International Coffee Organization
Organización Internacional del Café
Organização Internacional do Café
Organisation Internationale du Café

WP **Board 1054/08**

18 August 2008
Original: English

E

Projects/Common Fund

Executive Board/
International Coffee Council
22 – 26 September 2008
London, England

Characterization, enhanced utilization and conservation of *Coffea* germplasm diversity

Project proposal

Background

1. The present document has been submitted by the National Coffee Research Centre (Centro Nacional de Investigaciones Cafeteras, Colombia – CENICAFE) in collaboration with Cornell University, and contains the summary of a project proposal designed to facilitate genetic diversity characterization, preservation and utilization in *Coffea*, and to ensure long-term sustainability of coffee production (social, economic and environmental). Estimation of genetic diversity in cultivated crops is essential for breeding programmes and for the conservation of genetic resources. All genetic-resource conservation activities require the characterization of the diversity present in both the gene pools and the gene banks.

2. The proposal has been sent to the Virtual Screening Committee (VSC) for evaluation and will be examined by the Executive Board in September 2008.

Action

The Executive Board is requested to consider this proposal together with the recommendations of the VSC and, if appropriate, to recommend approval by the Council.

PROJECT SUMMARY

1.	**Project title:**	Characterization, enhanced utilization and conservation of *Coffea* germplasm diversity
2.	**Duration:**	Five years
3.	**Location:**	Worldwide through the International Coffee Genome Network (ICGN)
4.	**Nature of the project:**	Encourage poverty alleviation through research and development that secures future characterization of germplasm in *Coffea* for enhanced utilization and preservation to promote the sustainability of the sector.
5.	**Brief description:**	The main goal of this proposal is to contribute to the international efforts of the ICGN and the international coffee scientific community worldwide to facilitate genetic diversity characterization, preservation and utilization in *Coffea*, and to ensure long-term sustainability of coffee production (social, economic and environmental). Estimation of genetic diversity in cultivated crops is essential for breeding programmes and for the conservation of genetic resources. All genetic-resource conservation activities require the characterization of the diversity present in both the gene pools and the gene banks.
6.	**Estimated total cost:**	US$3,000,000
7.	**Financing sought from the Fund:**	US$3,000,000
8.	**Mode of financing:**	Grant
9.	**Co-financing:**	CENICAFE in collaboration with Cornell University has developed the microsatellite single sequence repeats (SSR) markers that will be used in this project with funding from the National Federation of Coffee Growers of Colombia and the Colombian Ministry of Agriculture. More than 2,000 microsatellite markers have been developed from several coffee genomic

libraries (small insert libraries and a large insert Bacterial Artificial Chromosome (BAC) library that has been sequenced to generate BAC end sequences (BES)) and will be available for this project. The *Coffea arabica* BAC library *Hind* III (10.6 x coverage) is composed of 114,816 clones and has been entirely fingerprinted and BAC end sequenced to generate 144 Mb of sequence (114,816 clones x 630 bp per BAC end x 2 ends). Markers for mapping and diversity studies, microsatellites or SSRs, have been derived from these BES. One of the distinguishing features of SSR loci is their hypervariability, which is associated with expansion/contraction potential of the SSR motif itself. This feature, in combination with the ease of polymerase chain reaction (PCR) amplification, the co-dominant profiles and the potential for automation make them ideal for genetic diversity studies in *Coffea*.

10. Mode of co-financing: n/a

11. Counterpart contribution: n/a

12. Project Executing Agency (PEA): CENICAFE, a technical agency of the National Federation of Coffee Growers of Colombia, and Cornell University

13. Supervisory body: International Coffee Organization (ICO)

14. Estimated starting date: Once funding is secured.

Specific project objectives:

1. Provide advanced molecular markers for evaluation of existing *in situ* and *ex situ Coffea* collections.
2. Develop advanced genomic tools for characterization of germplasm to enhance the use of *Coffea* germplasm by breeding programmes worldwide.
3. Contribute to developing and implementing cost-effective and reliable conservation strategies for *Coffea* germplasm in *ex situ* and *in situ* collections.

This project will contribute to preservation efforts in coffee by providing robust molecular markers and support automated genotyping to help systematically characterize germplasm diversity in *Coffea* and help create a platform for capturing diversity, e.g. through the development of genome tilling arrays for DNA polymorphisms.

Project components

1. Standardize the use of a set of SSR markers, automated genotyping and a common nomenclature system so that the results of genetic analysis can be readily interpreted and easily integrated into a genome database.
2. Standardize state of the art genotyping evaluation methods to characterize genetic diversity for conservation and develop an information system for *Coffea* genetic resources.
3. Support international collaborative efforts through the ICGN/ICO to improve evaluation and characterization of germplasm for coordinated worldwide diversity preservation efforts.

Germplasm evaluation will be performed in parallel using automated fluorescent labeled SSR markers. We will first select a set of SSR markers for coffee that can be reliably multiplexed for germplasm evaluation using a panel of wild accessions and some cultivated coffee varieties. Once the set of markers has been developed and adequately tested, they will be used to obtain genotypic data on coffee accessions from different germplasm collections. Multiplexed panels of fluorescent-labeled primers will be used to run the coffee microsatellites on the germplasm accessions using an ABI Prism 3700 DNA Analyser and Genotyper software. The data generated will be analysed to detect allelic diversity and conduct population structure analysis (linkage disequilibrium LD). SSRs will be used to identify the presence of genetically identifiable subgroups within a population by testing for LD to detect significant associations among alleles transmitted together for selective reasons.

A database will be developed with application of new algorithmic tools to promote and facilitate a systems approach to study coffee diversity and adaptation (develop data models and data structures for coffee). In addition, new sequencing strategies that use nanotechnology to greatly increase sequencing throughput by laying out millions of DNA

fragments in parallel will be adapted to coffee. The various technologies differ in the procedures used to array DNA fragments e.g. coated beads (454) vs. DNA attached directly to chips (Solexa). The shorter reads generated present major bioinformatics challenges particularly for genome assembly but their reduced costs and increased accuracy make them of great interest to target sequencing of *de novo* species such as coffee and to get a sense of natural variation in the genus *Coffea*. These new high throughput sequencing platforms will be used to capture diversity as a more systematic approach to identify genes for crop improvement and develop solid genomic tools to speed up selection and genotyping. This information will provide the foundation for detailed functional characterization of coffee genes. Generation of genomic tools for genotyping and breeding will help to overcome some of the major constraints in coffee breeding that have limited the utilization of germplasm within the genus *Coffea*.

The bulk of natural genetic variation in organisms is represented by small insertions or deletions known as single nucleotide polymorphisms (SNPs). Their high frequency (5 per 1,000 bases in humans; 4 per 1,000 bases in rice; and 1-2 per 100 bases in corn) make them ideal for fine mapping studies in candidate genes. Understanding the evolutionary dynamics of plant genomes involves assessing the level, pattern and distribution of SNPs. This information will be very useful to breeders since it will help them understand the relative value of different alleles for genes of agronomic importance in the germplasm pool. Novel sequencing technologies will be utilized for high throughput identification of relevant SNPs in complex genomes by re-sequencing non-repetitive portions of the genomes in a set of multiple accessions. Application of the SNP data generated for association genetics will require detailed and comprehensive phenotyping of the accessions for multiple traits (e.g. tolerance to abiotic and biotic stress, quality, etc.) under different environments.

Implementation of advance genomic tools for germplasm characterization in *Coffea* will allow intensified exploration of the reservoir of genes in collections to preserve them and facilitate their use in breeding programmes. Marker assisted breeding will be one of the first benefits to reduce costs, by selecting plants at an early stage, reducing costly field work.

Beneficiaries

The main beneficiaries of this project will be coffee growers around the world (70% of whom are smallholders) in more than 60 countries that produce coffee. In addition, this project will benefit coffee research institutions and ultimately coffee consumers. Better characterization and preservation of *Coffea* germplasm worldwide is one of the goals of the ICGN and will facilitate the development of new improved varieties. The use of coffee genetic resources is the most effective and environmentally sound technological innovation that can be proposed to overcome major constraints in coffee production worldwide. With the exception of Ethiopia, the genetic diversity of *Coffea arabica* varieties cultivated around the world is extremely limited due to the restricted source of materials originally introduced and also

because *Coffea arabica* coffee is predominantly self-pollinating. Characterization of germplasm diversity will facilitate selection programmes for new varieties with improved yield and quality coupled with good adaptation, disease resistance and tolerance to abiotic stress.

The forests of West and Central Africa, Madagascar and south-western Ethiopia as well as those of neighbouring countries, are the centres of origin of *Coffea*, and are the ultimate sources of coffee genetic diversity. Deforestation to use the land for agricultural activities, population pressures and economic hardships threaten all these reservoirs of great genetic diversity and, with them, comes the danger of significant erosion of the *Coffea* genepool. So far, very little of the genetic diversity contained in the wild or in germplasm collections around the world has been exploited in coffee improvement programmes. Characterization of genetic diversity within the genus *Coffea* will facilitate the use of germplasm to target major diseases, abiotic stress and more sustainable production worldwide. Enhanced and effective use of *Coffea* germplasm in breeding programmes can only be fully realized once the diversity has been thoroughly characterized and preserved.

Field germplasm banks require considerable inputs over long periods of time in terms of land, labour, maintenance, phenotyping, etc. This project will contribute to characterize diversity and prioritize germplasm to be preserved in *ex situ* and *in situ* germplasm banks to facilitate its future utilization. By coordinating this effort internationally and using advanced genomic technology, we can enhance the capacity to save, preserve and utilize *Coffea* germplasm that is currently disappearing in natural forest ecosystems. The great differences in size, morphology and ecological adaptation suggest that *Coffea* diversity is an untapped resource.