

8. 国立遺伝学研究所遺伝情報研究センター —DNAデータバンクのための計算機システム—

遺伝研では、1984年4月、国立大学共同利用研究所への移行に伴い遺伝情報研究センターが設置され、日本におけるDNAデータバンク(DNA Data Bank of Japan)を引き受けることになった。その業務はDNA塩基配列データの収集と入力、およびその解析手段の開発・提供である。研究遂行のため1987年3月に計算機が導入された。

遺伝研の計算機システム¹⁾は、特に他と異なる特徴を持ったシステムというわけではない。ただ、オンライン共同利用の計算機ではあるが、データバンク専用の計算機である関係上、幅広い分野の研究者を対象とした大学の計算センターとは異なり、生物科学の研究者が対象であり、その観点からシステムがデザインされた。

1. システムデザイン

計算機選定の際、データバンク業務からの要求事項として特に以下のことを考慮した。

- (1) データベースの構築、維持管理、データ解析プログラムの開発
- ① オペレーティングシステム(OS)としては、会話型OS、プログラム開発向き、かつ科学計算向きOSでなければならない。また限られた人員で管理しなければならないので、少人数で管理可能なOSであること(現在1~2名で運営している)。
- ② DNAデータバンクに関係した応用プログラムが国内外で数多く開発されているので、これらのプログラムの移植が容易なシステムであること。この分野のソフトウェアのほとんどはDEC社製のVAX/VMSで開発されている。
- (2) データベースの提供
データベースを多くの研究者に利用してもらうため、
- ① 公衆回線によるオンライン使用のサポート、特に
 - ・モデムを使用する通常の公衆電話回線
 - ・DDXパケット通信のサポートが要求される。特にパケット通信は、通信量に応じた料金体系であるため遠隔地からの利用者には便利である。また他の計算機と広域ネットワークを構築するためにも必要である。
- ② 多数のTSSオンライン使用、特にパーソナルコンピュータを用いての全2重通信によるオンライン使用が可能なシステムであること。ここで強調したいのは全

2重通信のサポートということである。端末インタフェースとして、全2重が半2重に比べて優れていることはいうまでもない。特にPCに慣れた(生物科学の)研究者にとって、人間に通信の方向切り替えを強いる半2重通信はなじめないであろう。生物科学研究者においては、メインフレームを使用したことがない人がほとんどである。

(3) 日本におけるDNAデータバンクのセンターとして国内外の関係機関とネットワークの構築が要求される。

- ① GenBank(アメリカDNAデータバンク)、EMBL(ヨーロッパDNAデータバンク)と広域ネットワークの構築。
- ② 国内の関係機関、例えば東京大学医科学研究所とのネットワーク構築。

現在、遺伝子のDNA配列決定は加速度的に増加し、各データバンクともデータ作成、入力が間に合わない状況である。現在は発表論文からデータを作成しているため、専門家の手が必要とされ時間がかかる。そこで各データバンク間でのデータ入力の国際分業が検討されている。このような国際分業にあたっては、データバンク間での密接な情報交換が必須である。

われわれは、以上のような観点から計算機およびオペレーティングシステム(OS)を検討した結果、UNIXシステムが最適であると考えた。UNIXシステムの解説²⁾は他に譲るとして、UNIXシステムを選択した理由は以下のとおりである。

- (1) **Portable OS** : UNIXは高水準言語Cで書かれているため移植しやすく、マイクロコンピュータからIBMの大型計算機、スーパーコンピュータGray-2にも移植され利用されている。NEC PC-9800版もある。またソースコードも安く入手できるため教育用としての効果も高い。
- (2) **優れたプロセス間コミュニケーション** : その良い例は、同期的にプログラムの出力を他のプログラムの入力とするパイプと呼ばれるものである。パイプを用いることにより大きなプログラムを個々の仕事に分け、フィルタと呼ばれる小さなプログラムの組み合わせとして構築することが可能となる。これはUNIXの大きな特徴の一つである。
- (3) **豊富なツール** : 代表的なツールはソフトウェア開発管理(sccs)、バージョンアップ管理(make)、言語処理系作成ツール(lex, yaccコンパイラコンパイラ)、英語ドキュメント作成支援プログラム(troff)等である。

(4) ネットワーク機能：4.2 BSDには階層的プロトコルとしてTCP/IPプロトコルが、また公衆電話回線を利用する uucp 機能がある。

(5) 各種のターミナルをサポートできる機能

その他、言語なみの機能を有するコマンドインタプリタ (sh, csh) など多くの優れた特徴を持ち、他の OS に多大の影響を与えた。MS-DOSもその特徴の多くをUNIXから得ている。以上のような点で、UNIXシステムは少なくとも現時点では他より優れていると考えた。また、VAX/VMSで開発されたプログラムも、VMSの拡張機能を利用しているものを除けば、移植もそれほど困難ではないと思われる。VMS、UNIXともに

MULTICSの影響を受けているので類似点も多い。VMSはシステム管理、ジョブ管理に優れ、UNIXはプロセスコミュニケーションの点で勝っていると思う。

UNIXには、研究開発向けのパークレー版4.2 BSDと、ビジネス向けのAT & T版System Vの二種がある。4.2 BSDをサポートし長い間UNIXのホストマシーンであったVAX 計算機は、アメリカはもとより世界中の大学、研究所で広く使われ、VAX用ソフトウェアの国際的な流通性は非常に高い。このことは、国際協力を必要とするこの分野では特に大事なことである。残念ながら遺伝研ではさまざまな事情から日本のメーカーのFACOM M-380Q/UTS(System V)を選定しなくてはならな

Ether net TCP-IP LAN (10 Mb /s)

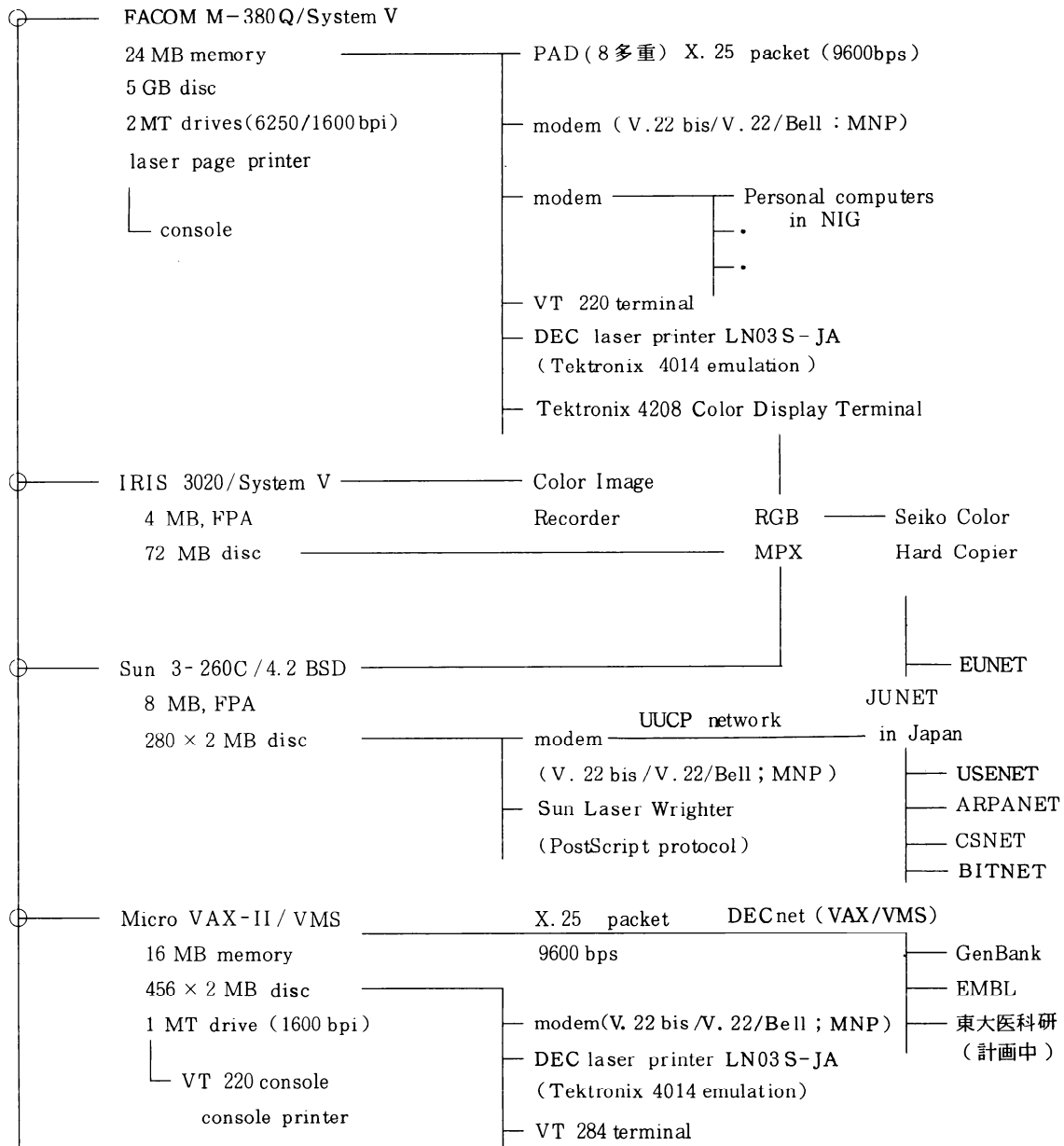


図-1 遺伝研システム構成図

った。蛇足ではあるが、国際化の進む現在このような障害が早く解消されることを切に望む。

2. システム構成

詳細は図-1を参照してもらうこととして、ここではどのような立場からシステムを構成したか簡単に述べたい。システムは小規模ではあるがネットワーク(Ethernet LAN)を構成している。ネットワークを構成している FACOM M-380Q, Sun 3/260 C WS, Micro VAX II/VMS, Iris 3-D Color WS は、バークレー版 UNIXで標準の Ethernet TCP/IP プロトコルを用い、相互に

- Virtual terminal; remote log in
- Remote command execution (between UNIX systems)
- File transfer

が可能である。VAX/VMSでは DARPA プロトコルのみがサポートされる。これらの機能は資源の有効な共有

を図る上で最も重要であり、ホストコンピュータに機器を単に端末として結合するシステムでは不可能な LAN の優れた利点の一つである。

SUN 3/260 C ワークステーションは研究環境向け UNIX 4.2 BSD をサポートし、優れたワークステーション機能(マルチウィンドウ機能ほか)を持つことで有名である。また AT & T 版 System V UNIX もサポートするので、System V である FACOM M-380 Q を補完し、ワークステーションのホストとしての役割を期待している。

国内外の関係機関とのネットワーク構築は、そのほとんどが VAX/VMS を使用しているため、VAX 計算機関のネットワークプロトコルである DEC net を用いるのが最も有効と考え、スーパーマイクロコンピュータ Micro VAX II/VMS を DEC net へのゲートウェイとしてシステムに組み込んだ。一方、これまで世界で開発された DNA 塩基配列解析ソフトのほとんどは VAX/VMS の上で開発されており、Micro VAX II/VMS の導入はそれらのソフトの使用に便を図るというメリッ

表-1 遺伝研で導入する DNA, 蛋白質配列解析, 分子構造解析ソフトウェア

1. DNA, 蛋白質配列解析ソフトウェア:

VAX/VMS で稼働:

- (1) NBRF のデータベース検索プログラム; PSQ, NAQ
- (2) UWGCG Univ. of Wisconsin Genetics Computer Group 検索, 解析プログラム
- (3) IDEAS Integrated Database and Extended Analysis System for Nucleic Acids and Proteins
- (4) Staden 解析プログラム

UNIX システムに移植予定のプログラム:

- (1) PHYLIP Phylogeny Inference Package by Dr. J. Felsenstein in Univ. Washington
- (2) Staden

2. 分子構造解析, シミュレーションプログラム

- (1) CHARMM A Program for Macromolecular Energy, Minimization and Dynamics Calculations from Dr. M. Karplus, Harvard Univ.

3. IRIS 3020 用 3-D 分子構造表示, シミュレーションプログラム

- (1) MIDAS Molecular Interactive Display and Simulation from Dr. R. Langridge, School of Pharmacy, Univ. of California, San Francisco
- (2) FRODO an interactive 3 D molecular modeling and refinement program from Dr. S. Oatley, Univ. of California, San Diego
- (3) GRID a program for detecting energetically favorable binding sites on molecules of known structure from Dr. S. Oatley, Univ. of California, San Diego
- (4) MMS UCSD Molecular Modeling System : a real-time, interactive, 3 D molecular display program from Dr. S. Dempsey, Univ. of California, San Diego
- (5) GRINCH an interactive program for the initial interpretation of protein electron density maps from Dr. M. Carson, Univ. of Alabama
- (6) Amber molecular mechanics / dynamics program from Dr. T. Darden, NIEHS and Dr. P. Kollman, UCSF

トもある(表-1参照)。

広域ネットワークはパケット通信と公衆電話回線により実現される。前者はNTTのDDX-Pに加入しKDDの通信網システムVENUS-Pを経てアメリカ国内(Telenet, TYMNET, ...), ヨーロッパ国内(DATEX-P)の通信パケット網に接続される。よって遺伝研の計算機と国内外の計算機との間で直接のアクセスが可能である。公衆電話回線を用いるネットワークはUNIXのネットワーク機能uucpにより実現される。日本でのuucpネットワークであるJUNETは、アメリカのuucpネットワークUSENET, ヨーロッパのネットワークEUNETにも接続されている。またアメリカの研究者間ネットワークARPANET, CSNET, またBITNETへも接続されている。このネットワークの電子メール機能は研究者間の研究連絡に活用されよう。しかし欧米のDNAデータバンクとの研究連絡は頻繁であり、またヨーロッパの機関は主にBITNETに加入しているため、専用線ネットワークBITNETへの加入も考えている。

IRIS 3-D ワークステーションは3次元分子構造表示用高性能カラーグラフィック装置として分子設計、蛋白質設計のような分野で活用されよう。カラーグラフィック端末としてはEvans & Sutherland PS-300が有名だが、端末という制約がある。IRIS 3020はラスタースキャン型のUNIXワークステーションである。多くの分

子構造表示ソフトウェアがEvans & Sutherland PS-300から移植され、また新たなソフトウェアが開発されつつある(表-1参照)。

3. 問題点

FACOM M-380 Q/UTSは数少ないメインフレームのUNIX専用機である。アメリカのAmdahl社開発のSystem Vで、4.2 BSDでないのが残念である。FACOM拡張の部分は残念ながら信頼性に欠ける。仮想メモリサイズがわずかに16MBであること、全2重通信制御の不備、SMTP機能のノンサポート、troffサポートのプリンタがないこと、日本語roffがないこと等多々問題がある。またファイルの共有を可能にするNFS(Network File System)はサポートしない方針らしい。UNIXソースコードは非公開である。UNIXの思想を理解していないように思える。UNIX流サポートの早い実現を望んでいる。

[宮澤三造]

文 献

- 1) 宮澤三造: DDBJ 計算機システムが決定, DDBJ ニュースレター, 6:20, 1987.
- 2) 石田晴久: UNIX システムの動向, 東京大学大型計算機センターニュース, 17 (No.11): 53, 1985.