



GenomeNet

ゲノムネットの紹介

W1

Introduction to the GenomeNet

荻原 淳、内山 郁夫、秋山 泰、金久 實

Atsushi Ogiwara, Ikuro Uchiyama, Yutaka Akiyama, Minoru Kanehisa

京都大学化学研究所

Institute for Chemical Research, Kyoto University

The GenomeNet is a computer network established in September 1991 for promoting academic genome research in Japan. It is supported by the grant-in-aid for scientific research on the priority area 'Genome Informatics' and by the Human Genome Center of the University of Tokyo under the Monbusho (Ministry of Education, Science and Culture) Human Genome Project.

The GenomeNet has been operated in close collaboration with the Todai International Science Network (TISN). It has a major link from University of Tokyo to Kyoto University and to Kyushu University. The link between Kyoto and Kyushu is operated in collaboration with the WIDE 'computer network. Other places such as Osaka University and Japan Foundation of Cancer Research are connected either Kyoto or Tokyo. The GenomeNet is expanding with more sites being added either by leased line connection or by utilization of the ISDN network.

There are two centers in the GenomeNet; one is the Supercomputer Laboratory of the Institute for Chemical Research, Kyoto University, and the other is the Human Genome Center of the University of Tokyo which will be operational in February 1993. Both centers provide database and computational services and coordinate exchange of biological data in laboratories involved in genome research.

In this session, we will show you demonstrations of BLAST/FASTA mail servers to search homologous sequences, DBGET mail server to retrieve sequence databases, anonymous FTP to download databases, and interactive access to Entrez and IRX servers to retrieve database entries. We also explain how to achieve daily updates of sequence databases in collaboration with the NCBI by using the GenomeNet.

[概要]

ゲノムネットは国内のゲノム研究を推進するために、1991年9月から運用を開始されたデータ通信のためのアカデミック運用ネットワークであり、国内のアカデミックネットワークの上では理学研究目的に組織された運用ネットワーク TISN (Toudai International Science Network) に加盟し、ここを通して国内ネットワークはもちろん、U.S. Internet など海外にも到達できるようになっている。ゲノムネットの主な目的は、国内のゲノム研究に関係する諸研究機関を接続してこれら相互及び関連研究センターへのアクセス手段を提供すると共に、ゲノム情報研究拠点となるセンターに関連データベースを整備し維持して行くこと、さらにはこれらのデータベース情報ををゲノムおよ

び生物学関係の一般研究者に様々な形態で提供すること等である。現在、ゲノムネットの整備とデータベースの充実に関する活動は文部省科学研究費重点領域研究「ゲノム解析に伴う大量知識情報処理の研究」および、東京大学医科学研究所ヒトゲノム解析センターの支援のもとに進められている。

[経緯]

ゲノムネットの構築に当たり、運用の約1年前より TISN への加盟準備や接続サイトでの調整などの準備を行い、1991年9月20日に東大理学部(TISN)～京大化学研究所～阪大細胞工学センター間が64KBPS専用回線として開通した。ついで同年11月7日に東大理学部～東大医科学研究所間が64KBPS専用回線で接続された。1992年8月20日には東大理学部～癌研究会リンクが64KBPS専用回線としてスタートし、同年9月30日には東大理学部～京大化研間のバックボーンが3倍の192KBPSに増強されると共に、京大化研から WIDE 京都 NOC (京都高度技術研究所) を経由して九州大学に至る線が192KBPSとして開通した。これらの回線のうち、WIDE 京都 NOC～九州大学間は WIDE プロジェクトとの共同運用になっている。また、同年10月からは ISDN (INS64) を用いた接続の試験を開始しており、現在蛋白質研究奨励会が接続されている。さらに今年度中には岡崎国立共同研究機構が TISN のノードとして64KBPS専用線化されるほか、ISDN 網を利用した接続がさらに拡大される予定である。

[ゲノムネットセンター]

ゲノムネットでは、京大化学研究所スーパーコンピューターラボラトリーと、東大医科学研究所ヒトゲノム解析センターをゲノムネットのセンターとして運用している。京大化研スパコンラボ(略称 SCL)は1992年1月より稼働をはじめ、スーパーコンピューター CRAY Y-MP/2E をメインとした構成で、FASTA 等によるホモロジーの高速検索や、ファイルサーバー SPARC 490 を利用した各種データベースのサポートが行われている他、ゲノムネットの NIC (Network Information Center) が設置され、ゲノムネットの管理も行われている。東大医科研ヒトゲノム解析センター(略称 HGC)は主システムが1993年2月に稼働開始の予定で、2台の SPARC 690 によるデータベースサーバーや HITAC メインフレームによる大容量ファイルサーバーを備え、さらに超並列計算機 CM-5 や nCUBE などが導入される。ここではデータベースサーバー上の Sybase を用いたデータベースサービスなどが予定されている。これらのセンターは以上のようなサービスや計算資源の提供のほかに、ゲノムネットを介して国内のゲノム研究、生物学研究用データベース維持サイトの1つとして、関係するデータベースセンターとの協力の下にデータベースの運用センターとしての役割も担っている。

[利用者サービス]

現在、一般利用者のための無記名サービス(登録がなくても利用できるサービス)として、電子メールによる検索、Anonymous FTP によるデータの取得、オンラインプログラムによるデータベース問い合わせなどのシステムが開放されている。それぞれのアドレスなどは末尾の付録に載せているが、メールサーバーなどは本文に help と書

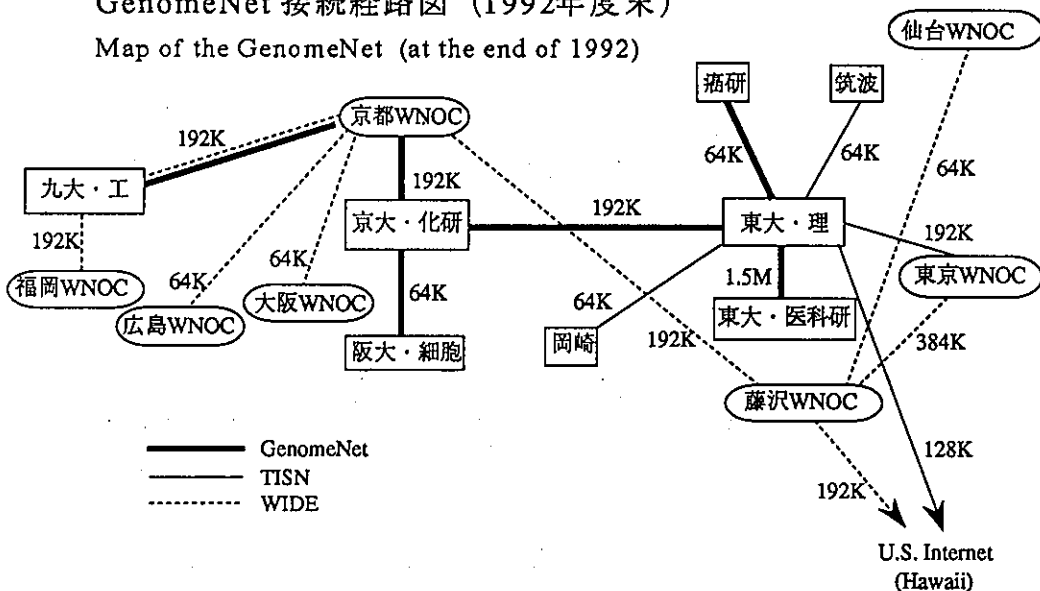
いて送ると使用法のマニュアルが送り返されてくる。FASTA および BLAST は共にホモロジー検索プログラムであるが、一般に BLASTの方がギャップに弱いものの高速である。DBGET はエントリー名またはアクセッション番号で指定されたデータベースエントリーをとりだして返送したり、与えられたキーワードをもつエントリー名リストを返送する。MotifFinder は我々が作成しているタンパク質のスーパーファミリーを特徴付けるモチーフを集めたモチーフ辞書を用いて、与えられたタンパク質配列をサーチし、モチーフが見つかった場合は関連する構造・機能部位を図示する PostScript ファイルと共に返送するものである。FTP では、他サイトで作成されているデータベースなどに対しては、毎週 1~2 回自動的に FTP を発行して作成元の原版と内容をあわせるような機構が取り入れられているため、NCBI repository や GDB 等も常に最新のものが入手可能である。オンラインプログラムでは Entrez (NCBI), IRX (NLM) が telnet を介して利用できる。Entrez ではキーワードをもとに GenBank, PIR, Medline の関連エントリーをつぎつぎに辿ることができ、IRX でも広範なキーワード検索が GenBank, PIR 等に対して行える。途中でセーブした結果は電子メールなどによって自分のマシンに返送させることも可能である。

[アップデート体制]

本年 10 月より GenBank の配付元が NCBI に移管されたのを受けて、ゲノムネットセンターでも NCBI からの日時更新データを毎日受けることで最新のデータベースを維持している。これらは自動化された FTP によって ASN.1 データ交換形式で用意されたファイルを転送し、重複などをチェックした後に最新リリース分として追加更新される。これらはさらに各検索プログラム用に変換されて、メールサーバー等でも利用できるようになっている。

GenomeNet 接続経路図 (1992年度末)

Map of the GenomeNet (at the end of 1992)



Appendix: List of Services on the GenomeNet

E-mail Services

電子メールアドレス：

blast@genome.ad.jp
fasta@genome.ad.jp
dbget@genome.ad.jp
motif@genome.ad.jp

サービス内容：

サーバー	プログラム	データベース
BLAST	blastp, blastn blastx, tblastn	GenBank, EMBL, PIR, Swiss-Prot
FASTA	fasta, tfasta	GenBank, GenBank-subdivisions, EMBL, EMBL-subdivisions, PIR, Swiss-Prot
DBGET	get, find	GenBank, EMBL, PIR, Swiss-Prot, PRF, PDB, PDBSTR
MotifFinder	(search), list, get	MotifDictionary (& PIR, PDB)

Anonymous FTP

FTP アドレス：

ftp.genome.ad.jp	(133.3.49.24)
------------------	---------------

Login: anonymous, Password: 自分のメールアドレス

サービス内容：

ディレクトリー	データベース
/db/local/icr	京大化研作成のデータベース (Amino acid index database, Enzyme reaction database, 他)
/db/local/prf	蛋白質研究奨励会作成のデータベース (Fragment peptide library, Mutation data, 他)
/db/ncbi	NCBI repository に登録されたデータベース (TFD, NGDD, EPD, LiMB, metabolism, REBASE, swiss-prot, prosite, enzyme, seqanalref, journals_toc, Eco, FlyBase, acedb, kabat, aids-db, estdb, CarbBank, blocks, t4phage, ECO2DBASE, pkinases, RLDB, 他)
/db/gdb	GDB が提供しているデータベース、その他 (GDB, OMIM)

Interactive retrieval

ホスト名	login 名	Program	動作に必要な条件
entrez.genome.ad.jp	entrez	Entrez (by NCBI)	X11R4 が動作する端末、 ワークステーション
irx.genome.ad.jp	irx	IRX (by NLM)	フルスクリーン 制御が可能な端末