A Sample Tracking Tool for Rice Genome Sequencing

Katsumi Sakata [1] **Kazunori Waki** [1] **Takuji Sasaki** [1]
`sakata@staff.or.jp` `waki@staff.or.jp` `tsasaki@abr.affrc.go.jp`
Michihiko Shimomura [2] **Masaki Kise** [2]
`shimomur@mi.mss.co.jp` `kise@mi.mss.co.jp`

[1] Rice Genome Research Program, NIAR/STAFF, 446-1 Ippaizuka, Kamiyokoba, Tsukuba, Ibaraki, 305-0854, Japan
[2] Mitsubishi Space Software Co., Ltd., 1-17-15 Sengen, Tsukuba, Ibaraki, 305-0047, Japan

1 Introduction

The Rice Genome Research Program (RGP) has entered a new stage this year with the goal of sequencing the entire genome. Consequently, the International Rice Genome Sequencing Project [1] is organized to perform this effort during a short period. Since the genome size of rice amounts to 430 Mb, this will requires an enormous number of samples to be sequenced (approximately 2000 samples per day at a constant rate). In order to keep the quality of experimental data reliable and stabilized, a sample tracking system has been prepared. The major features of this system include: (1) a server-client system, (2) web-based interfaces, and (3) bar-code data management.

Now, a prototype version of the system is completed and will be demonstrated. It is expected that by using this system the work load of research assistants will be reduced and their errors will be prevented. Further, useful information about the sequencing process will also be made available to the researchers.

2 System configuration

Fig. 1 shows the system configuration of the sample tracking tool. The type of client's computer is not limited because a general web-browser is used to display the data.

3 Functions

The following data are displayed and some of them can be edited on the web-based interface.

(1) Work history of sample
(2) Location of sample
(3) Time-line of process (displayed graphically)
(4) Accumulated number of bases being analyzed (displayed graphically)
(5) Quality information of assembling (displayed graphically)

This system is adapted to the assembling process which uses phred-phrap and consed [2, 3, 4] programs. Fig. 2 shows examples of work history and time-line displays. The above functions have been derived by customizing an application software, Sample Data Tracking Manager.

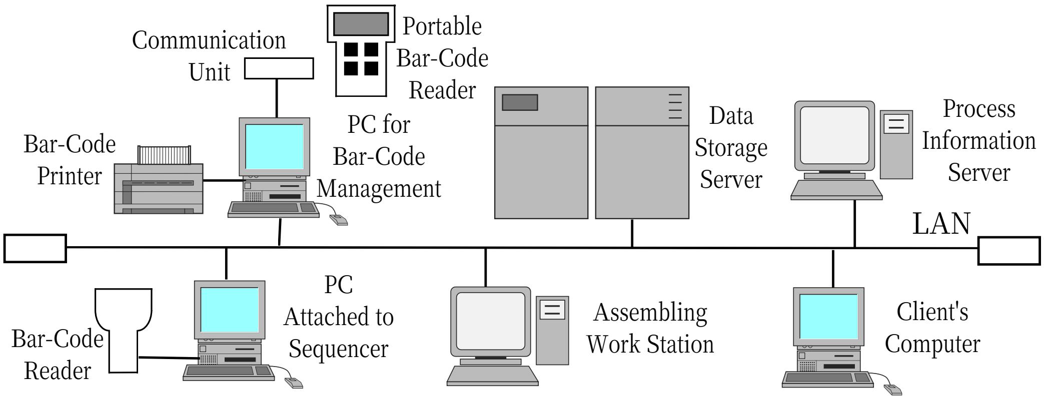
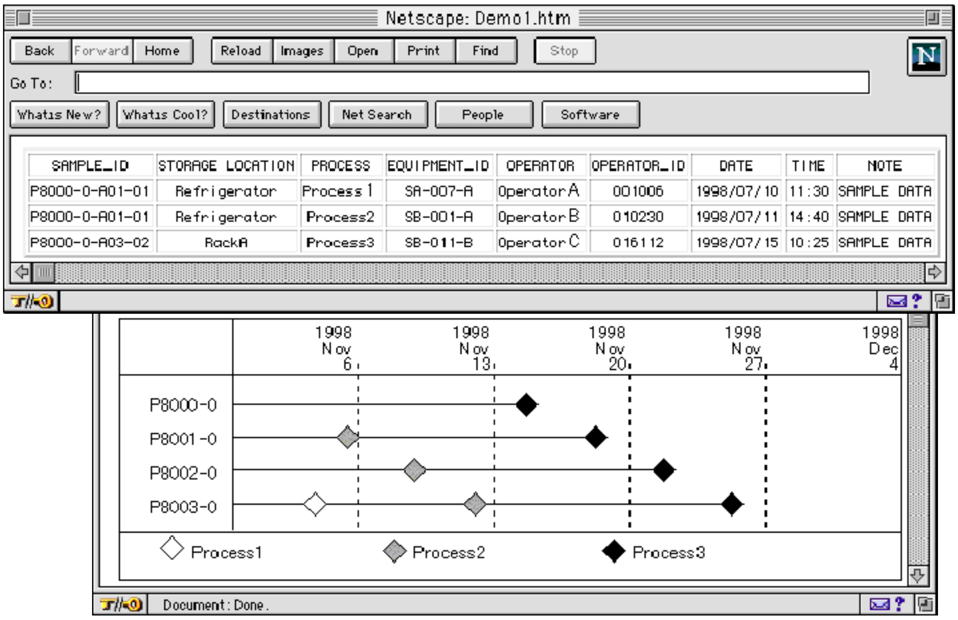


Figure 1: System configuration.



SAMPLE_ID	STORAGE LOCATION	PROCESS	EQUIPMENT_ID	OPERATOR	OPERATOR_ID	DATE	TIME	NOTE
P8000-0-A01-01	Refrigerator	Process1	SA-007-A	OperatorA	001006	1998/07/10	11:30	SAMPLE DATA
P8000-0-A01-01	Refrigerator	Process2	SB-001-A	OperatorB	010230	1998/07/11	14:40	SAMPLE DATA
P8000-0-A03-02	RackA	Process3	SB-011-B	OperatorC	016112	1998/07/15	10:25	SAMPLE DATA

Figure 2: Examples of work history and time-line displays.

References

[1] ftp://genome1.bio.bnl.gov/pub/maize/RiceProject.html

[2] Ewing, B., Hillier, L., Wendl, M., and Green, P., Basecalling of automated sequencer traces using phred. I. Accuracy assessment, *Genome Research*, 8:175–185, 1998.

[3] Ewing, B. and Green, P., Basecalling of automated sequencer traces using phred. II. Error probabilities, *Genome Research*, 8:186–194, 1998.

[4] Gordon, D., Abajian, C., and Green, P., Consed: A graphical tool for sequence finishing, *Genome Research*, 8:195–202, 1998.