

A data management system for microbial genome projects

Ki-Bong Kim, Hyeweon Nam, Hwajung Seo and Kiejung Park

Information Technology Institute, SmallSoft Co., Ltd.

Room# 4113, HTC KAIST, Hwaam-Dong 62-1, Yuseong-Gu, Daejeon, 305-348, Korea

Abstract

A lot of microbial genome sequencing projects is being done in many genome centers around the world, since the first genome, *Haemophilus influenzae*, was sequenced in 1995. The deluge of microbial genome sequence data demands new and highly automatic data flow system in order for genome researchers to manage and analyze their own bulky sequence data from low-level to high-level. In such an aspect, we developed the automatic data management system for microbial genome projects, which consists mainly of local database, analysis programs, and user-friendly interface. We designed and implemented the local database for large-scale sequencing projects, which makes systematic and consistent data management and retrieval possible and is tightly coupled with analysis programs and web-based user interface. That is, parsing and storage of the results of analysis programs in local database is possible and user can retrieve the data in any level of data process by means of web-based graphical user interface. Contig assembly, homology search, and ORF prediction, which are essential in genome projects, make analysis programs in our system. All but Contig assembly program are open as public domain. These programs are connected with each other by means of a lot of utility programs. As a result, this system will maximize the efficiency in cost and time in genome research.

CURRICULUM VITAE

KIM, KI-BONG

Senior Researcher/Chief, Department of Software Development,
Information Technology Institute, SmallSoft Co., Ltd.

URL : <http://www.smallsoft.co.kr>

E-mail : kbkim@bioinfo.smallsoft.co.kr

Phone No. : +82-42-864-2524

RESEARCH INTERESTS

Gene Structure Prediction (mathematical and statistical methods to identify functional elements in prokaryotic and eukaryotic genomes, especially the genes and their control and regulatory elements) Artificial Intelligence

EDUCATION

B.S., Microbiology, February 1992,
Kyungpook National University, Daegu, Korea

M.S., Computational Biology (Dep. of Microbiology), February 1997,
Kyungpook National University, Daegu, Korea

Entered Ph.D. course, Computer Engineering, March 1998,
AI Lab., Chungnam National Univ., Daejeon, Korea

PROFESSIONAL EXPERIENCE

May 1999 ~ Present	Senior Researcher/Chief, Department of Software Development, Information Technology Institute, SmallSoft Co., Ltd.
1994 ~ Feb. 1999	Scientific Researcher, Bioinformatics Team, Genome Research Center Korea Research Institute of Bioscience and Biotechnology(KRIBB)

RESEARCH ACTIVITIES AND PUBLICATIONS

Thesis

- M.S. "Development of the program to predict protein coding region in Human DNA sequences" (1997)

Program Registrations

- Program to predict the protein coding regions in Human DNA sequences
(Reg. No. 97-01-12-1456)
- Conversion Program to manage GenBank/PIR/SwissProt
(Reg. No. 97-01-12-6286)
- Development of an X-Window browser for GenBank/PIR search and BLAST homology search (Reg. No. 97-01-12-6286)

Projects involved

- Development of the Automatic System for the Functional Analysis of Massive Genomic Sequences Using the Computational Methods
(P. I. ; 1999.8 – 2000. 7)
- Development of analysis system for genomic information ('97. 1 – '98. 12)
- Construction of protein database and development of WWW service system ('97. 7. – '98. 6.)
- Construction of genome research system ('97. 1. – '97. 12)
- Development of integrated processing system for genomic sequences ('95. 1. – '96. 12)

Conference presentations

- K. B. Kim, C. K. Hur, H. W. Nam, K. J. Park, and S. M. Kim, Development of an X-window browser for GenBank/PIR/ database retrieval and BLAST homology search, 1997, The Korean Society for Molecular Biology
- K. B. Kim, E. B. Kong, S. M. Kim, Development of an ORF prediction program from DNA Sequences, 1998, The Korean Society for Molecular Biology