Present Status of Genetic Resource Information Management in Japan

By SHIGERU SUZUKI and KINEO KUMAGAI

Department of Physiology and Genetics, National Institute of Agricultural Sciences (Yatabe, Ibaraki, 305 Japan)

Genetic resources are the reservoir of genetic information accumulated for generations starting from outbreak of genetic diversity by various reasons and having survived against acute or chronic changes of evolutionary pressure at various localities. They are indeed a precious heritage through history on the earth, and of immeasurable value for the future use.

Rapid progress of farming practices especially the spread of newly bred high-yielding varieties can easily cause 'genetic erosion' or loss of diversity of genetic resources scattered and adapted to each restricted *niche* for long years. Unless a special care is taken, this may result in serious losses which can never be recovered again.

Importance of genetic resources in plant breeding is well accepted, and national as well as international plans for collection, preservation or utilization are going on at present. For the efficient use of genetic resources, proper management of their information is inevitable. In searching for breeding materials, breeders are usually concerned with characteristics they want to incorporate into their new varieties. Retrieval of genetic resources by combination of several defined character values is very important for their efficient utilization. Storage of genetic resources without characteristic information is in fact insufficient for the future use. Easy retrieval, updating and analysis of their information guarantee the practical value of resource conservation. Management of information is important primarily in finding out appropriate

materials from *index seminum*. Retrieval of strain name satisfying various conditions of characteristics is the basic requirement, but listing of other character values than those used for retrieval, or even statistical analysis of related character values or their mode of association in retrieved groups corresponding to the difference in some of the key characters is also of growing interest^{5,11}.

Process of plant breeding can be interpreted as follows: beginning with the examination of genetic resource information in order to pick up breeding materials, evaluating them by field observation data, selecting limited number of them for crossing of some sort, raising their progeny and selecting by some elaborate criterion through generations, and, finally selected lines are evaluated under a proposed range of environment. Thus, starting from genetic resources, breeders manipulate gametes and/or zygotes based on the result of processing information they have collected, in order to recombine the resources into a more desirable form ready for the use and record its characteristics as the information of newly established genetic resources. In all these processes of plant breeding, efficient management of genetic resource information is ultimately important^{2,9)}. An outline of present status concerning management of genetic resource information in this country with regard to the organization and function, mainly of recent trend combined with computer use, will be presented.

Information of genetic resources and computerization

Information of genetic resources has been utilized in plant breeding for long years. In old times it was perhaps compiled in a notebook of the breeder only for his personal use. Recently systematic collections or introductions of genetic resources concentrated on special crop species are often carried out at research stations or universities and the nomenclature information with or without the results of primary evalution have been compiled in books. They contain very important information for plant breeding, and therefore have been used quite intimately by breeders. They can be used efficiently when the number of included strains and evaluated characteristics are limited. However, when number of strains increases and subjects for the search become complicated, manual or visual retrieval from the printed list will become almost impracticable. Worldwide trend of the use of computerized data bases has been reported also in the field of genetic resource information management⁴⁾.

In the Ministry of Agriculture, Forestry and Fisheries (MAFF) a special committee to study the information management system for genetic resources was organized in 1979, and a project aiming at the efficient processing of genetic resource information by computer was initiated. With the participation of specialists in universities, the general plan was firstly proposed, and planning and testing of practical examples of various crop species were started by the working group with members shown in Table 1. Secretariat of Agriculture, Forestry, and Fisheries Research Council of MAFF is responsible for promoting the committee's activities, and breeders for each of important crop species or crop groups have been cooperating on the schedule shown in Fig. 1. The activities of the working group are centered for the moment on a few crop species with some former experience of related study.

As regards to the computer use, MAFF has

Table 1. List of participants of the project team for developing genetic resource information management system in MAFF

Organization							
Coordinator	Liaison and Coordination Division of AFF Research Council Secretariat						
Teshainal	Univ. of Tokyo, Lab. of Biometrics						
advisors	Computing Centre for Res. in AFF (CCRAFF)						
Members	Nat. Inst. Agricultural Sciences Div. of Genetics						
	Div. of Statistical Research						
	Central Agricultural Experiment Sta-						
	tion (Rice, Wheat, Barley)						
	Hokkaido N.A.E.S. (Sugar beet)						
	Tohoku N.A.E.S. (Soy bean)						
	Hokuriku N.A.E.S. (Rice)						
	National Grassland Research Institute (Forage crops)						
	Fruit Tree Research Station (Fruit tree)						
	Vegetable and Ornamental Crops Re- search Station (Vegetable, Flower)						
	National Research Institutee of Tea (Tea tree)						
	Sericultural Experiment Station (Mul- berry tree)						

AFF: Agriculture, Forestry, and Fisheries. N.A.E.S.: National Agricultural Experiment Station.

the Computing Centre for Research in Agriculture, Forestry and Fisheries (CCRAFF) at Tsukuba, with its central computer, ACOS 800 II with 4 MB core memory. It acts as the center of the on-line network system as shown in Fig. 2, having five A class translating stations with a card reader and a line printer which enables local remote batch use, and with display terminal for time-sharing use at the station and also with the communication control unit which accepts and transfers local telephone call from timesharing terminals at nearby research stations to the center. B class translating station with all capabilities except communication control facility as that of A class station is now established at three research stations in Kanto District and to be furnished in several other research stations year by year.



Fig. 1. System development plan of information management for plant genetic resources in MAFF. DBMS: Data base management system.

Through the activities of the committee and also supported by personal interest of some breeders, a dozen data bases composed of varieties or strains of each crop such as rice, soy bean, barley or others have been created in CCRAFF mainly for the search of breeding materials. Items are ranging from 800 to over 10,000, and descriptors are from 30 to 75 including agronomic characteristics and sometimes pedigree information, too.

Data base management systems

A data base management system EXIS, specially designed for the use of information retrieval and analysis of genetic resources at the University of Colorado, was introduced into Japan and adapted to the current operating system ACOS-6 of the CCRAFF. Numerical data, coded data and also word data like names can be processed in selecting descriptor options according to the detail of data structure. It has its own macro commands in natural English expression such as READ DATA BANK, HOW MANY ITEMS WITH...., or PRINT....FOR WITH...., for the convenient management of data. EXIS includes EXIR, a retrieval system, RPG, a report paper generator, MINITAB, an elementary statistical analyzer and NTSYS, a multivariate classificatory package program (Fig. 3). The data base is firstly created by



Fig. 2. On-line network of computer system in MAFF



Fig. 3. System flow diagram of EXIS

EXIR, defining necessary descriptors in literal, nominal or numerical option using source data of constituent items punched in cards. Selected items which satisfy the given complex conditions indicated by a query or combination of logical expressions are counted, and descriptor states of the indicated descriptors of the items responding to the above query are printed according to the given commands. If necessary, data retrieved by EXIR can be transferred by the interface program through work file facility for further A statistical package program analyses. named STATPAC 6 is also available for analyses of retrieved data using special interface program. When the user is interested in his proper way of handling retrieved data, he can store them into his own permanent file by appropriate Job Control Language for his late use.

In this paper we discuss mainly on the use of EXIS, but several other data base management systems are in operation among us now. Saito⁶⁾ established a data base management system SERIS for silkworm breeding project. Annual performance reports are processed by this system and past breeding records are available for retrieval and analysis by breeders. Ootsuka³⁾ reported a data base management system BIRS-T, named after Breeders' Information Retrieval System for family Tree reference. It is based on the TAXIR accesioner theory and is equipped with facilities specific to breeding, like family tree drawing and printing of agronomic character values paired with those of the standard variety.

Information management in the germplasm seed storage

Management of genetic resources of main seed crops for agriculture is officially taken over by the Germplasm Seed Storage Center (GSSC) which belongs to the Division of Genetics, National Institute of Agricultural Sciences (NIAS), as far as research activities in MAFF are concerned.

Routine passage of crop seeds for conservation as genetic resources, those of new varieties or strains harvested at breeding stations as well as introduced or collected by breeders or researchers, is outlined in Fig. 4. Mostly they are evaluated as to important agronomic characters at each stations, but sometimes data are insufficient for compilation as genetic resource information. Systematic data recording with standardization of descriptors and descriptor states is necessary and has been discussed for several crop species recently.

In Fig. 5 process of seed acceptance is shown with several directories compiled at the time of reception. The data base is up-dated by the information contained in the card of special format attached to the in-coming seed lots, and some classified lists of varieties in order of accession number, in alphabetical order of variety names, on harvested year or location and so on are printed. Some of these lists are used conveniently at the time of seed distribution, germinability test or seed



Fig. 4. Flow chart for preservation and utilization of the germplasm seed in MAFF ————flow of information ————flow of seed

multiplication.

Steps of action to be taken in reply to the seed request from breeders or researchers are shown in Fig. 6. The data base is referred at this time, too. After shipment, the data base is corrected and lists of statistics showing the mode of seed distribution is obtained when necessary.

The data base for the management of Germplasm Seed Storage has also been created in CCRAFF, separately for each crop species with several thousand accessions for the mo-



Fig. 5. Function of GSSC (1) Seed acceptance



Fig. 6. Function of GSSC (2) Seed distribution

ment. Altogether it will cover more than thirty thousand accessions in the storage and will be open to every breeders or researchers in MAFF for on-line reference.

Information management in GSSC is now partly in operation by the use of data bases. Printing of several classified lists of accessions convenient for the review of present collection and for design of the new introduction or expedition, data analysis of germinability test and supervising of inventory for multiplication or rejuvenation are the examples of the activity.

We are now introducing a new computer into GSSC for more efficient information management, and after its establishment most of the hitherto manual handling of information such as those concerning germinability test or actions concerning seed multiplication are to be supervised by computerized data bases in near future.

Use of EXIS in genetic resource information management

The old and newly developed two-rowed barley strains preserved at Kyushu National Agricultural Experiment Station were defined using 34 descriptors such as variety name, origin, mother name, culm length, panicle length, grain size, spring habit and so on (Table 2). Breeding materials were extracted by the queries formulating necessary conditions as to agronomic characters, and examined closely. Correlations among agronomic characters were determined for groups different in heading date as shown in Table 3. Character associations including culm length or panicle length were rather different between groups, indicating the multi-dimensional disposition of these two groups different in maturity. This is important in selecting breeding materials and crossing combination appropriate for the breeding purposes. Correlations among agronomic characters were also determined for several groups different in their origin, and compared each other to see the geographical effect on the feature of character combinations among the collection and their relative importance as breeding materials.

Making subdivisions of data corresponding to the difference of any (combination of) characters are not always easy in usual way

Descriptors		0	Option				
1	NUMBER	ORDER	FROM 1 TO 2000				
2	VARIETY NAME	TEXT	<32 letters				
3	ORIGIN COUNTRY	NAME	<100 cases				
4	ORIGIN PREFECTURE	CODEZ	HOKKAIDO (01), AOMORI (02) KAGOSHIMA (46), OKINAWA (47)				
	MOTHER NAME	TEXT	<24 letters				
20	CULM LENGTH	ORDER	FROM 0 TO 1000 BY 1 IN CM				
21	PANICLE LENGTH	ORDER	FROM 0 TO 50.0 BY 0.1 IN CM				
25	GRAIN SIZE	ORDER	FROM 1 TO 10 BY 1				
34	SPRING HABIT	ORDER	FROM 1 TO 99 BY 1				

Table 2. Barley data base

34 descriptors and 849 items.

Investigated and published at Kyushu National Agricultural Experiment Station.

Tata variation			Early varieties							
Late varieties			Cı	C ₂	C4	C ₅	C ₈	C ₁₂		
	mean	s.d.	plant type	days to heading	clum length	panicle length	grain size	1000 grain weight		
mean			2.29	18.6	84.3	8.30	2.23	49.0		
s.d.			1.20	3.25	11.3	1.34	0.805	5.51		
C1	3.39	1.12		0.39**	-0.10	-0.06	0.02	-0.03		
C ₂	39.4	1.88	0.13		0.14	0.25*	0.08	0.08		
C ₄	87.2	8.01	0.24	-0.14		0.24*	-0.04	-0.21		
C ₅	10.3	1.99	0.02	-0.06	0.17		-0.04	-0.12		
Cs	2.23	0.762	0.05	0.03	-0.33*	0.13	2200127621	-0.30*		
C12	47.7	4.40	0.21	-0.04	0.35*	-0.21	-0.75**			

Table 3. Retrieved results: difference of character values and their association between early and late varieties

 C_1 plant type: prostrate 1, erect 5

C2 days to heading: number of days from April 1 to heading

C4 culm length: cm

C₅ panicle length: cm

C₈ grain size: large 1, small 5 C₁₂ 1000 grain weight: g *: significant at 5% level

**: significant at 1% level

is quite easy and, in using EXIS, retrieved statistical methods available.¹⁾ data groups are simply transferred to the

of data handling. By data base technique it succeeding analysis with ample choice of

Another example is the reference of parent-



offspring relationship of rice varieties. Over 600 recommended rice varieties are defined with their old strain names before registration, mother name, father name, and some other breeding records. Parents of any variety are retrieved and varieties having the same parent or parental combination are listed interactively. Genealogical tree of any variety can be drawn also from this data base using EXIR with additional function newly developed by one of the present authors (Fig. 7).⁹⁾

Use of EXIS in analysis of adaptability

Tall fescue strains improved at the Hokkaido National Agricultural Experiment Station were tested over 18 locations throughout Japan for 1 to 3 years. Their forage production was measured by clipping and weighing 1 to 11 times per year, depending on strain and location. Climatic variables such as temperature or day length were also recorded. These data were compiled as a data base for 35 descriptors like test location, strain name, year of stand, clipping number, clipping date, dry matter weight, mean temperature, amount of precipitation and so on as 2065 items.8) For the analysis of adaptability many subsets of data, i.e., subgroups of environments and/or strains, are often subjected to the same type of analysis. Also several different types of multivariate statistical methods are applied to the same data set. Not only one character such as yield but also several other characters, or sometimes, climatic variables are included in the analysis. For this way of data handling use of data base system is quite efficient.¹⁰⁾ Wide adaptability of the variety Yamanami through Mainland Japan, and specific adaptability of another variety Hokuryo to Hokkaido were ascertained as indicated in Table 4. Differences in climatic response especially to accumulated temperature or day length among

Table 4. Regression analysis of dry matter yield on site mean for three regions

	Hokka	ido	North ce Japa	ntral n	Southern Japan		
Strain	D.m. yield kg/a	Regr. coef.	D.m. yield kg/a	Regr. coef.	D.m. yield kg/a	Regr. coef.	
Hokuryo	31.0	1.24	26.7	1.28	17.6	1.09	
Yamanami	27.5	.881	27.6	.761	18.3	.749	
Hokkai-No. 3	28.7	.947	27.5	1.02	16.6	1.10	
Hokkai-No. 4	28.2	.915	28.1	1.07	19.4	1.09	
Hokkai-No. 5	29.3	1.15	23.8	1.07	15.2	1.17	
Kentucky-31	25.0	.736	25.4	.808	16.0	.806	

Table 5. Regression analysis of dry matter yield on climatic conditions

Strain	D.m.	Partial regr. coef.			Mult	% of reg. ss.		
	yield kg/a	Gr. days	Acc. temp.	Day length	corr. coef.	Gr days	Acc. temp.	Day length
Hokuryo	23.50	.518**	0148**	4.07**	.626**	69.0	9.5	21.5
Yamanami	23.05	.360**	0108**	1.88**	.508**	70.3	17.1	12.7
Hokkai No. 3	23.92	.420**	0099**	2.15**	.558**	81.7	7.8	10.4
Hokkal No. 4	23.79	.424**	0100**	2.46**	.559**	80.8	6.7	12.5
Kentucky 31	20.84	.326**	0098**	1.69*	.490**	71.3	16.3	12.5

reg. ss.: regr. sum of squares, Acc. temp.: accumulated temperature

Gr. days: No. of growing days

*: significant at 5% level

**: significant at 1% level

bred strains were revealed by regression analyses (Table 5). Stepwise regression analysis of dry matter yield on 8 environmental variables for the test period in 3 regions of Japan indicated resemblance and difference among strains within and between regions. Relative contribution of dry matter yield into several largest principal components extracted from climatic records and performance data clearly indicated the difference in physiological response among these strains to climatic conditions. Hokuryo and Yamanami were registered by MAFF as new high-yielding varieties adaptable to Hokkaido or south-western Mainland, respectively, mainly for grazing heifers and beef cattle, and seed multiplication has been initiated. Results of these analyses are useful and important for further breeding in accumulating basic information as well as for evaluation of candidate strains.

In the rice breeding project 13 breeding stations distributed all over Japan has been developing new varieties aiming at various breeding purposes inherent to the localities. Finally selected candidate strains from these stations are delivered to the prefectural testing stations in the proposed area of diffusion to evaluate their adaptability. The data of this test, called as the 'Test for Deciding Varieties for Recommendation', are very precious but seemingly have not fully been utilized yet. It is proposed to create data bases by BIRS-T for printing the annual report very efficiently and also by EXIS for further analysis starting from this year.

As to the activities on genetic resource information management outside of MAFF, Sugawara, Tateno and Sakamoto⁷⁾ reported that The Institute of Physical and Chemical Research has been engaged in research and development of data bases of micro-organisms, plant, algae, animal and plant tissue culture, and newly established information system will serve for scientists with bibliographical and characteristic data of these laboratory organisms.

Future scope

EXIS can be used in several other ways by

the help of some devices in job setting and attached FORTRAN programs. Family tree drawing, data transformation or addition of new descriptors based on the manipulation of already existing descriptors and so on are being devised by one of the authors.9) Simpler steps to refer another data base by the retrieved results of one data base is also under way. A study group presided by Professor K. Saio of the University of Tokyo has been constructing a prototype of relational data base management system SIRA/GR especially designed for the more efficient use of information to help progress of plant breeding projects¹²⁾. For other stages of plant breeding, such as planning for selection procedure through generations, for instance, the appropriate use of data base by the conversational mode through computer terminals on the laboratory desk may be helpful in examining available data, reflecting possible choices and deciding which way to take. Wide use of data base technique in plant breeding making the best use of genetic resource information seems to be more and more important in future.

References

- Kumagai, K. & Uchiyamada, H.: Rice characteristic analysis. In Tokyo Scientific Center Report N; GE18-1882-0, Using Genetic Resources Data Bases. Tokyo, IBM Japan. p. 3-55 (1978) [In Japanese with English summary].
- Kushibuchi, K. & Ootsuka, Y.: Construction and use of a data base from rice breeding records. Jap. J. Breed., 28, 375-385 (1978) [In Japanese].
- Ootsuka, Y.: Breeders' information retrieval system(1) BIRS-T(77.3). Bull. Computing Centre for Research in Agriculture, Forestry and Fisheries., A 14, 213-285 (1978) [In Japanese with FORTRAN program list].
- Rogers, D.J. & Hersh, G.N.: Genetic resources communication information documentation system GR/CIDS. Proc. Int. Symp. on Wheat Genetic Resources., 65-75 (1975).
- Saio, K. et al.: On the retrieval and analyses system for description of varieties and strains. 1. Application test of the TAXIR theory. Jap. J. Breed., 26 (Suppl. 2), 27-28

(1976) [In Japanese].

- Saito, S.: Studies on the data processing method for the cross-bred comparison of silkworm by use of computer. *Tech. Bull. Sericultural Experiment Station*, 107, 53-90 (1978) [In Japanese].
- 7) Sugawara, H., Tateno, Y. & Sakamoto, N.: Development of a National Information System of laboratory organisms. 7th International CODATA Conference, Kyoto. (1980).
- Suzuki, S.: Examination of adaptability test of tall fescue. In Tokyo Scientific Center Report N: GE18-1882-0, Using Genetic Resources Data Bases. Tokyo, IBM Japan, 98-150 (1978) [In Japanese with English summary].
- 9) Suzuki, S.: On the use of a data base system for information processing in plant

breeding. 7th International CODATA Conference, Kyoto. (1980).

- Suzuki, S. & Kawabata, S.: The system for information retrieval and analysis of genetic resources. 6. Use of EXIS for the efficient evaluation of adaptability in the strain testing program. Jap. J. Breed., 29 (Suppl. 2), 8-9 (1979) [In Japanese].
- Suzuki, S. & Takano, Y.: Use of EXIR. In Tokyo Scientific Center Report N: GE18-1857-0, Gene Resources Information System. Tokyo, IBM Japan, 18-41 (1977) [In Japanese with English summary].
- Takano, Y. & Saio, K.: The system for information retrieval and analysis of genetic resources.
 A design of the system. Jap. J. Breed., 29 (Suppl. 2), 6-7 (1979) [In Japanese].

(Received for publication, February 10, 1981)

84