

## テンサイ育種におけるデータベースの利用(1)

誌名	てん菜研究会報 = Proceedings of the Sugar Beet Research Association
ISSN	09121048
著者	川勝, 正夫 安間, 舜
巻/号	28号
掲載ページ	p. 44-50
発行年月	1987年9月

テンサイ育種におけるデータベースの利用

第1報 GRIMS/CGSを用いたデータベースの構築

川勝正夫・安間 舜

(北海道農業試験場)

1. 緒 言

筆者らは農林水産省の農林水産研究計算センターの大型計算機を用い、総合情報管理プログラムであるGRIMS/CGS<sup>1,2,3)</sup>によって、テンサイ育種における各種育種試験の情報をデータベース化することを計画している。

この総合情報管理プログラムの基本的な機能であるデータベースの構築、修正、検索はEXIR<sup>\*</sup>と呼ばれるプログラムで行っている。このEXIR<sup>\*</sup>はコロラド大学を中心として開発されたデータベース管理システムを基とし、遺伝資源はもとより、より広汎な分野に適用できるシステムとするために若干の機能追加を行ったものである。

本報告は農林水産研究計算センターの全国オンラ

インネットワークとして、北海道農試に設置されているA端末を用いて、テンサイの遺伝資源に関する1つのデータベースを構築し、それについて追加、修正などを試み、育種試験成績をデータベース化する際の問題点摘出を図ったので、その結果を報告する。

本報告に当たり、農林水産省農業生物資源研究所遺伝資源部遺伝情報管理システム研究室、関島稔室長ならびに横山忠治主任研究官の御指導をいただいたので御礼申し上げる。

2. データベースの構築

1) EXIR<sup>\*</sup>で扱えるデータ

EXIR<sup>\*</sup>で扱えるデータは第1表のような2元表の形で与えられるデータである。2元表の縦にあたる

第1表 [GRIMS/CGS] 遺伝資源情報管理プログラム (EXIR<sup>\*</sup>のデータ構造)

ディスクリプタ (記述子) →

アイテム (項目) ↓	品種 番号	品 種 名	原産国	葉 数 枚	葉 長 cm	葉 身 長 cm	葉 幅 cm	根 長 cm	根 周 cm	茎葉重 t	根 重 t	
	1	Dounyuu 2	JPN	22.5	62.6	26.1	19.0	14.6	27.8	6.67	3.78	
	2	NK-152	JPN	22.0	55.7	22.8	17.2	14.6	32.0	5.92	4.82	
	3	Monohikari	JPN	23.0	58.1	25.0	15.7	15.5	32.4	6.04	5.51	
	4	TA-15	PLD	25.7	56.8	21.3	17.7	12.8	28.2	5.43	3.39	
	5	TA-73	NLD	31.2	47.3	19.1	17.2	15.8	26.9	5.74		
	6	TA-36	GFR	28.0	56.5	21.9	17.2	13.9	32.3	5.56	4.98	
	7	TA-103	CPR	25.2	57.1	22.8	19.0	13.0	28.4		3.56	
				ディスクリプタ状態								
				中			略			未知 (UNKNOWN)		
	46	HON-IKU 400	JPN	24.2	57.2	24.8	18.6	20.1	29.3	5.13	3.97	
	47	HON-IKU 401	JPN	30.1	49.9	22.4	16.7	17.6	27.3	5.68	2.78	

注) GRIMS/CGS (Genetic Resources Information Management System for Centre of Genetic Stock の略), EXIR (EXective Information Retrieval の略)

品種や系統をEXIR<sup>\*</sup>ではアイテム (Item:項目) と呼び、アイテムを構成する共通の記述をディスクリプタ (Descriptor:記述子) と呼んでいる。個々のデータをディスクリプタ状態 (Descriptor states) または、単に状態 (states) と言う。データは数値データだけでなく、文字や記号で表現される情報も処理することができる。なお、ディスクリプタの状態での欠測は、未知 (Unknown) というデータが入

っていると解釈され、ゼロとは識別される。

## 2) データベースの規模

ここでは1983年の二倍体テンサイ遺伝資源の特性調査およびテンサイ二倍体系統の組合せ能力検定の試験成績を使用してデータベースの構築を試みた。

ディスクリプタは第2表に示した。アイテムは品種番号の1番から3番までを抜粋して例示した。ディスクリプタの中には品種番号や品種名などアイテ

第2表 テンサイ品種特性データ

品種番号	試験番号	品種名	原産国	倍数性	胚数	胚軸色	葉姿	葉色	葉形	葉面縮	葉数
1	8301	導入2号	日本	二倍体	多胚	85.3	直立	中	楕円	中	22.5
2	8302	NK-152	日本	二倍体	多胚	91.5	直立	濃緑	皮針	中	22.0
3	8303	モノヒカリ	日本	二倍体	単胚	66.7	直立	濃緑	皮針	中	23.3

  

品種番号	葉重 <sub>A</sub>	葉面積	葉柄太さ	葉長	葉身長	葉柄長	葉幅	葉身長/葉長	葉幅/葉身長	根形	分岐根
1	50.0	411.9	1.63	62.6	26.1	36.5	19.0	0.42	0.73	円錐	中
2	38.5	322.0	1.42	55.7	22.8	32.9	17.2	0.41	0.75	円錐	少
3	36.0	313.4	1.38	58.1	25.0	33.1	15.7	0.43	0.63	円錐	少

  

品種番号	露肩	根長	根周	根周/根長	全重	茎葉重	葉重 <sub>B</sub>	茎重	根茎重	根重	葉重/全重
1	中	14.6	27.8	1.90	10.45	6.67	5.73	0.94	4.72	3.78	0.55
2	中	14.6	32.0	2.19	10.74	5.92	5.28	0.64	5.46	4.82	0.49
3	中	15.5	32.4	2.09	11.15	6.04	5.36	0.68	6.19	5.51	0.46

  

品種番号	茎重/全重	根重/全重	茎葉重/根重	根中糖分	糖量	N	Na	K	不純物価	可製糖分	可製糖量
1	0.09	0.36	1.76	16.46	623	196	198	1651	412	14.29	540
2	0.06	0.45	1.23	18.51	889	180	142	1312	301	16.74	807
3	0.06	0.48	1.10	17.65	972	121	157	1185	268	16.01	882

注) 品種番号1~3の部分抜粋したもの。

ムの識別のみに用いるものから、葉色、葉形などの観察値、葉重、根重などの測定値、根重/茎葉重のような計算値など種々の特性データまで、全部で45がこのデータベースでは定義されている。ここでは品種・系統すなわちアイテムは47である。このGRIMS/CGSではアイテムが37以上なくては構築できないことになっている。この例のデータベースではアイテムが47、ディスクリプタが45ということになる。

### 3) データベース格納ファイルの作成

次に2元表のデータをデータベース化する場合の構築の入力様式を説明する。なお、構築時のデータ入力はカードで行うことを前提とした。

農林水産省の職員が計算センターを利用しようと

するとき、計算センターの利用申請書を提出すると利用者番号、パスワードとともに個人ファイルの最大値として390L リンクス (0.5MB) のファイルスペースが登録される。これはカード約5900枚に相当するスペースといわれている。

GRIMS/CGSによってデータベースの構築に入る前に、そのデータベースを格納するためのファイル (格納しておく場所) を作成する必要がある。第3表のように各利用者のファイル領域範囲内 (390L リンクス) で、固有の名前を付けてこのデータベース格納のためにファイルスペースを確保する。

データベースを格納するためファイルの修飾名は利用者番号/ファイル名という形をとる。構築後端末からTSSを用いて検索をする場合を想定すれば

第3表 データベース格納ファイルの作成

カード の順番	1	8	16 (カラム)			
①	\$	JOB	<u>I0300131</u>	\$	<u>PASSWORD</u>	, <u>KAWAKATSU</u> , <u>A</u>
			利用者番号		パスワード	識別名 処理クラス
②	\$	FILSYS				
③	\$	CF	<u>I0300131</u>	/	<u>KOCHIKU</u>	, <u>LLINKS/1,13</u> / <u>R</u>
			利用者番号		ファイル名	ファイル容量の 読み取り許可 (8文字以内) 大きさの指定
④	\$	ENDJOB				
⑤	\$\$\$	EOS				

注) 下線部分は各利用者が指定する。

第4表 データベース構築の入力様式

カード の順号	1	.8	16 (カラム)			
①	\$	JOB	<u>I0300131</u>	\$	<u>PASSWORD</u>	, <u>KAWAKATSU</u> , <u>B</u>
			利用者番号		パスワード	識別名 処理クラス
②	\$	SELECT	GRIMS/SEL/F1 (データベース構築のセレクトファイル名)			
③	\$	PRMFL	04,W, S, <u>I0300131/KOCHIKU</u>			
			データベース格納ファイル修飾名			
④	\$	DATA	I*			
⑤		BATCH ON *	(バッチ/TSSの切替え) バッチ処理の宣言			
			注) 下線部分は各利用者が指定する。			
⑥		DEFINE DESCRIPORS	<u>45</u>			(ディスクリプタの定義)
						(定義するディスクリプタの総数)
⑦		VARIETY NO.	(1,ORDER FROM 1 TO 50 BY 1),			
			中			略
⑧		ORIGIN	(4,NAME, 5),			
⑨		POLYPLOIDY	(5,CODEZ, DIPLOID (DI),			
			TRIPLOID (TR),			
			TETRAPLOID(TE),			
			ANISOPLOID(AN)),			
			中			略
⑩		SUGAR CONTENT	(38,ORDER FROM 15.50 TO 20.50 BY 0.01 IN %),			
			省			略
⑪		DEFINE[AND PRINT]ITEMS FROM CARDS FIXED	(アイテム定義)			
⑫		1-2,4-7,9-19,21-23,25-26,,	省略,			
⑬		2/4-2/7,2/9-2/12,,	省略			
⑭		3/4-3/7,3/9-3/12,省略,3/57-3/59 *	} データ書式の指定			
⑮		1 8301 DOUNYUU 2 JPN DI MU 85.3 ER MO EL I	省略			
⑯		1 36.5 19.0 0.42 0.73 CO I L 14.6 27.8	省略			
⑰		1 0.55 0.09 0.36	省略	540		} データ
			省			略
⑱		END OF ITEMS	(データ入力の終了)			
		FORMER NO. OF ITEMS IN THE DATA BANK=	0			
		NO. OF ITEMS ADDED TO THE BANK	=	47		} EXIR*からのメッセージ
		CURRENT NO. OF ITEMS IN THE BANK	=	47		

⑱ CONTROL VOCABULARY\* (データベース定義内容の出力)

省 略

4. ORIGIN

NO. OF CHARACTERS IN LONGEST STATE = 3  
 OPTION = NAME NO. OF STATES = 5  
 NO. OF DELETED STATES = 0  
 NO. OF DICTIONARY ENTRIES RESERVED = 5

CPR

CFR

JPN

NLD

PLD

5. PLOIDY

NO. OF CHARACTERS IN LONGEST STATE = 10  
 OPTION = CODEZ NO. OF STATES = 4

- 1. DI - DIPLOID
- 2. TR - TRIPLOID
- 3. TE - TETRAPLOID
- 4. AN - ANISOPLOID

38. SUGER PERCENT

NO. OF CHARACTERS IN LONGEST STATE = 5  
 OPTION = ORDER NO. OF STATES = 501

FROM 15.50

TO 20.50

BY 0.01

IN %

省 略

⑳ WRITE DATA BANK\* (データベースファイルの作成)

DATA BANK WRITTEN TO LOGICAL UNIT 4-DATE 86/ 5/15

THE SIZE OF DATA BANK IS MEASURED BY THE FOLLOWING CRITICAL VALUES.

NO. OF DESCRIPTORS (NODESC) = 45  
 NO. OF NUMERIC DESCRIPTORS (FTBI) = 34  
 NO. OF DESCRIPTOR DICTIONARY ENTRIES (NODDE) = 45  
 NO. OF DESCRIPTOR-STATE SUBDICTIONARY RESERVATIONS (NODSSE) = 80  
 NO. OF CODEZ DESCRIPTOR-STATE SUBDICTIONARY ENTRIES (NCOSSE) = 19  
 NO. OF OVERFLOW ENTRIES (OVX) = 195  
 NO. OF BITS IN AN ITEM DEFINITION (NOIF) = 298  
 NO. OF BUFFERS OF IFILES (IBUF) = 2  
 NO. OF ITEMS IN THE LAST BLOCK OF THE LAST BUFFERLOAD (LBAIF) = 11  
 NUMBER OF ACTIVE ITEMS IN THE DATA BANK = 47  
 TOTAL NUMBER OF ITEMS IN THE DATA BANK = 47

㉑ END\* (プログラムの終了を宣言する)

TOTAL RUN TIME IN SECONDS

CENTRAL PROCESSOR = 1.827 (CPU時間)

POOL SIZE 20000 MAXIMUM USED 2039 SIZE AT END 2037 (POOL SIZE情報)

㉒ \$ ENDCOPY

㉓ \$ ENDJOB } JCL

㉔ \$\$\$EOS

ファイル名は8文字以内とする方がよい。

4) データベース構築の入力様式

第4表はデータベース構築のためのJCL (ジョブ制御文)カードとそのカード入力の順番を示したものである。

②のGRIMS/SEL/F1のカードはデータベースを自動的に構築するためのコントロールカードで

③、④とともに構築のためのJCLである。

⑥から⑩はデータを入力するためのディスクリプタの名前や順序、オプションを定義するためのものである。

ディスクリプタにそれぞれ固有の名前を付け、入力順序を決める(後述(第5表))。

⑪のDEFINE ITEMSコマンドでディスクリプ

タに対応するデータの入力様式を指定する。⑫から⑰でデータの書式を指定し、ディスクリプタの定義に用いたオプションに従って入力する。データカードの最後に⑱のEND OF ITEMS カードを入れれば、これで構築が完了する。

⑲のCONTROL VOCABULARY コマンドはEXIR\*によって作られたデータベースのディスクリプタの状態に関する概要が出力される。

⑳のWRITE DATA BANKコマンドを入れると構築されたデータベースをファイルに格納することを指示する。実行されると収録したデータベースの概要が出力される。

㉑のEND コマンドでEXIR\*プログラムの終了を宣言する。

以上で構築は完了する。

第5表はデータや情報を効率よく格納するためにディスクリプタの内容の性格を区分し、それぞれに最も適するコマンドオプションを選択しなければならないので、その対応を示した。普通、一定間隔をもつ数字データはORDER FROM TOを用い、文字、記号、数詞等はディスクリプタの状況、データベースの利用目的に応じてTEXT, NAME, CODE, CODEZ等を用いる。

第5表 データの性格とオプションの対応

1. 数値データ

葉数 葉長等 一定間隔を持つ数値データの場合：ORDER FROM...TO

(ORDER FROM  a  TO  b  BY  c  IN  単位 )  
下限値 上限値 増分 6文字以内

LEAF LENGTH(1, ORDER FROM 47.0 TO 65.0 BY 0.1 IN CM) ← 情報提供の場合

LEAF LENGTH(2, ORDER FROM 47.0 TO 65.0 BY 0.1) ← データ解析の場合

2. 文字、記号、数詞等で表されたデータ

1) もっぱら文字綴だけで扱われ、データの内容が構築時にわかっていない場合

a. データの内容も数も予測できない場合(文章的なもの)：TEXT

b. データの内容は予測できないが、その数の限度数はわかっている場合：NAME

2) データの内容がすべて構築時に明らかになっていて、コード化できたり、

固有あるいは指定した順序がある場合

完全な状態名	CODE	CODEZ	ORDER
DIPLOID	1	DI	DIPLOID
TRIPLOID	2	TR	TRIPLOID
TETRAPLOID	3	TE	TETRAPLOID
ANISOPLOID	4	AN	ANISOPLOID

a. 入力されるデータの内容が整数型のコードで表される場合：CODE

(1, CODE, DIPLOID, TRIPLOID, TETRAPLOID, ANISOPLOID)

b. ディスクリプタ状態が省略形で表されている場合：CODEZ

(2, CODEZ, DIPLOID(DI), TRIPLOID(TR), TETRAPLOID(TE), ANISOPLOID(AN))

c. ディスクリプタ状態が完全な状態名で表されている場合：ORDER

(3, ORDER, DIPLOID, TRIPLOID, TETRAPLOID, ANISOPLOID)

3. データベース構築後のアイテム及びディスクリプタの追加、修正

構築済みのデータベースに新しいアイテムやディ

スクリプタを追加する場合は第6表のAからCのどの部分で行うかで方法が異なり、同表に準じて行えばよい。但し、この追加修正は簡単なデータの追加

第6表 データベース構築後のアイテム及びディスクリプタの追加と修正

アイテム	ディスクリプタ →
	(旧データベース)
	DEFINE MORE DESCRIPTORS DEFINE ITEMS FROM CARDS FIXED COMPUTE (C)
	CORRECTION
	1. データが構築時と同じ書式の場合 DEFINE ITEMS FROM CARD SAME 2. データの順序がディスクリプタの順序と一致しない場合。 1) カラム位置の指定で操作 2) RESEQUENCE DESCRIPTORS コマンドで順序の変更。 ORIGINAL DESCRIPTORS コマンドで元に戻す。 (A) (A+B) (B)

修正であれば問題はないが、ディスクリプタの追加、修正などはかなり手間がかかり、新しいデータベースを作り出すのと大差ない。そのためデータベース構築は最初からディスクリプタの追加、修正などがないように考えておく必要がある。

#### 4. 摘 要

テンサイ育種試験から得られた情報を、GRIMS/CGSを用いて小さなデータベースを構築し、本格的なデータベースを構築する場合の問題点を摘出した。

- 1) 作成されたデータベースへのディスクリプタの追加や修正は極めて煩雑であること、大きな広範囲にわたるデータベースより、専門的な中、小型のデータベースのほうがより使い易いことがわかった。
- 2) ディスクリプタ名やディスクリプタ状態はデータベースの利用法によって変える必要がある。例えば、情報提供の場合は長くてもわかりやすい表示

が望ましく、データ解析の場合は簡単な表現が推奨できる。

データベースはテンサイ育種において遺伝資源の保存や育種計画の作成のために効果的に用いることができると思われる。

#### 引用文献

- 1) 農林水産技術会議事務局(1982) : GRIMS/CGS—遺伝資源情報管理プログラム—。農林水産研究計算センター利用者マニュアル14 : 1~130.
- 2) 清水矩宏(1982) : データベースシステム GRIMS/CGS を利用したデータ処理法。農業技術 37 : 496~499.
- 3) 草地試験場(1982) : 草地飼料作研究におけるデータベースシステム [GRIMS/CGS] を利用したデータ管理と解析方法。草地試験場資料 No 56-11 : 1~63.

# Utilization of Database on Sugarbeet Breeding

## 1. Construction of database using GRIMS/CGS

Masao KAWAKATSU and Shun AMMA

*Hokkaido National Agricultural Experiment Station, Sapporo 004*

### Summary

Information obtained from the past sugarbeet breeding was incorporated into a limited database by using GRIMS/CGS, and it was tested if any problems should be found in re-constructing it into a full scale database.

It was found to be highly complicated to make additions and/or corrections of descriptors to the constructed database. For such purpose, it was easier to use a database of a medium- or small-sized exclusive type than that of a large and extensive one.

Descriptor name and state should be chosen according to the utilization purpose of the database; for instance, in the case of offering information, a long and descriptive title is preferable, while in the case of data analysis, a simple expression is recommended.

A database will be very instrumental in sugarbeet breeding, concerning the germplasm preservation and designing of breeding procedures.