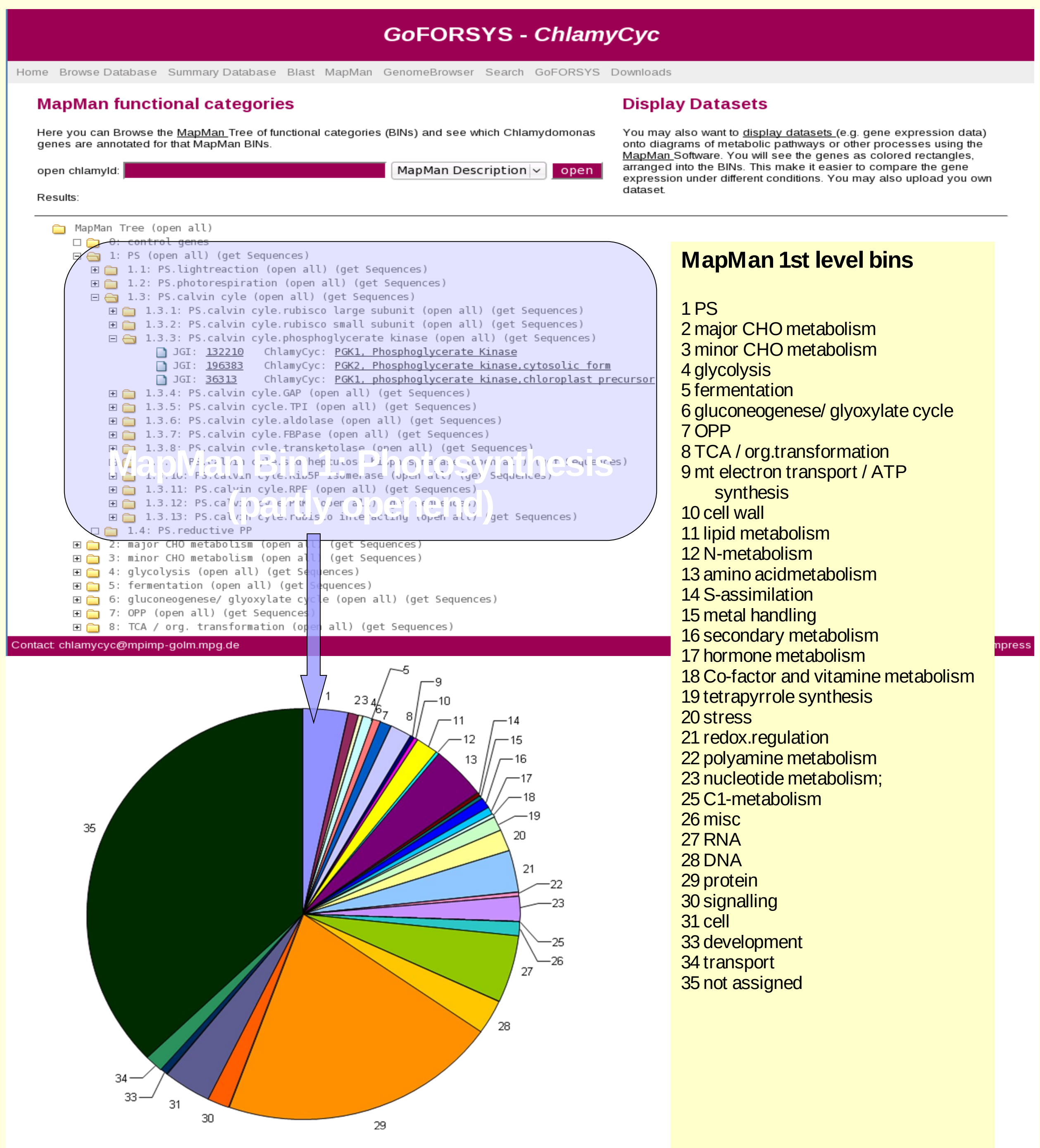
In addition to to pathway information, comprehensive gene-based annotation including sequence data, domain annotation, MapMan annotation, orthology and paralogy information, reactions, peptide support, links to genome browser, primer information and external databases is provided by gene centered report pages.

MapMan annotation

MapMan is an ontology developed to capture the functional capabilities of plants.

We could annotate 5.277 proteins (including 76 organelle-encoded) into 681 non-trivial MapMan classification bins.

The visualization of *Chlamydomonas* gene expression experiments with MapManWeb is directly linked.



You may also want to display datasets (e.g. gene expression data) onto diagrams of metabolic pathways or other processes using the MapMan Software. You will see the genes as colored rectangles, arranged into the BINs. This make it easier to compare the gene expression under different conditions. You may also upload you own dataset.

Annotation of *Chlamydomonas reinhardtii* proteins onto MapMan bins.

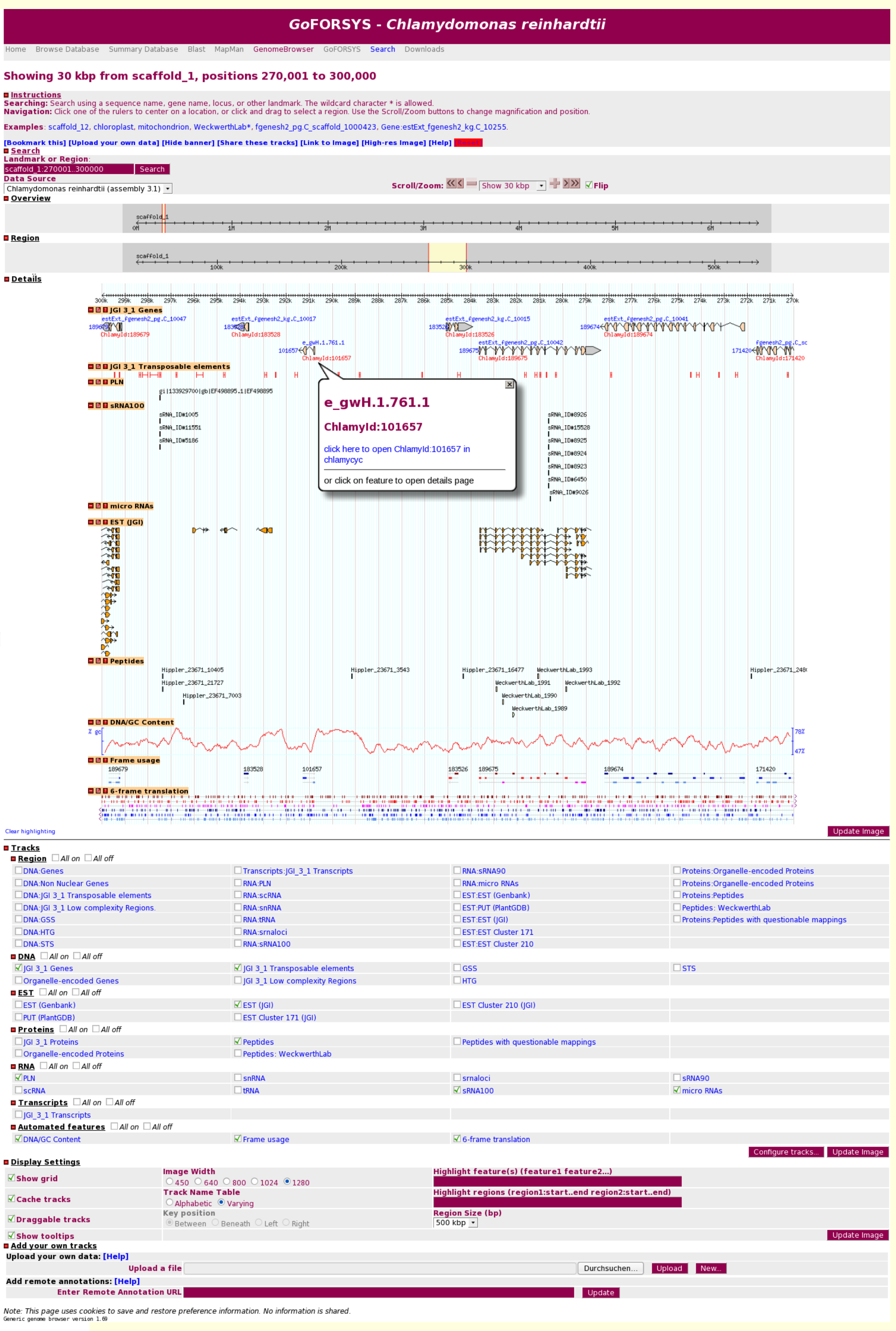
MapMan 1st level bins

1 PS
2 major CHO metabolism
3 minor CHO metabolism
4 glycolysis
5 fermentation
6 gluconeogenese/ glyoxylate cycle
7 OPP
8 TCA / org.transformation
9 mt electron transport / ATP synthesis
10 cell wall
11 lipid metabolism
12 N-metabolism
13 amino acidmetabolism
14 S-assimilation
15 metal handling
16 secondary metabolism
17 hormone metabolism
18 Co-factor and vitamine metabolism
19 tetrapyrrole synthesis
20 stress
21 redox.regulation
22 polyamine metabolism
23 nucleotide metabolism;
25 C1-metabolism
26 misc
27 RNA
28 DNA
29 protein
30 signalling
31 cell
33 development
34 transport
35 not assigned

Genome Browser

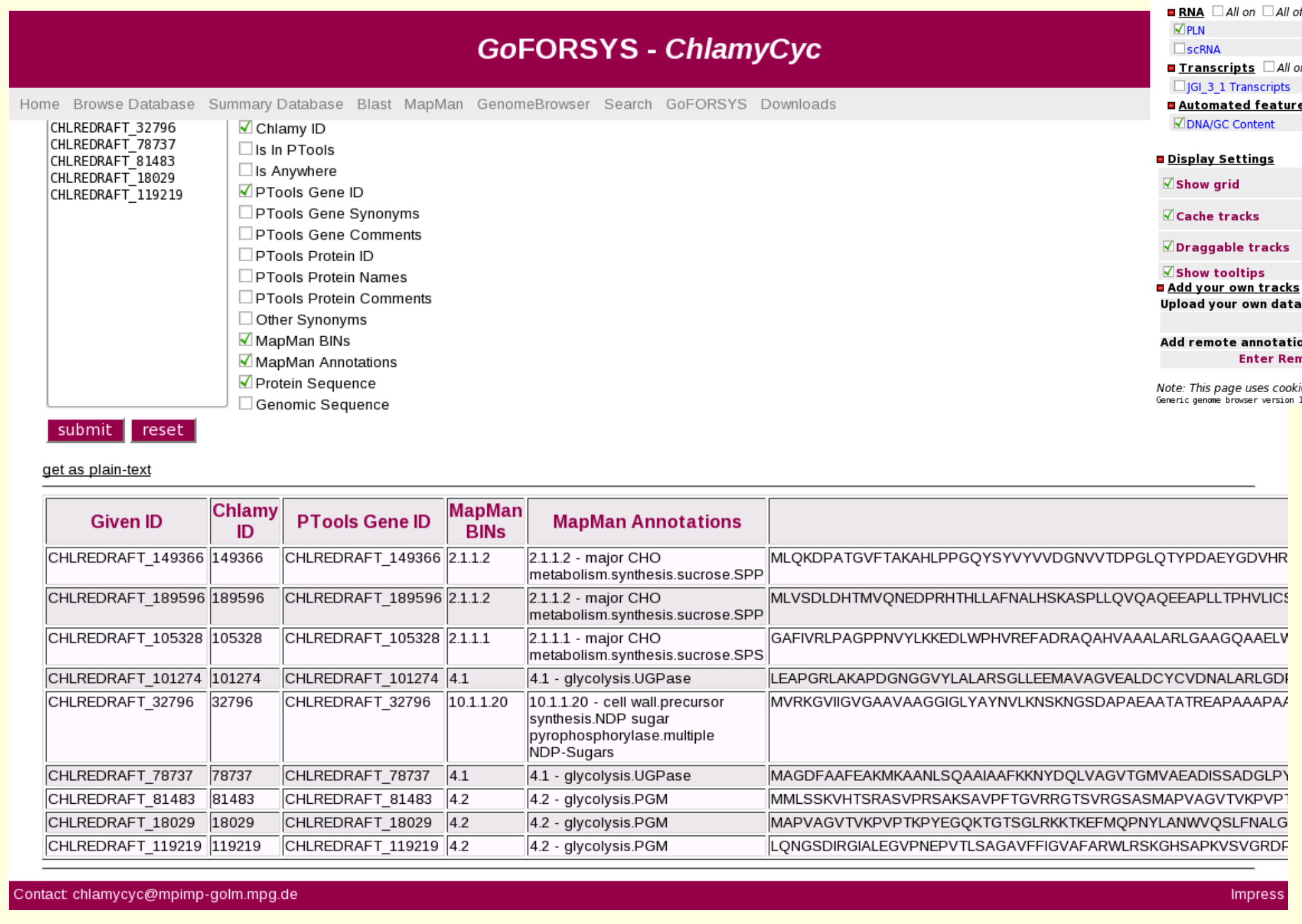
We have implemented a *Chlamydomonas reinhardtii* specific genome browser based on the GBrowse Software package [4] to provide a detailed view of the nuclear, plastid and mitochondrial genomes.

The GBrowse window can be used to display user defined combinations of 30 tracks including gene and protein annotation, various RNA-species as well as the EST and peptide coverage from inhouse proteomics experiments and external annotations.



BLAST search

We have implemented a web version of the standard BLAST [5] software to search against more than 40 sequence databases including those browseable with the genome browser.



Search and download

All data are searchable in various ways and are provided for download. E.g., one can easily download all sequences involved in a certain pathway or find all proteins annotated for a given MapMan annotation.

We present an integrated analysis of the molecular repertoire of *Chlamydomonas reinhardtii*. Bioinformatics annotation methods combined with GCxGC/MS-based metabolomics and LC/MS-based shotgun proteomics profiling technologies have been applied to characterize abundant proteins and metabolites. Based on the adopted MapMan annotation and high-throughput " Omics" data we reconstructed the metabolic network and built the pathway/genome database **ChlamyCyc**. The web resource combines network analysis and visualization, sequence analysis, and genome annotation.

References

1 Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M: " MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes" ; Plant J. 2004 Mar; 37(6):914-39
2 Karp PD, Paley S, Romero P: " The Pathway Tools software" ; Bioinformatics 2002; 18 Suppl 1:225-32
3 Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M. Paley S, Rhee SY, Shearer AG, Tissier C, Walk TC, Zhang P, Karp PD: " The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases" ; Nucleic Acids Res. 2008 Jan; 36 (Database issue): D623-31.
4 Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S: " The generic genome browser: A building block for a model organism system database." ; Genome Res. 2002; 12:1599-610
5 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: " Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." ; Nucleic Acids Res. 1997 Sep 1;25(17):3389-402

This research is funded by the FORSYS BMBF grant (GoFORSYS Grant Nr. 0313924) and the Max Planck Society.