Repères

# Genetic Diversity of Cultivated Tropical Plants

*Scientific editors*

Perla Hamon, Marc Seguin, Xavier Perrier
and Jean Christophe Glaszmann

CIRAD

SPI

# Genetic Diversity of Cultivated Tropical Plants

REPÈRES

# Genetic Diversity of Cultivated Tropical Plants

*Scientific editors*
Perla Hamon
Marc Seguin
Xavier Perrier
Jean Christophe Glaszmann

CIRAD
Centre de coopération internationale en recherche agronomique pour le développment, France

Science Publishers, Inc.
Enfield (NH) Plymouth, UK

Joint Publication: Science Publishers, Inc., USA
and CIRAD, France

Exclusive selling rights in North America and Asia: Science Publishers, Inc.

CIRAD (Centre de coopération internationale en recherche
agronomique pour le devéloppement)
Avenue Agropolis
34398 Montpellier Cedex 5, France
Internet site: http://www.cirad.fr

© CIRAD, Paris 2003
ISBN (papier) : 978-2-87614-541-2
ISBN (pdf) : 978-2-87614-825-3
DOI : 10.19182/agritrop/00212

# Preface

The development of agriculture marked the beginnings of plant improvement. It started from the Paleolithic, about ten to twelve thousand years ago, when communities that had till then drawn their subsistence from hunting, fishing, and gathering began to form settlements. Between that era and the middle of the 19th century, almost all of the plants cultivated today were domesticated. These plants were transformed from wild species to those of a traditional ecotype or variety by progressively acquiring a set of characteristics more or less different from those of their ancestors. This is called the 'domestication syndrome'. Domestication took place either within closely limited areas—as in the case of maize and potato, in pre-Colombian times—or across several continents, as in the case of rice and sorghum. Cultivated species thus evolved under the pressures of natural and human selection linked to cumulative as well as divergent processes, of permanent introgression and diversification. Humans exploited this diversity by improving first the resources of their immediate environment and then, from the 16th century onwards, those resources made accessible by the rapid development of trade and transport. An infinite number of cultivation systems were thus developed, each of which was an original response to the particular needs and constraints of a community.

Till the 1930s, modern improvement of varieties based on selection procedures relied solely on the use of traditional varieties or ecotypes and the notion of genetic resources was limited to cultivated plants. Following the works of Vavilov, this notion was extended to related wild species and then to increasingly distant genera. Nevertheless, the importance of genetic resources became apparent only with the threat of their extinction. This phenomenon was particularly marked in Europe after World War II. Indeed, the natural environmental balance was seriously disturbed by the considerable extension of urban areas at the expense of farmland, the cultivation of new areas that might be unsuitable or fragile, with or without deforestation, and the degradation of vast natural regions that had been preserved until that time.

Conscious of this risk, the international scientific community was mobilized during the 1960s and 1970s to collect and conserve as many genetic resources as possible, giving priority to food species, which have a major economic value on the global scale. A large number of collections were thus assembled throughout the world. The FAO published a synthesis on this subject, the *Report on the state of the world's plant genetic resources for food and agriculture* (1997), on the occasion of the International Conference at Leipzig in June 1996. At present, most of these collections have become so large that they are hard to maintain and the taxonomic characterization of their accessions has become difficult. Their management is now an acute problem. It has become necessary to find a balance between conservation, evaluation, and use of genetic resources. Without evaluation, no rational use is possible. Without utility, the conservation is not justified from the point of view of institutions that are not specifically mandated to maintain them.

To respond to these concerns, Frankel and Brown introduced in the 1980s the concept of *core collection*, which they defined as a limited sample of accessions taken from a larger collection, called the base collection, and chosen to best represent the existing spectrum of diversity. Such a core collection can serve several purposes. It allows scientists to identify the material that ought to be conserved as a priority, but also allows reasonably good access to the genetic diversity available in the base collection, which facilitates the search for new sources of useful characteristics for selection.

In general, the base collections contain passport data, including the geographic and ecological origin of the accessions. This information is complemented as needed by morphotaxonomic, agronomic, and genetic data generated from biochemical markers (e.g., isozymes, polyphenols) or molecular markers (e.g., RFLP, RAPD).

During the compilation of a core collection, the useful characteristics are most important for the breeder, but they are difficult to evaluate directly and may conceal complex genetic controls. Moreover, this evaluation may be burdened by a difficulty in predicting the constraints that may weigh on the culture in the future. On the other hand, molecular genetic markers have no direct utility, but they express general structures of diversity, which may in turn serve as a basis for constituting a core collection. Markers sometimes reveal groups of accessions partly isolated from each other, which may have fixed distinct alleles for useful characteristics, by foundation effect, by genetic drift, or even under the action of various pressures of selection.

The relationships between the two levels of variability—that of molecular genetic markers, probably mostly neutral, and those of agronomic characters, generally more complex and subject to natural or human selection—are poorly understood. They are of central concern to the scientific community involved in the management of genetic resources, as appears in the conclusions of the colloquium on plant and animal genetic resources and the methodologies of their study and management, organized by INRA and BRG at Montpellier, in September 1993.

These relationships probably vary as a function of the population structure. They are generally strong if there is a serious gametic disequilibrium. To proceed from this reflection, however, we must tackle specific questions: Are the different types of molecular markers equivalent? Are strong structures at the molecular scale, which control the gametic disequilibrium generalized on the genome as a whole, systematically associated with strong structures for agronomic characters? Do the structures at these two levels therefore coincide?

Several methods have been proposed to construct the core collections. As a general rule, they are based on a stratification of the base collection and then on a random sample within each group thus defined, according to various modalities. When data are available for morphoagronomic evaluation, it is useful to consider the structure of the genetic diversity established with the help of genetic markers in order to maximize the agronomic diversity and minimize the loss of alleles that are rare overall but occasionally significant.

The constitution of a core collection must therefore be founded on an excellent description of the populations and on a sound understanding of their structure. Statistical tools are indispensable here. For each type of marker, it is important to use a measure of difference between the relevant taxonomic units (individuals or populations) from the perspective of their mathematical properties and their interpretation in genetic terms. In the same way, the choice of algorithms of representation of dissimilarities must be based on an equilibrium between maximum efficiency and minimum complexity, in order to deal with large tables covering more than 100 individuals. By comparing the structures observed with the help of various types of descriptors, we can envisage overall the organization of a plant's genetic diversity. This question can be tackled in various ways, among others, by looking for structures common to two or several trees. In these conditions, how can we construct consensus trees or common minimum trees? What is their reliability and their biological significance?

This work provides the elements of an answer to these questions, beginning with the study of eleven types of plants, chosen so as to cover a wide range of biological characteristics (perennial, annual, autogamous, allogamous, etc.): asian rice, banana, cacao, cassava, citrus, coconut, coffee, pearl millet, rubber tree, sorghum, and sugarcane. Three methodological chapters complement these studies. The first is devoted to the use of biological and molecular markers to analyse the diversity of collections, the second addresses data analysis, and the third describes a method for constituting core collections based on maximization of variability.

# List of Abbreviations

| | |
|---|---|
| AAT | aspartate aminotransferase |
| ADH | alcohol dehydrogenase |
| AFLP™ | amplified fragment length polymorphism |
| AMP | aminopeptidase |
| BT | binary table |
| CA | correspondence analysis |
| CAT | catalase |
| cDNA | complementary DNA |
| cpDNA | chloroplastic DNA |
| DAF | DNA amplification fingerprinting |
| DNA | deoxyribonucleic acid |
| EST | esterase |
| FISH | fluorescent *in situ* hybridization |
| GISH | genomic *in situ* hybridization |
| GOT | glutamate oxaloacetate transaminase |
| HC | hierarchical clustering |
| ICD/IDH | isocitrate dehydrogenase |
| ISSR | inter-simple sequence repeat |
| LAP | leucine aminopeptidase |
| MCA | multiple correspondence analysis |
| MDH | malate dehydrogenase |
| MFA | multiple factorial analysis |
| mtDNA | mitochondrial DNA |
| PCA | principal components analysis |
| PCoA | principal coordinates analysis |
| PCR | polymerase chain reaction |
| PER | peroxydase |

| | |
|---|---|
| PGD | 6-phosphogluconate dehydrogenase |
| PGI | phosphogluco-isomerase |
| PGM | phosphoglucomutase |
| PIC | polymorphism information content |
| QTL | quantitative trait loci |
| RAPD | randomly amplified polymorphic DNA |
| rDNA | ribosomal DNA |
| RFLP | restriction fragment length polymorphism |
| RNA | ribonucleic acid |
| SKDH | shikimate dehydrogenase |
| SSR | simple sequence repeat |
| STMS | sequence tagged microsatellite site |
| UPGMA | unweighted pair group method with average |
| VNRT | variable number of tandem repeat |
| WPGMA | weighted pair group method using average |

# List of Contributors


Albar, Laurence, IRD, BP 64501, 34394 Montpellier Cedex 5, France

Anthony, François, CATIE, Apartado 59, 7170 Turrialba, Costa Rica

Baudouin, Luc, CIRAD, département des cultures pérennes, TA 80/03, 34398 Montpellier Cedex 5, France

Berthaud, Julien, IRD, BP 64501, 34394 Montpellier Cedex 5, France

Bezançon, Gilles, IRD, BP 64501, 34394 Montpellier Cedex 5, France

Bonnot, François, CIRAD, département des cultures pérennes, TA 80/03, 34398 Montpellier Cedex 5, France

Bourdeix, Roland, CIRAD, département des cultures pérennes, représentation du CIRAD, résidence les Acacias, 01 BP 6483, Abidjan 01, Côte d'Ivoire

Carreel, Françoise, CIRAD, département des productions fruitières et horticoles, station de Neufchâteau, Sainte-Marie, 97130 Capesterre-Belle-Eau, Guadeloupe

Chantereau, Jacques, CIRAD, département des cultures annuelles, TA 70/01, 34398 Montpellier Cedex 5, France

Clément-Demange André, CIRAD, département des cultures pérennes, TA 80/PS1, 34398 Montpellier Cedex 5, France

Colombo Carlos, Instituto Agronômico de Campinas, Caixa postal 28, 13001-970 Campinas, SP, Brazil

Combes, Marie-Christine, IRD, BP 64501, 34394 Montpellier Cedex 5, France

Courtois, Brigitte, CIRAD, département d'amélioration des méthodes pour l'innovation scientifique, TA 40/03, 34398 Montpellier Cedex 5, France

Deu, Monique, CIRAD, département des cultures annuelles, TA 70/03, 34398 Montpellier Cedex 5, France

D'Hont, Angélique, CIRAD, département des cultures annuelles, TA 40/03, 34398 Montpellier Cedex 5, France

Dubois, Cécile, CIRAD, département des productions fruitières et horticoles, TA 50/PS4, 34398 Montpellier Cedex 5, France

Dussert, Stéphane, IRD, BP 64501, 34394 Montpellier Cedex 5, France

Flori, Albert, CIRAD, département des cultures pérennes, TA 80/03, 34398 Montpellier Cedex 5, France

Ghesquière, Alain, IRD, BP 64501, 34394 Montpellier Cedex 5, France

Glaszmann, Jean Christophe, CIRAD, département d'amélioration des méthodes pour l'innovation scientifique, TA 40/03, 34398 Montpellier Cedex 5, France

Grivet, Laurent, CIRAD, département des cultures annuelles, TA 40/03, 34398 Montpellier Cedex 5, France

Hamon, Perla, Centre universitaire, rue du Dr Georges Salan, 30000 Nîmes, France

Hamon, Serge, IRD, BP 64501, 34394 Montpellier Cedex 5, France

Horry, Jean-Pierre, BP 153, 97202 Fort de France, Martinique

Jacquemond, Camille, INRA, station de recherche agronomique de San Giuliano, 20230 San Nicolao, France

Jacquot, Michel, CIRAD, département des cultures annuelles, TA 70/03, 34398 Montpellier Cedex 5, France

Jannoo, Nazeema, Centro de biologia molecular e engenharia genetica, Universidade Estadual de Campinas, Caixa Postal 6109-13083-970 Campinas SP, Brasil

Jenny, Christophe, CIRAD, département des productions fruitières et horticoles, station de Neufchâteau, Sainte-Marie, 97130 Capesterre-Belle-Eau, Guadeloupe

Lanaud, Claire, CIRAD, département d'amélioration des méthodes pour l'innovation scientifique, TA 40/03, 34398 Montpellier Cedex 5, France

Lashermes, Philippe, IRD, BP 64501, 34394 Montpellier Cedex 5, France

Lebrun, Patricia, CIRAD, département des cultures pérennes, TA 80/03, 34398 Montpellier Cedex 5, France

Legnaté, Hyacinthe, CNRA, 01 BP 6483, Abidjan 01, Côte d'Ivoire

Luce, Claude, CIRAD, département des cultures annuelles, Station de Roujol, 97170 Petit Bourg, Guadeloupe

Luro, François, INRA, station de recherche agronomique de San Giuliano, 20230 San Nicolao, France

Montagnon, Christophe, CIRAD, département des cultures pérennes, représentation du CIRAD, résidence les Acacias, 01 BP 6483, Abidjan 01, Côte d'Ivoire

Motamayor, Juan-Carlos, USDA-ARS, Subtropical Horticultural Research Station 13601 Old Cutler Road, Miami, FL 33158, USA

N'cho, Yavo-Pierre, CNRA, station Marc Delorme, 07 BP 13, Abidjan 07, Côte d'Ivoire

Noirot, Michel, IRD, BP 64501, 34394 Montpellier Cedex 5, France

Noyer, Jean-Louis, CIRAD, département d'amélioration des méthodes pour l'innovation scientifique, TA 40/03, 34398 Montpellier Cedex 5, France

Ollitrault, Patrick, CIRAD, département des productions fruitières et horticoles, TA 50/PS4, 34398 Montpellier Cedex 5, France

Perrier, Xavier, CIRAD, département des productions fruitières et horticoles, TA 50/PS4, 34398 Montpellier Cedex 5, France

Raffaillac, Jean-Pierre, IRD, CIRAD, département d'amélioration des méthodes pour l'innovation scientifique, laboratoire Ecotrop, 34398 Montpellier Cedex 5, France

Second, Gérard, mission IRD, Whimper 442 y Corua, apartado 17-12-857, Quito, Ecuador

Seguin, Marc, CIRAD, département des cultures pérennes, TA 80/03, 34398 Montpellier Cedex 5, France

Sounigo, Olivier, CIRAD, département des cultures pérennes, TA 80/PS3, 34398 Montpellier Cedex 5, France

Tézenas du Montcel, Hugues, CIRAD, département des productions fruitières et horticoles, BP 153, 97202 Fort-de-France, Martinique

Tomekpe, Kodjo, CIRAD, département des productions fruitières et horticoles, BP 2995, Douala, Cameroun

Trouslot, Pierre, IRD, BP 64501, 34394 Montpellier Cedex 5, France

# Contents


*Preface* *v*

*List of abbreviations* *ix*

*List of contributors* *xi*

Biochemical and molecular markers 1
Laurent Grivet and Jean-Louis Noyer

Methods of data analysis 31
Xavier Perrier, Albert Flori and François Bonnot

A method for building core collections 65
Michel Noirot, François Anthony, Stephane Dussert and Serge Hamon

Asian rice 77
Jean-Christophe Glaszmann, Laurent Grivet, Brigitte Courtois, Jean-Louis Noyer, Claude Luce, Michel Jacquot, Laurence Albar, Alain Ghesquière and Gérard Second

Banana 99
Christophe Jenny, Françoise Carreel, Kodjo Tomekpe, Xavier Perrier, Cécile Dubois, Jean-Pierre Horry and Hugues Tézenas du Montcel

Cacao 125
Claire Lanaud, Juan-Carlos Motamayor and Olivier Sounigo

Cassava 157
Gérard Second, Jean-Pierre Raffaillac and Carlos Colombo

Citrus 193
Patrick Ollitrault, Camille Jacquemond, Cécile Dubois and François Luro

Coconut 219
Patricia Lebrun, Yavo-Pierre N'cho, Roland Bourdeix and Luc Baudouin

Coffee (*Coffea canephora*) 239
Stéphane Dussert, Philippe Lashermes, François Anthony, Christophe Montagnon, Pierre Trouslot, Marie-Christine Combes, Julien Berthaud, Michel Noirot and Serge Hamon

Pearl millet 259
Gilles Bezançon

Rubber tree (*Hevea brasiliensis*) 277
Marc Seguin, Albert Flori, Hyacinthe Legnaté and André Clément-Demange

Sorghum 307
Monique Deu, Perla Hamon, François Bonnot and Jacques Chantereau

Sugarcane 337
Jean-Christophe Glaszmann, Nazeema Jannoo, Laurent Grivet and Angélique D'Hont

# Biochemical and Molecular Markers

Laurent Grivet and Jean-Louis Noyer

Biochemical and molecular markers have several applications in plant genetics. They allow us to observe closely the polymorphism of DNA sequences at a certain number of sites or loci spread across the genome. More precisely, biochemical markers reveal the polymorphism of sequences of certain proteins and thus, indirectly, the polymorphism of DNA sequences from which they are translated. Molecular markers directly reveal the polymorphism of DNA, the targeted sequences corresponding or not corresponding to the coding sequences.

Because of their properties, biochemical and molecular markers are a powerful tool to study the structure of genetic variability within a species and trace its evolutionary history. They are relatively unaffected by the environment or genetic basis. We can thus use them to compare individuals that were studied in different experiments or that are present in different collections. It is generally acknowledged that biochemical and molecular markers reveal a neutral polymorphism, i.e., one that is not subject to selection. They are relatively insensitive to homoplasy: there is little chance of observing two identical alleles that result from different mutational histories.

In this chapter we specify the characteristics of different biochemical and molecular markers to study the diversity of collections of plant genetic material. There are now more than ten techniques of genetic marking (see, for example, Weising et al., 1995; Karp et al., 1996, 1997, 1998; Santoni, 1996; de Vienne and Santoni, 1998). We describe here five widely used and promising techniques: isozymes, RFLP (restriction fragment length polymorphism), RAPD (random amplified polymorphic DNA, also called rapid), AFLP (amplified fragment length polymorphism), and microsatellites. The principle of each technique is briefly described in the appendix to this chapter. After summarizing the structure of the plant genome, we compare the five techniques in methodological terms (target sequences, nature and

level of polymorphism detected, genetic similarity) and practical terms (cost, quickness of result, infrastructure needed).

# ORGANIZATION AND VARIABILITY OF PLANT GENOMES

## The Organization of the Genome

The genome of a plant is distributed in three cell compartments: the mitochondria, the chloroplasts, and the nucleus.

The mitochondrial genome is composed of a circular molecule of master DNA of 200 to 2500 kilobases (1 kilobase = 1 kb = $10^3$ base pairs) depending on the species, carrying 100 to 120 genes. The chloroplast genome is also circular and its size is about 150 kb. It has about 100 genes. The chloroplast and mitochondrial genomes are, most often, inherited through the maternal side.

The nuclear genome is composed of a definite number of chromosomes comprising DNA linear molecules. The genes are spread on a non-coding DNA matrix, essentially constituted of repeat DNA. The size of the nuclear genome varies considerably from one species to another: it is about 400 megabases (1 megabase = 1 Mb = $10^6$ base pairs) in rice and about 16,000 Mb in wheat. The largest known nuclear genome among the angiosperms, that of *Fritillaria assyriaca*, contains 600 times more DNA than the smallest, that of *Arabidopsis thaliana* (Bennett and Smith, 1991). These variations are due to differences in ploidy and especially to differences in the quantity of non-coding dispersed repeat DNA. It is accepted that the number of genes carried by a nuclear genome on the plant is between 20,000 and 50,000. Therefore, the information contained in the nucleus, even if diluted, thus remains clearly more important than that contained in the cytoplasmic organelles. The nuclear genome has a biparental heredity via meiosis and fertilization.

## DNA Polymorphism

The polymorphism of DNA results from the accumulation of mutations, that is, of modifications of sequences under the action of endogenous or exogenous factors. The mutations may appear in the form of visibly large rearrangements on the cytogenetic scale (deletion, translocation, inversion) or in the form of occasional modifications of sequences.

Biochemical and molecular markers are essentially used to detect occasional variations in sequences. Two types of variations can generally be distinguished: mutations corresponding to tmsubstitution of one base for another and mutations by insertion or deletion of a short fragment of DNA.