

# SAKURA: A New Data Submission System of DDBJ to Meet Users' Needs in the Age of Mass Production of DNA Sequences

Hikaru Yamamoto <sup>3</sup>	Takuro Tamura <sup>3</sup>	Katsumi Isono <sup>2</sup>
hikyamam@genes.nig.ac.jp	tatamura@genes.nig.ac.jp	isono@sbgsp2.biol.kobe-u.ac.jp
Takashi Gojobori <sup>1</sup>	Hideaki Sugawara <sup>1</sup>	Ken Nishikawa <sup>1</sup>
tgojobor@genes.nig.ac.jp	hsugawar@genes.nig.ac.jp	knishika@genes.nig.ac.jp
Naruya Saitou <sup>1</sup>	Tadashi Imanishi <sup>1</sup>	Kaoru Fukami-Kobayashi <sup>1</sup>
nsaitou@genes.nig.ac.jp	timanish@genes.nig.ac.jp	kfukami@genes.nig.ac.jp
Kazuho Ikeo <sup>1</sup>	Yoshio Tateno <sup>1</sup>	
kikeo@genes.nig.ac.jp	ytateno@genes.nig.ac.jp	

<sup>1</sup> National Institute of Genetics, 1111, Yata, Mishima, Shizuoka, Japan 411

<sup>2</sup> Faculty of Science, Kobe University, 2-1, Rokoudai-cho, Nada-ku, Kobe, Hyogo, Japan 657

<sup>3</sup> Hitachi Software Engineering Co., Ltd., 5-79, Onoe-cho, Naka-ku, Yokohama, Kanagawa, Japan 231

The nucleotide sequence database created and maintained by the international collaboration of DDBJ/EMBL/GenBank is growing rapidly: the doubling time of the database has been steadily shortened and it is now only about a year. This fast growth of the database is largely attributable to the systematic genome projects, but the significant contribution of an increasing number of individual researchers around the world should not be underestimated. To make the data submission of the latter category easier and simpler, therefore, it is important to develop an efficient logistics system which facilitates faster data flow from submitters to the user community.

For this reason, we have developed a new data submission system based on the World Wide Web (WWW) and termed it SAKURA. The system has distinct advantages over the previous data submission by e-mail or floppy disk and the data acquisition process has become more reliable and efficient than before as described below.

### (1) Expandability

SAKURA is made flexible to accommodate changes in the data specifications. For this purpose, the parser, page design, format, menu and language information are stored in separate resource files independent of the computer programs that constitute SAKURA.

### (2) Languages

All data entries must be prepared in English, but users can choose other languages for communication with the server during their submission sessions. At present, only Japanese or English are options. However, since SAKURA is capable of supporting multiple languages if appropriate language resource files are prepared, we plan to support, for example, Korean and Chinese in the near future.

### (3) Pause and resume

SAKURA provides submitters with an option to 'pause and resume' his/her data submission session: the typed-in data will temporarily be held on the server for a period of up to a month, unless the submitter explicitly terminates the session. During this period, the submitter is able to retrieve the data from the server, modify and append them, if necessary. This capability will also enable submitters to copy and paste annotation lines when submitting multiple sets of data.

### (4) Multiple entries

Consecutive accession numbers will be issued if multiple sets of data are received from a single submitter. The process will be automatically performed by the SAKURA system.

### (5) Error check

Errors are more rigorously checked by Sakura than floppy disk or E-mail submissions. SAKURA classifies three types of errors, mandatory, illegal, or semantic, and will issue error messages. A 'mandatory' error message will be issued when any of the items that are absolutely required for data-processing are not supplied. An 'illegal' error message appears when the data includes illegal characters defined in the 'form' resource file. A 'semantic' error includes an erroneous biological description: for example, under the CDS feature identifier, the amino acid sequence that are experimentally determined and provided by the submitter must be identical to the one predicted from the DNA sequence. In addition, Sakura also issues warning messages whenever appropriate.

The first version of SAKURA was made available on the DDBJ WWW server in December 1995. The number of submissions using SAKURA has been steadily increasing since then and more than 70 % of the submissions were made with SAKURA in July this year. This clearly indicates that an easy-to-use interface for data submission is indispensable for both submitters and the database. Currently, we are planning to convert the gateway scripts of the server into JAVA language so as to improve the user interface further.