Development and characterization of a new set of 164 polymorphic EST-SSR markers for diversity and breeding studies in rubber tree (Hevea brasiliensis Mull. Arg.)

Philippe Cubry, Valérie Pujade-Renaud, Dominique Garcia, Sandra Espéout, Vincent Le Guen, Francoise Granet, Marc Seguin

Plant Breeding, **133**, 419–426 (2014)
doi:10.1111/pbr.12158

Development and characterization of a new set of 164 polymorphic EST-SSR markers for diversity and breeding studies in rubber tree (*Hevea brasiliensis* Müll. Arg.)

Philippe Cubry[1,2,5], Valérie Pujade-Renaud[3,4], Dominique Garcia[1], Sandra Espeout[1], Vincent Le Guen[1], Françoise Granet[2] and Marc Seguin[1]

[1]UMR-AGAP, CIRAD, F-34398, Montpellier, France; [2]Manufacture Française des Pneumatiques MICHELIN, F-63040, Clermont-Ferrand Cedex, France; [3]UMR-AGAP, CIRAD, F-63000, Clermont-Ferrand, France; [4]UMR 547 PIAF, Clermont Université, Université Blaise Pascal, F-63000, Clermont-Ferrand Cedex, France; [5]Corresponding author, E-mail: philippe.cubry@hotmail.fr

With 3 figures and 5 tables

Received August 7, 2013/Accepted December 14, 2013
Communicated by F. Ordon

Abstract

Despite its economic importance and recent genome release, the need for molecular tools for *Hevea brasiliensis* is high. In the frame of a disease resistance study, EST sequences were retrieved from public database or generated by sequencing SSH libraries. Sequences were trimmed and microsatellite motifs searched using an ad hoc bioinformatic pipeline, and pairs of primers for the amplification of candidate markers were generated. We found a total of 10 499 unigenes from both sources of sequences, and 673 microsatellites motifs were detected using the default parameters of the pipeline. Two hundred sixty-four primer pairs were tested and 226 (85.6%) successfully amplified. Out of the amplified candidate markers, 164 exhibited polymorphism. Relationships based on dendrograms using simple matching index and diversity statistics based on EST-SSRs were compared with Genomic SSRs, showing the potentialities of EST-derived microsatellites for resistance studies but also for population genetics approaches.

Key words: EST-SSR — microsatellites — rubber tree — *Hevea* spp. — *Microcyclus ulei*

Hevea brasiliensis (Willd. ex A. Juss.) Müll. Arg. (rubber tree) is the main source of natural rubber in the world, counting 11.3 million tons in 2012 (www.rubberstudy.com). It is a perennial cross-pollinating and monoecious plant of the Euphorbiaceae family. Genetic breeding of *H. brasiliensis* is difficult given its low rate of success for controlled pollinations and its long life cycle. Propagation of a new cultivar requires at least 20–25 years of breeding experiments (Le Guen et al. 2009).

Molecular tools are useful for the characterization of genotypes and can improve and speed up the breeding scheme. They have been extensively used for the last two decades to study genetic diversity and genetic mapping in *H. brasiliensis* (Besse et al. 1994, Lespinasse et al. 2000, Seguin et al. 2003). Molecular markers used in *Hevea* encompass isozymes, randomly amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP™) and, more recently, microsatellites and single-nucleotide polymorphism (SNP) (Triwitayakorn et al. 2011, Li et al. 2012).

Microsatellite markers, also known as simple sequence repeats (SSR), are particularly interesting as they are codominant, relatively cheap to use, repeatable and highly polymorphic. Despite the growing availability of SNP markers, SSRs are still the most used molecular markers by small and intermediate breeding companies that might not have access to expensive SNP-genotyping platforms (Tyrka et al. 2008).

However, development of SSR-enriched libraries for genomic SSRs (gSSR) is still costly and time consuming. More than 450 gSSR markers have already been published (Seguin et al. 2003, Le Guen et al. 2011). Souza et al. (2009) and Pootakham et al. (2012) published additional sets of 27 and 90 polymorphic gSSRs, respectively.

Expressed sequence tag (EST)-derived microsatellites (EST-SSRs) are developed from expressed sequences and therefore can be found in genes involved in agronomically important traits. EST-SSRs are less polymorphic and more transferable across taxa than gSSRs (Varshney et al. 2005, Ellis and Burke 2007). They can be used for genetic mapping, functional diversity and comparative mapping. The development of EST-SSRs is mainly based on the study of genomic sequences that are not specifically obtained for SSR research (e.g. publicly available EST sequences, construction of cDNA libraries). This development is rapid and cost effective (Ellis and Burke 2007). SSR markers are becoming more important for breeding purposes when they are closely linked or diagnostic to important traits such as *Sbm1* in wheat (Perovic et al. 2009), and EST-SSRs should fall into this class.

Sets of polymorphic EST-SSRs from rubber tree have been published (Feng et al. 2009, Triwitayakorn et al. 2011), but none has been specifically developed to study resistance to biotic stresses. *H. brasiliensis* is susceptible to the Ascomycota *Microcyclus ulei* (Henn.) Arx 1962, the pathogen causing the South American leaf blight (SALB), a devastating disease affecting most of the currently cultivated varieties. Despite its economic impact, few molecular tools are available to study the sources of resistance of *H. brasiliensis* to *M. ulei.*

The objective of this study was to develop a new set of EST-SSRs from disease-related candidate genes that could be used in resistance studies. These markers will also help saturating existing genetic maps, as this saturation with different types of molecular markers is highly important for genetic studies and breeding purposes (Gupta et al. 2013). We chose to address two kinds of EST sources: (i) disease resistance and defence-related genes from publicly available EST databases and (ii) sequences

obtained from suppression subtractive hybridization (SSH) libraries built as part of a global team project to study resistance genetics (Garcia et al. 2011).

Materials and Methods

Plant material: A set of 19 *Hevea* accessions (Table 1) was chosen to assess the polymorphism and diversity of newly developed primers to amplify microsatellites detected in EST sequences.

Sampling of individuals: We first used genotyping data from previous studies (Le Guen 2008, Le Guen et al. 2009) to sample a set of nine genotypes encompassing the whole known diversity of the *H. brasiliensis* species. We chose the genotypes using the 'max-length subtree' heuristic procedure (Perrier et al. 2003) implemented in DARWIN 5.0 software (Perrier and Jacquemoud-Collet 2006). This stepwise procedure searches for a subset of units while pruning redundant units and limiting diversity loss. Starting from a dendrogram, distances between units in the tree are calculated, the pair of units of minimal distance is selected, and the unit with the smallest external edge is removed, until the sample size chosen by the user is reached (Perrier et al. 2003). For our sampling, we used a neighbour-joining dendrogram based on a dissimilarity matrix for 264 genotypes and 18 genomic microsatellite markers, and a final sample size of nine genotypes.

In addition to this diversity set, we selected nine genotypes of particular interest to breeders (stated as improved material) or involved in present or future mapping progeny (complete list Table 1). We also included as outgroup, a *Hevea benthamiana* accession (F4542).

Sixteen of the 19 selected genotypes had already been used in a previous diversity study (Le Guen et al. 2009). Genotyping data for 18 genomic microsatellite markers derived from enriched libraries were retrieved for comparison purposes.

DNA extraction: DNA extraction was carried out from either ground fresh or frozen leaves according to the procedure by Le Guen et al. (2009). DNA concentration was evaluated using a Nanoquant plate read with an Infinite® 200 PRO NanoQuant device (Tecan Group Ltd., Männedorf, Switzerland), then standardized at 5 ng/µl as working solution.

EST sequence retrieving, microsatellite motif detection, primer design: We used two different sources of EST sequences: 1) SSH libraries built from infected leaves developed in a resistance study (Garcia et al. 2011, Berger et al. in prep) and 2) latex EST sequences of *H. brasiliensis* retrieved from public databases (GenBank). Both EST sources were analysed separately through the ESTtik pipeline (Argout et al. 2008).

SSH libraries: For SSH sequences, Sanger chromatograms were included in the pipeline. Base calling and vector trimming were carried out, and sequence quality was evaluated. Low-quality sequences were discarded, and clustering was achieved using default parameters of the ESTtik pipeline (Argout et al. 2008). An automated search for SSR motifs using MISA programme was carried out following the default parameters for the minimal number of repeats used in ESTtik pipeline (six for dinucleotide and five for tri-, tetra-, penta- and hexanucleotide). Primer design was performed with PRIMER3 software (Rozen and Skaletsky 2000) with default parameters for PCR amplification. We summarized the results of the detection by computing frequencies concerning the microsatellite motif classes. We also analysed the motif types for di- and trinucleotide microsatellites.

H. brasiliensis **cDNA public libraries:** We retrieved all *H. brasiliensis* cDNA sequences from EMBL/GenBank databases. Nearly all of the available sequences were from cDNA libraries built from latex, the cytoplasm of laticiferous cells. All the sequences were imported in ESTtik pipeline, but primer design was performed only for resistance- or defence-related genes. The putative function of the genes was established by ontology using Blast2GO (Conesa et al. 2005). Evaluation of sequence quality was carried out on the basis of sequence lengths because no chromatograms or sequence quality scores were available. We retained only sequences of more than 100 base pairs length. Clustering was performed, and the same procedure as above was applied for SSR motif search and primer design. For these latex EST sequences, we also performed an extra MISA search of microsatellite motifs using less stringent parameters (minimal number of repeats in the extended SSR search from latex sequences: five for dinucleotide, four for trinucleotide and three for tetra-, penta- and hexanucleotide). The results for both sets were summarized following the same procedure as above.

Genotyping experiments: PCRs were performed in a volume of 10 µl, composed of 25 ng of total DNA, 0.2 µM of each dNTP, 0.08 µM of forward primer tailed with an M13 sequence, 0.1 µM of reverse primer, 0.1 µM of a fluorescently labelled M13 primer, 10 mM Tris, 50 mM KCl and 2 mM $MgCl_2$. PCR amplifications were performed with an Eppendorf MasterCycler device (Eppendorf AG, Hamburg, Germany) as follows: an initial denaturation step of 5 min at 95°C, followed by a touch-down step of 15 cycles (95°C for 45-s denaturation, 59°C for

Table 1: List of the genotypes used in the study

Code	Species	Set	Inferred diversity group according to Le Guen et al. 2009 & Seguin et al. 2003
PB260	*Hevea brasiliensis*	Mapping progeny parent	Wickham
RRIM600	*H. brasiliensis*	Mapping progeny parent	Wickham
IRCA109	*H. brasiliensis*	Mapping progeny parent	Wickham
MDF180	*H. brasiliensis*	Mapping progeny parent	East Acre – Madre de Dios (MDF)
RO38	*H. brasiliensis* × *H. benthamiana*	Mapping progeny parent	Interspecific hybrid
FX2784	*H. brasiliensis*	Improved material	Intraspecific Wickham × Amazonian hybrid
FDR5597	*H. brasiliensis*	Improved material	Intraspecific Wickham × Amazonian hybrid
FDR5788	*H. brasiliensis*	Improved material	Interspecific Wickham × *H. non-brasiliensis* hybrid
PFB5	*H. brasiliensis*	Improved material	Interspecific *H. brasiliensis* × *H. non-brasiliensis* hybrid
MT/IT/16/24	*H. brasiliensis*	Diversity set	Mato Grosso – Rondonia/Pimenta Bueno (RO/PB)
AC/B/15/25	*H. brasiliensis*	Diversity set	East Acre – Madre de Dios (MDF)
AC/S/12/86	*H. brasiliensis*	Diversity set	East Acre – Madre de Dios (MDF
AC/B/19/46	*H. brasiliensis*	Diversity set	East Acre – Madre de Dios (MDF)
RO/C/9/27	*H. brasiliensis*	Diversity set	Rondonia – Mato Grosso/Villa Bella (MT/VB)
RO/JP/3/37	*H. brasiliensis*	Diversity set	Rondonia – Mato Grosso/Villa Bella (MT/VB)
MT/C/2/46	*H. brasiliensis*	Diversity set	Mato Grosso – Rondonia/Pimenta Bueno (RO/PB)
RO/I/0/89	*H. brasiliensis*	Diversity set	Rondonia – Mato Grosso/Villa Bella (MT/VB)
Wickham seedling	*H. brasiliensis*	Diversity set	Wickham
F4542	*H. benthamiana*	Outgroup	*H. non-brasiliensis* species

1-min hybridization (− 0.5°C per cycle), 72°C for 1 min 30 elongation); this touch-down step was followed by 25 cycles without touch-down (95°C for 45-s denaturation, 52°C for 1-min hybridization, 72°C for 1 min 30 elongation) and a final elongation step (72°C for 5 min). Primer pairs that gave bad amplification or no product were tested again with the following PCR programme: 95°C for 5 min as initial denaturation step, 10 cycles of touch-down cycles (95°C for 45-s denaturation, 55°C for 1-min hybridization (−0.5°C per cycle), 72°C for 1 min 30 elongation), followed by 25 cycles without temperature decrement (95°C for 45-s denaturation, 50°C for 1-min hybridization, 72°C for 1 min 30 elongation) and a final elongation step (72°C for 5 min).

PCR products were denatured at 95°C for 3 min before being separated and revealed onto a 6.5% polyacrylamide gel on an automated Li-Cor 4300 DNA analyzer (Li-Cor Biosciences, Lincoln, Nebraska, USA). Gel images were retrieved into dedicated SAGA^{GT} software (Li-Cor Biosciences, Lincoln, Nebraska, USA), and allele calling was carried out manually. Results were then exported into MS EXCEL® software and formatted for the various softwares used.

Statistical analysis:

Diversity and polymorphism description: Polymorphism scoring of successfully amplified products was performed manually. Allele number, gene diversity (unbiased Nei's statistic), observed heterozygosity (H_o) and polymorphism information content (PIC) were computed with POWERMARKER 3.25 software (Liu and Muse 2005). A dissimilarity matrix was computed with DARWIN 5.0 and a simple matching distance measure:

$$d_{ij} = 1 - \frac{1}{L}\sum_{l=1}^{L}\frac{m_l}{\pi},$$

where d_{ij} is the dissimilarity between individuals i and j, L the number of loci, π the ploidy [here = 2, *H. brasiliensis* and *H. benthamiana* being diploid species, (Lespinasse et al. 2000)] and m_l the number of matching alleles for locus l. This dissimilarity matrix serves as a basis for computing a dendrogram using a neighbour-joining (NJ) algorithm (Saitou and Nei 1987).

Comparison between EST-SSRs and genomic SSRs: We used the data obtained by Le Guen et al. (2009), who used a stratified random selection of 18 genomic SSR markers among hundreds of polymorphic ones. They were chosen genetically independent and covering the whole *Hevea* genome (1 SSR per chromosome pair).

For comparison purposes, we also randomly selected 18 EST-SSR markers among the 164 polymorphic ones. We then drew NJ hierarchical trees for 16 individuals common to the Le Guen et al. (2009) study and our work on both types of markers.

Results

EST sequences from SSH libraries

A total of 16 079 sequences were used as input for ESTtik pipeline. After PHRED quality score evaluation, 12 466 were considered of good quality for further analyses. Clustering resulted in 6992 unigenes (1483 contigs and 5509 singletons). The automated search for microsatellite motifs found 289 motifs in 280 unigenes. The most abundant types detected were dinucleotide (53.3%, Table 2), followed by trinucleotide (36.7%). Two hundred and thirty-two primer pairs corresponding to unique unigenes were successfully designed.

GenBank EST sequences

We retrieved 10 849 latex-derived EST sequences from public databases (GenBank) and incorporated them into ESTtik pipeline. Following quality evaluation, 10 365 sequences were retained for further analysis. After clustering, 3507 unigenes (1301 contigs and 2206 singletons) were determined. Three hundred and eighty-four microsatellite motifs were detected using the default parameters for MISA. Di- and trinucleotide microsatellites were the most abundant (Table 2). We also observed a high proportion of compound microsatellites (20.6%). Resulting microsatellite regions were blasted against the ones obtained from SSH libraries to check for redundancy. As no redundant microsatellite was detected, we retained the whole set for further analysis.

Extended search for microsatellite motifs using less stringent parameters resulted in 757 microsatellite motifs, among which trinucleotide was the most abundant (35.4%, Table 2), followed by dinucleotide (34.3%). A selection on this 757 candidate

Table 3: Polymorphism of the successfully amplified EST-derived microsatellites

Type of repeat	Number of sequences containing microsatellites	Number of polymorphic markers	Percentage of polymorphic markers
Dinucleotide	116	90	78
Trinucleotide	84	59	70
Tetranucleotide	7	3	43
Pentanucleotide	5	4	80
Hexanucleotide	4	3	75
Compound	10	5	50
Total	226	164	73

Table 2: Microsatellite types detected in the EST sequences

	SSH libraries EST-SSRs		Latex EST-SSRs with default EsTtik research parameters[1]		Total default parameters[1]		Latex EST-SSRs with extended research parameters[1]	
Microsatellite	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
Dinucleotide	154	53.3	186	48.4	340	50.5	260	34.3
Trinucleotide	106	36.7	109	28.4	215	31.9	268	35.4
Tetranucleotide	10	3.5	4	1.0	14	2.1	86	11.4
Pentanucleotide	2	0.7	2	0.5	4	0.6	23	3.0
Hexanucleotide	3	1.0	4	1.0	7	1.0	28	3.7
Compound	14	4.8	79	20.6	93	13.8	92	12.2
Total	289	100.0	384	100.0	673	100.0	757	100.0

[1]Default parameters stand for using default length of core motif parameters for the search of microsatellite motifs using ESTtik pipeline (dinucleotide: 6, trinucleotide: 5, tetra- penta- and hexanucleotide: 5), extended parameters stand for using modified parameters (dinucleotide: 5, trinucleotide: 4, tetra- penta- and hexanucleotide: 3) for microsatellite detection using ESTtik pipeline.

markers list, based on putative gene functions of plant resistance or defence inferred from automatic ontology annotation and blasting, enabled us to retain 32 SSR-containing sequences with successfully designed primer pairs for testing (see Table S3).

Type of microsatellite repeats

The most abundant type of repeats among dinucleotide motifs using default parameters was AG/TC (77.6% of the dinucleotide repeats, Table 2), followed by AT/TA (18.8%) and CA/TG (3.5%). Among trinucleotide motifs, AAG/TTC was the most common (35.3% of the trinucleotide repeats), followed by AAT/TTA (15.8%) and ACC/TGG (12.6%). Tetra-, penta- and hexanucleotide motifs accounted for <5%.

Efficiency of the amplification and polymorphism of EST-SSR candidate markers

Of the 264 primer pairs tested, 226 (85.6%) gave amplification products. Polymorphism appeared relatively high within the

Table 4: Summary statistics for EST-SSR markers on 19 individuals

	Number of alleles		Gene diversity		Observed heterozygosity		PIC	
Type of repeat	Mean	SD[1]	Mean	SD[1]	Mean	SD[1]	Mean	SD[1]
Dinucleotide	7.4	0.4	0.62	0.02	0.45	0.02	0.61	0.02
Trinucleotide	3.9	0.3	0.42	0.03	0.33	0.03	0.4	0.03
Tetranucleotide	5.0	0.9	0.54	0.12	0.36	0.05	0.53	0.12
Pentanucleotide	3.8	1.0	0.49	0.03	0.42	0.1	0.43	0.04
Hexanucleotide	3.7	0.7	0.37	0.05	0.4	0.06	0.35	0.04
Compound	6.8	1.5	0.61	0.08	0.54	0.08	0.59	0.08
Whole	5.9	0.3	0.54	0.02	0.4	0.02	0.52	0.02

SD, standard deviation.
[1]SD is calculated based on 5000 iterations of a bootstrap procedure.

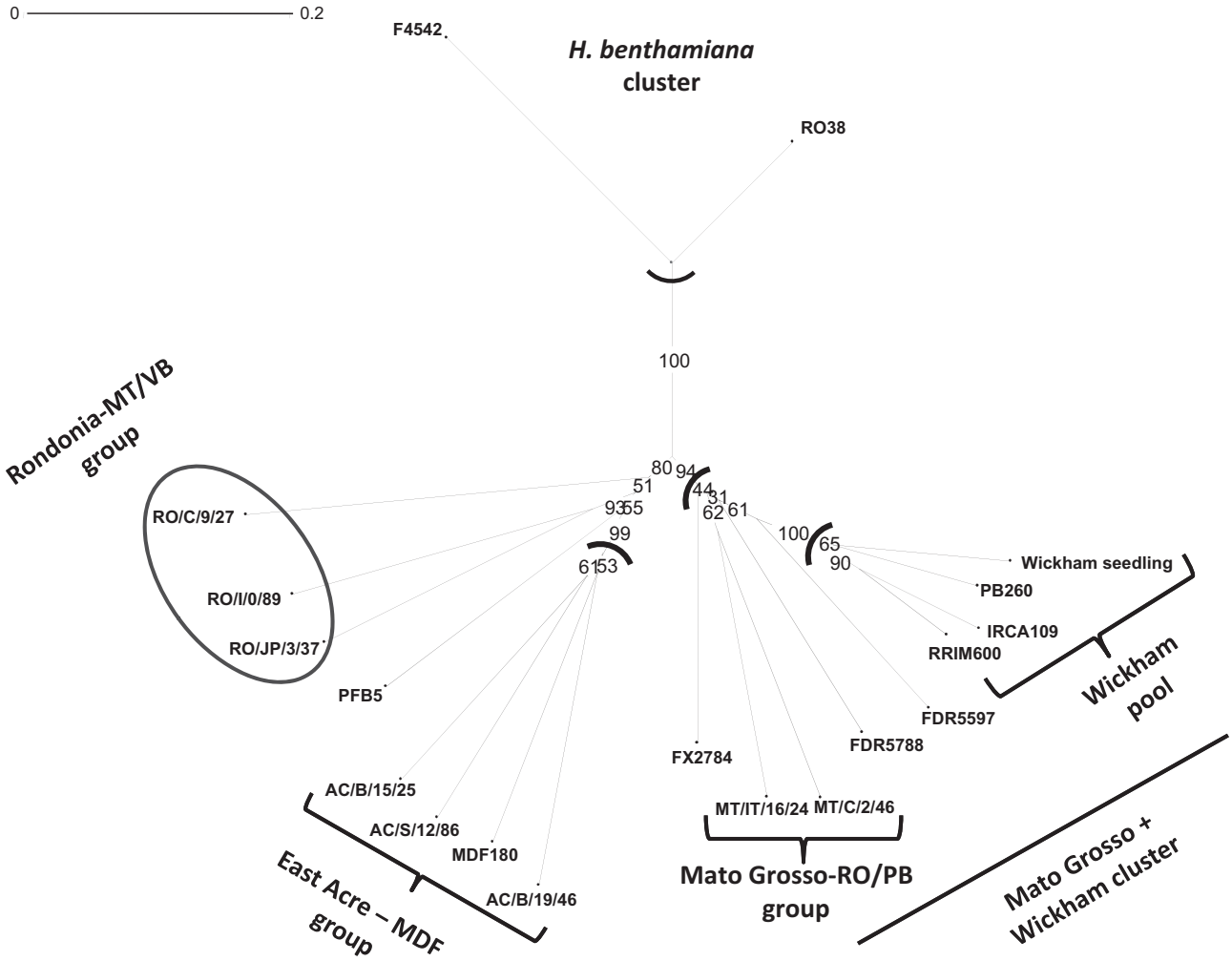

Fig. 1: Neighbour-joining tree based on dissimilarity matrix for 19 genotypes and 164 polymorphic EST-SSR markers. Bootstrap values are calculated on 5000 iterations

amplified sequences as 164 (72.6%) were polymorphic. In summary, of the 264 primer pairs, 62.1% were converted into polymorphic markers. A list of the 164 polymorphic markers is provided as Table S1. Additional information for microsatellites motifs detected including blast-derived annotation is provided in Table S2 for SSH libraries and in Table S3 for the 32 selected GenBank ESTs-derived SSRs.

Polymorphism appeared to depend on the length of the core repeat, as dinucleotide microsatellites exhibited 77.6% of polymorphic markers among the amplified products, whereas trinucleotides had 70.2% (Table 3).

Summary statistics and diversity of EST-SSRs for 19 genotypes

The summary statistics calculated from the 164 polymorphic EST-SSRs are shown in Table 4. Individual descriptive statistics per marker, including availability (rates of successful amplifica-

Table 5: Comparison of summary statistics for genomic and EST-derived microsatellite markers based on data for 16 genotypes

	Number of alleles		Gene diversity		Observed heterozygosity		PIC	
Marker set	Mean	SD[1]	Mean	SD[1]	Mean	SD[1]	Mean	SD[1]
164 EST-SSR markers	5.8	0.3	0.54	0.02	0.41	0.02	0.54	0.02
18 randomly chosen EST-SSR markers	6.2	0.6	0.65	0.03	0.47	0.04	0.63	0.04
18 genomic SSR markers	12.8	0.7	0.83	0.02	0.64	0.04	0.85	0.03

SD, standard deviation.
[1]SD is calculated based on 5000 iterations of a bootstrap procedure.

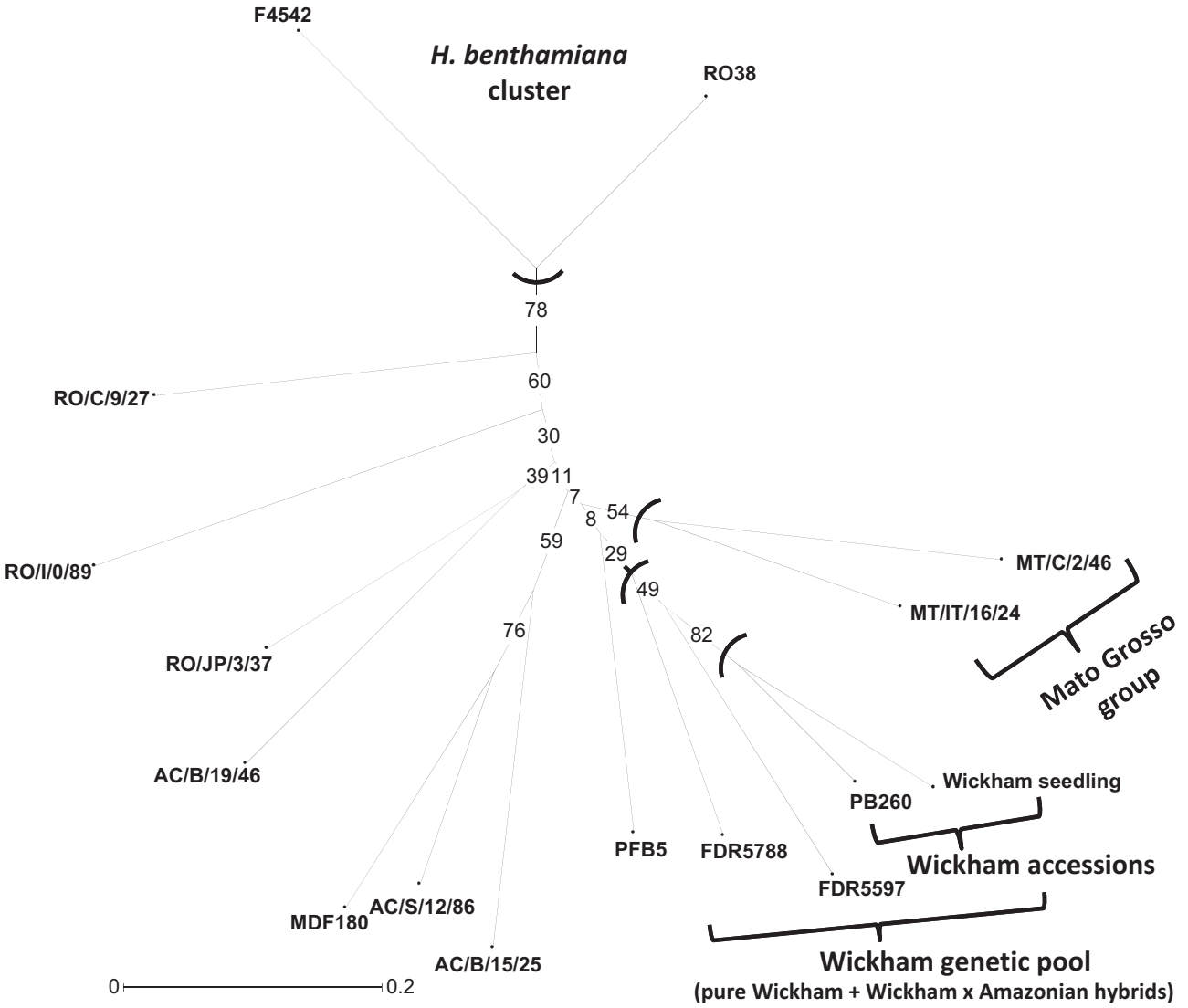

Fig. 2: Neighbour-joining tree based on dissimilarity matrix for 16 genotypes and 18 EST-SSR markers. Bootstrap values are calculated on 5000 iterations

tions), allele number, gene diversity, observed heterozygosity (H_O) and exact test of Hardy–Weinberg equilibrium, are provided as Table S1. The average number of alleles ranges from 3.69 to 7.42 for the hexanucleotide and dinucleotide EST-derived microsatellites, respectively, with a mean of 5.94 for all the markers. The dinucleotide microsatellites appeared to be the most polymorphic with the highest mean number of alleles, gene diversity and PIC values. With regard to the observed heterozygosity, only compound microsatellites had a higher value than dinucleotide, but this result should be taken with caution given the small number (10) of compound microsatellites.

The NJ tree based on the dissimilarity matrix calculated with the whole set of 164 markers (Fig. 1) displays a well-defined genetic structure within *H. brasiliensis* germplasm. This clustering is consistent with known genetic structure of Amazonian populations (Le Guen et al. 2009). It includes the three main groups (Mato Grosso, Acre and Rondonia) described, with Wickham genetic pool correctly placed within Mato Grosso origin. Nodes are supported by high bootstrap values regarding the use of microsatellite markers.

Summary statistics and diversity of EST-SSRs for 16 genotypes and comparison with genomic SSRs

The results of all the statistics of the EST-SSR data sets for 16 individuals were close to those obtained with 19 individuals (Table 5). When considering the genomic markers, we observed an important increase in all the statistics, in particular with a number of alleles twice that of the EST-SSR (12.8 and 5.8, respectively). Gene diversity, observed heterozygosity and PIC values of genomic SSRs were also significantly higher than those of EST-SSRs.

The dendrograms of Figs 2 and 3 for both types of markers show a comparable global view of genetic diversity, with the different groups clearly separated. However, EST-SSRs present a more robust tree with well-supported nodes. Some of the separations are more accurate and closer to the whole diversity study conducted by Le Guen et al. (2009) when considering EST-SSRs rather than gSSRs, except for the Acre vs. Rondonia separation.

Discussion and Conclusion

Our newly developed frameset of EST-SSR markers in the frame of a study of *H. brasiliensis* resistance to *M. ulei* appears to amplify better than those developed by Feng et al. (2009). This could be explained by the absence of mononucleotide repeats in our study, and differences in the parameters used for SSR detection. These markers may be valuable tools for resistance studies. Indeed, either they originate from SSH libraries built from the identification of resistance-related genes or they were selected based on gene ontology in relation to resistance.

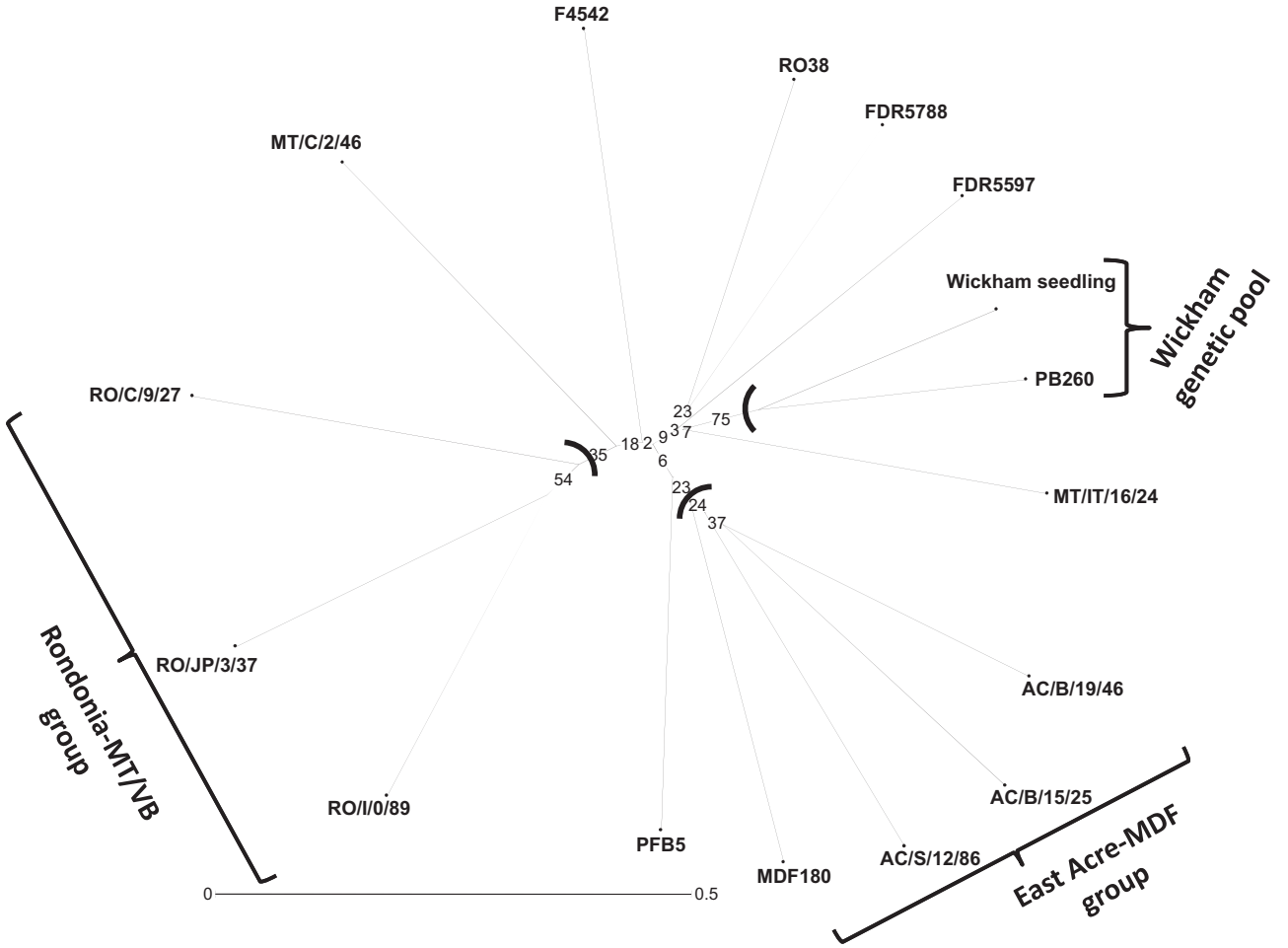

Fig. 3: Neighbour-joining tree based on dissimilarity matrix for 16 genotypes and 18 genomic microsatellites, data retrieved from Le Guen et al. 2009. Bootstrap values are calculated on 5000 iterations

As EST-SSR markers appear to have better transferability than gSSRs (Varshney et al. 2005, Kumar Yadav et al. 2010, Wen et al. 2010), they may also prove useful in comparative genomics, evolutionary studies, population genetics and association studies.

Our study based on a core representation of the whole genetic diversity of *H. brasiliensis* species validates these markers as powerful tools for population genetics, whether they are closely linked to or part of expressed genes. The relatively low number of alleles observed in EST-SSR markers compared with gSSR markers (twice as low) indicates a lower mutation rate and better suitability for such studies and association mapping. These studies perform better with a fewer number of alleles, both in terms of statistical power (Gupta et al. 2005, Rafalski 2010) and extent of linkage disequilibrium (Cubry et al. 2013). The lower mutation rate of EST-SSRs compared with gSSRs also reduces the risk of homoplasy when studying the demographic history of rubber tree populations.

We compared the genetic structure in both types of markers using data on gSSRs from Le Guen et al. (2009). The NJ tree obtained with EST-SSRs was more robust than the one obtained with gSSRs. The lower resolution of Acre vs. Rondonia separation with EST-SSRs may be due to the close origin of both populations, better characterized by markers exhibiting a higher mutation rate. At species level, EST-SSRs seem to provide clearer clusters of diversity, but at a population level, gSSRs can identify a very fine structure more accurately.

Distribution of the core motifs was consistent with previous studies on different plants, including rubber tree or tomato (Feng et al. 2009, Ohyama et al. 2009). Higher diversity indices were obtained with dinucleotide core-length microsatellites. This result has also been reported in several studies, for example Vigouroux et al. (2005).

This new set of markers provides molecular breeders and researchers with powerful tools, giving the opportunity to map EST-SSRs potentially related to resistance on already developed progeny. They will also contribute to filling in the gaps that still exist in genetic maps.

Acknowledgements

This work was performed as part of the CIRAD – Michelin collaborative project 'Genesalb' supported by a grant from the French National Research Agency (Agence Nationale pour la Recherche; contrat ANR/ Génoplante n° GPLA07017C). We thank the Montpellier Languedoc-Roussillon Grand plateau technique régional for hosting the SSR genotyping activities. ESTtik database was made available by the SouthGreen bioinformatics platform at http://southgreen.cirad.fr/. We thank Ronan Rivallan for technical help, Xavier Argout for computational procedures and data retrieving, Gérald Oliver for keeping some vegetal material in a greenhouse in Montpellier, Jean-Marc Thevenin and Fabien Doaré for providing us with some fresh material from CIRAD's Pointe Combi Centre (French Guyana) collection.

References

Argout, X., O. Fouet, P. Wincker, K. Gramacho, T. Legavre, X. Sabau, A. Risterucci, C. Da Silva, J. Cascardo, M. Allegre, D. Kuhn, J. Verica, B. Courtois, G. Loor, R. Babin, O. Sounigo, M. Ducamp, M. Guiltinan, M. Ruiz, L. Alemanno, R. Machado, W. Phillips, R. Schnell, M. Gilmour, E. Rosenquist, D. Butler, S. Maximova, and C. Lanaud, 2008: Towards the understanding of the cocoa transcriptome: production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* L. generated from various tissues and under various conditions. BMC Genomics **9**, 512.

Besse, P., M. Seguin, P. Lebrun, M. H. Chevallier, D. Nicolas, and C. Lanaud, 1994: Genetic diversity among wild and cultivated populations of *Hevea brasiliensis* assessed by nuclear RFLP analysis. Theor. Appl. Genet. **88**, 199—207.

Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon, and M. Robles, 2005: Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21**, 3674—3676.

Cubry, P., F. De Bellis, K. Avia, S. Bouchet, D. Pot, M. Dufour, H. Legnate, and T. Leroy, 2013: An initial assessment of linkage disequilibrium (LD) in coffee trees: LD patterns in groups of Coffea canephora Pierre using microsatellite analysis. BMC Genomics **14**, 10.

Ellis, J. R., and J. M. Burke, 2007: EST-SSR as a resource for population genetic analyses. Heredity **99**, 125—132.

Feng, S. P., W. G. Li, H. S. Huang, J. Y. Wang, and Y. T. Wu, 2009: Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). Mol. Breed. **23**, 85—97.

Garcia, D., N. Carels, D. M. Koop, L. A. de Sousa, S. J. de Andrade Junior, V. Pujade-Renaud, C. R. Reis Mattos, and J. C. de Mattos Cascardo, 2011: EST profiling of resistant and susceptible *Hevea* infected by *Microcyclus ulei*. Physiol. Mol. Plant Pathol. **76**, 126—136.

Gupta, P. K., S. Rustgi, and P. L. Kulwal, 2005: Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol. Biol. **57**, 461—485.

Gupta, S., K. Kumari, M. Muthamilara, A. Subramanian, and M. Prasad, 2013: Development and utilization of novel SSRs in foxtail millet [Setaria italica (L.) P. Beauv.]. Plant Breeding **132**, 367—374.

Kumar Yadav, H., A. Ranjan, M. H. Asif, S. Mantri, S. V. Sawant, and R. Tuli, 2010: EST-derived SSR markers in *Jatropha curcas* L.: development, characterization, polymorphism, and transferability across the species/genera. Tree Genet. Genomes **7**, 207—219.

Le Guen, V., 2008: Exploration de la diversité des résistances génétiques à la maladie sud-américaine des feuilles de l'hévéa (*Microcyclus ulei*) par cartographie et génétique d'association au sein de populations naturelles. Université Montpellier II, Montpellier, France.

Le Guen, V., F. Doaré, C. Weber, and M. Seguin, 2009: Genetic structure of Amazonian populations of *Hevea brasiliensis* is shaped by hydrographical network and isolation by distance. Tree Genet. Genomes **5**, 673—683.

Le Guen, V., C. Gay, T. C. Xiong, L. M. Souza, M. Rodier-Goud, and M. Seguin, 2011: Development and characterization of 296 new polymorphic microsatellite markers for rubber tree (*Hevea brasiliensis*). Plant Breeding **130**, 294—296.

Lespinasse, D., M. Rodier-Goud, L. Grivet, A. Leconte, H. Legnate, and M. Seguin, 2000: A saturated genetic linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite, and isozyme markers. Theor. Appl. Genet. **100**, 127—138.

Li, D., Z. Deng, B. Qin, X. Liu, and Z. Men, 2012: De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). BMC Genomics **13**, 192.

Liu, K., and S. V. Muse, 2005: PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics **21**, 2128—2129.

Ohyama, A., E. Asamizu, S. Negoro, K. Miyatake, H. Yamaguchi, S. Tabata, and H. Fukuoka, 2009: Characterization of tomato SSR markers developed using BAC-end and cDNA sequences from genome databases. Mol. Breed. **23**, 685—691.

Perovic, D., J. Förster, P. Devaux, D. Hariri, M. Guilleroux, K. Kanyuka, R. Lyons, J. Weyen, D. Feuerhelm, U. Kastirr, M. Röder, and F. Ordon, 2009: Mapping and diagnostic marker development for Soilborne cereal mosaic virus resistance in bread wheat. Mol. Breed. **23**, 641—653.

Perrier, X., and J. P. Jacquemoud-Collet, 2006: DARwin software: http://darwin.cirad.fr/darwin. 5 edn. Cirad, Montpellier, France.

Perrier, X., A. Flori, and F. Bonnot, 2003: Methods of data analysis. In: P. Hamon, M. Seguin, X. Perrier, and J. C. Glaszmann (eds), Genetic

Diversity of Tropical Crops, 31—64. CIRAD & Science Publ. Inc., Montpellier, France & Enfield (NH), USA.

Pootakham, W., J. Chanprasert, N. Jomchai, D. Sangsrakru, T. Yoocha, S. Tragoonrung, and S. Tangphatsornruang, 2012: Development of genomic-derived simple sequence repeat markers in Hevea brasiliensis from 454 genome shotgun sequences. Plant Breeding **131**, 555—562.

Rafalski, J. A., 2010: Association genetics in crop improvement. Curr. Opin. Plant Biol. **13**, 174—180.

Rozen, S., and H. J. Skaletsky, 2000: Primer3 on the WWW for general users and for biologist programmers. In: S. Krawetz, and S. Misener (eds), Bioinformatics Methods and Protocols: Methods in Molecular Biology, 365—386. Humana Press, Totowa, NJ.

Saitou, N., and M. Nei, 1987: The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**, 406—425.

Seguin, M., A. Flori, H. Legnate, and A. Clément-Demange, 2003: Rubber tree. In: P. Hamon, M. Seguin, X. Perrier, and J. C. Glaszmann (eds) Genetic Diversity of Tropical Crops, 277—305. CIRAD & Science Publ. Inc., Montpellier, France & Enfield (NH), USA.

Souza, L. M., C. C. Mantello, M. O. Santos, P. Souza Gonçalves, and A. P. Souza, 2009: Microsatellites from rubber tree (Hevea brasiliensis) for genetic diversity analysis and cross-amplification in six Hevea wild species. Conserv. Genet. Res. **1**, 75—79.

Triwitayakorn, K., P. Chatkulkawin, S. Kanjanawattanawong, S. Sraphet, T. Yoocha, D. Sangsrakru, J. Chanprasert, C. Ngamphiw, N. Jomchai, K. Therawattanasuk, and S. Tangphatsornruang, 2011: Transcriptome sequencing of *Hevea brasiliensis* for development of microsatellite markers and construction of a genetic linkage map. DNA Res. **18**, 471—482.

Tyrka, M., D. Perovic, A. Wardyńska, and F. Ordon, 2008: A new diagnostic SSR marker for selection of theRym4/Rym5 locus in barley breeding. J. Appl. Genet. **49**, 127—134.

Varshney, R. K., A. Graner, and M. E. Sorrells, 2005: Genic microsatellite markers in plants: features and applications. Trends Biotech. **23**, 48—55.

Vigouroux, Y., S. Mitchell, Y. Matsuoka, M. Hamblin, S. Kresovich, J. S. C. Smith, J. Jaqueth, O. S. Smith, and J. Doebley, 2005: An analysis of genetic diversity across the maize genome using microsatellites. Genetics **169**, 1617—1630.

Wen, M., H. Wang, Z. Xia, M. Zou, C. Lu, and W. Wang, 2010: Development of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha Curcas* L. BMC Res. Notes **3**, 42.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. List of the 164 newly developed microsatellites markers with repeat types, sequence origin, primer sequences.

Table S2. List of SSR markers detected in SSH libraries with original sequence and blast annotation.

Table S3. List of SSR markers detected in GenBank-retrieved ESTs with original sequence and blast annotation.