Consolidate microsatellite data on coconut diversity

Final report

LOA APO/08/007

Luc Baudouin

CR.UPR29/2009-2
Mars 2009

BIOVERSITY REPORT FACT SHEET

Library Report No. (Library use only)	

FULL TITLE of REPORT/PROJECT	Consolidate microsatellite data on coconut diversity	
AUTHOR of Commissioned Organization or Contractor (Name and title of person)	Luc Baudouin (Coconut geneticist (CIRAD)	
NAME/ADDRESS of Commissioned Organization or Contractor	CIRAD Biological Systems Department Boulevard de la Lironde TA A-DIR / PS3 34398 Montpellier, Cedex 5 France	
DATE REPORT SUBMITTED	20/03/09	
TYPE OF REPORT i.e. progress/final	Final	
PREVIOUS REPORTS (Please fill in dates and add more lines if necessary) To be completed by BIOVERSITY Coordinator	1st Progress Report	
	Annual Progress Report	
BIOVERSITY LETTER OF AGREEMENT To be completed by BIOVERSITY Coordinator	LOA APO/08/007	
BIOVERSITY PROJECT CODE **To be completed by BIOVERSITY Coordinator**	F03	
BIOVERSITY CONTACT	Maria Luz George	
ABSTRACT (Minimum 100 words)	Microsatellite protocols and data on coconut diversity were consolidated. A database based on 2 marker sets was constructed and its content is described. A survey of reports and scientific publication shows that the microsatellite kit developed in 2002 is effectively adopted by a growing number of research teams. Recommendations are made for improving the kit and for making the database publicly available. The use of the microsatellite kit should be encouraged as a tool for efficient coconut germplasm conservation and use through breeding. Information about the kit and the database should be disseminated through institutional bulletins as well as scientific publications.	
KEYWORDS	Country/Region: Sri Lanka, Ivory Coast, the Philippines, India and Mexico Crop(s): Coconut Subject: Microsatellite data	

Introduction

Efficient germplasm conservation and utilization through plant breeding require an accurate identification of the plant material, in coconut as well as in any other crop. From a breeding point of view, the genetic distance between the intercrossed varieties conditions heterosis (in the F_1) and the creation of variability to select from (in the subsequent generations). From a conservation point of view, coconut is a recalcitrant tree crop, meaning that its seed cannot be conserved on a long period. Conservation in the field is the only practical option (possibly with embryo cryopreservation as a backup). It is highly consuming in terms of space and gene banks cannot afford conserving and using improperly identified germplasm. In particular, rejuvenating collections accessions require proper isolated pollination techniques, since collections have, by definition, a very mixed pollinic environment. The efficiency of these techniques is a key point for an accurate reproduction of the initial germplasm. Finally, even with good isolation techniques, the question of genetic losses during reproduction deserves consideration: a good crossing plan is also necessary.

Molecular markers can address the above issues since they provide a large amount of information usable to identify a cultivar. In the case of the self-pollinating Dwarf cultivars, this information is simply the list of alleles they have at each locus. But Tall cultivars are cross-pollinating and the genotypes vary from tree to tree. However, under traditional management, allele frequencies are more or less stable from generation to generation, thus allowing identifying a cultivar based on its multilocus allele frequency array. Microsatellites were identified as convenient for this purpose due to their high polymorphism, co-dominant nature, reasonable cost and their good repeatability, even with medium quality DNA.

Going back to the issues raised in the first paragraph, microsatellite markers are considered to provide more reliable estimates of genetic distances since they are not influenced by environmental factors (Barker *et al.* 2000). In connection with appropriate software, they are efficient in assigning individuals or groups of individuals to their most probable population of origin (Baudouin *et al.* 2004). They can also be used to detect contamination due to the arrival of unwanted pollen (Baudouin *et al.* 2008a; Lebrun *et al.* 2008). Finally, they allow an estimation of genetic diversity parameters in different generation of the same population, providing an estimate of diversity losses across generations.

Cirad has long been involved in the field of coconut germplasm collection strategies (Bourdeix *et al.* 1999), genebank documentation and information management (Hamelin *et al.* 2005) and coconut molecular markers including RFLP (Lebrun *et al.* 1999), microsatellites and AFLP (Teulat *et al.* 2000). In a project sponsored by IPGRI and COGENT [IPGRI LOA 99/079], conducted in 2000-2002, Cirad developed an initial set of 14 markers, which were applied to 610 individuals (Baudouin and Lebrun 2002). A training session to the use of the resulting kit was held in Montpellier in April 2002 and CIRAD supervised a series of post-training projects conducted 9 COGENT member countries from 2002 to 2005. In a subsequent project (GCP Tier 2), a new set of 30 markers was applied to 1215 individuals. There is a partial overlap between the two studies: 13 of the 14 initial markers were included in the second study and most of the coconut genotypes involved in the first study were also used in the second one.

The main benefit of developing a microsatellite kit for coconut cultivar identification is the opportunity to capitalize the results obtained in different laboratories. The result is an increased accuracy and discriminating power as the data accumulate.

For these reasons, the COGENT coordinator requested CIRAD's expertise in order to consolidate microsatellite data on coconut diversity. The expected outputs of the present consultancy are the following:

Expected outputs

1. Compilation of the microsatellite protocol for coconut (laboratory manual) developed by CIRAD in collaboration with COGENT, including information on the sequences of the 13 or 14 microsatellite markers in the kit, summary of standard alleles, control DNA and molecular weight markers for eventual public dissemination.

2. A data base of microsatellite data related to coconut diversity studies obtained with the microsatellite kit (initial 14 loci or extended to 30 loci);

3. An assessment of this database in terms of cultivars studied so far and of number of individuals;

4. Summary information made available as soon as possible about germplasm characterized elsewhere with other microsatellites;

5. Suggestions about the marker set to be used in further studies;

6. An estimate of numbers of individuals to sample necessary for the following studies;

a) Fingerprinting an accession;

b) Checking the reliability of controlled pollination;

c) Assessing genetic losses during rejuvenation.

Project activities conducted

Planed activities were:

1. Consolidate the microsatellite procedure and data obtained at Cirad on coconut diversity;

Done: See appendices A, B, C and D.

2. Summarize this information in terms of cultivars/accession studied and number of palms already analyzed

Done: See appendix E.

3. Evaluate data obtained through the "post training microsatellite kit project"

Done: See appendix F.

4. Contact Sri Lanka, Cote d'Ivoire, the Philippines, India and Mexico to request summary information about the germplasm characterized in their laboratory.

A questionnaire (see appendix G) was sent to L. Perera (CRI – Sri Lanka, V Arunachalam (CPCRI – India), R. Rivera (PCA – the Philippines), J.L. Konan (CNRA – Côte d'Ivoire), C. Oropeza (CICY – Mexico), Bee Gunn (National

University of Australia) and H. Gomez (Generation Challenge Programme Genotyping Support Services – Mexico). In spite of two reminders, only the latter two responded, resulting in 1 useful response (the GSS had no coconut genotyping activities). Part of the information requested was however obtained from a literature review.

Outputs actually achieved,

The experimental protocols and information on microsatellite markers

Output 1: Compilation of the microsatellite protocol for coconut (laboratory manual) developed by CIRAD in collaboration with COGENT, including information on the sequences of the 13 or 14 microsatellite markers in the kit, summary of standard alleles, control DNA and molecular weight markers for eventual public dissemination.

The marker sets

The available marker set for coconut cultivar characterization results from two projects conducted in 2002 and in 2007. The initial set was made up of 14 markers and the second version had 30 markers: 13 of the initial markers and 17 new markers. Marker CnCir A3 was dropped because its very large allele number made it difficult to score. The markers were selected based on the presence of diversity, both within and between cultivars, on ease of scoring and on repeatability. Contrarily to the initial set, the extended set was chosen among mapped markers and their positions are given in appendix A.

Appendix B presents annotated gel pictures of 26 of the 30 markers. One important point is to ensure comparability of the result from one laboratory to another. This required basing fragment length estimation on control DNA samples rather than simply using a “ladder” as a reference. In fact, different laboratories may use different ladders and, due to variations in migration conditions and in ribonucleotide sequences, the estimated fragment size may differ somewhat from the actual size, which is determined by sequencing. Using the control DNAs T1 and T2 is thus necessary to compensate for these differences.

Experimental protocols

The manual established in 2002 for the microsatellite kit is provided in appendix C. Note that the method presented in the manual uses silver staining. We added a protocol for fluorescence tagging.

Genetic analysis software

Bayesian software *Geneclass 2* (Piry *et al.* 2004)dedicated to population assignment was designed for the coconut microsatellite kit. It can be downloaded at the following URL:

http://www1.montpellier.inra.fr/URLB/

This is the outcome of collaboration between CIRAD and INRA and allows identifying the most probable origin of an individual, among a set of reference populations, each of them being represented by a (possibly small) number of individuals (Baudouin *et al.* 2004). Other software may be useful. Among them,

Structure, like Geneclass 2 is Bayesian assignment software but the philosophy is different: it deduces the information from the data rather than from reference sets. Useful when the number of populations is low.

http://pritch.bsd.uchicago.edu/structure.html

NewHybrids is also very powerful Bayesian assignment software. It is especially useful to detect genetic contamination in rejuvenation.

http://ib.berkeley.edu/labs/slatkin/eriq/software/software.htm#NewHybs

DARwin is convenient to represent genetic diversity in the form of factorial analyses and dendrograms (tree graphs).

http://darwin.cirad.fr/darwin/Home.php

Arlequin performs a lot of population genetics calculations

http://lgb.unige.ch/arlequin/

Popgene may also be used for the same purpose

http://www.ualberta.ca/~fyeh/index.htm

The next version of the database should include specific output formats in order to transfer data to these programs.

The coconut microsatellite database

Output 2: A data base of microsatellite data related to coconut diversity studies obtained with the microsatellite kit (initial 14 loci or extended to 30 loci)

The most important outcome of this consultancy is the construction of a relational database of available coconut microsatellite data. The first version of this database constructed using *Microsoft Access* is described here and a further version will be made available to the public. Its public release will require improvements in order to make it more accessible to users that are not familiar with *Access* and to protect the integrity of the data. In order to prevent inadvertently modifying the original or improper use of unwarranted results, permissions and passwords will need to be installed. A user-friendly front end, links to genetic software and various formatted reports will need to be included. An access through the internet is highly desirable. We suggest transferring the contents of the database to the Tropgene DB platform, managed by CIRAD, which already includes data of various crops, including some coconut data.

Overview of the structure of the database

In its present form, the database is a relational database with 11 tables which are described in appendix D. The *Identity* table is central and allows identifying individuals analyzed in several laboratories. We distributed repeated information relative to the individuals in various tables. Three branches are visible in the figure of appendix D and address three types of questions:

- What type of use can be done with the data of a particular tree is dealt with by sub-table *use,*
- how the sample did reach the laboratory, from where and with which *a priori* information is dealt with by the sub-table *batch*,
- what is the actual place of the individual in the coconut diversity space is dealt with by three sub-tables *origin 1*, *origin 2* and *classification*.

Information contained in these sub-tables was checked by crossing *a priori* information and molecular data.

The table *marker* and three further sub-tables describe all markers used in the data base.

The microsatellite data are presented in two forms: the *data matrix* table has one record per individual and two columns per marker. It is convenient for data inspection. In contrast, the *data list* table has one record for each individual-allele combination. The "amount" column indicates whether the allele is homozygous (1) or heterozygous (0.5). This form is convenient to perform various calculations.

Overview of the contents of the database

Output 3: An assessment of this database in terms of cultivars studied so far and of number of individuals;

Several COGENT steering committee meetings have proposed that CIRAD would act as a repository for microsatellite data. As such, we felt that the database should accept a great variety of results, not only those that are intended to serve for cultivar characterization. This is the purpose of the *use* descriptor, which acts as a filter to select individuals for a particular purpose. This is also why the data base includes more individuals than what is necessary for cultivar characterization purpose. The content of the database in terms of individuals is summarized in table 1, based on the various "use" categories. The columns KIT and GCP corresponds to the numbers of individuals studied respectively with the initial 14 marker set and the extended 30 marker set.

The most important use category is "Diversity", which corresponds to the reference set for cultivar characterization. The detail of this category is given in appendix E where the individuals are grouped into operational taxonomic units (OTU). In most cases, an OTU corresponds to a cultivar, but in a few cases, several very similar cultivars are grouped in an OTU. In other cases a single cultivar may be split into several cultivars: for example, the Panama Tall from Aguadulce and Monagre are indistinguishable on a morpho-agronomic basis (this is thus a single cultivar) but can easily be distinguished on a molecular basis.

There are in average 17 individuals/OTU in the data base. This number varies however greatly due to the history of the molecular marker research at CIRAD: The first studies aimed at studying the genetic relationships of coconut cultivars collected from many places. As a result, the planed sample size was only 5 individuals for Tall accessions and 2 for the self-pollinating Dwarf accessions. This made it possible to study a large number of accessions at a reasonable cost. The present state results from two processes: Firstly, accessions of the same cultivars were grouped whenever possible, secondly, studies conducted at CIRAD for various purposes contributed to adding many individuals from a number of cultivars. Further discussion on the number of individuals per cultivar according to the objectives will be found below. It is however clear that a microsatellite kit and a centralized public database is the best way to take advantage of the progress made on coconut diversity studies in the world.

Table 1: Summary description of the genotypic data

Use	Description	GCP	KIT	Total
Breeding	Improved genotypes from VTT		10	10
Check	Genotypes whose identity needs to be checked further	2	19	21
Diversity	Genotypes usable for germplasm characterization studies. See detail in appendix D (tables D1 to D16).	**1198**	**660**	**1858**
Diversity2	Cultivars represented by only one genotype. These data may be used when more data are available. See detail in appendix D (table D17)	1	10	11
Duplicate	8 duplicate genotypes	4	4	8
Ethnobotany	Genotypes from Vanua Lava, used for ethnobotanical studies		246	246
Excluded	Genotypes which don't correspond to their purported cultivar of origin	10	80	90
Hybrid	Hybrid genotypes		48	48
Insufficient data	These individuals have less than 8 markers genotyped in the initial marker set		178	178
Mapping	Selfed RIT used for gene mapping		85	85
Individuals with data		**1215**	**1342**	**2557**
None	Genotypes collected and extracted, but no genotyping data obtained (a small number of them will be included in the database at a later stage.)			498
Individual identified in the data base				**3055**

Adding more genotypes to the database

As an illustration of the above, we examined the results provided by the 9 member countries involved in the "post-training" study conducted during the 2003/2005 period. Some 380 more individuals representing 40 collection accessions and 16 farmer's varieties could be introduced into the database.

To arrive at this figure, we reviewed the reports and the data files, trying to answer the following questions:

- Did the institution in charge obtain molecular data?
- Was the genetic origin of the tested trees clearly identified?
- Were the results consistent with what is usually found in similar varieties? (Direct examination of allele frequencies, use of Geneclass 2.)
- Was the number of marker used sufficient? (An insufficient number would not prevent to introduce the data in the database, but their effective use for cultivar characterization.)

Our comments are given in appendix E.

As a result, we identified data from CPCRI (India), CRI (Sri Lanka), CICY (Mexico) and probably EMBRAPA (Brazil) for introduction into the database. The results from CNRA (Côte d'Ivoire) and NCDP (Tanzania) can unfortunately not be includes without further examination because of discrepancies between the results provided and those obtained at Cirad for the same (or similar) cultivars. These difficulties may result from the problem mentioned in the "marker sets" section above.

The kit is also being used in independent studies such as the one conducted in the Andaman Islands by CPCRI (Rajesh *et al.* 2008) where 100 individuals represent 26 populations. Another unpublished study was also conducted by the Washington University (St Louis, USA) and the Australian National University fills important gaps in Seychelles Islands and in Madagascar. The 113 individuals involved in this study are already in the database.

In conclusion, the initial objective of making the kit a shared tool for the study of coconut genetic diversity has become a reality: in the next future, the database will include some 2300 useful individuals, of which almost 30% were studied outside of CIRAD.

Other microsatellite markers

Output 4. Summary information made available as soon as possible about germplasm characterized elsewhere with other microsatellite.

In addition to the present microsatellite kit, several microsatellite marker sets were developed:

A set of 41 markers was developed at IACR-Long Ashton Research station (Rivera *et al.* 1999). These markers are identified by codes such as "CNZ#" or "CN#*L*#" where #stands for a number and *L* for a letter. The kit presented here includes three markers of this set: CNZ03, CNZ40 and CNZ 42.

A set of 19 markers was developed at the Scottish Crop Research Institute (Perera *et al.* 2000; Perera *et al.* 2003). These markers are identified by codes such as "CAC#". A subset of these markers were used by (Meerow *et al.* 2003).

More markers were used in linkage mapping studies (Baudouin *et al.* 2006; Lebrun *et al.* 2001).

The characteristics of these markers are given in the papers in reference or in electronic supplementary material. Part of them are already in TropGENE DB.

Optimizing the microsatellite marker set

Output 5. Suggestions about the marker set to be used in further studies.

During the GCP study, we identified a subset of 13 markers (out of 30), which was more discriminating than the initial kit. These markers were selected as follows.

list the markers in decreasing order of "mutual information" between markers and populations

Discard markers which are difficult to score or have null alleles

Whenever two markers map on the same chromosome (at a short distance), chose only the one with the highest mutual information.

In a comparison between the different marker subsets, the percentage of correctly identified individuals rose fro 67% (13 markers of the initial set without CnCir C3') to 72% (13 markers of the new set) with the same number of markers. The percentage was 80% with the whole 30 marker set. The primary aim of this exercise was to show how choosing markers based on mutual information can improve discriminating power at the same cost. On the other hand, one should think twice before discarding markers from a standard kit that has already been used for a large number of individuals. We propose a compromise solution in the "recommendation" section.

Table 2: Assessment of the values of markers in respect to population characterization

The markers sets are the initial set (KIT) and the additional markers of the GCP study (GCP). Data given here were calculated on the present database and differ slightly from the data obtained with the GCP data. The diagnostic alleles are alleles that are especially rare in the Indo-Atlantic cultivars and frequent in the Pacific group (A) or reciprocally (A).

marker	number of cultivars studied	Mutual information	Marker set	Diagnostic alleles	Remarks
Markers identified in the GCP for an improved kit					
CnCir C5	87	2.12	GCP		
CnCirB12	103	1.92	KIT		
Cncir H11	87	1.91	GCP		
CnCirE2	103	1.97	KIT		
CNZ40	87	1.89	GCP	152 (B)	
CnCir 206	87	1.71	GCP		
mEgCIR2739	87	1.70	GCP		
CnCirC7	103	1.56	KIT	157 (A)	
CnCirF2	103	1.56	KIT	193 (B)	
CnCirB6	103	1.46	KIT		
CnCirC12	103	1.42	KIT	167 (B)	
CnCirG11	103	1.32	KIT		
CnCirA3	103	1.24	KIT	228 (B)	
Other markers					
CnCirC3'	102	2.30	KIT		Was in the initial kit, but was considered too difficult to score due to the large number of alleles
CnCir F3'	83	2.25	GCP		Presence of null alleles suspected
CnCir 119	87	1.53	GCP		Closely linked to mEgCIR2739
mEgCIR3400	87	1.40	GCP		
CnCir 2	87	1.37	GCP		
CnCir 215	87	1.34	GCP		Presence of null alleles suspected
CnCir 147	87	1.33	GCP		Closely linked to CnCir 206
CNZ42	87	1.24	GCP		Linked to CnCir 2
CnCir E1	87	1.24	GCP	225 (B)	Closely linked to mEgCIR3400
CnCirH7	103	1.23	KIT		
CnCirA9	103	1.26	KIT		
CnCirE10	103	1.25	KIT	244 (B)	
mEgCIR3750	87	0.94	GCP	125 (B)	Closely linked to CnCir H11
CnCirH4'	103	0.81	KIT		Closely linked to B12
CnCirE12	103	0.99	KIT	174 (B)	
CNZ03	87	0.79	GCP	89 (A)	Closely linked to CnCir H11
CnCir 126	87	0.28	GCP		
CnCir I4	87	0.25	GCP		

The mutual information quantity between a marker and the populations evaluates the discriminating power of the marker. For locus L, this quantity is equal to:

$$I_L(pop;all) = h(pop) + h_L(all) - h_L(pop,all)$$

(Battail 1997) where

- $h(pop) = -\sum_i p_i \log p_i$. Where p_i is the probability of population i. We consider populations as equiprobable, thus, p_i is $1/n$, where n is the number of populations and $h(pop)=\log n$.

- $h(all) = -\sum_j p_{Lj} \log p_{Lj}$ where p_{Lj} is the frequency of allele L_j of in the whole database.

- Finally, $h_L(pop,all) = -\sum_i \sum_j \left(p_{Lij}/n\right) \log p_{Lij}/n$ where p_{Lij} is the frequency of allele L_j in population i. The base of the logarithms is 2, and mutual information is thus measured in "Shannons" (Sh).

Table 2 above lists the performances of the markers in terms of mutual information. Using mutual information, it is possible to tailor specific marker sets for a specific purpose. To illustrate this, we included "diagnostic alleles" whose frequencies vary greatly according to the genetic group (Indo-Atlantic and Pacific). To discriminate the two main groups, it is sufficient to consider only 3 to 5 loci with such alleles.

Number of individuals to be sampled

Output 6. An estimate of numbers of individuals to sample necessary for the following studies:

a) Fingerprinting an accession;

b) Checking the reliability of controlled pollination;

c) Assessing genetic losses during rejuvenation.

Fingerprinting an accession

The Bayesian assignment procedures such a *Geneclass 2* rely on multi-dimensional information relative to several markers rather than on an extreme precision on each marker. As a result, more than 2/3 of the individuals of the database are already assigned to their own population in the present state. Cumulating the information of several individuals from the same populations improves greatly the precision. Based on our experience, a sample of 10 to 15 individuals is sufficient to characterize a population accurately.

From another point of view, the number of individuals from a cultivar in the database is at least as important as the number of individuals sampled from the tested accession. There is thus certainly a need to reinforce the representativeness of the database. A reasonable objective would be to have 10

individuals for a Tall cultivar and 5 for a Dwarf. To meet these standards, about 260 more coconut trees should be sampled and genotyped.

Checking the reliability of controlled pollination

Software NewHybrids is efficient in identifying population hybrids when there are just 2 putative parent populations (for example the accession in collection and the local Tall) and if they differ sufficiently from each other. A rule of the thumb is that if one wants to detect accessions with a percentage of illegitimates above 1/*n* the sample size should be at least 3 times *n* (for example 60 individuals are needed to have a probability 0.95 to detect at least one illegitimate tree if the illegitimacy rate is just 5%).

Assessing genetic losses during rejuvenation

Rejuvenation of an accession inevitably results in a random deviation from the genetic structure of the original accession and in loss of genetic diversity. The main factors responsible for this loss are the number of genitors involved, their degree of relatedness and the choice of the crossing plan. The deviation can be measured in terms of variations of allele frequencies, the genetic loss can be measured in terms of reduction of Nei's gene diversity $h = 1 - \sum p_{ij}^2$ and in terms of loss of alleles. We suggest assessing genetic losses in a small number of cases. Such study can be associated with the test of the reliability of controlled pollinations. It involves genotyping an equal number (60) of individuals from the initial and from the rejuvenated accession.

Constraints and Opportunities and how these were addressed

The author's computer hard disk crashed at the end of January. Although the coconut microsatellite data themselves were not lost, the database had to be reconstructed and the assessed again.

A questionnaire was sent to 7 selected researchers. In spite of two reminders, only two responses were obtained (from non COGENT members). To obtain the requested information, the author had to rely on a literature review.

Recommendations

An important effort of coconut germplasm conservation has been undertaken by the COGENT member institutions. In order to maximize the benefit of this effort, it is necessary to ensure the representativeness of the conserved accessions, to have a better knowledge of their genetic relations and to assess the quality of the multiplication techniques. The microsatellite kit for coconut cultivar characterization has been designed as a tool to address these issues. A wide dissemination of the results obtained with the kit may be an asset for the coconut genetic research community. In consequence, we recommend to

- **Make the contents of the database presented here publicly accessible**

For convenience, we developed this database using *Microsoft Access*. In its present state, it can be used by persons having a good knowledge of *Access* and

of its structure. To make it accessible to a wider public, it would certainly possible to create a user-friendly interface under *Access*. A better solution would be to use a platform accessible through the internet. We thus recommend incorporating its contents into TropGENE DB, which manages data for banana, cocoa, coffee, cotton, oil palm, rice, rubber tree and sugarcane. In its present version, accessible at the URL http://tropgenedb.cirad.fr/index.html , some coconut data are already present. However, the interface is being modified to enhance the access to molecular data.

- **Introduce 480 more genotypic data into the database**

The post-training data from CPCRI, CRI, CICY and EMBRAPA can be introduced into the database.

- **Optimize the marker kit**

The choice of a new marker set needs to be considered carefully: There is indeed a benefit in using an optimal marker set for *future analyzes*. The benefit is however less obvious for past analyzes conducted with the initial kit: A number of cultivars were analyzed only with the initial kit and would have only 8 loci in common with the new analyzes if the set presented in table 2 was adopted. A good compromise would be to discard markers CnCirC3' (too difficult to score and CnCirH4' (closely linked to CnCirB12, with a low polymorphism) and to add three highly discriminating markers from the GCP set: CnCir C5, Cncir H11 and CNZ40. The new kit would thus have 15 markers.

- **Promote further characterization of a selected set of cultivars that are still under-represented in the database**

Five Dwarf cultivars have less than 5 individuals in the database and 51 Talls have less than 10. Updating data would require about 260 analyses.

- **Promote genetic studies relating the genetic relationships between cultivars as shown by microsatellites and useful traits (production, quality, stress resistance)**

The distributions of molecular markers and of genetically determined traits result from the history of its diversification and dissemination. They are thus probably related and some of these relations are already known. For example, it is certainly not by chance that the *niu kafa* predominates in the Indo-Atlantic predominates in the Indo-Atlantic cultivars and the *niu vai* in the Pacific ones (Lebrun *et al.* 2005). Likewise, paralleling microsatellite studies and field testing strongly suggests a South-East Asian origin for most resistance factors to Lethal Yellowing Diseases (Baudouin *et al.* 2008b). Similar studies on various traits would help devising sounder breeding strategies.

- **Encourage the use of good pollination practices through assessing trueness-to-type in a number of rejuvenated accession**

Such studies should be conducted in at least 10 cases (e.g. one Tall and one Dwarf cultivar in each ICG) and would consist in identifying possible off-types in a sample of 60 progenies.

- **Assess diversity losses (genetic erosion) in a limited set of rejuvenated accessions**

Such studies should be coupled with trueness-to-type studies: 60 individuals of the parental populations would be sampled. They would be conducted in no more than 5 Tall cultivars.

- **Disseminate information on the microsatellite kit and on the reference database**

This information should be disseminated through Bioversity International and/or COGENT publications as well as in at least one peer-reviewed article. The objective is to describe encourage the use of the kit as a common tool for coconut geneticists, germplasm managers and plant breeders. It is also to foster feedback to the database in the form of new genotypes studied.

Bibliography.

Barker, J. H. A., B. Teulat, P. Lebrun, L. Baudouin, A. Karp *et al.*, 2000 Coconut microsatellite markers and their use in genebank organisation and genetic mapping, pp. in *Plant and Animal Genomes Conference. 8, 2000-01-09/2000-01-12,*. s.n., San Diego, USA.

Battail, G., 1997 *Théorie de l'information, Application aux techniques de communication*. Masson, Paris.

Baudouin, L., and P. Lebrun, 2002 The development of a microsatellite kit for use with coconuts : Final report IPGRI n° 99/079, pp. 52 p. in *The development of a microsatellite kit for use with coconuts*. CIRAD-CP, Montpellier.

Baudouin, L., P. Lebrun, A. Berger, W. Myrie, B. Been *et al.*, 2008a The Panama Tall and the Maypan hybrid coconut in Jamaica : Did genetic contamination cause a loss of resistance to Lethal Yellowing? Euphytica **161:** 353-360.

Baudouin, L., P. Lebrun, J.-L. Konan, E. Ritter, A. Berger *et al.*, 2006 QTL analysis of fruit components in the progeny of a Rennell Island Tall coconut (Cocos nucifera L.) individual. Theoretical and applied genetics **112:** 258-268.

Baudouin, L., R. Philippe, S. K. Dery, R. N. Quaicoe and M. Dollet, 2008b Survol général de la recherche et de l'expérimentation sur les variétés résistantes/tolérantes au JM [Résumé] = General overview of genetic research & experimentation on Coconut varieties tolerant/resistant to LY, pp. 1 p. in *International Workshop on Lethal Yellowing Disease on Coconut* edited by F. C. M. CIRAD. s.n., s.l.

Baudouin, L., S. Piry and J. M. Cornuet, 2004 Analytical bayesian approach for assigning individuals to populations. Journal of heredity **95:** 217-224.

Bourdeix, R., L. Baudouin, J. Ollivier and J.-P. Labouisse, 1999 Consultancy report on coconut collecting strategies : ADB-CGRNAP report, pp. 110 p. CIRAD-CP, Montpellier.

Hamelin, C., R. Bourdeix and L. Baudouin, 2005 The International Coconut Genetic Resources Database, pp. 427-438 in *Coconut Genetic Resources*, edited by P. Batugal, V. Ramanatha Rao and J. Oliver. IPGRI-APO, Serdang.

Lebrun, P., L. Baudouin, R. Bourdeix, J.-L. Konan, J. H. A. Barker *et al.*, 2001 Construction of a linkage map of the Rennell Island Tall coconut type (Cocos nucifera L.) and QTL analysis for yield characters. Génome **44:** 962-970.

Lebrun, P., L. Baudouin, W. Myrie, A. Berger and M. Dollet, 2008 Recent lethal yellowing outbreak: Why is the Malayan Yellow Dwarf Coconut no longer resistant in Jamaica? Tree genetics and genomes **4:** 125-131.

Lebrun, P., A. Berger, T. Hodgkin and L. Baudouin, 2005 Biochemical and Molecular Methods for characterizinng coconut diversity, pp. 225-250 in *Coconut Genetic Resources*, edited by P. Batugal, V. Ramanatha Rao and J. Oliver. IPGRI-APO, Serdang.

Lebrun, P., L. Grivet and L. Baudouin, 1999 Use of RFLP markers to study the diversity of the coconut palm, pp. p.73-87 in *Current advances in coconut biotechnology*, edited by V. J. L. e. Oropeza C. (ed.), Ashburner G.R. (ed.), Cardena R. (ed.), Santamaria J.M. (ed.); CICY. Kluwer Academic, Dordrecht.

Meerow, A. W., R. J. Wisser, J. Steven Brown, D. N. Kuhn, R. J. Schnell *et al.*, 2003 Analysis of genetic diversity and population structure within Florida
coconut (Cocos nucifera L.) germplasm using microsatellite DNA,
with special emphasis on the Fiji Dwarf cultivar. Theoretical and applied genetics **106:** 715-726.

Perera, L., J. R. Russell, J. Provan and W. Powell, 2000 Use of microsatellite DNA markers to investigate the level of genetic diversity and population genetic structure of coconut (Cocos nucifera L.). Genome **43:** 15-21.

Perera, L., J. R. Russell, J. Provan and W. Powell, 2003 Studying genetic relationships among coconut varieties/populations using microsatellite markers. Euphytica **132:** 121-128.

Piry, S., A. Alapetite, J. M. Cornuet, D. Paetkau, L. Baudouin *et al.*, 2004 GENECLASS2: A software for genetic assignment and first-generation migrant detection. Journal of heredity **95:** 536-539.

Rajesh, M., P. Nagarajan, B. Jerard, V. Arunachalam and R. Dhanapal, 2008 Microsatellite variability of coconut accessions (*Cocos nucifera* L.) from Andaman and Nicobar Islands. Current Science **94:** 1627-1631.

Rivera, R., K. J. Edwards, J. H. A. Barker, G. M. Arnold, G. Ayad *et al.*, 1999 Isolation and characterization of polymorphic
microsatellites in *Cocos nucifera* L. Genome **42:** 668-675.

Teulat, B., C. Aldam, R. Trehin, P. Lebrun, J. H. A. Barker *et al.*, 2000 An analysis of genetic diversity in coconut (Cocos nucifera) populations from across the geographic range using sequence-tagged microsatellites (SSRs) and AFLPs. Theoretical and Applied Genetics **100:** 764-771.