

Mapping of coconut genetic diversity

R Bourdeix¹, L Guarino², PN Mathur³ and L Baudouin⁴

¹Coconut Breeder and ⁴Geneticist, Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Montpellier, Cedex 5, France

²Plant Genetic Resources Advisor, Secretariat of the Pacific Community (SPC), Suva, Fiji

³Scientist, International Plant Genetic Resources Institute - South Asia Office, New Delhi, India

Introduction

Mapping of coconut genetic diversity means representing any characteristic of coconut populations on maps, be it their phenotypic or molecular traits; and then studying the links between these traits and any other useful spatial information. According to the needs, the cultivars may be related to their site of origin or to the genebank where they are conserved. This type of analysis can improve the effectiveness of collecting, conservation, management and use of coconut genetic diversity.

The mapping studies conducted so far have used data on accessions already collected and conserved in germplasm banks around the world. The latitude and longitude of collection sites have been entered into databases and checked. Then the localities of collection sites are mapped to locate under-represented areas, i.e. areas in which the coconut palms can potentially grow, but where collecting has been inadequate or has not occurred at all. Subsequently, it is possible to identify hotspots of diversity and investigate the geographic distribution of specific traits or combinations of traits using information such as characterization and evaluation, including molecular markers, of the available genetic resources.

Georeferencing coconut accessions

From 1995 to 2002, coconut researchers of Brazil, China, India, Indonesia, Ivory Coast, Jamaica, Mexico, Papua New Guinea, Philippines, Sri Lanka, Thailand, Vanuatu and Vietnam were trained (Bourdeix 1996; 1997a; 1997b; 1998; Bourdeix *et al.* 1999; Baudouin 2002) in gathering and inputting data into the Coconut Genetic Resources Database (CGRD) (see Hamelin *et al.*, Chapter 7). Most of the determination and recording of the geographical location of the collecting sites were done in the framework of this work (Bourdeix *et al.* 1999).

This preliminary work was conducted without any sophisticated geographical information system (GIS). It consisted mainly in marking sites of collection by hand on easily available commercial hard-copy maps, preferably with the aid of the researchers who were in charge of the collecting in each country.

The method of linear approximation was used to determine, as precisely as possible, the longitude and latitude of each collection site on the maps. Later, the Encarta electronic atlas was also used to obtain geographic information more quickly. All the data were systematically entered into the CGRD and then extracted for further geographical analysis.

Table 1 presents the status of the geographical localization of coconut accessions according to countries of conservation in the CGRD (version 5.0, December 2002).

Table 1. Georeferencing of coconut accessions in CGRD (version 5.0)

Country	Number of conservation sites	Total number of coconut accessions	Number of accessions with geographical coordinates	Percentage of accessions with geographical coordinates
Benin	1	4	4	100
Bangladesh	1	40	40	100
Brazil	1	16	16	100
China	1	17	17	100
Ivory Coast	1	99	99	100
Mexico	1	20	20	100
Philippines	3	224	224	100
Vietnam	1	31	31	100
Indonesia	4	156	151	97
Vanuatu	1	79	71	90
Papua New Guinea	2	57	51	89
Sri Lanka	1	78	65	83
Fiji	1	11	9	82
Thailand	2	124	97	78
India	1	212	115	54
Solomon Islands	1	21	11	52
Malaysia	2	89	18	20
Jamaica	1	60	4	7
Ghana	1	16	0	0
Pakistan	1	32	0	0
Tonga	1	7	0	0
Western Samoa	1	9	0	0
Total	30	1402	1043	Average 74

A total of 1043 accessions in the CGRD are localized by longitude and latitude. Because of duplicates, these 1043 accessions refer to only 710 distinct cultivars or population names. Only 579 collecting sites have a unique combination of latitude and longitude, to the level of minutes.

Currently, the CGRD database gives only the geographical localization of the factual female parent, i.e. the location of the palms where the seednuts of the accession have been collected. Many accessions have been moved from one research institute to another; others are rejuvenations of the original population within the same institute. In these two situations, as the female parent is located in a research institute, the geographic coordinates of the institute is given in the CGRD as the 'collection site'. There is no direct information regarding the collecting site of the original sample of each accession.

Let us take a practical example. A researcher wants to know the real

origin of the accession 'NCDP-D9'. This accession is a Tacunan Green Dwarf (originating from the Philippines) but planted in Tanzania in 1989:

- 1) Looking at the passport data entered in Tanzania, the researcher will see that the accession 'NCDP-D9' came from 'Station Marc Delorme', a research centre in Côte d'Ivoire. For the accession 'NCDP-D9' the latitude and longitude given as collection site indeed refers to the research station in Côte d'Ivoire, the place where the seednuts were collected. Unfortunately, both fields 'male parent' accession number and 'female parent' accession number have yet to be entered in Tanzania's collection record. Therefore, the researcher would not know the parental accession of 'NCDP-D9'.
- 2) However, looking at all the data from Côte d'Ivoire, a researcher will see that there is only one accession of 'Tacunan Green Dwarf' in Côte d'Ivoire: it is the accession 'SMD NVP3' planted in 1982. Therefore, she/he will conclude that 'NCDP-D9' is the progeny of 'SMD NVP3'. Looking further at the passport data of Côte d'Ivoire germplasm collection, she/he will see that the accession 'SMD NVP3' came from the Davao Research Centre in the Philippines. However, no accession from the Davao Research Centre is yet registered in the CGRD database.
- 3) Nevertheless, if the researcher is clever and persistent, she/he will check all coconut accessions available in the Philippines. She/he will finally find that the original accession of Tacunan Green Dwarf is also available at the Zamboanga Research Centre, as accession 'ZRC PD1' planted in 1977. Based on CGRD, the original collection site of 'ZRC PD1', and therefore 'NCDP-D9' is the village of Tacunan, Davao, 007°04'N, 125°36'E.

This search process will take a researcher at least 15 minutes, and it will require some luck for complete success. Just one piece of information missing in the whole line and, it becomes difficult to make the links. Another option available in CGRD consists in searching directly all the accessions of 'Tacunan Green Dwarf' registered worldwide, but even this does not solve the problem. Both descriptors list and dedicated software have been conceived for managing complete information, and therefore become inefficient if information is incomplete. Data regarding the collection site of the original sample should be recorded in the passport data of each accession.

Nevertheless, the status of georeferencing of coconut accessions (Table 1) compares favourably with other crops. In the CGRD, 74% of the accessions have latitude and longitude information. As there are 1402 accessions but only 710 distinct cultivars/population names, it can be said that more than 80% of the coconut cultivars of the world's germplasm

banks are geo-referenced. As comparison, only 9% of the accessions of six major genebanks of the United State Department of Agriculture have coordinates, although 50% have a locality description (Greene and Hart 1996). Therefore, in the case of the coconut palm, a large amount of geographical information exists. The challenge is to improve its reliability, to make it more easily available, and to use it to improve coconut genetic resources conservation.

Mapping collection sites

Mapping of the locations where accessions were collected was done using GIS tools. This may be defined as a database management system which can simultaneously handle spatial data i.e., maps, or the 'where'– and related, logically-attached, non-spatial, attribute data, and the labels and descriptions of the different areas within a map, or the 'what' (Guarino *et al.* 2002). It is a tool for managing information of any kind according to where it is located (Trewick 1999). The main elements of a GIS are as follows (Guarino 1995; Guarino *et al.* 1999):

- Data input, verification and editing
- Data storage, retrieval and management
- Data manipulation and analysis
- Output

The first mapping was done using the Map module of the Corel Quattro-Pro Software (Bourdeix *et al.* 1999). The data obtained were checked to detect and correct abnormal localizations. As many island countries were involved, a very convenient test was to detect errors such as accessions which appear to have been collected in the open sea ('sea coconuts'). Sometimes the commercial maps were inaccurate and had to be changed. Lists of errors and corrections were exchanged a number of times with most of the countries in order to reach an acceptable level of precision.

This work has not yet been systematically conducted on the entire database. The number of coconut accessions registered in the CGRD database increased from 665 in 1994-1995 to more than 1400 as of 2003. No checking of geographical coordinates was done after 1999 at the database level. There is thus a need to continue this work more efficiently.

The International Plant Genetic Resources Institute (IPGRI) and the International Potato Center (CIP) collaborated and developed specialized GIS software called DIVA-GIS that could be downloaded free from Internet at <http://diva-gis.org/>. DIVA-GIS is dedicated to the analysis of genebank and herbarium databases to elucidate genetic, ecological and geographic patterns in the distribution of crops and wild species.

The maps in Figures 1, 2 and 3 have been made using the 579 locations where coconut accessions have been collected. Climatic information from

various sources can be used in conjunction with these georeferenced accessions to determine the zone of cultivation of the coconut palm. A DIVA-GIS module uses the Food and Agricultural Organization (FAO) Ecocrop database of crop climatic and other environmental requirements together with world climatic surfaces to predict the level of crop adaptation, particularly coconut, over geographic areas. By superimposing the theoretical coconut growing area and the location of collecting sites, it is possible to visualize covered geographical regions. Some areas remain clearly under-represented in the national and international coconut

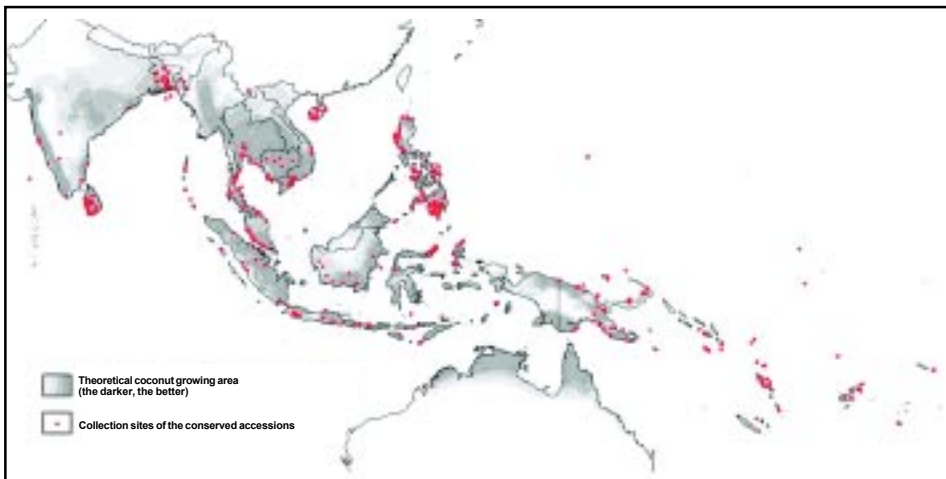


Figure 1. Mapping of the locations where coconut accessions were collected in Asia and the Pacific regions.

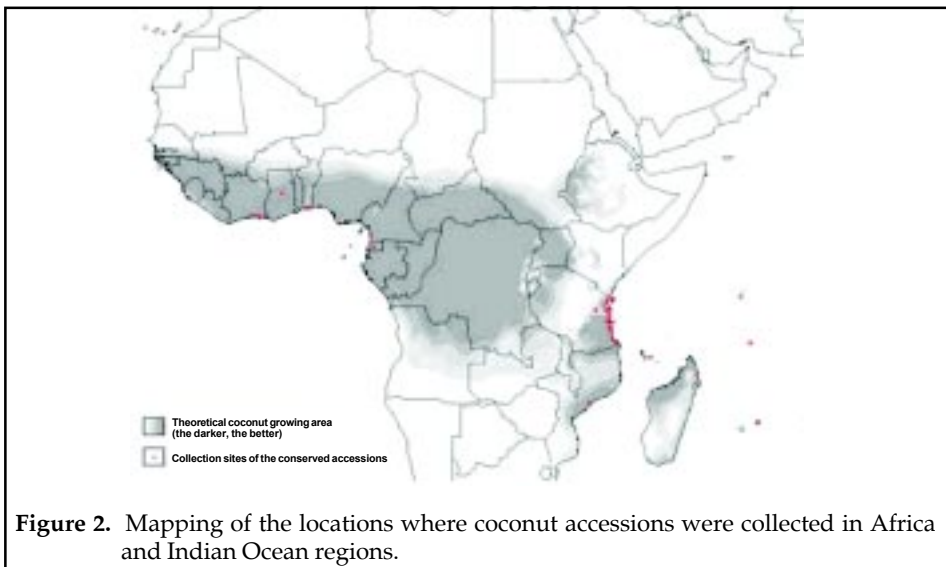


Figure 2. Mapping of the locations where coconut accessions were collected in Africa and Indian Ocean regions.

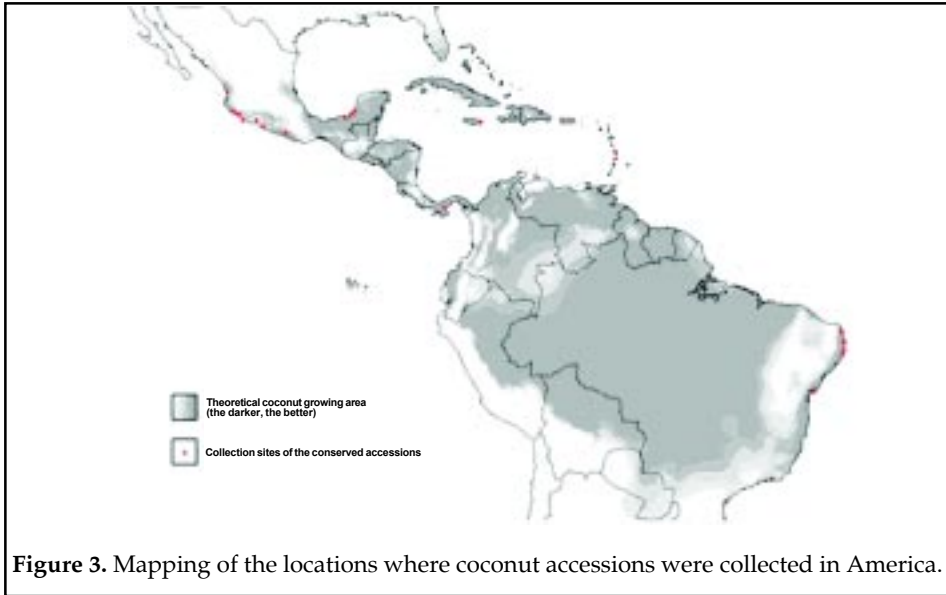


Figure 3. Mapping of the locations where coconut accessions were collected in America.

germplasm centres. These are areas where coconut palm can grow, but where there are no accession points recorded. For details of this analysis see Chapter 2 on 'Status, gaps and strategy in coconut germplasm collecting'.

DIVA-GIS can also be used to check the coordinates of the collection sites in relation to an administrative boundaries database. In effect, this would automate the data-checking process (including locating 'sea coconuts') which used to be done manually. It would therefore be useful in the future to create an interface between CGRD and DIVA-GIS. This would require a module allowing the export of data from the CGRD in a format easily readable by DIVA-GIS.

Mapping morphometric characteristics

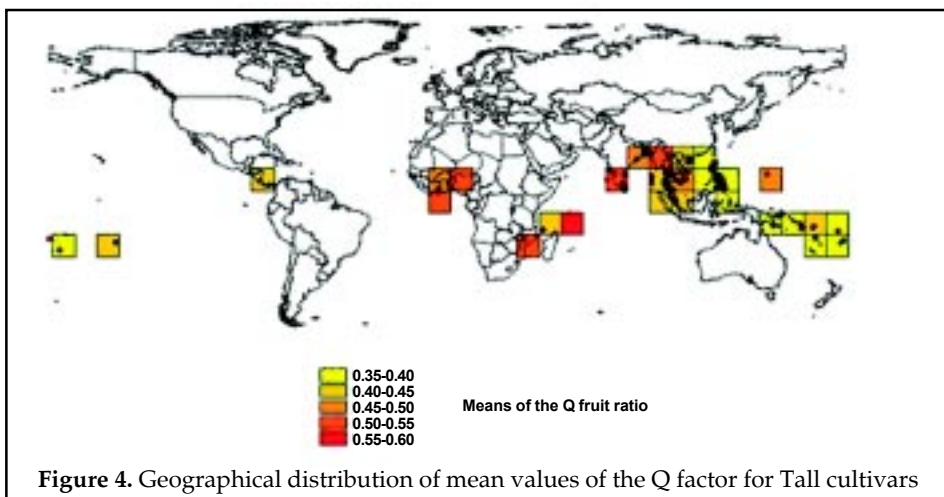
Further analytical functions implemented in DIVA-GIS include mapping the distribution of specific traits and mapping of richness and diversity. Genetic diversity mapping usually begins by dividing the target area (or strata within the target area, e.g. climate zones) into a number of smaller zones, for each of which a measure of diversity can be calculated (Guarino *et al.* 2002). Different geometric, political or socioeconomic spatial units have been used, but ideally, areas of equal shape and size (to reduce the effect of the area on diversity measures) should be employed, for example square grid cells (Nabhan 1991; Ferguson *et al.* 1998).

One of the important parameters describing the quality of the coconut fruit is the Q factor. The Q factor can be defined as the weight of husk

divided by the weight of the fruit without free water (coconut water inside the nut). The quantity of free water is quite variable according to environmental factors (such as rainfall) and the degree of maturity of the fruit. This is the reason why the Q factor is calculated without taking into account the free water. The larger the Q factor, the higher the proportion of husk in the fruit.

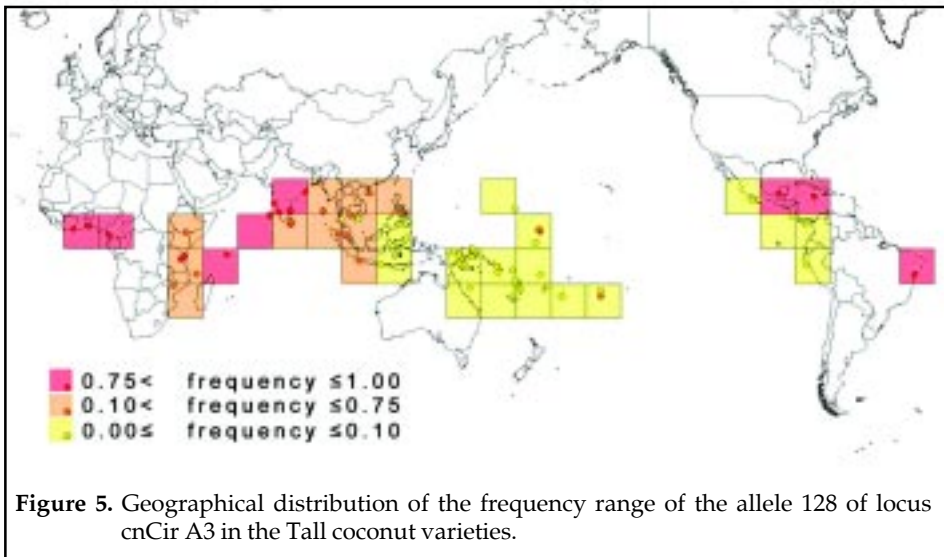
Normally, to get a good estimate of fruit composition, a sample of two fruits is analyzed six times a year over four years on each palm, and this must be done on 30 palms to characterize an accession. In CGRD version 5.0, only 32% of the accessions have data for fruit component analysis. But in practice, the data are even less complete. It can be estimated that at least 20% of the fruit component data available in CGRD 5.0 were derived from the analysis of a single fruit sample harvested at one point. This again reinforces the importance of having complete data in order to carry out an analysis that can be really useful in germplasm conservation, management and use. In any case, all the available georeferenced accessions with fruit analysis data were used to carry out some spatial analyses. The first results obtained were not convincing, because they included both Tall and Dwarf cultivars, which have distinct fruit characteristics.

Nevertheless, there was a clear geographical pattern, based on the geographical distribution of mean values of Q factor, for Tall cultivars (Figure 4). Accessions originating from India and Africa show a higher Q value than accessions from Southeast Asia and the Pacific region.



Mapping molecular markers

New functionality in version 2 of the DIVA-GIS software includes options of mapping based on molecular markers (DNA) data that are illustrated here by an example. Figure 5 presents the geographical repartition of allele 128 of locus *CnCir A3* in the Tall coconut varieties (see Lebrun *et al.*, Chapter 4). Allele 128 is one of the four alleles whose presence at a significant frequency is characteristic of Indo-Atlantic coconuts. Similar (but not identical) pictures could be obtained with any of these markers. Allele 128 is found at a high frequency in an area extending from the Indian sub-continent to the Atlantic coast of America. It is rare or absent in the eastern part of Southeast Asia, in the South Pacific and on the Pacific coasts of America.



The average frequency is based on the samples studied, which do not necessarily reflect the relative abundance of the cultivars. Nevertheless, if we consider the picture in more detail, the intermediate frequencies observed in East Africa result most probably from the introduction of coconuts from South Asia by Austronesian navigators, whose language is still spoken in Madagascar. There is also a transition zone in the western part of Southeast Asia, demonstrating some genetic exchange with South Asia. This may have involved floating, but the activity of Arab merchants who have crossed it for several centuries is probably the cause of most of the exchanges across the Indian Ocean.

The apparently medium frequency of this allele in Sri Lanka is actually an artefact in the Sri Lanka Tall, by far the most dominant cultivar of

this country. The allele 128 frequency is 0.82, but other cultivars, with a round nut were also sampled, which could be the result of hybridization of local coconuts with planting material imported at different periods from Southeast Asia.

Finally, three cultivars with allele 128 frequencies between 0.10 and 0.16 are found in the Pacific Ocean. Considering the three other characteristic (marker) alleles, they are probably not related to the Indo-Atlantic group. The presence of allele 128 at a low frequency in this area is probably a case of homoplasy (i.e., the fact that similar traits – here, fragments with the same length – appear independently by mutation in different genetic groups). Homoplasy is not infrequent in microsatellites.

Mapping for collecting genetic diversity

The genetic diversity mapping described in this Chapter concerns mainly coconut palms that are already in genebanks and their characteristics, such as fruit composition and molecular marker profiles. Another possible field of investigation could be to study coconut populations *in situ* and to map their diversity before the collecting of seednuts or embryos. According to Pernes (1984), the best germplasm collecting programmes are carried out in two stages, with a first exploration and preliminary survey used as a basis for studies that will permit better planning of the second, more systematic campaign. Such a two-step programme was done in Mexico (Zizumbo Villarreal *et al.* 1993) where fruit analyses were made first at 47 localities along the Atlantic and Pacific coasts and in the narrowest part of the country (the States of Oaxaca and Veracruz). Collecting was then carried out in only 19 localities, 90% on the Pacific coast, where the greatest fruit variability was found.

However, most budget appropriations seldom permit such organized programme. It is also necessary to underline the great sensitivity of the coconut to environmental variations. The phenotype of a palm at a certain moment is usually not representative of its genetic value. There is much variation linked to heterogeneity of the field, for example in soil fertility or water availability. The growth of the coconut palm is controlled by various rhythms, depending on internal and external factors. Although it may produce continuously, yields are often irregularly distributed over the year. Biannual cycles are superimposed on this variation. These phenomena are rather frequent in young palms or following droughts (Bourdeix *et al.* 1994). For a proper evaluation, vegetative and yield characteristics of a sufficient number of palms should be measured over a period of several years. Nucé de Lamothe and Wuidart (1982) have emphasized that it is difficult to conduct such a study outside a research station. However, it will be possible to collect leaf samples from surveys, to analyze their DNA and then to sample the field origin where genetic

diversity is greatest. Up to now, we have no example of such a strategy, but it may happen in the near future (see related articles in Chapter 2).

Conclusion

Recent developments in GIS technology for mapping genetic diversity is expected to contribute significantly to identify and fill the gaps in coconut collections, enhance the effectiveness of collecting, better manage them in field genebanks, effectively select parents from geographically distinct regions and expand coconut production through site-genotype matching. Figure 6 provides an analogous map of the number of accessions registered in the germplasm banks of the COGENT member countries.

An additional effort is needed to make coconut descriptors list and related software better adapted to management of incomplete data. The new version of DIVA-GIS software (Version 4.0) can be downloaded free from the internet (<http://www.diva-gis.org>), and is easy to learn and use, and is tailor-made for genetic resources applications. Country-level GIS databases can also be downloaded from this site and these databases can be used together with the genetic resources databases that are being mapped and analyzed. These are files with data on administrative boundaries, country boundaries, and first and second level administrative subdivisions for most countries. For all countries, grids are available for altitude, land cover and population density. DIVA-GIS can be used to check existing coordinates and carry out analyses of characterization and evaluation data. Specific software devoted to the management of genetic resources, such as CGRD for coconut, could be improved by allowing easy exporting of data to DIVA-GIS.

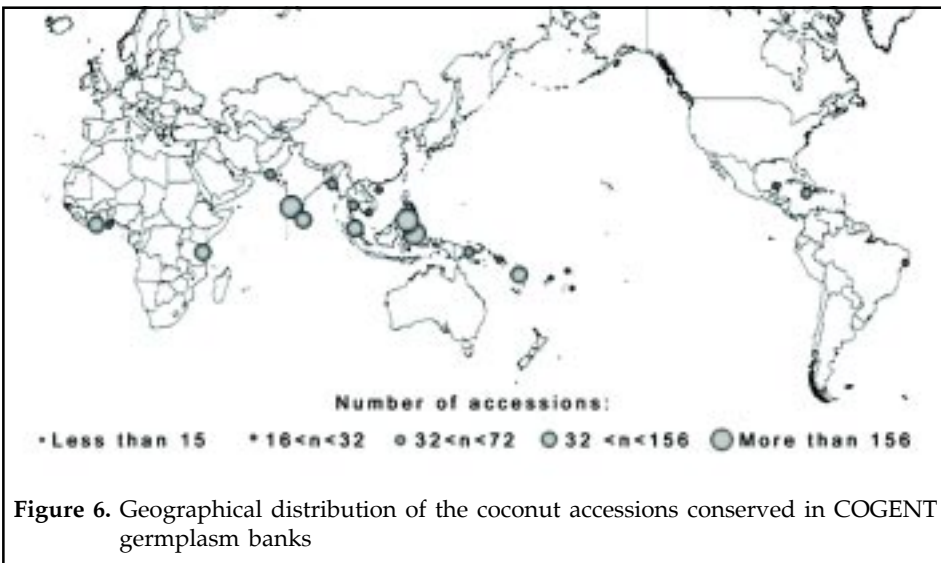


Figure 6. Geographical distribution of the coconut accessions conserved in COGENT germplasm banks

References

- Baudouin, L. 2002. Study on genetic resources status in Hainan, China (visit 12-18 December 2000). CIRAD Report (CP SIC no. 1505), France. 50p.
- Bourdeix, R. 1996. Coconut germplasm in Jamaica, Mexico and Brazil. Mission Report, March 1997. Doc CIRAD-CP no. 548, France. 53p.
- Bourdeix, R. 1997a. Coconut germplasm in Tanzania, Sri Lanka and India. Mission Report, March 1997. Doc CIRAD-CP no. 739, 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, 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, France. 52p.
- Bourdeix, R., Y.P. N'Cho and A. Sangare. 1994. Rythmes de production chez le cocotier Nain (*Cocos nucifera* L.): Etude de l'alternance castration-production comme mode de gestion des champs semenciers. *Agronomie Africaine* VI, 2:77-162.
- Bourdeix, R, L Baudouin, J Ollivier and JP Labouisse. 1999. Consultancy report on coconut collecting strategies (submitted to COGENT/IPGRI). IPGRI, Rome.
- Ferguson, ME, BV Ford-Lloyd, LD Robertson, N Maxted and HJ Newbury. 1998. Mapping the geographical distribution of genetic variation in the genus *Lens* for the enhanced conservation of plant genetic diversity. *Molecular Ecology* 7:1743-1755.
- Greene, SL and T Hart. 1996. Plant genetic resources collections: An opportunity for the evolution of global data set. Third International Conference/Workshop of integrating GIS and Environmental Modeling Santa Fe, New Mexico, USA.
- Guarino, L, A Jarvis, RJ Hijmans and N Maxted. 2002. Geographical Information Systems (GIS) and the conservation and use of plant genetic resources. Pp. 387-404. *In*: JMM Engels, V Ramanatha Rao, AHD Brown and MT Jackson (eds). *Managing plant genetic diversity*. CAB International, Wallingford.
- Guarino, L. 1995. Geographic information systems and remote sensing for the plant germplasm collector. Pp 315-328. *In*: L Guarino, V Ramanatha Rao and R Reid (eds). *Collecting plant genetic Diversity. Technical Guidelines*. CAB International, Wallingford.
- Guarino, L, N Maxted and M Sawkins. 1999. Analysis of geo-referenced

- data and the conservation and use of plant genetic resources. Pp. 1-24. *In*: SL Greene and L Guarino (eds). Linking genetic resources and geography: Emerging strategies for conserving and using crop biodiversity. CSSA Special Publication No. 27. ASA and CSSA, Madison, U.S.A.
- Nabhan, GP. 1991. Wild *Phaseolus* Ecogeography in the Sierra Madre Occidental, Mexico. Systematic and ecogeographic studies of crop gene pools No. 5. IBPGR, Rome, Italy.
- Nuce de Lamothe, MW de and F Wuidart. 1982. L'observation des caractéristiques de développement végétatif, de floraison et de production chez le cocotier. *Oléagineux* 37:290-296.
- Pernès J. 1984. Gestion des ressources génétiques des plantes. Tome 2 : Manuel. Agence de coopération culturelle et technique, Paris ISBN 92-9028-043-3, éditions Lavoisier, France.
- Treweek, J. 1999. Ecological impact assessment. Blackwell Science, Oxford, U.K.
- Zizumbo-Villarreal, DF, R Hernandez and HC Harries. 1993. Coconut varieties in Mexico. *Economic Botany* 47(1): 65-78.