

## 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—

平川美夏<sup>\*1</sup>, 鈴木政彦<sup>\*1</sup>, 諏訪秀策<sup>\*1</sup>, 宮沢三造<sup>\*2</sup>, 清水信義<sup>\*3</sup>

## GDB (Genome Data Base)

### —The construction of database in Human Genome Project—

HIRAKAWA Mika, SUZUKI Masahiko, SUWA Shusaku, MIYAZAWA Sanzo, SHIMIZU Nobuyoshi

[著者抄録] GDB(Genome Data Base)は、国際的ヒトゲノム解析プロジェクトの研究成果の蓄積・流通を担うデータベースである。米国 Johns Hopkins 大学で、国家および国際的助成の下で作成・維持されている。世界6か国にノードが設置され、日本では JICST が公式ノードを開設する。ゲノム地図データベースとして、遺伝子座、地図、多型情報とその実験材料や文献情報および、病気に関わる遺伝子のカタログ OMIM(Online Mendelian Inheritance in Man)を収録している。

[著者付与キーワード] GDB (ゲノムデータベース), ゲノム地図, ヒトゲノム解析プロジェクト

[Author Abstract] The GDB(Genome Data Base) is a database that is responsible for acquisition and distribution of the information generated by research of the international Human Genome Project. It is developed and maintained at Johns Hopkins University in federal and international funds. The six of international nodes have been established and the Japan node is open at JICST. The GDB is collecting genome mapping data such as loci, maps, polymorphisms, research materials, citations and OMIM(Online Mendelian Inheritance in Man).

[Keywords Author] GDB(Genome Data Base), genome map, Human Genome Project

- 
- \* 1 日本科学技術情報センター技術開発部 (〒100 千代田区永田町 2-5-2) Tel. 03(3581)6411  
JICST Technology Research and Development Department  
(5-2, Nagatacho 2-chome, Chiyoda-ku,100)
  - \* 2 群馬大学工学部工学基礎教室 (〒376 桐生市天神町 1-5-1) Tel. 0277(22)3181  
Gunma University, Faculty of Engineering (5-1, 1-chome Tenjincho, Kiryu-shi, 376)
  - \* 3 慶応大学医学部分子生物学教室 (〒160 新宿区信濃町35) Tel. 03(3353)1211  
Keio University, School of Medicine (35, Shinanomachi, Shinjuku-ku, 160)

## 1. はじめに

生物が生まれ、生きていくために、それぞれの生物ごとにあらかじめ用意されているすべての情報、それが「ゲノム」である。「ゲノム」の情報は、DNA (デオキシリボ核酸) の配列として、世代を越えて受け継がれ、あらゆる生物の存在と生存の基本となっている。物質レベルでは、細菌も植物も動物も、情報の媒体として等しくDNAを利用しているが、「ゲノム」という単位では各々の生物固有の構造と大きさを持ち、これこそが地球上の生物の多様性の根源を成すものである。

現在、様々な生物の「ゲノム」解析プロジェクトが世界中で進行中である。なかでも人間を生物である「ヒト」として、30億塩基のDNAからなるゲノム情報を解読することを目的とする「ヒトゲノム解析プロジェクト」は、国際レベルの協力のもとに生物学史上初めての巨大プロジェクトとしてめざましい進展を見せている。米国では、プロジェクトの達成が思いのほか順調であるとして、先頃当初の5か年計画を更新している。主要な成果は、ゲノムの地図の作成において上げられており、それを収集、管理しているデータベースがGDB (Genome Data Base) である。

## 2. ゲノムプロジェクトの成立

### 2.1 ゲノムとゲノム「地図」

すべての生物の生存を支える情報は、遺伝子としてDNAのアデニン、グアニン、シトシン、チミン (それぞれA, G, C, Tの文字として記述される) の4種の塩基の配列で表されている。さらにそれを元に蛋白質を合成する機構から、その蛋白質の機能そのものまで、多くの同じシステムを共有している。したがって、生物の多様性やヒトをヒトたらしめている情報は、固有の遺伝子にあるというより、ゲノムとしての構造にあり、遺伝子が正しくヒトのゲノムの構造を再現したとき、そのDNA配列は、ヒトを作り出すと考えられる。しかし、約10万個と予想されているヒト遺伝

子は、全ゲノムDNA中の5%を占めるに過ぎず、残りの95%はこれまでの遺伝子のDNA解析においては、ほとんど目を向けられてこなかった。ゲノムの解析とは、この未知の95%を含めた全体像を明らかにすることを目的としており、それを集積する手段が、「ゲノム地図」である。

ゲノムは、生物を構成するほとんどすべての細胞の核中に染色体として存在し、各生物は固有の染色体構成を示す。ヒトでは、ゲノムDNAは22本の常染色体とX, Yの性染色体に分かれている。ゲノムの地図とは、どの染色体上のどの位置にどのような遺伝子が存在し、各遺伝子間にはどのような構造が挟み込まれているかを表すものである。

### 2.2 ヒトゲノムプロジェクト

1970年代後半、遺伝子組換え、DNA塩基配列解析技術の進歩と普及に伴い、個々の遺伝子の塩基配列の解読が広く生物学の分析手法として用いられるようになった。また、1980年、日本の理化学研究所の和田が、DNA解析の全自動化の可能性を指摘したことが、DNAの大量解析時代への幕開けとされている。

1986年に、DulbeccoがScience誌上でヒトゲノムの全解読の重要性を論じ、その直後のサンタフェの会議でヒトゲノム解析の可能性が、議論され、ヒトゲノム全解読の実施がいよいよ現実のものとなった。1988年には、ヒトゲノムプロジェクトの始動に併せて、プロジェクトの国際協力を唱える研究者の自発的組織、HUGO (Human Genome Organization) が結成された。

アメリカでは、1990年、国立衛生研究所(NIH) / エネルギー省 (DOE) の協力の下で、ヒトゲノムプロジェクトの15か年計画を発表し、当初5か年の目標を、ゲノムの地図作りに据えた。その第一段階として、染色体のバンドレベルの解像度の地図作成を実施してきたが、技術開発と精力的な取組により進展が早まっているとして、さらに5年後を見据えた新規計画を発表した。地図作成は順調に進んでおり、さらに今後解像度を上げ、成

果の効果的利用法を検討する時期であるとしている。特にデータベースと情報解析については、プロジェクトの成功を左右するとして、重要性を指摘している。

英国、フランスをはじめとする欧州各国や日本などは、アメリカの動きに注目しながら、独自の方針でヒトゲノム解析を推進している。実際の研究の現場には国境を越えた協力体制が浸透しており、理化学研究所とローレンスリバモア国立研究所やNIHとフランスのCEPH（多型解析センター）のように、各国の主要研究機関が共同研究を実施している。また、30余国の研究者が参加するHUGOが調整役を演じ、各種ワークショップなどを通じて国際的交流がなされている。

### 3. GDB (ゲノムデータベース)

#### 3. 1 GDBの経緯

ヒトゲノムプロジェクトのような大規模なDNA解析が実施された背景には、1980年代のDNA解析技術の普及と自動化を含む高速化が挙げられるが、ヒトの遺伝子を地図化する必要があるという動きは、1973年に開始された、HGMW (Human Gene Mapping Workshop) まで遡ることができる。隔年で開催されたこの会合の成果は、HGML (Human Gene Mapping Library) としてYale大学でデータベース化されていたが、ヒトゲノム解析プロジェクトの発生に伴う重要性の増加から、より集中的なデータ管理体制が求められ、1989年に現在のGDBに再編成を開始した。ゲノム研究を早くから助成してきたハーワードヒューズ医科学研究所と英国の王立がん研究基金の支援の下、McKusickのヒト遺伝病に関する文献データベースOMIM (Online Mendelian Inheritance in Man) の作成元であるジョンズホプキンス大学 (JHU) に移行し、リレーショナルデータベース Sybase 上に移植され、1990年9月にリリース1を公開した。

#### 3. 2 GDBのデータ

##### 3. 2. 1 研究と発生データ

ゲノム地図には、一般に遺伝的地図と呼ばれる「連鎖地図」と「物理的地図」がある。「連鎖地図」は、世代間の染色体の組換えの頻度から、遺伝子間の距離を算出する方法であり、家系図中に現れる表現形質と遺伝子マーカーとの関係、染色体上での分布が得られる。データとして、個人を区別するのに有効な多型を有するマーカーが蓄積する。遺伝学で地図とは、「連鎖地図」を指す場合が多いが、GDBのデータとしては、「連鎖地図」作成の国際的機関であるCEPHのデータを導入するまで、「物理地図」データが主流であった。

「物理地図」には、ボトムアップとトップダウンの2方向からのアプローチがある。染色体は、DNAを凝縮して折り畳んだ構造をしているが、ボトムアップは、断片化したDNAから染色体構造へと組み立てる方法であり、トップダウンは、染色体の位置にDNAを特定していく方法である。したがって、ボトムアップでは、DNA断片とその性質、順序関係が、トップダウンでは、特定の遺伝子やマーカーなどと染色体上の位置関係が明らかになる。データとしては、順序が決められたDNA断片のセットや染色体座に位置づけられた遺伝子やDNA断片が発生する (図1)。

##### 3. 2. 2 研究とデータベースデータ

GDBでは、位置とその位置に対応する遺伝子等のDNA領域とブレイクポイントなどの構造を「ローカス (遺伝子座)」として、その実験に使用されたDNAを「プローブ」、位置や順序によってまとめられたローカスのセットを「マップ (地図)」として蓄積する。

GDBのデータの特徴に、参照するデータが、研究に直結していることがあげられる。プローブ情報には、実験法と検出法が記載され、持ち主と直接接するのための、連絡先も登録されている。多型情報は、検出するための制限酵素と発生する断片長が示され、検定集団の国、人種情報もデータ化されている。ローカスの位置決定は、同じ遺伝

子についても、より詳細な解析を繰り返して、解像度を上げていくため、このようなデータを共有することは、研究推進上有効である。

### 3.3 GDBの構造

GDBは、ヒトゲノム解析研究の過程および、結果として発生するデータをデータの種別別に「マネージャー」という単位で扱うようになっている。

HGMLの遺伝子座、引用文献、多型、プローブ、プローブの持主の連絡先の各情報にOMIMを加えた六つのマネージャーで最初のGDBはリリースされた。地図の情報を搭載したバージョン4.0は、1991年7月、ロンドンで開催されたHGM11を目指して作成され、対立遺伝子座情報も合わせて、OMIMを含む9マネージャーになった。バージョン5.2は、12種のマネージャーから構成されている(図2)。

それぞれのマネージャーは、ウィンドウ型の検索画面と表示画面からなり、検索や画面切り替え

はメニューを指定して行う。

### 3.4 GDBの動作環境

GDBは、Sunワークステーションで、リレーショナルデータベース Sybase 上に構築されている。JHUでの開発環境はSPARCサーバー690、OSはSunOS4.1.2を使用している。データベースは、検索ソフトウェア開発ツールキット DAT を介して、ユーザインターフェースから検索される。JHUでは、現在、そのほかSQLサーバー、FTPサーバー等の5種類のサーバーを用意して、GDBを提供している。OSをSolarisに、SybaseをSystem10にバージョンアップする計画が、発表されている。

### 3.5 GDBのデータ量

GDBは、1993年10月現在、4千を超える遺伝子を含む、約3万のローカス情報と710種の地図情報を登録している。データのディスク容量は、500MBを越え、予定より半年以上早く到達したとしている。この量は、ローカス数が1万を越えたバージョン4.2の約2倍であり、1年数か月の間にデータ量が倍増した計算になる。

### 3.6 国際的共同

#### 3.6.1 GDBコミュニティ

GDBは、開発当初より英国のMRC(医科学研究審議会)と開発および資金面での協力を開始し、DKFZ(ドイツ癌センター)や米国内の各ゲノムセンターと協同開発を行っている。開発やデータ収集などの方針は、各国の研究者からなる諮問委員会の場で審議し、1データベース機関に留まらない国際的規模の実施体制を敷いている。

また、国際ワークショップHGMWのデータ蓄積を目的としてきた経緯を踏まえ、データ入力にはHGMWを初期から支えてきた参加者とHUGO委員に選出された染色体エディタが行い、ワークショップの場で内容を評価、検討している。この伝統は、現在も残されているが、データや参加者の増加から、数日間のワークショップで25染色体

地図の作成法

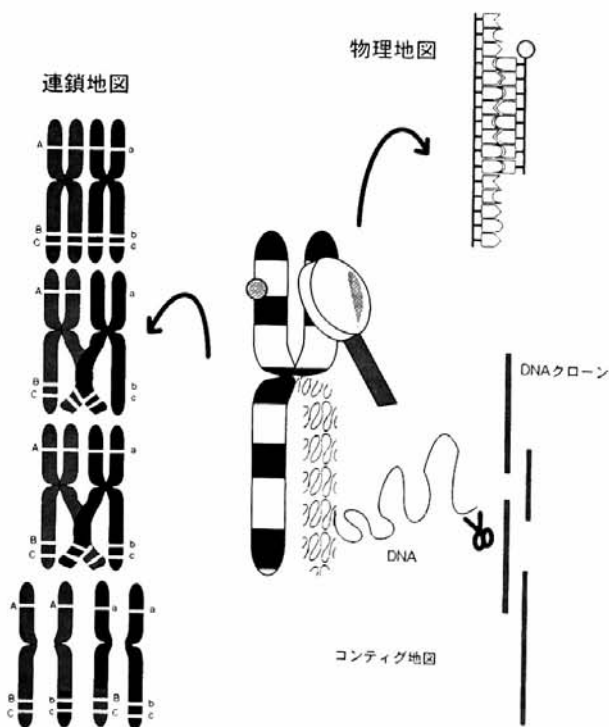


図1 ゲノム地図の作成法

(22常染色体, 2性染色体とミトコンドリアDNA)および周辺の問題について議論を尽くすことが困難になり, 染色体の討論は, 染色体個別の会議 SCW (Single Chromosome Workshop) に譲って, HGM ワークショップの性質は発表会形式へと変化しつつある。

### 3. 6. 2 国際ノード

GDB は, 各国がその国の利用者に対し, 国内回線を利用して GDB を提供し, 母国語によるサポートを行うように国別にノードを開設することを奨励している。現在, 英国, ドイツ, オーストラリア, オランダ, スウェーデンに公式ノードがすでに設置されており, フランスは事実上運用を実施し, カナダ, イタリアが予算措置を準備中である。日本では, 科学技術庁が共同分担金を出資し, 理化学研究所と日本科学技術情報センターの協力体制の下で1994年2月より, 公式ノード運用を開始する。

## 4. GDB の構築

### 4. 1 データ源

GDB データ源は, 主として学術論文である。基本的には, NLM が作成する医学文献ファイル Medline から遺伝子座決定や地図作成, 多型の検出などに主題を置いた文献を抽出したものを, JHU が染色体毎に各エディタに配布し, それを参照しながらエディタがデータを作成する。GDB の文献情報の多くは, データとして, 書誌情報の他, キーワード, アブストラクトも Medline から取りこんでいる(図3, 4)。GDB は, OMIM をデータベースシステムを含めて統合しているが, データのみの導入は酵素番号, 制限酵素名, ATCC など材料バンク情報, ジャクソン研究所 GBase に由来するマウスゲノム情報などがある。最新のバージョンでは, Medline の関連データベース情報も取り込まれ, 文献中に出現した化合物について, CAS のレジストリー番号を示している。また, 研

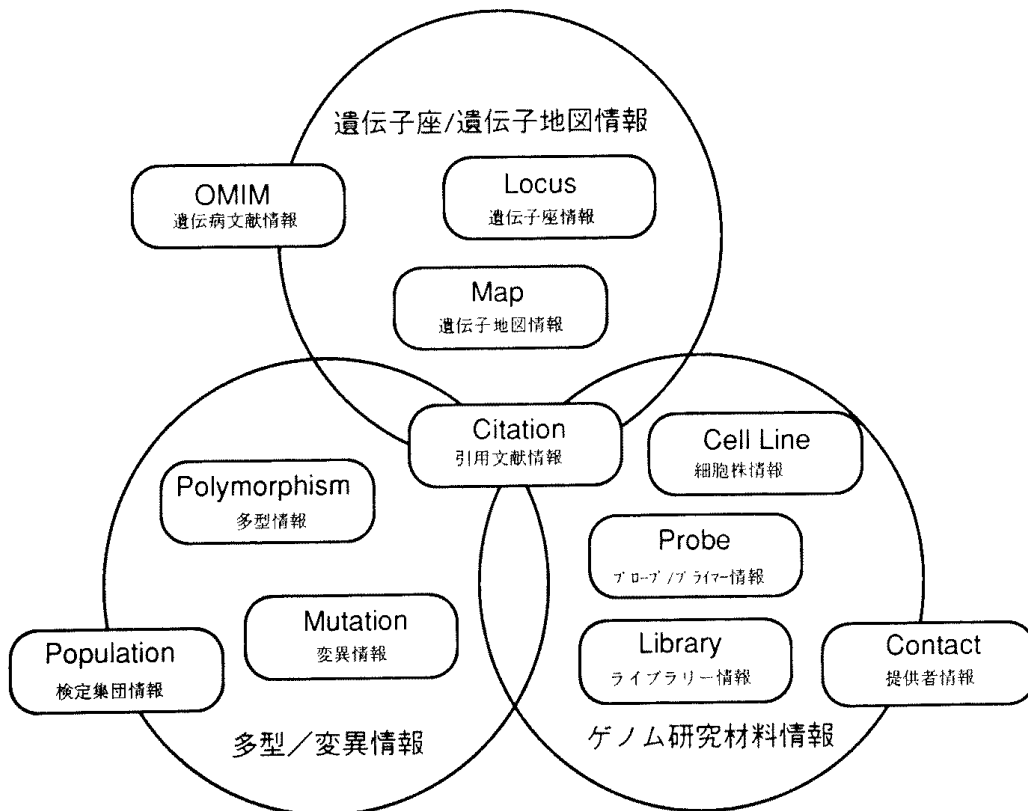


図2 GDB のマネージャー構造

研究所内の実験データベースとしては、ローレンスリバモア研究所のDNAクローンを連結させて作成されるコンテイング地図データが、ネットワーク経由でGDBに転送されている。フランスのCEPHやGenetonからも連鎖地図情報や多数の材料情報を導入し、MRCのプロジェクトともデータ交換を実施している。

#### 4.2 データ定義

GDBは、ゲノムの地図情報のデータベースとして、遺伝子や染色体上の位置決定に有効なマーカー配列を特定の位置と結び付け、それを効果的に利用するための情報を蓄積することを、使命としている。その機能の一つとして、遺伝子、DNA部位、ブレイクポイント等の遺伝子座すべてに、

**Citation Manager**

Go Edit View Call Select Retrieve Tool Help Exit

---

Locus: CMT1B

[ ] Type: Journal Citation 3 of 6 of set 0

Authors: Lebo, RV; Dyck, PJ; Chance, P; Bird, TD; Golbus, MS; Lynch, ED; Redila-Flores, MT; Anderson, LA; King, MC; Hall, JM; et al.

Title: Mapping the Duffy-linked Charcot-Marie-Tooth Disease locus.

Journal: (CYTOGENETICS AND CELL GENETICS )

Address:

Issue:	Vol: 51	Pages: 1030
Suppl:	Review: No	Language: Eng

Abstract: Yes      Published: Yes

Date: 1989

Medline Id:                      GDB Id: G00-001-368

Created: 15 Aug 90 16:03 Last Mod: 7 Apr 91 11:07

図3 GDBの文献情報 (Citation マネージャー)

**View Abstract**

Tool Help Exit

---

Citation: Lebo, RV Cytogenet Cell Genet 51:1030 1989

Charcot-Marie-Tooth Syndrome (CMT) are neurologic disorders that begin with peroneal muscle weakness and atrophy and advance insidiously to other distal leg and arm muscles. This worldwide disease has an incidence of about 1/3000 in Spain and Norway, 1/20,000 in England, and 1/10,000 in the eastern United States. Of the 76 families we studied, 68 are consistent with autosomal dominant inheritance, 6 with X-linked inheritance, and 2 have a single >

図4 Citation マネージャーの抄録画面

命名委員の与えたローカス（遺伝子座）シンボルを付与している。しかし、プロジェクトで発生する主要なデータであるコンティグ地図を構成するクローンは、特定の遺伝子の部分でも特異性の高いマーカーでもない、ランダムに切断されたDNAの1断片である場合が多く、さらに染色体バンドレベルの解像度では、別なコンティグに由来する断片同士の関係の特定は難しい。当初、ローカスとして認めてシンボルを付与するには、厳しい基準が設けられていたが、今では、発生した時点で命名し、関連情報が明らかになる度に追加する方式をとっている。そのため、ローカス同士が重視する場合が発生し、本来領域をカバーし得るクローン数や長さよりも、多くなる可能性が出てきている。しかし、GDBは、情報の網羅的収集と迅速な公開のため、この点に関しては、緩やかな対応をとっている（図5, 6）。

#### 4.3 GDB Idとデータ提出

データの集中と速やかな収集のため、GDBはGenBankなどの例に習い、論文投稿時に雑誌側で掲載条件として、GDB Idの取得を義務づける制度を発足させた。GDBのデータは、DNAの配列データに比べ、多彩で複雑なため、GenBankのアクセッション番号のような1配列1データに準ずる遺伝子座のシンボルによる特定化に留まらず、データほとんどすべてに特有のIdが授けられる。GDBバージョン5.1から設けられた、GDB Idマネージャーは、19種のデータタイプ別にIdを管理している（図7）。また、著者自らデータ提出を行う手段として、現在は専用のフォーマットの用紙を使用しているが、DNAデータベースのAuthorInに匹敵するような、データ提出ツールの開発が進んでいる。また、この点に関しては、各国が自国の利用者の使いやすい環境を提供し、データ収集に貢献することが期待されている。

#### 4.4 データ入力の現状と展望

エディタのデータ入力、研究の合間のボランティア的な活動であるため、ワークショップ直前

に集中的に行われることが少なくない。そのためゲノム解析プロジェクトで膨れあがったデータに対応するために、作成元であるJHUのスタッフが日常的にデータ入力を行っている。また、命名委員の所属機関やGDB-PEN（多型情報入力サイト）など外部入力の分担体制もあり、次第にその割合は増加している。そこで、JHUは、データ入力は専任のアシスタントを採用し、エディタの役割としては、データの評価と品質管理を担当することを提案している。しかし、データの著者入力が増えていくことが予想される現在、データの監視体制の確立が必要とされている。

### 5. 今後の展開

#### 5.1 GDBの改造計画

GDBは、ゲノム解析プロジェクトの各種のデータに対応するために改良を繰り返している。米国ゲノムプロジェクトの目標の見直し、DOEの生物情報の統合、高速化の方針を受けて、かなり大きな見直し計画されている。それは、現在の12のマネージャーをマップを中心としたメインのデータベースと多型などの関連データベース、文献データベースに分類し、再構築するというものである。文献データベースは、CitDBとして他の遺伝子関係データベースとともにゲノム文献情報データベースとして再作成されるため、OMIMの重要度が高まっている。

#### 5.2 OMIMプロジェクト

OMIMは、1966年からジョンズホプキンス大学のV.McKusick博士の手で作成され、10版を数える遺伝病やヒトの表現形質に関わる遺伝子の膨大なリストである。ゲノムプロジェクトの進展に連れ、各遺伝子と疾患の関係が明らかになり、それを医療現場に適用していく場合、データベースとしてより充実した機能が求められるため、GDBに匹敵するプロジェクトとして再編成が開始された。特に、遺伝子と疾患の関係は、1遺伝子が複数の疾患の原因となるケースや、一つの疾患に複

**Locus Manager**

Go Edit View Call Select Retrieve Tool Help Exit

[ ] Locus 1 of 1 of set 0

Locus Symbol: CMT1B Previous HGM Symbol: CMT1  
 Locus Name: Charcot-Marie-Tooth neuropathy 1B (greatly reduced nerve  
 Cyto Location: 1q  
 MIM # Disorder  
 118200 Charcot-Marie-Tooth neuropathy, slow nerve conduction type Ib  
 Assign. Mode: (L ) EC Number:  
 GenBank: Probes: No  
 Polymorphic: No Het: Allele Set:  
 Ref. Marker: No

Citations: Cytogenet Cell Genet 58:1854 1991  
 Cytogenet Cell Genet 37:3-21 1984

Annotation:  
 Created: 1 Jan 86 00:00 Last Modified: 3 Oct 91 10:05  
 GDB Id: G00-120-594

図5 GDBの遺伝子座情報 (Locus マネージャー)

**Map Manager**

Go Edit View Call Select Retrieve Tool Help Exit

[>>] Map 19 of 21 of set 0

Map Symbol: C6M38 Type: (Primary ) No. of Elements: 55  
 Method: (CEPH-Average )  
 Location: 6p23-q

Map: D6S344\_(D6S309,D6S296/D6S277/D6S263)\_D6S259\_(D6S289/D6S260)\_D6S288\_D6S274\_D6S285\_D6S299\_(D6S306/D6S276/D6S265/D6S258)\_D6S273\_D6S291\_(D6S271/D6S282)\_(D6S269,D6S272)\_(D6S295,D6S294)\_D6S257\_D6S313\_D6S280\_(D6S284/D6S286)\_(D6S275,D6S300)\_D6S283\_D6S301\_D6S268\_D6S278\_D6S302\_(D6S261/D6S266)\_(D6S303/D6S267/D6S304/D6S287)\_D6S262\_D6S270\_D6S292\_(D6S308/D6S310/D6S314/D6S279)\_D6S311\_D6S290\_D6S305\_D6S264\_D6S297\_D6S281

Citation: Nature 359:794-801,1992

Annotation:  
 Created: 26 Apr 93 14:14 Last Modified: 5 May 93 11:20  
 Release Date: GDB Id: G00-138-311

図6 GDBの地図情報 (Map マネージャー)



数の遺伝子が関わるケースなど複雑な関係を表現する必要があり、現状のテキスト形式では対応しきれない部分があるため、リレーショナルデータベース化と SGML の利用が計画されている。

### 5.3 関連データベースとの関係

ゲノム解析とその成果である地図情報のデータベース化は、他の生物についても進行中であり、それぞれの研究者のコミュニティが独自のデータベースを構築している。一方で、生物ごとの区別をせずに蓄積されてきた DNA の配列データベースや蛋白質構造のデータベースとゲノム地図データの統合利用が期待されている。関連する生物情報データベースを多型情報等と同様の関連データベースとして、地図データの周辺に対等に連結する計画があるが、個々のデータベースは、それぞれの方針で開発やデータの蓄積を実施しているので、相互に検索するために共有キーをポインターとして与えておく必要があることが認識されている。現在、GDB と GenBank は協力体制を敷き、

GenBank のアクセッション番号と GDB のシンボルや Id との擦り合わせが行われているが、プロジェクトの管轄の違いから、完全な共同作業までは至っていない。

### 5.4 GUI 開発と協調

GDB は、各ノードや共同開発機関に対し、GDB を活用するための GUI (グラフィカルユーザーインターフェース) の開発を奨励しており、すでに英国と EC の開発中の IGD (Integrated Genome Database) や SIGMA (ロスアラモス国立研究所)、ゲノムトポグラファー(コールドスプリングハーバー研究所)などのプロトタイプが完成し、システムとして一部提供が開始されている。いずれも地図の画像情報上に DNA 断片の整列やローカスの位置を表示し、GenBank などの配列データベースのデータを地図上の遺伝子と結び付けるシステムであり、この種の要求の多さを実感することができる。現状の機器の処理能力では、データベースを融合し、膨大なデータを突き合わせ整合性をとり画像化することは実際にむずかしい面もある。その上でプロジェクトの進展に伴う種々の要求に答えるため、画像化、データ処理能力、操作性の各面における優秀なシステムが生み出されることが、期待されている。

### 5.5 ネットワークの利用

現在注目されているのは、ネットワークを介して、それぞれのデータベースから関連データを自由に引き出す方法であり、JHU ではすでに WAIS や Gopher のサーバーを提供している。データの巨大化は避けられない現実であり、データの複雑化はしばしばデータベースの動作に負担を与える。GDB の将来計画にもあるように、データはデータ種ごとに分類し、その性質にふさわしいデータ構造を作り、適切な機関でデータ収集を実施して専門データベースとしての充実をはかることになるであろう。また、ネットワークの強化が欠かせない課題である。個々のデータベースは、それぞれの機能を尊重し、別な場所で作成されて

GDB Id Data Types	
Allele	
Allele Set	
Allele Set Pop	
Amp conditions	
Breakpoint sets	
Cell Line	
Citation	
Contact	
Cycle sets	
Detect Meth	
Library	
Locus	
Map	
Map Distance	
Mouse locus	
Mutation	
Polym	
Population	
Probe	

図7 GDB Id の付与されているデータ種

も、ネットワークを介し、WWW (World Wide Web) のように統合検索を実現する環境を利用すれば、あたかも一つのデータベースのように検索することができる。これを、仮想統合データベースと呼び、今後のゲノムプロジェクト関連データベースのあり方の一つとして挙げている。

## 6. おわりに—日本の現状と展望

ゲノムプロジェクトは、すでに走り出しており、その成果は、遺伝子の位置やその構造、疾患遺伝子座の同定と機能の解析として、最終結果を待たずして、日々の研究の中の随所に現れる。その成果が成果を生み、各データ間の関係が明確になるにつれ、ヒトとしての立体的な構造が再現されてくる。したがって、現状のデータは現状のまま有効に利用できる環境を整える必要があり、同時に将来を睨んだ、柔軟で安定した蓄積システムも構築しなければならない。現状では、GDBがそれを一手に担い、世界中のゲノム研究者コミュニティの知識がここに集中している。また、世界の関連する機関の情報関係者が注目し、参加の意志を示している。

日本では、科学技術庁の振興調整費や文部省の科学研究費の研究班において、ゲノム関連データの処理や表示を目的とした、システムの開発が行われている。また、ゲノム研究者コミュニティの交流と情報交換を目的とした、ネットワーク整備が進められている。GDBの日本における拠点としての公式ノードがJICSTに開設され、この分野における日本の貢献もますます大きくなるであろう。

1993年は、GDBのデータ検討会議CCM93とHGM93が日本で開催され、日本のゲノムプロジェクトの活性化の年となった。米国を中心とする国際的動きに対応しながら、GDB日本ノードが、日本の研究者コミュニティの連帯とデータ貢献の一助になれるよう、環境整備に努める所存である。

### 謝辞

本稿は、科学技術振興調整費研究課題「ヒト染色体地図作成技術の開発」及び、理化学研究所「GDB (ゲノムデータベース) の開発・提供に関する研究」の共同研究による。

## 参 考 文 献

- 1) Wada,A. Automated High-Speed DNA Sequencing, Nature325 : 771-772,1987.
- 2) Duibecco,R. A Turning Point in Cancer Research : Sequencing the Human Genome Science231 : 1055-1056,1986.
- 3) 松原謙一, 榊佳之 ヒトゲノムプロジェクトの成立 蛋白質核酸酵素 vol.38, 209-211 1993.
- 4) A.J.Cuticchia et al. Evolving Data Acquisition Strategies for the Human Genome Data Base1993 (GDB position paper)
- 5) OTA Report, Mapping Our Genes, April 1988.
- 6) GDB Fourm vol.1,number1
- 7) GDB Fourm vol.2,number1
- 8) GDB Fourm vol.2,number2
- 9) Human Genome News vol.4,number7
- 10) Gallas et al. A New Five Year Plan for the U.S. Human Genome Project Science vol.262 43-46 1993.
- 11) A.J.Cuticchia et al. Managing All Those bytes:The Human Genome Project Science vol.262 47-48 1993.