

# バイオ研究者向け支援サービス

独立行政法人理化学研究所 情報基盤センター横浜チーム

寺井 優晃

## 1. はじめに

バイオ分野では、核酸やタンパク質の配列を、既知配列であるデータベース中から検索することによって、遺伝子予測やタンパク質の立体構造予測等が行われている。情報基盤センターでは、このような遺伝子配列解析の急増を背景にバイオ研究者向け支援サービスを展開している。ここでは、バイオ公共 DB ミラーリングサーバ、RIKEN Super Combined Cluster, Bio Portal を支援サービスの柱としてインフラ整備を行っており、その概要について紹介する。

## 2. バイオ研究者向けサービスの概要

### - バイオ公共 DB ミラーリングサーバ (公共 DB)

バイオ分野の研究では、NCBI, EMBL-EBI, PDB 等で公開されているデータベースを利用する。しかし、データベースサーバの管理状況や回線状況は各公開元に依存しており、安定したデータベースへのアクセス環境が望まれていた。そのため、研究者の利便性向上を目的として、世界中で公開されているデータベースを理研内に設置されたサーバにミラーリングするサービスを2003年8月より行っている(図1)。

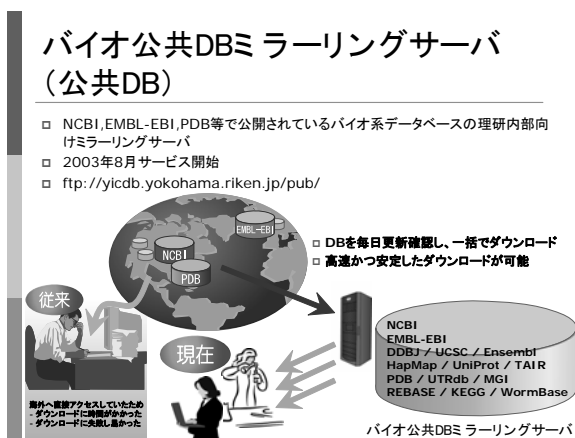


図1. バイオ公共ミラーリングサーバ

ミラーリングするデータベースは図2に示されるようなDBをユーザの要望により順次追加しており、データベース自身の増加と合わせると、指数的な容量増加を示している(図3)。そのため、ディスク追加を必要に応じて実施している。なお、次期リプレース予定は、2010年9月である。

## 公共DB - 公開DB一覧

2008/1/7現在

- バイオ系データベース
    - 配列データベース
    - 遺伝子発現データベース
    - 代謝経路データベース
    - 化合物データベース
    - 相互作用データベース
    - 文献情報データベース
  - ミラーリングされている代表的なDB
- 無償で公開されている代表的なものはほぼ網羅

DB名	データ容量	DB名	データ容量
BLASTDB(NCBI)	約118GB	genomes(NCBI)	約225GB
DDBJ Release	約139GB	GEO	約2376GB
DDBJ New	約14GB	GO	約2GB
EMBL Release	約100GB	goldenPath	約932GB
EMBL New	約8GB	KEGG	約159GB
Ensembl	約1029GB	PDB	約133GB
GenBank Release	約153GB	PFAM	約93GB
GenBank New	約3GB		

図2. 公開DB一覧

## 公共DB - ストレージ利用状況

- ミラーリング状況
  - http://yicdb.yokohama.riken.jp/public/status/
- DB容量の指数的な増加 ⇒ 空き容量の圧迫
- ディスクの追加を必要に応じて実施

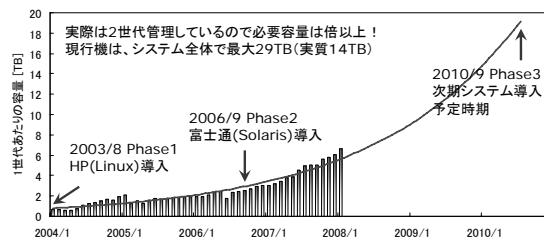


図3. ストレージ利用状況

### - RIKEN Super Combined Cluster (RSCC)

ここでは、バイオ系ユーザ向けの内容にのみ特記する。バイオ系アプリケーションのほとんどは並列化さ

