Payao: a community platform for SBML pathway model curation

Yukiko Matsuoka[1,2], Samik Ghosh[1], Norihiro Kikuchi[3] and Hiroaki Kitano[1,4,5,*]

[1]The Systems Biology Institute, Tokyo, [2]JST ERATO Kawaoka Infection-induced Host-response Network Project, Tokyo, [3]Mitsui Knowledge Industry Co. Ltd, Tokyo, [4]Okinawa Institute of Science and Technology, Okinawa and [5]Sony Computer Science Laboratories, Tokyo, Japan

Associate Editor: Jonathan Wren

ABSTRACT

Summary: Payao is a community-based, collaborative web service platform for gene-regulatory and biochemical pathway model curation. The system combines Web 2.0 technologies and online model visualization functions to enable a collaborative community to annotate and curate biological models. Payao reads the models in Systems Biology Markup Language format, displays them with CellDesigner, a process diagram editor, which complies with the Systems Biology Graphical Notation, and provides an interface for model enrichment (adding tags and comments to the models) for the access-controlled community members.

Availability and implementation: Freely available for model curation service at http://www.payaologue.org. Web site implemented in Seaser Framework 2.0 with S2Flex2, MySQL 5.0 and Tomcat 5.5, with all major browsers supported.

Contact: kitano@sbi.jp

1 INTRODUCTION

Creating an extensive model of gene-regulatory and biochemical networks with the latest data is a painstaking task. Curation is essential to creating an accurate model. Yet as science and technology advances rapidly, once curated models soon become out-of-date and need to be revised constantly. Many pathways and networks are now available online via pathway databases, such as Reactome, BioModels.net, Panther Pathways and many pathway editors are available (Bauer-Mehren *et al.*, 2009). What is needed is a framework to facilitate tracking and update mechanism for modelers and researchers in the community to contribute to the collaborative model building and curation process.

WikiPathways (Pico *et al.*, 2008) is an effort for such a collaborative platform in the Wiki style. While the Wiki system has its strength in collaborative editing and version tracking, it does not provide access control or explicit community tagging mechanisms. In a community-driven model enrichment environment, it is effective to differentiate privileges to special interest group (SIG) members for curation activities—commenting on existing tags, adding tags to models, annotating individual component inside a model and validating the annotations. In view of the complexity of biological pathways and the expertise of biologists in different areas,

a community platform for biology requires an exquisite balance of federated resource sharing and quality control of information by a SIG of experts in the particular pathway or process. An access control privilege system allows the community to share and disseminate the knowledge, while enabling a dedicated SIG to maintain high-quality, curated information.

To provide such a curation framework, we have developed a system called 'Payao'. The system is named after a fish aggregating device, an artificial floating raft where fish congregate and popular in Okinawa/Philippine area. Payao aims to become a biological knowledge aggregating system, which enable a community to work on the same models simultaneously, insert tags as pop-up balloon to the parts of the model, exchange comments, record the discussions and eventually update the models accurately and concurrently.

The current workflow for pathway curation has two phases working in a cyclical manner, as shown in Figure 1: pathway editing using biological pathway editors (CellDesigner) and community-driven pathway enrichment and knowledge sharing. Payao serves for enrichment phase of the curation. Payao is a web-based platform, providing an interface for adding tags and comments to the components (such as Species, Reactions and specified area) of the model, as well as community management functionality. The information on the users and tag data is stored in a relational database (RDBMS) on the server. Payao adopts community standards, accepting Systems Biology Markup Language (SBML; Hucka *et al.*, 2003) format models and displays them in Systems Biology Graphical Notation (Le Novere *et al.*, 2009) compliant CellDesigner (Funahashi *et al.*, 2008) graphical notation. Curation data on Payao can be easily reintegrated into the original model via CellDesigner.
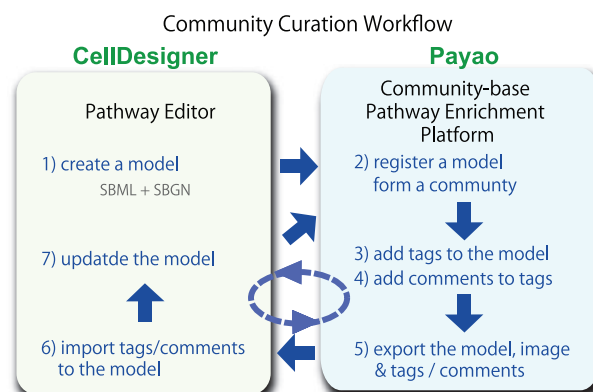

Fig. 1. Workflow of community pathway curation.

2 PLATFORM

Payao consists of the server application, client user interface and database. The server application has been implemented in Java on Seaser Framework 2.0 with S2Flex2. Tomcat 5.5 was used as the servlet container to build the web application. The client user interface has been implemented in ActionScript on Flex framework 2.0.1, which allows us to build a Rich Internet Application to visualize SBML models. The server communicates with the client via the Action Message Format (AMF3) protocol on S2Flex2, which enables us to translate between Java Objects in Seaser and Action Script Object in Flex. The server can handle CellDesigner models including SBML models and the visual information using CellDesigner API ver.4.0. It parses SBML files sent from client to create CellDesigner models, and provides the information as CellDesigner Plugin classes. The client application receives the model information and draws the model. MySQL 5.0 is used to store information on user, model and tags/comments in the database.

3 FEATURES

As Payao accepts pathway models stored in SBML format and uses CellDesigner APIs for visualization, the most suitable SBML editor for Payao is CellDesigner. In SBML format, models can capture details of biochemical process descriptions, not only protein–protein interactions. Adopting SBML format enables the models to be easily used as the base of computational data analysis or simulation of dynamic behaviors. The Payao platform enriches the model curation process by providing a host of features for user management, tagging and model updates [detailed are available for reference in (Payao User guide, 2009)].

3.1 Community management

Forming a community is an important step for curation. Different expertise groups can contribute variety of information to the model. As web-based Payao can be accessible from all physical locations, it enables experts across the world to communicate in a collaborative curation effort.

Community is formed around a pathway model. It is the model owner who sets access control over the registered model. In the Payao system, access controls can be set by specifying the privileges to individuals as well as to user categories, such as guest, login user and model user (who are invited to access the model by the model owner). This enables a user to stage the curation process, initiate the curation within a small group (e.g. SIG) and then switch the access control of the model for public viewing.

3.2 Model management

The model owner registers and manages the SBML model. Upon registration, the model owner specifies the basic model information including, thumbnail image, references and copyright. The owner sets the access privilege to the user in three levels (browsing, adding tags, adding comments) by user categories or by individual users. The registered models can be sorted by Register Date and by Popularity. Popularity is measured by the activity level (number of tags and comments) and ranked in the list. All the registered models are listed with the thumbnails in the top screen in the right panel (Fig. 2). Registered models are stored in the database.
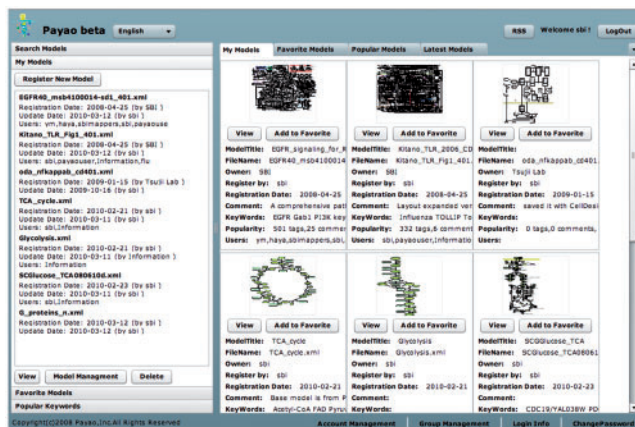

Fig. 2. Payao Model Browser screen—browse the registered models in the list format (left) and the thumbnails with summary data and statistics (right).

3.3 Community tagging and commenting

The tagging on the visually represented pathways is a characteristic of Payao, which makes the curators easy to grasp the nature of the pathway while discussing on the specific component of the pathways. Like Google Maps, tags are displayed in a bubble form attached to the items (Species, Reactions or any specified area), and click to expand and display the content of the information in the tag. Tags can specifically be keywords, links, PubMed IDs as well as free text, as shown in Figure 3. A TagSet groups a set of tags and can be color coded for ease of viewing. User-defined TagSets also allow access control features (browsing, editing and editing tags) to be set by a user, in the same way as for model access privilege settings. Thus, a user can set permissions to a 'My Tagset', which hides comments and tags from the community members. Inside the tag, comments can be added in the free text format. While tags anchor the points for annotation in the model, comments function serves as the discussion space.

3.4 Model update: tag data export and import

Data in the tags can be exported from the Payao system, evaluated outside the system and integrated into the Payao system or to the SBML model file. All the information, including the model file (CellDesigner format), Tags/Comments and model image (png/jpg/pdf format) in a ZIP file format are exported in a batch. Information stored in Tags is exported as a table in .csv format, which can be edited and imported back to the system. The tag information can also be imported into the base CellDesigner-format model as 'notes' using a CellDesigner plugin (import notes) and stored in SBML 'notes' annotation. Once the base model has been updated using CellDesigner, the model owner can reregister the model onto Payao for further curation.

4 FUTURE DEVELOPMENT

Payao will facilitate the evolution of CellDesigner from a simple tag-based pathway curation tool to a more versatile, comprehensive platform. The next version of CellDesigner will adopt support for Minimum Information Required in the Annotation of Models (MIRIAMs; Le Novère *et al.*, 2005) standard for annotation and
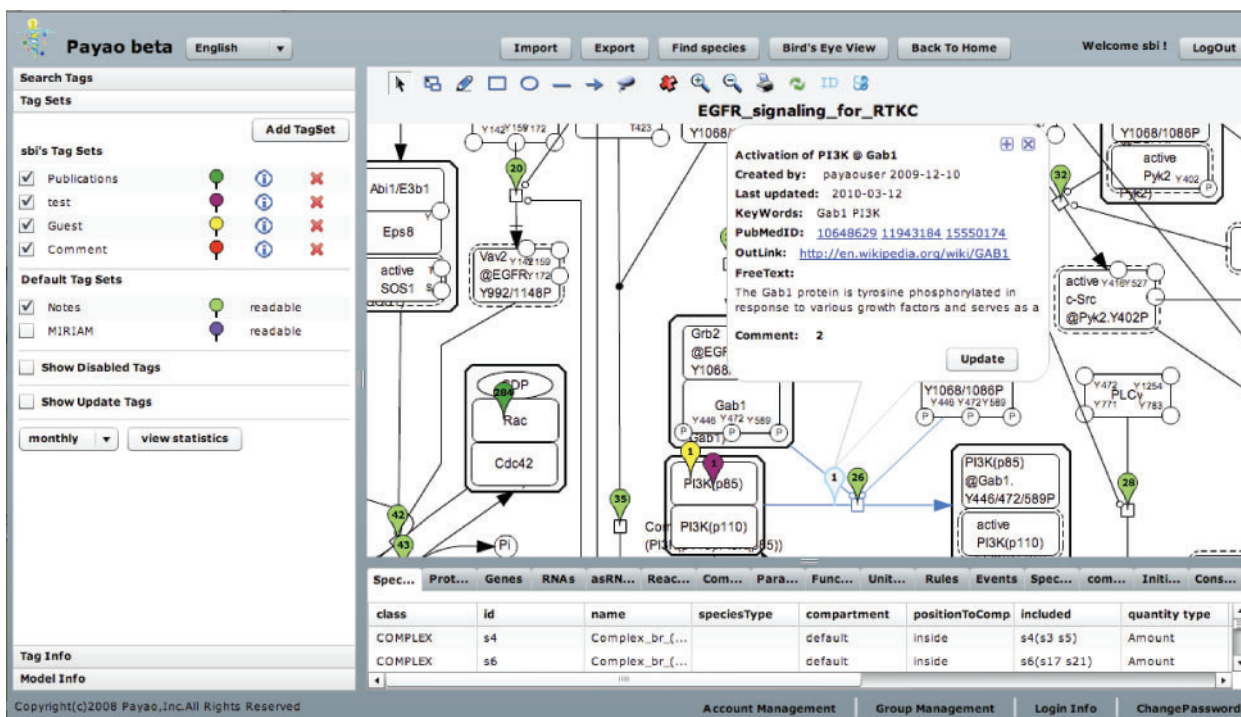
Fig. 3. Payao Model Viewer screenshot—model view panel and tag example. In the left panel, tagsets are listed to identify by specified color. In the right panel, under the tagset, assigned model users can add/edit tags associated with the nodes/reactions. Other users can add comments to the tag for providing further information or raising an issue against the tag contents.

allow MIRIAM annotations to be viewed on Payao. Modules to support import of different pathway exchange formats (BioPax), visual mapping of experimental data on pathway components in Payao, would be developed in future versions. Future enhancements also include integration of Payao with biological text mining techniques to facilitate literature-driven knowledge enrichment of existing pathway models. Tracking model updates using an RSS (Really Simple Syndication) feed will assist curators as well as public viewers. Statistics for measuring contribution of individuals to the model curation process as well as community interest (measured by count of tags and comments on a model) would be integrated into Payao in future releases. In the long run, we envisage this system to be a platform for aggregation, dissemination and community exchange for biological pathway models.

Conflict of Interest: none declared.

REFERENCES

Bauer-Mehren,A. *et al.* (2009) Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol. Syst. Biol.*, **5**, 290.

Funahashi,A. *et al.* (2008) CellDesigner 3.5: a versatile modeling tool for biochemical networks. *Pro. IEEE*, **96**, 1254–1265.

Hucka,M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

Le Novère,N. *et al.* (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.*, **23**, 1509–1515.

Le Novère, N. *et al.* (2009) Systems biology graphical notation. *Nat. Biotech.*, **8**, 735–741.

Pico,A. *et al.* (2008) WikiPathways: pathway editing for the people. *PLoS Biol.*, **6**, e184.

Payao User Guide. (2009) Available at http://celldesigner.org/payao/payaopreview.html (last accessed date April 1, 2010)