

Version 2.0
Date: 7/15/2004
Report # 12

Pathway Database System 2.0

Overview

Center for Computational Genomics

EECS Department
Case Western Reserve University
Cleveland, OH 44106

Department of Genetics
Case Western Reserve University
Cleveland, OH 44106

1. Introduction

Biological pathways are used for presentation and modeling genomic data, studying genome sequences in the context of biology, and in functional studies of attributing functions to genes. Pathways are the sequential and cumulative action of genetically distinct, but functionally related molecules. Each reaction in a pathway is a biochemical step from specific substrates (input molecules) to products (output molecules) that are chemically modified substrates. Each step may also use various combinations of molecules as cofactors, activators, inhibitors, and regulators, and usually involves at least one genetically unique gene product that catalyzes the reaction step.

Pathways, in general, illustrate the functional relations between molecules. Functional annotations include, for example, the identity of the substrate(s), product(s), cofactors, activators, inhibitors, enzymes or other processing molecules, RNA and protein expression patterns, reaction kinetics, and associated phenotypic variation and diseases. Ultimately, many other kinds of information (or knowledge) can be incorporated. Such information forms a rich research resource that integrates genomic and biological information which can be managed, analyzed, queried and displayed in dynamic ways at various levels of biological and genetic detail to provide insight into diverse biological processes in health and disease.

Pathways Database System is an integrated software system for storing, managing, analyzing, and querying biological pathways at different levels of genetic, molecular, biochemical and organismal detail. At the computational level, Pathways Database System allows users to visualize pathways in multiple abstraction levels, and to pose predetermined as well as ad hoc queries using a graphical user interface. Pathways are represented as graphs, and implemented as a relational database. Pathways Database System has multiple levels, with multiple tools at each level. Pathway Database System 2.0 includes a set of tools (Figure 1):

- **Metabolic Pathway Browser** – a tree viewer to browse metabolic pathways at pathway, process and molecular entity levels (Figure 2-a)
- **Signaling Pathway Browser** – a tree viewer to browse signaling pathways at pathway, signaling step and molecular entity levels (Figure 2-b) (*new for release 2.0*)
- **Pathway Viewer** – a graphical tool to view and edit pathway graphs and genes (Figure 3) (*Genome Viewer is new for release 2.0*)
- **Pathway Explorer** – a tree-structured query interface (Figure 4) (*new for release 2.0*)
- **Pathway Editor** – a plug-in tool to create new pathways and update pathways (completely revised for release 2.0)
- **Pathway Web Services** – an XML web service to query the pathway database (Figure 6)

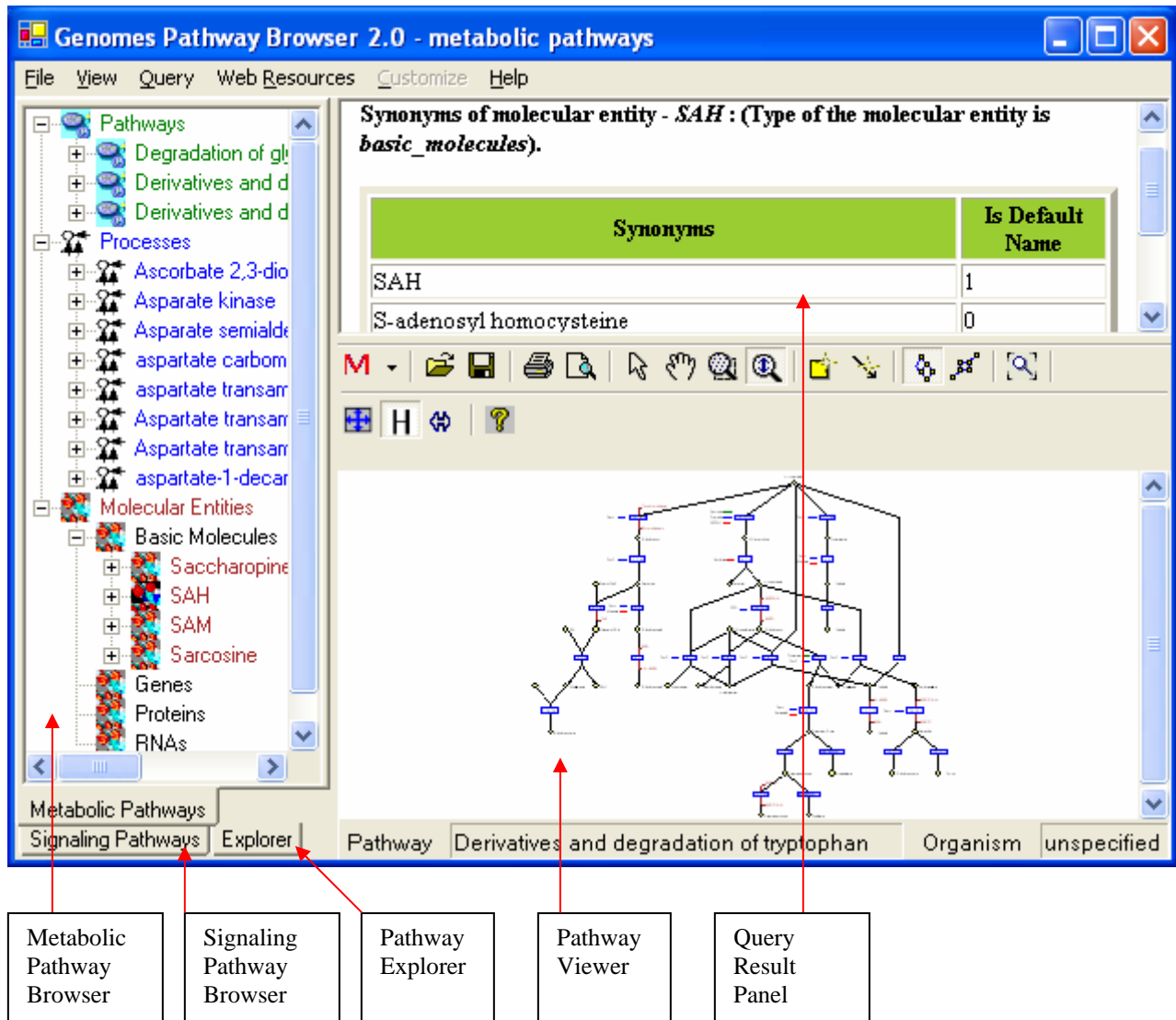
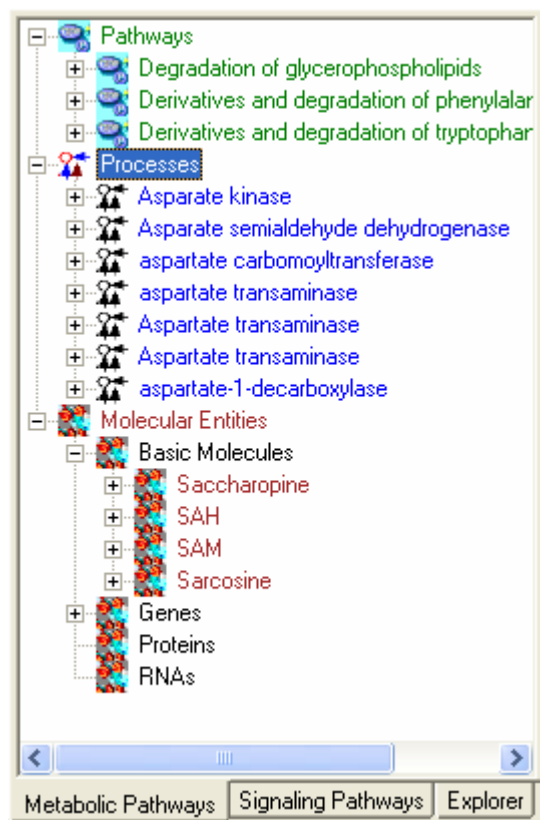


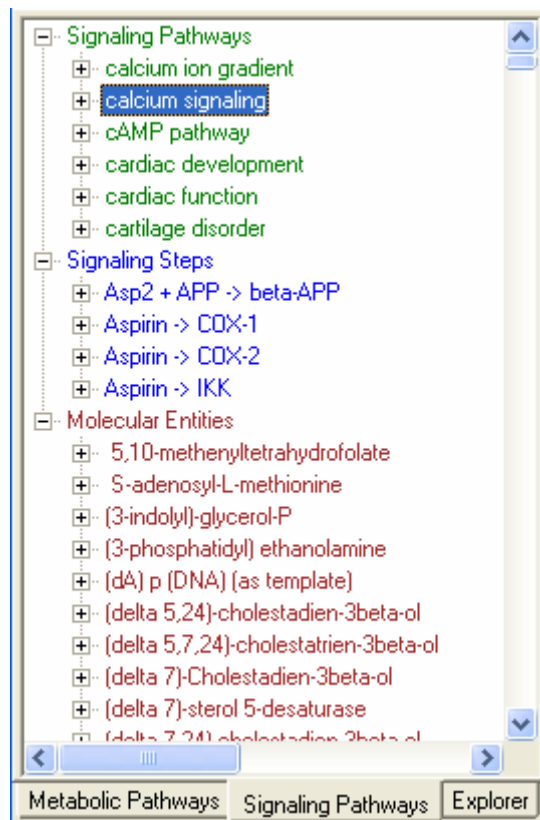
Figure 1. Snapshot of the Pathway Database System main GUI.

2. Pathway Browser

Pathway Browser provides a tree-structured view of the metabolic pathways, and signaling pathways available in the database. Metabolic pathways, processes, and molecular entities can be browsed in different orders (Figure 2-a), and a set of useful queries can be specified about pathways, processes and molecular entities, from predefined simple queries to parameterized path selection and neighborhood queries. Similarly, for signaling pathways, signaling steps, and signaling and signaled molecules and their components can be browsed; a set of useful queries for signaling pathways are presently being developed (Figure 2-b). When one right-clicks on a tree node, dynamic context menus will popup, which leads to a set of functionalities. For more details about the Pathway Browser, see [KL03].



a



b

Figure 2. Metabolic Pathway Browser (a) and Signaling Pathway Browser (b).

3. Pathway Viewer

Pathway Viewer presently visualizes a single metabolic pathway (Figure 3) or multiple metabolic pathways in an organism-specific or organism-independent manner, provides a large number of functionalities including collapsing-zooming, hiding-expanding, editing, and node property-setting. The tool supports ad-hoc querying of the pathways data as well as menu-driven queries. It also provides a seamless transition from metabolic pathways to signaling pathways, and from signaling steps that belong to metabolic pathway processes into the relevant pathways. Pathway viewer for signaling pathways is currently under construction. For more details on the Pathway Viewer, see [MY04].

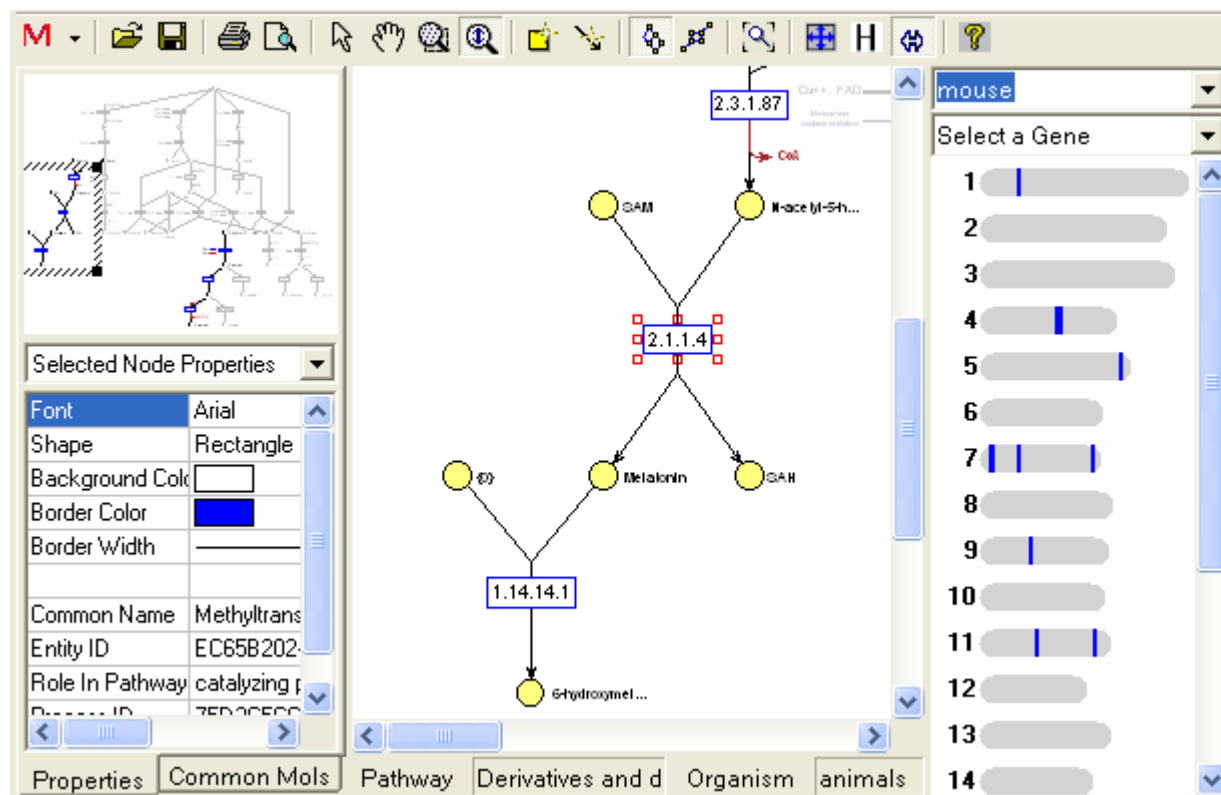


Figure 3. Pathway Viewer. The middle pane is the main display window for pathway graph. The right pane is the Genome Viewer, which displays genes for the pathway in the middle pane. The left pane is the Overview window and Property window for graph objects. The genome Viewer and the Overview/Property panel can be displayed or hidden as needed.

4. Pathway Explorer

Pathway Explorer (Figure 4) provides a generic user interface for posing a large class of queries about pathways using tree-structured views. Entities involved in the pathways database can be better viewed by generalization/specialization hierarchies, e.g., classes of pathways, super pathways, functional classifications in arbitrary granularities, etc.. In addition, pathway components, i.e., processes in a pathway, and molecular entities involved in each process can naturally be viewed in the form of a tree-structured (hierarchical) view. Motivated by this observation, Pathways Explorer is a query interface which allows users to query the database using tree-structured views. As an example, consider the tree-structured (hierarchical) schema Pathways-Process-Molecular entity, where a pathway consists of several processes, and each process is associated with several molecular entities. This may also be visualized as a logical schema of a nested relation (Pathways (Process (Molecular_Entity)*)*), or an XML view. Thus, Pathways Explorer can also be considered as a query interface for querying tree-structured views for a relational database.

Given the schema of an XML view, or a nested relation view of the underlying relational database, query interface (template) structure follows that of the view schema. The schema declares that pathway elements contain process elements that contain molecule elements, so the tree in the user-interface would contain only these nodes, along with the node attributes as children. Each attribute's possible values may then be seen

individually by opening its drop-down box. Choosing a value for an attribute effectively creates a query constraint declaring that only nodes with the chosen attribute value are of interest.

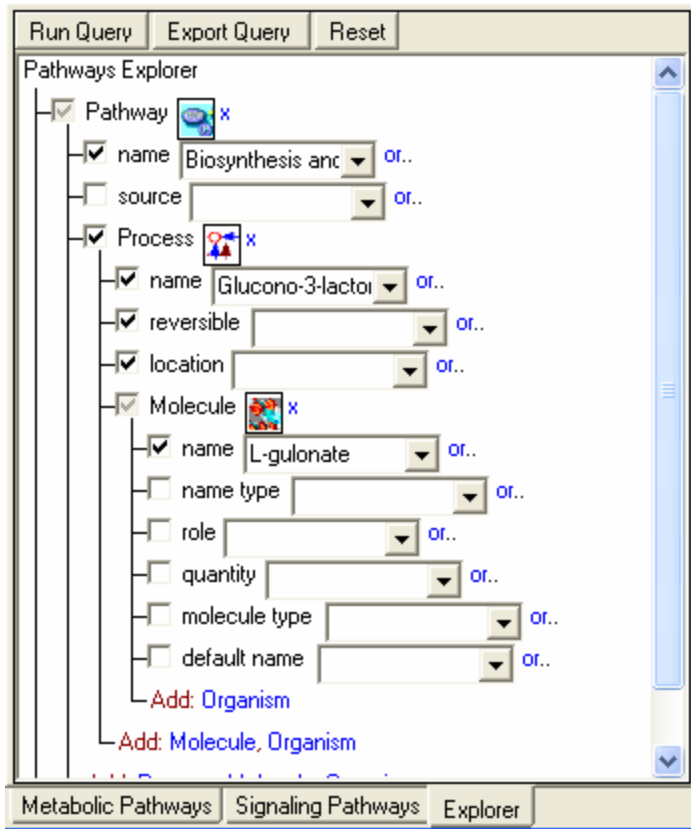


Figure 4. A snapshot of Pathway Explorer.

5. Pathway Editor

Pathway Editor allows view-based database updates; that is, users can insert/update/edit pathways via the high-level user views of the pathway viewer. The pathway editor, while enforcing the integrity of the existing database (sometimes interactively with user input), maps them as updates to the underlying relational database. This component is currently functional for metabolic pathways. Pathway Editor is initiated from context menu of the tree node in Pathway Browser.

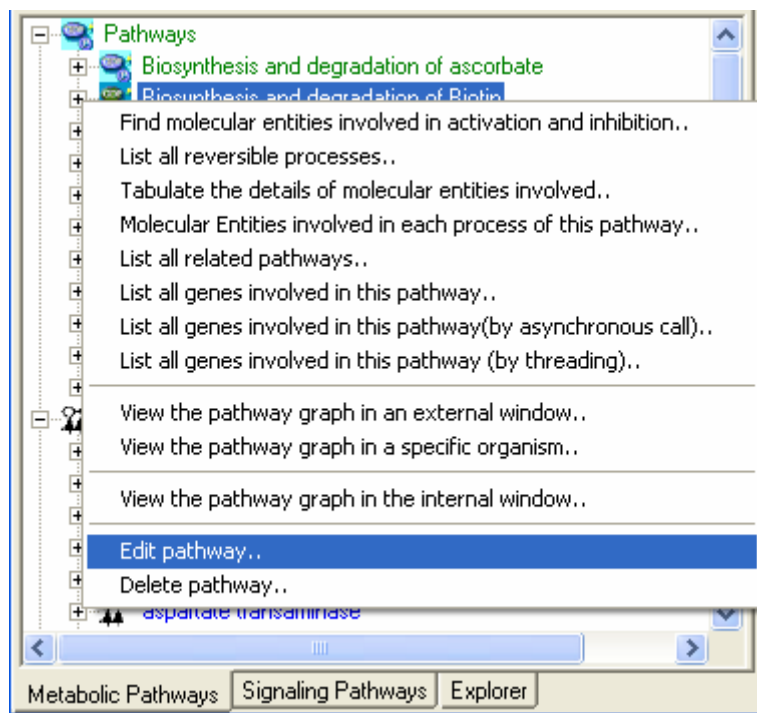


Figure 5. Snapshot to launch Pathway Editor.

6. Pathway Web Services

The Pathway web services expose querying functions to Internet. Each function implements a query. Pathway tools in any machine can use SOAP (Simple Object Access Protocol) to invoke these functions, and receive query results. We have implemented over thirty functions. Querying functions are categorized into four types: molecular entity queries, reaction (process) queries, pathway queries and path computations. These queries are illustrated in Figure 6, and can be invoked from the website http://nashua.cwru.edu/pathways_metabolic_1/service1.asmx

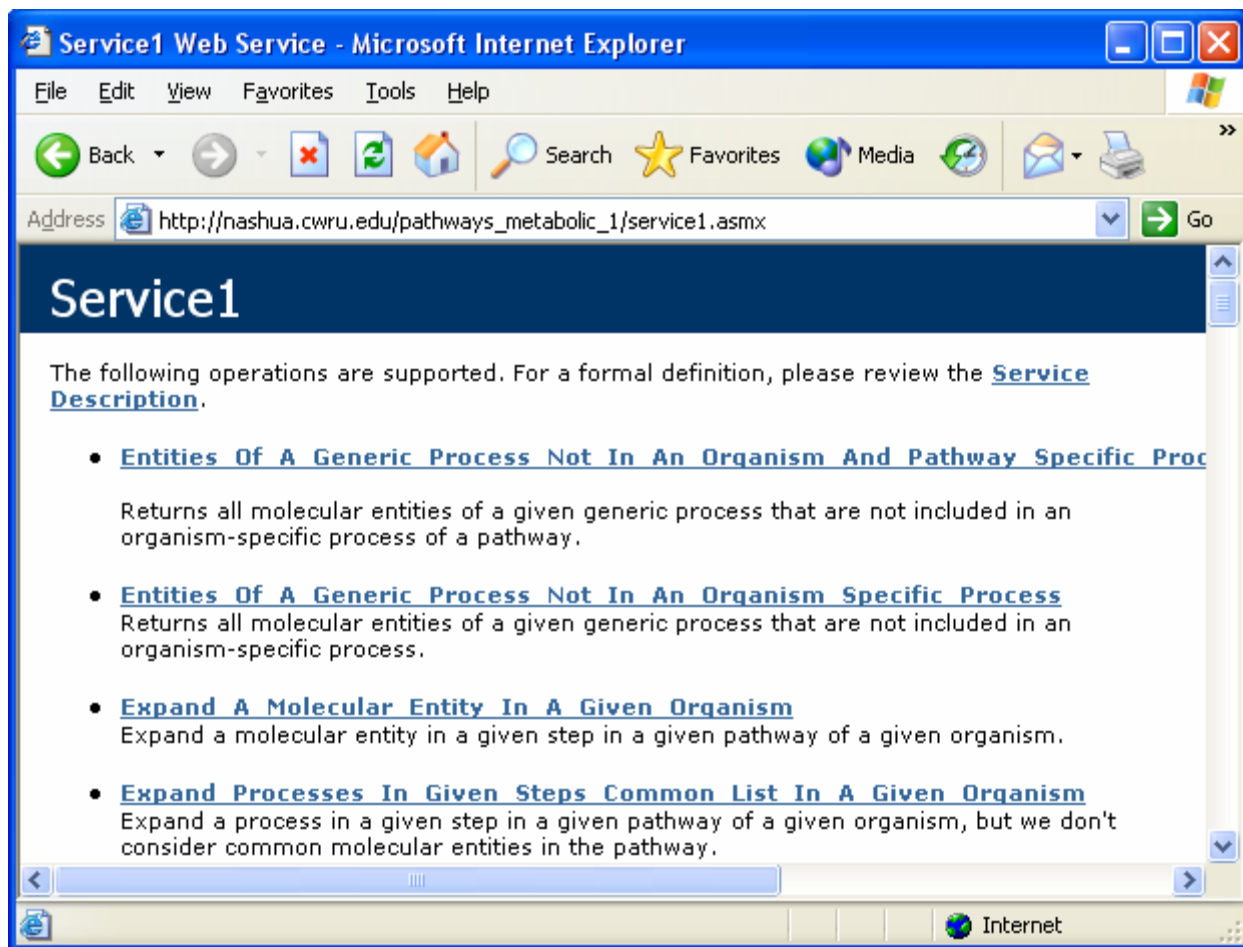


Figure 6. Pathway web services.

7. What is New for Release 2.0?

Since the release of version 1.0 on July 2003, more tools and features are added in the current release. We redesigned the database to capture gene and organism information in an integrated manner. The new release offers more reliable operations, better performance and enhanced services.

- The Genome Pathway Browser takes the organism information into account. Organism can be viewed in different levels. In the application, users can set the organism information from three components (Figure 7).
 - a) Set organism from the Pathway Browser,
 - b) Set organism from the queries in the main menu,
 - c) Set organism from the Pathway Viewer.

In Pathway Browser, users can set organism information at different levels. The organism setting of the top-most tree node controls its descendant organism settings. The organism settings can stay the same or be changed from more general to less general going from ancestors to descendants in the hierarchy. (For example, when the top-most tree node, say Pathways, organism property is set to "unspecified", users can set its descendant nodes, say processes, to a specific organism, say mouse, to browse the processes in mouse that belong to the given general pathway).

- Genome Viewer is added and gene information is updated. Our current database stores thousands of human and mouse genes and proteins with cross links to Locus Link and Swiss-Prot. Users can query gene paralogs and orthologs from the Pathway Browser. With the Genome Viewer, genes for a pathway or a process can be drawn on the corresponding chromosome. From the genes in the Genome Viewer, users can initiate queries such as highlighting the processes on the graph that the selected gene is involved in. Currently, only Folate and Homocysteine pathway has complete human and mouse gene data [KT03]. The remaining pathways data is from the Gene Atlas [MG98] where instead of specific organisms, organism groups, (e.g. animal, plants) are provided.
- Pathway Explorer is a new tool for this release.
- Signaling pathway data is available to the browser. Signaling Pathway Browser is ready to use. However, the functionalities for drawing a signaling pathway as well as queries are not yet available at current release.
- *Better performance.* Improved database design helped significantly for a better overall performance. For example, by exploiting the improved database design, the response time for the query “get the connected pathways” in the data viewer is now less than 2 seconds as compared to an average of more than 10 seconds in the previous version.
- *Easy installation.* Utilizing the .NET setup makes this release a no-touch deployment. With no-touch deployment, the applications can be downloaded, installed, and run directly on the users' machines without any alteration of the registry or shared system components.

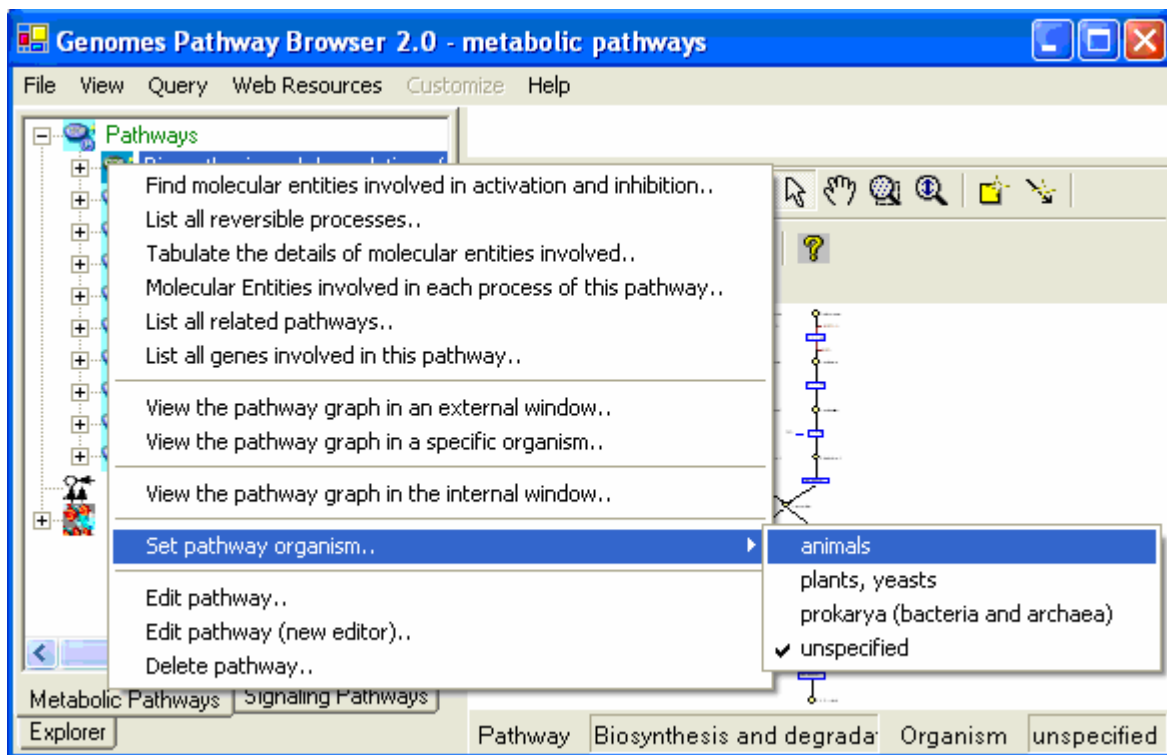


Figure 7.a. Choosing Organism from Pathway Browser

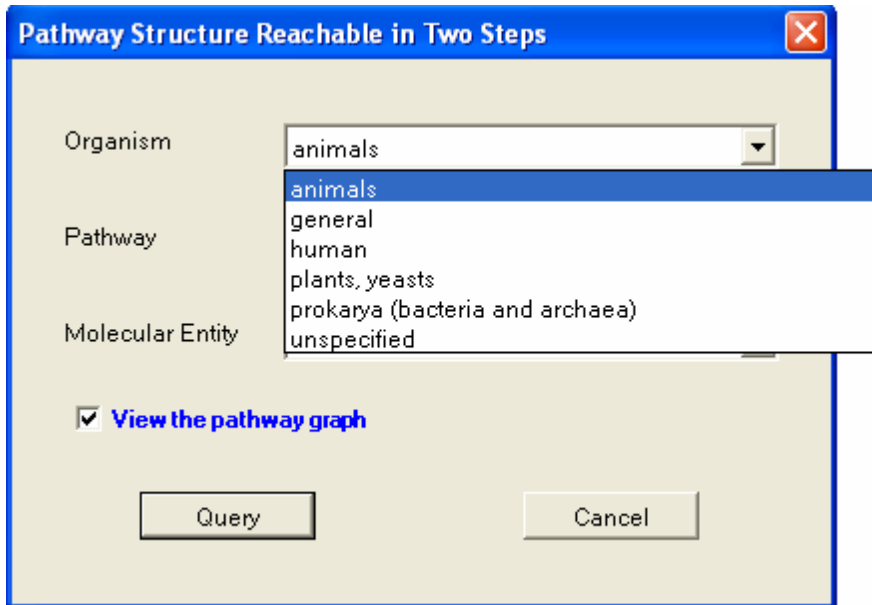


Figure 7.b. Organism in queries from the main menu

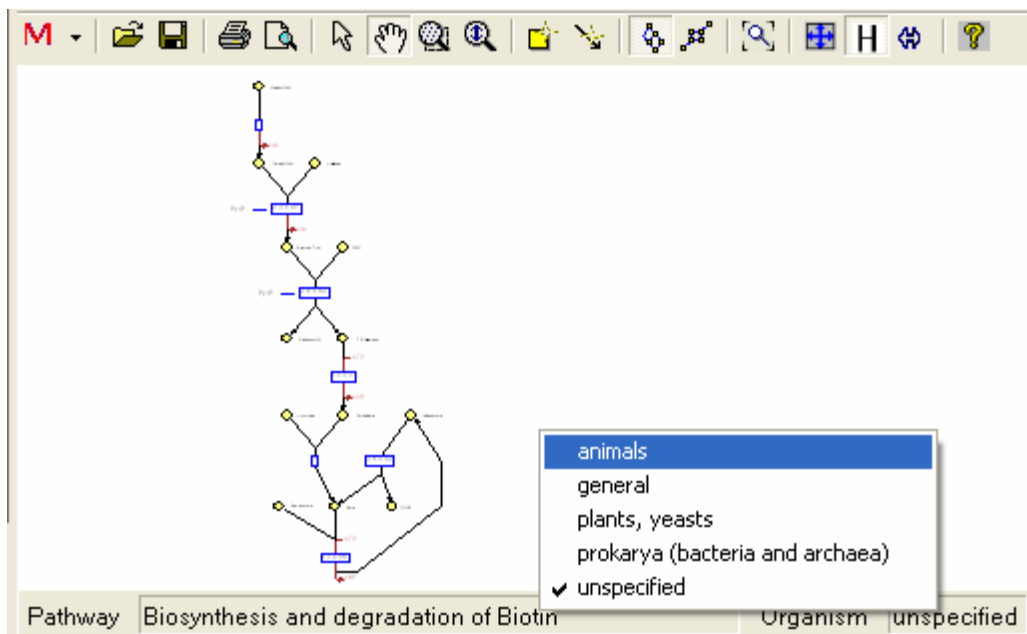


Figure 7.c. Choosing Organism from Pathway Viewer

Figure 7. Choosing organism from Pathway Browser and Pathway Viewer

References

[KL03] Khrishnamurthy, L., et. al., Pathways Database System: An integrated set of tools for biological pathways, SAC 2003, March 9 - 12, 2003, Melbourne, Florida, USA.

[MY04] Mei, Y., "Pathway Viewer Tool--Release #13", users manual, Computational Genomics Center, EECS and Genetics Departments, CWRU, July 2004, available at <http://nashua.cwru.edu/Pathways>

[KT03] KITAMI, T., Data for Folate and Homocysteine Pathways in Mouse and Human, Department of Genetics, CWRU, Unpublished Manuscript, 2003.

[MG98] Michal, G., Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, Wiley-Spektrum, 1998, ISBN: 0471331309.