Development, characterization, validation, and mapping of SSRs derived from Theobroma cacao L.–Moniliophthora perniciosa interaction ESTs

Article *in* Tree Genetics & Genomes · October 2010

DOI: 10.1007/s11295-010-0282-1

CITATIONS
23

READS
182

10 authors, including:

KARINA PERES GRAMACHO
Comissão Executiva do Plano da Lavoura Cacaueira
184 PUBLICATIONS **2,658** CITATIONS
SEE PROFILE

José Luis Pires
Comissão Executiva do Plano da Lavoura Cacaueira
38 PUBLICATIONS **575** CITATIONS
SEE PROFILE

Didier Clement
Cirad - La recherche agronomique pour le développement
35 PUBLICATIONS **1,508** CITATIONS
SEE PROFILE

Uilson Vanderlei Lopes
Comissão Executiva do Plano da Lavoura Cacaueira
17 PUBLICATIONS **193** CITATIONS
SEE PROFILE

Tree Genetics & Genomes
DOI 10.1007/s11295-010-0282-1

ORIGINAL PAPER

Development, characterization, validation, and mapping of SSRs derived from *Theobroma cacao* L.–*Moniliophthora perniciosa* interaction ESTs

Lívia Santos Lima · Karina Peres Gramacho · José Luis Pires · Didier Clement · Uilson Vanderlei Lopes · Nicolas Carels · Abelmon da Silva Gesteira · Fernanda Amato Gaiotto · Júlio Cézar de Mattos Cascardo · Fabienne Micheli

Received: 18 August 2009 / Revised: 9 December 2009 / Accepted: 9 February 2010

Abstract In this study, we report results of the detection and analysis of SSR markers derived of cacao–*Moniliophthora perniciosa* expressed sequence tags (ESTs) in relation to cacao resistance to witches' broom disease (WBD), and we compare the polymorphism of those ESTs (EST-simple sequence repeat (SSR)) with classical neutral SSR markers. A total of 3,487 ESTs was used in this investigation. SSRs were identified in 430 sequences: 277 from the resistant genotype TSH 1188 and 153 from the susceptible one Catongo, totalizing 505 EST-SSRs with three types of motifs: dinucleotides (72.1%), trinucleotides (27.3%), and tetranucleotides (0.6%). EST-SSRs were classified into 16 main categories; most of the EST-SSRs belonged to "Unknown function" and "No homology" categories (45.82%). A high frequency of SSRs was found in the 5'UTR and in the ORF (about 27%) and a low frequency was observed in the 3'UTR (about 8%). Forty-nine EST-SSR primers were designed and evaluated in 21 cacao accessions, 12 revealed polymorphism, having 47 alleles in total, with an average of 3.92 alleles per locus. On the other hand, the 11 genomic SSR markers revealed a total of 47 alleles, with an average of 5.22 alleles per locus. The

Communicated by J. Davis

Electronic supplementary material The online version of this article (doi:10.1007/s11295-010-0282-1) contains supplementary material, which is available to authorized users.

L. S. Lima · K. P. Gramacho (✉) · J. L. Pires · D. Clement · U. V. Lopes
CEPLAC/CEPEC/SEFIT,
Molecular Plant Pathology Laboratory,
Rodovia Ilhéus-Itabuna, km 22,
45600-970 Ilhéus, Bahia, Brazil
e-mail: karina@cepec.gov.br

J. L. Pires · U. V. Lopes
Plant Breeding Section,
Rodovia Ilhéus-Itabuna, km 22,
45600-970 Ilhéus, Bahia, Brazil

N. Carels · A. da Silva Gesteira · F. A. Gaiotto · J. C. de Mattos Cascardo · F. Micheli
UESC, DCB,
Rodovia Ilhéus-Itabuna, km 16,
45650-000 Ilhéus, Bahia, Brazil

A. da Silva Gesteira · F. A. Gaiotto · J. C. de Mattos Cascardo · F. Micheli
Laboratório de Genômica e Expressão Gênica,
Rodovia Ilhéus-Itabuna, km 16,
45650-000 Ilhéus, Bahia, Brazil

N. Carels
Núcleo de Biologia Computacional e Gestão de Informações Biotecnológicas,
Rodovia Ilhéus-Itabuna, km 16,
45650-000 Ilhéus, Bahia, Brazil

A. da Silva Gesteira
Embrapa Mandioca e Fruticultura Tropical,
Rua Embrapa, s/nº,
44380-000 Cruz das Almas, Bahia, Brazil

D. Clement · F. Micheli
CIRAD, UMR DAP,
Avenue Agropolis TA96/03,
34398 Montpellier cedex 5, France

Published online: 09 March 2010

association of EST-SSR with the genomic SSR enhanced the analysis of genetic distance among the genotypes. Among the 12 polymorphic EST-SSR markers, two were mapped on the F_2 Sca 6×ICS 1 population reference for WBD resistance.

Keywords Witches' broom disease · Microsatellite · Resistance · Genetic diversity

Abbreviations

AFLP	Amplified fragment length polymorphism
EST	Expressed sequence tag
HWE	Hardy–Weinberg equilibrium
LD	Linkage disequilibrium
LG	Linkage group
MAS	Marker-assisted selection
ORF	Open reading frame
PCA	Principal component analysis
PIC	Polymorphic information content
QTL	Quantitative trait locus
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
Sca 6	Scavina 6
SD	Standard deviation
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeats
TSH	Trinidad selected hybrid
UTR	Untranslated region, WBD, witches' broom disease
WBD	Witches' broom disease

Introduction

Cacao (*Theobroma cacao* L.) is a tropical sub-canopy tree originally from the rain forest of the Amazon basin. It is cultivated primarily to provide cacao liquor, butter, and powder for the chocolate industry, essentially for its flavor properties. Unfortunately, the cacao production is threatened by many pathogens such as *Moniliophthora* (*Crinipellis*) *perniciosa* (Stahel) Aime and Phillips-Mora (2005), the causal agent of the witches' broom disease (WBD) which has spread throughout Brazil, destroying cacao plantations and leading to important economical and social changes in affected areas (Trevizan and da Silva 1995; Luz et al. 2005). The main strategy for WBD control is the use of resistant varieties, and the major source of resistance is the Peruvian clones Scavinas (Pound 1938 and 1943). However, the fungus is adapting to this source (Rios-Ruiz 2001; Pires 2003; Albuquerque 2006; Paim 2006). Therefore, selection of new sources of resistance and accumulation of genes (pyramiding) in varieties became the priority of cacao breeding programs in producing countries.

With advancements in molecular marker technology, marker-assisted selection combined with conventional breeding has been one way in which the cacao resistance to WBD can be improved. To date, many types of DNA markers, including restriction fragment length polymorphisms, random amplified polymorphic DNAs, amplified fragment length polymorphisms, simple sequence repeats (SSR), and single nucleotide polymorphisms have been developed for cacao research (N'Goran et al., 1994; Lanaud et al. 1999; Flament et al. 2001; Motamayor et al. 2002, 2003; Clement et al. 2003; Borrone et al. 2004; Pugh et al. 2004; Borrone et al. 2007). However, only recently markers related to express genomic regions associated to WBD resistance have been developed (Lima et al. 2008). But, more markers are still needed for saturating the existing genetic maps and for marker-assisted breeding and germplasm collection management (Flament et al. 2001; Pugh et al. 2004).

Microsatellites (SSRs) consist of a variable number of tandem repeats of a simple motif sequence, typically a mono-, di-, tri-, or tetranucleotide repeat. Polymorphisms are detected by polymerase chain reaction (PCR) amplification with specific flanking primers and subsequent size sieving in agarose or denaturing polyacrylamide gels. SSR markers are the most favored for a variety of applications in plant genetics and breeding because of their multi-allelic nature, reproducibility, codominant inheritance, high abundance, and extensive genome coverage (Gupta and Varshney 2000). Recent studies have revealed that gene transcripts can also contain repeat motifs and the abundance of expressed sequence tags (ESTs) constitutes an attractive source of SSR markers (Kantety et al. 2002). Most of the microsatellites developed from ESTs, popularly known as EST-SSRs or genic SSRs, represent functional markers as a "putative function," which can be identified either by database search or by *in silico* approaches. Furthermore, EST-SSR markers are expected to possess high interspecific transferability as they belong to relatively conserved genic regions of the genome. With the increase of functional genomic studies, large datasets of ESTs have been developed, and by using bioinformatic tools it is possible to identify and develop EST-SSR markers at a large scale in a time and cost-effective manner (Han et al. 2006; Aggarwal et al. 2007).

In the last few years, breeders around the world have used genomic information to identify resistance genes to WDB in cacao. EST elicited with the pathogen *M. perniciosa* (Gesteira et al. 2007; Leal et al. 2007; Argout et al. 2008) and with non-biotic agents (Jones et al. 2002; Verica et al. 2004) have been obtained and analyzed. These ESTs constitute a potential source of new genes related to WBD resistance in cacao, as well as of non-neutral polymorphism (e.g., polymorphism occurring in or in vicinity of the coding sequence and directly related to its

function). We report here results on: (a) the detection and analysis of SSR markers on cacao-*M. perniciosa* ESTs; (b) the characterization of polymorphism SSRs for the cacao genotypes in relation to resistance to WBD; (c) the comparison of the polymorphism level in genic (ESTs) and genomic DNA using EST-SSR and classic neutral SSR markers, respectively.

Material and methods

Plant material

Cacao (*T. cacao* L.) genotypes used for EST-SSR validation were selected to include non-related Scavina genotypes aimed at the possibility of pyramiding resistance genes in future breeding programs. Based on the available information, they also correspond to different cacao genetic groups (Pires 2003). Twenty-three genetically distinct cacao genotypes, being 20 resistant clones from different origins and showing resistance in the field and in greenhouse (Table 1), one susceptible and two bulked DNA generated from six resistant and six susceptible plants from a segregating F_2 Sca 6×ICS 1(Faleiro et al. 2006) population for WDB resistance were used. This same F_2 population was used to carry out the genetic segregation of the polymorphic alleles from Sca 6 and ICS 1 and positioned in the map the corresponding EST-SSR markers.

EST sequencing

In this study, 3,487 EST sequences from cacao–*M. perniciosa* interaction obtained from functional study projects from the Cacao Research Center (CEPEC-BA, Brazil; CNPq research grant no. 471274/2006-2) and the Universidade Estadual de Santa Cruz (UESC, Brazil; Gesteira et al. 2007) were used. Briefly, these ESTs were obtained from cacao pods of TSH 1188 (resistant genotype) and from cacao meristems of Catongo (susceptible genotype) and TSH 1188, all elicited with *M. perniciosa*. In the preparatory step, for each EST generated, (a) the largest sequenced stretch with Phred quality ≥10 was extracted (allowing 1% nucleotide with Phred quality <10) using a Perl script (Ewing et al. 1998), (b) the plasmid vector sequence with cross-match (-minmatch 20, -minscore 5) was removed, (c) the "X" introduced by cross-match in the insert sequence with original nucleotides was substituted, and (d) the poly(A) tail was removed. After this trimming process, only sequences longer than 90 bp were included in the dataset and considered for further analysis. Finally, the redundancy was eliminated by contig assembling with CAP3 (Huang and Madan 1999) and the codon distribution was obtained from the remaining sequence pool.

Table 1 Cacao genotypes used for the diversity study related to resistance to witches' broom disease

Clone number	Genotype	Origin
1	R01	Peru
2	R02	Peru
3	R03	Trinidad
4	R04	Peru
5	R05	Ecuador
6	R06	Brazil
7	R07	Ecuador
8	R08	French Guiana
9	R09	Ecuador
10	R10	Brazil
11	R11	Ecuador
12	R12	Brazil
13	R13	Venezuela
14	R14	Ecuador
15	R15	Trinidad
16	R16	Nd
17	R17	Trinidad
18	R18	Colombia
19	R19	Brazil
20	Sca 6[a]	Peru
21	Catongo[b]	Brazil

Genotypes were chosen according to genetic differences and other characteristics such as resistance in the field or in greenhouse under natural or controlled conditions of cacao inoculation with *M. perniciosa* (Pires 2003)

nd non-determined

[a] Standard source of resistance

[b] Susceptible genotype

Putative sequence function analysis

For putative function determination and annotation, EST sequences were compared with the public sequence database (http://www.ncbi.nih.gov/BLAST/) using BLASTX and TBLASTX. Alignments showing similarity with an expected value $\leq 1.10^{-4}$ were considered significant. Additional information about the putative function of the ESTs was obtained using the ProDom (Corpet et al. 2000), the NRDL3D, and the Pfam softwares. Also the GO software (http://www.geneontology.org/) was used to produce a control vocabulary of the annotations (Harris et al. 2004). EST clusters and associated predicted proteins were manually inspected and annotated as described by Journet et al. (2002) and Gesteira et al. (2007).

EST-SSR identification and primer design

SSR marker identification was made on the cacao ESTs using a PEARL and SAS software (SAS 1988). The criteria

adopted in the SSR search were repeated stretches having a minimum of four repeat units. In our case, di-, tri-, and tetranucleotide SSRs. Primer pairs for EST-SSRs were designed using either Primer Design Report or PRIMER3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi; Table 2). The polymorphism and genetic diversity EST-SSR primers were compared to the 11 genomic DNA SSR primers (Table 2) chosen from previously studies in mapping studies (Lanaud et al. 1999; Faleiro et al. 2006). These primers were chosen based on their relation with the quantitative trait locus for WBD resistance localized in chromosome 9.

Location of the EST-SSR in relation to the coding sequence of the cDNA

Samples of ESTs containing SSRs were randomly chosen for localization of the EST-SSR in relation to the coding sequence of the cDNA. The open reading frame (ORF) was determined using the ORF Finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and the EST-SSR was localized in relation to the ORF. Three locations were possibly obtained: in the 5' untranslated region (5'UTR), in the ORF, or in the 3' untranslated region (3'UTR). In some cases, due to the EST sequence length or quality, it is not possible to clearly determine the ORF and consequently the location of the SSR in the cDNA.

DNA extraction, PCR amplification, and electrophoresis

Cacao genomic DNA was isolated from young leaves as described by Doyle and Doyle (1990). PCR amplifications were made on a Programmable Thermal Controller (MJ Research, Inc.) under the following conditions: final volume of 20 µl containing 30 ng of DNA, 0.2 mmol l^{-1} of each primer, 2 mmol l^{-1} of $MgCl_2$, 0.2 mmol l^{-1} of each dNTP (Ludwig Biotecnologia Ltda), 1× buffer, and 1 U of *Taq* DNA Polymerase (Ludwig Biotecnologia Ltda). PCR quality was checked on 1% TBE-BET agarose gel. Polymorphism evaluation was made by electrophoresis on 6% denaturing TBE acrylamide gel. Microsatellites polymorphisms were visualized by silver staining method according to Creste et al. (2001). The 10 bp molecular marker from Invitrogen was used as a reference to score the bands.

Table 2 *Theobroma cacao* microsatellite primers (SSR): SSR type (genomic or ESTs), locus identifier, forward (F) and reverse (R) primer sequences, GenBank accession number, and reference

SSR type	Marker analyzed	Primer Sequence (5' – 3')		Accession number	Reference
EST-SSR	msEstTsh-1	F: CACGAAGAAGTGGACGAT	R: CACATGGCTTGACTGGAA	AM851097	Lima et al. (2008)
	msEstTsh-2	F: ATTCCCTGCCCTCTTACG	R: CCAGATGTGGATGCGGAT	AM851098	Lima et al. (2008)
	msEstTsh-3	F: CGGGGAATCTCACACATA	R: ATCCTGGTTGGTGAGCTA	AM851096	Lima et al. (2008)
	msEstTsh-4	F: ATATCTCCACCACCACAG	R: CCGGAGAATGTAGAACCT	AM851099	Lima et al. (2008)
	msEstTsh-5	F: ACGACTTTAGGAGCTGACC	R: AACTTCAACACCAAGACCAT	AM851100	Lima et al. (2008)
	msEstTsh-6	F: ATGAATATTGTGGAGGAGGTT	R: TAGCAGTGCTTACAGCTCAA	AM851101	Lima et al. (2008)
	msEstTsh-7	F: GGAGCTGTTAGGAGAATGC	R: AGACCAGGAAAGAAGAGTCC	AM851102	Lima et al. (2008)
	msEstTsh-8	F: AACCCTTCATGAGACAATGA	R: CAGTCCCTTCTCTTCTGTGA	AM851103	Lima et al. (2008)
	msEstTsh-9	F: CACTTTTGACACTTCAAGCA	R: TCAAATCTTGACCCCATAAC	AM851104	Lima et al. (2008)
	msEstTsh-10	F: ACCCCTCAATCTCACACATA	R: GCTTGGCGCTCTTAGTATC	AM851105	Lima et al. (2008)
	msEstTsh-11	F: GGAGAAACACCTCTCATGTC	R: CTTTCTTCAAAGAAGGAAACAT	AM851106	Lima et al. (2008)
	msEstTsh-12[a]	F: GTCACTCCTGCAAGGCTAT	R: CAGATGGAAGAACGGATCTA	–	–
Genomic SSR	mTcCIR12	F :TCT GAC CCC AAA CCT GTA	R: ATT CCA GTT AAA GCA CAT	Y16986	Lanaud et al. (1999)
	mTcCIR26	F: GCA TTC ATC AAT ACA TTC	R: GCA CTC AAA GTT CAT ACT AC	Y16998	Lanaud et al. (1999)
	mTcCIR30	F: TGAAGATCCTACTGTTGAG	R: TGATAATAACTGCTTAGTGG	AJ271823	Lanaud et al. (1999)
	mTcCIR37	F: CTGGGTGCTGATAGATAA	R: AATACCCTCCACACAAAT	AJ271942	Lanaud et al. (1999)
	mTcCIR58	F: TTTTTGGTGATGGAACTAT	R: TGGTTAAGCAACACTAAACT	AJ271957	Lanaud et al. (1999)
	mTcCIR157	F: ACTAATGCTGTTGGCTTC	R: TCACTCGACTCGACTGTC	AJ566488	Pugh et al. (2004)
	mTcCIR166	F: ATGAACCACTATGTAAGACC	R: ATTCCAAAGGATTAGCAG	AJ566495	Pugh et al. (2004)
	mTcCIR215	F: GCTTCAACTCCAAATCAC	R: TAGCATCCCGTATTGTG	–	Pugh et al. (2004)
	mTcCIR251	F: TCTATGGGATTTGATGAG	R: AGATACAGCAGGAACACA	AJ566572	Pugh et al. (2004)

[a] Validated in this paper

Statistical analysis

The cacao genotypes used in this study were treated as a single population. The amplified SSR DNA bands representing different alleles were scored at different genotypes. A principal component analysis, conducted on the allele frequency data, the average observed allele number, observed heterozygosity (H_O), expected heterozygosity (H_E), and total heterozygosity (H_T) were determined with the GENETIX software (Ver. 4.05.2; Belkhir et al. 1999), Hardy–Weinberg equilibrium (HWE), and linkage disequilibrium were assessed through exact tests, by using the program FSTAT version 2.9.3.2 (Goudet 2002). Polymorphic information content (PIC) was obtained for each locus according to Anderson et al. (1993). Comparison of the allele number averages and PIC averages from EST-SSR and genomic SSR were obtained by Student's *t* test using the SPSS program v.10.

Genetic mapping

The genetic mapping of the SSR-EST markers was carried on a genetic linkage map of 136 individuals in the F_2 Sca 6×ICS 1 population (Faleiro et al. 2006) using the JoinMap. 4 software (Van Ooijen 2006).

Results

Characteristics of EST-SSR derived from cacao–*M. perniciosa* interaction

A total of 3,487 ESTs of cacao–*M. perniciosa* interaction was used in this investigation to screen for SSR markers. We detected 430 sequences, 277 from TSH 1188 and 153 from Catongo containing one or several SSRs (Electronic supplementary material). Sixty-five ESTs contained more than one SSR, and a total of 505 EST-SSRs were identified (Table 3). Among the 505 EST-SSRs identified, three types of motifs were found: dinucleotides (72.08%), trinucleotides (27.33%), and tetranucleotides (0.59%) (Fig. 1a). The motifs were found with a minimum of four and a maximum of 15 repeats (Fig. 1 and Table 3). The most frequent numbers of repeats were four (67.92%), five (13.07%), and six (7.52%) (Fig. 1b and Table 3). The other numbers of repeats (seven to 15) were found in low frequency from 0.20% to 3.56% (Fig. 1b). Among the EST-SSRs identified, 58.21% correspond to AG (18.61%), CT (16.83%), TC (12.87%), and AT (9.9%) repeats (Table 3). The other 47 repeat motifs were found with a total frequency of 41.79% (Table 3).

Putative function of ESTs containing SSR

To explore the potential utility of the EST-SSR markers in research of cacao structural genomes, the 505 ESTs containing SSR were compared to the Genbank database and classified into 16 main categories (Fig. 2). Most of the EST-SSRs were placed in the "Unknown function" and "No homology" categories (45.82%). Also, 15.12% of the EST-SSRs were found in "Protein synthesis and processing," 6.74% in "Primary metabolism," 5.81% in "Abiotic stimuli and development," 5.58% in "Gene expression and RNA metabolism," and 4.88% in "Signal transduction and post-translational regulation." The remaining 16.05% EST-SSRs were dispatched in nine other categories (Fig. 2). Genes with important biological function potentially related to cacao resistance/susceptibility to WBD, and containing SSRs, were identified such as bZip transcription factor (Verica et al. 2004), caffeine synthase (Jones et al. 2002), ABC transporter (Smart and Fleming 1996), pathogenesis-related protein (Coram and Pang 2005), thaumatin (Cheong et al. 2002), and AVR9/Cf-9 rapidly elicited protein (Romeis et al. 1999; Electronic supplementary material).

Location of the SSR in relation to the coding sequence of the ESTs

The position of each EST-SSR was obtained in relation to the ORF of the corresponding cDNA (Fig. 3). Analyzing all the EST-SSRs, we observed that their distribution pattern in cDNAs was quite similar between the three libraries (Fig. 3a). A high frequency of SSRs was found in the 5'UTR and in the ORF (about 27% each) and a low frequency was observed in the 3'UTR (about 8%). About 35% of the SSRs were not precisely located in relation to the ORF. Analyzing the location of the 12 polymorphic EST-SSRs (see "Results" section below and Table 5), the highest EST-SSR frequency was observed in the ORF (36.4%), in the 5'UTR (18.2%) and in the 3'UTR (9.1%) (Fig. 3b). For 36.4% of the EST-SSRs, no successful location in the cDNA was obtained (Fig. 3b).

Primer validation and detection of polymorphism

A total of 49 designed primer pairs was used for the genetic validation of the EST-SSR markers, being seven sequences of the susceptible genotype (Catongo) and 42 sequences of the resistant genotype (TSH 1188; Table 4). Thirty-two EST-SSRs were obtained from pod sequences and 17 EST-SSRs from the meristem sequences. Only EST-SSRs with four to 14 repeats were considered (Table 4).

These EST-SSRs developed were compared with genomic SSRs to verify the effectiveness of those two types of markers in 21 cacao accessions. The analysis using 12

Table 3 Frequency of different SSR types identified in 430 expressed sequence tag of the interaction cacao–*M. perniciosa* ESTs

Repeat motif	Number of repeat units												Total repeats	Frequency (%)
	4	5	6	7	8	9	10	11	12	13	14	15		
AG	74	4	7	–	–	1	–	4	2	–	2	–	94	18.61
CT	64	7	4	–	2	4	2	–	1	–	–	1	85	16.83
TC	33	14	5	5	5	–	1	1	–	1	–	–	65	12.87
AT	32	10	3	2	1	2	–	–	–	–	–	–	50	9.90
TA	19	4	1	–	1	–	–	1	–	–	–	–	26	5.15
AGA	5	4	3	5	–	–	–	–	–	–	–	–	17	3.37
GA	5	4	2	1	1	1	1	–	–	2	–	–	17	3.37
GAA	12	1	–	–	–	–	–	–	–	–	–	–	13	2.57
GAT	9	–	1	–	–	–	–	–	–	–	–	–	10	1.98
TG	9	–	–	–	1	–	–	–	–	–	–	–	10	1.98
TCT	7	1	1	–	–	–	–	–	–	–	–	–	9	1.78
CTG	6	–	1	1	–	–	–	–	–	–	–	–	8	1.58
AC	7	–	–	–	–	–	–	–	–	–	–	–	7	1.39
ACC	2	–	4	–	–	–	–	–	–	–	–	–	6	1.19
AGC	5	1	–	–	–	–	–	–	–	–	–	–	6	1.19
ATA	1	1	1	2	–	–	–	–	–	–	–	–	5	0.99
CTT	2	1	–	1	1	–	–	–	–	–	–	–	5	0.99
GT	5	–	–	–	–	–	–	–	–	–	–	–	5	0.99
ATC	2	–	2	–	–	–	–	–	–	–	–	–	4	0.79
CTC	2	2	–	–	–	–	–	–	–	–	–	–	4	0.79
GAG	1	3	–	–	–	–	–	–	–	–	–	–	4	0.79
TGC	3	–	–	1	–	–	–	–	–	–	–	–	4	0.79
AGG	1	2	–	–	–	–	–	–	–	–	–	–	3	0.59
CA	2	1	–	–	–	–	–	–	–	–	–	–	3	0.59
CAA	3	–	–	–	–	–	–	–	–	–	–	–	3	0.59
GCT	3	–	–	–	–	–	–	–	–	–	–	–	3	0.59
TAA	3	–	–	–	–	–	–	–	–	–	–	–	3	0.59
TCC	1	2	–	–	–	–	–	–	–	–	–	–	3	0.59
ACA	2	–	–	–	–	–	–	–	–	–	–	–	2	0.40
ACG	1	1	–	–	–	–	–	–	–	–	–	–	2	0.40
CAC	2	–	–	–	–	–	–	–	–	–	–	–	2	0.40
CG	2	–	–	–	–	–	–	–	–	–	–	–	2	0.40
GTT	2	–		–	–	–	–	–	–	–	–	–	2	0.40
TAC	1	–	–	–	–	–	1	–	–	–	–	–	2	0.40
TAG	2	–	–	–	–	–	–	–	–	–	–	–	2	0.40
TCA	2	–	–	–	–	–	–	–	–	–	–	–	2	0.40
TGA	2	–	–	–	–	–	–	–	–	–	–	–	2	0.40
TGG	1	1	–	–	–	–	–	–	–	–	–	–	2	0.40
AGAA	1	–	–	–	–	–	–	–	–	–	–	–	1	0.20
ATGT	1	–	–	–	–	–	–	–	–	–	–	–	1	0.20
ATT	–	1	–	–	–	–	–	–	–	–	–	–	1	0.20
CAG	1	–	–	–	–	–	–	–	–	–	–	–	1	0.20
CAT	–	1	–	–	–	–	–	–	–	–	–	–	1	0.20
CGA	–	–	1	–	–	–	–	–	–	–	–	–	1	0.20
CGC	1	–	–	–	–	–	–	–	–	–	–	–	1	0.20
CGG	1	–	–	–	–	–	–	–	–	–	–	–	1	0.20
GAAA	1	–	–	–	–	–	–	–	–	–	–	–	1	0.20
GAC	1	–	–	–	–	–	–	–	–	–	–	–	1	0.20
GTG	–	–	1	–	–	–	–	–	–	–	–	–	1	0.20
TAT	1	–	–	–	–	–	–	–	–	–	–	–	1	0.20
TGT	–	–	1	–	–	–	–	–	–	–	–	–	1	0.20
Total	343	66	38	18	12	8	5	6	3	3	2	1	505	100

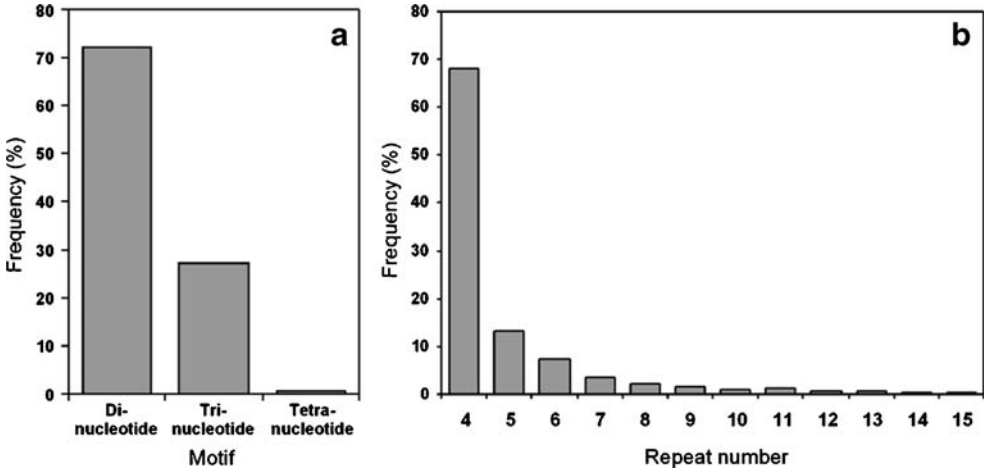


Fig. 1 Frequency of SSR with different motif and repeat number. **a**. Frequency of SSRs with di-, tri-, and tetranucleotide motif. **b** Frequency of SSRs with four to 20 repeat number

EST-SSR loci revealed 100% polymorphism and a total of 47 alleles. The number of alleles per locus ranged from two to six, with an average of 3.92 alleles per locus (Table 5). The H_O varied among loci from 0 to 0.67, with an average of 0.30, and H_E varied among loci from 0.05 to 0.70, with an average of 0.49. All loci showed a significant deviation from HWE and were independent after Bonferroni correction for multiple tests. The 12 polymorphic EST-SSR markers revealed allelic diversity with PIC values ranging from 0.05 to 0.71 (mean 0.49; Table 5).

In comparison with EST-SSRs, the 11 genomic SSR markers revealed 82% polymorphism (only nine were polymorphic), revealing a total of 47 alleles. The number of alleles per locus ranged from three to eight, with an average of 5.22 alleles per locus (Table 5). The H_O varied between loci from 0.15 to 0.78, with an average of 0.44. The H_E varied between loci from 0.2 to 0.8, with an average of 0.61. All loci showed a significant deviation from HWE and were independent after Boferroni correction for multiple tests. The nine polymorphic genomic SSR markers showed allelic diversity with PIC values ranging from 0.2 to 0.8 (average 0.62; Table 5). Although, in terms of allelic diversity and based on a *t* test, there was no significant difference either in the allele number or PIC between the tested SSR types (ESTs vs. genomic; Table 5). The association of EST-SSR with the genomic SSR enhanced the analysis of genetic distance among the genotypes thereby showing a tendency for clusterization according to the origin of the genotypes (Fig. 4).

Genetic mapping of the EST-SSRs

Among the 12 polymorphic EST-SSR markers mentioned in the Table 5, only two could be mapped on the 136 individuals of the F_2 Sca 6×ICS 1: the SSR msEstTsh-6 and msEstTsh-12 were positioned in the linkage group 2 (LG 2; Fig. 5).

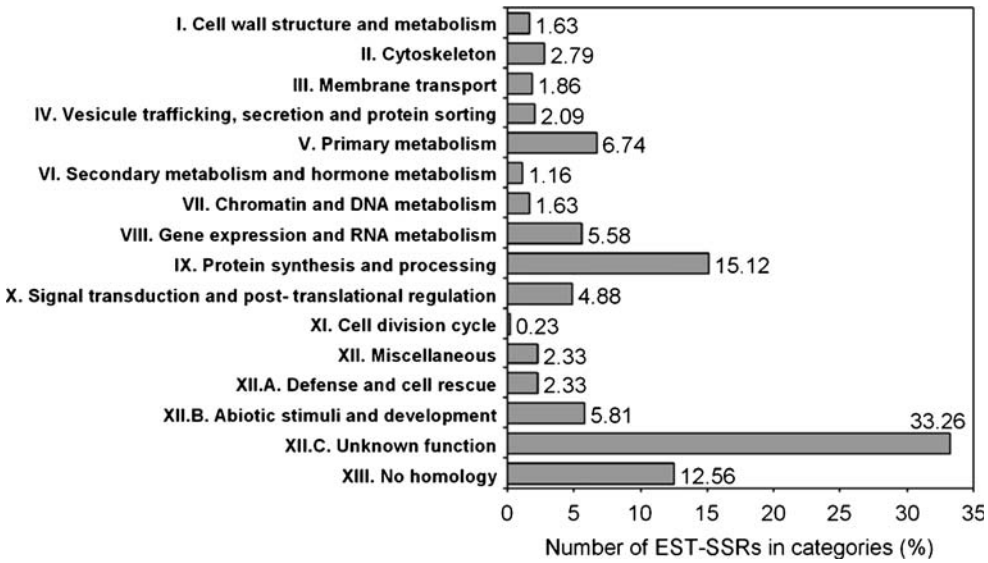


Fig. 2 Distribution of the 592 ESTs containing SSRs into functional classes. The 16 broad categories that were used for classification during the semi-automatic annotation are indicated, as well as the number of corresponding sequences (in percent). Only one class was assigned to each sequence

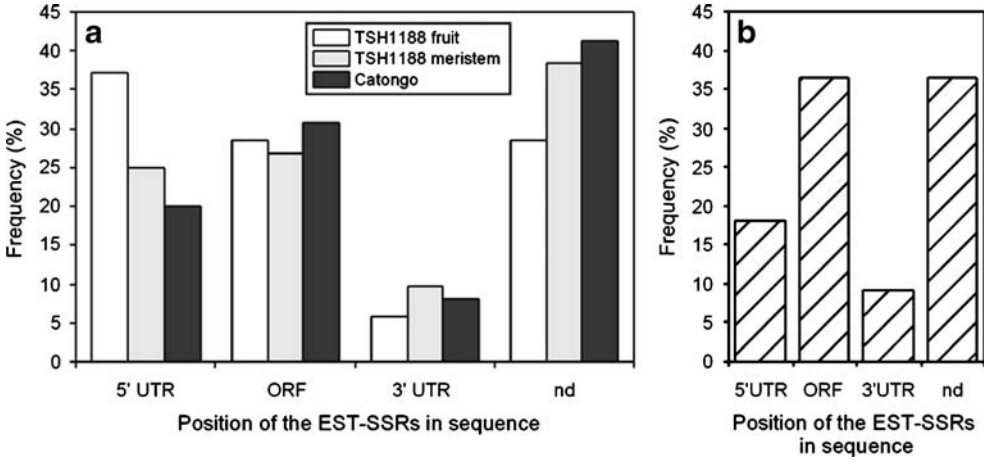


Fig. 3 Position of the EST-SSRs in sequence. **a** Frequency of the overall EST-SSRs in the different sequence parts. **b** Frequency of the 11 polymorphic EST-SSRs (see Table 5) in the different sequence parts. The 11 polymorphic EST-SSRs belong to the TSH 1188 libraries (meristem or fruit). *ORF* open reading frame, *UTR* untranslated region

Discussion

As demonstrated in various plant species, EST sequences could be a very good source of SSR markers (Varshney et al. 2005), therefore could be related to agronomically important traits. In order to improve the number of microsatellite markers on cacao and to map non-anonymous loci, we investigated SSRs present in cacao–*M. perniciosa* interaction ESTs. SSRs were detected in 12.33% of the ESTs analyzed. Lower frequency of SSRs in ESTs have been reported in other crops (4.7% in rice and 1.5% in maize, Kantety et al. 2002; 7.5% in wheat, Nicot et al. 2004; 7.15% in cotton, Han et al. 2006), whereas other published SSR analysis showed higher frequency, such as 18.5% (Aggarwal et al. 2007). The frequency of SSRs found in the present study for cacao–*M. perniciosa* ESTs appeared as highly abundant. This difference of our results and most published results may be attributed to the SSR search criterion that in our study was defined minimum size of the SSR as 8 bp (e.g., AT_4) while in other studies usually minimum sizes of 10 to 12 bp were adopted (Sanwen et al. 2000; Nicot et al. 2004).

In this study the highest proportion of EST-SSRs identified was dinucleotides (75.6%) followed by trinucleotides (22.6%), in contrast with the majority of studies which report trinucleotides as the most abundant class of SSRs in ESTs in plants (Kantety et al. 2002; Nicot et al. 2004). Our results, however, are in agreement with recent studies in coffee (Aggarwal et al. 2007), *Actinidia* (Fraser et al. 2004) and *Picea* species (Rungis et al. 2004), in which dinucleotides were found to be the most abundant class of EST-SSRs. As supported by other authors (Varshney et al. 2005; Aggarwal et al. 2007), this difference is probably due to the search criteria used for EST mining, which affect the relative estimates of frequency of EST-SSRs. In fact, in our study, as well as in the Aggarwal et al. (2007) the minimum number of repeat units considered for all types of SSR (di-, tri-, and tetranucleotides) was four, whereas in others the minimum number of repeat units for SSR identification was higher (six to dinucleotides and five to tri- and tetranucleotides). When we apply the same criteria when mining our ESTs, we obtain almost the same abundance for the di- and trinucleotides (58.5% and 41.5%, respectively), reducing drastically the frequency difference between these two SSR categories (data not shown).

SSRs were found in all functional categories of ESTs; some categories containing more SSRs than others (Fig. 2). The percent of SSRs present in each category is directly related to the class size as described by Gesteira et al. (2007) and Zaidan et al. (2005). In Gesteira et al. (2007) the most prevailed categories were “No homology” (23.9%), “Unknown function” (22.9%), “Protein synthesis and processing” (12.1%), and “Primary metabolism” (11.1%), which is in accordance with our results. The other categories also followed the same corresponding pattern between EST and SSR number. Because the studied EST libraries contained 1,357,131 bp (data not shown), the number of SSR/bp detected in this study corresponded to 3.7×10^{-4}. The EST-SSRs were found in the coding sequence as well as in the UTR regions (Fig. 3a). Interestingly, a similar number of SSRs was observed in the ORF and in the 5’UTR, while the 3’UTR presented a smaller number of detected markers. The difference of SSR number between the two UTR regions may be related to a longer 5’UTR sequence due to the directional sequencing made from the 5’end for all the ESTs to avoid the polyA tail sequencing (Gesteira et al. 2007). In these conditions, the 5’UTR was systematically sequenced, while the 3’UTR either partially or not sequenced depending of the ORF length and sequencing capacities. The high frequency of EST-SSRs (including the polymorphic ones; Fig. 3b) in ORFs was unexpected and contradicted some works showing that the UTRs have a higher frequency of SSRs than the rest of the genome (Morgante et al. 2002; Feingold

Table 4 ESTs and corresponding SSRs chosen for primer design. CAT: Catongo; P: pod; M: meristem; TSH: TSH 1188

Genotype	Organ	Functional annotation of the EST-SSR	EST size (bp)	Repeat motif (number of repeat)
CAT	M	ADP-ribosylation factor	368	TC(10)
CAT	M	High-glycine tyrosine keratin-like protein	399	CT(9)
CAT	M	Late embryogenesis abundant protein Lea5-D	698	AT(9)
CAT	M	Latex profilin	516	AGA(7)
CAT	M	Ubiquitin conjugating protein	439	GA(13)
CAT	M	Unknown protein	267	GA(10)
CAT	M	Unknown protein	509	AG(11)
TSH	M	ABC transporter	420	CTT(4)
TSH	M	Casein kinase	330	TAC(10)
TSH	M	Hypothetical protein	802	CT(10)
TSH	M	Hypothetical protein	709	CGA(6)
TSH	M	Rat TTF-1 mRNA for thyroid nuclear factor 1	453	CTT(7)-CTG(4)
TSH	M	Unknown function	457	TC(13)
TSH	M	Unknown function	421	AGA(7)
TSH	M	Unknown protein	405	TGC(7)
TSH	M	Unknown protein	497	ACC(6)
TSH	M	Zinc finger (C3HC4-type RING finger) family protein	215	GAT(6)
TSH	P	26 S proteasome regulatory subunit S5A	438	CT(4)
TSH	P	ATP-dependent clp protease ATP	489	AG(4)
TSH	P	Bzip transcription factor ATB2	803	CT(4)N(7)CT(4)
TSH	P	Caffeine synthase	415	GA(4)-AT(7)[a]
TSH	P	Calmodulin-binding family protein	507	CTG(4)
TSH	P	Copper chaperone (CCH)-related	494	GAAA(4)
TSH	P	DNA polymerase-related	388	AT(9)
TSH	P	Double-stranded DNA-binding family protein	449	CT(4)-TGA(4)[a]
TSH	P	Drought responsive family protein	428	CT(4) N(5) TC(4)
TSH	P	Expressed protein	410	TA(5)-AC(4)
TSH	P	Expressed protein	373	CT(9)
TSH	P	Expressed protein	494	CTG(4) N(5)CTG(7)
TSH	P	Glutaredoxin family protein	427	TCT(4)
TSH	P	GTP-binding nuclear protein RAN1	521	CT(4)
TSH	P	Methionyl aminopeptidase	449	AT(5)
TSH	P	Peroxiredoxin	575	TC(4)
TSH	P	Polyphenol oxidase	696	CT(4)
TSH	P	Profilin 1	560	AGA(7)
TSH	P	Proline rich protein	790	TGG(5)
TSH	P	Protein F2D10.18	786	CT(9)
TSH	P	Rac GDP dissociation Inhibitor	323	TC(4)
TSH	P	Ring box protein related	352	AG(4)
TSH	P	Transformer serine/arginine-rich ribonucleoprotein	378	AG(4)
TSH	P	Translationally controlled tumor protein homolog (TCTP)	467	CT(4)
TSH	P	Translationally controlled tumor protein homolog (TCTP)	430	AT(4)
TSH	P	Universal stress protein (USP) family protein	544	AG(14)[a]
TSH	P	Zinc finger (AN1-like) family protein	325	TC(7)
TSH	P	Zinc Finger (C3HC4 type ring finger)	459	TC(4)
TSH	P	Zinc finger (C3HC4)	459	CAT(5)

[a] Two primer pairs were designed

Table 5 Primer characteristics for twelve and nine microsatellite loci from cacao–*M. perniciosa* interaction ESTs and cacao genomic DNA, respectively

SSR type	Marker analyzed	Putative gene function	Position	Ta (°C)	Repeat type	Marker validation					
						Allele number	Size range (bp)	H_O	H_E	H_T	PIC
EST-SSR	msEstTsh-1	DNA polimerase related	nd	56.8	$(AT)_9$	3	192–200	0.00	0.50	0.50	0.48
	msEstTsh-2	Expressed protein	5' UTR	62	$(CT)_9$	4	90–96	0.33	0.65	0.51	0.65
	msEstTsh-3	Protein F2D10.18	ORF	56.8	$(CT)_9$	6	102–112	0.31	0.49	0.17	0.50
	msEstTsh-4	Expressed protein	ORF	59.5	$(TA)_5(AC)_4$	3	173–179	0.00	0.16	0.68	0.16
	msEstTsh-5	Unknown protein	ORF	TD60-48	$(ACC)_6$	4	290–302	0.14	0.26	0.46	0.26
	msEstTsh-6	Unknown protein	nd	TD60-48	$(AGA)_7$	4	212–229	0.58	0.70	0.17	0.71
	msEstTsh-7	Thyroid nuclear factor 1	nd	TD60-48	$(CTT)_7$ $(CTG)_4$	4	149–161	0.25	0.62	0.60	0.62
	msEstTsh-8	Unknown protein	ORF	TD60-48	$(TGC)_7$	2	190–193	0.05	0.05	0.00	0.05
	msEstTsh-9	Unknown protein	nd	TD60-48	$(TC)_{13}$	3	206–210	0.67	0.57	0.17	0.57
	msEstTsh-10	Hypothetical protein	5' UTR	TD60-48	$(CT)_{10}$	5	156–174	0.47	0.70	0.32	0.71
	msEstTsh-11	Casein kinase	3' UTR	TD60-48	$(TAC)_{10}$	4	209–218	0.41	0.65	0.37	0.65
	msEstTsh-12	Unknown protein	ORF	TD60-48	$(GAA)_8$	5	207–222	0.41	0.51	0.38	0.52
	Mean of polymorphic locus	–	–	–	–	3.92(47)[b]	–	0.30	0.49	0.36	0.49
	SD of polymorphic locus	–	–	–	–	1.08	–	–	–	–	0.23
Genomic SSR	mTcCIR12	–	–	46	$(CATA)_4N18(TG)_6$	7	188–212	0.78	0.77	0.01	0.77
	mTcCIR26	–	–	46	$(TC)_9C(CT)_4TT(CT)_{11}$	5	294–302	0.55	0.60	0.08	0.60
	mTcCIR30	–	–	46	$(CA)_{11}$	6	178–192	0.42	0.73	0.42	0.73
	mTcCIR37	–	–	46	$(GT)_{15}$	6	144–154	0.57	0.74	0.23	0.74
	mTcCIR58	–	–	51	$(GT)_{40}$	8	254–274	0.53	0.80	0.34	0.80
	mTcCIR157	–	–	49.6	$(AG)_9$	3	151–155	0.42	0.47	0.11	0.47
	mTcCIR166	–	–	48.2	$(CT)_9(CA)_8$	3	215–219	0.22	0.20	0.10	0.20
	mTcCIR215	–	–	49.1	$(AG)_{13}$	3	197–201	0.15	0.52	0.71	0.52
	mTcCIR251	–	–	46.8	$(CT)_7(CA)_{12}$	6	186–196	0.31	0.70	0.55	0.71
	Mean of polymorphic locus	–	–	–	–	5.22 (47)[b]	–	0.44	0.61	0.28	0.62
	SD of polymorphic locus	–	–	–	–	1.86	–	–	–	–	0.19
Student's *t* test[a]	–	–	–	–	–	−2.00	–	–	–	–	−1.36

The putative gene function, size range in bp for each locus is given. H_O, H_E, and H_T represent the number of observed, expected, and total heterozygosities, respectively, per locus and genotypes

nd non-determined, *ORF* open reading frame, *PIC* polymorphic information content, *SD* standard deviation, *UTR* untranslated region

[a] The Student's *t* test was used to compare the EST-SSR and genomic SSR data (allele number and PIC). ta=2.101 for dl=18 and α=0.05

[b] Total number of alleles

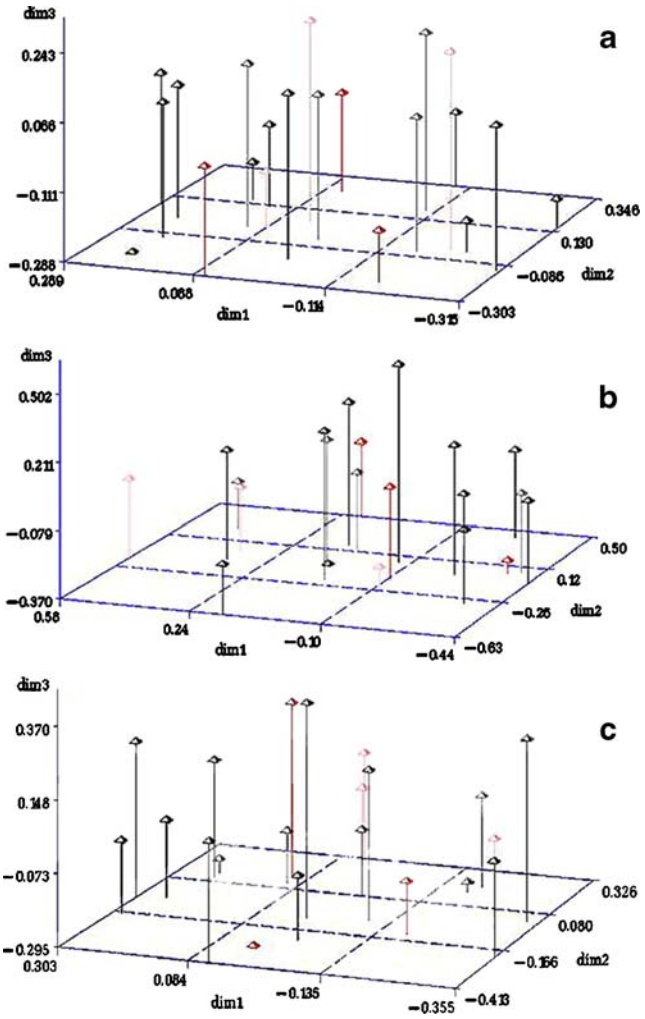
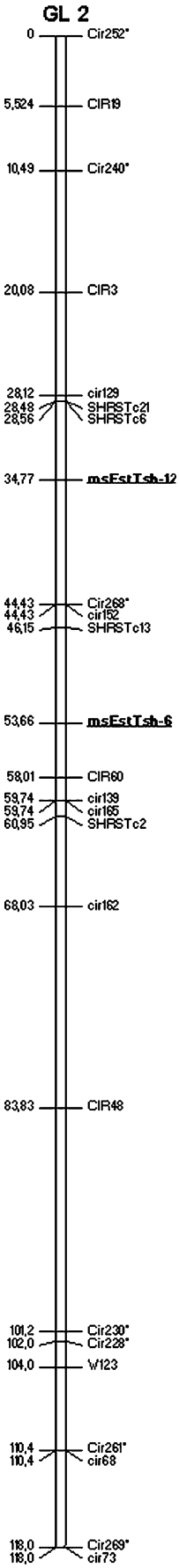


Fig. 4 Analysis of genetic distance among 21 cacao genotypes revealed by EST-SSRs the genomic SSRs. **a** Genetic distance revealed for genomic SSRs. **b** Genetic distance revealed for EST-SSRs. **c** Genetic distance revealed for combination de EST-SSRs and genomic SSRs. Individuals from the high Amazon and from the low Amazon are indicated in *black* and *gray*, respectively. Trinitarios are represented in *red*. Hybrids are represented in *pink*

et al. 2005), and that the SSRs derived from UTRs have the potential of higher polymorphism than those derived from coding regions (Feingold et al. 2005). It has been also described that for polymorphism detection, EST-SSRs derived from 3'UTR are superior to those derived from 5'UTR (Varshney et al. 2005 for review), which was not confirmed by our analysis (Fig. 3b). Because increasing the number of trinucleotide motifs did not lead to a change of the reading frame (and may avoid the appearance of a stop codon), it has been presumed that the SSR motif type present in ORF may be preferentially trinucleotide (Nicot et al. 2004; Morgante et al. 2002). However, in our study, trinucleotides (44.7%) were present in ORFs in a very similar proportion as dinucleotides (55.3%; data not shown). On the other hand, the dinucleotides may vary

Fig. 5 Genetic linkage map of TSH516 linkage group 2. Markers are indicated to the *right* and distances between markers to the *left* of the linkage group. The markers in each linkage group have been ordered and arranged to determine their most probable position along the linkage group. CEPEC EST-SSR markers are designated *msEstTsh followed by a number*, CIRAD SSR markers are *designated CIR followed by a number*, and USDAARS, SHRS SSR markers are designated *SHRSTc followed by a number*



following a step of three motifs (or a multiple of three) which do not modify the reading frame (three motifs of two nucleotides=6 bp=2 codons) as stated by Morgante et al. (2002). For these reasons definitive conclusions cannot be obtained regarding a direct correlation between motif type and polymorphic EST-SSR presence in ORF. The presence of SSR in the transcripts of genes suggests that they might have a role in gene function. It has also been shown that variation in repeat units of SSRs: (a) in 5'UTR affects gene transcription and/or translation; (b) in coding region inactivate or activate genes or truncated protein; and (c) in 3'UTR might be responsible for gene silencing or transcript slippage (Varshney et al. 2005).

EST-SSR primers have been reported to be less polymorphic than genomic SSR in crop plants because of high conservation in transcribed regions (Varshney et al. 2005, for review). In this respect, because no statistic difference between EST-SSR and genomic SSR polymorphism was observed in our study, our results are not in agreement with published data from other plant and animal species (Cho et al. 2000; Liewlaksaneeyanawin et al. 2004; Eujayl et al. 2001), but are in accordance with the work of Cupertino (2007) showing that *Eucalyptus* spp. EST-SSR markers are as polymorphic as genomic ones.

In conclusion, SSRs are currently an important option available to plant breeders for foreground selection in marker-assisted backcrossing programs because their ease of use, codominant inheritance, high levels of polymorphism, and reasonably even distribution across nuclear genome. Therefore, development of additional markers is a valuable objective for the cacao community, especially if these can be a target at expressed regions known to contribute to the control of economically important, genetically complex traits such as disease resistance.

Acknowledgments The present study was carried out in the Molecular Plant Pathology Laboratory (FITOMOL) of the Cocoa Research Center, Bahia, Brazil. This paper is part of an effort for searching novel genes to WBD resistance, coordinated by José Luis Pires and Karina Peres Gramacho at CEPEC/CEPLAC, and is funded by the Brazilian agency of Bahia–FAPESB (FAPESB; research grant no. APR160/2007) and for providing a research fellowship to N. Carels. This research was also supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; research grant no. 471274/2006-2 and the Ministère des Affaires Etrangères Français (MAE). LSL (Doctor Student) was granted by the Brazilian agency of Bahia–FAPESB.

References

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellites for genome analyses of coffee and related species. Theor Appl Genet 114:359–372

Aime MC, Phillips-Mora W (2005) The causal agents of witches' broom and frosty pod rot of cacao (chocolate, *Theobroma cacao*) form a new lineage of Marasmiaceae. Mycologia 97:1012–1022

Albuquerque PSB (2006) Mapas de ligação e Identificação de locos controladores de características quantitativas (QTLs) associados à resistência a *Crinipellis perniciosa* em acessos de cacaueiro (*Theobroma cacao*) originários da Amazônia Brasileira. Tese (Doutorado em Agronomia)–Escola Superior de Agricultura Luis de Queiroz. Universidade de São Paulo, Piracicaba, p 133

Anderson JA, Churchill GA, Sutrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186

Argout X, Fouet O, Wincker P, Gramacho K, Legavre T, Sabau X, Risterucci AM, Da Silva C, Cascardo J, Allegre M, Kuhn D, Verica JA, Courtois B, Loor RG, Regis B, Sounigo O, Ducamp M, Guiltinan MJ, Ruiz M, Alemanno L, Machado R, Phillips W, Schnell R, Gilmour M, Rosenquist E, Butler D, Maximova S, Lanaud C (2008) Towards the understanding of the cocoa transcriptome: production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* generated from various tissues and under various conditions. BMC Genomics 30:512

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1999) GENETIX 4.04, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

Borrone JW, Meerow AW, Kuhn DN, Whitlock BA, Schnell RJ (2007) The potential of the WRKY gene family for phylogenetic reconstruction: an example from the Malvaceae. Mol Phylogenet Evol 44:1141–1154

Borrone JW, Kuhn DN, Schnell RJ (2004) Isolation, characterization, and development of WRKY genes as useful genetic markers in *Theobroma cacao*. Theor Appl Genet 109:495–507

Cheong YH, Chang H, Gupta R, Wang X, Zhu T, Luan S (2002) Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis1. Plant Physiol 129:661–677

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic licraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Clement D, Risterucci AM, Motamayor JC, N'Goran JAK, Lanaud C (2003) Mapping quantitative trait loci for bean traits and ovule number in *Theobroma cacao* L. Genome 46:103–111

Coram TE, Pang ECK (2005) Isolation and analysis of candidate ascochyta blight defence genes in chickpea. Part II. Microarray expression analysis of putative defence-related ESTs. Physiol Mol Plant Pathol 66:201–210

Corpet F, Servant F, Gouzy J, Kahn D (2000) ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. Nucleic Acids Res 28:267–269

Creste S, Tulmann Neto A, Figueira F (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. Plant Mol Biol Report 19:299–306

Cupertino FB (2007) Marcadores microssatélites como ferramentas úteis para estimação de diversidade genética e certificação de paternidade em híbridos de *Eucalyptus spp*. Master thesis, Universidade Estadual de Santa Cruz, UESC, Ilhéus-BA, Brazil

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Eujayl I, Sorrells M, Baum M, Wolters P, Powell W (2001) Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. Euphytica 119:39–43

Ewing B, Hillier L, Wendl M, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–185

Faleiro FG, Queiroz VT, Lopes UV, Guimarães CT, Pires JL, Yamada MM, Araújo IS, Pereira MG, Schnell R, Souza Filho GA, Ferreira CF, Barros EG, Moreira MA (2006) Mapping QTLs for Witches' Broom (*Crinipellis perniciosa*) Resistance in cacao (*Theobroma cacao* L.). Euphytica 149:227–235

Feingold S, Lloyd J, Norero N, Bonierbale M, Lorenzen J (2005) Mapping and characterization of new EST-derived microsatellites for potato (*Solanum tuberosum* L.). Theor Appl Genet 111:456–466

Flament MH, Kebe I, Clement D, Pieretti I, Risterucci AM, N'Goran JAK, Cilas C, Despréaux D, Lanaud C (2001) Genetic mapping of resistance factors to *Phytophtora palmivora* in cocoa. Genome 44:79–85

Fraser LG, Harvey CF, Crowhurst RN, De Silva HN (2004) EST-derived microsatellites from Actinidia species and their potential for mapping. Theor Appl Genet 108:1010–1016

Gesteira AS, Micheli F, Carels N, Da Silva AC, Gramacho KP, Schuster I, Macedo JN, Pereira GAG, Cascardo JCM (2007) Comparative analysis of expressed genes from cacao meristems infected by *Moniliophthora perniciosa*. Annals of Botany doi:10.1093/aob/mcm092

Goudet J (2002) **FSTAT**: a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). University of Lausanne, Department of Ecology & Evolution, Lausanne

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Han Z, Wang C, Song X, Guo W, Gou J, Li C, Chen X, Zhang T (2006) Characteristics, development and mapping of Gossypium hirsutum derived EST-SSRs in allotetraploid cotton. Theor Appl Genet 112:30–439

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R (2004) Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 32:258–261

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Jones PG, Allaway D, Gilmour DM, Harris C, Rankin D, Retzel ER, Jones CA (2002) Gene discovery and microarray analysis of cacao (*Theobroma cacao* L.) varieties. Planta 216:255–264

Journet E-P, van Tuinen D, Gouzy J, Crespeau H, Carreau V, Farmer M-J, Niebel A, Schiex T, Jaillon O, Chatagnier O, Godiard L, Micheli F, Kahn D, Gianinazzi-Pearson V, Gamas P (2002) Exploring root symbiotic programs in the model legume *Medicago truncatula* using EST analysis. Nucleic Acids Res 30:5579–5592

Kantety RV, Rota ML, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet A, Lagoda PJL (1999) Isolation and characterization of microsatellites in *Theobroma cacao* L. Mol Ecol 8:2141–2152

Leal GA, Albuquerque PSB, Figueira A (2007) Genes differentially expressed in *Theobroma cacao* associated with resistance to witches' broom disease caused by *Crinipellis perniciosa*. Mol Plant Pathol 8:279–292

Liewlaksaneeyanawin C, Ritland CR, El-Kassaby YA, Ritland K (2004) Single-copy, species transferable microsatellites markers developes from loblolly pine. Theor Appl Genet 109:361–369

Lima SL, Gramacho KP, Gesteira AS, Lopes UV, Gaiotto FA, Zaidan HA, Pires JL, Cascardo JCM, Micheli F (2008) Characterization of microsatellites from cacao–*Moniliophthora perniciosa* interaction expressed sequence tags. Mol Breeding 112:399–406

Luz EDMN, Sgrillo RB, Santos Filho LP (2005) Estimativas de danos e perdas causadas por doenças no cacaueiro. In: Workshop de epidemiologia de doenças de Plantas, 1 Viçosa 2004. Proceedings. Editora UFV, Viçosa, pp 67–79

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with non repetitive DNA in plant genomes. Nat Genet 30:194–200

Motamayor JC, Risterucci AM, Heath M, Lanaud C (2003) Cacao domestication II: progenitor germplasm of the Trinitario cacao cultivar. Heredity 91:322–330

Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, Moreno A, Lanaud C (2002) Cacao domestication I: the origin of the cacao cultivated by the Mayas. Heredity 89:380–386

N'Goran JAK, Laurent V, Risterucci AM, Lanaud C (1994) Comparative genetic diversity of *Theobroma cacao* L. using RFLP and RAPD markers. Heredity 73:589–597

Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P (2004) Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). Theor Appl Genet 109:800–805

Paim VRLM, Luz EDMN, Pires JL, Silva, SDVM, Souza JTS, Bevilaqua A, Santos Filho LP (2006) Sources of resistance to Crinipellis perniciosa in progenies of cacao accessions collected in the Brazilian Amazon. Scientia. Agrícola 63:572-578

Pires JL (2003) Avaliação quantitativa e molecular de germoplasma para o melhoramento do cacaueiro com ênfase na produtividade, qualidade de frutos e resistência a doenças. Tese (Doutorado em Genética e Melhoramento). Universidade Federal de Viçosa, p 328

Pound FJ (1938) Cocoa and witches' broom disease (*Maramius pernicious*) of South America with notes on other species of *Theobroma*. Cocoa Res 1:20–72

Pound FJ (1943) Cocoa and witches' broom disease: report on a recent visit to the Amazon territory of Peru. In: Toxopeus H (ed) The archive of cocoa research, vol 1. ACRI- IOCC, London, United Kingdom

Pugh T, Fouet O, Risterucci AM, Brottier P, Abouladze M, Deletrez C, Courtois B, Clement D, Larmande P, N'Goran JAK, Lanaud C (2004) A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. Theor Appl Genet 108:1151–1161

Rios-Ruiz RA (2001) Melhoramento para resistência a doenças. In: Dias L.A.S. (ed.), Melhoramento genético do cacaueiro. Funape-UFG, Editora Folha de Viçosa Ltda, p 289–324

Romeis T, Piedras P, Zhang S, Klessig DF, Hirt H, Jones JDG (1999) Rapid Avr9- and Cf-9–dependent activation of MAP kinases in tobacco cell cultures and leaves: convergence of resistance gene, elicitor, wound, and salicylate responses. Plant Cell 11:273–287

Rungis D, Berube Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K (2004) Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. Theor Appl Genet 109:1283–1294

Sanwen H, Baoxi Z, Milbourne D, Cardle L, Guimei Y, Jiazhen G (2000) Development of pepper SSR markers from sequence databases. Euphytica 117:163–167

SAS Institute INC (1988) SAS/STAT User's Guide. Release 6.03. SAS Institute Inc, Cary, NC, p 1028

Smart CC, Fleming AJ (1996) Hormonal and environmental regulation of a plant PDR5-like ABC transporter. Am Soc Biochem Mol Biol 271:19351–19357

Trevizan SDP, Da Silva MF Jr (1995) Impactos sociais, econômicos e ambientais relacionados a vassoura-de-bruxa. Congresso Brasileiro de Economia e Sociologia Rural, vol 33. Anais, Curitiba, PR, p 1409

Van Ooijen JW (2006) JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotech 23:48–55

Verica JA, Maximova SN, Strem MD, Carlson JE, Bailey BA, Guiltinan MJ (2004) Isolation of ESTs from cacao (*Theobroma cacao* L.) leaves treated with inducers of the defense response. Plant Cell Report 23:404–413

Zaidan HA, Gesteira A, Micheli FFL, Ceita GO, Cascardo J, Carels N, Braz NG, Serra WO, Gramacho KP, Lopes UV (2005) Caracterizaçao molecular da resistência do cacaueiro (*Theobroma cacao* L.) a vassoura de bruxa. In: XXXVIII Congresso Brasileiro de Fitopatologia, Brasília, Brasil