

The international coconut genetic resources database

C Hamelin¹, R Bourdeix² and L Baudouin³

¹Scientist, ²Coconut Breeder and ³Geneticist Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Montpellier, Cedex 5, France

Background

During the International Workshop on Coconut Genetic Resources, held in Cipanas (Indonesia) in October 1991, two major recommendations were made for coconut: the first was to set up the International Coconut Genetic Resources Network (COGENT), and the second, concerning network information and documentation, stated as follows:

“Participants agree on the need for a single central coconut database to be developed for the initial stage of the network. The offer by CIRAD Montpellier to act as a host for this database was gratefully accepted. [...]”

A meeting was then organized in Montpellier (France) from 19 to 22 May 1992, with representatives from national collections to clarify the status of existing collections, define how the database was to be organized and draw up the list of descriptors to be taken into account, along with precise standardized ways of observing those descriptors. It was decided that the database would be developed in several stages. The development of an IT (Information Technology) application was initiated in 1992 with the presentation of a database mock-up called the Coconut Genetic Resources Database (CGRD), which eventually became the formal title for the database. It has continued since then on an annual basis.

Each stage involved adding/improving management system functionalities, and increasing the number of accession sites contained in the database. Furthermore, in order to take into account the progress made in IT techniques, the application, which initially functioned under the MS-DOS system, was completely rewritten for the Microsoft Windows system, so that photos and graphs could be displayed, and to provide a more user-friendly interface. Lastly, the progress made in molecular techniques, and their use on coconut to characterize genetic diversity, revealed the need to add the possibility of storing molecular information from these techniques in the database.

Table 1 shows how the database developed over the years in terms of its functions and content. It can be seen that the successive increases in the number of accessions over time has been irregular. The growth rates that can be calculated vary between 1 and 34%. From 1995 to 2002,

Table 1. Stages in the development of CGRD

Date	Version	Functions added/Improved	No. of accessions
1993	Mock-up	Passport data	0
	Version 1.0	Functions to query the database and to create reports	500
1994	Version 2.0	Characterization and evaluation data	669
		Backup function	
1996	Version 2.1	DIP (Data Interchange Protocol) format for data export	738
		Restoration function	
1997	Version 2.2	New structure	
		Improved software	
1998	Version 2.2 improved	New DIP (Data Interchange Protocol) format for data export	936
1999	Version 2.2	New reports	1225
2000	Version 3.0 for Windows	Migration to Windows	
2001	Version 4.0 for Windows	Improved functions Export in delimited ASCII files (to use in statistical software)	1352
2002	Version 5.0 for Windows	Introduction of coconut molecular data	1369
2003	Version 5.1 for Windows	Improved backup function	1416

Brazil, China, India, Indonesia, Ivory Coast, Jamaica, Mexico, Papua New Guinea, Philippines, Sri Lanka, Thailand, Vanuatu and Vietnam (Bourdeix 1996; 1997a; 1997b; 1998; Bourdeix *et al.* 1999; Baudouin 2002) were visited and local researchers were trained in gathering and inputting data into the database. Such visits also provided an excellent opportunity to test the software in real situations and to detect items that needed improvements.

Objectives

The work on CGRD was initiated for the main purpose of providing the COGENT members with an easily consultable computerized catalogue of accessions representing a large number of cultivars spread throughout the coconut growing zone, in order to gain a clearer picture of coconut genetic diversity and thereby promote exchange of germplasm. This continues to be the main objective.

Another purpose of this database is for the COGENT country members to establish a list of passport descriptors and standardize characterization/assessment descriptors specific to coconut to be used by all partners. In addition, the database was created to speed up cultivar characterization and evaluation. The members of COGENT supplying the information contained in the database are regularly encouraged to add new information or complete the information already recorded requiring them to make a considerable effort to gather information and thereby improve their knowledge of the cultivars planted in their genebanks.

Organization

The data entered into the base are the values of passport and characterization descriptors for coconut accessions defined by COGENT. It also includes photos of the palms, along with results and diagrams from microsatellite molecular analyses. Based on an analysis of these data, and of the relations and constraints existing between them, a conceptual model of database organization was established. It provided a formal description of the database (Figure 1) using entities (symbolized by rectangles) corresponding to the natural objects identified in the system (sites, accessions, cultivars, photos, etc.), and relations (symbolized by circles) between those entities. This model was modified during development, when molecular data or photos were integrated.

In the model, accessions and cultivars form the core of the database. The database structure respects certain management rules, such as:

- It is compulsory for an accession to belong to a collection site and its number is unique in the database;
- A cultivar can be represented by several accessions; and
- Photographs and molecular data are attached to the cultivars.

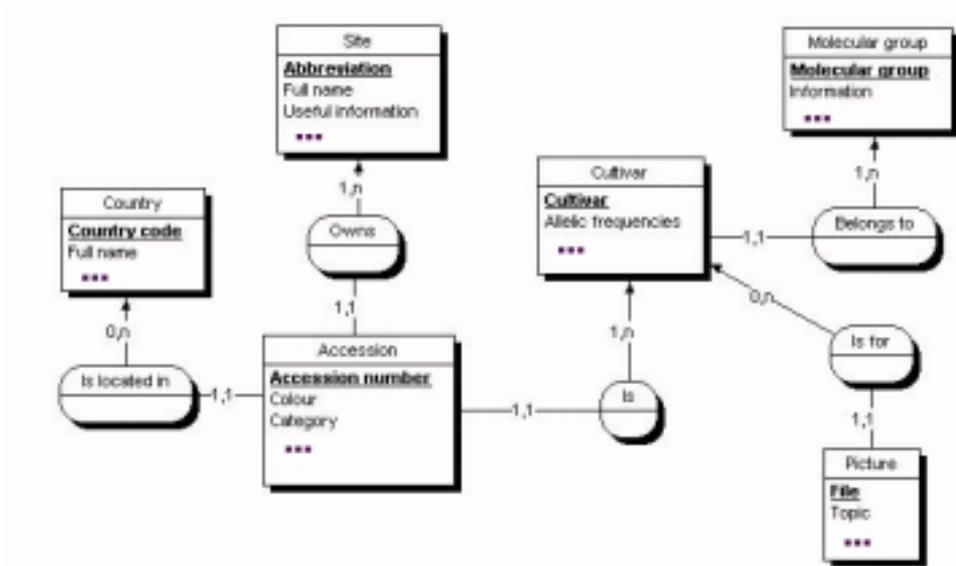


Figure 1. Simplified conceptual model of the CGRD

The conceptual model was translated into a relational type logical model (Figure 2) consisting of tables (symbolized by rectangles) linked to each other, derived from the entities and relations of the conceptual model. A relational type organization was chosen because it is a widely used

organization, which has proven its worth and enables the use of a very powerful query language. Lastly, a very large number of commercial database management systems function with this type of organization. This logical model was implemented in the chosen relational database management system.

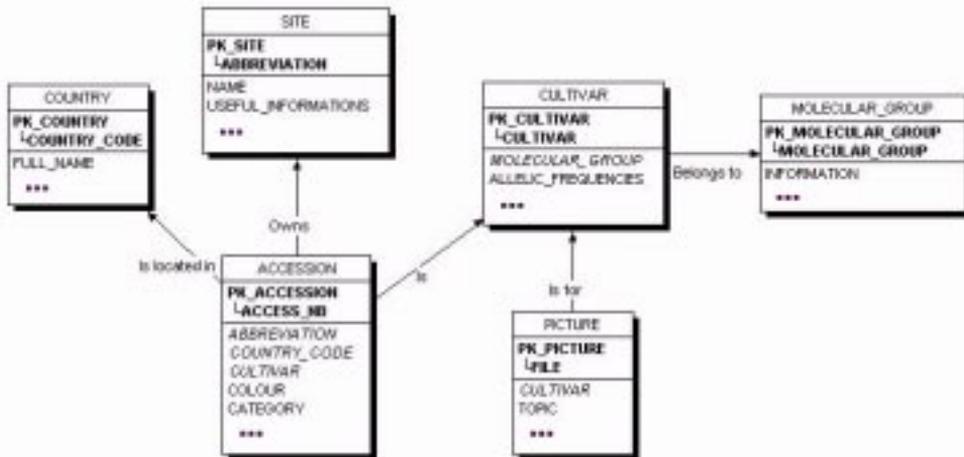


Figure 2. Simplified relational logical model of CGRD database

The CGRD was designed to function in a semi-centralized way. All COGENT members have a version of the database on their machine, which can be used to consult the entire catalogue, and to add or update data on the accessions at their site. In the latter case, a basic functionality enables data backup per site, for transmission to CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement), the organization that centralizes the data from all the collection sites. CIRAD checks data coherence before entering them in the database, which is distributed annually to COGENT members.

Functions

The CGRD has a management system endowed with functions that can be used to carry out all the necessary operations on the data it contains. Among the functions available, which are listed below, there are those that are found in conventional database management systems, but there are also specific functions, given the nature of the database:

- Entering/consulting information on collection sites
- Entering/consulting data on individual accessions
- Selections in the database using criteria
- Creation of various types of reports

- Backing up of the database on an external medium
- Restoring database from a previous backup to the computer hard disk (when the database files are damaged on the hard disk)
- Exporting accession data in D.I.P. format (Data Interchange Protocol, for introduction into generalized genetic resources databases accepting this format)
- Consulting pictures of coconut cultivars on various topics
- Consulting the molecular group-based coconut classification
- Obtaining information on microsatellite markers and viewing the electrophoretic profile
- Online help

The database has a Microsoft Windows type graphic interface with a menu from which operations to be implemented on the data can be chosen.

Figures 3 to 5 show how data are displayed in CGRD. The screen displayed in Figure 3 is a part of the main screen of CGRD where pictures can be seen and from which descriptors data can be entered by clicking on the Passport or Characterization and evaluation data tabs.

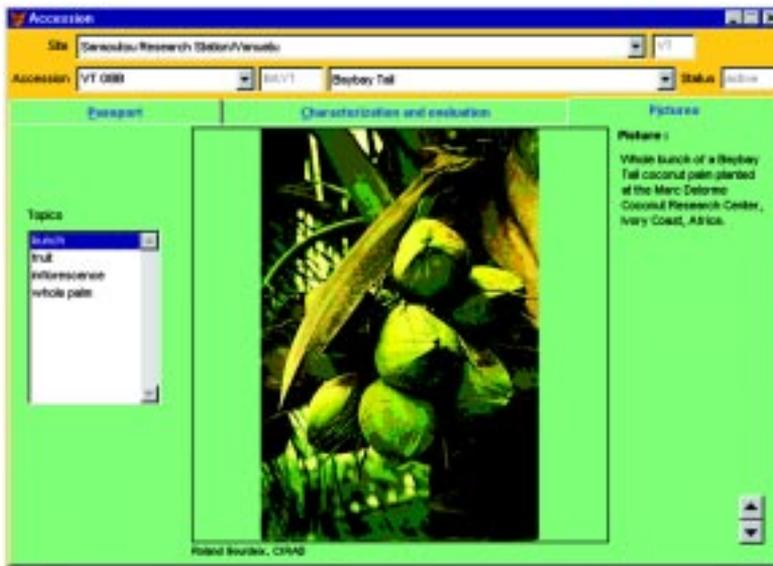


Figure 3. The tab picture of the accession screen in CGRD

Figure 4 shows a list of accessions retrieved after a successful search in the database. By clicking on the Accessions details tab when an accession is selected, all previously entered information on passport or on characterization data will be displayed on the screen.

Date	Accession number	Cultivar name	Abbreviation	Colour1	Colour2
BARI / Bangladesh	B-00001	Rahmatpur Tall-A	R-ALT	Green	Green-Yellow
BARI / Bangladesh	B-00002	Rahmatpur Tall-B	R-BTM	Green	Red-Green
BARI / Bangladesh	B-00003	Rahmatpur Tall-C	R-RTS	Green	NULL
BARI / Bangladesh	B-00005	Khairata Tall	KHT	Green	NULL
BARI / Bangladesh	B-00006	Bhakhal Tall	BKT	Green	NULL
BARI / Bangladesh	B-00007	Hafizari Tall	HTZT	Green	Green-Red
BARI / Bangladesh	B-00008	Shundi	HTZTS	Yellow	Red-Yellow
BARI / Bangladesh	B-00009	Shundi	PHTR	Red	NULL
BARI / Bangladesh	B-00010	Jamarai Tall	JMT	Green	Green-Yellow
BARI / Bangladesh	B-00011	Deshi Narkal	SDT	Green	Green-Yellow
BARI / Bangladesh	B-00012	Polanpur Tall	PLRPT	Green	Green-Red
BARI / Bangladesh	B-00013	Rangsh Tall	RACHT	Green	Green-Yellow
BARI / Bangladesh	B-00014	Khagchari Tall	KHAGT	Green	Green-Yellow
BARI / Bangladesh	B-00015	Babagan Tall	BBOT	Green	Yellow-Green
BARI / Bangladesh	B-00016	Utpar Tall	ULT	Green	Green-Yellow
BARI / Bangladesh	B-00017	Aganera Tall	AGT	Green	Green-Yellow
BARI / Bangladesh	B-00018	Swankathi Tall	SWT	Green	NULL
BARI / Bangladesh	B-00019	Jhalakathi Tall	JLKT	Green	Yellow-Green
BARI / Bangladesh	B-00020	Kalapara Tall	KPAT	Green	NULL
BARI / Bangladesh	B-00021	Labakhal Tall	LBUT	Green	NULL
BARI / Bangladesh	B-00022	Bhola Tall	BHOLT	Green	NULL
BARI / Bangladesh	B-00024	Baghapara Tall	BOPT	Green	NULL

Figure 4. The screen resulting from a successful selection of accessions in CGRD

Figure 5 shows the Classification Tree of the coconut cultivars annotated with the molecular groups identified in the species. A click on a letter of the molecular group, displays the list of cultivars of the same group on the right of the screen.

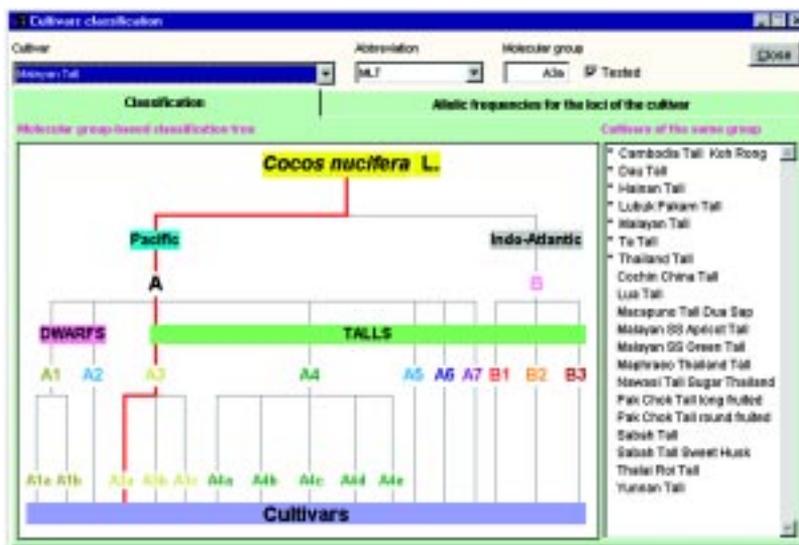


Figure 5. The Classification Tree in CGRD

Contents

Each accession in the database is characterized by 145 descriptors, of which there are 72 passport descriptors and 73 characterization and evaluation descriptors. Passport descriptors are divided into two groups: the accession data group, which contains accession characteristics (category, colour, cultivar, parental origin, etc.) and the collection data group, which contains data on the site and the original population from which the accession was sampled.

The characterization and evaluation descriptors are generally quantitative traits, whose values are means, calculated from values measured on several palms. A standard deviation is associated with each of the means, to have an idea of the variability of the trait within the accessions. To obtain these values, a data management system has to be put in place for every palm in the field. This is not trouble-free, because it has to be carried out over a long period following a regular schedule. In order to help researchers to follow this management protocol, COGENT has committed CIRAD in 1996 to develop dedicated software called CDM (Coconut Data Management). This software was designed to manage experimental data observed on collections and experimental fields of coconut and other crops. The version 3 delivered in March 2000 is able to manage the palm identification characters along with data on observations during the vegetative phase, leaf morphology, stem measurements and state of the palms. It is possible to execute powerful queries on the database, to export data into external file, and to make statistical analysis of widely used experimental designs.

The CGRD characterization and evaluation descriptors are divided into:

- Site descriptors, information about the site at which the accession is to be found and about the people assessing it;
- Germination descriptors, germination rates and percentages;
- Stem descriptors, stem morphology (height, circumference, number of internodes, etc.);
- Leaf descriptors, leaf morphology (petiole, rachis, leaflets, etc.);
- Inflorescence descriptors, to characterize inflorescences (peduncle, spikelets, number of female flowers, etc.);
- Flowering descriptors, such as the length of male and female phases, information on overlapping of these phases, spathe emission date, inflorescence opening date, etc.;
- Fruit descriptors, nut characteristics (shape, weight of different compartments, dry matter weight, etc.);
- Yield descriptors, number of bunches, number of nuts, quantity of copra; and

- Oil descriptors, oil/nut characteristics (quantity of oil in fresh matter and in dry matter form).

In 2003, the main statistics about the database were as follows (Table 2a and 2b):

- The accessions of 28 genebanks located in South Asia, Southeast Asia, the South Pacific, Africa, and the Caribbean–Central America zone figure in the database.
- Some countries, such as Indonesia, Malaysia and Papua New Guinea, have at least two sites where collections are maintained, with a maximum of four for Indonesia.
- A little over 60% of the accessions are in the South and Southeast Asia, primarily in India, the Philippines and Indonesia.
- Not all the descriptors are filled in; a little over half of the accessions (all sites combined) have values for 25% to 50% of their passport and evaluation descriptors.
- Very few accessions have values for 100% of their passport descriptors or evaluation descriptors.
- The database lists 599 Tall cultivars, 111 Dwarf cultivars, 1 semi-Tall cultivar plus a few others of small size but cross-fertilizing.
- Some cultivars such as the Malayan Tall, Pakistan Tall are described very often in the database, represented by 49 and 32 accessions, respectively, whereas more than 70% of the cultivars are only represented by one accession.
- The database contains 754 photos representing a little over 20% of the cultivars. The aspects illustrated are the whole plant, the crown, the inflorescence, the fruit, genetic diversity, and cultural aspects of the coconut palm.
- Photos of inflorescence, bunch and fruit are the most numerous.
- The molecular data contained in the database come from microsatellite electrophoresis studies carried out on cultivars. On average, 14 microsatellite loci have been analyzed per cultivar.
- For each cultivar and each locus, the frequencies of the different alleles found at the locus are indicated.
- Based on the allelic frequencies of the microsatellite loci studied, cultivars have been assigned to different molecular affiliation groups. The molecular group of each cultivar is also recorded in the database.

A cultivar classification tree based on the molecular groups is proposed in one of the database modules. It is possible to move along the different branches of the tree, bringing up a list of the cultivars attached to the selected molecular group in each case. The microsatellite data of 119

cultivars are stored in the database. Lastly, data per microsatellite locus (SSR locus type, allele size, number of alleles, chromosome etc.), along with images of the microsatellite electrophoretic profiles and profile interpretation diagrams are also recorded in the database.

Table 2a. Number of accessions per site or per region according to the percentage of passport descriptors (P) filled in

Site	Number of accessions	P=0	0<P≤25	25<P≤50	50<P≤75	75≤P≤100
CNRA Marc Delorme Research Station, Côte d'Ivoire	99	0	0	28		7
Coconut Programme, OPRI, Ghana	16	0	16	0	0	0
CRC Sémé Podji, Benin	4	0	0	4	0	0
National Coconut Development Programme, Tanzania	72	0	1	63	8	0
AFRICAN REGION	191	0	17	95	72	7
Centro de Investigacion Cientifica de Yucatan, Mexico	20	0	0	20	0	0
Coconut Industry Board, Jamaica	60	0	44	15	1	0
EMBRAPA, Brazil	16	0	0	0	16	0
LATIN AMERICA-CARIBBEAN REGION	96	0	44	35	17	0
BARI, Bangladesh	40	0	22	16	2	0
Coconut Research Institute, Sri Lanka	78	0	0	46	32	0
CPCRI, India	212	0	71	74	67	0
RS, Pakistan	32	0	32	0	0	0
SOUTH ASIAN REGION	362	0	125	136	101	0
Cocoa and Coconut Research Institute, PNG	3	0	3	0	0	0
Ministry of Agriculture, Tonga	7	0	7	0	0	0
Saraoutou Research Station, Vanuatu	79	0	6	26	45	2
Stewart Research Station, PNG	54	0	23	31	0	0
Taveuni Coconut Centre, Fiji	11	0	2	0	8	1
RS, Western Samoa	9	0	9	0	0	0
RS, Yandina, Solomon Islands	21	0	17	4	0	0
SOUTH PACIFIC REGION	184	0	67	61	53	3
RS, China	17	0	2	15	0	0
Department of Agriculture, Sabah, Malaysia	45	0	22	23	0	0
MARDI, Hilir, Perak, Malaysia	44	0	10	34	0	0
Bone Bone Experimental Garden, S. Sulawesi, Indonesia	41	0	6	35	0	0
Mapanget Experimental Garden, N. Sulawesi, Indonesia	74	0	0	63	11	0
Pakuwon Experimental Garden, W. Java, Indonesia	25	0	3	21	1	0
Sikijang Experimental Garden, Indonesia	30	0	0	10	20	0
Philippine Coconut Authority, Philippines	224	0	3	220	1	0
Chumphon Horticultural Research Centre, Thailand	52	0	10	37	5	0
Dong Go Experimental Center, Vietnam	31	0	0	1	29	1
SOUTHEAST ASIAN REGION	583	0	56	459	67	1
TOTAL FOR ALL REGIONS	1416	0	309	786	310	11

Table 2b. Number of accessions per site or per region according to the percentage of evaluation descriptors (E) filled in

Site	Number of accessions	E=0	0<E≤25	25<E≤50	50<E≤75	75≤E≤100
CNRA Marc Delorme Research Station, Côte d'Ivoire	99	0	28	28	43	0
Coconut Programme, OPRI, Ghana	16	12	0	4	0	0
CRC Sémé Podji, LOCATION Benin	4	0	0	4	0	0
National Coconut Development Programme, Tanzania	72	0	3	69	0	0
AFRICAN REGION	191	12	31	105	43	0
Centro de Investigacion Cientifica de Yucatan, Mexico	20	0	19	1	0	0
Coconut Industry Board, Jamaica	60	0	2	58	0	0
EMBRAPA, Brazil	16	0	0	16	0	0
LATIN AMERICA-CARIBBEAN REGION	96	0	21	75	0	0
BARI, Bangladesh	40	0	3	3	34	0
Coconut Research Institute, Sri Lanka	78	0	14	51	13	0
CPCRI, India	212	0	1	138	73	0
RS, Pakistan	32	0	32	0	0	0
SOUTH ASIAN REGION	362	0	50	192	120	0
Cocoa and Coconut Research Institute, PNG	3	0	0	3	0	0
Ministry of Agriculture, Tonga	7	0	6	1	0	0
Saraoutou Research Station, Vanuatu	79	0	6	10	1	62
Stewart Research Station, PNG	54	0	0	54	0	0
Taveuni Coconut Centre, Fiji	11	0	4	3	4	0
RS, Western Samoa	9	0	0	9	0	0
RS Yandina, Solomon Islands	21	0	0	21	0	0
SOUTH PACIFIC REGION	184	0	16	101	5	62
RS, China	17	0	0	17	0	0
Department of Agriculture, Sabah Malaysia	45	0	15	30	0	0
MARDI Hilir Perak, Malaysia	44	0	5	2	37	0
Bone Bone Experimental Garden, S. Sulawesi, Indonesia	41	0	0	41	0	0
Mapanget Experimental Garden, N. Sulawesi, Indonesia	74	0	29	45	0	0
Pakuwon Experimental Garden, W. Java, Indonesia	25	0	0	25	0	0
Sikijang Experimental Garden, Indonesia	30	0	0	30	0	0
Philippine Coconut Authority, Philippines	224	0	5	138	81	0
Chumphon Hort. Research Centre, Thailand	52	0	0	52	0	0
Dong Go Experimental Center, /Vietnam	31	0	15	12	4	0
SOUTHEAST ASIAN REGION	583	0	69	392	122	0
TOTAL FOR ALL REGIONS	1416	12	187	865	290	62

Technical references

The CGRD management system for the relational structure is Microsoft Visual FoxPro software, which operates under Microsoft Windows. Tables are in Dbase (DBF) format, which is a universally recognized format.

Conclusion

The CGRD is a very useful tool for coconut genetic resources management, and can help planting material exchanges. It can be considered that the development of the database is now mostly complete. However, extra outputs for dedicated software such as the new geographical information system DIVA-GIS, which is tailor-made for genetic resources, should be included. It is also expected that the coconut descriptors list is a dynamic one and will evolve as new traits are added or the existing ones are modified. The CGRD has been designed to deal with such changes. For example, some morphological descriptors, such as the nut germ pore size and shape, may be added in the future. Moreover, the existing description of yield is often considered as insufficient by researchers. Similarly, the information about the pedigree of the accessions needs to be clarified and simplified in order for researchers to locate more easily the original collection site of the accessions.

Data for a large number of accessions on some of the descriptors is incomplete. Some key data, such as the inventory/counting of the living coconut palms, need to be updated much more often. It would also be worth adding photos, as too few varieties are illustrated. So far, molecular data have only involved microsatellite markers, but it is possible to add results from other types of molecular markers.

It is important to ensure that the number of accessions, photos and molecular data are increased significantly. This will have to be carried out by the national curators. Regular visits by technical experts to participating countries should be promoted to keep the momentum going and make the CGRD effort sustainable.

Lastly, it would be worthwhile if a version either complete or, only consultative, or even restricted to passport data of this database could be made accessible via the internet, as is the case for some other plants such as banana (Arnaud and Horry 1997). This would greatly help the global collaborative effort on coconut genetic resources conservation and their utilization. But this would require cooperation and agreement of the COGENT member countries.

References

- Arnaud, E and JP Horry. 1997. *Musalogue*: a catalogue of *Musa* germplasm. Papua New Guinea collecting missions 1988-1989. <http://www.ipgri.cgiar.org/system/page.asp?theme=4#Bananasandpla>.
- Baudouin, L, R Bourdeix, F Bonnot, Ch Hamelin and A Rouzière. 2000. COGENT establishes an international coconut genetic resources database (CGRD). COGENT Newsletter 3:1-2.
- Baudouin, L. 2002. Study on genetic resources status in Hainan – China (visit 12-18 December 2000). Doc CIRAD-CP SIC no. 1505. CIRAD, Montpellier, France. 54p.
- Bourdeix, R. 1996. Coconut germplasm in Jamaica, Mexico and Brazil. Mission Report, March 1997, Doc CIRAD-CP no. 548. CIRAD, Montpellier, France. 53p.
- Bourdeix, R. 1997a. Coconut germplasm in Tanzania, Sri Lanka and India. Mission Report, March 1997, Doc CIRAD-CP No. 739. CIRAD, Montpellier, France. 49p.
- Bourdeix, R. 1997b. Actions de formation en Afrique et Amérique Latine/ Caraïbes. Cours de formation des formateurs à l'utilisation du manuel Stantech (Techniques Standardisées de Recherches pour l'Amélioration du Cocotier). Mission Report, August 1997, Doc CIRAD-CP No. 857. CIRAD, Montpellier, France. 49p.
- Bourdeix, R. 1998. Coconut germplasm and breeding in Papua New Guinea and Vanuatu. CIRAD 1997 Annual Review Report (March 1998), Doc CIRAD-CP No. 998. CIRAD, Montpellier, France. 50p.