# Integrative Systems Biology Visualization with MAYDAY

**Stephan Symons†, Christian Zipplies, Florian Battke and Kay Nieselt**

Center for Bioinformatics Tübingen, University of Tübingen, Sand 14, 72076 Tübingen, Germany, http://www-ps.informatik.uni-tuebingen.de

**Summary**

Visualization is pivotal for gaining insight in systems biology data. As the size and complexity of datasets and supplemental information increases, an efficient, integrated framework for general and specialized views is necessary. MAYDAY is an application for analysis and visualization of general 'omics' data. It follows a trifold approach for data visualization, consisting of flexible data preprocessing, highly customizable data perspective plots for general purpose visualization and systems based plots. Here, we introduce two new systems biology visualization tools for MAYDAY. Efficiently implemented genomic viewers allow the display of variables associated with genomic locations. Multiple variables can be viewed using our new track-based ChromeTracks tool. A functional perspective is provided by visualizing metabolic pathways either in KEGG or BioPax format. Multiple options of displaying pathway components are available, including Systems Biology Graphical Notation (SBGN) glyphs. Furthermore, pathways can be viewed together with gene expression data either as heatmaps or profiles.
We apply our tools to two 'omics' datasets of *Pseudomonas aeruginosa*. The general analysis and visualization tools of MAYDAY as well as our ChromeTracks viewer are applied to a transcriptome dataset. We furthermore integrate this dataset with a metabolome dataset and compare the activity of amino acid degradation pathways between these two datasets, by visually enhancing the pathway diagrams produced by MAYDAY.

## 1 Introduction

One main focus of today's biology and life science research is to obtain a systems based view of organisms and biological processes. Systems biology focuses on interactions in biological systems with the ultimate aim to model so-called emergent properties of these systems. While gene expression analysis is a valuable tool to monitor biological processes, it measures them on a basic level. Measuring metabolite concentrations on the other hand addresses the final effects of biological processes. Techniques like gas chromatography combined with mass spectrometry (GC-MS) allow to measure metabolite concentrations in a high throughput approach. With this technological advance, measuring all the metabolites that are present within a cell, tissue or organism during a certain physiological status has become feasible, and therefore metabolomics has received much attention in the last years. Though the number of metabolites that can be identified and quantified is still significantly lower than with techniques allowing to measure the proteome and transcriptome, a number of publications have actually dealt with analysis of the 'omics' data from common samples (see [7] for a review).

A general workflow for any 'omics' data encompasses (i) quality control and normalization of the raw data, (ii) statistical analysis (striving to find differentially abundant species) and

†Corresponding author: symons@informatik.uni-tuebingen.de

(iii) higher level analyses aiming to find relevant relationships in the data. One of the biggest challenges for systems biology comes in terms of the data integration that will be necessary in order to cross-correlate information obtained from the different 'omics' approaches. One important aspect is to offer software for pattern recognition by visual inspection and multivariate statistics. We have presented and continue to develop MAYDAY, an integrative tool for transcriptome based systems biology [4, 8]. Data-agnostic by design, MAYDAY can also work on general 'omics' data. Here, we introduce two new systems biology visualization tools for MAYDAY, complementing the data processing and visual analytics features already available.

In systems biology, it is common to interpret measured gene expression in its genomic neighborhood. Various tools have been implemented for this task. The UCSC Genome Browser [13] is among the most commonly used web-based genome visualization tools. It provides a track-based chromosome view, displaying one feature per track. It is mostly used for sequence analysis. ChromoViz [14] is a web-based genome visualization tool for displaying gene expression data. It also employs a track-based view and allows some user interaction. However, only ideogram-based chromosome navigation is possible. The Bluejay genome browser provides an integrative way of visualizing multiple gene expression datasets either from one or several organisms in a genomic context [5]. A stand-alone tool for genome visualization is Genome Projector [2]. It features several modes of displaying prokaryotic gene expression data. Instead of several tracks, Genome Projector displays one aspect per view, with the whole chromosome arranged as a stack of lines. Data integration, including the addition of new genomes, requires external software and considerable effort.

We have developed a new genome-based visualization tool, ChromeTracks, for MAYDAY. This track-based genome browser is interactive and scalable, and it can visualize expression data together with any metadata with genomic context.

When studying biological processes, the analysis of metabolic pathways has been proven to be a valuable tool, especially when a mapping of gene expression and general 'omics' data is possible. For this purpose, several applications with very different focuses have been implemented. Among the most distinguished products is Cytoscape [21], a general purpose tool for visualizing molecular interaction data including metabolic pathways. It is plugin based and provides a wide range of functionality, including mapping of measured data to the pathways. Ingenuity Pathway Analysis [11] is an extensive tool for exploring and analyzing metabolical, signaling and disease-related pathways. However, it is limited to selected model organisms. ProMeTra [18] is a web based tool for mapping systems biology data on pathway images. It relies on annotated SVG images created from other pathway sources. Other web-based tools include KaPPA-View [22] and MapMan [20] which are specifically designed for certain organisms, and Pathway Projector [15] which displays an overall map of metabolic pathways but has very limited analytical features. Most the these tools have their distinction for specific purposes, but they often lack data integration and analysis capabilities, extensibility and user interaction features. Besides visualizing metabolic pathways, systems biology software should offer a visual representation of changes during a time series or a physiological switch, thereby integrating data from different 'omics' sources. Our new pathway viewer visualizes metabolomic data within metabolic pathways either in KEGG or BioPax format together with expression data. Being a visual analytics tool it provides scalability and interactivity.

With our new visual analytics tools in MAYDAY we have continued our development of a framework for contextualizing high throughput 'omics' datasets. We aim at closing the gap

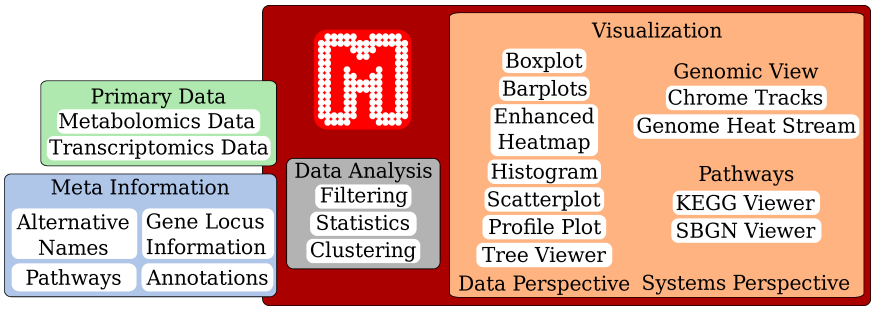


**Figure 1: Visualization concept of MAYDAY. MAYDAY combines various analytical and visualization methods, working on a wide range of data. For details on the different views and concepts, see text.**

between the data analysis workbench and the specific visualization tool. With our new features, MAYDAY allows the user to analyze multi-omics data without switching the application, thus minimizing the data conversion overhead.

## 2 Methods

**General features of Mayday** MAYDAY ("Microarray Data Analysis") is an application for the analysis and visualization of transcriptomics and other 'omics' data. It provides a plugin based interface for all aspects of data processing, including filtering, clustering, statistical analyses, and visualization. MAYDAY can import transcriptomics data from several sources, such as the Affymetrix platform. Omics data can also be imported from any tabular text file, as well as meta information such as gene names or probe genomic coordinates. The data model used in MAYDAY is independent of the data itself and summarizes measured species (genes, metabolites, etc) as probes. Meta information can be associated with all components of the data model and is structured hierarchically in meta information groups. MAYDAY is implemented in Java and freely available at `http://www-ps.informatik.uni-tuebingen.de/mayday`.

Our visualization approach is trifold (see figure 1). Generally, data (pre)processing precedes visualization. Secondly, for investigating general data properties (data perspective), we provide multiple purpose views tightly interconnected in a common framework. Thirdly, a systems based view is achieved with tools for inspecting data in a genomic context and mapped on metabolomic pathways (systems perspective).

For data analysis, MAYDAY offers a wide range of methods, including hierarchical and partitioning clustering, filtering and statistics. Statistical tests include parametric and nonparametric methods. Calculated test statistics and $p$-values are stored as meta information and can be used for various types of visualization and filtering. Furthermore, any statistical method available in R can be used via MAYDAY's RLink, which allows to operate directly on MAYDAY data using R. For details on MAYDAY's data analysis features see [4].

Data perspective plots show general properties of the current dataset. Among these are the profile plot, used to view temporal patterns of activity and the enhanced heatmap that offers the integration of supplemental data from different sources for the visual exploration of microarray data. Hierarchical clusterings can be directly viewed in a tree viewer or stored for later use. The enhanced heatmap is also capable of displaying trees above or at the side, inducing an ordering

on the data. To inspect distributions of experimental values and meta information, MAYDAY provides scatterplots, histograms and box plots. A uniform data model supports multiple simultaneous views of the same data and communicates selections between plots. Most plots can use meta information, allowing to display continuous, discrete and categorical meta information values with the plotted data. All plots rely on a unified framework for mapping primary or meta information to a wide range of color gradients and categorical data is mapped to distinct colors. All plots created with MAYDAY can be exported to several bitmap formats as well as the scalable vector graphics (SVG) format for high-quality publication-ready output.

To complement the data analysis and data perspective by a systems perspective we have now added a set of scalable genome visualization tools based on a library of efficient data structures for storing large genomic feature data. Furthermore, we introduce a framework for visualization of metabolic pathways, featuring viewers for KEGG pathway data and BioPax data in SBGN notation.

**Genome Visualization** Many types of 'omics' data are associated with genomic loci, such as measurements of transcription, epigenetic modifications, and protein levels. These genomic coordinates associated with each measurement can provide further insights, for instance on chromosomal clusters of co-regulated genes or larger mutations such as deletions. We have implemented ChromeTracks, a track-based genome browser for data associated with genomic coordinates. This tool is fully integrated with MAYDAY's visualization framework described above.

ChromeTracks displays data in one or more horizontal panels (tracks) which are aligned such that the same genomic coordinate is displayed at the same horizontal position in each track (see figure 2). Each track thus represents a linear view of the chromosome with associated data. Currently we provide three main types of track renderers. The heatmap track renderer can display either one strand or both strands of the chromosome. It allows to visualize two variables associated with each position using color and transparency, respectively. In a systems biology study, for instance, color could be used to represent the expression strength at a certain locus while transparency can be used to indicate the abundance of the corresponding protein. The stem track renderer adds a third variable, using its values to infer the vertical extent of colored, transparent boxes aligned with the chromosome. A third renderer can display locus data such as genes or exons, imported from tabular files, PTT, GFF or GenBank files. Further track renderers can be implemented as MAYDAY plug-ins.

Users can swiftly navigate the genome, jump or scroll to new positions, zoom horizontally (changing the number of base pairs per pixel displayed) and vertically (resizing individual tracks to take a closer look). Tracks can be added, removed or reconfigured at any time and the order of the tracks can be changed by the user. A region of the chromosome can be selected to limit the visible range, probes can be selected (synchronous with other MAYDAY plots). For increased resolution a second visualization option is offered: The heat stream visualization splits the linear chromosome into a number of rows stacked vertically, both strands are visualized together (see supplement for an example). The width of the visualizer window determines the horizontal extent of each row and the total number of rows needed is determined by the length of the chromosome and the zoom factor, configurable from single base-pair resolution to a view fitting the whole chromosome into the current window. Each row is composed of a ruler indicating genomic positions, and two rows of boxes, one for each DNA strand. Each

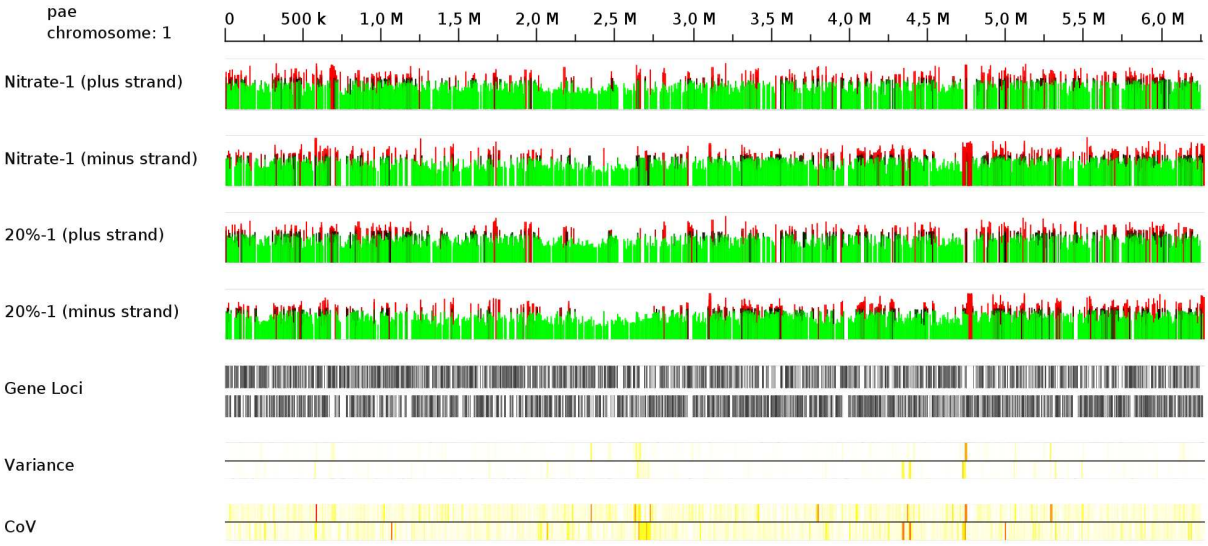


**Figure 2: MAYDAY ChromeTracks view of *P. aeruginosa*. From top to bottom tracks show expression data of the probes located on the plus and minus strand under the anaerobic nitrate respiration, followed by the expression data of the plus and minus strand under the oxygen aerobic conditions. These tracks use a green-black-red color gradient, with green indicating low and red indicating high expression. The two grey tracks show the genomic location of gene probes of the GeneChip used for this study on the plus and minus strand respectively. The next tracks show two statistical values derived from the expression data. The first track visualizes the expression variance of each probe across the experiments followed by the coefficient of variation. Here yellow indicates low variance and low coefficient of variation, respectively.**

box represents a certain number of base pairs and is colored according to an associated value (primary or meta information). The stacked view allows to show a larger genomic region than a track-based view, simplifying the exact selection of individual probes. Since most genomic data only covers the genome in a sparse manner, the heat stream offers a "condensed" view wherein bases that are not covered by the data are hidden.

We have implemented efficient data structures for the representation of chromosomes sparsely populated by features, complementing MAYDAY's data model with locus information. For most operations these data structures allow to work in near-constant time. For memory efficiency, we use native (non-object) data types and specialized containers wherever possible and map genome coverage to `long` values, each covering 64 bases. Depending on the zoom level, we employ different strategies for multi-threaded rendering. Generally, a priority-queue based rendering strategy is used which ensures that the currently visible area is always rendered with preference while invisible areas are rendered in the background for later display. At higher zoom levels, we restrict buffering to a large region centered around the currently displayed area. This provides a smooth scrolling experience while the buffered region is shrunk/expanded during scrolling. The third strategy is used at the highest zoom levels, when rendering directly to screen is faster than buffering. All rendering is done in background threads, keeping the user interface responsive at all times.

**Pathway Viewer** For a general systems-focused view on the data, metabolic pathways are a valuable tool. Whole transcriptome expression studies and upcoming large scale metabolic fingerprinting studies provide the underlying data for new insights in life sciences. Graphically laid out pathways have been used in publications and textbooks for a long while and are intu-

itively understandable [16]. Recently, SBGN, the Systems Biology Graphical Notation, a new standard for the visual markup of pathways has been introduced [16]. Another quasi standard for the visual presentation of pathways has been defined by KEGG [19]. We have implemented a framework for directed graphs as a foundation of the pathway visualization features of MAYDAY. Based on this, MAYDAY can visualize metabolic pathways from different sources and in two different styles.

We have added a lightweight library for directed graphs based on adjacency lists to MAYDAY. It features some common algorithms, e.g. for traversing graphs and finding connected components. For visualization, a model-view approach was used. Each graph is encapsulated by a model and can be visualized by several views, which display each node as an independent component. Each component has an individual renderer allowing to display different nodes in several ways, for example as shapes or displaying probe expression values. This allows to visualize pathways highlighting either the general structure, the type of the graph component, or species abundance.

The KEGG pathway viewer visualizes KEGG pathways specific for the selected organism, restricted to the enzymes and compounds actually present in the selected organism (see figure 3 for an example). Pathways are laid out as defined in the KEGG pathway file. Optionally, references to neighboring pathways and connections from pathway components can be displayed or hidden for clarity of the plot. Users can browse the whole pathway landscape by using pathway icons as hyperlinks. The KEGG viewer uses locally stored pathways in KGML format and KEGG annotation files. Annotations and pathways can be directly retrieved from KEGG when required. Enzymes and compounds are displayed as boxes, with the type of component indicated by a specific icon.

A common format for pathway definitions is BioPax [3]. It is used by all major providers of metabolic and signaling pathways including MetaCyc, Reactome, Pathway Commons and KEGG. We have implemented a lightweight parser to extract necessary information from BioPax `.owl` files. From these components, a graph for each pathway is constructed, using the SBGN notation as a conceptual framework. These graphs are laid out automatically and presented in either SBGN notation or displaying quantitative information (see figure 6).

Pathway data from both formats are internally represented as graphs using our lightweight graph library. They are subsequently mapped to probes using annotations derived from KEGG and BioPax files. Our graph layout algorithm is based on a recursive scheme [12]: Primitive types of pathways are laid out according to their type. This encompasses circular, linear and branched pathways which are laid out using naive algorithms and a hierarchical algorithm [9]. First, pathways are dissected into strongly connected components. Each component is then analyzed and rendered according to its class (circular, linear or branched). Complex components that can not be further dissected are rendered using a force-based algorithm [10]. If the pathway can be fragmented, each strongly connected component is laid out separately and collapsed into a single vertex. The overall layout is done recursively on the reduced graph. For laying out pathways, we first layout the pathway reaction backbone. The side components, including enzymes and reactants, are then placed heuristically. Any mentioned layout algorithm can be applied separately (with or without heuristically placing side nodes), allowing the user to find an alternative view if desired. Quantitative information can be rendered on any component for which a mapping between the names and references in the biopax file is available.

In both pathway viewers, visualization of ‘omics’ data or meta information is possible in differ-

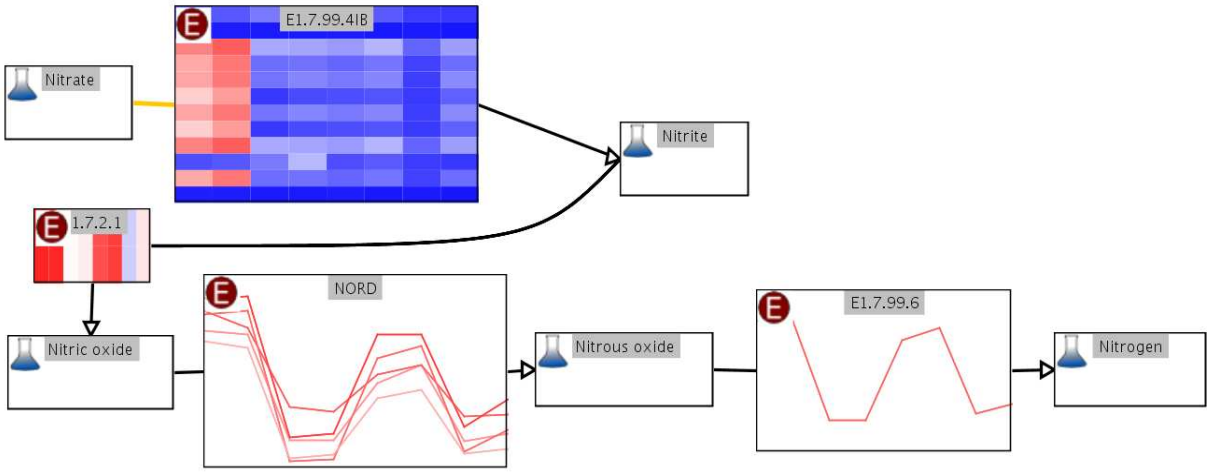


**Figure 3: Part of the nitrogen metabolism (KEGG id pae00910) visualized with overlaid enzyme expression values. All probes associated with the relevant enzymes are shown, either as profiles or as heatmap rows. Expression values are encoded as colors (low=blue, high=red) and as profiles. Conditions are sorted in ascending order by oxygen concentration (anaerobic, 0.4%, 2%, 20%), with both replicates per condition shown as separate adjacent experiments. Enzymes are marked with the letter "E", metabolites are marked with the image of an Erlenmeyer flask. Components are labelled with names assigned by KEGG.**

ent ways. Components are rendered as boxes or other simple shapes, displaying basic information via color. For nodes which can not be mapped to one or more probes or meta information, a simple renderer states the name and type of the object. Concentrating on a single value, the whole component is colored using a customizable color gradient. If several experiments are to be inspected, a single-row heatmap is displayed instead. A multi-row heatmap rendering is used to display several probes and experiments. The same data can be displayed by a profile plot. Users can interactively change renderers for all or a single component (see figure 3). An enlarged view of each node can be opened interactively. Furthermore, any component can be moved and resized in any dimension. Edges are drawn either as straight lines, or as bezier curves, heuristically and efficiently bundling edges.

## 3 Systems Biology of *Pseudomonas aeruginosa*

For an example of our integrative visualization approach we chose *Pseudomonas aeruginosa* as a case study for which poly-omics data is available. We downloaded microarray gene expression data from ArrayExpress (Accession id E-GEOD-6741) studying the growth of *P. aeruginosa* under different conditions: anaerobic nitrate respiration, and 0.4%, 2% and 20% oxygen aerobic conditions [1]. The data contains two independently prepared samples for each condition and measures the transcriptome of using the Affymetrix *P. aeruginosa* GeneChip.

Within MAYDAY we imported and normalized the raw data using RMA. For each probe the genomic coordinates were acquired from GenBank and read into MAYDAY via a tab-separated file. Further imported meta data included the common names for protein coding genes (PAxxxx) and a mapping to MetaCyc identifiers. Furthermore, we acquired pathway files for *P. aeruginosa* from KEGG and MetaCyc.

First, the expression variance and coefficient of variation (CoV) of each probeset across the

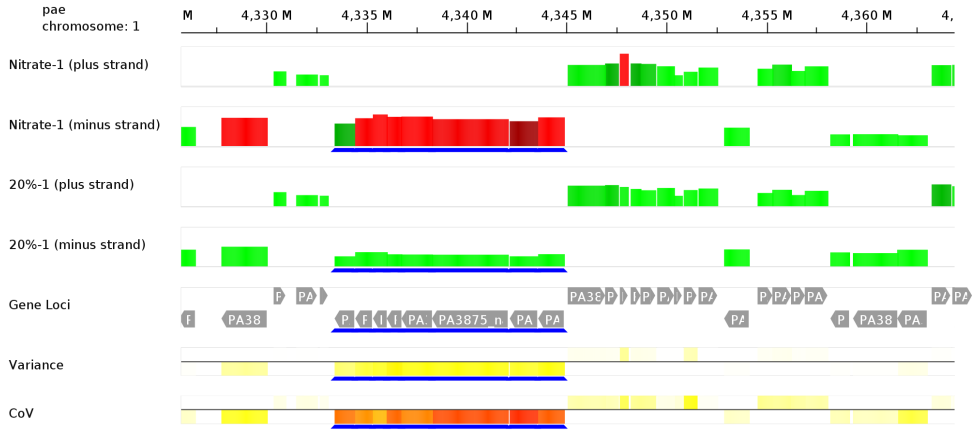


**Figure 4: MAYDAY ChromeTracks showing a selected region of the *P. aeruginosa* chromosome, highlighting a cluster of highly variant probes. Among the visible genes, most genes of the highly variant cluster between 4.335M and 4.345M belong to the nitrogen metabolism and are distinctly highly expressed (fold change $> 4$; data not shown) in the anaerobic samples.**

experiments were calculated. We used ChromeTracks to display these two deduced statistical parameters together with the experimental values and gene loci as aligned tracks (see figures 2 and 4). From the plot it can nicely be seen that we can observe genomic clusters of high variance that correlate with expression level differences between anaerobic and aerobic conditions. Filtering with CoV $> 0.1$ yielded 642 probes. Using ChromeTracks, we interactively inspected and extracted large clusters of genomically adjacent probes. We found many of the highly variant genes to be involved in nitrogen metabolism. Furthermore, many genes in the immediate genomic vicinity of nitrogen-related genes show a high variance and distinct expression profile (see figure 4). While genes involved in the oxidation of nitrate to nitrite exhibit a strongly regulated behaviour, some genes involved the reduction of nitric oxide are distinctly stronger expressed in anaerobic and 2% oxygen samples, albeit to a lower extent in the latter. For visualization of this process, we used the pathway from KEGG (accession id pae00910) and analyzed the nitrate metabolism part of it (see figure 3). While the energy metabolism related reactions of the nitrate metabolism show a clearly visible reaction to anaerobic conditions, this can not be said of the amino acid related reactions of this pathway (see supplement).

Other highly variant genes are involved in the degradation of several amino acids. Visually inspecting the valine and leucine degradation pathway, we found that most genes in these pathways were less active in the 20% oxygen condition, while being highly expressed in the other conditions (see figure 5). In contrast, most genes in the tyrosine degradation were less expressed under the anaerobic condition. Genes involved in the histidine degradation pathway were found to be highest expressed in 0.4% oxygen condition.

For sake of exposition, we compared these finding with a metabolomics dataset acquired under similar, though not identical conditions. Metabolomics data studying *P. aeruginosa* in various growth phases and oxygen concentrations are available from the Systomonas project [6]. The data contained measurements for anaerobic and aerobic conditions (10 samples each) during the exponential growth phase, which is comparable to the transcriptomics data. The metabolomics dataset is available from `http://www.systomonas.de/` (series 6 and 8). The data was imported in MAYDAY, where it was transformed ($\log_{10}$) and missing values were imputed. Species with more than 30% missing values were excluded from the analysis.

Using MAYDAY, we conducted a t-test to find metabolites which are differentially abundant in

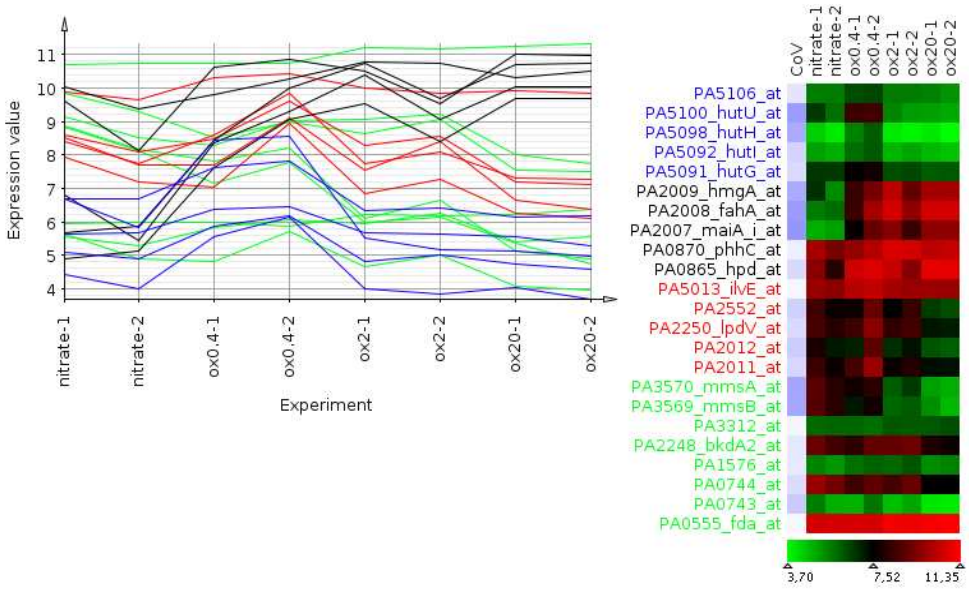


**Figure 5: Profile plot (left) and enhanced heatmap (right) showing genes involved in histidine (blue), tyrosine (black), leucine (red) and valine (green) degradation. The heatmap additionally displays the coefficient of variation as a column using a blue gradient.**

both conditions. We found that in fact, tyrosine and leucine had a FDR-corrected $p$-value $< 0.05$. Their concentrations are decreased in the aerobic samples. Also valine abundance is decreased in the aerobic condition ($p = 0.07$). Since the above gene expression data and the metabolomics data are not directly comparable, we used $z$-scores to transform the values to be visually comparable. We partitioned the transcriptome dataset into two classes combining conditions with similar growth rates [1]: class one contained experiments with 0% and 0.4% $O_2$ and class two those with 2% and 20% $O_2$. We visualized selected pathways using MAYDAY (see figure 6).

We found that the other metabolites in the degradation pathways of tyrosine (see figure 6), leucine and valine had a more or less constant concentration in both conditions, while the basic metabolite is decreased. The findings agree well with the fact that *P. aeruginosa* cultures grow up to 2.5 times faster under aerobic than under anaerobic conditions, consuming amino acids for proteins. The increased tyrosine degradation activity is in concordance with the fact that tyrosine is less commonly used than leucine and valine, and that the tyrosine degradation yields glutamate in the first step which is a substrate for many other biosynthesis pathways.

## 4 Discussion

Within MAYDAY we have implemented new visual analytics tools to integrate various 'omics' data. As a comprehensive platform for analyzing and visualizing systems biology data, MAYDAY provides tools for most of the necessary steps in one common environment. Our genome viewer visualizes data in its genomic context. The track-based concept allows to compare several measurements as well as meta data. Using the heat stream view, a single variable such as gene expression or a statistical value can be highlighted in a genomic context.

Our new metabolic pathway viewer is an interactive visualization tool that allows researchers to gain insights into the functional neighborhood of genes and proteins. Using automatically laid out pathways from common sources, MAYDAY can be used to visualize pathways from any

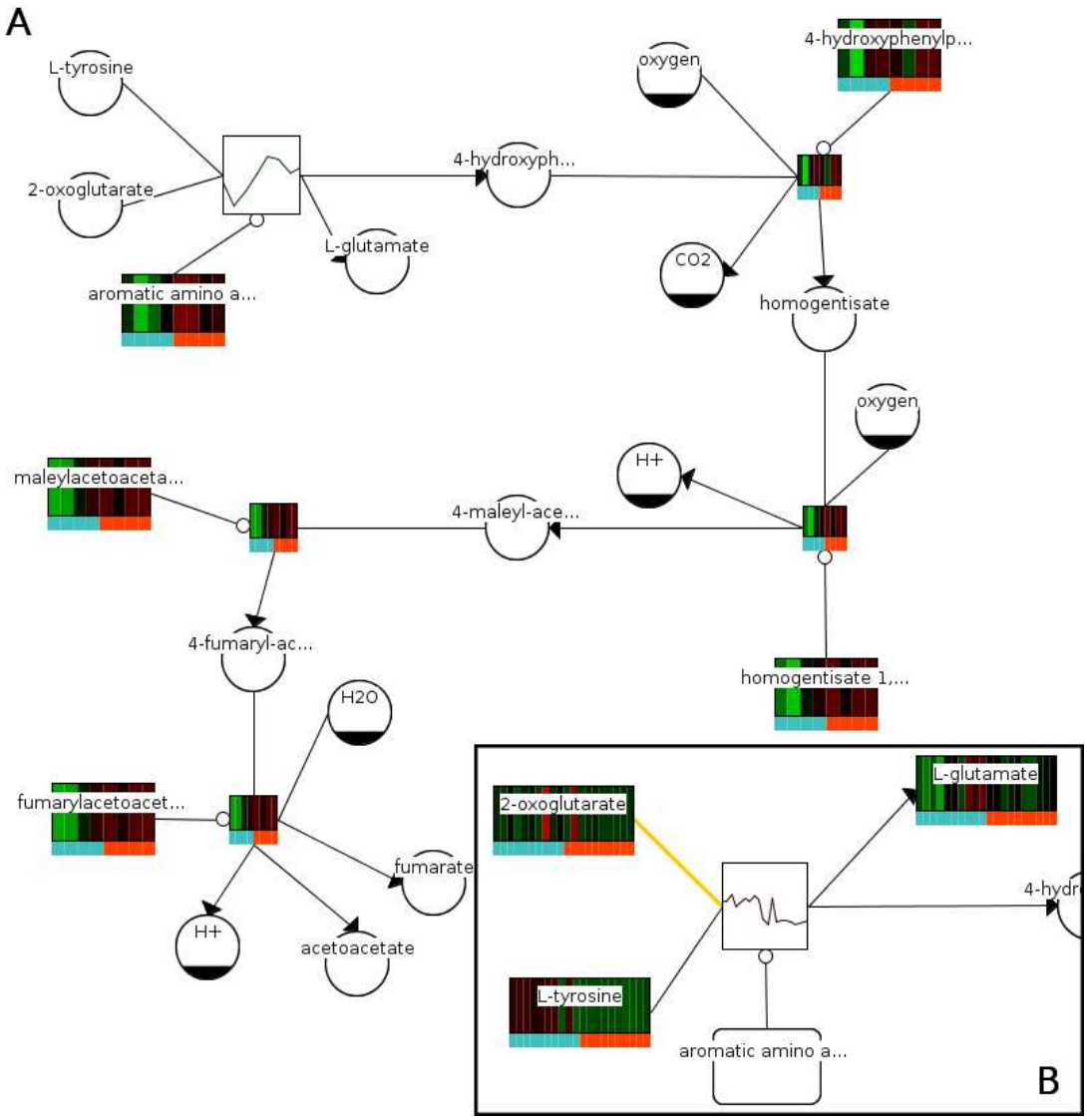


**Figure 6: Tyrosine degradation pathway of *Pseudomonas aeruginosa* with integrated transcriptomic and metabolomic data. For comparability data has been z-score transformed. Experiments have been divided into two classes indicated by the color bars below the heatmaps: anaerobic and low oxygen conditions (light blue), aerobic conditions (orange). (A) The transcriptome data; (B) the metabolomics data.**

organism. The layout and drawing of biological pathways is a subject of continued research. As it is known from general graph layout, optimal solutions are difficult to acquire algorithmically. Despite not necessarily being optimal, our automatically drawn layouts are flexible and fully interactive. In the future we will expand the use of SBGN as a basis for analyzing biological relationships.

We have demonstrated our new tools with poly-omics data from *P. aeruginosa*. The application of data analysis, data perspective visualization and systems perspective visualization shows how MAYDAY can be used to integrate and interpret systems biology data. Using ChromeTracks we identified clusters of high expression variance and distinctive expression profiles. These pointed to amino acid degradation pathways which were further analyzed in both the transcriptome and metabolome dataset. Furthermore, an indirect comparison with respect to the class labels was conducted, yielding insights into the differences between aerobic and anaerobic conditions.

Currently published data sets are not yet exploiting the full potential of these tools, as very few multi-omics data sets are freely available. We are well aware that only limited conclusions can be drawn from related, but not identical studies. Furthermore transcriptome analysis poorly predicts enzyme concentration and activity. However, the case study used here as a proof of concept, shows that the visualization strategy of MAYDAY allows the user to analyze data of different sources and to compare pathways between datasets by introducing class labels. MAYDAY allows cross-dataset poly-omics analysis within an intuitive and efficient framework. The ChromeTracks genome viewer scales well even on large eukaryotic chromosomes, integrates expression and other data easily. The pathway viewer allows use of SBGN which to our knowledge is currently rarely used. Visual analytics features are provided by connection to other views, data transformation and summary features.

The size of systems biology datasets makes analysis and visualization challenging. MAYDAY meets this challenge by a tight integration of analysis and visualization tools. Making most of the data is often a matter of the right visualization. In systems biology, both overviews and detailed views are necessary. In general, MAYDAY can cope with large datasets, with up to 10 million data points, given enough memory available. The largest dataset analyzed in MAYDAY encompassed 110 million data points. ChromeTracks can easily render large eukaryotic chromosomes, including the human chromosome 1 with 55 million bases mapped to genes. In general, the number of mapped base pairs, rather than the length of the chromosome defines the limitation of ChromeTracks. The graph framework underlying the pathway viewers can work on tens of thousands of nodes and edges. Rendering and interaction speed depends on the complexity of the visual representation and the density of the graph. In general practice, neither of these limitations are of concern when working on common datasets.

Our new genomic visualizations are implemented very efficiently and we have already successfully tested them with large expression datasets for the human genome. These plots will be especially useful in the context of the recently developed ultra-high throughput DNA sequencing technologies for measuring gene expression (RNA-Seq) [23] and protein-DNA interactions (ChIP-Seq) [17]. We are currently working on an integration of these new type of data into MAYDAY.

## References

[1] C. Alvarez-Ortega and C. Harwood. Responses of Pseudomonas aeruginosa to low oxygen indicate that growth in the cystic fibrosis lung is by aerobic respiration. *Molecular Microbiology*, 65(1):153, 2007.

[2] K. Arakawa, S. Tamaki, et al. Genome Projector: zoomable genome map with multiple views. *BMC Bioinformatics*, 10(1):31, 2009.

[3] G. Bader, E. Brauner, et al. BioPAX–Biological Pathways Exchange language. *BioPAX Workgroup*, 2005.

[4] F. Battke, S. Symons, and K. Nieselt. Mayday - Integrative Analytics for Expression Data. *BMC Bioinformatics*, in press, 2010.

[5] K. Borowski, J. Soh, and C. W. Sensen. Visual comparison of multiple gene expression datasets in a genomic context. *Journal of Integrative Bioinformatics*, 5(2):97, 2008.

[6] C. Choi, R. Munch, et al. SYSTOMONAS–an integrated database for systems biology analysis of Pseudomonas. *Nucleic Acids Research*, 35(Database issue):D533, 2007.

[7] M. Cuperlovic-Culf, N. Belacel, and A. Culf. Integrated analysis of transcriptomics and metabolomics profiles. *Expert Opinion on Medical Diagnostics*, 2(5):497–509, 2008.

[8] J. Dietzsch, N. Gehlenborg, and K. Nieselt. Mayday - a Microarray Data Analysis Workbench. *Bioinformatics*, 22(8):1010–1012, 2006.

[9] P. Eades, X. Lin, and R. Tamassia. An algorithm for drawing a hierarchical graph. *International Journal of Computational Geometry and Applications*, 6(2):145–156, 1996.

[10] T. Fruchterman and E. Reingold. Graph drawing by force-directed placement. *Software-Practice and Experience*, 21(11):1129–1164, 1991.

[11] Ingenuity Systems. Ingenuity Pathway Analysis. http://www.ingenuity.com/, Dec. 2010.

[12] P. Karp and S. Paley. Automated drawing of metabolic pathways. 1994.

[13] W. Kent, C. Sugnet, et al. The human genome browser at UCSC. *Genome Research*, 12(6):996, 2002.

[14] J. Kim, H. Chung, et al. ChromoViz: visualization of gene expression data onto chromosomes using scalable vector graphics. *Bioinformatics*, 20:1191–1192, 2004.

[15] N. Kono, K. Arakawa, et al. Pathway Projector: Web-Based Zoomable Pathway Browser Using KEGG Atlas and Google Maps API. *PloS One*, 4(11):e7710, 2009.

[16] N. Le Novère, M. Hucka, et al. The Systems Biology Graphical Notation. *Nature Biotechnology*, 27(8):735–741, 2009.

[17] T. Mikkelsen, M. Ku, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448(7153):553–560, 2007.

[18] H. Neuweger, M. Persicke, et al. Visualizing Post Genomics Data-sets on customized Pathway Maps by ProMeTra- aeration-dependent gene expression and metabolism of Corynebacterium glutamicum as an example. *BMC Systems Biology*, 3(1):82, 2009.

[19] H. Ogata, S. Goto, et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*, 27(1):29–34, Jan 1999.

[20] S. Rhee et al. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant Journal*, 37:914–939, 2004.

[21] P. Shannon, A. Markiel, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498, 2003.

[22] T. Tokimatsu, N. Sakurai, et al. KaPPA-View. A web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. *Plant Physiology*, 138(3):1289, 2005.

[23] Z. Wang, M. Gerstein, and M. Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63, 2009.

# Supplement to “Integrative Systems Biology Visualization with Mayday”

## Genome HeatStream Visualization

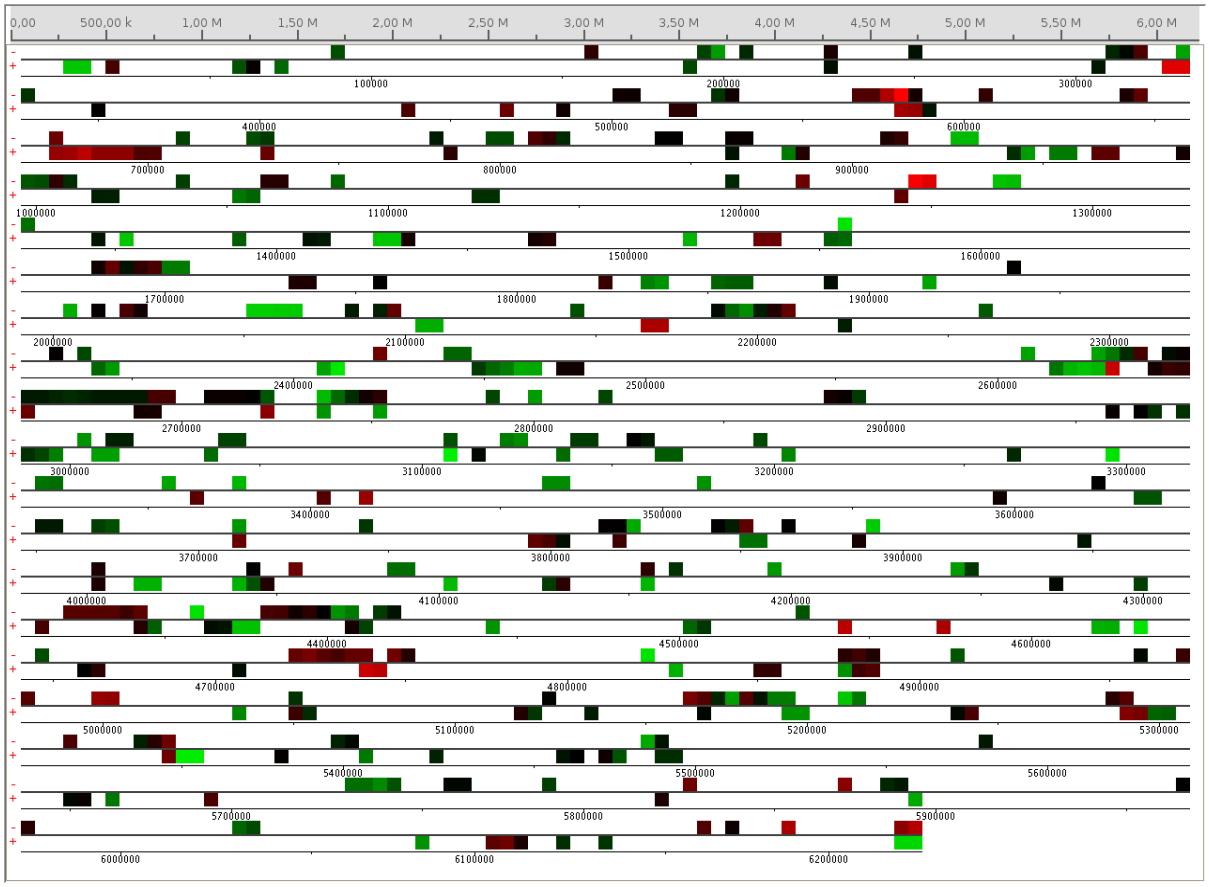


**Figure 1: Genome Heat stream visualization of highly variant genes in *Pseudomonas aeruginosa* (variance across all experiments $> 0.5$).**

## Nitrogen Metabolism Pathway

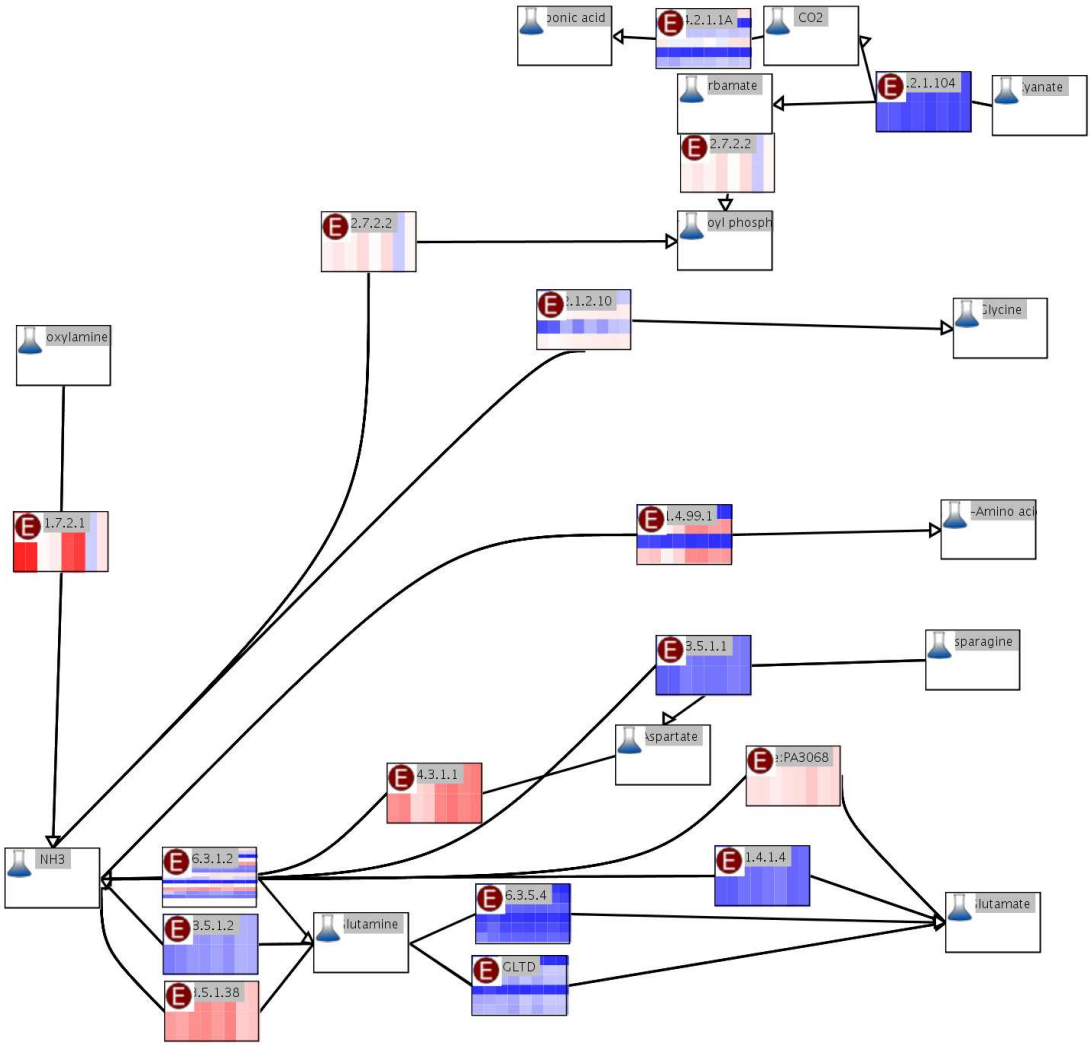


**Figure 2: Amino acid-related part of the nitrogen metabolism of *Pseudomonas aeruginosa* (KEGG pathway id pae00910). All probes associated with the relevant enzymes are shown, either as profiles or as heatmap rows. Expression values are encoded as colors (low=blue, high=red) and as profiles. Conditions are ordered ascendingly by oxygen concentration (anaerobic, 0.4%, 2%, 20%), with both replicates per condition shown as seperate adjacent experiments. Enzymes are marked with the letter "E", metabolites are marked with the image of an Erlenmeyer flask. Components are labelled with names assigned by KEGG.**