

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

梅原正道, 武田尚人, 竹谷勝, 川田真佐枝, 服部紳二

農林水産省  
農業生物資源研究所遺伝資源第1部

植物遺伝資源の特性データの管理は、取り扱う植物の種類が多数であること、植物ごとに評価する特性項目が異なることから、多数の表が必要となる。現在、データ管理のために運用している表の数は、600余りになる。これらの表を作成するSQL文を手作業で作成すること、また表ごとにデータ管理プログラムを作成することは、容易でない。

上記の問題を解決するため、植物遺伝資源の特性項目についてのデータ辞書を整備し、データ辞書から表作成のSQL文を生成するシステムを開発した。つづいて、データ辞書のデータを用いて、ユーザインタフェース画面を生成し、動的SQLによって全ての表のデータ管理を行うプログラムを開発した。

Automatic creation and total management of tables  
for the data management of evaluated characteristics  
of the plant genetic resources

Masamichi UMEHARA, Hisato TAKEDA, Masaru TAKEYA, Masae KAWADA, Shinji HATTORI

Department of Genetic Resources I  
National Institute of Agrobiological Resources  
Ministry of Agriculture, Forestry and Fisheries

For the data management of evaluated characteristics for the plant genetic resources, many tables are required. Plant species involved are large in number and characteristics to be evaluated are specific for each plant species. Currently, about 600 tables are in use. It requires laborious efforts to create SQL statements for 600 tables by hand and to develop 600 programs for each tables.

A data dictionary system was developed for the characteristics of the plant genetic resources to solve the problem. Further, a program was developed to generate SQL statements for creating tables. Finally, a data management program was developed to handle whole tables for the evaluated characteristics using the data dictionary and dynamic SQL.

## 1. はじめに

農業は最も古い産業の一つであり、生産の対象となる作物は、野生の植物から長い年月をかけて改良されてきたものである。稲、小麦等の作物の品種は、近代の育種技術及び育種組織が確立するまでは、農民自身の手によって絶え間なく改良が行われてきた。これらの在来品種は、数千年の年月と地理的隔離、人為淘汰の蓄積により、遺伝的な多様性を拡大しつつ、生産力を向上させてきた。しかし、近代農学の進歩に伴った育種技術の発展と育成品種の普及は、生産力を大幅に向上させた反面、長い年月をかけて成立してきた在来種の作付の激減をもたらし、生物種としての遺伝的多様性の保持を危うくしている。元々、植物遺伝資源の収集は、多様な遺伝的特性を持つ育種材料の確保を目的としていたが、育種の成果である新品種が、その元となる遺伝子プールを減少させかねないという結果を招きつつある。このような現状で、将来とも必要な遺伝子プールを確保するため、急速に栽培面積が失われつつある在来種を中心に、遺伝資源を収集し、保存することはきわめて重要な意義を持っている。

米国、旧ソ連等では、19世紀から遺伝資源の重要性に着目し、世界的に収集を行ってきており、米国では現在までに約50万点、旧ソ連では約36万点の遺伝資源を収集しているといわれている。

日本では、遺伝資源の重要性に対する認識はやや遅れ、約40年前頃から態勢整備が開始されたが、その重要性に鑑み、1984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、遺伝資源の収集、評価、保存及び配布を行っている。

## 2. ジーンバンク事業の植物遺伝資源部門の概要

農林水産省が実施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、分類・評価、増殖、保存及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、毎年海外に4隊の探索隊を送るほか、国内の未収集の遺伝資源の探索収集も行っている。その他に海外の研究機関との遺伝資源の交換も行われており、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。

分類評価は、収集した遺伝資源を分類し（種の同定）、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だけの量がないので、まず増殖を行い、その後保存を行う。保存には、永久保存と、配布用保存の2種類がある。永久保存は、その名の通り原則として永久に保存するもので、種子の場合、 $-10^{\circ}\text{C}$ 、湿度30パーセントの環境下で保存する。配布用の種子は、 $-1^{\circ}\text{C}$ 、湿度30パーセント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。配布によって在庫が僅少になった遺伝資源については、増殖を行う。センターバンクである農業生物資源研究所では、種子で保存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については、保存施設のある全国のサブバンクに分散し、保存している。

このジーンバンク事業により、現在では約18万の遺伝資源が保存され、育種、栽培、生理等の研究分野で利用されている。

## 3. 植物遺伝資源データベースの概要

ジーンバンク事業を推進するため、リレーショナルデータベースを用いた遺伝資源データベースを構築し、利用している。

データベーススキーマは、次の三つの主要な部分から構成される。

(1) 個々の遺伝資源の戸籍に相当する「パスポートデータ」に関する部分。

パスポートデータは、遺伝資源の分野では古くから使われている用語で、遺伝資源の名前、来歴、所在等のデータをあらわし、個々の遺伝資源を識別する目的で使用される。遺伝資源は収集された段階で、原産地、名前、植物の種（後で同定する場合もある）等のパスポートデータが作成される。遺伝資源は、このパスポートデータが遺伝資源データベースに登録されることによって、固有の整理番号（accession number）が与えられ、遺伝資源として登録されたことになる。この部分は、パスポートデータ表を中心に約10個の表から構成される。

(2) 遺伝資源の増殖、保存、配布等、遺伝資源の管理者に必要な「在庫管理」に関する部分。

「在庫管理」関係は、種子庫中のアドレスの管理表、配布申請を記録する表等約20個の表から構成される。この部分は、種子の配布や増殖等、ジーンバンク事業の実施の上からきわめて重要なものである。

(3) 遺伝資源の利用者が最も関心を持つ、遺伝資源の「特性評価データ」に関する部分。

「特性評価データ」は、本論文の主題であり、次に詳しく述べる。

データの性格として、パスポートデータと在庫管理データは、ビジネスデータであるのに対し、特性評価データは科学技術データであるという特徴を持ったデータベースである。またデータベースの利用者の立場からみると、遺伝資源の管理者に重要なものは、パスポートデータと在庫管理データであり、一般利用者に重要なものは、パスポートデータと特性評価データである。

#### 4. 特性評価データ

「特性評価データ」は、育種材料またはその他の実験に使用する遺伝資源を選択するために、必要なものである。育種材料としては、病気に強い形質、蛋白含量が高い形質、また逆に蛋白含量が低い形質等、育種目的である形質の遺伝子を持っている遺伝資源が必要である。このような遺伝資源として評価することが必要な形質を、遺伝資源の特性という。

この遺伝資源の特性は、植物の種ごとに必要とされるものが異なる。例えば、「稲」に必要な特性と、「きく」に必要な特性とが異なることは、明らかであろう。現在、約100種の植物について、遺伝資源として評価が必要な特性を定めている。また、実際の特性評価は、一の試験ですべての特性を評価することはできないので、形態的特性を評価する1次特性評価、生理的特性を評価する2次特性評価、経済的特性を評価する3次特性評価に分類し、実施している。また1次から3次特性のそれぞれに、必須調査項目と、選択調査項目を設けているので、データベース化にあたっては、植物ごとに6つの表が必要になる。

このような理由で、「特性評価データ」の管理には非常に多くの表を必要とし、現在作成している表の数は、600余りになる。そして、この数は将来、対象植物の増加とともに増加することが必至である。

植物遺伝資源データベースの開発は、1987年に開始し、データ管理システムとしては、1992

年によやく一応の完成をみたものである。このように長期間を要したのは、特性評価部分の開発になかなか着手できなかったからである。それ以外の「パスポートデータ」及び「在庫管理」部分は、基本的に1年程度でデータ管理システムの開発を行い、運用を行いながら改良を図ってきた。

この原因は、統一的な特性評価の基準を作成することが、かなり困難な作業であったからである。特性評価を、全国の試験研究機関で統一した基準で実施することは、きわめて重要である。各機関ごとに独自の基準で調査を行うとすると、現在でも600も必要としている表の数が、いくつ必要になるか見当もつかなくなる。更に、稲だけで異なる様式の表が多数できると、検索の手間も大変になる。

この基準の作成作業はかなり難航し、1986年に1次特性の調査マニュアルの未定稿を作成後、一時作業が中断していた。センターバンクでは、この作業を、1989年より、各植物の専門家の協力を得て、再開した。そして1990年に「植物遺伝資源特性調査マニュアル」（未定稿）として刊行し、2年間の試行を経て、1992年に定稿として刊行した。これによって、「特性評価データ」の管理システムの開発を可能とする環境が整備された。

## 5. データ辞書の開発と特性調査マニュアルの作成

上記の「植物遺伝資源特性調査マニュアル」の策定作業を開始するにあたって、印刷に必要なデータをすべて「特性評価データ」のデータ辞書として整備する方針を立て、まずそのためのシステム開発を行った。データ辞書は、下記に示す表から構成される。各表は第3正規形である。

### ①最も主要な表として、「特性項目」

すべての植物の特性評価項目についての属性を収めるもので、植物を表すコードとして特性植物番号、特性項目の順番として項目番号があり、この組がプライマリーキーとなる。その他、特性評価項目の名称、調査基準に係わる諸項目がある。ここでデータ種別が階級値とされるものは、次の表「区分」によってその内容が定義される。

### ②特性評価結果の階級値を定義する「区分」

特性評価の結果を、0から9までの階級値で評価する項目について、階級値と評価値との対応付けの表である。具体的には、抵抗性等の場合

2	3	4	5	6	7	8
極強	強	やや強	中	やや弱	弱	極弱

のように、階級値と評価結果の対応付けが定義されている。

### ③参照用の外部表

特性植物番号等のコードとの対応をとるための表

このデータ辞書のデータ管理システムを1990年に開発し、表「特性項目」には、約6,000件、表「区分」には、約25,000件のデータを入力した。

次に、このデータ辞書を用いて、「植物遺伝資源特性調査マニュアル」を作成するプログラムを開発した。この出力を用いて、データ辞書の校正を行った。このマニュアルは5分冊、延べ700ページのものであるが、図1にその1ページを示す。

作物の種類		稲		01001		1次必須項目											
項目番号	項目名	調査数	方法	分級・単位										調査方法等			
				0	1	2	3	4	5	6	7	8	9				
1	稈長	5個体	測定	cm (小数第1位を四捨五入)										生育が中熟の個体の最長稈の地際から穂首までの長さ			
2	穂長	5個体	測定	cm (小数第2位を四捨五入)										生育が中熟の個体の最長稈の穂首から穂の先端(芒を除く)までの長さ			
3	穂数	5個体	測定	本/個体 (小数第2位を四捨五入)										遅れ穂を除く穂の数			
4	ふ先色	区	観察	黄白 ~ 黄	黄褐	褐	赤褐	淡赤	赤	淡紫	紫	黒紫	黒紫	成熟期(出穂後3週間目)におけるふ先の色			
5	穂長	5粒	測定	mm (小数第2位を四捨五入)										投影機又はダイヤアルゲージ等により測定した穂の長さ			
6	穂幅	5粒	測定	mm (小数第2位を四捨五入)										投影機又はダイヤアルゲージ等により測定した穂の幅			
7	玄米長	5粒	測定	mm (小数第2位を四捨五入)										投影機又はダイヤアルゲージ等により測定した玄米の長さ			
8	玄米幅	5粒	測定	mm (小数第2位を四捨五入)										投影機又はダイヤアルゲージ等により測定した玄米の幅			
9	うるち(腹)・もち(精)の別	5個体	観察	腹										直接又はヨード・ヨードカリ反応による玄米の観察			
10	出穂期	区	観察	月 日										全個体の50%が出穂を始めた日(「出穂」とは穂の一部分(芒を除く)が葉鞘から現れたことをいう)			
11	顕色	区	観察	黄白	黄	黄金	赤黄	褐	赤褐	褐	紫	黒紫	その他	成熟期(出穂後3週間目)における顕色			
12	芒の有無と多少	区	観察	無	稀	極少	少	中	やや多	多	極多	甚		少:10%, 中:25%, 多:40%			
13	芒長	10粒	測定	極短 短 中 中 長 極長										有芒穂について芒の長い方から30%の穂について測定した平均。短:2cm, 中:4cm, 長:6cm			

図1 植物遺伝資源特性調査マニュアルの例

データ辞書として、作成したスキーマを次に示す。なお、このスキーマは、使用しているDBMSがデータ制約定義を実現してから改訂したもので、第2版になる。

create table 特性項目

```
(
  特性植物番号      char(5) not null references 特性植物,
  項目番号          smallint not null,
  回数              char(1) not null check (回数 between "1" and "3"),
  必須項目          char(1) not null check (必須項目 between "0" and "1"),
  項目名            char(50) not null,
  調査数            char(20),
  測定法コード      char(1) not null references 測定法,
  データ種別コード char(1) not null references データ種別,
  単位              char(20),
  全桁数            smallint,
  小数桁数          smallint,
  調査方法          char(200),
  primary key (特性植物番号,項目番号)
);
```

create table 区分

```
(
  特性植物番号      char(5) not null,
  項目番号          smallint not null,
  区分番号          smallint not null check(区分番号 0 and 10),
  区分名            char(20) not null,
  primary key (特性植物番号,項目番号,区分番号),
  foreign key (特性植物番号, 項目番号) references 特性項目
);
```

## 6. 表の生成

特性評価データのデータベース化に必要な約600の表を、人手でSQL文を入力して作成することはかなり大変な作業で、ミスも多いと考えられる。そのため、データ辞書から表を作成するSQL文を生成するプログラムを作成し、表の作成を自動化した。生成されるSQL文の一部を以下に示す。

```
CREATE TABLE t01001h1 (
  整理番号  CHAR(8) NOT NULL REFERENCES passport(整理番号),
  機関コード CHAR(5) NOT NULL,
  部室コード CHAR(5) NOT NULL,
  年度      DATETIME YEAR TO YEAR NOT NULL CHECK (年度 > 1970),
```

```

h1_001 INTEGER CHECK(h1_001 > 0 AND h1_001 < 999 ),
h1_002 DECIMAL(4,1) CHECK(h1_002 > 0 AND h1_002 < 99.9 ),
h1_003 DECIMAL(4,1) CHECK(h1_003 > 0 AND h1_003 < 99.9 ),
h1_004 CHAR(1) CHECK(h1_004 BETWEEN "1" and "9" ),
h1_005 DECIMAL(5,1) CHECK(h1_005 > 0 AND h1_005 < 999.9 ),
h1_006 DECIMAL(4,1) CHECK(h1_006 > 0 AND h1_006 < 99.9 ),
h1_007 DECIMAL(5,1) CHECK(h1_007 > 0 AND h1_007 < 999.9 ),
h1_008 DECIMAL(4,1) CHECK(h1_008 > 0 AND h1_008 < 99.9 ),
h1_009 CHAR(1) CHECK(h1_009 IN("2","8" )),
h1_010 CHAR(5),
h1_011 CHAR(1) CHECK(h1_011 BETWEEN "1" and "9" ),
h1_012 CHAR(1) CHECK(h1_012 BETWEEN "0" and "9" ),
h1_013 CHAR(1) CHECK(h1_013 BETWEEN "2" and "8" ),
PRIMARY KEY(整理番号,機関コード,部室コード,年度),
FOREIGN KEY(機関コード,部室コード)
REFERENCES bushitu(機関コード,部室コード) );

```

生成される表は、特性植物番号と試験の区分による表名が付けられるので、特性植物番号表を検索することによって、どの植物を対象とした表であるかを知ることができる。生成規則は、

t+特性植物番号+(h|s)+(1|2|3)

である。稲の場合は、

t01001h1, t01001h2, t01001h3, t01001s1, t01001s2, t01001s3

の6つの表を生成する。また表の構造は、同一属性のプライマリーキーを持つ他は、すべてデータ辞書から機械的に生成した属性を持つ。なお、データ辞書から定義可能なデータ制約については定義を行っている。生成した属性名は、一見無意味なものであるが、データ辞書を検索することによって、その定義を知ることができる。

## 7. 表の一括管理

生成した表のデータ管理のためには、現在のDBMSを使う限り、通常なんらかのアプリケーションが必要である。簡易なフォームを使用したとしても、600のアプリケーションを作成、管理することは容易ではない。そのため、データ辞書を用いて、すべての表を一括管理するプログラムを開発した。

このプログラムは、管理対象とする表を3段階の階層メニューで選択する。最初のメニューでは、図2に示す、麦類、野菜、果樹等の植物種類を選択する。次に、図3に示すように、その植物種類に属する植物を検索して表示し、個々の植物の選択を行う。最後にその植物の6つの個別表を表示し、データ管理の対象となる表を決定する。(図4)

この後、プログラムはデータ辞書から該当する表に関するデータを検索する。このデータを用いて、まずオペレータとのインタフェース画面(図5)を生成する。この画面では、無意味な属性名の代わりに、データ辞書から検索した特性項目が表示される。続いて、入出力用のバッファをメモリ上に確保し、オペレータの入力を待つ。この後は、オペレータの入力に従い、動的SQLを生成して、検索、挿入、

特性データ管理システム

- ← 1 : 01 稲類
- |         |          |
|---------|----------|
| 2 : 02  | 麦類       |
| 3 : 03  | 豆類       |
| 4 : 04  | いも類      |
| 5 : 05  | 雑穀・特用作物  |
| 6 : 06  | 牧草・飼料作物  |
| 7 : 07  | 果樹       |
| 8 : 08  | 野菜       |
| 9 : 09  | 花き・緑化植物  |
| 10 : 10 | 茶        |
| 11 : 11 | 桑        |
| 12 : 12 | 熱帯・亜熱帯作物 |
| 13 : 13 | その他の植物   |

選択して下さい ( 終了 E )

図2 植物種類選択メニュー

特性データ管理システム

- |                  |                  |
|------------------|------------------|
| 1 : 07001 りんご    | 11 : 07011 にほんなし |
| 2 : 07002 せいやなし  | 12 : 07012 もも    |
| 3 : 07003 マルメロ   | 13 : 07013 すもも   |
| 4 : 07004 おうとう   | 14 : 07014 あんず   |
| 5 : 07005 きいちご   | 15 : 07015 うめ    |
| 6 : 07006 カーランツ  | 16 : 07016 ぶどう   |
| 7 : 07007 グーズベリー | 17 : 07017 かき    |
| 8 : 07008 ブルーベリー | 18 : 07018 いちじく  |
| 9 : 07009 くるみ    | 19 : 07019 かんきつ  |
| 10 : 07010 くり    | 20 : 07020 やまもも  |

選択して下さい ( 終了 E )

図3 植物選択メニュー



特性データ管理システム

- 1 : t07001h1 りんご 1次必須 307 件
- 2 : t07001h2 りんご 2次必須 43 件
- 3 : t07001h3 りんご 3次必須 41 件
- 4 : t07001s1 りんご 1次選択 264 件
- 5 : t07001s2 りんご 2次選択 0 件
- 6 : t07001s3 りんご 3次選択 0 件

選択して下さい ( 終了 E )

図4 表選択メニュー

特性:  検索 U/更新 N/次 P/前 F/最初 L/最後 J/指定 A/追加 R/削除 S/画面 E/終了  
データを検索します 1/1 面 最表示/CTRL+L

検索数/ 84件 現在/ 1番目

りんご 1次必須

整理番号  EIKAN  
機関コード  果樹試 盛岡 育研 年度

枝条の色	<input type="text" value="3"/> 褐	葉身の大きさ	<input type="text" value="32"/>
葉身の形	<input type="text" value="164"/>	葉縁の鋸齒	<input type="text" value="1"/> 鈍鋸齒
成葉の毛茸	<input type="text" value="7"/> 多	托葉の形	<input type="text" value="4"/> 鎌形
花の大きさ	<input type="text" value="56"/>	果実の大きさ	<input type="text" value="210"/>
果実の形	<input type="text" value="1"/> 円	果実の色	<input type="text" value="1"/> 赤

検索終了

図5 インタフェース画面

更新を該当表に対して行う。

上記のメニューは階層構造を自由に上下に移動可能で、稲の1次必須特性の検索を行った後、大豆の2次必須特性を検索することも可能である。

## 8. おわりに

この開発の結果は以下の2点に要約される。

- ①「植物遺伝資源特性調査マニュアル」を刊行し、統一的なデータ生産を可能にしたこと。
- ②植物遺伝資源の特性評価データの表の作成と管理をルーチン化したこと。

その結果、約1年間の運用で10万件ほどのデータ入力を行うことができた。現在の特性調査の計画によれば、10年後には60万件程度のデータが蓄積されることになり、植物遺伝資源の利用に大いに資するものと期待される。

システム開発としてみると、このデータを利用するアプリケーションは別として、データ管理の目的で開発したプログラムは、基本的に

- ①データ辞書管理プログラム
- ②マニュアル印刷プログラム
- ③SQL文生成プログラム
- ④特性評価データ表管理プログラム

の4本であり、規模としては、総計2万ステップに満たないものである。このシステムで600の表の管理を可能としており、きわめて効率が高いと評価している。これは、植物遺伝資源の特性評価データという、同質的な表構造を対象としたデータ辞書であったため、データ辞書の構造が単純化できたことが寄与している。

また、このシステムについて最も強調したい点は、新しい植物の特性評価データをデータベース化するためには、

- ①データ辞書にその植物の特性の定義データを追加し、
- ②SQL文の生成プログラムによって、表を作成する。

という作業を行えば、データ管理プログラムが、追加して作成した表を認識し、データ管理が可能になる点である。すなわち、どれだけ植物を追加しても、システム開発は一切不要で、関連するデータを定義する作業だけでよいということである。

残されている問題は、現在使用している動的SQLは、SQL2で規格化される以前のDBMSのインプリメントに依存したもので、今後SQL2規格にあわせていく予定である。

【謝辞】本稿の投稿を勧めいただき、貴重なアドバイスをいただいた、図書館情報大学教授増永良文氏に深謝いたします。

## 参考文献

梅原、武田、服部、「農林水産ゾーンバンクにおけるRDBMS」、  
アドバンスト・データベースシステム・シンポジウム(1991)

松尾、「植物遺伝資源集成」、講談社(1989)

池橋・梅原、「遺伝資源と種苗産業」、計測と制御、Vol. 26, No. 9