

“CICSJ Bulletin, Vol. 13, No. 1, 1995”  
Published Bimonthly by  
Division of Chemical Information and Computer Sciences,  
The Chemical Society of Japan  
原稿 01/08/95

# 日本ゲノムデータベースネットワーク (GDBnetJP)

群馬大学工学部共通講座 宮澤三造  
smiyazaw@smlab.eg.gunma-u.ac.jp

生物関連科学におけるインフラストラクチャとしてのインターネットの役割について述べ、ゲノム研究の国際的な研究協力や学際的な研究協力の発展に資することを目的としたネットワーク機関である日本ゲノムデータベースネットワーク (GDBnetJP) について紹介する。

## 1 生物関連科学におけるインターネットの役割

TCP/IP プロトコルを利用した広域ネットワークであるインターネットは、近年米国の情報ハイウエー政策にも触発されインターネットの参加機関およびその利用は急速に増加し、また広域ネットワークにおける情報検索ツールの開発が引金となり、いまや全世界のほとんどの研究者にとって研究に欠かせないツールとなったばかりでなく、一般社会のなかでも確固とした地位を築きつつある。遺伝子解析に関連した分子遺伝学、生物情報学、蛋白質工学を始めとする生物、化学、医学の諸分野もその例外ではない。解析技術の進歩に伴う DNA/蛋白質の配列/構造データ等の指数関数的な増加、また最近では人間の全塩基配列までも解析しようとするゲノム解析計画が米国、欧州をはじめ日本でも発足したことによる遺伝子地図データを始めとするゲノム関連データの急増により、研究に欠かせないデータ、ソフトウェアを up-to-date に入手するためにインターネットは必須なものになった。一方データの入手ばかりでなく、配列解析を始めとするデータ解析、配列/構造データベースやゲノムデータベースへのアクセスにおいてもインターネットはいまや欠かせないインフラストラクチャである。以下では生物関連科学におけるインターネットの利用について簡単に述べる。

### 1.1 データベースアクセス、データ解析のためのサーバー - クライアントシステム

従来データベースアクセス、解析等のための計算機利用は、ホストコンピュータを計算機ネットワークを経て遠隔地よりリモートログインし、ホストコンピュータの上でプログラムを使用することにより実現された。しかしネットワーク関連技術の進歩と共に、サーバー - クライアントコンピューティング環境、つまりクライアントソフトからの要求に基づき一定の処理を行い結果をクライアントソフトに返送するサーバーソフト、結果の表示、入力等ユーザーとのインターフェースを担当しサーバーの処理要求を送信するクライアントソフトが、互いに決められたプロトコルで交信し協調して仕事を遂行するシステム、を実現する手法が整備され、リモートログインすることなくデータベースへのアクセスや配列解析を提供するサーバー - クライアントシステムが開発され運用されている。

### 1.2 広域ネットワーク情報検索ツール

インターネットに接続される計算機が増えるのつれ、また通信回線の帯域も高速になるにつれ、接続される計算機においてネットワークを利用して公開したい情報また公開することを目的に作成された情報が急速に増大してきた。インターネット上で公開される情報の増加とともに利用者がインターネット上で必要な情報を効率良くアクセスするための汎用プロトコルおよびそのプロトコルに基づくツールが開発されつつある。そのような情報検索システムはほとんどがクライアントサーバー概念に基づいて設計されている。汎用広域ネットワーク情報検索プロトコル、ツールとして欠かせないものとして WAIS (Wide Area Information Servers), Gopher, WWW (World Wide Web) がある。この種のツールは汎用であるが故に、世界中に広がったインターネット上に分散している各種リソースを同一の GUI (Graphic User Interface) でアクセス可能という優れた特徴を持つ。また利用が広範囲に渡るため、質の高いソフトウェアが開発されるという面も無視できない利点である。特に NCSA で開発された WWW のクライアント Mosaic ではハイパーテキストという見ために綺麗な GUI で WAIS・Gopher へもアクセス可能なためわずか 1-2 年で世界中で使用されるようになった。また Gopher・WWW サーバーの立ち上げも比較的容易なため世界中のあちこちで Gopher/WWW サーバーが稼働している。このような利点か

ら、配列 / 構造データベース、各種ゲノムデータベース、配列 / 構造解析プログラムも WAIS, Gopher, WWW のプロトコルで世界各地で提供されている。

表 1 構造生物学関連のインターネット上の主なリソース†

URL	備考
<a href="http://www.pdb.bnl.gov">http://www.pdb.bnl.gov</a>	PDB, Gif and RGB images
<a href="http://www.nih.gov/molecular_modeling/mmmhome.html">http://www.nih.gov/molecular_modeling/mmmhome.html</a>	molecular modeling, Molecules R US
<a href="http://scop.mrc-lmb.cam.ac.uk/scop">http://scop.mrc-lmb.cam.ac.uk/scop</a>	structural classification; SCOP
<a href="http://www.gdb.org/Dan/proteins/nrl3d.html">http://www.gdb.org/Dan/proteins/nrl3d.html</a>	protein structure database; NRL3D
<a href="http://www.embl-heidelberg.de/Services/index.html#5">http://www.embl-heidelberg.de/Services/index.html#5</a>	protein sequence-structure analyses
<a href="http://expasy.hcuge.ch/pub/Graphics">http://expasy.hcuge.ch/pub/Graphics</a>	Images of proteins

† 分子生物学関連のリソースについては安永博士の論文を参照願いたい。

## 2 GDBnetJP とは

ゲノム計画の一環として国際共同プロジェクトとなったヒト遺伝子地図、物理地図データからなるゲノムデータベース (GDB) の日本ノードが必要であるとの研究者からの要望を背景に、1991 年頃から科学技術庁のサポートの得、日本科学技術情報センター (JICST) に GDB の日本ノードを引き受けてもらうべく関係者が努力してきた。その過程で JICST が GDB の日本ノードとして機能するにはインフラストラクチャとして米国のノード機関とのネットワーク接続が必須であるとの認識から、インターネット接続のために努力してきた。

当時、学術情報センターのネットワークは大型計算機センターを結ぶ N 1 ネットワークのみであり、実用的なレベルでインターネットへのアクセスを可能にするには TISN (国際理学ネットワーク) か WIDE といったネットワーク機関に加入する以外に方法はなく、大学、研究所で接続されていた機関は限られていた。これらのネットワーク機関に加入するには、ネットワーク研究のためか、もしくは当然ながら国際通信回線費用の負担をせねばならず、研究室レベルでの接続は不可能であった。

当時のこのようなネットワーク環境、および将来予想されるネットワーク需要を考慮すると、国際回線を所有し独自の立場から研究機関、研究者をインターネットに接続する権利を有するネットワーク機関として機能した方が、関連研究者の研究活動をサポートする上で効果があるとの結論にいたり、1993 年 10 月以来日本ゲノムデータベースネットワーク (GDBnetJP) は国際回線を所有しゲノム研究を主目的とした国内・国外の研究機関の LAN 及び地域 WAN を相互に接続するネットワーク機関として、当該学問分野の国際的な研究協力や学際的な研究協力の発展に資することを目的とし活動している。運営母体は日本科学技術情報センターで、ネットワーク委員会が参加機関と共に運営に当たっている。

現在では、TISN、WIDE に加え学術情報センターも SINET と呼ばれるインターネットを構築し多くの大学が接続されている。また地域ネットワークが各地に設立され、ネットワーク接続の手助けをしている。一方、営利を目的とするネットワーク機関も複数設立され、国内のネットワーク接続機関も急速に増加した。更に、米国の NII に触発されたためか多くの省庁がネットワークの構築に関心を持ち活動を始めている。このような状況変化のなかで、関連研究室のネットワーク接続の便を図るという GDBnetJP 設立目的の一つは若干意味を失いつつある。しかし営利機関を含めると未だネットワーク未接続の機関は多く潜在的要望は高い。一方、回線の帯域幅は国内幹線で 512kbps - 6 Mbps 程度で、T3 回線 (45Mbps) を用いている米国とは比較にならない。このような貧弱な状況で接続機関が増加し、一方 WWW の普及で画像データが増えたこと、また近年音声、映画を multicast で送る MBONE が試みられるようになってから、過負荷状態にある回線が多々見受けられるようになった。特に国際回線は 1.5Mbps 程度の回線が数本しか利用できない現状では混雑するいっぽうである。ネットワーク回線におけるこのような現状は我々研究者にとって研究上の障害になりつつある。その意味で、研究のインフラストラクチャとして高速通信回線を提供する GDBnetJP の存在意義は増すばかりである。

## 3 GDBnetJP の活動

以下 GDBnetJP の活動を簡単に紹介する。詳細は <ftp.gdbnet.ad.jp> から入手できる文書を参照してもらいたい。

### 3.1 ネットワーク接続

参加機関は現在 9 機関である。米国カルフォルニア州 Stockton と東京芝浦間に 1.5Mbps の国際回線を引き、参加機関は現時点ではスタートポロジで芝浦にあるルーターに接続されている。他ネットワークとは 学術情報センターが運営している JIX(Japan Internet Exchange) 経由で接続されている。また、CBInet (計算機と化学・生物学の会運営のネットワーク) と芝浦で接続されている。CBInet は大坂まで回線を延長することを計画しているが、その折は CBInet と協力して大坂にアクセスポイントを設ける計画である。経路制御はネットワーク間は BGP、GDBnetJP 内は RIP で行っている。IGP は近い将来に RIP から OSPF に移行する計画である。

Genome Data Base Network in Japan

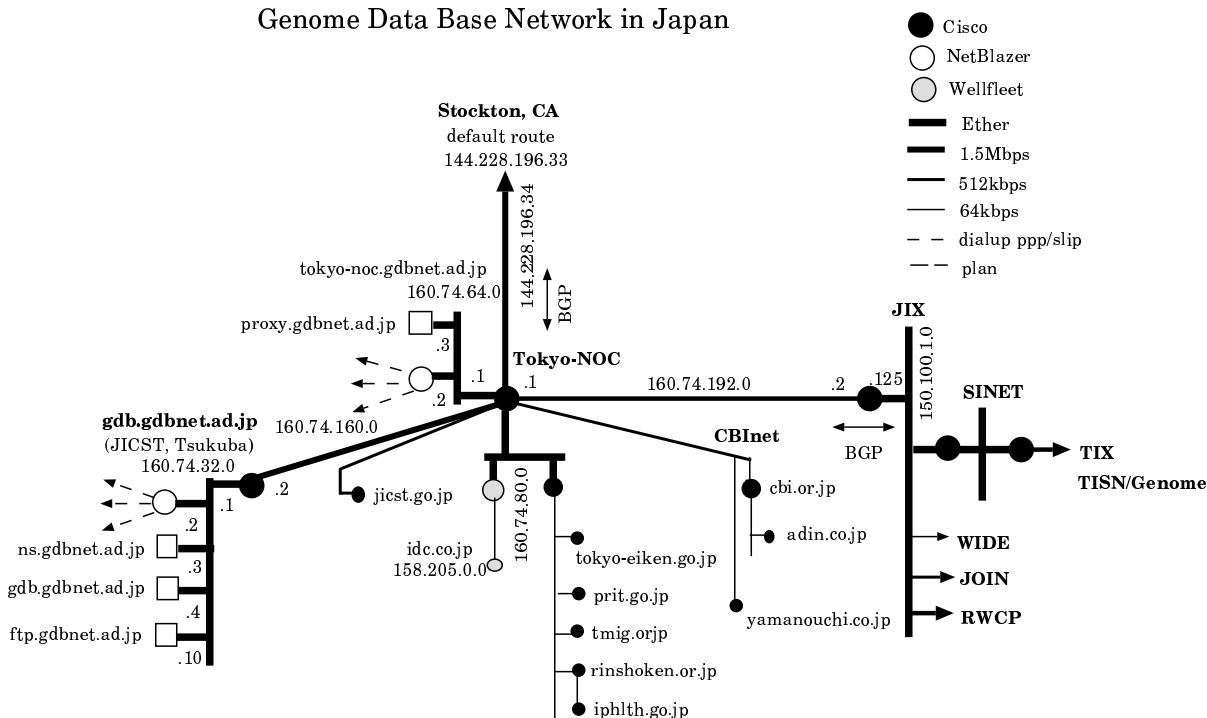


図 1

#### 3.1.1 セキュリティー

個々の組織での セキュリティー への対処の労苦を軽減するため、組織からの依頼があれば GDBnetJP 側 で以下のような代行処置をおこなっている。

- packets の filtering。  
現在、GDBnetJP 組織外からの sunrpc は通過させない設定を default としている。
- firewall の構築。  
firewall の構築のために socks と CERN proxy WWW server を芝浦で稼働させている。CERN WWW sever はまた caching server として運用し回線の効率良い使用に役立っている。

### 3.2 GDBnetJP 国際通信回線の提供

国内のネットワーク利用の拡大に伴う国際通信回線の混雑はゲノム分野の研究者にとっても、研究上の障害になりつつある。しかし GDBnetJP としては、現時点では IP source address による routing が不可能なため、日本国内に広く薄く散らばった研究室、研究者に等しく手を差し伸べることは困難である。そこで、不十分ではあるが、混雑している国際回線に悩まされている研究者に以下のような方法による GDBnetJP の国際回線の利用の機会を提供している。

1. socks サーバーの提供: proxy.gdbnet.ad.jp

ご承知のように socks はセキュリティー上 直接 IP reachable にしたくない組織において外部に接続されている (firewall 上の) 計算機で稼働され直接 IP reachable でない計算機に TCP による 外部アクセスを提供するために開発されたソフトウェアである。しかしソフトウェア自体は TCP packets の単なる relay だから、firewall 上だけでなく、インターネットに接続されているが GDBnetJP には直接接続されていない計算機が GDBnetJP の国際回線を使用して海外をアクセスするための TCP relay として使用することが可能である。インターネットに接続されているどの計算機からもこの socks サーバーをへて GDBnetJP の国際回線を利用することができる。

2. CERN WWW proxy server および DeleGate sever の提供: proxy.gdbnet.ad.jp

WWW の proxy server も HTTP relay として通常は firewall 上で稼働したり、または効率よい回線利用を目的に caching sever として運用するために利用されるが、ここでは海外の GDB 等関連する機関の WWW サーバーの負荷を減らし高速アクセスを可能にする caching sever として、また GDBnetJP の国際回線を使用するための proxy sever として提供されている。

### 3.3 PPP/SLIP dialup access の提供

GDBnetJP は 日本国内に広く薄く散らばった関連の研究室、研究者に IP access を提供する目的で PPP/SLIP dialup の回線 (ISDN および一般公衆回線) を開放している。ただし 現在利用できる回線は東京、筑波の 2 地点のみである。

### 3.4 主要な ftp サーバーのミラー: ftp.gdbnet.ad.jp

ゲノム研究に関連した必須なアイテムは言うまでもなく、ゲノム遺伝子地図、DNA/蛋白質配列/構造等のデータベースである。ここ数年データベースも WWW でアクセス可能となったが、データベースの構築、データ解析にあたって素データが必要なことが多い。現在データベースに納められている DNA は 約 240 k 遺伝子、230 Mbases、ファイル容量で 700Mbytes にのぼる。毎日平均して約 2 Mbytes のデータがリリースされている。また構造データも総量で約 1GB、平均 2 エントリー/日リリースされている。米国、欧州に散在する ftp サーバーの負荷を軽減し国際回線を効率良く利用するために、これら各種データベースを提供している ftp サーバーのうちメジャーなサイトのデータ (総量約 7GB) を ftp.gdbnet.ad.jp でミラーしている。(表参照)

表 2 ftp.gdbnet.jp でミラーを行っている bioinformatics 関連の ftp サーバー

ftp サーバー	備考
ncbi.nlm.nih.gov	GenBank, SwissProt, Acedb, Entrez, BLAST, ...
ftp.ebi.ac.uk	EMBL, SwissProt, ...
ftp.embl-heidelberg.de	Software, ...
ftp.bchs.uh.edu	PIR, ...
ftp.bio.indiana.edu	PHYLP, ...
ftp.pdb.bnl.gov	PDB, Gif and RGB images of proteins
ftp.gdb.org	Genome Data Base, OMIM
ftp.genethon.fr	Human genetic and phisical maps
ceph-genethon-map.genethon.fr	CEPH map
ftp.chlc.org	Cooperative Human Linkage Center

## 4 ドキュメント

GDBnetJP に関係したドキュメントは ftp.gdbnet.ad.jp の ftp/pub/GDBnetJP 以下のディレクトリに置いてある。WWW も www.gdbnet.ad.jp で公開予定である。

## 5 問い合わせ先: gdbnet-query@gdbnet.ad.jp