Agrotrópica 29(2): 111 - 118. 2017.
Centro de Pesquisas do Cacau, Ilhéus, Bahia, Brasil

DEVELOPMENT OF SNP MARKERS PRESENT IN EXPRESSED GENES OF THE PLANT-PATHOGEN INTERACTION: *Theobroma cacao - Moniliophtora perniciosa*

Lívia Santos Lima Lemos[1,2], Karina Peres Gramacho[1], José Luis Pires[1], Rogério Mercês Ferreira Santos[1,2], Rodrigo Sousa Ganem[1], Marcio Gilberto C. Costa[2], Fabienne Micheli[2]

[1]Cocoa Research Center, CEPLAC/CEPEC, Itabuna, Bahia, Brazil, e-mail: liviasllemos@gmail.com; gramachokp@hotmail.com, zezorga@gmail.com, rogeriomerces@gmail.com, rodrigoganem@live.com;[2]Universidade Estadual de Santa Cruz, Campus Soane Nazaré de Andrade, Ilhéus, Bahia, Brazil, marciogc.costa@gmail.com, fabienne.micheli@cirad.fr;[*]Current address: Universidade Federal do Sul da Bahia/ IHAC, Campus Paulo Freire, Praça Joana Angélica, 250, Bairro São José, 45988-058, Teixeira de Freitas-Bahia, Brazil.

We report the detection, validation and analysis of SNPs in the plant-pathogen interaction between cacao and *Moniliophthora perniciosa* ESTs using resequencing. This analysis in 73 EST sequences allowed the identification of 185 SNPs, 57% of them corresponding to transversion, 29% to transition and 14% to indels. The ESTs containing SNPs were classified into 14 main functional categories. After validation, 91 SNPs were confirmed, categorized and the parameters of nucleotide diversity and haplotype were calculated. Haplotype-based gene diversity and polymorphic information content (PIC) ranged from 0.559 to 0.56 and 0.115 to 0.12; respectively. Also, it was the advantage when considering haplotypes structure for each locus in place of single SNPs. Most of the gene fragments had a major haplotype combined to a series of low frequency haplotypes. Thus, the re-sequencing approach proved to be a valuable resource to identify useful SNPs for wide genetic applications. Furthermore, the cacao genome sequence availability allow a positional selection of DNA fragments to be re-sequenced enhancing the usefulness of the discovered SNPs. These results indicate the potential use of SNPs markers to identify allelic status of cacao resistance genes through marker-assisted selection to support the development of promising genotypes with high resistance to witch's broom disease.

Key words: Cacao, ESTs, plant-pathogen interaction, resistante genes, single nucleotide polymorphism.

Desenvolvimento de marcadores SNPs presentes em genes expressos da interação planta-patógeno: *Theobroma cacao - Moniliophtora perniciosa.* Foi reportado aqui a detecção, validação e análise de SNPs em ESTs da interação planta-patógeno entre cacau e *Moniliophthora perniciosa*, utilizando re-sequenciamento. Esta análise em 73 sequencias ESTs permitiu a identificação de 185 SNPs, 57% deles correspondendo a transversão, 29% a transição e 14% a inserções e deleções. As ESTs contendo SNPs foram classificadas em 14 principais categorias funcionais. Através da validação, 91 SNPs foram confirmados, categorizados e os parâmetros de diversidade de nucleotídeos e haplótipos foram calculados. A diversidade genética baseada em haplótipos e o conteúdo informativo polimórfico (PIC) variaram de 0,559 a 0,56 e 0,115 a 0,12, respectivamente. Além disso, foi apontado a vantagem de considerar estrutura de haplótipos para cada locus no lugar de um único SNPs. A maioria dos fragmentos de genes apresentou um haplótipo principal acompanhado por uma série de haplótipos de baixa frequência. Assim, a abordagem de re-sequenciamento provou ser eficiente para identificar SNPs úteis de ampla aplicação genética. Além disso, a disponibilidade da sequência genômica de Cacau permite uma seleção posicional de fragmentos de DNA a serem resequenciados, aumentando a utilidade dos SNPs descobertos. Estes resultados indicam a potencialidade do uso dos marcadores SNPs para identificação do estado alélico dos genes de resistência do cacau, através de seleção assistida por marcadores, para apoiar o desenvolvimento de genótipos promissores com alta resistência à vassoura-de-bruxa.

Palavras-chave: Cacau, ESTs, interação planta-patógeno, genes de resistência, polimorfismo de nucleotídeo unico.

Recebido para publicação em 10 de abril de 2017. Aceito em 28 de julho de 2017.
DOI: 10.21757/0103-3816.2017v29n2p111-118

Introduction

Moniliophthora (*ex Crinipellis*) *perniciosa* (Stahel) Singer (Aime and Phillips-Mora, 2005) is a hemibiotrophic basidiomycete responsible for witches' broom disease (WBD) on cacao trees (*Theobroma cacao* L.) (Frias et al. 1991). This disease brought severe economic losses in all cacao-growing regions to which it has spread, leading up to 90% of yield reductions. Even with the introduction of resistant varieties, Cacao production in South America reflects the severity of this pathogen, as the yields in most of the infected regions have not returned to pre-outbreak levels. Several functional genomics studies of the cacao-*Moniliophthora perniciosa* plant-pathogen interaction have been developed (Gesteira et al. 2007; Argout et al. 2008, CEPLAC projects granted by FAPESB and CNPq). These programs allowed the identification of ESTs involved in resistance of cacao to WBD. The data provides a background to polymorphisms detections (e.g. SNPs) needed for further studies, such as breeding strategies, gene pyramiding and marker-assisted selection (MAS).

Single nucleotide polymorphisms (SNPs) correspond to the most commonly found class of genetic variation in genomes (Cho et al. 1999; Picoult-Newberg et al. 1999; Rafalski, 2002; Liu et al. 2015; Zhou et al. 2016). Therefore, it quickly became a marker of choice in agricultural research, especially in high-throughput marker-assisted breeding (Rafalski, 2002; Fang et al. 2014). The identification of SNPs in ESTs allow to estimate allele frequencies and association to interesting phenotypes, providing a powerful resource for genetic association studies (Rafalski, 2002). Several technologies amenable to automation have been used for SNP discovery in plants (Ching and Rafalski 2002; Jander et al. 2002; Gotoh and Oishi 2003; Pacey-Miller and Henry 2003; Schwarz et al. 2004). Among them, the resequencing candidate gene approach is commonly used to identify a polymorphism by aligning the amplified fragment in different individuals (Pavy et al. 2008; Deleu et al. 2009). This technique has an advantage, because depending on the number of individuals used in the study, it does not requires validation (Munerato, 2005).

A large number of SNPs has been also identified in EST collections (Argout et al. 2008, Lima et al. 2009). In 2009, Lima and collaborators reported the detection and analysis of 71 SNPs in ESTs from the cacao-*M. perniciosa* interaction using *in silico* analysis, but these SNPs have not been validated. In the present study, we report the discovery, validation and analysis of SNPs from cacao-*M. perniciosa* interaction ESTs using resequencing of candidate genes. These SNPs are potential markers for subsequent use in MAS breeding programs of cacao for WBD resistance.

Materials and Methods

Identification and haplotype diversity of SNPs from the ESTs collection

Seventy-three EST sequences corresponding to plant-pathogen interaction related to resistance genes were selected from cacao-*M. perniciosa* interaction libraries of different genotypes, as follow: 53 genes from TSH 1188 (resistant genotype; 40 from pods and 13 from meristem; Gesteira et al. 2007, Zaidan et al. 2005; CEPLAC projects granted by FAPESB and CNPq), 6 from Scavina 6 (Sca 6, resistant genotype; Argout et al. 2008), and 14 from Catongo (susceptible genotype; Gesteira et al. 2007) (Table 1). The software package Primer3 was used to design primers flanking each candidate gene (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Once primer pairs were designed, they were optimized on DNA progenitors of the F_2 population, the Sca 6 and ICS1 (Faleiro et al. 2006). Initially, SNPs detection were made on the Sca 6 and ICS 1 progenitors and on the F1 TSH 516 population segregating for WDB resistance (Faleiro et

Table 1. Characteristics of the cDNA libraries used in this study for putative SNP identification

Cultivar	Phenotype[a]	Tissue	Library type	Reference
TSH1188	R/inoculated	Meristem	Full-length	Gesteira et al. 2007
	R/inoculated	Pod	Full-length	Zaidan et al. 2005
Scavina 6	R/inoculated	Pod	SSH	Argout et al. 2008
Catongo	S/inoculated	Meristem	Full-length	Gesteira et al. 2007

[a] R = resistant; S = susceptible; inoculated: with *M. perniciosa* spores

al. 2006). The validation of the polymorphic alleles (1% of the individuals) from Sca 6, ICS 1 and TSH 516 was carried out on 68 individuals of the F_2 (Sca 6 × ICS 1) population. DNA was extracted using the method of Doyle and Doyle (1990), with minor modifications. The polymerase chain reaction (PCR) (20 ì l) was performed as follows: 10 ng of DNA, 0.2 mmol.L^{-1} of each primer, 2.0 mmol.L^{-1} of $MgCl_2$, 0.2 mmol.L^{-1} of each dNTP (Ludwig Biotecnologia Ltda), buffer 1X and 1 U of *Taq* DNA Polimerase (MBI Fermentas). Amplified PCR products (6 ì l) were treated with 0.5 ì l of ExoSAP-IT reagent (USB Corporation) in a 7 ì l final volume. Sequencing was done directly from the PCR product on an ABI 3100 sequencer (Applied Biosystems). ABI Electropherograms trace files were aligned and mutations were identified using ABI Prism SeqScape Software version 2.0 (Applied Biosystems). Sequences were aligned with BioEdit v. 7.0.9.0 software (Hall, 1999) and the parameters of nucleotide and haplotype diversity for the validated SNPs were calculated using the DNASP version 5 (Librado and Rozas, 2009). Haplotype based gene diversity (Hd) and the haplotype-based polymorphic information content (PIC) values for each gene fragment were computed as described by Botstein et al. (1980): $PIC = 1 - \sum Pi^2 - \sum 2Pi^2Pj^2$, where $\sum Pi^2$ is the sum of each squared haplotype frequency. Each haplotype was considered to be an allele.

Putative sequence function analysis and data processing

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).

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

Samples of ESTs containing SNPs were randomly chosen for localization of the EST-SNP 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-SNP was localized in relation to the ORF.

Results and Discussion

From the 73 ESTs sequences selected related to cacao-*M.perniciosa* interaction, only 38 sequences were enable to obtain the annealing temperature and 11 sequences shawing good quality for sequencing were analyzed from PCR products, identifying a total of 185 putative SNPs. The PCR products corresponded to PCR amplification with primer pairs designed from the genes: ABC transporter, Acinus protein, Beta 1,3 glucanase, Disease resistance protein, Ferredoxin family, Jasmonic Acid 2, Leucine rich repeat kinase, Pathogenesis related protein, Pathogenesis related protein 4B, Phenylalanine ammonia lyase and Resistance protein candidate. Among the putative SNPs, 57% were substitutions and transversion types, 29% of the substitutions were transitions type. The last 14% was Indels with 13 insertions and 14 deletions. All genes presented both types of substitutions, except the Pathogenesis related protein and Jasmonic acid 2 which had only one transition and one transversion, respectively. The transversion type appeared with a higher frequency in the ABC transporter (84.7%) gene, followed by the Acinus protein gene (73.6%). Whereas the transition type showed a higher frequency in the genes Ferredoxin family protein (55%) and Resistance protein candidate (50%). The most representative putative SNPs for transition type corresponded to the substitution of cytosine for thymine (C/T; 12%), followed by substitution of thymine for cytosine (T/C; 8%) and guanine for adenine (G/A; 7%). Among the putative SNPs of the transversion type substitution, 12% were substitutions of thymine for adenine (T/A) and 11% of cytosine for adenine (C/A). The other 50% corresponded to A/G transitions and G/T, T/G, A/C, A/T, C/G and G/C transversions.

To explore the potential utility of the putative SNPs in research of cacao structural genome, 11 genes used for discovery SNPs were compared to the Genbank database and classified into 14 main categories as described for Gesteira et al. (2007). However, the SNPs were found only in 6 categories (Figure 1). Most of the sequences containing SNPs were placed in the defense and cell rescue, gene expression and RNA metabolism categories (56%). While, 13% of the sequences were found in membrane transport and 11% in the categories primary metabolism, miscellaneous and abiotic stimuli and development. SNPs were also found in very

important defense genes, these potentially related to cacao resistance/susceptibility to WBD, such as pathogenesis related protein, which is believed to participate in the phenomenon of systemic acquired resistance (SAR) (Durrant and Dong 2004; Coram and Pang 2005). The validation of the polymorphic alleles from Sca 6, ICS 1 and TSH516 using the 68 individuals from the Sca 6 x ICS 1 F_2 population, allowed to confirm 91 SNPs,which were in the â-1,3-glucanase (17 SNPs; 3 Indels), Phenylalanine ammonia lyase (17 SNPs; 7 Indels), Pathogenesis related protein 4B (9 SNPs; 3 Indels), Disease resistance protein (32 SNPs; 3 Indels) genes (Table 2). In this study, the frequency of SNP corresponds to an average of 1 SNP per 31.25 bp in coding regions, and 1 SNP per 15.5 bp in the non-coding regions (Table 2). Similar high frequency of SNPs has been found in other crops, for example in maize wherein 1 SNP per 31 bp in non-coding region and 1 SNP per 124 bp in the coding region, these were obtained based on the analysis of 18 genes in 36 inbred lines (Ching et al. 2002). Lima et al. (2009) performing an *in silico* SNPs analysis from cacao-*M. perniciosa* ESTs, found a frequency of 1 putative SNP per 71 bp in the coding regions, and 1 putative SNP per 223 bp in non-coding regions. Therefore, *in silico* analysis which used very stringent criteria for SNPs identification may have caused loss of good SNPs and finally appeared to be less efficient compared to the analysis done herein. Thus, the high efficiency of the re-sequencing approach provides a powerful tool for genetic analysis in *T. cacao*. Nevertheless, caution comparing both methods (*in silico vs* re-sequencing) is needed, because the analyzed genes were not the same. The frequency of Indels per bp was on average of 1 per 74.75 bp in coding regions, and 1 per 34 bp in the non-coding regions (Table 2). Substitutions in the coding region that lead to alterations in the amino acid sequences and/or early termination of the translation process may cause an altered phenotype (Klug et al. 2005). If the altered phenotype is valuable for breeding, then these coding SNPs could be considered useful genetic markers for cacao, allowing the association between genotypic and phenotypic variation (Rafalski 2002a; McCallum et al. 2000).

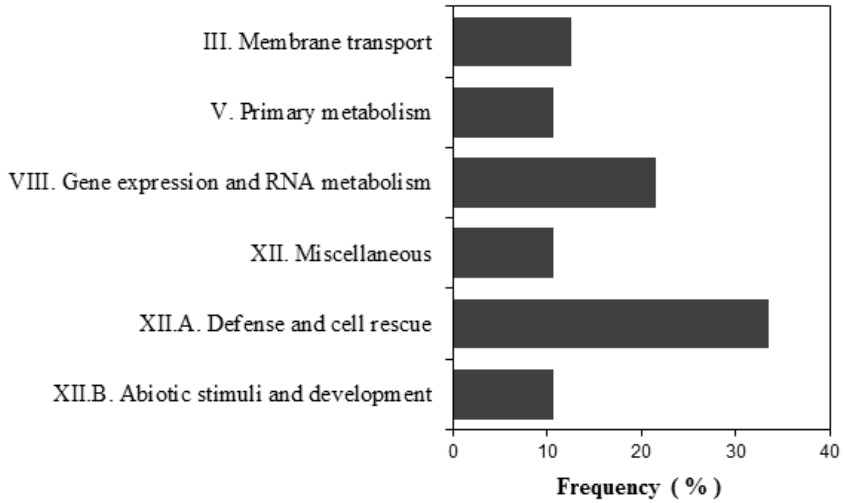

Figure 1. Frequency of putative SNPs into functional classes. The categories that were used for classification are indicated, as well as the frequency of putative SNPs in the each category.

Table 2. Summary of validation of SNPs identified in cacao-*M. perniciosa* ESTs

Parameters	Gene			
	BG	PAL	PR4B	DRP
Number of individual used in the validation	68	68	68	68
Size of sequence (bp)	317	478	479	509
Number of SNPs	17	17	9	32
Frequency of polymorphic sites per bp	1/19	1/28	1/53	1/16
Frequency of polymorphic sites per bp (coding)	1/40	1/18	1/47	1/20
Frequency of polymorphic per bp (non-coding)	1/16	1/33	-	1/13
Number of Indels	3	7	3	3
Overall Indel frequency	1/106	1/68	1/159	1/170
Frequency of Indels per bp (coding)	-	-	1/211	1/88
Frequency of Indels per bp (non-coding)	1/27	1/53	1/56	-

BG: Beta 1,3-glucanase; PAL: Phenylalanine ammonia lyase; PR4B: Pathogenesis related protein 4B; DRP: Disease resistance protein

In the Figure 2, we observed the position of the 75 SNPs regarding to the open reading frame (ORF) of the corresponding ESTs (1a), and the classification in non-synonymous (leading to amino acid alteration) or synonymous (without amino acid alteration) (1b). High

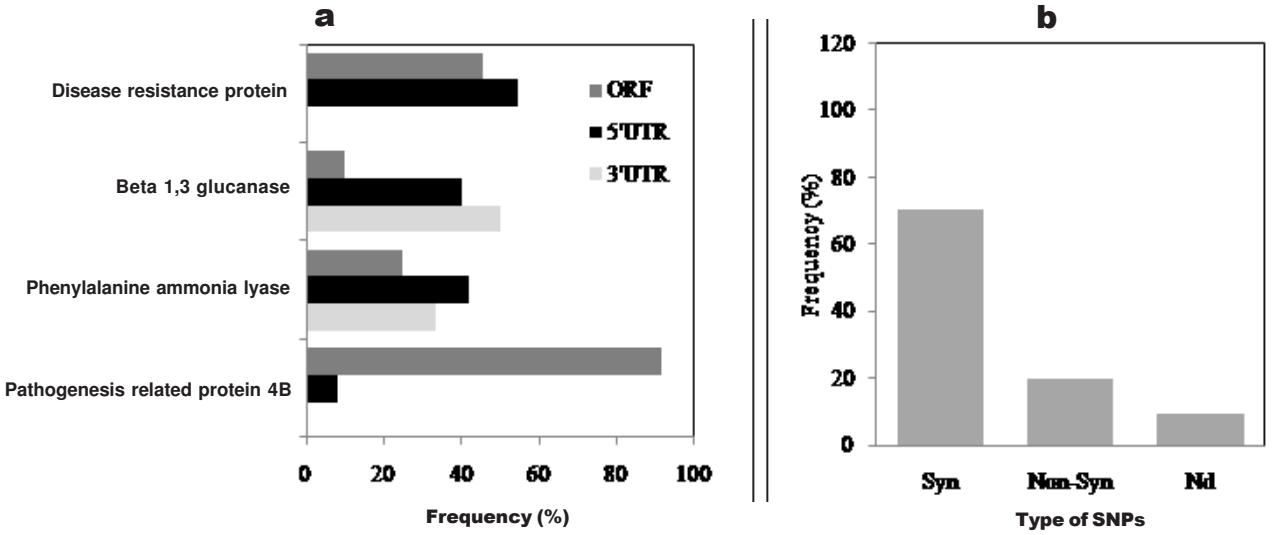

Figure 2. Frequency of SNPs validated in four cacao-*M. perniciosa* ESTs, ...elation to position in ORF and protein synthesis. a. Frequency of SNPs in the different sequence parts. ORF: open reading frame; UTR: untranslated region. b. Frequency of synonymous (Syn.) and non-synonymous (Non-Syn.) SNPs, Nd: not determined.

frequency or no SNP was found in the ORF and in the 5'UTR (untrans-lated region), low frequency or no SNP was observed in the 3'UTR, in the four genes in which the SNPs were validated, as follows: Disease resistance protein (46% in ORF, 54% in 5'UTR), â-1,3-glucanase (10% in ORF, 40% in 5'UTR, 50% in 3'UTR), Phenylalanine ammonia lyase (25% in ORF, 42% in 5'UTR, 33% in 3'UTR), and Pathogenesis related protein (92% in ORF, 8% in 5'UTR). The presence of a great amount of SNPs in ORFs was unexpected and contradicted previous works showing that the UTRs have a higher frequency of SNPs than the rest of the genome (Zhu et al. 2001; Ching et al. 2002), but agrees with Lima et al. (2009) work, in which SNPs were found more frequently in the ORF (44%) and in the 5'UTR (untranslated region) (32%) than in the 3'UTR. The condition of forward sequencing may explain the low frequency or absence of SNPs in the 3'UTR; the 5'UTR was sequenced, and the 3'UTR was, either partially sequenced or not sequenced depending on the length of the ORF, making it difficult to detect SNPs in this region. We also found more frequent synonymous SNPs (71%) than non-synonymous (20%), and 9% were impossible to determinate. When synonymous, the SNP can change the structure and stability of the messenger RNA, whereas non-synonymous SNP can change the structure and function of the protein and consequently affect the amount of produced protein and the presence of deleterious mutations.

Diversity values (expected heterozygosity) for SNPs were generally low due to their bi-allelic nature. Here, nucleotide diversity varied among the gene fragments, from 1.88 x 10^{-3} in the Phenylalanine ammonia lyase gene to 6.24 x 10^{-3} in Disease resistance protein gene. This potential handicap of SNP can be overcome either by using larger number of markers or by considering haplotypes structure for each locus in place of single SNPs. Values from Hd or heterozygosity and PIC were high in the Disease resistance protein gene (Hd = 0.559; PIC = 0.56), and low in the Phenylalanine ammonia lyase gene (Hd = 0.115; PIC = 0.12). The number of haplotypes varied among the gene fragments, from 4 in Phenylalanine ammonia lyase to 16 in Disease resistance protein gene (Table 3). The distribution of SNPs along gene fragments was examined to understand if nucleotide variability was

Table 3. Total number of haplotypes, haplotype diversity, nucleotide diversity and haplotype-based polymorphism information content of validated cacao genes

Gene	Haplotypes	Hd ± SD	Nucleotide diversity ± SD	PIC
BG	6	0.251±0.085	4.48 x 10^{-3}± 2.03 x 10^{-3}	0.26
DRP	16	0.559±0.088	6.24x 10^{-3}±1.71x 10^{-3}	0.56
PAL	4	0.115±0.061	1.88 x 10^{-3}±1.26 x 10^{-3}	0.12
PR4B	6	0.316±0.101	3.34x 10^{-3}±1.37 x 10^{-3}	0.30

BG: Beta 1,3 glucanase; DRP: Disease resistance protein. Hd: haplotype based gene diversity; PAL: Phenylalanine ammonia lyase; PR4B: Pathogenesis related protein 4B; PIC: polymorphic information content; SD: standard deviation.

distributed randomly or organized in haplotypes. Based on our results, most of the gene fragments had a major haplotype accompanied by a series of low frequency haplotypes (Figure 3). Data on phenotypic resistance to *M. perniciosa* obtained through artificial inoculations of the population used to validate the SNPs have being obtained to detect specific associations between haplotypes and resistance. This type of association has been found in Arabidopsis; Hao et al. (2009) suggested that the accession 163av has a specific haplotype that generates a phenotypic change related to drought tolerance.

The identification of SNPs from ESTs seems to be a potent approach to investigate markers in coding sequences and to relate markers to cellular function. The re-sequencing approach appears to became determinant to identify useful SNPs for wide genetic applications. Furthermore, the availability of the whole genome sequence allow a positional selection of DNA fragments to be re-sequenced, enhancing the usefulness of the discovered SNPs. Therefore, given the above, we recommend this functional co-dominant marker system for large-scale for the identification of allelic status of the cacao resistance genes through MAS, to support the development of the promising genotypes with high resistance durability to WBD.

Acknowledgements

The present study was carried out in the Molecular Plant Pathology Laboratory (FITOMOL) of the Cocoa Research Center, Bahia, Brazil. This research was

Nucleotide positions in sequence of gene

BG	nº of individuals		50	51	52		152	153	154	155	156	157		214	215	216	217	218		
Hap 1	68	*	A	**T**	G	*	C	C	**A**	**A**	T	**A**	*	**T**	**C**	T	**A**	G	*	*
Hap 2	6	*	A	**T**	G	*	C	C	**A**	**A**	T	**A**	*	**C**	**C**	T	**A**	G	*	*
Hap 3	2	*	A	**T**	G	*	C	C	**A**	**T**	T	**G**	*	**T**	**A**	T	**T**	G	*	*
Hap 4	2	*	A	**T**	G	*	C	C	**A**	**A**	T	**A**	*	**T**	**C**	T	**A**	G	*	*
Hap 5	1	*	A	**A**	G	*	C	C	**G**	**A**	T	**T**	*	**T**	**C**	T	**A**	G	*	*
Hap 6	1	*	A	**T**	G	*	C	C	**A**	**A**	T	**A**	*	**C**	**C**	T	**T**	G	*	*

PAL	nº of individuals		34	35	36	37	38	39	40		118		154	155	156		221		258	
Hap 1	1	*	**A**	G	C	**C**	**C**	A	T	*	**T**	*	**A**	**A**	**G**	*	**T**	*	**C**	*
Hap 2	64	*	**A**	G	C	**C**	**C**	A	T	*	**T**	*	**A**	**A**	**G**	*	**T**	*	**T**	*
Hap 3	1	*	**G**	G	C	**A**	**C**	A	T	*	**T**	*	**G**	**G**	**A**	*	**A**	*	**T**	*
Hap 4	2	*	**A**	G	C	**C**	**T**	A	T	*	**G**	*	**A**	**A**	**T**	*	**T**	*	**T**	*

PR4B	nº of individuals		89		178	179	180	181	182	183		278	279	280	281	282		295	296	
Hap 1	1	*	**C**	*	G	G	T	**G**	C	**C**	*	**A**	**A**	**C**	T	G	*	A	**T**	*
Hap 2	60	*	**C**	*	G	G	T	**G**	C	**C**	*	**A**	**T**	**T**	T	G	*	A	**C**	*
Hap 3	3	*	**C**	*	G	G	T	**A**	C	**C**	*	**A**	**T**	**T**	T	G	*	A	**C**	*
Hap 4	1	*	**A**	*	G	C	T	**A**	C	**C**	*	**A**	**T**	**T**	T	G	*	A	**C**	*
Hap 5	2	*	**C**	*	G	G	T	**G**	C	**T**	*	**A**	**T**	**T**	T	G	*	A	**C**	*
Hap 6	1	*	**C**	*	G	G	T	**G**	C	**C**	*	**T**	**T**	**T**	T	G	*	A	**C**	*

DRP	nº of individuals		15	16		55	56		102	103	104		210	211		254	255	256		
Hap 1	50	*	**A**	G	*	G	**A**	*	T	**T**	**G**	*	**A**	**A**	*	**A**	**C**	**G**	*	*
Hap 2	1	*	**A**	G	*	G	**C**	*	T	**A**	**G**	*	**A**	**A**	*	**A**	**C**	**G**	*	*
Hap 3	1	*	**A**	G	*	G	**A**	*	T	**T**	**G**	*	**A**	**A**	*	**A**	**C**	**G**	*	*
Hap 4	1	*	**A**	G	*	G	**A**	*	T	**T**	**G**	*	**A**	**A**	*	**A**	**A**	**G**	*	*
Hap 5	1	*	**G**	C	*	A	**A**	*	T	**T**	**G**	*	**C**	**C**	*	**A**	**C**	**G**	*	*
Hap 6	1	*	**A**	G	*	G	**A**	*	T	**T**	**G**	*	**C**	**A**	*	**A**	**C**	**G**	*	*
Hap 7	1	*	**A**	G	*	G	**G**	*	T	**T**	**G**	*	**A**	**A**	*	**C**	**C**	**G**	*	*
Hap 8	1	*	**A**	G	*	G	**A**	*	T	**G**	**G**	*	**C**	**A**	*	**A**	**C**	**T**	*	*
Hap 9	2	*	**A**	G	*	G	**A**	*	T	**T**	**G**	*	**C**	**A**	*	**A**	**C**	**G**	*	*
Hap 10	1	*	**A**	G	*	G	**A**	*	T	**T**	**A**	*	**A**	**A**	*	**A**	**C**	**G**	*	*
Hap 11	1	*	**A**	G	*	G	**A**	*	T	**T**	**G**	*	**A**	**A**	*	**A**	**C**	**G**	*	*
Hap 12	3	*	**A**	G	*	G	**A**	*	T	**T**	**G**	*	**A**	**A**	*	**A**	**C**	**G**	*	*
Hap 13	1	*	**A**	G	*	G	**A**	*	T	**T**	**G**	*	**A**	**A**	*	**A**	**C**	**G**	*	*
Hap 14	1	*	**A**	G	*	G	**A**	*	T	**T**	**G**	*	**A**	**A**	*	**A**	**C**	**G**	*	*
Hap 15	1	*	**A**	G	*	G	**A**	*	T	**T**	**G**	*	**C**	**A**	*	**A**	**C**	**G**	*	*
Hap 16	1	*	**A**	G	*	G	**A**	*	T	**T**	**G**	*	**A**	**A**	*	**A**	**C**	**G**	*	*

Figure 3. Haplotypes detected for four analyzed ESTs in 68 individuals of the F_2 (Sca 6×ICS 1) population. The SNPs are indicated in bold. The * represents interruption in base sequence. BG: Beta 1,3 glucanase; DRP: Disease resistance protein; HAP: haplotype; PAL: Phenylalanine ammonia lyase; PR4B: Pathogenesis related protein 4B. The rectangle in the first line indicates the haplotypes group.

supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Ministère des Affaires Etrangères Français (MAE).

Literature Cited

AIME M. C., 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.

ARGOUT, X. et al. 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.

BOTSTEIN, D. et al. 1980. Construction of genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32:314-331.

CHING, A. et al. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3:19.

CHING, A.; RAFALSKI, A. 2002. Rapid genetic mapping of ESTs using SNP pyrosequencing and indel analysis. Cellular and Molecular Biology Letters 7:803-810.

CHO, R. J. et al. 1999. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nature Genetics 23:203-207.

CORAM, T. E.; PANG, E. C. K. 2005. Isolation and analysis of candidate ascochyta blight defence genes in chickpea. Part II. Microarray expression analysis of putative defence-related ESTs. Physiological and Molecular Plant Pathology 66: 201-210.

CORPET, F. et al. 2000. ProDom and ProDom-CG: Tools for protein domain analysis and whole genome comparisons. Nucleic Acids Research 28:267-269.

DELEU W. et al. 2009. A set of EST-SNPs for map saturation and cultivar identification in melon. BMC Plant Biology 9:90.

DOYLE J. J.; DOYLE J. L. 1990. Isolation of plant DNA from fresh tissue. Focus 12:13-15.

DURRANT, W. E.; DONG, X. 2004. Systemic Acquired Resistance. Annual Review of Phytopathology 42:185-209.

FALEIRO, F. G. 2006. Mapping QTLs for witches' broom (*Crinipellis perniciosa*) resistance in cacao (*Theobroma cacao* L.). Euphytica 149:227-235.

FANG, W. et al. 2014. Varietal identification of tea (*Camellia sinensis*) using nanofluidic array of single nucleotide polymorphism (SNP) markers. Horticultural Research 1:14035

FRIAS, G. A. et al. 1991. Infection biology of *Crinipellis perniciosa* on vegetative flushes of cacao. Plant Disease 75:552-556.

GESTEIRA, A. et al. 2007. Exploring the cacao-*Moniliophthora perniciosa* interaction using EST analysis. Annals of Botany 100:129-140.

GOTOH, K.; OISHI, M. 2003. Screening of gene-associated polymorphismsn by use of in-gel competitive reassociation and EST (cDNA) array hybridization. Genome Research 13: 492–495.

HALL, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis. http://www.mbio.ncsu.edu/BioEdit/bioedit.html.

HAO, G. P. et al. 2009. Nucleotide variation in the NCED3 region of *Arabidopsis thaliana* and its association study with abscisic acid content under drought stress. Journal of Integrative Plant Biology 51:175-83.

HARRIS, M. A. et al. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research 32:D258-D261

JANDER, G. et al. 2002. Arabidopsis map-based cloning in the post-genome era. Plant Physiology 129:440-450.

JOURNET, E. P. et al. 2002. Exploring root symbiotic programs in the model legume *Medicago truncatula* using EST analysis. Nucleic Acids Research, 30:5579-5592.

KLUG, W. S.; CUMMINGS, M. R.; SPENCER, C. 2005. Concepts of genetics.

LIBRADO, P.; ROZAS J. 2009. DnaSP v. 5, A software for comprehensive analyses of DNA polymorphism data. Bioinformatics 25:1451-1452.

LIMA, L. et al. 2009. Single nucleotide polymorphisms from *Theobroma cacao* expressed sequence tags associated with witches' broom disease in cacao. Genetics Molecular Research 8:799-808.

LIU, W. et al. 2015. Identifying litchi (*Litchi chinensis* Sonn.) cultivars and their genetic relationships using single nucleotide polymorphism (SNP) markers. PLoS ONE 10(8).

MCCALLUM, C. M. et al. 2000. Targeted screening for induced mutations. Nature Biotechnology.

MUNERATO, P. 2005. Curso: Identificação de SNPs e genotipagem através de eletroforese capilar e sistema de detecção de seqüência. ABI EXPERT Training Center-Applied Biosystems. São Paulo, SP. 51p.

PACEY-MILLER, T.; HENRY R. J. 2003. SNP detection in plants using a single stranded pyrosequencing protocol with a universal biotinylated primer. Analytical Biochemistry 317:165-170.

PAVY, N. et al. 2008. Enhancing genetic mapping of complex genomes through the design of highly multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. BMC Genomics 9:21.

PICOULT-NEWBERG, L. et al. 1999. Mining SNPs from EST databases. Genome Research 9:167-174.

RAFALSKI, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 5:94-100.

RAFALSKI, J. A. 2002a. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Science 162:329-333.

SCHWARZ, G. et al. 2004. Determination of detection and quantification limits for SNP allele frequency estimation in DNA pools using real time PCR. Nucleic Acids Research 32:24.

ZAIDAN, H. A. et al. 2005. Caracterização molecular da resistência do cacaueiro (*Theobroma cacao* L.) à vassoura de bruxa. In: Congresso Brasileiro de Fitopatologia, 38. Brasilia, DF, Brasil.

ZHU, Y. L. et al. 2001. Single Nucleotide Polymorphisms (SNPs) in Soybean Genes, cDNAs, and Random Genomic Sequence. In: Plant and Animal Genome Conference, 9, San Diego.

ZHOU, L. et al. 2016. Developing Single Nucleotide Polymorphism (SNP) Markers for the Identification of Coffee Germplasm. Tropical Plant Biology 9:82-95.

●