MICROSATELLITES IN *Coffea arabica*: CONSTRUCTION AND SELECTION OF TWO GENOMIC LIBRARIES.

Vascotto F., Degli Ivanissevich S.[1], Rovelli P., Anthony F.[2], Anzueto F.[3], Lashermes P.[4], Graziosi G.

Dipartimento di Biologia, Università di Trieste, 34143 Trieste, Italy
[1] DPVTA. Università di Udine, Italy
[2] PROMECAFE Network, Guatemale City, Guatemala
[3] CATIE, Turrialba, Costa Rica
[4] IRD, Montpellier, France

INTRODUCTION.

Almost the entire world production of coffee is obtained by the cultivation of two species: *Coffea arabica* and *Coffea canephora*. The two species provide beans with distinct organoleptic qualities and *C. arabica* is generally preferred for its smoother taste and reduced caffeine content. Moreover *C. arabica* accounts for about 70% of the total world coffee production (Clifford and Willson, 1985).
The commercial importance of *C. arabica* is met by intense research activities at the agronomic level by selection and production of new varieties and new cultivation techniques as for instance the "Cafe adensado" (Sera et al., 1994). Nevertheless the genetic improvement of the varieties as well as the controlled introgression of valuable genes from other species is somewhat hindered by the low knowledge on the genetics of *C. arabica*. Although some mendelian inherited loci are known, as for instance *Maragogype* (large beams and leaves) or *Caturra* (short stature), they are too few for building up a genetic map and the lack of a genetic map has obvious effect on breeding programmes.
Polymorphic DNA loci represent the obvious choice as marker for constructing a genetic map and indeed a number of polymorphisms are available as RAPD (Lashermes et al., 1996; Orozco-Castillo et al., 1994) and AFLP (Lashermes et al., 1999). Nevertheless the polymorphic bands detected by these two methods show a dominant behaviour: one band is either present or absent while intensity variations are generally unreproducible. Therefore these methods do not allow for distinguishing between the dominant homozygote and the heterozygote, both characterised by the presence of the same bands.
More informative are the polymorphisms based the microsatellites: they show codominant bands and therefore they allow for the identification of all possible genotypes as well as easy transfer of markers in different crosses.
These characteristics are of great relevance in constructing a genetic map (Paillard et al., 1996) as well as in marker assisted breeding programmes.
We undertook this research project in view of identifying an characterising polymorphic microsatellites.

MATERIAL AND METHODS.

Plant material.

The genomic libraries were constructed from DNA extracted from leaves of a plant of C. arabica var. Caturra which is part of the IRD (Montpellier) collection. The polymorphisms were assessed on a F_2 population of 12 plants obtained from the cross of Caturra x Wild Ethiopia (ET30) and selfing of the F_1. Unfortunately the parental and the F_1 plants are presently unaccessible and they could not be tested for heterozygosity.

DNA extraction and purification.

The DNA was extracted from young leaves, 2-3 cm, which were lyophilised immediately after collection. Since the quality of a genomic library depends very much on the quality of the starting DNA, particular attention was paid for the initial DNA extraction and purification. We adopted a modification of the extraction technique of Murray and Thompson (1980) and Orozco-Castillo et al (1994), as reported in detailed in the protocol below.
- Grind 0.12 g of lyophilised leaf in a mortar, add 3ml lysis buffer (4% CTAB, 100mM Tris-HCl pH8.0, 1.4M NaCl, 20 mM EDTA, 10.5µl ß-Mercaptoethanol for 15 ml of buffer, to be added at the time of the extraction)
- Following incubation for 1 h at 65°C, extract with 2.25 ml CHL/IAA (24:1) for 5'
- Centrifuge 5' at 12000 rpm and transfer the supernatant into a clean tube; precipitate with 0.6 volumes of isopropilic alcohol 10' at room temperature (1.8 ml for 3 ml of buffer)

- Centrifuge 15' at 12000 rpm, drain supernatant and dry the pellet very well (eventually at 65°C)
- Resuspend in 250µl of TE and incubate at 65°C for 15' with 2.5µl RNAase A (1 mg/ml)

Further purification of the extracts was performed as reported below.

- Add 45 µl 5M NaCl and shake by inversion; add 300 µl 2% CTAB, 100 mM Tris pH 8.0, 1,4 M NaCl, 20 mM EDTA and shake by inversion
- Extract with 500 µl CHL/IAA wait 5' and centrifuge 5' at 12000 rpm then transfer the supernatant in a clean Eppendorf
- Precipitate for 1 h with the addition of 1.5 volumes of 1% CTAB, 50 mM TRIS pH 8.0 at room temperature and centrifuge at 12000 rpm for 10'
- After drying very well the pellet, remove the CTAB from the tube walls and resuspend in 75 µl 5M NaCl and 300 µl TE (incubate 10' at 56°C) and, if necessary, centrifuge 5' at 12000 rpm to eliminate undissolved particles
- Precipitate the DNA with 950 µl ethanol 100% (wait 10' at RT), centrifuge at 12000 rpm for 10'; wash the pellet with 500 µl ethanol 70%, wait 5' and centrifuge 5' at 12000 rpm
- Dry the pellet at 37°C and dissolve in 50 µl TE (eventually at 56°C)

Construction and enrichment of the genomic libraries.

Two genomic libraries were prepared as described by Rafalski et al. (1996), Morgante et al. (1998) with modifications as reported below. One library was enriched in $(TG)_n$ and the other one was enriched in $(ATC)_n$. The main steps of the procedure were the following:

Caturra DNA digestion with *Tsp509I*

DNA fragment selection (250-700 bp from agarose gels)

Fragment ligation to an adapter containing an *EcoRI* site and adequate primer sequences.

Enrichment in TG or ATC by DYNABEADS-DYNAL conjugated with biotin-strepavidin oligonucleotides complementary to TG and ATC;

PCR Amplification of the two libraries by priming the adapters

Control of the success of the selection procedure by Southern blot using the TG and ATC oligonucleotides as probes.

Purification of the two libraries by the Wizard-Promega kit.

Restriction of the genomic fragments by *EcoRI.*

Selective recovery of the digested fragments (250-700 bp) from agarose gel on DEAE NA45 filters.

Fragment cloning in the *EcoRI* site of -ZAPII (Stratagene, La Jolla, Calif.).

Phage packaging and infection of XL1-Blue MRF' strain.

Phage count and plating of the libraries.

Lifting on nylon membranes.

Screening of the libraries for the TG/ATC positives clones.

Sequencing and primer design.

The sequencing of the positive clones was performed by Thermo Sequenase dye terminator cycle sequencing pre-mix kit (Amersham Pharmacia Biotec) following the producer indications.

Pairs of primers were designed on the regions flanking the microsatellites. The on line programmes Primer3 (Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA) or *Primers!* (Williamstone Enterprises) were used, choosing a Tm of a 58°C .The constant tail KS (5'-TCGAGGTCGACGGTATC-3') was added to one of the primers for each primer pair.

Amplification and analysis of the microsatellites.

The microsatellites underwent touchdown PCR as reported by Hecker e Roux, 1996. A three primer system was developed: two primers were the locus specific primers, while the third primer was complementary to the constant tail of one of the primers and was tagged by a fluorochrome, either 6-FAM (6-carbossifluoresceine) or JOE (2',7'-dimetossi-4',5'-dicloro-6-carbossifluoresceine).

The amplification conditions were the following: 6 cycles, denaturation 45s at 94°C, elongation 45s at 72°C, annealing 45s with decreasing temperature from 60°C to 55°C; 34 cycles, denaturation at 94°C for 30 s, annealing at 55°C for 30 s, elongation at 72°C for 30s (8 mins for last cycle).

The amplified fragment were run on sequencing acrilamide gels in an automatic sequencer ABI 373A. The alleles identification was performed by the GENESCAN 672 (Perkin Elmer) software.

RESULTS

Constructing the genomic libraries.

The probability of finding polymorphic microsatellite depends very much on the procedure adopted for the construction of the libraries as well as on the enrichment strategy. Here we account for some relevant points of the strategy adopted.

As reported in the Material and Method section, we constructed two genomic libraries. The first library was enriched for the dinucleotide monomer sequence TG which was chosen on the

basis of the experience of various research groups on barley and other plants. The second library was enriched for the trinucleotide monomer sequence ATC which was chosen following hybridisation experiments. The DNA was digested with a few restriction enzymes including *Tsp509I* and hybridised with the following oligonucleotides: $(ATC)_{10}$, $(ACC)_{10}$, $(AAC)_{10}$, $(AAG)_{10}$. The best results were obtained with *Tsp509I* digestion for the oligonucleotide $(ATC)_{10}$, in fact the Southern blot showed a uniform profile where the majority of DNA fragments was below 1 Kbp.
The critical enrichment step was performed by conjugating the magnetic beads coated by streptavidin with the biotinilated oligonucleotide (ATC)10 or (TG) 13, the unbound DNA was discarded. The efficiency of enrichment was evaluated by hybridisation of the -ZAPII colonies with the chemiluminescent 5'-(DIG)3(ATC)10 or 5'-(DIG)3(TG) 13 oligomers. About 9% of the colonies gave a strong positive signal. We picked up only the positive colonies, thus introducing a second degree selection procedure.

The sequencing of the positive clones.
Most of the clones were short enough to be sequence in one run. About 70% of the (ATC)10/(TG) 13 positive clones contained a SSR. Here are reported some examples of the microsatellite sequences of the TG library:
GTGTGTGTGTGTGTGAGAGAGAGAGAGAG
AGAGAGAGAGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTG
CACACACACACACACACAAGCTCTCTCTC
TCTCTCTCTCT
TGTGTGTCTATGTATGTATGTATATATGTG
TGTGTGTATATGTGTGT
CACACACACACACACACAAGCTCTCTCTC
TCTCTCTCTCT
Besides the expected TG and the complementary AC repeats, compound microsatellites were obtained where the repeats could be: TA, CT or GA. Thus the sequencing to the clones allowed for the identification of the monomer sequence of microsatellites which were not specifically selected for. This result could be use in future projects for microsatellite production.

Primer designed.
Even if about 70% of the sequenced clones contained a microsatellite sequence, the number of primer pairs designed was much lower. Many clone sequences could not be used for designing primers for one or more of the following reasons:
No repeated sequence could be identified;
The insert was a chimera, i.e. composed by two or more DNA fragments of independent origin.
The microsatellite covered the entire insert and there was no space for designing the primers;
The microsatellite was too short, unlikely to be polymorphic;
The T_m of the two primers was too different;
The amplification conditions did not meet the requirements established "a priori". We fixed some amplification parameters which must be common to all primers to allow for the development of multiplex when analysing large sample population.
Because of all these limitations only 30% of the sequenced clones allowed for primer designed.

Identification of the polymorphic microsatellites.
The identification of the polymorphic loci was performed analysing DNA extracted from the cultivars Mundo Novo and Bourbon, from the 12 F_2 plants of the segregating population and from two plants of the same parental lines (ET30 and Caturra) of the F_2. The identification of the alleles was performed on sequencing gels because we could expected alleles differing only for 2 base pairs.
When using the same amplification conditions for all primers as reported in the Material and Method section, about one third of the primer pairs gave no amplification product. Presumably these primers require adjustments for the amplification conditions and they have been momentarily abandoned. Many of the remaining primer gave monomorphic bands. Only 25% of the primers which gave an amplification products eventually gave polymorphic bands either on the varieties or on the F_2 population.
Even if we analysed a limited number of varieties, frequently we obtained multiple alleles, from 2 to 5 alleles. Usually each single plant had either one or two alleles, except the locus E12-3CTG who gave 3 or 4 bands per plant.

DISCUSSION.

The construction of the genomic library enriched in microsatellite sequences proved to be rather complex. Nevertheless the results so far obtained are very promising and there are high probabilities of finding a reasonable number of polymorphic microsatellites. In facts more than 3% of the sequenced clones allowed for the identification of polymorphic loci or, if we refer to the genomic libraries, we expect that about 0.2% of the clones contain a polymorphic sequence. We estimated

that our enriched genomic libraries should have about 50,000 clones still to be screened and therefore we could expect about a hundred polymorphisms to be identified.

An open question is the redundancy of our genomic libraries. Until now, we had very few clones sharing the same sequence or partial sequence identity. Nevertheless it is reasonable to foresee an increase of clones containing sequences already known. Thus the total number of potential polymorphisms should be corrected for the redundancy but, for the moment, we cannot evaluate the correction value.

Coffea arabica is an autogamous plant and a reasonable percentage of homozygosity is expected. Moreover, it has been reported that the genetic base of the cultivated varieties is rather narrow (Bertraud and Charrier, 1988). Under these circumstances we would expect a reduced variability even at the microsatellite level. Our data do not allow for a comparison with other vegetal species on the

Acknowledgements: The research project has been supported by EU, INCO Contract n. ERBIC18CT970181.

REFERENCES.

Bertraud J., and A. Charrier. 1988. Genetic Resources of Coffea. In "Coffee", Vol. 4, Agronomy, R.J. Clarke and R. Macrae Ed.s, Elsevier Applied Science, London pp1-41

Clifford M.N., and K.C. Willson. (Editors) 1985. Coffee; botany, biochemistry and production of beans and beverage. London, Croom Helm,

Hecker K.H., and K.H. Roux. 1996. High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. Biotechniques 20: 478-485.

Lashermes P., S. Andrzejewski, B. Bertrand , M.C. Combes , S. Dussert , G. Graziosi, P. Trouslot, and F. Anthony. 1999 Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L.) Theor. Appl. Genet. (in print).

Lashermes, P., P. Trouslot, F. Anthony, M.C. Combes, and A. Charrier. 1996. Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*. Euphytica 87:59-64

Morgante, M., A. Pfeiffer, I. Jurman, G. Paglia , and A.M. Olivieri. 1998. Microsatellite markers in plants. In: Karp A, Isaac PG, Ingram DS (eds) Molecular tools for screening biodiversity. Plants and animals. Chapman and Hall, pp. 206-208,

Murray, M.G., and W.F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 8: 4321-4325.

Orozco-Castillo, C., K.J. Chalmers, R. Waugh, and W. Powell. 1994 Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. Theor. Appl. Genet. 87: 934-940.

Paillard, M., P. Lashermes, and V. Pétiard. 1996. Construction of a molecular linkage map in coffee. Theor. Appl. Genet . 93: 41-47.

Rafalski, J.A., J.M. Vogel, M. Morgante, W. Powell, C. Andre, and S.V. Tingey. 1996. Generating and using DNA markers in plants. In: Birren B, Lai E (eds) Nonmammalian genomic analysis: a practical guide. Academic Press, London New York, pp.75-134.

Sera, T., F.Androcioli, R.M.L. Cardoso, M.C.L.L. Dias, A. Guerreiro, and E. Silva. 1994. International Symposium on High Density Planting of Coffee Tree. p293-294 IAPAR

Vascotto F., Degli Ivanissevich S., Rovelli P., Anthony François, Anzueto F., Lashermes Philippe, Graziosi G. (2000).

Microsatellites in Coffea arabica : construction and selection of two genomic libraries.

In : Riede C.R. (ed.), Sera T. (ed.), Soccol C.R. (ed.), Roussos Sevastianos (ed.). Anais do 3 Seminario internacional sobre biotecnologia na agroindustria cafeeira = Proceedings of the 3rd international seminar on biotechnology in the coffee agroindustry.

Londrina, PR (BRA) ; Montpellier : IAPAR ; IRD, p. 125-128.

SIBAC : International Seminar, 3., Londrina, PR (BRA), 1999/05/24-28.