

DNA, 蛋白質配列フラットデータベースのためのデータ作成、
検索システムと電子郵便ネットワークデータベースサーバー¹⁾²⁾

宮澤三造
国立遺伝学研究所
遺伝情報研究センター
遺伝情報分析研究室

要 旨

現在 DNA Data Bank of Japan (DDBJ) で使用している DNA 及び蛋白質配列フラットデータベースのための 1) データベース作成システム (入力、更新) 2) 検索システム 及び 3) 電子郵便ネットワークデータベースサーバーを報告する。作成システムではその容易性から UNIX システムに付属する SCCS (Source Code Control System) を用い履歴管理及び排他管理 (更新に関する排他性) を行っている。検索システムは簡単な機能を果たす様々なツールからなる。基本ツールの例は、

- 指定されたタイプのレコードを出力する。
- 指定された文字列を含むエントリーのエントリー名を出力する。
- エントリー名からなるセットに関する and, or, xor。
- 指定されたエントリーを出力する。

等である。ほとんどのツールは UNIX におけるフィルターとして働く。このようなツールを UNIX のパイプで組み合わせることにより、著者名、論文名、生物種、遺伝子名、キーワード等による検索が可能である。文字列は通常 UNIX の正規表現で指定する。よってあいまいな文字列による検索が可能である。また特異な塩基配列をもつ遺伝子の検索においても塩基配列を正規表現で表現できる。この検索システムのコマンドの一部は電子郵便によっても利用可能となっている。現在 DDBJ は EMBL, GenBank と毎日入力したデータを電子郵便で交換し DNA データベースを更新しているので、Junet, Internet, Bitnet のサイトではこの電子郵便ネットワークデータベースサーバーを使用し最新のデータを入手することが可能である。この検索システムがサポートするフォーマットは現在 GenBank, EMBL フォーマット及び SwissProt, PIR, PRF フォーマット等である。このソフトウェアは、UNIX システムなら移植可能であると言う利点を持つ。

文献

- 1) Miyazawa, S.: The Interface between Computational Science and Nucleic Acid Sequencing, Santa Fe Institute Studies in the Sciences of Complexity, Eds. G. Bell and T. Marr (Reading, MA: Addison-Wesley), vol. VIII, 1989
- 2) Miyazawa, S.: The manual of the FLAT Database and Sequence analysis system for DNA and Proteins, version 1.1, (Mishima, DDBJ), 1989