

ヒト遺伝子地図データベースシステム *GeneView^{PLUS}*

平川美夏¹, 鈴木政彦¹, 諏訪秀策¹, 神田利彦¹, 灰原聡², 雪竹正治²,
菟島伸生³, 満山進³, 清水信義³

¹日本科学技術情報センター (JICST) 技術開発部

²(株) システムアーキテクチュア

³慶應義塾大学医学部 分子生物学教室

国際的規模で実施されているゲノム解析プロジェクトの重要な目標は、未知の遺伝子を網羅的に探し出し、塩基配列から機能を予測することである。そのためには、未知の遺伝子を含むDNAクローンのゲノム上での距離、相互の位置関係等を明らかにするマップの作成が必須である。プロジェクトも5年を経た今日GDB (Genome Data Base) に膨大なマップ情報が蓄積され、有効な活用に注目が集まりつつある。我々は、GDBのデータの検索表示及び、生の実験データを含めた多様なデータを模式図で簡便に扱うため、ヒトゲノムマッピングデータベースシステム *GeneView^{PLUS}* 作成しているので報告する。

GeneView^{PLUS} - A Human Genome Data Base System

Mika HIRAKAWA¹, Masahiko SUZUKI¹, Shusaku SUWA¹, Toshihiko KANDA¹,
Satoshi HAIBARA², Masaharu YUKITAKE²,
Shinsei MINOSHIMA³, Susumu MITSUYAMA³, Nobuyoshi SHIMIZU³

¹The Japan Information Center of Science and Technology

²SYSTEM ARCHITECTURE, Co., Ltd.

³Keio University School of Medicine

The ultimate goal of human genome project is to find all the genes and to define their function through DNA sequencing. To achieve this goal, it is necessary to create the map that shows the locations of DNA clones containing genes. During the past five years, huge amounts of mapping data have been accumulated in the GDB (Genome Data Base), and the efficient use of this information resource has become a critical matter of the project. *GeneView^{PLUS}* is a such system that can quickly search GDB data and enter raw mapping data with the aid of chromosome graphics familiar to genome researchers.

1. まえがき

ヒト全遺伝子は約10万と推定されているが、現在機能が明らかなものは5000余りにすぎない。ヒトゲノム解析プロジェクトでは、ヒトのゲノムDNAの30億塩基対の配列を決定することを通じて、未知の遺伝子を発見し、その構造的特徴等との関連から機能を明らかにする。しかし、全ゲノムDNA中の遺伝子として機能する部分は、わずか10%とされており、その効率的な解明には、個々の遺伝子やクローンを特定する指標としての「マップ」が必須である。そこで、ヒトゲノム解析プロジェクトの第一段階として、マーカーとなるDNAクローンのゲノム上の位置やクローン相互の位置関係、距離などを明らかにするマッピングが行われてきた。ヒトゲノムDNAは24種類の染色体に分配されており、個々の染色体は大きさ、形、バンドパターンで識別することができ、マッピングにおける位置の規準とされている。

国際的なヒトゲノム解析プロジェクトにおいては、マッピングデータはGDB (Genome Data Base) に文字情報として集約されている。GDBは米国ジョーンズ・ホプキンス大学に本部が置かれているが、国際協力で運営されており、各国に設置されたノードにアクセスして利用することができる。その一方で、日々データを生み出す研究現場での生データを取り扱う有用な計算機システムは少なく、マッピングデータの記録、マップの構築とデザイン、データの比較参照に適したツールの開発が望まれている。

*GeneView^{PLUS}*は、生データの入力及びマップ作成を支援するシステムと、染色体の模式図(イデオグラム)を検索のキー入力及び結果表示に用いるシステムを持つヒトゲノムマッピングデータベースである。物理的地図の入力・表示モジュール *Physical Mapper* は、*GeneView^{PLUS}* から必要に応じて呼び出して利用する場合と、単体のマップドローツールとしての利用を想定して設計している。

2. ヒトゲノムデータベース GDB

2.1 概要¹⁾²⁾

GDB (Genome Data Base) は、世界のヒトゲノム解析プロジェクトから発生したデータを収集

する唯一の国際データベースである。米国ジョーンズホプキンス大学で作成、提供されていたが、現在は世界各国の国際協力で運営されており、世界9カ所にあるノードに置かれた最新のコピー(毎日更新)にアクセスすることもできる。日本科学技術情報センター(JICST)は1994年2月より、公式日本ノードとして国内のゲノム研究者にGDBの提供を行っている³⁾。GDBには、SCW (Single Chromosome Workshop) 等の会議の場で議論を尽くし⁴⁾、HUGO (Human Genome Organization) に認定された染色体エディタのエディティングを施した公的データが登録されており、扱われるデータ項目は100種類にのぼる。GDBは関係データベース Sybase を DBMS (Data Base Management System) としており、ネットワーク又は電話回線を通じてオンラインアクセスし、VT100 端末または X-Window 環境で Sybase のフロントエンドタイプの画面から検索する。ゲノムプロジェクトの進展に伴うデータの増加、計算機の改良、ネットワークの普及等、ヒトゲノム解析情報を取りまく環境の変化を受けて、WWW、Gopher、WAIS、SQL、Mac用検索ソフト(GDB/Accessor)、PC版 CD-ROM (試作) など多様な利用形態が提示されている。

2.2 マッピングデータ

ヒトゲノム解析プロジェクトの目標はヒトゲノム全塩基配列の解析であり、その効率化と有効な解析材料の同定のために、DNAクローンのマッピングが率先して行われ、そのためGDBのデータは急増し種類も多様化している。DNAクローンは主に、FISH (Fluorescence In Situ Hybridization) 法で染色体上にマッピングされる。また、マーカーとの位置関係や、クローン間の重なりやその距離をもとに多数のクローンが整列され、コンテイングマップが作成される。これらのデータは、遺伝子やDNAクローンの染色体上の位置情報の検索表示を行う Locus (座位) マネージャー、マップセットとして順序づけられたクローンの並びを示す Map (地図) マネージャーで管理されている。また、クローンの保存状態、STSのプライマーなどは、Probe (プローブ) マネージャーに保持している。

GDBは、現在のところ文字情報のみのデータ

ベースであり、クローンの染色体上の位置関係を比較したり、クローン間の距離と並び順をグラフィカルに見ることはできない。そのため、マップセットのクローン間の関係は独自のシンタックスを使用してテキスト表現している。*GeneView^{PLUS}*は、GDBのデータ資源を生かし、さらに実験データを含めた多様なデータをモード図で扱うために、全面的にグラフィックスを用いている。特に、*Physical Mapper*では、クローン間の距離や重なりなどを自由な拡大率でリアルに表現することにより、マップセットの独自のシンタックスを視覚的に見ることができる。

3. *GeneView^{PLUS}* のシステム

*GeneView^{PLUS}*はC言語で書かれ、画面表示部分にはMotifライブラリーが用いられている。SunワークステーションのX-Window上で稼働するフルカラーのGUI (Graphical User Interface) を持つシステムである。

基本的な操作は全てマウス指示で行えるよう、実行はメニューとボタンで行い、図中をクリックする等の操作で範囲指定や拡大指定ができる。この操作性はゲノム研究者にとって身近なマッキントッシュをイメージしている。

データベースとしての*GeneView^{PLUS}*は、専門の研究者及び、海外の研究者のために英語表示を基本とするが、医療分野、教育分野での活用を考慮して日本語に切り換えて表示することもできる。さらに、言語データベースを用意さえすれば、他の多数の言語にも対応できるマルチリンガルな構造設計となっている。

4. *GeneView^{PLUS}* のデータ

GDBから公知データとして、遺伝子記号・染色体番号・領域・MIM番号、遺伝形質・分析方法等のマップデータ、および文献のタイトル・著者・ジャーナル・発行年等の文献情報を取り入れている。データは、扱い易さを考慮し、タブで分割したテキストとして蓄積されている。GDBに由来するデータは、フラットファイルから独自に開発したソフトウェアツールを使用して*GeneView^{PLUS}*タイプのデータに変換している。日本語データは、翻訳後に同フォーマット中に導入している。

*Physical Mapper*では、マップサイズ、クローンサイズ、クローン間の重なり、制限酵素切断部位などGDBから、直接得られないデータを扱っているため、GDBからデータを直接取り込むことはない。*GeneView^{PLUS}*を通じて、GDBに由来する物理的地図データを表示することを予定している。

5. *GeneView^{PLUS}* の仕様⁹⁾

5.1 マルチリンガル GUI マッピングデータ検索システム *GeneView^{PLUS}* の特徴

公共データベースとして流通しているヒトゲノムマッピングデータベースは事実上GDBのみであり、ヒトゲノム解析に関するデータの根幹をなしている。しかし、GDBは、テキストベースのデータベースであり、内容を熟知した研究者でなければデータやデータ間の関係を理解し利用することは難しい。最近ではGDBの多様な利用方法が、作成元のジョーンズホプキンス大学からも提案され、WWWを通じて染色体の模式図を検索に利用することもできるが、検索結果がテキストデータであることには変わらない。研究に直結したイメージを喚起させるためには、遺伝学の慣例に沿った方式で、検索キーの与え方、検索結果の出力共に図式化されていることが望ましい。日常的に参照するマッピングデータとその出典などのデータを簡便に取り扱え、さらに馴染みやすく日本語も導入したマッピングデータベースシステムとして*GeneView^{PLUS}*を開発した。研究室で発生した生データの記録や検索、表示にも利用できるよう、*Physical Mapper*のようなデータ入力ツールと統合したシステムを目指している。

5.2 マッピングデータの入力、マップ構築ツール *Physical Mapper* の特徴

染色体上にマッピングされた実験データの検証のための図示や発表のための作図は、手書きやドローソフトウェアで行われてきた。しかし、ドローソフトで描かれた図では、データの追加、修正が容易でなく、系統的なデータの記録に適していない。また、長さの比率等をあらかじめ計算して設計する必要がある、スケールの違うデータを比較表示することは難しい。そこで、塩基数をスケールとした、物理地図専用のマッ

作成・表示ツール *Physical Mapper* の設計・開発を開始した。*Physical Mapper* は、コンテイングマップ、制限酵素地図、STS マップなど多様な物理地図を、ほとんどマウス操作のみで容易に作成し、それらのマップは同一画面上に表示することができる。スケールは染色体 DNA 全長 (~250Mb) からプラスミド (~5kb) までの広い範囲を任意に選択することができ、ズーム機能により自由に拡大・縮小しながら、マップの作成・表示が可能である。*Physical Mapper* は、統合ソフトウェアである *GeneView^{PLUS}* の中で物理地図の入力や表示機能を担うモジュールであるが、単体として制限酵素地図やクローンの記載など広範囲に利用することも想定して設計している。

6. *GeneView^{PLUS}* の機能

6.1 *GeneView^{PLUS}* の特徴

GeneView^{PLUS} は、メニュー、ボタン、項目などを日本語で表示し、マウスによる操作を基調とした分かりやすい GUI を用いたデータベースで

ある。

GeneView^{PLUS} を起動するとオープニング画面が開き、そこから検索の種類を選んで検索画面を開く。マップ検索では、遺伝子記号、遺伝形質、存在領域、遺伝子等のオブジェクトから検索し、検索結果は一覧表で表示される (図 1 *GeneView^{PLUS}* の検索画面、一覧画面)。

そこからマップ上に表示したい遺伝子等を指定し、位置を遺伝学で用いられる縦型の染色体イデオグラムに沿った複数のバーで表示する。表示する遺伝子数に制限はなく、ディスプレイの幅を超えたデータはスクロールして見るが、折り畳みによって空きスペースを有効に利用する機能もある。また、イデオグラムの任意の範囲を指定し、拡大表示することができる。一方、イデオグラム中のバンドを指定して、その領域を検索キーとして、そこにマップされている遺伝子を検索、表示することも可能である (図 2 *GeneView^{PLUS}* のマップ画面)。

それぞれの遺伝子等に対する詳細データは、一覧表をドラッグまたは、マップ表示上のエン

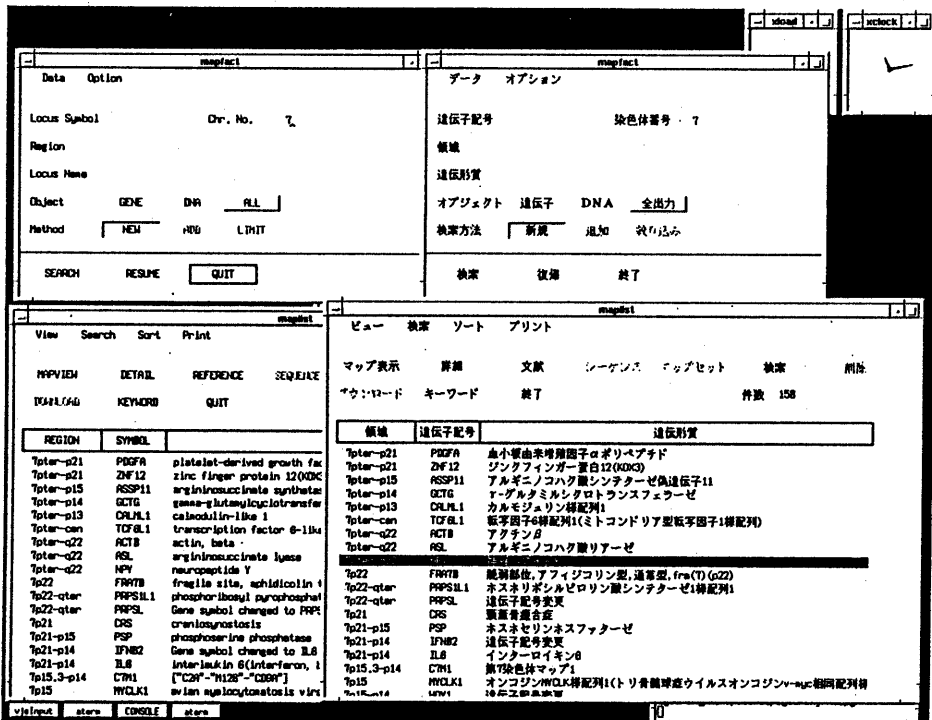


図 1 *GeneView^{PLUS}* の検索画面、一覧画面

トリーのクリックで文字を反転し、詳細ボタンで表示する。詳細表示ウィンドウ内のボタン(次遺伝子、前遺伝子)をクリックして、選択した前後のエントリーの詳細表示に移動できる。

文献データは、マップデータにリンクした検索と著者、文献タイトル等の文献に付随した情報からの検索の両方が可能である。前者はマップ情報の詳細データ表示と同様、エントリーをクリックで反転し文献ボタンを選んで行う。後者は、オープニング画面から文献検索メニューを選んで行う。次遺伝子、前遺伝子ボタンは文献表示と詳細表示の内容を連動して移動する(図3 *GeneView^{PLUS}*の詳細データ、文献表示ボックス)。

6.2 *Physical Mapper* の特徴

Physical Mapper は、ヒト染色体DNAの特定範囲を設定し、その領域における制限酵素地図やコンティグマップを描くツールである。常用する制限酵素をあらかじめ登録しておくことができる。扱うDNAクローンはYAC、PI、コスミド、プラスミドなどより選択し、これらはマップ上でもカラーで区別して表示できる。

Physical Mapper を起動すると、オープニング画面が現れ、そこで新規データの入力またはすでに記録されたデータの編集を選択する。マップを記録する画面には、そのマップの全長を表示するバーとそのマップの染色体上での位置が表示されており、5種類の制限酵素地図を記録する領域とScale、Basepointのバー、整列クローンを記入する領域が平行なバーで示される(図4 *Physical Mapper* の画面)。

画面はボタンを用いてズームイン、ズームアウトして、マップ領域のバーに表示された部分を拡大、縮小したり、左右にスクロールすることができる。データの入力を行う時はEditメニューから入力データの種類を選択する。ウィンドウが開き、制限酵素地図、コンティグ地図にそれぞれ対応したデータ項目を記入する。複数のデータを連続して入力するとグループとして取り扱われる(図5 *Physical Mapper* の画面と入力ダイアログボックス)。

Physical Mapper のもう一つの大きな特徴はマップオブジェクトの位置をそのマップ内での絶対座標以外に他のオブジェクト(制限酵素切

断部位、クローンの左端または右端等)からの相対的距離で表すことができることである。あるオブジェクトの位置を相対距離で表す場合は、基準点を示すためにBasepointスケールが用いられる。この方法によると、2種のマップを規準となるクローンとの位置関係を通じて互いに連結することができる。例えば、別なコンティグの島として登録されていた2個のコンティグマップCMAとCMBを想定する。CMAの右端とCMBの左端に存在したギャップを埋めるマップ情報(新規クローン、FISH等による距離情報等)が得られたとすると、CMAの右端のオブジェクトを新たにCMBの左端のオブジェクトの基準点に設定し直すことにより、他のオブジェクトについて何の修正もなく、一つの長いコンティグに連結することが可能である。

7 *GeneView^{PLUS}* の評価

GeneView^{PLUS} は、画像表示、日本語、迅速な検索などの特徴を持った世界的にも画期的なシステムである。このプロトタイプは、*GeneView*として設計・作成された⁹⁾が、その後、大幅な機能拡張に伴い全面的に新しいプログラムに書き直し、現在の*GeneView^{PLUS}*に至っている。学会等のデモンストレーションを通じて多くの研究者の関心を集めており、また、プリントアウトであるマップデータ集は、日本語版ヒトゲノムマップのアトラスとして利用され始めている。

日常的なデータの記録や整理はデータベースや表計算ソフトウェアを利用し、発表データを作成する際にはドローソフトを用いるという、二重の手間と苦勞を強いられてきたのが研究者の実状である。*Physical Mapper* は、そのようなマップ入力・作図の現状を改善する有用なツールとして作成されている。*GeneView^{PLUS}* の一部分としてだけでなく、分子生物学における作図ツールとしても十分に機能すると考えられる。

Physical Mapper は、開発途中であり、予定されている機能のうち現時点では実装されていないものも多い。それらの作成を継続する一方、動作確認及びテストデータ入力を通じて、改善すべき点、追加すべき機能を検討中である。

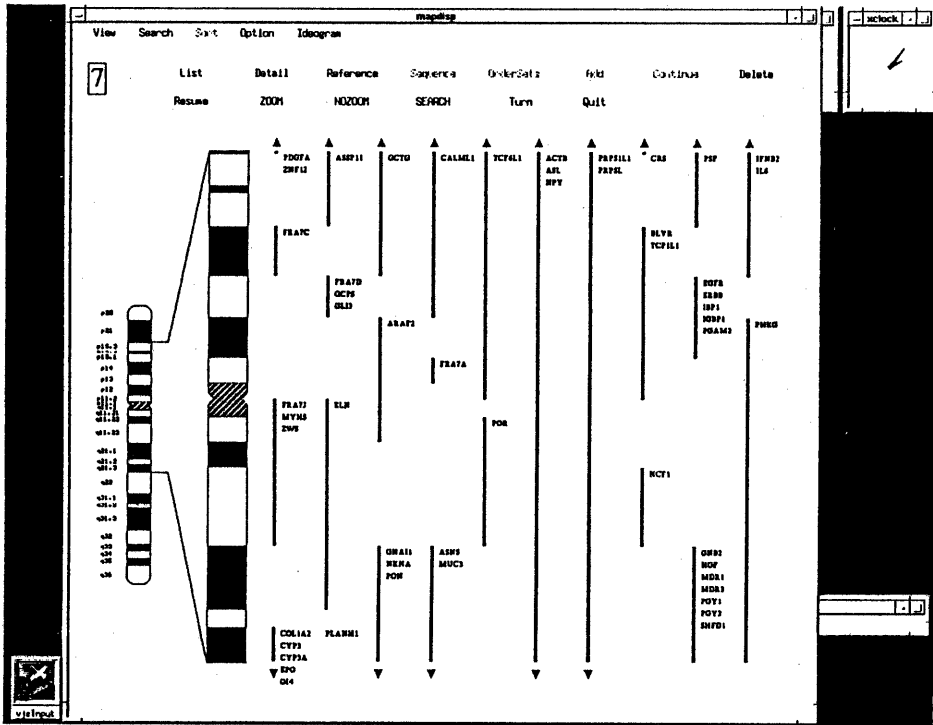


図2 GeneView^{PLUS}のマップ画面

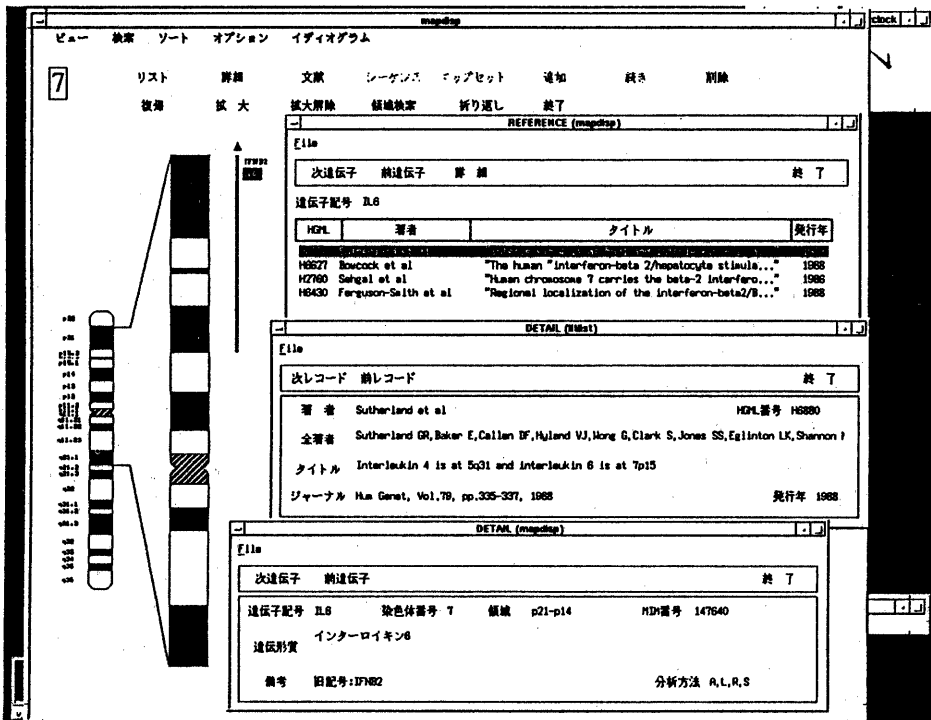


図3 GeneView^{PLUS}の詳細データ、文献表示ボックス

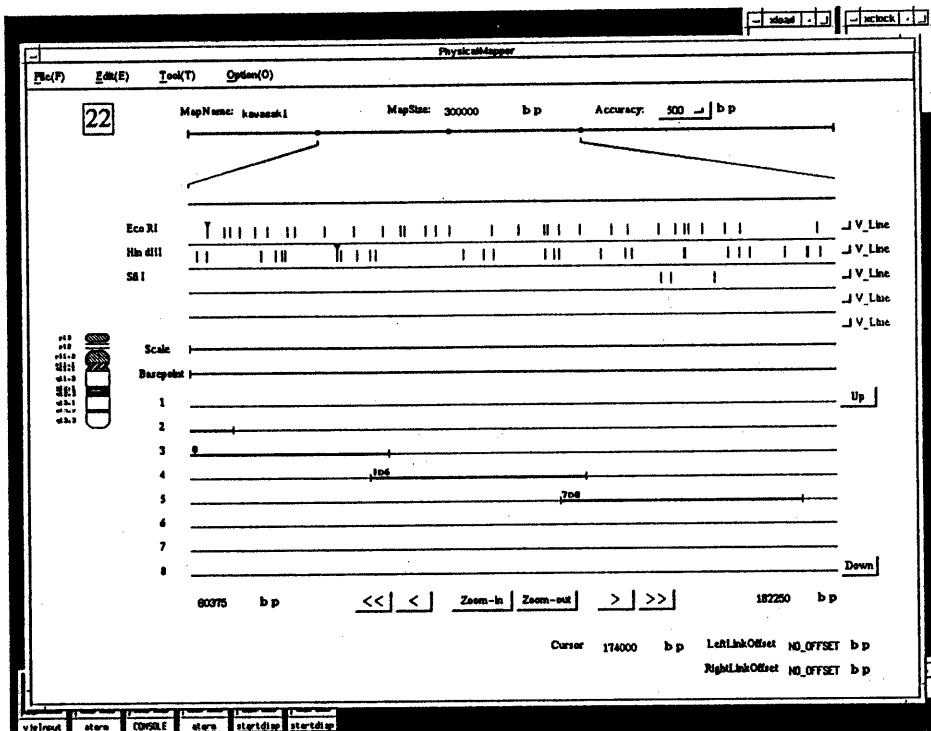


図 4 Physical Mapper の画面

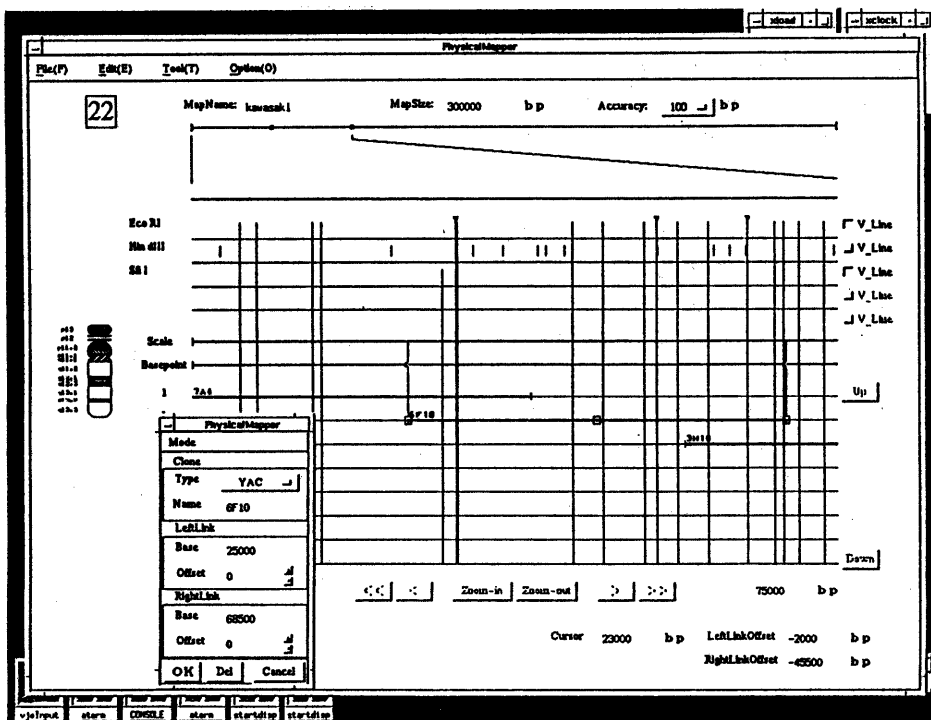


図 5 Physical Mapper の画面と入力ダイアログボックス

8 .GeneView^{PLUS}の課題

*GeneView^{PLUS}*は、*Physical Mapper*のような簡便な物理地図作成・表示ツールとGDBデータを日本語や画像表示などで分かりやすく検索できるデータベースから成るヒトゲノムマッピングデータベースシステムである。現在は、それぞれの部分を独立に開発した段階であり、データを相互に利用し合い表示することはできない。将来的には、*Physical Mapper*上のクローンやコンティグ等のデータから*GeneView^{PLUS}*の詳細データや文献データを参照したり、*GeneView^{PLUS}*のマップの遺伝子から、周辺のクローンやその位置関係や重なり、距離を*Physical Mapper*で表示することを目指している。染色体レベルのマップは、*GeneView^{PLUS}*で縦型の染色体上に必要数だけ表示することができ、*Physical Mapper*でクローン間の位置や距離を水平なバーとして表されるため、慣例に従ったイメージでデータを取り扱うことができる。また、統合システムとして使用することにより、GDBの公知データと実験で発生したデータを相互に参照し、研究の指針を立てることが可能であろう。

研究上発生し*GeneView^{PLUS}*に入力されたデータは、GDBに登録されるべきであり、その作業が自動化されることは当然の要求と言える。現在GDBに直接データを登録する場合は、登録用紙に記入しFax送付するか、テキスト状のフォーマットに記入し電子メールで送っている。ジョンズホプキンス大学でもデータ入力の効率的方法を検討中であるため、協調的方法を考案しGDBへのデータ提出を行うための仕組みを実装する予定である。また、同一インターフェース上からネットワークを経由して直接GDBを検索し、最新データや関連するその他のデータを検索する機能も懸案事項である。*GeneView^{PLUS}*は、GDBのデータ資産をフル活用して研究を進展させ、さらにGDBのデータを産み出すという循環を通じて、ヒトゲノム解析に大いに貢献するシステムとなるであろう。

謝辞

本研究は、科学技術庁科学技術振興調整費・研究課題「ヒト遺伝子地図作製技術の開発に関する研究」として援助を受けた。

参考文献

- 1) 平川美夏他, 「GDBTM (ゲノムデータベース) システムの概要」第30回情報科学技術研究会発表論文集, 131-137, 1993
- 2) 平川美夏他, 「GDB (ゲノムデータベース) - ヒトゲノム解析プロジェクトにおけるデータベースの構築 -」, 情報管理 Feb. 36 (11): 1023-1032, 1994
- 3) GDB Adds Nodes in Israel, Japan, and France, Human Genome News, May 6 (1):11, 1994
- 4) Shimizu, N. General report of the CCM93 meeting in Tsukuba, Human Genome Mapping 1993, The Johns Hopkins University Press, 1993
- 5) Shimizu, N. et al., Three Computer Softwares to Manage Various Genome Mapping Data, Report of the Fifth International Workshop on Human Chromosome 21 Mapping 1994, In press
- 6) Dohi, H. et al., *GeneView*: multi-language human gene mapping library with a graphical user interface, CABIOS, 9 (4):459-464, 1993