Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops

Norbert Billotte[a]*
Pierre J.L. Lagoda[a]
Ange-Marie Risterucci[a]
Franc-Christophe Baurens[b]

[a] Cirad (Centre de coopération internationale en recherche agronomique pour le développement), BP 5035, 34032 Montpellier cedex 1, France

[b] Vitropic SA, ZAE « Les Avants », 34270 Saint-Mathieu-de-Tréviers, France

Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops.

Abstract — Introduction. Tropical crops are of great interest for developing countries but can be considered as orphan crops in terms of molecular markers. Microsatellite markers have many advantages compared to the other main molecular marker techniques such as isozyme, RFLP and RAPD analyses. Our purpose was to develop a method for building microsatellite-enriched libraries using DNA from tropical crops as starting material. A quick and simple technique for the construction of $(GA)_n$ microsatellite-enriched libraries of oil palm is described here. **Microsatellite-enriched library building procedure**. The hybridization with biotin-labeled oligoprobe followed by the capture of the selected sequences with streptavidin-coated magnetic beads are the two principles of the technique. The different steps of the method are discussed and an optimized protocol is proposed. **Characterization of the enriched libraries**. The libraries, of several thousand clones each, exhibit a percentage of positive clones, effectively having a microsatellite, greater than 70%. The global redundancy rate of positive clones mainly depends on the DNA shearing method. Redundancy was 20 and 60% with *Pst*I endonuclease-digested fragments and with sonicated fragments respectively. **Conclusion**. This study gives a technical basis for easy microsatellite marker development in tropical crops, until the international scientific community starts up molecular projects applied to the genetic improvement of these traditionally orphan crops. (© Elsevier, Paris)

France / methods / molecular biology / genetic markers / microsatellites / databases / nucleotide sequence

Banques enrichies en microsatellite : méthode pour isoler des marqueurs SSR dans les productions tropicales.

Résumé — Introduction. Les productions tropicales, primordiales pour les pays en voie de développement, peuvent cependant être considérées comme le parent pauvre des marqueurs moléculaires. L'utilisation des marqueurs microsatellites a de nombreux avantages par rapport à d'autres techniques moléculaires comme les analyses d'isozymes, de RFLP ou de RADP. Notre objectif a donc été de développer une méthode permettant de constituer des banques enrichies en microsatellites en utilisant, comme matériel de départ, de l'ADN de plantes tropicales. Une construction simple et rapide de banque enrichie en séquences microsatellites $(GA)_n$ du palmier à huile est décrite ici. **Procédure de construction de banques enrichies en microsatellites**. L'hybridation avec une sonde microsatellite biotinylée puis la capture des séquences ciblées grâce à des particules magnétiques recouvertes de streptavidine sont les deux principes de la technique utilisée. Les différentes étapes de la méthode sont discutées et un protocole optimisé est proposé. **Caractérisation des banques enrichies**. Les banques de plusieurs milliers de clones ont une proportion de clones positifs, contenant un microsatellite, supérieure à 70 %. Le taux de redondance des clones positifs dépend du mode de préparation de l'ADN, avec, respectivement, 20 et 60 % pour des fragments obtenus par restriction par *Pst*I ou par sonication. **Conclusion**. L'étude présentée donne une technique pour développer facilement des banques de marqueurs microsatellites spécifiques des plantes tropicales en attendant que la communauté internationale entreprenne des projets moléculaires appliqués à l'amélioration génétique de ces cultures traditionnellement oubliées. (© Elsevier, Paris)

France / méthode / biologie moléculaire / marqueur génétique / microsatellite / banque de données / séquence nucléotidique

* *Author for correspondence* (billotte@cirad.fr)

Received 30 April 1999

Fruits, 1999, vol. 54, p. 277–288
© Elsevier, Paris

***Resumen Español**, p. 287*

1. introduction

Simple sequence repeat (SSR) length polymorphism was first used for mammalian genome mapping in the early nineties [1–3]. Since then, it has been increasingly used in a wide range of genetic plant studies such as genetic mapping and QTL detection of agronomic traits in cultivated annual or perennial crops [4–9], diversity and pedigree analyses [10–13], phylogenetic analyses [8, 14]. Microsatellite markers have many advantages compared to the other main molecular marker techniques such as isozyme, RFLP and RAPD analyses [15–20]. However, costly and work-intensive development of such markers is a key obstacle preventing a more generalized use for many laboratories worldwide. Tropical crops are of great interest for developing countries but can be considered as orphan crops in terms of molecular markers. Our purpose is to develop a method for building microsatellite-enriched libraries using DNA from tropical crops as starting material.

Since the early nineties, microsatellite loci have been obtained by screening plant genomic libraries. Unfortunately, even good genomic libraries contain only 0.5 to 10% of microsatellite inserts, as reported by previous authors [5, 16, 17, 21–23]. This low percentage means it is necessary to check an enormous number of clones for the development of several hundred microsatellite markers. Therefore, the use of enriched libraries has been proposed [24, 25]. Different methods are available to construct a microsatellite-enriched library, all based on the hybridization of a synthetic oligonucleotide microsatellite sequence onto genomic DNA. These libraries increase the number of microsatellite sequences from at least 20 to 90% [24–27] and allow for cost reduction at the screening stage. We describe here an improved method for developing microsatellite-enriched libraries on oil palm and which can be applied to any tropical plants.

2. microsatellite-enriched library building procedure (*figure 1*)

2.1. total genomic DNA extraction

Total genomic oil palm DNA was extracted from a freeze-dried leaf sample [28], and then purified on anion exchange microcolumns.

2.1.1. shearing of DNA

10 µg of genomic DNA was used for shearing by endonuclease digestion by *Pst*I using 10 enzyme units per µg of total genomic DNA in a final volume of 500 µL. The size of the DNA fragments obtained was checked by agarose gel electrophoresis without size selection.

2.1.2. "polishing" DNA fragment ends

Cohesive ends of DNA fragments were "polished" by the action of T4 DNA polymerase to obtain blunt ends as indicated by the supplier (Gibco-BRL). The DNA fragments were first treated using the exonuclease activity of T4 DNA polymerase (2.5 units for 1 min at 37 °C in a final volume of 30 µL) and then, for polymerase activity, by adding 33 µM of dNTP (30 min at 37 °C).

2.1.3. adaptator ligation and PCR preamplification

Blunt end DNA fragments were ligated by T4 DNA ligase (Gibco-BRL) to *Mlu*I self-complementary adaptators RSA21 5'-CTCTTGCTTACGCGTGGACTA-3' and RSA25 5'-AGTCCACGCGTAAGCAAGAGCACA-3' according to Edwards et al. [26] with the following modifications: 1 µg of sheared DNA, 0.05 µM of each adaptator, 4 units of T4 DNA ligase, 2 h at 20 °C in a final volume of 200 µL. The ligation was checked by PCR amplification. Five ng of ligated DNA were amplified in a final volume of 25 µL with 1 µM of RSA21 primer in a buffer containing 10 mM TrisHCl pH 8, 100 mM KCl, 0.05% w/v gelatin and 1.5 mM $MgCl_2$, and following the PCR program:

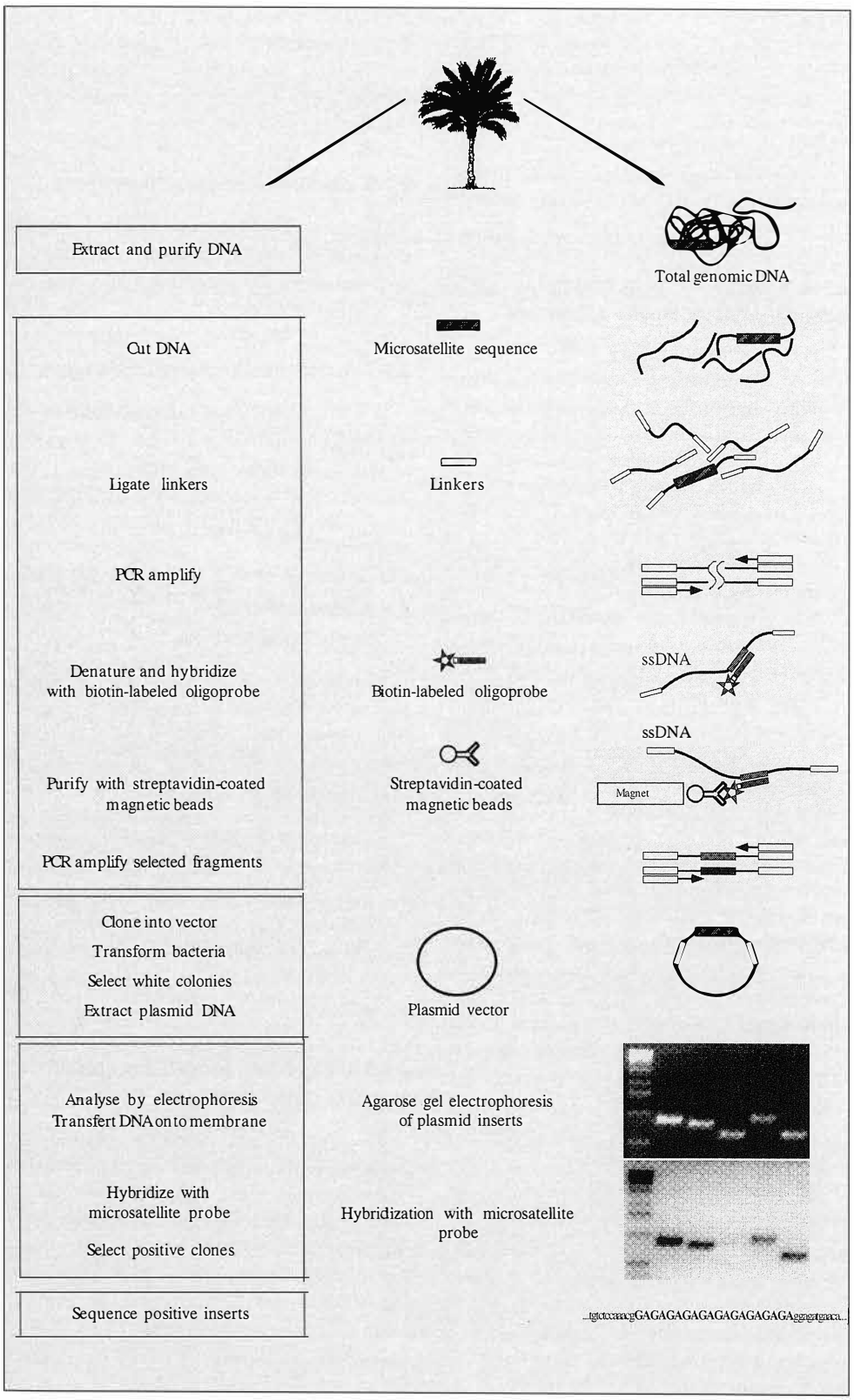

Figure 1.
Microsatellite-enriched library building procedure.

denaturation at 95 °C for 1 min and 28 cycles of [94 °C for 40 s, 60 °C for 60 s, 72 °C for 120 s]. The size of the amplified ligated fragments was checked by agarose gel electrophoresis of 10 µL of the PCR product. The rest of the PCR product was purified on anion exchange microcolumns, and then resuspended in 100 µL of sterile water for the next step.

2.1.4. selection of DNA fragments containing (GA) microsatellite sequences

The selection of microsatellite sequences was performed using biotin-labeled microsatellite oligoprobe and streptavidin-coated magnetic beads, following a hybridization-based capture methodology adapted from Kijas et al. [27]. Magnetic bead-based selection was performed using the Magnetosphere Magnetic Separation Products kit (Promega Madison, USA). Our oligoprobe includes one biotynilated inosine (I*) at the 5' end followed by 4 inosines (I) and by eight repetitions of a TC dinucleotide: 5'-I*IIIITCTCTCTCTCTCTC-3'.

Purified PCR product was denaturated at 95 °C for 15 min in 500 µL before adding 3 µL of 50 µM biotynilated oligoprobe and 13 µL of 20X SSC. Hybridization was performed 20 min at room temperature. A quantity of 600 µg of pre-washed streptavidin-coated magnetic particles, according to the manufacturer's specifications, was resuspended in 100 µL of 0.5X SSC and added to the hybridization mix. After 10 min at room temperature, the magnetic beads were washed 4 times with 300 µL of 0.1X SSC, then eluted twice with 100 µL and 150 µL respectively. The 250 µL final elution fraction can be stored at –20 °C before further use.

2.1.5. PCR amplification of the selected fragments

One twenty-fifth of the eluted fraction was PCR-amplified in a final volume of 100 µL reaction mix with primer RSA21, as previously described for the ligation step. The sizes of the amplified fragments were checked by running a 10 µL aliquot of PCR products on agarose gels. The quality of the amplification was evaluated by comparison with an amplification of unselected fragments. The rest of the PCR product was purified by anion exchange column prior to the cloning step.

2.1.6. cloning of selected fragments

The purified PCR products were cloned into pGEM-T Easy vector, as indicated by the supplier (Promega, Madison, USA), in a final volume of 20 µL.

2.1.7. transformation by cloned fragments

A 2 µL aliquot fraction of the ligation mix was used to transform 100 µL of competent DH5α *Escherichia coli* strain [29]. Transformed bacteria were plated onto 2YT, Xgal, Ampicillin media and incubated overnight at 37 °C [30].

2.1.8. characterization of the enriched libraries

The size of the selected inserts was estimated by agarose gel electrophoresis of PCR amplification products using standard M13 and reverse primers flanking each pGEM-T Easy plasmid polylinker. Electrophoresis gels were stained with ethidium bromide and photographed. The electrophoresed PCR products were then alkaline-Southern transferred onto Hybond N+ nylon membranes (Amersham). RFLP hybridization was performed with a ^{32}P radiolabeled $(GA)_{15}$ synthetic microsatellite probe according to Cregan et al. [31].

2.1.9. sequencing template preparation and data gathering

Plasmids from positive clones containing a microsatellite sequence were extracted by plasmid DNA minipreparation using the High Pure Plasmid DNA Extraction kit (Roche). Sequencing was performed using the universal T7 and M13 reverse sequencing primers and the dideoxy dye terminator method. Sequence data analysis was performed on a random sample of 50 sequences per enriched library.

2. 2. modifications

2.2.1. alternative method for shearing DNA: sonication

Step 2. A quantity of 10 µg of total genomic DNA was treated for 2 h in a 500 µL volume of 3X SSC in a sonicator (Bransonic B52). The length of the sonication step was adapted to obtain DNA fragments below 600 bp in length. The 3' and 5' phosphate groups of sonicated DNA fragments were eliminated by the action of Shrimp Alcaline Phosphatase (SAP). Reaction was performed on 1 µg of DNA at 37 °C for 1 h. The SAP was inactivated at 65 °C for 15 min. DNA fragments were precipitated, resuspended in 20 µL of sterile water and treated according to the standard protocol starting at step 3.

Step 3. This step is ended by supplementary phosphorylation of the 5' fragment ends by the T4 polynucleotide kinase (Gibco-BRL).

2.2.2. biotin-labeled microsatellite oligoprobe

Step 5. Other microsatellite motives can be selected using the same primer design. For (CA) repeats the primer 5'-I*IIIITGTGTGTGTGTGTG-3' can be used.

3. characterization of the enriched libraries

3.1. *Pst*I enriched library

A total of 254 tested clones contained inserts with a fragment length ranging from 300 pb to 1200 pb. About 74% of them were positive by RFLP hybridization with the microsatellite probe. The sequencing data confirmed that all of the latter included a perfect $(GA/TC)_n$ microsatellite locus, sometimes directly flanked by a $(CA/GT)_n$ or a $(CAT/GTA)_n$ microsatellite sequence (in 6 and 1 cases, respectively). Only 20% of redundant clones were detected. Non-redundant sequences had an average length of 605 pb. Up to 87.5% of these sequences had two flanking regions exceeding 50 pb. The number of (GA) repeats of non-redundant sequences ranged from 8 to 28 with an average of 17.

3.2. sonicated library

Analysis of 486 white colonies shows an average insert size of 450 pb, and 345 of them (71%) gave a positive signal by hybridization with the $(GA)_{15}$ radiolabeled microsatellite probe. Sequence analyses confirmed that 100% of the positive clones contain a microsatellite sequence. One chimeric sequence was detected with one microsatellite flanking region from one clone and the other flanking region from another clone. Normal non-redundant sequences (40%) have an average total size of 288 pb. All microsatellite loci have a perfect $(GA/TC)_n$ locus except one which is a $(GGA/CCT)_9$ repetitive sequence. Both flanking regions of the microsatellite locus have a size of more than 50 pb for 76% of the non-redundant sequences. The average repetition number of the (GA) dinucleotide is 20 with a rather "small" standard deviation of 4, translating a range varying between 15 and 31 repetitions.

3.3. choice of the nature of SSR for library enrichment

A higher abundance of $(GA)_n$ microsatellites compared to other dinucleotide SSR is currently observed in plant genomes as in rice [4, 6, 16], Arabidopsis [32], Norway spruce [17], rapeseed [23] or banana [22]. The choice of a (GA) enrichment was preferred for a first library of fragments, on the basis of comparisons with other plants and because of the technically easier hybridization with GA or GT probes compared to AT or GC probes due to self-complementarity and probe annealing temperature. Moreover, previous studies on plant GA microsatellite show that they are often well-distributed throughout the genome, ensuring good coverage [15, 33]. In fact, the choice of microsatellite types mainly depends on the application wanted for the marker. Good coverage of the genome and high abundance are interesting for genetic mapping and diversity studies. On the other

hand, high abundance may not always be the right choice for development of fingerprinting [11].

3.4. design of the oligonucleotide probe and enrichment step

Actually, only two systems of interaction between magnetic beads and oligonucleotide probes are commercially available, both of them involving protein / protein or protein / ligand interactions (digoxigenin / anti-digoxigenin or biotin / streptavidin). As the size of the protein coated onto the magnetic beads is quite high compared to the size of the probe, we added an inosine tail at the 5' end of our probe to separate the repeated sequence from the bead in order to favor better hybridization with DNA fragments by decreasing steric hindrance. The size of the tail was empirically fixed to 5 inosine units. For the sonicated fragment library, the average size of the microsatellite flanking region selected on the biotin side of the probe is half the size of the opposite flanking region (data not shown). A possible explanation of this observation may be that the proximity of the streptavidin / biotin complex as well as the magnetic bead itself introduce a strong bias in the selected fragments through steric interactions. If this hypothesis is to be verified, the size of the inosine tail will have to be increased in order to decrease this effect against the flanking region length. In contrast, the *Pst*I fragment library shows no measurable difference between the length of the 5' and 3' flanking regions. Problems of short flanking regions depend partly on the DNA preparation technique and may not be totally avoidable.

3.5. shearing of DNA

The first step for selecting microsatellite-containing inserts is the shearing of the genomic DNA. DNA molecules can be fragmented by enzymatic or physical procedures: hydrolysis or repeated pipetting through a thin nozzle, repeated thawing, thermal agitation, sonication, ionizing or UV radiation, or chemical disruption. Previous works have shown that microsatellite contents in partial genomic libraries depend on the choice of the endonuclease. For example, in banana, partial genomic libraries contain (GA) microsatellite sequences at a frequency of less than 1% for *Sau*3A1, 1.2% for *Taq*I [22], 1.3% for *Eco*RI (J.L. Noyer, personal communication) and 7% for *Pst*I [22]. The *Pst*I enzyme is very sensitive to methylation and is a frequent cutter in HTF (*Hpa*II Tiny Fragments) "methylation-free islands" corresponding in great majority to transcriptionally active sequences [34]. These results suggest that microsatellites are preferentially located close to genes or sequences involved in transcription modulation [35]. Thus the choice of a restriction enzyme sensitive to methylation is preferred to increase the abundance of microsatellite sequences in the starting material and to fall within or near coding regions of the genome.

The choice of an endonuclease is not made at random. In fact, the use of restriction endonucleases of different recognition site sizes (4 bp, 6 bp or other) will statistically lead to DNA fragments differing in size average. As best results for cloning are obtained with relatively small fragments, between 300 and 2 500 bp (data not shown), exploratory assays should be performed to evaluate an adequate fragment size distribution per plant species. On the other hand, depending on available information on the nucleotidic composition of the studied genome, the GC content of the recognition site of the restriction endonuclease should be taken into account in the definition of the endonucleases to be tested in the exploratory assay. For example, for the 4 bp cutters, *Mse*I (TTAA) could be considered as targeting "AT-rich regions"; *Taq*I (TCGA), *Sau*3A1 (GATC) and *Rsa*I (GTAC) could be considered as "intermediate", and *Hae*III (GGCC) could be considered as targeting "GC-rich regions".

The objective of physical methods, sonication in particular, is to generate DNA fragments which are well-distributed throughout the whole genome as has been shown in rice [16]. Shearing of DNA by sonication is a safe physical method compared to ionizing radiation. DNA from all tropical crop species we studied were sheared to a few

hundred bp fragments in less than 2 h (data not shown). In order to allow for a good ligation step, the fragment ends need to be treated by phosphatase / kinase.

Restriction endonucleases are not one hundred percent efficient due to hidden partial DNA segments [36]. On the other hand, our data of *Pst*I restriction indicate that the sonication could introduce differences not only with regard to the average location of the (GA) SSR region (closer to the 5' sequence end), but also to microsatellite flanking region length variability. Indeed, the 5' flanking region of sonicated fragments appears smaller than the 3' flanking region (average of 83 pb against 165 pb). The total length of the sequence is not significantly correlated to the 5' flanking region length (ρ = –0.26 with 18 df), while it is positively correlated to the 3' flanking region length (ρ = 0.88** with 18 df). On the other hand, correlation is significantly positive for *Pst*I restricted fragments (ρ = 0.73** or 0.31* respectively with 39 df). This means that, statistically, selecting for longer sonicated fragments would not give a better internal microsatellite location starting from the 5' end as can be expected for *Pst*I restricted fragments. But such assumptions require complementary sequence data for confirmation.

3.6. microsatellite locus sampling

The number of microsatellite loci sampled amongst their global population is estimated from the genome copy number represented in our starting material. Basically, we estimate that the DNA restriction and "polishing" steps were completed with a maximum efficacy of 1. The ligation of restricted fragments to adaptators is the first factor reducing the DNA quantity. The blunt end ligation efficacy was estimated by comparison of DNA quantities before and after ligation of DNA linkers in the AFLP technique [37] which was found to be 10^{-4}. The other reducing factor is sampling of the starting material at step 4 (0.05%). Thus, the final number of loci present in the library can be calculated by knowing the size of the oil palm genome which is 4 pg / 2C [38]. The quantity of the starting DNA material represents up to $2.5 \cdot 10^6$ copies of the oil palm genome and decreases to about 0.1 copies at the first PCR amplification. Because the next steps of our protocol are no longer limiting in terms of the genome copy number, the enriched library samples about 10% of the total number of the genome microsatellite loci. The oil palm (GA) microsatellite content can be estimated approximately at 10,000 to 40,000 different loci, based on other monocotyledons (banana genome microsatellite content data and review in [22]). Our enriched library may represent between 1,000 to 4,000 different microsatellite loci.

3.7. PCR preamplifications

3.7.1. redundancy of a given clone

Two PCR amplification steps were added to our protocol to allow for low DNA starting quantities. Pragmatically, we chose to amplify the DNA because the most critical step of this protocol is the DNA quantity prior to cloning. However, PCR amplifications raise the question of clone redundancy. Redundancy (R) of a given clone is calculated by multiplying the efficacy of each different step of our protocol starting from the first PCR amplification in step 4. The PCR efficacy (E) is evaluated at 0.47 according to Siebert and Larrick [39] and to molecular data from Baurens et al. [40]. Capture and purification efficacy (E_c and E_p) is fixed at 1 and 0.75, respectively. The one base sticky end ligation efficacy (E_l), measured by ligating a standard insert into the pGEM-T Easy vector, was $8 \cdot 10^{-5}$.

Knowing that the DNA template of the first PCR amplification is lower than one genome equivalent, the redundancy of a given clone is:

$$R = (1+ E)^n \times E_p \times E_c \times 1/25 \times (1+ E)^n \times E_p \times E_l,$$

where n is the number of PCR cycles and 1/25 is the initial dilution in step 1.

We assumed that clone redundancy is homogeneous in the library and that it depends mainly on the technical steps of our protocol, i.e., the number of PCR cycles, the PCR efficacy and ligation efficacy. Optimizing one of these parameters would dras-

tically enhance redundancy. According to our computing model and data, about 400 redundant sequences per clone have been generated just prior to the bacteria transformation step.

Our model for redundancy computing predicts that decreasing the number of PCR amplification cycles from 28 to 20 reduces the redundancy to single sequences in the library. In this case, DNA quantity prior to bacteria transformation decreases from 2 pg to 4 fg. Thus, poor bacteria transformation efficacy that does not exceed 10^{-3} transformants per molecule of plasmid should be improved by no less than 2 orders of magnitude.

3.7.2. relative sizes and global redundancy rates of *Pst*I and sonicated enriched libraries

The global redundancy percentage of the *Pst*I enriched library is 20%. The global frequency of redundant clones (60%) for the sonicated enriched library is significantly higher than that observed for the *Pst*I enriched library (χ^2 = 20.8** with df = 1). But nearly half (47%) of the redundant sonicated fragments concerned one particular locus only. Despite this fact, different global redundancies should reflect a real difference between the two libraries, in terms of total number of microsatellite loci represented per library, assuming that equivalent experimental conditions ensure the same copy number for any given locus. Thus, the *Pst*I library should sample more loci of the genome than the sonicated enriched library does. This is in accordance with previously exposed data suggesting that clonable DNA fragments issuing from a *Pst*I restriction contain larger amounts of microsatellite loci compared to those obtained by sonication or by restriction with other enzymes.

We feel the *Pst*I enriched library should be conserved and extensively used for microsatellite marker construction. On the other hand, the sonicated library should be trashed and the building of a new one starting from step 6 reconsidered. In fact, the extensive use of a library and the total number of sequences needed for obtaining a given number of different microsatellite loci depend on the global redundancy rate that modulates cost-effectiveness when identifying previously characterized clones. Statistical investigations must be taken into account to optimize such choices.

3.7.3. presence of chimeric clones

Moreover, preamplification could lead to chimeric probes by PCR jumping [41]. In the first cycles of PCR amplification, as all sequences contain the same microsatellite repeat, it is possible that two different newly produced, but partially synthesized, fragments hybridize onto this complementary microsatellite repeat and produce a new double-stranded DNA molecule containing the left flanking region of one microsatellite locus and the right flanking region of another unrelated one. Due to the quite high annealing temperature of the RSA21 primer (60 °C) in the preamplification step, PCR jumping should not appear with $(GA)_n$ microsatellites below 10 repeats. The risk increases for higher numbers of repeats, but this phenomenon should not be of major concern, as we only found one chimeric probe in 100 loci analyzed.

4. conclusion

Our method described here should be suitable for enrichment of all types of microsatellite. The technical steps should not be affected by switching microsatellite types. We used dinucleotide repeats (GA), but we assume that other dinucleotide, trinucleotide or tetranucleotide repeat types are also of interest. This study gives a technical basis for easy microsatellite marker development in tropical crops, until the international scientific community starts up molecular projects applied to the genetic improvement of these traditionally orphan crops.

Protocol for developing microsatellite-enriched libraries

Step 1. Shearing of DNA

- purified total genomic DNA solution (250 ng.µL^{-1})
- *Pst*I endonuclease and 10X reaction buffer
- 40 mM spermidin
- sterile distilled water
- 1XTBE buffer (89mM tris boric acid, 1mM EDTA)
- agarose electrophoresis grade
- electrophoresis tank
- transformator
- water bath

1. Aliquot 40 µL of purified total genomic DNA solution into an eppendorf.
2. Mix 100 units of *Pst*I endonuclease (10 units·µg^{-1} of DNA), 10 µL of reaction buffer, 10 µL of 40 mM spermidin, sterile distilled water to 100 µL.
3. Incubate overnight at 37 °C.
4. Inactivate the endonuclease at 65 °C for 15 min in the water bath.
5. Prepare a 1% agarose gel.
6. Load 10 µL of restricted DNA onto the gel and run.
7. Analyse shearing of DNA (the distinct band of DNA in the limited mobility region should disapear into lower molecular weight DNA smear).

Step 2. 'Polishing' DNA fragment ends

- T4 DNA polymerase and 5X reaction buffer
- sterile distilled water
- 2.5 mM dNTP
- water bath

1. Aliquot 10 µL (1 µg) of restricted DNA.
2. Mix 2.5 units of T4 DNA polymerase, 6 µL of reaction buffer, steril distilled water to 30 µL.
3. Incubate in a water bath at 37 °C for 30 min.
4. Add 1 µL of dNTP.
5. Incubate for 30 min at 37 °C.
6. Inactivate for 10 min at 70 °C.

Step 3. Adaptator ligation and PCR preamplification

Adaptator ligation

- RSA21 primer (10 µM)
- RSA25 adaptator (10 µM)
- T4 DNA ligase and 5X reaction buffer
- sterile distilled water

1. Mix 1 µg of "polished" restricted DNA, 1 µL of RSA21, 1 µL of RSA25, 40 µL of reaction buffer, 4 units of T4 DNA ligase, sterile distilled water to 200 µL.
2. Incubate for 2 h at 20 °C.

PCR preamplification

- *Taq* DNA polymerase and 10X reaction buffer with $MgCl_2$
- RSA21 primer (10 µM)
- 2.5 mM dNTP
- sterile distilled water
- thermocycler
- anion exchange centrifuge microcolumns

1. Aliquot 1 µL of the ligated DNA solution.
2. Mix 2.5 µL of RSA21, 2 µL of dNTP, 2.5 µL of reaction buffer, 1 unit of *Taq* DNA polymerase, sterile distilled water to 25 µL.
3. Add 2 drops of mineral oil.
4. Run the PCR program: 95 °C for 1 min, 20 cycles of [94 °C for 40 s, 60 °C for 60 s, 72 °C for 120 s], 72 °C for 5 min.
5. Electrophorese 10 µL of PCR products on 1% agarose gel.
6. Analyse fragments size (smear should be similar to that obtained in step 1).
7. Purify the rest of PCR products through microcolumns and resuspend in 100 µL of water.

Step 4. Selection of DNA fragments containing (GA) microsatellite sequences

- streptavidin-coated magnetic beads (Magnetosphere Magnetic Separation Products kit, Promega Madison, USA)
- 50 µM biotin-labelled microsatellite oligoprobe
- 20X SSC, 0.5X SSC, 0.1X SSC
- heating block
- magnet
- sterile distilled water

1. Mix 400 µL of water with the 100 µL of purified PCR products.
2. Incubate at 9 °C for 15 min.
3. Add 3 µL of oligoprobe and 13 µL of 20X SSC.
4. Leave at room temperature for 20 min.
5. Wash three times 600 µL of magnetic beads with 300 µL of 0.5X SSC and resuspend in 100 µL of 0.5X SSC.
6. Mix the 100 µL of prewashed magnetic beads with 516 µL of hybridization mix.
7. Incubate 10 min at room temperature.
8. Magnetize and discard the liquid.
9. Resuspend in 300 µL of 0.1X SSC.
10. Repeat step 8 and 9, three times.
11. Elute with 100 µL of water, and finally with an additional volume of 150 µL.
12. Store if necessary at 20 °C.

Step 5. PCR amplification of the selected fragments

- *Taq* DNA polymerase and 10X reaction buffer with $MgCl_2$
- RSA21 primer (10 µM)
- 2.5 mM dNTP
- sterile distilled water
- thermocycler
- anion exchange centrifuge microcolumns

1. Aliquot 10 µL of the eluted fraction from step 4.
2. Mix 10 µL of RSA21, 8 µL of dNTP, 10 µL of reaction buffer, 4 unit of *Taq* DNA polymerase, sterile distilled water to 100 µL.
3. Add 2 drops of mineral oil.
4. Run the PCR program: 95 °C for 1 min, 20 cycles of [94 °C for 40 s, 60 °C for 60 s, 72 °C for 120 s], 72 °C for 5 min.
5. Prepare a 1% agarose gel.
6. Load 10 µL of PCR products onto the gel and run.
7. Analyse fragment sizes (discrete bands should not appear through the smear compared to unselected fragments smear from step 3).
8. Purify the rest of PCR products through a microcolumn and resuspend in 15 µL of water.
9. Microsatellite enriched fragments are ready to be cloned.

references

[1] Love J.M., Knight A., McAleer M., Todd J.A., Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites, Nucleic Acids Res. 18 (1990) 4123–4130.

[2] Serikawa T., Kuramoto T., Hilbert P., Mori M., Yamada J., Dubay C.J., Lindpainter K., Gantezn D., Guenet J.L., Lathrop G.M., Beckman J.S., Rat gene mapping using PCR-analysed microsatellites, Genetics 131 (1992) 701–721.

[3] Weissenbach J., Gyapay G., Dib C., Vignal A., Morissette J., Millaseau P., Vaysseix G., Lathrop M., A second generation map of the human genome, Nature 359 (1992) 794–801.

[4] Akagi H., Yokozeki Y., Inagaki A., Fujimura T., Microsatellite DNA markers for rice chromosomes, Theor. Appl. Genet. 93 (1996) 1071–1077.

[5] Brondani R.P.V., Brondani C., Tarchini R., Grattapaglia D., Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla,* Theor. Appl. Genet. 97 (1998) 816–827.

[6] Chen X., Temnykh S., Xu Y., Cho Y.G., McCouch S.R., Development of a microsatellite framework map providing genome-wide coverage in rice (*Orysa sativa* L.), Theor. Appl. Genet. 95 (1997) 4 553–567.

[7] Condit R., Hubbel S.P., Abundance of DNA sequence of two-base repeat regions in tropical tree genomes, Genome 34 (1991) 66–71.

[8] Jarne P., Lagoda P.J.L., Microsatellites, from molecules to populations and back, Trends Ecol. Evol. 11 (10) (1996) 385–444.

[9] Taramino G., Tarchini R., Ferrario S., Lee M., Pe' M.E., Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*, Theor. Appl. Genet. 95 (1997) 66–72.

[10] Chowdari K.V., Venkatachalam S.R., Davierwala A.P., Gupta V.S., Ranjekar P.K., Govila O.P., Hybrid performance and genetic distance as revealed by the $(GATA)_4$ microsatellite and RAPD markers in pearl millet, Theor. Appl. Genet. 97 (1998) 163–169.

[11] Dehmer K.J., Friedt W., Evaluation of different microsatellite motifs for analysing genetic relationships in cultivated sunflower *Helianthus annuus* L., Plant Breeding 117 (1) (1998) 45–48.

[12] Grapin A., Noyer J.L., Carreel F., Dambier D., Baurens F.C., Lanaud C., Lagoda P.J.L., Diploid *Musa acuminata* genetic diversity assayed with sequence-tagged-microsatellite sites, Electrophoresis 19 (1998) 1374–1380.

[13] Sefc K.M., Steinkellner H., Glössl J., Kampfer S., Regner F., Reconstruction of a grape wine pedigree by microsatellite analysis, Theor. Appl. Genet. 97 (1998) 227–231.

[14] Goldstein D.B., Pollock D.D., Launching microsatellites: a review of mutation processes and methods of phylogenetic inference, J. Hered. 88 (1997) 335–342.

[15] McCouch S.R., Chen X.L., Panaud O., Temnykh S., Xu Y., Cho Y.G., Huang N., Ishii T., Blair M., Microsatellite marker development, mapping and application in rice genetics and breeding, Plant Mol. Biol. 35 (1997) 1–2, 89–99.

[16] Panaud O., Chen X., McCouch S.R., Development of microsatellite markers and characterization of simple sequence lenght polymorphism (SSLP) in rice (*Oryza sativa* L.), Mol. Gen. Genet. 252 (1996) 597–607.

[17] Pfeiffer A., Olivieri A.M., Morgante M., Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.), Genome 40 (4) (1997) 411–419.

[18] Röder M.S., Plaschke J., Köning S.U., Börner A., Sorrells S.E., Tanksley S.D., Ganal M.W., Abundance, variability and chromosomal location of microsatellites in wheat, Mol. Gen. Genet. 246 (1995) 327–333.

[19] Smith J.S.C., Chin E.C.L., Shu H., Smith O.S., Wall S.J., Senior M.L., Mitchell S.E., Kresovich S., Ziegle J., An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.), comparisons with data from RFLPs and pedigree, Theor. Appl. Genet. 95 (1997) 163–173.

[20] Terauchi R., Konuma A., Microsatellite polymorphism in *Dioscorea tokoro*, a wild yam species, Genome 37 (1994) 794–801.

[21] Crouch H.K., Crouch J.H., Jarret R.L., Cregan P.B., Ortiz R., Segregation at microsatellite loci in haploid and diploid gametes of *Musa*, Crop. Sci. 38 (1) (1998) 211–217.

[22] Kaemmer D., Fischer D., Jarret R.L., Baurens F.C., Grapin A., Dambier D., Noyer J.L., Lanaud C., Kahl G., Lagoda P.J.L., Molecular breeding in the genus *Musa*: a strong case for STMS marker technology, Euphytica 96 (1997) 49–63.

[23] Kresovich S., Szewc-McFadden A.K., Bliek S.M., McFerson J.R., Abundance and characterization of simple sequence repeats (SSRs) isolated from a size-fractionated library of *Brassica napus* L. (rapeseed), Theor. Appl. Genet. 91 (1995) 206–211.

[24] Karagyozov L., Kalcheva I.D., Chapman V.M., Construction of random small insert genomic libraries highly enriched for simple sequence repeats, Nucleic Acids Res. 21 (16) (1993) 3911–3912.

[25] Ostrander E.A., Jong P.M., Rine J., Duyk G., Construction of small-insert genomic libraries highly enriched for microsatellite repeat sequences, Proc. Natl. Acad. Sci. USA 89 (8) (1992) 3419–3423.

[26] Edwards K.J., Barker J.H.A., Daly A., Jones C., Karp A., Microsatellite libraries enriched for several microsatellite sequences in plants, BioTechniques 20 (1996) 758–760.

[27] Kijas J.M.H., Fowler J.C.S., Garbett C.A., Thomas M.R., Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles, BioTechniques 16 (4) (1994) 656–660,662.

[28] Gawel N.J., Jarret R.L., A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea,* Mol. Biol. Rep. 9 (1991) 262–266.

[29] Hanahan D., Studies on transformation of *Escherichia coli* with plasmids, J. Mol. Biol. 166 (1983) 557.

[30] Sambrook J., Fritsch E.F., Maniatis T., Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press, New York, NY, 1989.

[31] Cregan P.B., Bhagwat A.A., Akkay M.S., Romgwen J., Microsatellite mapping and fingerprinting in soybean, Method. Mol. Cell. Biol. 5 (1994) 49–61.

[32] Depeiges A., Goubely C., Lenoir A., Cocherel S., Picard G., Raynal M., Grellet F., Delseny M., Identification of the most represented repeated motifs in *Arabidopsis thaliana* microsatellite loci, Theor. Appl. Genet. 91 (1) (1995) 160–168.

[33] Ma Z.Q., Roder M., Sorrells M.E., Frequencies and sequence characteristics of di-,tri-, and tetra-nucleotide microsatellites in wheat, Genome 39 (1) (1996) 123–130.

[34] Cheung W.Y., Moore G., Money T.A., Gale M.D., *Hpa*II library indicates 'methylation-free islands' in wheat and barley, Theor. Appl. Genet. 84 (1992) 739–746.

[35] Lagoda P.J.L., Hypervariable sequenzen als genetisch marker in der onkogenese, thesis, Institut für Humangenetik, Homburg/Saar 1992, 142 pp.

[36] Benzinger E.A., Emerek E.A., Grigsby N.L., Duewer D.L., Lovekamp M.L., Deadman H., Thompson J.L., Sallee P.J., Riech A.K., Products of partial digestion with HaeIII, Part 1. Characterization, casework experience and confirmation of the theory of three-, four- and five-banded RFLP pattern origins using partial digestion, J. Forensic Sci. Sep. 42 (5) (1997) 850–863.

[37] Vos P., Hogers R., Bleeker M., Reijans M., van de Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M., Zabeau M., AFLP: A new technique for DNA fingerprinting, Nucleic Acids Res. 23 (1995) 4407–4414.

[38] Bennett M.D., Smith J.B., Nuclear DNA amounts in angiosperms, Philosophical Transactions of the Royal Society of London B 334 (1991) 309–345.

[39] Siebert P.D., Larrick J.W., PCR MIMICs: Competitive DNA fragments for use as internal standards in quantitaties PCR, BioTechniques 14 (1993) 244–249.

[40] Baurens F.C., Noyer J.L., Lanaud C., Lagoda P.J.L., Use of competitive PCR to assay copy number of repetitive elements in banana, Mol. Gen. Genet. 253 (1996) 57–64.

[41] Huang L.M., Jeang K.T., Long-range jumping of incompletely extended polymerase chain fragments generates unexpected products, BioTechniques 16 (2) (1994) 242–246.

Genotecas enriquecidas en microsatelites: metodología aplicada para el desarrollo de marcadores SSR en cultivos tropicales.

Resumén — Introducción. Los cultivos tropicales son de gran importancia para los países en desarrollo pero pueden ser considerados como cultivos huérfanos en terminos de su abundancia en marcadores moleculares. Los marcadores microsatelites presentan muchas ventajas comparados a otros marcadores de uso corriente (isoenzimas, RFLP y RAPD). Nuestro objetivo fué el de desarrollar un método para construir genotecas enriquecidas en microsatelites a

partir del ADN de cultivos tropicales. Se describe una técnica simple y rápida para la construcción de genotecas enriquecidas en microsatelites (GA) en la palma aceitera. **Procedimiento para la obtención de genotecas enriquecidas en microsatelites**. Los dos principios de la técnica utilizada son la hibridación con sondas oligonucleotídicas biotinadas y la captura de las secuencias seleccionadas mediante partículas magnéticas cubiertas de streptavidina. **Caracterización de las genotecas enriquecidas**. Las genotecas con miles de clones tienen una proporción de clones positivos (conteniendo un microsatelite) superior al 70%. La tasa de clones redudantes dependió del método de preparación del ADN fué de 20 % cuando se obtuvieron los fragmentos de ADN mediante restricción por *Pst*I y de 60% para el método de fragmentación del ADN por sonicación. **Conclusión**. Este estudio aporta las bases técnicas para un método facil de desarrollo de marcadores microsatelites en plantas tropicales, este método puede ser de importancia hasta que la comunidad científica internacional se decida a desarrollar proyectos en biología molecular aplicados al mejoramiento genetico de estos cultivos tradicionalmente abandonados. (© Elsevier, Paris)

Francia / métodos / biología molecular / marcadores genéticos / microsatelites / bases de datos / secuencia nucleotidica