

## 植物遺伝資源データベースにおける画像データ管理

竹谷勝、川田真佐枝、梅原正道

農林水産省

農業生物資源研究所遺伝資源第一部

種子カードを画像データとして植物遺伝資源データベースに統合するためのシステムを開発した。このシステムでは、DBMSが画像データを直接管理するので、データの一貫性を保持することが可能となった。また、GUIを用いたシステムを構築することによって、データ操作の簡易化が図られた。画像データ管理システムについて、概要を報告する。

## Image Data Management of Plant Genetic Resources Database System

Masaru TAKEYA, Masae KAWADA, Masamichi UMEHARA

Department of Genetic Resources I

National Institute of Agrobiological Resources

Ministry of Agriculture, Forestry and Fisheries

We developed a new system to integrate plant genetic resources database with image data. This system has a consistency of image data and an ease of data manipulation. This paper presents an outline of image data management of plant genetic resources.

## 1. ジーンバンク事業の植物遺伝資源部門の概要

地球環境の悪化が国際的な論議となっている今日、生物種の多様性と種内の遺伝的多様性を保全することはきわめて重要な意義を持っている。農林水産省では、昭和60年からジーンバンク事業として、植物、微生物、動物、林木及び水産の5部門にわたって、遺伝資源の確保と利用を推進している。

ジーンバンク事業の植物遺伝資源部門は、農業生物資源研究所をセンターバンクとして、各地域農業試験場等をサブバンクとする体制で、遺伝資源の探索・収集、評価、保存および配布を行っている。参画研究室は100以上にのぼり、多数の研究者が関与している。

探索・収集としては、毎年国外へ4隊と国内に6ないし7隊の探索隊を送って、各植物種の分布と変異を調査し、遺伝資源の収集を行っている。その他に海外の研究機関との遺伝資源の交換も行っており、毎年5千ないし1万の新しい遺伝資源が収集されている。

評価は、収集した遺伝資源を形態に着目した解析法や、遺伝子レベルでの解析法等で分類・同定し、遺伝資源としての様々な特性を評価するものである。遺伝資源の特性評価については、「植物遺伝資源特性評価マニュアル」が定められており、これに従って全国統一の基準で評価を行う。

収集した遺伝資源は配布に必要な量を得るために、まず増殖を行い、その後保存を行う。センターバンクでは、種子での保存が可能な遺伝資源のみを保存しており、種子繁殖でない遺伝資源は、保存施設のあるサブバンクに分散し、保存している。また保存には、永久保存と、配布用保存の2種類がある。永久保存用の種子は、-10度、湿度30%の環境下で保存し、配布用の種子は、-1度、湿度30%の環境下で保存している。

配布は、配布申請に基づいて、試験研究用の材

料として、遺伝資源を配布するものであり、毎年約1万品種程度配布されている。配布によって在庫量が僅少になった遺伝資源は増殖を行う。

このジーンバンク事業により、現在では約20万の遺伝資源が保存され、育種等の研究に利用されている。

## 2. 植物遺伝資源管理システムにおける種子カード

植物遺伝資源のデータ管理は、リレーショナルデータベースマネジメントシステムを用いたデータ管理システムによって行っている。現システムを開発する前は、図1の種子カードによって遺伝資源の管理を行っていた。種子カードに記載された主要なデータは、植物番号や、品種名、原産地等の品種の来歴を表すパスポートデータと、種子庫の位置や、種子量、発芽率等の在庫管理のためのデータである。これらのデータは現システムでは27列のパスポートデータと各種の在庫管理テーブルに登録した。その結果、現行のシステムでは種子カードを用いずに遺伝資源の管理を行っている。しかし、パスポートデータおよび昔の在庫データの確認のため、種子カードが必要となることがある。また、種子カードを保存するためのシステムファイル(写真1)の耐用年数が過ぎており、これを更新するか、別途システムを開発するかの必要がある。システムファイル自体が高額なうえかなりの設置スペースを必要とすることから、種子カードを画像データとしてデータベースに統合することによって、種子カードを廃棄できるシステムを開発した。図2に画像データを含む植物遺伝資源管理システムの概要を示す。

種子カード

作物名	オオムギ	科名	イネ科	114125
登録番号	JENA 54-1	0-19	JENA 54-1	220257
原産地	1985	原産地	NEPAL	22/
採集者	岡山大学農学研究所	採集者	農学博士 渡辺 隆夫	22/
採集年月日	岡山大学 山西八郎 1986年 1-10月	採集地	ネパールの山岳地帯	
採集場所	ネパールの山岳地帯	採集方法	種子採取	
採集方法	種子採取	採集場所	ネパールの山岳地帯	
登録年月日	86.10.22	登録番号	004126	
登録場所	1 0435	登録量	32	4 0947
登録者	農学博士 渡辺 隆夫	登録者	農学博士 渡辺 隆夫	
登録内容	作物分二部 農学博士 渡辺 隆夫	登録内容	作物分二部 農学博士 渡辺 隆夫	
登録内容	作物分二部 農学博士 渡辺 隆夫	登録内容	作物分二部 農学博士 渡辺 隆夫	
登録内容	作物分二部 農学博士 渡辺 隆夫	登録内容	作物分二部 農学博士 渡辺 隆夫	

注) 大枠内に記入、詳細は裏面参照 農林水産省農産生物資源研究所 遺伝資源部 生産員保存管理室

図1 種子カード

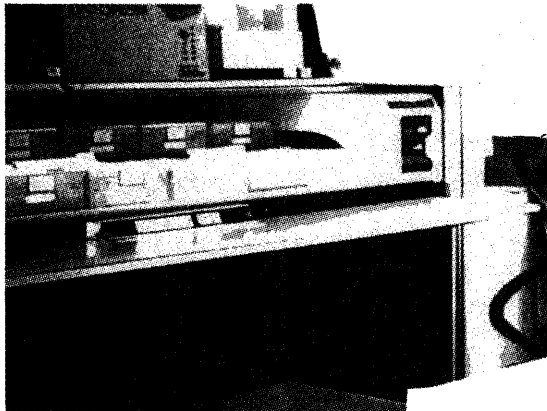


写真1 システムファイル

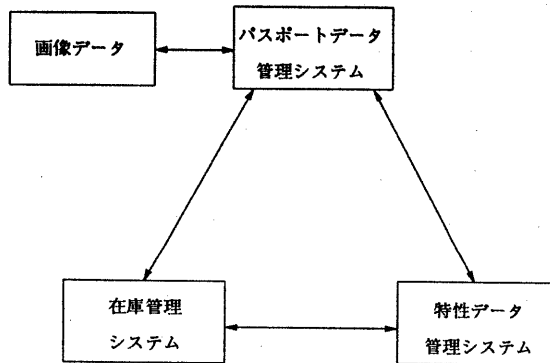


図2 植物遺伝資源管理システム

### 3.ハードウェアの構成

図3にハードウェアの構成の概要を示す。スキャナーにはHP ScanJetを使用して、HP9000/720とX端末に接続した。HP9000/720はこのシステムのアプリケーションサーバである。

スキャナーから取り込まれたデータはHP9000/720で圧縮され、データベースサーバに送られる。逆にデータベースサーバから検索されたデータは、HP9000/720で伸張され、端末の画面に表示される。

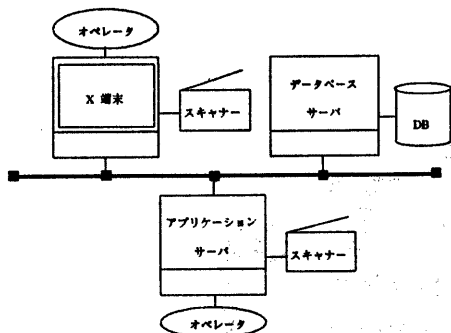


図3 ハードウェア構成

### 4.画像処理

アプリケーションはC言語によって記述し、図4の6つのモジュールから構成されている。各モジュールはデータの受け渡しを行うので、順番に実行される。データを追加する場合は、初めに種子カードをスキャナーで取り込み、次に画像の圧縮、最後にデータベースへの格納が実行される。また、データを画面に表示させる場合は、初めにデータベースからの検索を実行し、次に画像の伸張、最後に画面への表示を行う。ここでは画像処理に関する4つのモジュールの概要を示す。

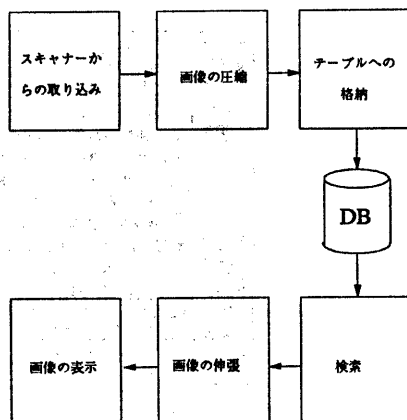


図4 アプリケーションの概要

#### 1)取り込みと表示

スキャナーの操作は、本来C言語から制御したほうがアプリケーションとしての統一性はとりやすい。しかし、本スキャナーについては、インターフェイスの仕様が未公開で、制御のためのライブラリも提供されていないため、直接C言語から制御するのは困難であった。このため、スキャナーと一緒に提供されているコマンド形式のアプリケーションをシステム関数で呼び出してスキャナーを制御する方法をとった。オーバーヘッド等を考えるとあまり望ましくないが、現状ではもっとも現実的な解決法であった。

#### 2)圧縮と伸張

1)の処理によって取り込まれた画像データは、TIFF形式である。鉛筆で細かく書かれた文字を判読できる程度に、スキャナーの解像度を設定しなければならない。しかし、仮に解像度160dpiとした場合、19×26cmサイズの種子カードは、グレイスケールで約1.4MBのファイルとなる。今回は約10万枚の種子カードの表裏の画

像をデータベースに格納しなければならないので、圧縮しなければ約280GBの容量が必要となる。本システムではJPEGによってTIFF形式のデータを圧縮している。

次に、解像度と圧縮率の値をどのように決定したのかを述べる。伸張した画像が判読可能で、かつファイルサイズを最小にする解像度と圧縮率の組合せを特定する必要がある。そこで今回は植物遺伝資源データベースの利用者8名に協力を依頼して、判読できる画像の範囲についてのアンケートを実施した。アンケートは、解像度と圧縮率を伏せた9枚の画像(同一の画像を伸張したもの)を各自に配って、評価を記入してもらう方法で行った。その結果、全員が「A」と評価した画像の仕様(解像度160dpiで取り込み、TIFF形式のファイルを6.65%の大きさに圧縮)で処理することとした。なおアンケートの結果だけでなく読み取りにくい特定の文字の結果も考慮に入れている。

## 5. スキーマ

遺伝資源の管理では、DBMSにINFORMIX-OnLineを用いている。INFORMIXでは、画像などのバイナリデータの列をテキストデータの列と総称してBLOB(Binary Large Object)と呼んでいる。BLOBのデータを効率よく扱うためには、ディスクにBLOB領域というBLOBデータ専用の領域を作成する。バイナリデータを含む表を作成する場合は、バイナリデータの列をBLOB領域に作成するように指定する。

今回、画像データを格納するために作成した表は、「seed\_card」という。この表は7つの列を持つ。以下にスキーマを示す。

```
create table seed_card
|
|
| 画像番号  serial primary key,
| 整理番号  char (8) not null
|          references passport,
| 型式      char (1) not null
|          check ( 型式 BETWEEN "1" AND "4"),
| 表裏      char (1) not null
|          check ((表裏 = "1") OR (表裏 = "2")),
| 枚目      char (1) not null,
| 登録日    date default today,
| 種子カード byte in blobdbs,
|          unique (整理番号, 型式, 表裏, 枚目)
|
```

整理番号は、パスポートデータの整理番号で、品種を特定するためのキーである。種子カードには年代により4つのタイプがあるので、列「型式」はそのタイプを入力する。列「表裏」は、取り込む種子カードが「表」であるか、または「裏」であるかを入力するための列。列「枚目」は、取り込む種子カードが同一タイプの中で何枚目であるかを入力するための列である。列「種子カード」が画像データを格納する列となる。

## 6. データ操作

データベースに数値や文字を格納するためには、SQLのINSERT文を用いる。しかしBLOBデータを処理するには、INFORMIX-4GLや、INFORMIX-ESQL/Cのような埋め込みSQLをサポートする言語でプログラムを作成しなければならない。今回はINFORMIX-ESQL/Cで記述し、ロケータ構造体を通して、BLOBデータを操作した。ロケータ構造体とは、BLOBデータのサイズと位置についての情報が納められている構造体である。

## 7.GUI

本プログラムではOSF/Motif上の簡単な操作で使えるグラフィカルなユーザインターフェースを作成した。C言語から直接Motifのライブラリを使うのは効率が悪いので、UIM/XというGUIビルダーを使用した。UIM/Xを用いるとプッシュボタンやスクロールボタンなどのウィジェットをマウスによって簡単に生成できて、その移動や大きさの変更もマウスの操作によって行うことができる。またウィジェットが実行されたときに、どのように反応するか(コールバック)をC言語で記述することができる。

本アプリケーションにおいて、画像データを表に追加する場合、オペレータはスキャナーに種子カードをセットした後、図5のメインメニューに「作物番号」と「品種番号」を入力する。次にオプションメニューから「型式」と「表裏」の該当する項目を選択し、「枚目」を記入して、最後に「追加実行」ボタンをクリックする。

図5 メインメニュー

一連の処理が終了すると、システムから次の種子カードの入力を許可するメッセージが図6のように返ってくるので、スキャナーに次の種子カードをセットして、同様の操作を行う。

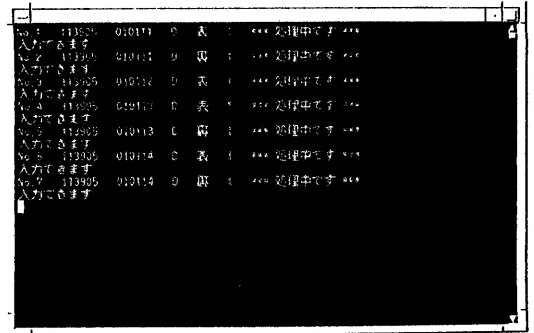


図6 メッセージ画面

画像を表示させる場合は、必要な項目をすべて入力してから「画像表示」ボタンをクリックする。画像の表示に必要な時間は12秒である。

「追加実行」と「画像表示」ボタンに設定したコールバックは、図4の上部と下部の三つのモジュールを順番に処理するためのものである。

UIM/Xでは、作成したインターフェースがC言語のソースファイルに変換され、main関数とMakefileを自動的に生成することができるので、カスタマイズが容易である。

## 8.結果と考察

平成6年12月21日現在、1463件の画像データがテーブルに格納され、容量は101MBである。1件あたりの容量は平均71KBとなる。したがって、20万件の画像データを格納するためには、約14GBの容量が見積もられる。また、作業時間については、1時間あたり平均33件の画像データが追加処理できることから、日に6時間、月に20日作業を行うとして、4年ほどの日数を要する。

解像度と圧縮率の値を決定する場合、アンケートの結果では、多くの者から高い評価をうけた画像でも、特定の文字に着目すると判読できないことがある。欠落した文字にも重要な情報が含まれていることがあるので、アンケートだけでなく、読み取りにくい特定の文字の結果も考慮しなければならない。

本システムは、過去の種子カードを処理するためのものであった。今後は、植物の画像をデータとして処理するためのシステムを開発する予定である。

## 9.参考文献

1)梅原,武田,服部(1991):農林水産ジーンバンクにおけるRDBMS,

Proceeding of Advanced Database System Symposium' 91 P37~46

2)梅原,武田,竹谷,川田,服部(1993):植物遺伝資源の特性データ管理における表作成の自動化と一括管理,情報処理学会研究報告Vol. 93, No96 P55~64

3)梅原(1987):SQL,ESQL/C,Cによるデータベースプログラミング, 第6回電子計算機利用研究発表会講演要旨集 P48~63