CyanoEXpress: A Comprehensive Database for Evaluating the Transcriptional Response in *Synechocystis*

Miguel A. Hernández-Prieto, and Matthias E. Futschik
IBB - CBME, University of Algarve, Faro, Portugal

Background

The early availability of the *Synechocystis* sequenced genome has resulted in the accumulation of numerous genome-wide expression measurements for a large variety of environmental and genetic perturbations. In fact, over **600 individual microarray** measurements are stored in different public repositories. This massive data set can provide us with a plethora of new insights into the regulation of single genes, as well as into the coordination of cellular processes.

Why CyanoEXpress?

At present, researchers seeking to utilize accumulated expression data face several difficulties due to:

- use of different microarray platforms
- variety of data formats
- different data processing approaches

To assist researchers to overcome these difficulties, we have developed CyanoEXpress, a web-accessible database. CyanoEXpress enables users to browse expression data from many individual experiments in an intuitive way (Figure 1). Special care was taken in curation and pre-processing of the microarray data to avoid database errors and to obtain a large set of distinct expression patterns.

Methods

Raw expression data were collected from the NCBI Gene Expression Omnibus [1], EBI ArrayExpress [2] and KEGG Expression database [3]. Data processing was carried out using the R/Bioconductor platform [4]. Datasets were individually processed and normalized by OIN (Optimised Intensity-dependent Normalisation) to correct for potential dye bias [5, 6]. Log2 fold changes were calculated with respect to the controls used in individual experiments (primarily *Synechocystis* wild-type cultures grown under standard conditions).

Processed expression data were filtered to eliminate genes for which less than 80% of the measurements were available.

To support identification of co-expression and potential co-regulation, hierarchical clustering of genes was carried out using the software Cluster 3.0 [7]. Clustering was based on complete linkage, with the Spearman correlation as similarity measure. The clustered expression matrices can be visualized and queried through a Common Gateway Interface (CGI)-based application, which is a modified version of the GeneXplorer software [8].

Data information

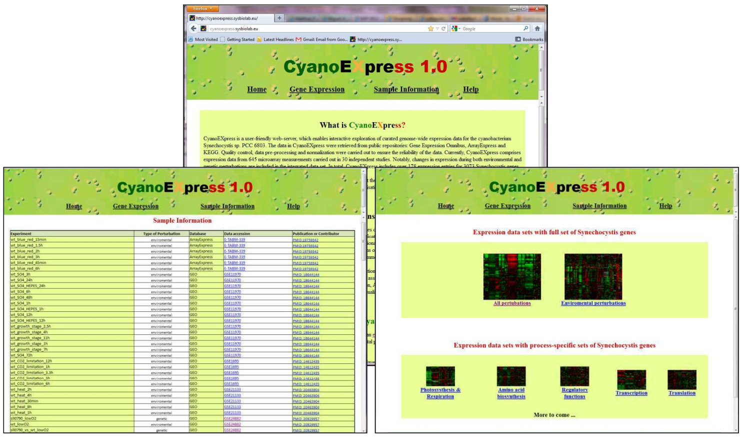
The integrated data stored in CyanoExpress contains expression data for **3 165 protein-coding genes,** derived from **651 microarrays,** used in **28 individual experiments** covering **166 different conditions** (i.e., environmental or genetic perturbations).

To facilitate the access of the user to information about the experimental conditions used in the original experiment we have added a table containing information about the microarray studies included in our database, as well as links to the corresponding raw data and publication (Figure 1).

http://cyanoexpress.sysbiolab.eu/

Acknowledgements
This work is supported by national Portuguese funding through FCT – Fundação para a Ciência e a Tecnologia, PTDC/BIA-MIC/101036/2008

CBME Centro de Biomedicina Molecular e Estrutural
FCT Fundação para a Ciência e a Tecnologia
Ministério da Educação e Ciência


http://cyanoexpress.sysbiolab.eu/

What is CyanoEXpress?

CyanoEXpress is a user-friendly web-server, which enables interactive exploration of curated genome-wide expression data for the cyanobacterium Synechocystis sp. PCC 6803. The data in CyanoEXpress were retrieved from public repositories: Gene Expression Omnibus, ArrayExpress and KEGG. Quality control, data pre-processing and normalization were carried out to ensure the reliability of the data. Currently, CyanoEXpress comprises expression data from 645 microarray measurements carried out in 30 independent studies. Notably, changes in expression during both environmental and genetic perturbations are included in the integrated data set. In total, CyanoEXpress includes over 178 expression entries for 3073 Synechocystis genes

Sample Information

Expression data sets with full set of Synechocystis genes

Expression data sets with process-specific sets of Synechocystis genes

More to come ...

Figure 1 CyanoEXpress 1.0. From the homepage the user can access to the visualization tool (Gene Expression), as well as to different information on CyanoEXpress itself (*Home*) or the data analysed (*Sample Information*). To assist the user we have also created a step by step Help page (*Help*).

Applications

The clustering of genes facilitates identification of coherent expression patterns, which can indicate co-regulation across different conditions. For example, we found that transcripts of several subunits of NDH-1 complexes tend to be tightly clustered and display specific differential regulation upon environmental perturbations, despite originating from different operons (Figure 2A). This observation suggests the existence of common regulatory mechanisms.

Furthermore, CyanoEXpress can be used to assign unannotated genes to functions based on their co-expression with genes of known functions. This feature is advantageous given that numerous *Synechocystis* genes are yet to be assigned functions. For instance, the expression of *slr0006*, for which no functional annotation has been given, is correlated with genes encoding for NDH-1 complex subunits, and is similarly strongly up-regulated under CO_2 limitation. Such co-expression indicates a potential functional association of *slr0006* with the NDH-1 complex under CO_2 limiting conditions.

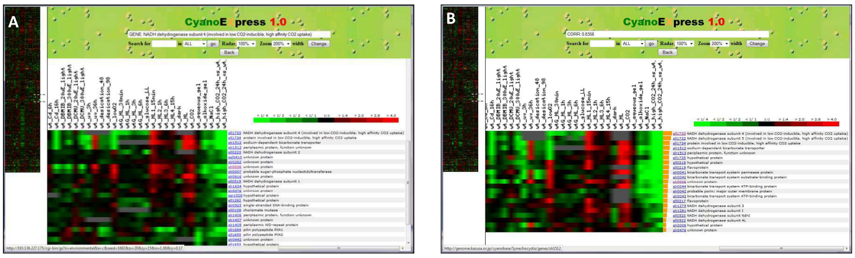
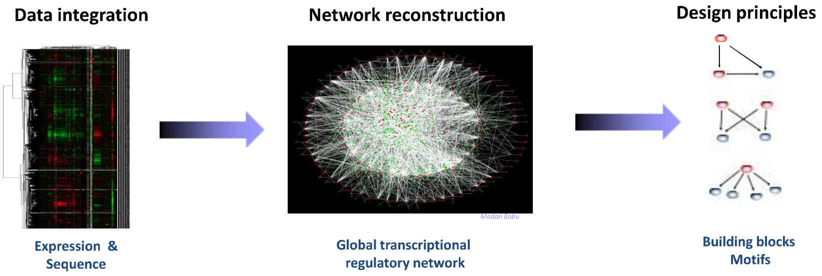


Figure 2 Visualization of expression levels of selected genes. (A) Using expression data obtained from *Synechocystis* wild type grown under different environmental conditions we selected an area of genes highly expressed under CO_2 limitation. (B) By clicking on a row of the selected heatmap a new page opens containing genes co-expressed with the chosen gene. The orange bar indicates the Spearman correlation coefficient with respect to the top gene, in this case *sll1733*.

Future Directions:

Solving the puzzle of transcriptional regulatory networks



Contact information

http://www.sysbiolab.eu/

E-mail: mprieto@ualg.pt

Literature cited

[1] http://www.ncbi.nlm.nih.gov/geo/
[2] http://www.ebi.ac.uk/arrayexpress/
[3] http://www.genome.jp/kegg/expression/
[4] http://www.bioconductor.org/
[5] Futschik M & Crompton T, *Genome Biol. 2004 5: R60*
[6] Futschik ME & Crompton T, *Bioinformatics. 2005 21: 1724*
[7] de Hoon MJ *et al. Bioinformatics. 2004 20: 1453*
[8] Rees CA *et al. BMC Bioinformatics. 2004 5: 141*