

Visualization and analysis of large-scale biochemical data with VANTED

Christian Klukas und Björn Junker

Network Analysis Group

Leibniz Institute of Plant Genetics
and Crop Plant Research, Gatersleben

Friedrich-Schiller-Universität Jena
(2006-01-17)



VANTED - Visualization and Analysis of Networks containing Experimental Data

Overview

- Motivation
- VANTEDs Features
 - Data Input
 - Data-Visualization
 - Statistic Functions
 - Extensibility
- Demo
- Discussion

Motivation

- ❑ Massively-parallel techniques generate more and more data
 - A top-down view on the biochemistry of a organism is made possible
- ❑ The amount of work needed to evaluate the data increases
 - New tools need to be evaluated or developed
- ❑ Goals
 - ☑ Show large amounts of data in a readable and understandable form
 - ☑ Consider related networks
 - ☑ Fast data evaluation with the help of statistic functions like t-test or correlation analysis, and clustering algorithms

Motivation



Low internal O₂ levels restrict storage activity of crop plant seeds

IPK
Institut für Pflanzen-genetik und Kulturpflanzenzüchtung (IPK), D-06506 Gatersleben, Germany

H. Rolletschek, H. Weber, U. Wobus, L. Borisjuk

Introduction
The role of low internal oxygen (hypoxia) as potential growth and yield limiting factor has become a hot topic in seed research. Internal hypoxia can affect metabolic activity, gene expression and, finally, biosynthetic fluxes. O₂ is delivered by either diffusion or photosynthesis. To determine its potential role for energy and storage metabolism of seeds we analyzed O₂ gradients using microsensors, ATP-distribution using bioluminescence imaging, metabolite profiles by LC-MS and biosynthetic fluxes by isotope tracer studies.

Results and discussion

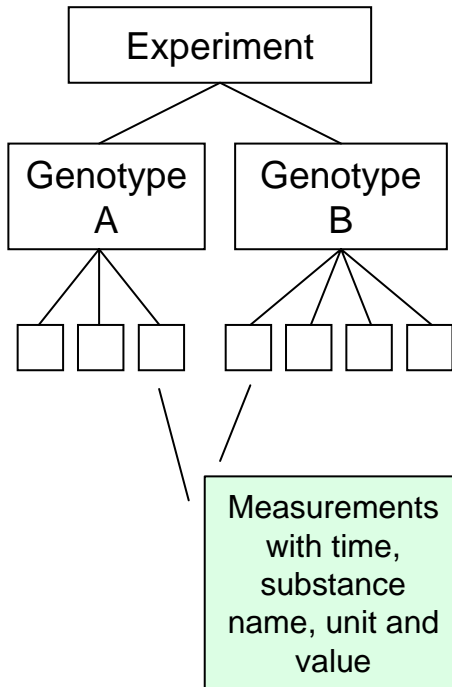
- 1. Oxygen mapping shows hypoxic regions within developing seeds**
- 2. ATP mapping shows metabolically-active regions and relationship to photosynthesis**
- 3. O₂ supply and photosynthesis affects metabolite levels and flux towards storage products**

A collection of scientific figures including micrographs of seeds, ATP distribution maps, and metabolic pathway diagrams. The figures illustrate the results of the research, showing hypoxic regions within developing seeds, metabolically-active regions, and the relationship between O₂ supply and photosynthesis to metabolite levels and flux towards storage products.

Key Features

- Data sources
 - Measurement data
 - Excel Files (VANTED-template)
 - Text Files (**J-Express** format)
 - DBE-Database (→ VANTED-DB)
 - **FLAREX** (Array experiment database at the IPK)
 - Pathway data (GML, Pajek-.NET, SBML)
- Data transformation and evaluation
 - t-test, U-test, Pearson- and Spearman correlation, SOM-data clustering, various layout commands, search and filter operations, extensible with script commands
- Data export
 - Image files (JPG, PNG, PDF, SVG)
 - Print out
 - Graph files (GML, Wilmascope-.XWG, DOT)

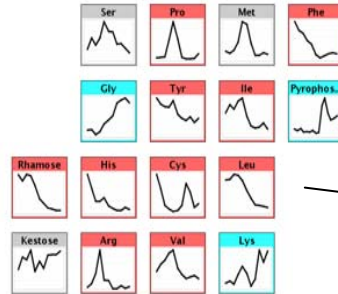
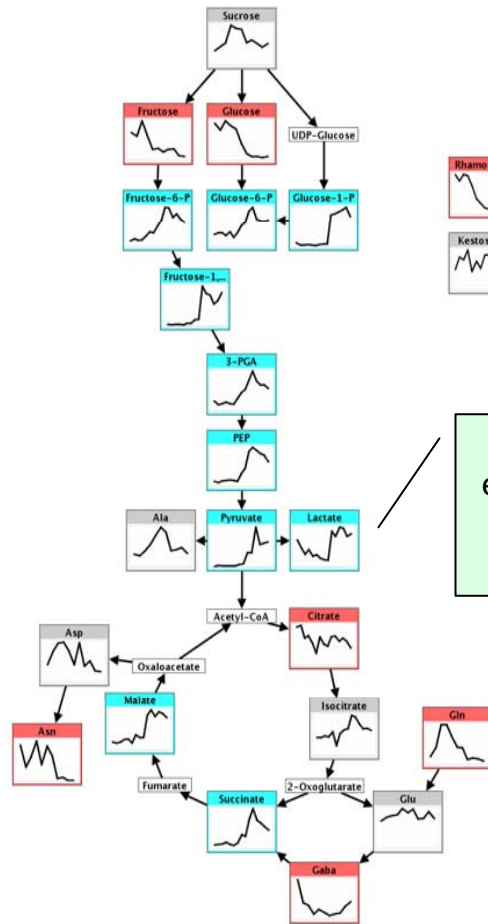
Excel Input-Form



Experiment			
Start of Experiment (Date)	08.03.2004	General information about the experiment	
Remark*	GPTas-Linien	optional input	
Experiment Name (ID)	GPTas-Transgene	contain numbers as 1, 2, 3, .	
Coordinator	Hardy Rolletschek	*** These cells must correlate to the numbers in	
Sequence-Name*		- The Experiment Name must be unique in the wf	
Plants/Genotypes**			
	1	2	3
Species	Vicia narbonensis	Vicia narbonensis	Vicia narbonensis
Variety*			
Genotype	wild type	GPTas9	GPTas13
Growth conditions*			
Treatment*			
Measurements			
			Substance
			Meas.-Tool*
			Unit
Plant/Genotype***	Replicate #	Time*	Unit (Time)*
1	1		Asp HPLC
1	2		Glu HPLC
1	3		Detector response
1	4		Detector response
1	5		Detector response

Measurement values

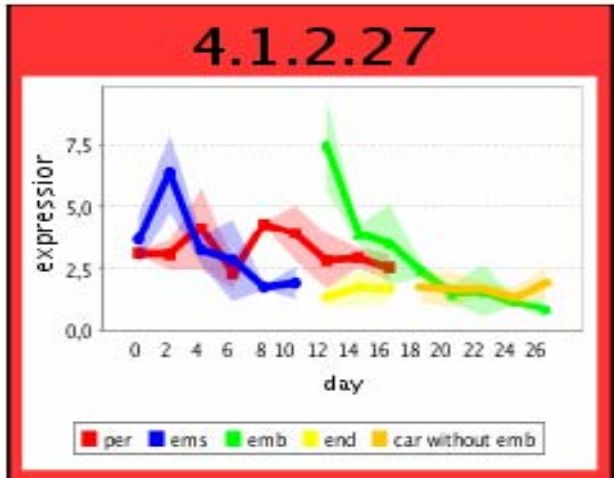
Data-Visualization



Clustering of data → Node-colour

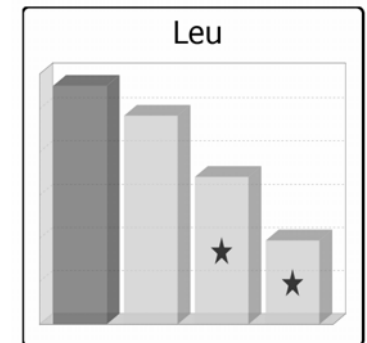
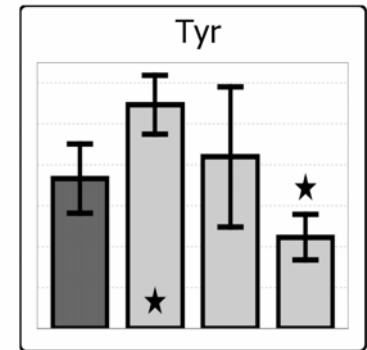
Level of detail

Mapping of experimental data onto pathways → bar or line charts



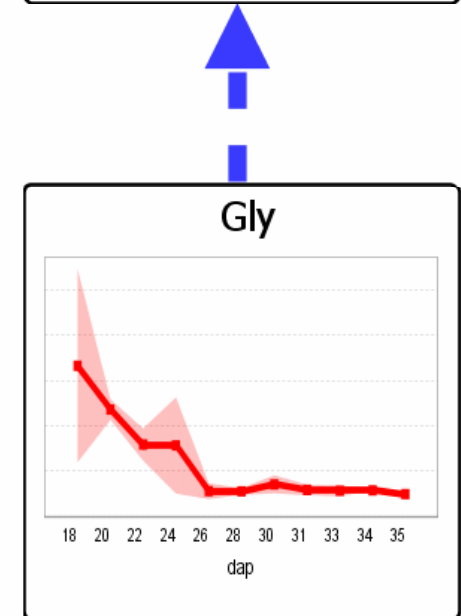
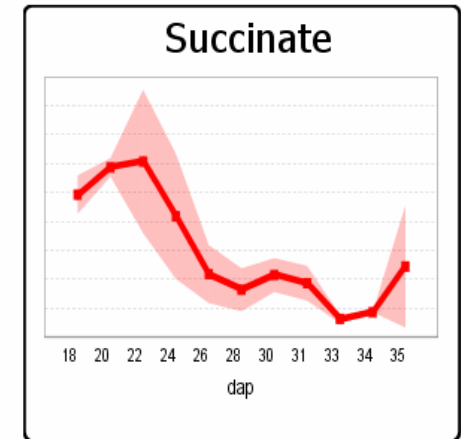
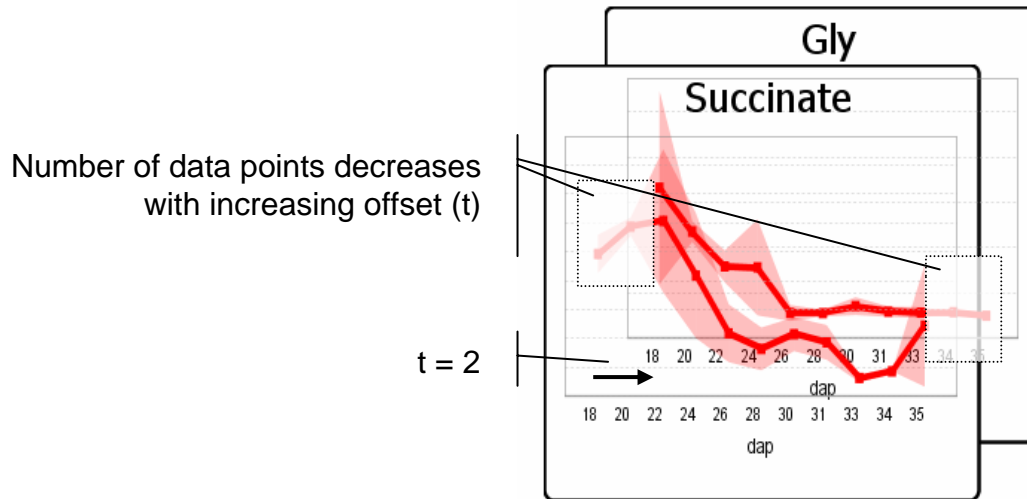
Statistic Tests

- Analyze data samples...
 - Check for normal distribution
 - ☑ David quick test
 - ☐ Chi-square test
 - Detect/Remove outliers
 - ☑ Grubbs test
 - Detection of significant mean differences with
 - ☑ t-test (2 variants)
 - ☑ U-test (rank-sum test)

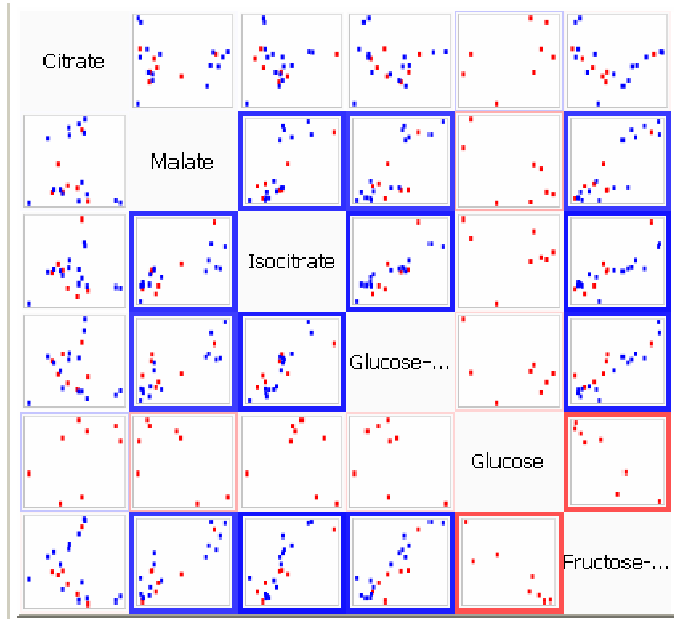


Correlation Analysis (1/2)

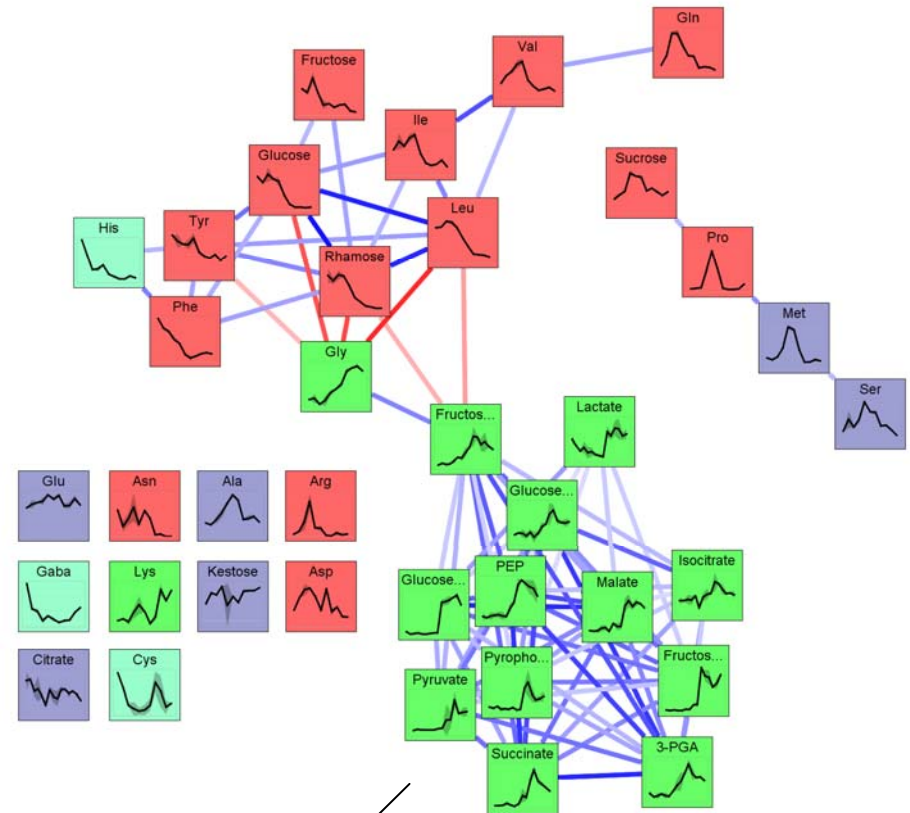
- Calculation of the Pearson (linear) or Spearman (rank-order) correlation
- Detection of correlations, shifted in time:
 - Repeated correlation calculation (r_i) for multiple time-offsets ($i=-3\dots3$, $t_{-3}\dots t_{+3}$)
 - Using $\max|r_i|$ for data visualization
- Test of significance with approximation to the t-distribution



Correlation Analysis (2/2)



Scatter Matrix



Correlation network

Extensibility (Example BSH Script)

Calculation of the average sample standard deviation

- Calculate average standard deviation for all clones (graph nodes)

```
//@Nodes:Calculate Average Sample StdDev$
series = node.getMappedSeriesData();
stddevs = new ArrayList();
for (SeriesData sd : series)
    stddevs.addAll(sd.getStdDevValues());
double sum = 0;
int i=0;
for (Double stddev : stddevs)
    sum += stddev;
node.setAttributeValue("script", "avg_stddev",
    new Double(sum/stddevs.size()));
```

Summary & Outlook

□ Website

- <http://vanted.ipk-gatersleben.de>

□ Publications

- Borisjuk, Hajirezaei, Klukas, Rolletschek, Schreiber: *Integrating data from biological experiments into metabolic networks with the DBE information system*. In *Silico Biology* (2004)
- Rolletschek, Radchuk, Klukas, Schreiber, Borisjuk: *Oil storage in soybean seeds: evidence for a key role of photosynthetic oxygen release*. *New Phytologist* (2005)
- Junker, Klukas, Schreiber: *VANTED: A System for Advanced Data Analysis and Visualization in the Context of Biological Networks*. *BMC Bioinformatics* (under revision)

□ Outlook

- Improve analysis and visualization of array data
 - Based on discussions with colleagues and feedback from users of the system

Acknowledgements

- Ljudmilla Borisjuk, Mohammad-Reza Hajirezaei, Hardy Rolletschek, Nese Sreenivasulu, Winfriede Weschke, Ruslana Radchuk, Dirk Koschützki, Falk Schreiber
 - > Discussion of system features and data provision
- Matthias Lange, Uwe Scholz, Andreas Stephanik, Karl Spies
 - > Database services and SOAP access to FLAREX

Software Demo

