YAYOI: Taxonomy Database System over International Computer Networks

Hajime Kitakami[1] Yoshio Tateno[2] Takashi Gojobori[2]
kitakami@its.hiroshima-cu.ac.jp ytateno@ddbj.nig.ac.jp tgojobor@ddbj.nig.ac.jp

[1] Hiroshima City University
151-5 Ozuka, Numata-Chou, Asa-Minami-Ku, Hiroshima-Shi 731-31, Japan

[2] National Institute of Genetics
1111 Yata, Mishima-Shi, Shizuoka-Ken 411, Japan

Abstract

We newly developed a repair system which is needed to effectively remove inconsistencies in each data bank and mismatches among data banks over international computer networks. This paper describes search functions to be useful for effectively removing both inconsistencies and mismatches from the databases. These functions are implemented in a relational database management system, SYBASE.

1 Introduction

DDBJ is constructing the DNA database with its European and American counterparts through mutual exchange of data over international computer networks[1]. The DNA database of each data bank includes a taxonomy database in a relational format. The taxonomy database is stored in a binary relation using Sybase or Oracle. Our final goal is to integrate taxonomy databases in the distributed computer environments. However, we have two problems to solve before achieving integration. The first one is that each database is inconsistent within itself. These inconsistencies include invalid pointers and names as well as redundant data. The inclusion of such inconsistencies will result in confusion with unification. The second problem is that the databases are inconsistent with each other. This means mismatches between any two databases which can be detected as different lineages between the same nodes.

2 Search Functions

Recursive join is a useful mechanism for analyzing consistencies in the taxonomy database. The taxonomy is logically represented by a tree structure and the tree structure is stored through binary relations with both child and parent columns. The search for a path from a given node to the root node is achieved by a recursive join. We implemented three local search functions[2] for analyzing the causes of inconsistencies. The first one is a lineage search function for a path from a given node to the root node in the tree structure. The second one is a posterity search function for all paths from a given node to its leaf nodes. However, the number of nodes found by the processing is so large that the system can not be visualized in any single window system at a single time. Thus we created this function to show only a given node and its child nodes. The third one is a homology search function for nodes on the same level as a given node. Moreover, we implemented a global search function to display the whole structure of any subtree. It includes a depth-first search mechanism[3] . It is useful for checking whether revisions are successful in each other.

[1]北上始：広島市立大学情報科学部 〒731-31 広島市安左南区沼田町大塚 151-5
[2]舘野義男、五條堀孝：国立遺伝学研究所 〒411 静岡県三島市谷田 1111

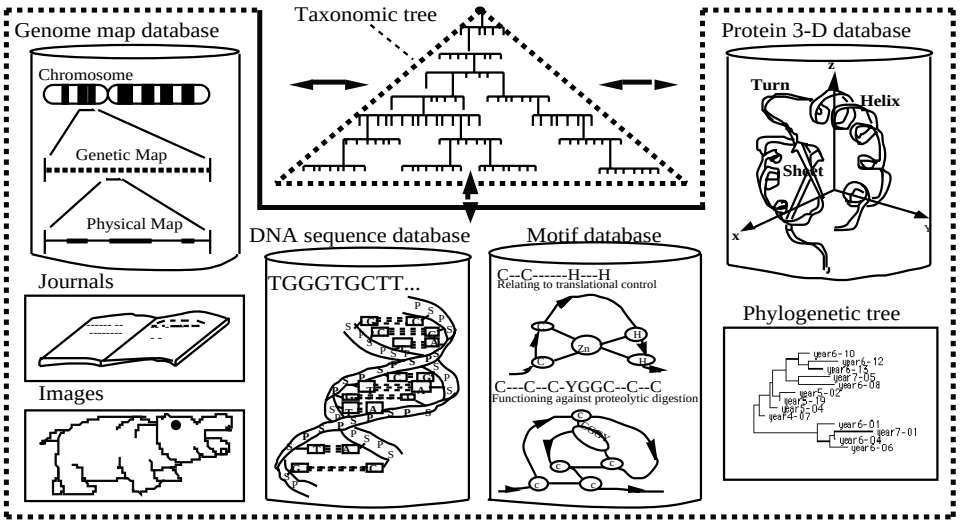

Figure 1: Taxonomic Tree and Other Databases

3 Relationships between Taxonomy Database and Other Databases

A unified database is constructed by integrating the existing taxonomy databases. If we can achieve the construction of such a database, we can obtain the largest taxonomy database among the international DNA data banks. Such a database would be useful in comparing many research results, and investigating future research directions from existing research results. Moreover, it would be useful in connecting the taxonomy and other databases in the field of Genome research including such databases as Genome map, journal, image, DNA sequence, motif, protein 3-D, and phylogenetic tree databases. In particular, a unified taxonomy database would be useful in comparing relationships between phylogenetic trees inferred from molecular data and taxonomic trees constructed from morphological data.

4 Conclusions

To summarize, we developed a new system, YAYOI, with the intelligent interface for effectively repairing the DDBJ-taxonomy database. We introduced two types of error detecting procedure, local search and global search functions using recursive join. YAYOI has been used to repair the taxonomic tree since early autumn 1993. DDBJ-staff have repaired 30 percent of about 4,000 errors found by the two types of error detectingprocedures.

References

[1] Hajime Kitakami, Yukiko Yamazaki, Kazuho Ikeo, Yoshihiro Ugawa, Tadasu Shini, Naruya Saitou, Takashi Gojobori, and Yoshio Tateno "Building and Search System for a Large-scale DNA Database" *Digest of One-Day Colloquim "Molecular Bioinformatics", The Institution of Electrical Engineers (IEE) Press, London*, pp. 61-69, 1994/2.

[2] Hajime Kitakami, Yoshio Tateno and Takashi Gojobori "Toward Unification of Taxonomy Databases in a Distributed Computer Environment" *ISMB-94 at Stanford University, AAAI Press*, pp. 227-235, 1994/8.

[3] Hajime Kitakami, Yasuma Mori, Arikawa Masatoshi, Yoshio Tateno and Takashi Gojobori "Recursive Query Processings for Taxonomy Databases" *Technical Report of IEICE (DE94-60), The Institute of Electronics Information and Communication engineers*, pp.39-46, 1994/9 (in Japanese).