

## BioPlace: A Web-Based Collaborative Environment for Effective Genome Research

AHN, GEON-TAE<sup>1</sup>, JIN-HONG KIM<sup>1</sup>, KYUNG-MI KANG<sup>2</sup>, MYUNG-JOON LEE<sup>1</sup>, AND IN-SEOB HAN<sup>3\*</sup>

<sup>1</sup>School of Computer Engineering & Information Technology, University of Ulsan, Ulsan 680-749, Korea

<sup>2</sup>Department of Biology Education, Seoul National University, Seoul, Korea

<sup>3</sup>Department of Biological Science, University of Ulsan, Ulsan 680-749, Korea

Received: October 29, 2003

Accepted: March 12, 2004

**Abstract** Genome research has become very popular in most nations. In order to enhance the efficiency of collaboration among genome research groups, ways to store and share data, communicate with each other, be guided through right research strategies, and to easily use well-established databases. In addition, since techniques and softwares for genome research groups are well established, a similar research road map could commonly be applied. In this study, we developed a web-based work place for effective genome research, named 'BioPlace.' From the beginning of writing a proposal, research members can work on the same environment with convenient aid to share files or data. BioPlace provides various ways of collaboration methods among genome researchers. The BioPlace system supports two types of workplaces, namely 'Personal Workspace' and 'Team Workspace.' For each BioPlace user, a Personal Workspace is provided, while a Team Workspace is provided for each group with the same purpose. In addition, BioPlace provides a 'General Research Road Map' for genome research, and several Korean user interfaces for BLAST, PDB, and Primer3. We expect that BioPlace may facilitate collaboration of genome research among the experienced scientists and help beginners in many different ways as well.

**Key words:** Genome research, research road map, web-based work place, K-BLAST, K-PDB

Genome research has become very popular in most nations, because comparative analyses of organisms' genomes begin to uncover many useful informations. Since techniques and softwares for genome research groups are well established, a similar strategy should commonly be applied. By the time the human genome had been sequenced, the complete sequence was already known for over 30 bacterial genomes

plus that of a yeast (*Saccharomyces cerevisiae*; 1996), the fruit fly (*Drosophila melanogaster*; 2000), a nematode (*Caenorhabditis elegans*; 1998), and a plant (*Arabidopsis thaliana*; 2000).

Genome research requires laborious and painstaking work, which includes cloning, sequencing, and application of experimental methods to uncover genes embedded in large genomic fragments. In order to enhance genome research, a collaborative environment for research members to communicate their data, tools, and other informations may be required [2]. Also, a genome research needs a data store to accumulate new data and share information, and research groups should be guided through feasible research strategies and have access to well-established databases. For example, when the goal to sequence the entire genome of the budding yeast *Saccharomyces* (15Mb) was set, European nations constructed a consortium and segmented the task. Indeed, it was the result of a multicenter collaboration [4].

In this study, we developed a web-based workplace for cooperative genome research, named 'BioPlace.' The World Wide Web (WWW) is a way of using the Internet through a web browser. A useful feature of the web is its complete accessibility [3]. From the beginning of writing proposals, research members can work on the same workplace with convenience of sharing files. BioPlace contains two data storage places, named Personal Workspace and Team Workspace, which are the main components of the system, a schedule management scheme among the related group members. In addition, it also includes a genome research road map and a set of sequence search programs to enhance the efficiency of research activities such as K-BLAST, K-PDB, and K-Primer, which are Korean versions of BLAST, PDB, and Primer3, respectively.

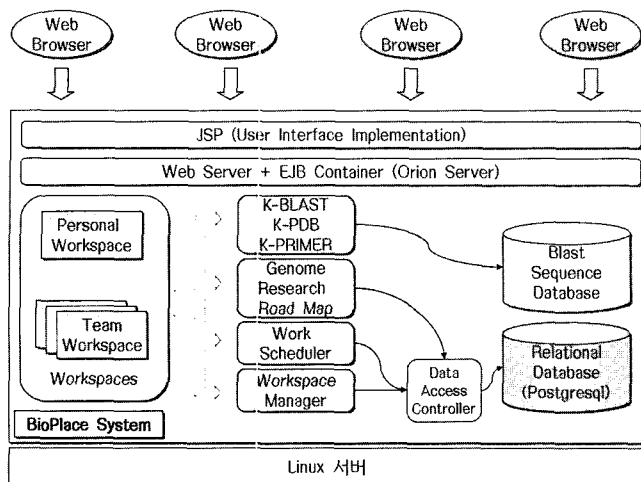
### Systems and Methods

Genome research requires useful tools to analyze enormous amounts of biological data and to communicate among

\*Corresponding author

Phone: 82-52-259-2352; Fax: 82-52-259-1694;

E-mail: hanis@ulsan.ac.kr



**Fig. 1.** BioPlace architecture.

researchers from different fields, for example, molecular biologist, bioinformaticist, and programmer. BioPlace is a collaborative system designed mainly for this purpose. Users are able to access BioPlace through a web browser. The workspaces in BioPlace are implemented by the EJB (Enterprise Java Beans) technique, which is a typical distributed architecture for web-based server systems [1]. As shown in Fig. 1, BioPlace contains workspaces, work schedulers, K-BLAST, K-PDB, and K-Primer.

### Workspace

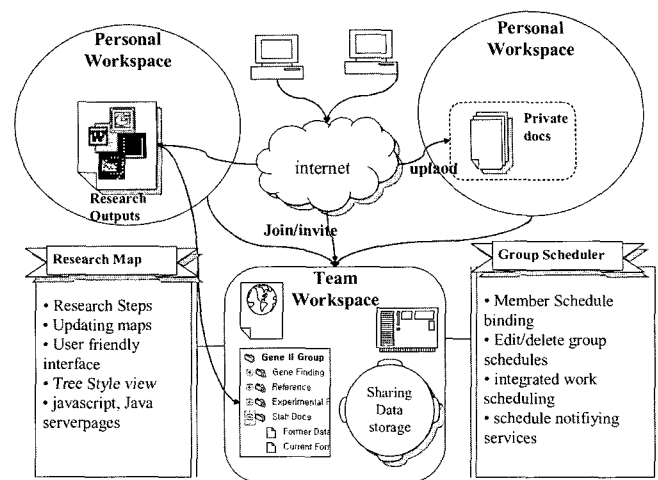
To conveniently share data and effectively communicate, Workspace supports a common working environment to genome researchers. Workspace includes Personal Workspace and Team Workspace. While Personal Workspace was designed for researchers to manage their individual information, we designed Team Workspace for researchers to share their own data with other members in the same group. In addition, Research Road Map was included for users to follow a track, depending on their purpose.

### Personal Workspace

Once a user registers as a valid BioPlace user, Personal Workspace is allotted to the user in BioPlace. Thus, users can directly upload their own material or data and manage it using a terminal computer. Since the uploaded file can always be downloaded to a terminal computer, Personal Workspace is useful for backup. Also, users can link to research-related sites cited in this Workspace and also use memo function.

### Team Workspace

A member, who belongs to a team, can move his data from his own Personal Workspace to one of his Team Workspaces, or upload new data by himself from his local computer. A user constructs a new collaborative team that



**Fig. 2.** Close relationship among the workspaces.

is required to be sufficient for research aims. It is possible for each team to include a variety of members depending on their own purpose. The person who constructs a team can be a manager and invite members from users of BioPlace to join the team. Team Workspace functions as an environment for discussion within the team and supports a function to monitor events happening at Workspace.

Figure 2 contains a diagram to illustrate the relationship between Personal Workspace and Team Workspace. Members within a team are connected with each other to work for their specific purpose. In Team Workspace, users can share various objects information that can be used in Workspace. For example, the objects are folder, file, memo, or URL. Upon requirement, these objects information can be transferred among different Team Workspace members and can be either downloaded to or uploaded from a local computer.

### Research Road Map

Genome research is generally carried out on the basis of tools to analyze high throughput data and informational materials of a specific research area. Although many tools are suggested, they are difficult for researchers in general to manipulate. Research Road Map is built in BioPlace to solve this problem. When a research group wishes to perform a genome research, the group members can follow the ways provided in Research Road Map. Research Road Map is mainly a stepwise procedure containing storage place and linked sites. Since each step in the procedure is hyperlinked to a sub procedure, a user can easily follow his research step by simply selecting one. As shown in Fig. 3, it is designed for beginners and experienced scientists to work together.

### Work Scheduler

Work Scheduler is a system to deliver research schedules to team members. Classified schedule items, such as date

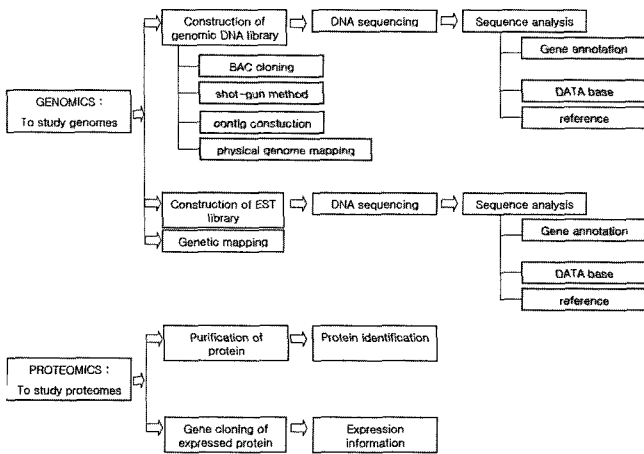


Fig. 3. Diagram outlining the steps used for genome Research Road Map.

of online meeting, time of data exchange, experiment plan, and so on, can be provided to each member. While users can change their own schedules, team schedules are managed only by a team manager when needed. Since Work Scheduler has a list of certain team members and can automatically distribute the team schedule to each member, the user gets an accumulated schedule suggested by different teams in Personal Workspace, instead of Team Workspace. It suggests that a researcher can efficiently organize his task, although he may belong to several teams. For example, if a researcher is simultaneously involved in an EST sequencer team and a genome sequencer team, two different team schedules are automatically incorporated into his schedule in his Personal Workspace.

**K-BLAST, K-PDB, and K-Primer**

BLAST algorithm is one of the most well known and commonly used tools for searching sequence databases. From a group of specified databases, BLAST finds subsequences that are similar to subsequences in the nucleotide and protein query sequence. This BLAST program is localized in BioPlace after transforming to a Korean version, named K-BLAST. K-BLAST has a Korean user interface and a handy brief manual.

K-PDB [5] and K-Primer3 were developed from PDB and Primer3 by converting to Korean versions of the associated user interfaces. PDB is the single international repository for public data on the 3-dimensional structures of biological macromolecules. The primary goals of PDB are to enable users to locate structures of interest and obtain information on these structures. PDB has often been used in the post-genome era. Primer 3 is a program to design PCR (polymerase chain reaction) or nucleic acid sequencing primers. A primer design program is necessary for convenient usage, because thousands of primers have to be synthesized to complete a genome project.

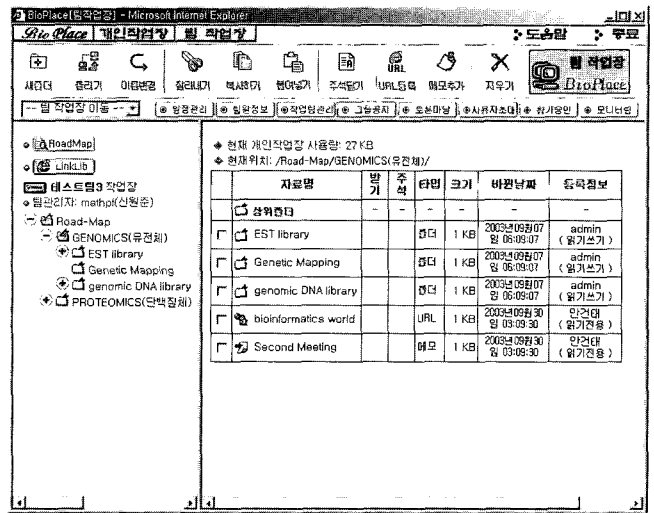


Fig. 4. An example of Workspace user interface.

**System Implementation Results**

BioPlace is the collaborative environment to support tools and programs for a genome research group. The first step in developing BioPlace was to design and implement workspace for both individual and collaborative work. Figure 4 shows an example of the user interface of the Workspace. A public database, PostgreSQL (<http://www.postgresql.org>), was used for constructing the server for workspaces. It was implemented by Java technology in order to independently operate from a computing environment. When a user clicks the item in Research Road Map, he can have a brief explanation, data storage place, and link sites for experimental procedures. In association with the item, it can be connected to his Personal Workspace or Team Workspace.

The second step was the implementation of Work Scheduler for organizing an experiment and discussion schedule.

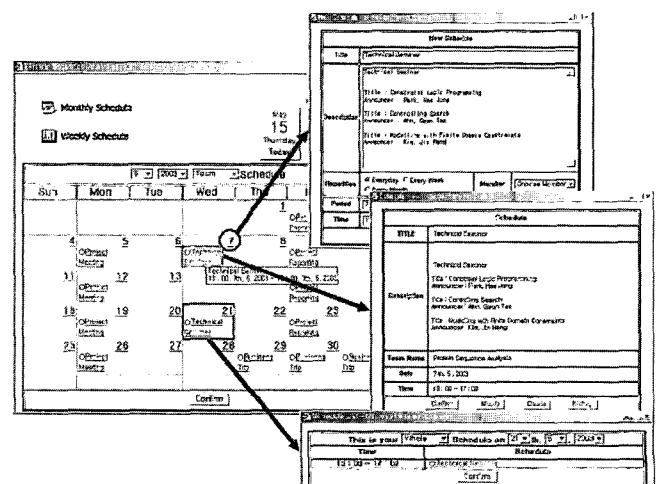


Fig. 5. User interfaces used for schedule management in Work Scheduler.

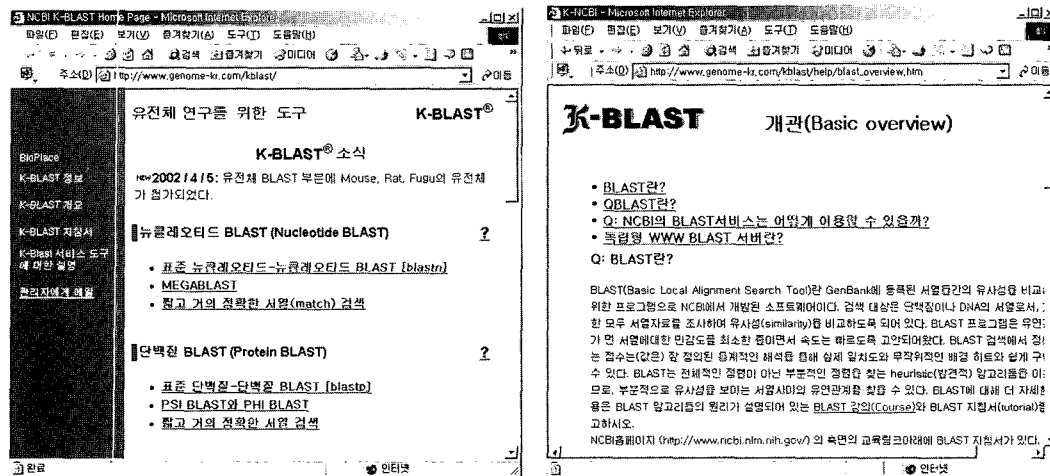


Fig. 6. Examples of user interfaces in K-BLAST.

Work Scheduler has functions of “add,” “delete,” “modify,” and “browser.” When a user clicks one of the links in his Personal Workspace or Team Workspaces, the system displays Work Scheduler interfaces (Fig. 5), and the users can organize his schedule on it.

The third step was the construction of K-BLAST, K-PDB, and K-Primer3 that have a user-friendly-designed Korean interface (Fig. 6). Although BLAST, which is serviced by NCBI, is the most important program in genome research, it is impossible for local users to use it located in the NCBI server, if a network is partitioned, and also it is difficult for Korean beginners to use it. Therefore, K-BLAST which provides a localized version of BLAST to Korean user interfaces was performed in this study, and it was then built in BioPlace. In addition, the Beginner’s Manual of K-BLAST, which is not available in BLAST of NCBI, was included.

All these subsystems are integrated into BioPlace, which is a web-based collaborative environment for genome research.

**DISCUSSION**

With the advances of several whole-genome sequencing projects, including human, mouse, *Drosophila*, *Arabidopsis*, and so on, more genome researches have increasingly been followed or are in planning. The completed and ongoing projects have revealed a great deal of information on how genomes are organized, including a number of unexpected discoveries that have been taken molecular biologists by surprise. These studies have successfully been achieved in various organisms by Korean research groups as well [6–9]. The genome research requires not only tools to analyze sequences and find new discoveries,

but also a common work place for communicating within a research group, which may include a molecular biologist, a computer scientist, and an operator of analytic instruments.

The BioPlace has been developed as a collaborative environment for genome research. Because BioPlace is a web-based program, it is convenient for research members who are not working in the same place. By uploading data to Team Workspace, all members of the team can share the data instead of individual contact. The uploaded data can be downloaded to any member in the same group and uploaded back after modification or transforming to another format. Such function of BioPlace enhances the speed of research and decreases the chances of data to be lost in the process of distributing to members.

Since most of the frequently used tools for genome research are integrated in BioPlace, it is not necessary for researchers to access to each web site for using genome research-related tools. Korean user interfaces of these tools allow both beginners and experienced scientists to easily use them. Additional tools can be linked by a simple click in Research Road Map.

We made Research Road Map that includes a general scheme of genome research. Therefore, unless a special research scheme is required, users can follow the stage suggested in Research Road Map. Since each stage provides a data storage place, users can share their data on the stage that they are working on. If he or she is a team chief, he or she can trace the stage that each member is working on and then manage the team project in a proper way. In virtue of Research Road Map, BioPlace can be used as an educational material in class. Students can learn the scheme of a genome project and gain knowledge about genome research tools and methods.

## Acknowledgments

This study was supported by the Basic Information and Communication Research Grant from the Ministry of Information and Communication and the 2003 Research Fund of the University of Ulsan.

## REFERENCES

1. Beca, L., G. C. Fox, and M. Podgorny. 1999. Component architecture for building web-based synchronous collaboration systems. *Proceedings of IEEE 8<sup>th</sup> International Workshops*.
2. Bentley, R., W. Appelt, U. Busbach, E. Hinrichs, D. Kerr, S. Sikkil, J. Trevor, and G. Woetzel. 1997. Basic support for cooperative work on the world Wide Web. *J. Human-Computer Studies* **46**: 827–846.
3. Berners, L. T., R. Cailliau, A. Luotnen, N. H. Frystyc, and A. Secret. 1994. The World Wide Web. *Communications of the ACM* **37**: 76–82.
4. Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. 1996. Life with 6000 genes. *Science* **274**: 546, 563–567.
5. Helen, M. B. and E. B. Philp. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28**: 235–242.
6. Jhon, D. Y. and S. H. Lee. 2003. Complete DNA sequence and analysis of a cryptic plasmid isolated from *Lactobacillus bifementans* in Kimchi. *J. Microbiol. Biotechnol.* **13**: 1018–1020.
7. Kim, J. S., S. W. Kwon, F. Jordan, and J. C. Ryu. 2003. Analysis of bacterial community structure in bulk soil, rhizosphere soil, and root samples of hot pepper plants using FAME and 16S rDNA clone libraries. *J. Microbiol. Biotechnol.* **13**: 236–242.
8. Kim, S. H., K. B. Lee, J. S. Lee, and Y. H. Cho. 2003. Genome diversification by phage-derived genomic islands in *Pseudomonas aeruginosa*. *J. Microbiol. Biotechnol.* **13**: 783–788.
9. Kim, Y. T., B. K. Park, E. I. Hwang, N. H. Yim, N. R. Kim, T. H. Kang, S. H. Lee, and S. U. Kim. 2004. Investigation of possible gene transfer to soil microorganisms for environmental risk assessment of genetically modified organisms. *J. Microbiol. Biotechnol.* **14**: 498–503.