Data Base Design of DNA Base Sequences

Sanzo MIYAZAWA

DNA base sequencing have become so easy that DNA base sequences analyzed are rapidly growing. This recent trend prompts biologists to use computer to analyze DNA base sequences. Typical examples of common analyses are those such as homology search, restriction enzyme site map, and open reading frame search. Search and retrieval of specific base sequences from DNA sequence data are prerequisite to such analyses. Thus, a DNA data base is demanded to efficiently manipulate sequence data. The unix operating system has been chosen as a target system to develope a data base, because of a portable operating system. By taking advantage of superior tools for interprocess communication in unix such as shared memory and pipeline, basic tools for data manupulation are being implemented as filters at the shell level.

Basic tools are provided such as

– to output specified entries,
– to output specified types of records in specified entries,
– to output the name of entries with strings specified in regular expression,
– to output base sequences specified in regular expression with their entry names by searching a whole data base, and also
– to convert formats.

These tools are used with unix tools such as "sort", "uniq" and so on to search and retrieve entries by keywords such as author names, the titles of papers, species names, locus names, and others. In addtion, a wide variety of application programs for sequence analyses will be incorporated into the data base system.

This type of simple search and retrieval system has an advantage to be portable among unix systems whether computer is a large, mini or personal computer. Although portability is lost, an alternative method to build data bases is to use one of data base management systems that are commercially available. Relational data base is advantageous, because it is easy to build and also it may allow detailed search and retrieval. We are planning to build a relational data base to manage data entry in the DNA Data Bank of Japan as well as for its use as a research tool.

PUBLICATIONS

AKETAGAWA, J., MIYATA, T., OHTSUBO, S., NAKAMURA, T., MORITA, T., HAYASHIDA, H., MIYATA, TAKASHI, and IWANAGA, S.: Primary structure of limulus anticoagulant anti-lipopolysaccharide factor. J. Biol. Chem., **261**: 7357–7365, 1986.

AKETAGAWA, J., MIYATA, T., OHTSUBO, S., NAKAMURA, T., MORITA, T., HAYASHIDA, H., MIYATA, TAKASHI, IWANAGA, S., TAKAO, T. and SHIMONISHI, Y.: Primary structure of anticoagulant, anti-lipopolysaccharide factor, from the limulus *Tachypleus tridentatus*. J. Biol. Chem. (in press).

AOKI, K.: A stochastic model of gene-culture coevolution suggested by the "culture historical hypothesis" for the evolution of adult lactose absorption in humans. Proc. Natl. Acad. Sci. USA, **83**: 2929–2933, 1986.

AOKI, K.: Stable polymorphic equilibria in a toy model of group selection. Jpn. J. Genet. **61**: 481–490, 1986.

Aota, S. and Ikemura, T.: Diversity in $G+C$ content at the third position of codons in vertebrate genes and its cause. Nucl. Acids Res. **16**: 6345–6355 & 8702, 1986.

BEHRENS, P. Q., NAKASHIMA, H., YOKOTA, E. and RIGGS, A. F.: The structure of hemocyanin II from the horseshoe crab, *Limulus polyphemus*. J. Biol. Chem. **261** (23): 10520–10525, 1986.

CARSTEN, A. L., BENZ, R. D., COMMERFORD, S. L., HUGHES, W., ICHIMASA, Y., IKUSHIMA, T., and TEZUKA, H.: Further studies on the genetic damage to bone marrow and other somatic effects following exposure to low level tritium. NIRS-M-52 (Proc. 2nd Workshop on Tritium Radiobiology and Health Physics), 258–275, 1985.

FUERST, P. A. and MARUYAMA, T.: Fate of genetic variation in captive 300 populations. Zoo Biology **5**: 171–179, 1986.

FUJII, T. and TANO S.: Mutagenic activities of EMS on somatic (M_1) and recessive (M_2) mutations in the soybean test system. Environ. Exp. Botany **26**: 191–195, 1986.

FUKUDA, R, and HATADA, E.: RNA synthesis of influenza A viruses: Temperature-sensitive mutants defective in RNA synthesis. (in Japanese) Uirus (Virus) **36**: 203–219, 1986.

FURUSATO, T., TAKANO, J., JIGAMI, Y., TANAKA, H., and YAMANE, K.: Two tandemly located promoters, artificially constructed, are active in a *Bacillus subtilis* α-amylase secretion vector. J. Biochem., **99**: 1181–1190, 1986.

GLASS, R. E., JONES, S. T. and ISHIHAMA, A.: Genetic studies on the β subunit of *Escherichia coli* RNA polymerase, VII. RNA polymerase is a target for ppGpp. Mol. Gen. Genet. **203**: 265–268, 1986.

GLASS, R. E., JONES, S. T., NENE, V., NOMURA, T., FUJITA, N. and ISHIHAMA, A.: Genetic studies on the β subunit of *Escherichia coli* RNA polymerase, VIII. Localisation of a region involved in promoter selectivity. Mol. Gen. Genet. **203**: 487–491, 1986.

GLASS, R. E., HONDA, A. and ISHIHAMA, A.: Genetic studies on the β subunit of *Escherichia coli* RNA polymerase, IX. The role of the carboxyterminus in enzyme assembly. Mol. Gen. Genet. **203**: 492–495, 1986.