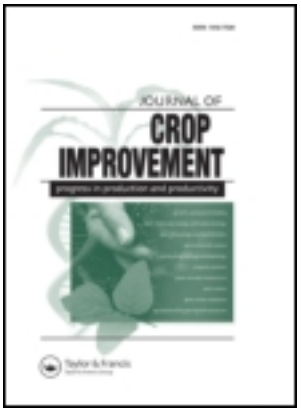

Journal of Crop Improvement
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wcim20

Conversion of Microsatellite Markers to Single Nucleotide Polymorphism (SNP) Markers for Genetic Fingerprinting of Theobroma cacao L.

Abu M. Dadzie * [a], Donald S. Livingstone III* [b], Stephen Y. Opoku [a], Jemmy Takrama [a], Francis Padi [a], Samuel K. Offei [d], Eric Y. Danquah [d], Juan Carlos Motamayor [c], Raymond J. Schnell [c] & David N. Kuhn [b]

[a] Cocoa Research Institute of Ghana (CRIG), Tafo-Akim, Ghana

[b] USDA-ARS, SHRS, Miami, Florida, USA

[c] Mars Inc., Hackettstown, New Jersey, USA

[d] Department of Crop Science, College of Agriculture and Consumer Sciences, University of Ghana, Accra, Ghana

Version of record first published: 18 Jan 2013.

To cite this article: Abu M. Dadzie * , Donald S. Livingstone III* , Stephen Y. Opoku , Jemmy Takrama , Francis Padi , Samuel K. Offei , Eric Y. Danquah , Juan Carlos Motamayor , Raymond J. Schnell & David N. Kuhn (2013): Conversion of Microsatellite Markers to Single Nucleotide Polymorphism (SNP) Markers for Genetic Fingerprinting of Theobroma cacao L., Journal of Crop Improvement, 27:2, 215-241

To link to this article: http://dx.doi.org/10.1080/15427528.2012.752773

Journal of Crop Improvement, 27:215–241, 2013

ISSN: 1542-7528 print/1542-7536 online
DOI: 10.1080/15427528.2012.752773

Taylor & Francis
Taylor & Francis Group

Conversion of Microsatellite Markers to Single Nucleotide Polymorphism (SNP) Markers for Genetic Fingerprinting of *Theobroma cacao* L.

ABU M. DADZIE*[1], DONALD S. LIVINGSTONE III*[2], STEPHEN Y. OPOKU[1], JEMMY TAKRAMA[1], FRANCIS PADI[1], SAMUEL K. OFFEI[4], ERIC Y. DANQUAH[4], JUAN CARLOS MOTAMAYOR[3], RAYMOND J. SCHNELL[3], and DAVID N. KUHN[2]

[1]*Cocoa Research Institute of Ghana (CRIG), Tafo-Akim, Ghana*
[2]*USDA-ARS, SHRS, Miami, Florida, USA*
[3]*Mars Inc., Hackettstown, New Jersey, USA*
[4]*Department of Crop Science, College of Agriculture and Consumer Sciences, University of Ghana, Accra, Ghana*

The majority of the world's cacao for chocolate manufacture is produced in West Africa. Cocoa breeding programs in West Africa need genetic markers to reduce the time needed for improving cocoa by screening seedlings for the presence of the markers rather than mature plants for the phenotypic traits (i.e., marker-assisted selection [MAS]). For MAS to be successful, the breeder must have both access to markers linked to desired traits and a convenient marker-assay system that can be performed locally. In this study, microsatellite markers that flanked disease resistance quantitative trait loci (QTL) but could not be assayed conveniently in West Africa were converted using a genome walking method into single nucleotide polymorphism (SNP) markers that could be assayed

Received 20 September 2012; accepted 21 November 2012.

*Contributed equally to research and writing of manuscript.

This work was supported by the USDA-ARS Subtropical Horticulture Research Station (SHRS) Project 6631-21000-017-26 Genetic Improvement of Cacao, which is a Mars Inc. trust to the USDA. Abu Dadzie was also supported by the World Cocoa Foundation and the Norman Borlaug fellowship program. We thank Eric Brenyah and Clement Aryee of the Physiology/Biochemistry Division of CRIG for technical assistance. We also want to thank all staff of SHRS for their support and assistance, especially Cecile Tondo, Kathleen Cariaga, Barbie Freeman, Dr. Stefan Royaert, and Wilber Quintanilla.

This work is published with permission of the executive director of CRIG.

Address correspondence to Dr. David N. Kuhn at USDA-ARS, SHRS, 13601 Old Cutler Road, Miami, FL 33158, USA. E-mail: david.kuhn@ars.usda.gov

locally. The SNP and microsatellite markers were equally effective in identifying off-types in two different mapping populations of cacao. Also, SNPs cast doubt on whether all microsatellite markers are identical by descent.

KEYWORDS cacao, marker-assisted selection (MAS), microsatellite markers, single nucleotide polymorphism (SNP) markers, genome walking, identity by descent

INTRODUCTION

Cacao (*Theobroma cacao* L.) is a tropical understory tree whose beans are the source of cocoa and cocoa butter for chocolate manufacturing and cosmetics. Cacoa beans, the fermented, dried seeds of cacao, serve as a major source of income for both small- and large-scale farmers in the humid tropics and are the main cash crop cultivated in West Africa. Cacao is believed to have originated from the Upper Amazon forest (Cheesman 1944; Motamayor et al. 2008), but 70% of the world's cocoa is now produced in West Africa (International Cocoa Organization 2010) . World cocoa production is negatively impacted by several factors, including diseases caused by *Phytophthora* species and cocoa swollen shoot virus (CSSV) (Ploetz 2007). It is estimated that yield loss caused by *Phytophthora megakarya* alone ranges between 80% and 100% in regions of Ghana, threatening the livelihood of many cocoa farmers (Opoku et al. 1997). In addition, misidentified genotypes used in breeding programs may result in the distribution of unimproved planting materials to farmers (Motilal and Butler 2003; Takrama et al. 2005).

To improve the yield and disease resistance of cacao, breeding programs in West African countries have been active since the establishment of the West African Cocoa Research Institute (now Cocoa Research Institute of Ghana [CRIG]) in 1938. The generation time of cacao is approximately five years, which slows progress in breeding programs. In general, to accelerate improvement in crops through breeding, it is advantageous to associate favorable agronomic traits with genetic markers (Rafalski 2002; Collard and Mackill 2008), a process known as marker-assisted selection (MAS). In cacao, there have been efforts to identify genetic markers associated with traits, such as yield and disease resistance (Clement et al. 2003a, 2003b; Schnell et al. 2005; Brown et al. 2007; Lanaud et al. 2009). By genotyping seedlings from crosses, breeders can screen thousands of progeny and only retain those with the favorable alleles for evaluation in the field, which indeed shortens the breeding cycle. However, this then requires genetic markers that can be assayed conveniently at the site of the breeding program. For cacao breeding, the majority of the genetic markers currently in use are microsatellite markers; these markers have proved difficult to assay in West

African cocoa-producing countries because of limited access or unavailability of high-resolution microsatellite assay platforms and the concomitant cost of reagents for these platforms.

Recently, a SNP assay has been developed (Livingstone et al. 2011; Livingstone et al. 2012) that has been successfully used at CRIG (Takrama et al. 2012). However, most of the maps and QTL-associated markers for cocoa are based on microsatellites (Pugh et al. 2004; Brown et al. 2007; Schnell et al. 2007; Brown et al. 2008; Lanaud et al. 2009). Therefore, a means to convert microsatellite markers into SNPs for MAS in West Africa was needed.

The microsatellite markers that have been associated with the quantitative trait loci (QTL) through mapping are assayed by amplifying the microsatellite region using primers that flank it. An assumption about these microsatellite markers is that two microsatellite alleles of the same length are inherited from a common ancestor and are identical by descent, rather than being identical by chance or state (Schroeder et al. 2009). Two alleles that are identical by chance (i.e., do not share the same sequence in the flanking regions and did not inherit the allele from a common ancestor) are considered to be an example of size homoplasy, which means that nothing can be inferred from the association of an allele of that size with a trait (Estoup et al. 2002). In association studies, only alleles that are identical by descent have value for MAS.

Using a method known as genome walking (Parker et al. 1991), more sequence data can be obtained from the region that flanks the microsatellite region at both the 5' and 3' end. This additional sequence can be searched for SNPs by amplifying the parents of the mapping populations and identifying nucleotide positions where one of the parents is heterozygous. The genome walker technique was employed for microsatellite markers that flanked black pod QTL to convert eight microsatellite markers into SNP markers, which will allow association of a particular allele with the phenotype in the progeny. Here we describe the conversion of QTL-associated microsatellite markers into SNP markers and their use in determining off-types in two related mapping populations.

MATERIALS AND METHODS

Plant Materials

The plant materials used in this study were fresh cacao leaves obtained from progeny of two crosses ('Pound 7' × 'P 30' and 'PA 7' × 'P 30') established in 1981 at Apedwa Research Station in the Eastern region of Ghana. These progeny segregate for black pod-disease resistance (Adomako 2006). The first population ('Pound 7' × 'P 30') consisted of 85 individuals, whereas the second population ('PA 7' × 'P 30') consisted of 105 individuals.

The female parents were selected based on recorded resistance to black pod disease (Adomako 2006). They belong to the Upper Amazon genetic group, now sub-grouped into Nanay ('Pound 7') and Marańoń ('PA 7'), while the male parent ('P 30'), which is a local selection, belongs to the Amelonado group (Motamayor et al. 2008).

DNA Extraction

At the molecular biology laboratory at CRIG, genomic DNA was extracted from fresh mature leaf samples of the individual trees in the two populations as described in Opoku et al. (2007) and was given a laboratory identifier beginning with the letters TC for easy tracking, which also identified their exact location in the field. At USDA-ARS Subtropical Horticulture Research Station (SHRS), genomic DNA of 'Matina 1–6' was isolated and prepared for genome walking, as described in Livingstone et al. (2011). In addition, previously isolated DNA from the following cacao cultivars, 'SCA 12' (Contamana), 'CAB 0331 PL4' (Iquitos), 'Las Brisas 17' (Nacional), 'BE 4PL3' (Amelonado), 'PA 7' (Maranon), 'Coc 3335' (Curaray), 'Pound 7' (Nanay), 'GU 124A' (Guiana), 'UF 273' (Trinitario hybrid), 'Peru' (unknown), and 'TSH 516' ('SCA 6' × 'ICS 1'), was used to identify SNPs.

DNA Quantification

The DNA was quantified using Sybr Green (Lonza, CA, USA) and the FLX 800 micro-plate fluorescence reader (Bio-Tek Instruments, Winooski, VT, USA), as described in Livingstone et al. (2009). The samples were then diluted to a working concentration of 4 ng/μl for polymerase chain reaction (PCR) amplification.

PCR and Microsatellite Analysis

Six polymorphic microsatellite markers (mTcCIR12, mTcCIR21, mTcCIR42, mTcCIR87, mTcCIR141, and mTcCIR128) were used to fingerprint progeny of 'PA 7' × 'P 30' and 'Pound 7' × 'P 30' and their parental clones. These markers, designed by Centre de Cooperation Internationale en Recherche Agronomique pour le Development (CIRAD), Montpelier, France (Lanaud et al. 1999), are highly polymorphic and as such are recognized as part of the international marker set for genetic fingerprinting of cacao germplasm (Saunders et al. 2004; Zhang et al. 2009; Irish et al. 2010; Zhang et al. 2012). A further set of eight markers (mTcCIR24, mTcCIR37, mTcCIR61, mTcCIR76, mTcCIR160, mTcCIR200, mTcCIR211, and mTcCIR222) that flanked QTL regions was also used to genotype the progeny of both populations and their parental clones after off-types had been removed. A DNA Engine Tetrad 2,

Peltier Thermal cycler (BioRad, Hercules, CA, USA) was used to perform PCR amplification in a 10 μl reaction mixture, as described in Schnell et al. (2005).

Construction of GenomeWalker DNA Libraries

A genome walking protocol by Clontech Laboratories Inc. (GenomeWalker, Mountain View, CA, USA) was followed in the construction of genomic DNA libraries from a 'Matina 1–6' clone. In brief, four restriction enzymes (DraI, EcoRV, PvuII, and StuI) were used to prepare four blunt end libraries. Each library was set up by combining 'Matina 1–6' genomic DNA (0.1 μg/μl) with 10 units/μl restriction enzyme, 10μl of 10X restriction enzyme buffer, and de-ionized water. The reaction tube was gently inverted to ensure thorough mixing without shearing the genomic DNA. The mixture was incubated at 37°C for two hours. The mixture was then vortexed slowly and incubated at 37°C overnight (16 hours). From each of the reaction tubes, 5 μl of the reaction was removed and run on a 0.6% agarose/EtBr gel to determine whether digestion was complete. Digested DNA was phenol:chloroform purified according to manufacturer's protocol (Clontech Laboratories, Inc., Mountain View, CA, USA), and genome walker adaptors were ligated onto the digested DNA by the following reaction: 4 μl of digested, purified DNA combined with 1.9 μl genome walker adaptor (25 μM), 1.6 μl 10X ligation buffer, and 0.5 μl T4 DNA ligase (6 units/μl). The reaction was incubated overnight at 16°C.

Amplification of DNA Flanking Microsatellite Markers

The available sequences of microsatellites mTcCIR24, mTcCIR37, mTcCIR61, mTcCIR76, mTcCIR160, mTcCIR200, mTcCIR211, and mTcCIR222 were retrieved from the National Center for Biotechnology Information (NCBI) database. According to the published cacao composite map (Brown et al. 2007), these microsatellite markers flank black pod and witches' broom resistance QTL on linkage groups 1, 4, 8, 9, and 10 (Table 1). As per the GenomeWalker protocol, a pair of nested gene specific primers was designed to amplify upstream from the microsatellite repeat, whereas a second pair was also designed to amplify downstream of the repeat. Each set of nested primers was used in conjunction with a set of adaptor-specific primers to amplify the DNA surrounding each particular marker.

Nested gene-specific primers were designed for each microsatellite marker that was examined. All gene-specific primers were designed with Primer 3 version 0.4.0 software web interface (Rosen and Skaletsky 2000) and were between 26 and 30 bases long with GC content of 40%–60% (Table 2). Adaptor primer sequences were obtained from the GenomeWalker manual (Clontech Laboratories Inc., Mountain View, CA, USA). Nested PCR reactions were performed as per standard protocols (GenomeWalker manual), and PCR success was evaluated on a 0.6 % agarose gel.

TABLE 1 Microsatellite loci and their associated disease-resistance quantitative trait loci (QTL) regions identified in cacao and converted into single nucleotide polymorphism (SNP) markers

Converted SSR locus	Linkage group	QTL flanked	Number of SNPs identified	Evaluated SNP marker	SNP
mTcCIR24	1	WB	2	None	–
mTcCIR37	10	BP[a]	7	mTcCIR37s112	C/G
mTcCIR61	10	BP[a]	8	None	–
mTcCIR76	4	BP[a]	4	None	–
mTcCIR160	9	WB	6	mTcCIR160s384	A/G
mTcCIR200	8	BP[a]	5	None	–
mTcCIR211	8	BP[b]	5	mTcCIR211s1036	A/T
mTcCIR222	4	BP[a]	11	mTcCIR222s296	C/T
				mTcCIR222s316	G/T

(BP[a]) Black pod QTL identified by Brown et al. (2007) in F1progeny of 'Pound 7' × 'UF 273'; (BP[b]) black pod QTL identified by Brown et al. (2005) in F2 progeny of 'Sca6' × 'ICS1'; (WB) witches' broom QTL by Brown et al. (2005), in F2 progeny of 'Sca6' × 'ICS1'; (None) SNPs in these loci were monomorphic for the populations evaluated or not evaluated due to missing data in mapping population parents.

Sequencing of Amplified DNA Fragment

From the flanking regions of microsatellites mTcCIR24, mTcCIR37, mTcCIR61, mTcCIR76, mTcCIR160, mTcCIR200, mTcCIR211, and mTcCIR222, the amplified PCR products were sequenced. The additional sequence generated did not exceed 2,000 bases from the microsatellite repeat regions. New primers (Forward Genome Walker [FGW]; Reverse Genome Walker [RGW]) were designed from the newly added microsatellite flanking sequence (Table 2).

These primers were then used to amplify DNA of 'Pound 7', 'P 30', 'PA 7', and nine other cacao genotypes, each representing the major genetic groups of cacao as described by Motamayor et al. (2008) The amplified products were sequenced with an ABI 3730 genetic analyzer (Applied Biosystems, Foster City, CA, USA) and aligned with Phred, Phrap, Polyphred, and Consed software for sequence comparison and SNP detection (Ewing and Green 1998; Ewing et al. 1998; Gordon et al. 1998; Stephens et al. 2006). Each SNP site detected was named after the mTcCIR microsatellite marker (locus) from which it was identified, followed by the distance in nucleotides of the SNP from the 5' end of the sequence. All SNPs identified per locus are presented in Tables 3A and 3B. In the cases where there were too much missing data or the parents of the mapping populations were homozygous and identical, probes were not designed (Table 1, Table 3B). For the microsatellites converted, a total of 13 SNP probes and primer pairs were designed with (ABI) Primer Express version 3.0 and TaqMan MGB allelic discrimination software (Applied Biosystems, Foster City, CA, USA) using the consensus sequence of the respective locus. The probes were designed to be complementary to the SNP nucleotide position, whereas the primers amplified the fragment that contained the SNP (Table 4).

TABLE 2 List of designed sequencing and nested GenomeWalker primers and properties

PRIMER	SEQUENCE	%GC	°C TEMP	LENGTH
CIR_24_F_1	GTTCACCAAAAGACGAGACAGATCACAAAGTG	43.8	68.9	32
CIR_24_F_2	GAAAGGTAATGGCTTAACATTGTACACG	37.9	63.9	29
CIR_24_R_1	CCTCACATGCATATGAGAGACCTGAAGGAG	50.0	68.4	30
CIR_24_R_2 **(AJ748501)**	ATGGAAGGAGAAGAGTCAGAGGACATAAAG	43.3	62.2	30
CIR_37_F_1	ATT GGG TGT GGA ATC ATA TGG AAA CTC AGG	43.3	60.2	30
CIR_37_F_2	GAA CGA GAA CTG GTC CAC GTT TTA GAG	48.1	59.0	27
CIR_37_R_1	TTT CTC AGT CAC CAT CCA TGT TTC CTT TAC	40.0	58.7	30
CIR_37_R_2	CCA TTA GCT GAA CGA GCT CTA CTG CGT	51.9	61.9	27
CIR_37_FGW **(AJ748506)**	TCT CAT GGT CAC GAT GAA AGT GCG TGT GAA	46.7	63.0	30
CIR_61_RGW	GCTGGGAACAAACATGTCCCTGAGCGATT	51.7	64.1	29
CIR_61_FGW	AGT GTG CTT TGG TAC CGG GTC AGT TTT A	46.4	61.7	28
CIR_61_F_1	GTG GGA AAG AAA AGC AAA GAA ACT CTC CAC	43.3	59.8	30
CIR_61_F_2	AAT CTA CCT TCA TGT GCA GGC ACC TCA GGT	50.0	64.3	30
CIR_61_R_1	GAC TAT TTA TCT TGT TTC AGA CTG CCT GAA	36.7	56.8	30
CIR_61_R_2 **(AJ271959)**	CCT GAA ATT TGA ATT TGA GTG GCA ATG AGC	40.0	59.1	30
CIR_76_FGW	GTA TAC GTG AAA TTT GAT ACA GTG TGT TGA	33.3	55.6	30
CIR_76_F_1	CAC ACT CTT CCT CTT CGA AAA ATC TAG GGT	43.3	59.5	30
CIR_76_F_2	TCC CTC TTC GCC TCC CTC CTC TGA AAT TC	55.2	64.2	29
CIR_76_R_2	TTC TTC AAT GGA ATT ACG ATA ATG GGC TTG	36.7	57.5	30
CIR_76_R_1 **(AJ566422)**	GAA CGA GAA AGA GAT CTT AAA CGC GTC GTT	43.3	60.2	30
CIR_160_F_1	GTTGCTGTTGACATGTTTCAATGAAAGC	39.3	65.5	28
CIR_160_F_2	AAACTGCACACCCATCCGCACACCTTCA	53.6	74.6	28

(*Continued*)

TABLE 2 (Continued)

PRIMER	SEQUENCE	%GC	°C TEMP	LENGTH
CIR_160_R_1 **(AJ566490)**	GTGTACATGGTTACTTAGATGCACTTGACTG	41.9	63.1	31
CIR_200_R_1	TAGAAAGAGCCCAATCAGGGTAAGATAC	42.0	57.6	28
CIR_200_R_2	GGCTAGCAACTCAGTGCGTGGATTCAA	51.9	62.7	27
CIR_200_FGW **(AJ566526)**	TGA TTG TAG GAT ATG GTT TAG TGA TTA GGC	36.7	56.3	30
CIR_211_F_1	CTT CTT GTT TGG CCT CTG CAT AGA TCA TCT	43.3	60.1	30
CIR_211_F_2	CTC TCT CAC TCT CTC TCA CAT TCT TTT TTG	40.0	57.0	30
CIR_211_R_1	AGA GGA GCC TCT AAG ATT TGA GTT AGC ACC	46.7	60.8	30
CIR_211_R_2	GAA GTG CAA TCC CTC GCT CAA CTG TAG G	53.6	62.4	28
CIR_211_FGW **(AJ566534)**	ACC GCC TGG CGA TGG CTA TCC CTG ACA AG	62.1	68.3	29
CIR_222_F_1	GAG AGA GAA GCA AAA ATG TGC AAC CCC TTA	43.3	60.6	30
CIR_222_F_2	TGT AGC AGT GCC TTC AAC ATA CTC TGT TGC	46.7	61.9	30
CIR_222_RGW **(AJ566543)**	TTG TTC GTC CTA TGG TAT CAA CAC TGT TGG	43.3	60.0	30

F_1, F_2, R_1, and R_2 represent the nested forward and reverse primers designed to amplify the published sequence of the mTcCIR loci; FGW and RGW represent the primers designed to amplify the added sequence of the locus after genome walking; AJxxxxxx are the NCBI accession numbers of the published sequence of the loci used.

TABLE 3A SNP positions identified for each converted microsatellite locus: mTcCIR37, mTcCIR160, mTcCIR211, and mTcCIR222

Converted Marker	mTcCIR37							mTcCIR160					
	SNP Position							SNP Position					
	11	14	15	16	20	21	23	38	39	56	58	62	66
Clone Name													
'Pound 7'	C/G	T	C	G	G	G	C	A	G	T	C	G	G
'Pa7'	G	T	C	G	G	G	C	G	G	T	A	A	A
'P30'	G	T	C	G	G	G	C	A	G	T	C	G	G
SCA 12	G	T	C/G	G	G	A/G	C/T	A	G	T	A	A	A
COC 3335	G	T	C	G/T	G	G	C/T	A	G	C	A	A	A
UF273 type1	G	T	C	G/T	G	G	C/T	A	G	T	A/C	A/G	A/G
GU124A	G	T	C	G	G	G	C	G	G	T	A	A	A
LasBrisas1717	G	T	C	G/T	G	G	T	A	G	T	A	A	A
BE4 PL3	G	T	C	G	G	G	C	A	G	T	C	G	G
CAB 0331 PL4	G	T	C	G	G	G	C	A	C/G	T	A/C	A/G	A/G
THS516	G	T	G	G	G	A	C	A	G	T	A	A	A
PERU	G	G/T	C/G	G	A/G	A/G	C	A	G	T	A	A	A

Converted Marker	mTcCIR211					mTcCIR222									
	SNP Position					SNP Position									
	30	44	49	97	10	29	31	51	54	56	57	59	61	62	65	72
Clone Name																
'Pound 7'	T	G	C	T	A/T	T	T	A	C	A	C	C	T	C/T	A	C
'Pa7'	T	G	C	T	T	C/T	G/T	A	C	A/C	C	C	T	T	A	C
'P30'	T	G	C	T	A	T	G	A	C	A	C	C	T	T	A	C
SCA 12	A	G	C	T	T	T	T	A	C	A	C	C	T	T	A	C
COC 3335	T	G/T	C/T	T	T	T	T	A/G	C	A	T	C	T	T	A	C
UF273 type1	T	G	C	T	A/T	T	G/T	A/G	C	A	C/T	C	T	T	A	C
GU124A	T	G	C	–	–	T	G	A	C	A	C	C	T	T	A	C
LasBrisas1717	T	G	C	T	T	T	T	G	C	A	T	C	T	T	A	C
BE4 PL3	T	G	C	T	A	T	G	A	C	A	C	C	T	T	A	C
CAB 0331 PL4	T	G	C	T	T	T	T	A	C	A/C	C	C	T	T	A	C
THS516	A/T	G	C	T	A/T	T	G/T	A	C	A	C/T	C	T	T	A	C/T
PERU	A	G	C	T	T	T	T	A	C/T	A	C/T	C	A/T	T	A	C

Single letters represent a homozygous allele for the nucleotide listed. Nucleotide pairs divided by a slash (/) represent heterozygous alleles for the nucleotides listed. Dashes (-) represent missing data. Clones in this table are representatives of commonly used breeding parents including 'Pound 7', 'PA 7', and 'P30'. SNP position is calculated as number of nucleotides from 5' end of amplified expanded microsatellite locus.

TABLE 3B SNP positions identified for each non-converted microsatellite locus: mTcCIR24, mTcCIR61, mTcCIR76, and mTcCIR200

Converted Marker	mTcCIR24		mTcCIR61							
	SNP Position		SNP Position							
	767	902	33	50	53	10	10	11	11	12
Clone Name										
'Pound 7'	G	A	A	A	G	C	A	C	T	G
'Pa7'	–	–	–	–	–	–	–	–	–	–
'P30'	G	A	A	A	G	C	A	C	T	G
SCA 12	T	A	A	–	–	C	A	C	T	G
COC 3335	–	–	–	–	–	–	–	–	–	–
UF273 type1	G	A	A/G	A/G	C/G	C	A/G	C/T	C/T	G
GU124A	G	A	A	A	G	C	A	C	T	G
LasBrisas1717	G	A	G	A/G	C	–	–	–	–	–
BE4 PL3	G	A	A	A	G	C	A	C	T	G
CAB 0331 PL4	–	–	A	A	G	C	A	C	T	G
THS516	–	–	A	A	C/G	–	A	C	–	–
PERU	G	G	A	A	C	–	A	C	T	–

Converted Marker	mTcCIR76				mTcCIR200				
	SNP Position				SNP Position				
	13	15	15	17	47	50	51	58	60
Clone Name									
'Pound 7'	T	T	T	C	T	C	A	A	G
'Pa7'	–	–	–	–	–	–	–	–	–
'P30'	–	–	–	–	T	C	A	A	G
SCA 12	–	–	–	–	–	–	–	–	–
COC 3335	–	–	–	–	–	–	–	–	–
UF273 type1	–	–	–	C	T	C	A	A	G
GU124A	T	T	T	C	C/T	C/T	A/T	A/G	G
LasBrisas1717	–	–	–	–	–	–	–	–	–
BE4 PL3	–	–	–	–	–	–	–	–	–
CAB 0331 PL4	T	T	T	C	T	C	–	–	–
THS516	C	C	C	C	–	–	–	–	–
PERU	T	T	T	T	–	–	–	–	–

Single letters represent a homozygous allele for the nucleotide listed. Nucleotide pairs divided by a slash (/) represent heterozygous alleles for the nucleotides listed. Dashes (-) represent missing data. Clones in this table are representatives of commonly used breeding parents including 'Pound 7', 'PA 7', and 'P30'. SNP position is calculated as number of nucleotides from 5' end of amplified expanded microsatellite locus.

The designed primers and probes were used in a TaqMan-based SNP assay (Applied Biosystems, Foster City, CA, USA) with DNA from progeny of 'Pound 7' × 'P 30' and progeny of 'PA 7' × 'P 30'. All SNP assay reactions were performed as described in Livingstone et al. (2010).

TABLE 4 List of designed SNP assay primers and probes

SNP Marker	PRIMER/PROBE	% GC	°C TEMP	LENGTH
mTcCIR37s112 (AJ748506)				
mTcCIR37s112_FW	AAAGTGCGTGTGAAGAGTTCCTATC	44	58.9	25
mTcCIR37s112_RV	GCATGGAACGATCCAAGTTAGTC	48	58.6	23
mTcCIR37s112_pG	**(FAM)** ATAATGGAAGAgAACTTGGT	35	67	20
mTcCIR37s112_pC	**(VIC)** ATAATGGAAGAcAACTTG	33	67	18
mTcCIR37s153 (AJ748506)				
mTcCIR37s153_FW	AAAGTGCGTGTGAAGAGTTCCTATC	44	58.9	25
mTcCIR37s153_RV	CTTTATAACCCTTAGGCCATGCA	43	58.3	23
mTcCIR37s153_pG	**(FAM)** ACTAACTTgGATCGTTCCA	42	67	19
mTcCIR37s153_PA	**(VIC)** ACTAACTTcGATCGTTCC	39	65	18
mTcCIR37s214 (AJ748506)				
mTcCIR37s214_FW	TGATGATGGGGATATGATTTTAGAAGA	33	59.9	27
mTcCIR37s214_RV	TTACAATGAGGTCAGCCCTTACC	48	58.2	23
mTcCIR37s214_pG	**(FAM)** TGTGGTTAATAgTGAGATCA	35	67	20
mTcCIR37s214_PA	**(VIC)** TGTGGTTAATAaTGAGATC	32	65	19
mTcCIR37s236 (AJ748506)				
mTcCIR37s236_FW	TGATGATGGGGATATGATTTTAGAAGA	33	59.9	27
mTcCIR37s236_RV	TCTCAGTCACCATCCATGTTTCC	48	59.8	23
mTcCIR37s236_pC	**(FAM)** AGTGGTAAGGGcTGAC	56	66	16
mTcCIR37s236_pT	**(VIC)** AAGTGGTAAGGGtTGAC	41	65	17
mTcCIR160s384 (AJ566490)				
mTcCIR160s384_FW	ATGATGGTGACAACAGCAAGAAA	39	58.2	23
mTcCIR160s384_RV	ATGCCTATTAATCACCTAGGTGAGACT	41	58	27
mTcCIR160s384_pA	**(FAM)** CAAGGATCaTTTTTGCT	35	66	17
mTcCIR160s384_pG	**(VIC)** AAGGATCgTTTTTGCT	38	65	16

(*Continued*)

TABLE 4 (Continued)

SNP Marker	PRIMER/PROBE	% GC	°C TEMP	LENGTH
mTcCIR160s399 (AJ566490)				
mTcCIR160s399_FW	ATGATGGTGACAACAGCAAGAAA	39	58.2	23
mTcCIR160s399_RV	CTAGGTGAGACTGACATATGTAAGTAATAGATAGAC	36	58.5	36
mTcCIR160s399_pG	**(FAM)** TGCTCTTATTgCATATC	35	65	17
mTcCIR160s399_pC	(**VIC**) TGCTCTTATTcCATATCT	33	67	18
mTcCIR211s305 (AJ566534)				
mTcCIR211s305_FW	TGGAAATGTCTCAAATTTTGGCTT	35	58.6	23
mTcCIR211s305_RV	CGAGCTTTCTTTATGCTATATCTCTTTCT	34	58.5	29
mTcCIR211s305_pT	**(FAM)** CTCACTTtCGGATGAT	44	65	16
mTcCIR211s305_pA	**(VIC)** CTCACTTaCGGATGATGA	44	67	18
mTcCIR211s493 (AJ566534)				
mTcCIR211s493_FW	TGGAAATGTCTCAATTTTGGCTT	35	58.6	23
mTcCIR211s493_RV	TCCATGAAGCATGACAAATAGCA	39	59.2	23
mTcCIR211s493_pC	**(FAM)** CTTCTCCcTTTAACCTT	41	67	17
mTcCIR211s493_pT	**(VIC)** TCTCCtTTTAACCTTTG	35	65	17
mTcCIR211s1036 (AJ566534)				
mTcCIR211s1036_FW	ACCTTAATTTTATGGGAAACGAGGT	36	58.4	25
mTcCIR211s1036_RV	CCAAACAAAATCTTATTGCACTGTG	36	58.7	25
mTcCIR211s1036_pT	**(FAM)** AATCtGTGCTGACTGAT	41	67	17
mTcCIR211s1036_pA	**(VIC)** CAATCaGTGCTGACTG	50	65	16
mTcCIR222s296 (AJ566543)				
mTcCIR222s296_FW	AGCAGTGCCTTCAACATACTCTGT	46	58.2	24
mTcCIR222s296_RV	CCAGTTGGCTCAAAAGTTTGG	48	58.3	21
mTcCIR222s2s96_pT	**(FAM)** CATTCCGGTtTTAAGCT	41	67	17
mTcCIR222s296_pC	**(VIC)** TCCATTCCGGTcTTAA	44	67	16

mTcCIR222s316 (AJ566543)				
mTcCIR222s316_FW	AGCAGTGCCTTCAACATACTCTGT	46	58.2	24
mTcCIR222s316_RV	CCAGTTGGCTCAAAAGTTTGG	48	58.3	21
mTcCIR222s316_pG	**(FAM)** CACTTTgCCAAGAGA	47	66	15
mTcCIR222s316_pT	**(VIC)** CCACTTTtCCAAGAGA	44	66	16
mTcCIR222s565 (AJ566543)				
mTcCIR222s565_FW	CTAGTGCAGCAGCCATAGTCCC	59	59.9	22
mTcCIR222s565_RV	GTCAACTCGATTGCTGGAAGTAAG	46	58.2	24
mTcCIR222s565_pA	**(FAM)** CGTCTAAAaCTATTGGAG	39	65	18
mTcCIR222s565_PC	**(VIC)** CGTCTAAAcCTATTGGAGT	42	66	19
mTcCIR222s628 (AJ566543)				
mTcCIR222s628_FW	CTCTTAACCAAGCAGCAACTGC	50	58	22
mTcCIR222s628_RV	CCTTGCTCACCCAAGTTACCTG	55	58.2	22
mTcCIR222s628_pC	**(FAM)** CTCCATATcGTGAATCA	41	66	17
mTcCIR222s628_pT	**(VIC)** TCTCCATATtGTGAATCA	33	65	18

AJxxxxxx: NCBI accession numbers for the mTcCIR locus sequence converted to SNPs. Probe sequences are labeled as either FAM or VIC depending on the fluorescent tag molecule attached to the probe. SNP variants are presented as lowercase letters in probe sequence.

RESULTS

DNA and Microsatellite Analysis

Six highly polymorphic microsatellite markers (mTcCIR12, mTcCIR21, mTcCIR42, mTcCIR87, mTcCIR128, and mTcCIR141) were used for off-type detection in the progeny of the mapping populations. These markers provided a unique allelic profile of all individual plants tested. Allele calls of progeny with common parents were consistent, except for off-types. Off-types are individual samples whose genotype or allelic configuration does not conform to the allelic configuration of their supposed parents. Individuals were considered off-types if detected by at least two microsatellite markers (Tables 5A and B). Out of 85 total individuals in the 'Pound 7' × 'P 30' cross, five were identified as off-types; four of these samples contained alleles that were not present in either parent (Table 5A). However, alleles from off-type Tc11581 suggested that it had self-pollinated because it possessed only maternal alleles yet the allelic configuration differed from the mother, eliminating the possibility of a mislabeled maternal clone.

For the 'PA 7' × 'P 30' cross, nine off-types were identified out of 105 samples (Table 5B). Eight of the off-type samples contained alleles that were not present in either of the reported parents. The remaining off-type sample, Tc11680, contained alleles only found in the maternal parent, suggesting a possible self-pollination event. Off-types from both populations were removed, leaving 80 individuals in the 'Pound 7' × 'P 30' cross and 96 individuals in the 'PA 7' × 'P 30' cross for further studies.

Genome Walking and SNP Identification

Eight important microsatellite markers were converted into SNP markers using genome walking and a diversity panel to obtain sequence data and identify SNPs. These loci and the QTL they flank can be seen in Table 1. Primers were designed to expand the flanking regions around these microsatellite loci (Table 2). These primers were used to amplify a panel of genetically diverse cacao cultivars, and SNPs were identified from the sequenced amplicons (Tables 3A and 3B).

Several SNPs were identified along the entire length of the newly generated sequence for each locus. The SNP position in Tables 3A and 3B is determined by distance from the 5' end of the sequenced amplicon. In all, 48 SNP markers were identified across the diversity panel in the flanking sequences of the eight microsatellite loci expanded by genome walking (Tables 3A and 3B). Primers and probes for TaqMan assays were designed for a subset of 13 SNP markers (Table 4) from the converted mTcCIR37, mTcCIR160, mTcCIR211, and mTcCIR222 loci. The four remaining microsatellites (Table 3B) showed no evidence of polymorphism in the parents of the populations being evaluated ('Pound 7', 'PA 7', and 'P 30'). Therefore, none of

TABLE 5A Microsatellite off-typing of the 'PA 7' × 'P30' population

SSR marker	mTcCIR12		mTcCIR21		mTcCIR42		mTcCIR87		mTcCIR128		mTcCIR141	
Parents	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2
'Pound7' (female)	199	211	149	153	231	231	130	134	135	135	211	215
'P30' (male)	250	250	153	153	231	231	144	144	151	151	219	219
Progeny												
TC11548 (O)	**187**	**203**	149	153	231	**234**	130	144	135	151	211	219
TC11562 (O)	**201**	250	153	**155**	231	**234**	130	144	135	151	215	219
TC11567 (O)	199	**211**	153	**163**	**225**	231	**124**	134	135	**135**	211	**212**
TC11572 (O)	**187**	211	149	153	**201**	231	**124**	134	135	151	**209**	211
TC11581 (O)	199	**211**	149	153	**231**	231	134	**134**	135	**135**	211	**211**
TC11554 (T)	199	250	153	153	231	231	130	144	135	151	215	219
TC11555 (T)	199	250	149	153	231	231	134	144	135	151	215	219
TC11558 (T)	211	250	149	153	231	231	134	144	135	151	211	219

TABLE 5B Microsatellite off-typing of the 'Pound 7' × 'P30' population

SSR marker	mTcCIR12		mTcCIR21		mTcCIR42		mTcCIR87		mTcCIR128		mTcCIR141	
Parents	Allele1	Allele2	Allele1	Allele2	Allele1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele1	Allele 2
PA 7' (female)	199	211	153	159	213	225	124	124	135	135	209	209
P 30' (male)	250	250	153	153	231	231	144	144	151	151	219	219
Progeny												
TC11633 (O)	211	**219**	153	153	231	**234**	**140**	144	**143**	151	209	**209**
TC11635 (O)	**218**	250	**149**	**149**	213	231	124	144	135	151	209	219
TC11662 (O)	**187**	**219**	153	**163**	**201**	231	**131**	–	**130**	**143**	209	**209**
TC11663 (O)	**203**	211	153	**163**	**201**	**234**	**140**	144	**130**	151	209	**209**
TC11664 (O)	**187**	**219**	153	**163**	**201**	**234**	**140**	144	**130**	**143**	209	**215**
TC11665 (O)	–	–	153	**163**	**201**	**234**	**140**	144	151	**151**	209	**209**
TC11672 (O)	**187**	**219**	–	–	**157**	**164**	**231**	**235**	**184**	**186**	**140**	**144**
TC11680 (O)	199	**199**	153	159	213	**213**	124	**124**	135	**135**	209	**209**
TC11709 (O)	**214**	**219**	153	**163**	**201**	231	**140**	144	**143**	151	209	**209**
TC11628 (T)	211	250	153	153	213	231	124	144	135	151	209	219
TC11658 (T)	211	250	153	159	213	231	124	144	135	151	209	219
TC11661 (T)	199	250	153	159	225	231	124	144	135	151	209	219

(O) off-type samples with non-parental alleles in bold type and underlined; (T) samples representative of true-to-type individuals.

the SNPs identified in mTcCIR24, mTcCIR61, mTcCIR76, and mTcCIR200 was converted into TaqMan assays. However, the SNP identification data have been included (Table 3B) as SNP assays useful in other mapping populations could be designed for these converted markers.

Five SNP markers (mTcCIR37s112, mTcCIR160s384, mTcCIR211s1036, mTcCIR222s296, and mTcCIR222s316) out of the 13 were used to fingerprint all the progenies of 'Pound 7' × 'P 30' and 'PA 7' × 'P 30' to assess their discrimination potential (Tables 6A and 6B). In both of these populations, off-types previously identified by microsatellites (Tables 5A and 5B) were also identified by at least one of the five converted SNP markers (Tables 6A and 6B). In the 'Pound 7' × 'PA 30' cross, two of the converted SNP markers (mTcCIR160s384 and mTcCIR222s296) were monomorphic and did not identify any off-types. In contrast, mTcCIR37s112 was monomorphic in the 'PA 7' × 'P 30' cross (Table 6B), but it was able to identify an off-type.

Microsatellite analysis suggested possible selfing events in each of the populations in samples TC11581 (Table 5A) and TC11680 (Table 5B). The SNP marker data for TC11581 are consistent with that hypothesis, as all the genotypes could have been generated by self-pollination of the maternal parent (Table 6A). However, for TC11680, mTcCIR37s112 showed an allelic configuration that would not be possible by self-pollination of the mother (Table 6B). Thus, in this case, the biallelic SNP genotype provides more information than the more highly polymorphic microsatellite genotype.

Because the SNPs were within 2,000 nucleotides of the microsatellite region, it is unlikely that recombination occurred between the SNP and the microsatellite. Thus, the SNP haplotypes for the microsatellite alleles for the two populations could be used to analyze whether the inheritance of a particular size allele was by descent or an example of size homoplasy of the amplified microsatellite region (Estoup et al. 2002). There were three microsatellite loci where this could be determined, mTcCIR37, mTcCIR 211, and mTcCIR222, as both populations had been genotyped with these microsatellite markers; mTcCIR160 had not been used to genotype the two populations. By comparing the SNP haplotype to the microsatellite genotype, a nucleotide haplotype could be assigned to each microsatellite allele (Tables 7A and 7B). For mTcCIR37, the nucleotide for microsatellite allele 139 was "C" in the 'Pound7' parent (Table 7B) and "G" in the 'PA7' parent (Table 7A), suggesting size homoplasy rather than identity by descent as the SNP was only 112 nucleotides from the microsatellite region. For mTcCIR211, all three parents were homozygous 182/182, but the 182 alleles differed between the 'PA 7' parent ("T") and 'P 30' ("A") (Table 7A), whereas 'Pound 7' (Table 7B) was heterozygous ("T/A") for the SNP that was 1,036 nucleotides from the microsatellite region. For mTcCIR222, two SNPs were available. 'PA7' and 'Pound7' shared the 220 allele, which had the identical haplotype ("TT") in both populations. This was the only evidence for identity by descent in the microsatellite loci that we could analyze.

TABLE 6A SNP off-typing in 'Pound 7' × 'P30' population

SNP markers	mTcCIR37s112	mTcCIR160s384	mTcCIR211s1036	mTcCIR222s296	mTcCIR222s316
Parents					
'Pound 7' (female)	C/G (139/146)	A	A/T (182/182)	T (218/220)	T (218/220)
'P 30' (male)	G (164/164)	A	A (182/182)	T (216/216)	G (216/216)
Progeny					
TC11548 (O)	***C***	A	A/T	T	G/T
TC11562 (O)	***C***	A	***T***	T	G/T
TC11567 (O)	C/G	–	***T***	T	***T***
TC11572 (O)	***C***	–	A/T	T	G/T
TC11581 (O)	C/G	A	***T***	T	***T***
TC11554 (T)	G	A	A/T	T	G/T
TC11555 (T)	G	A	A/T	T	G/T
TC11558 (T)	C/G	–	A	T	G/T

TABLE 6B SNP off-typing in 'Pound 7' × 'P30' population

SNP markers	mTcCIR37s112	mTcCIRl60s384	mTcCIR211s1036	mTcCIR222s296	mTcCIR222s316
Parents					
'PA7'(mother)	G (139/144)	G	T (182/182)	C/T (218/220)	G/T (218/220)
'P30' (father)	G (164/164)	A	A (182/182)	T (216/216)	G (216/216)
Progeny					
TC11633 (O)	G	***A***	A/T	***C***	G
TC11635 (O)	G	A/G	***T***	T	G
TC11662 (O)	G	***A***	A/T	***C***	G
TC11663 (O)	G	***A***	A/T	***C***	G
TC11664 (O)	G	***A***	A/T	***C***	G
TC11665 (O)	G	***A***	A/T	***C***	G
TC11672 (O)	G	***A***	A/T	***C***	G/T
TC11680 (O)	***C/G***	***G***	***T***	T	G/T
TC11709 (O)	G	***A***	A/T	T	G
TC11628 (T)	G	A/G	A/T	T	G/T
TC11658 (T)	G	A/G	A/T	C/T	G
TC11661 (T)	G	A/G	A/T	C/T	G

(O) off-type samples with non-parental alleles in bold type and underlined; (T) samples representative of true-to-type individuals.
Microsatellite genotype in parentheses.

TABLE 7A SNP haplotypes of microsatellite alleles in 'PA7' × 'P30' population

SNP markers							
Sample	mTcCIR37 genotype	mTcCIR37 s112	mTcCIR211 genotype	mTcCIR211 s1036	mTcCIR222 genotype	mTcCIR222 s296	mTcCIR222 s316
Parents							
'PA7' (female)	139/144	G	182/182	T	220/224	C/T	G/T
	139 allele	G	182 allele	T	220 allele	T	T
	144 allele	G	182 allele	A	224 allele	C	G
'P30' (male)	164/164	G	182/182	A	216/216	T	G
Progeny							
TC11622	144/164	G	182/182	A/T	216/220	T	G/T
TC11627	144/164	G	182/182	A/T	216/220	T	G/T
TC11628	139/164	G	182/182	A/T	216/220	T	G/T
TC11658	139/164	G	182/182	A/T	216/224	C/T	G
TC11659	139/164	G	182/182	A/T	216/224	C/T	G
TC11660	139/164	G	182/182	A/T	216/220	T	G/T
TC11661	139/164	G	182/182	A/T	216/224	C/T	G
TC11666	139/164	G	182/182	A/T	216/220	T	G/T
TC11667	139/164	G	182/182	A/T	216/224	C/T	G
TC11673	139/164	G	182/182	A/T	216/220	T	G/T
TC11674	144/164	G	182/182	A/T	216/220	T	G/T

TABLE 7B SNP haplotypes of microsatellite alleles in 'Pound7' × 'P30' population

Sample	mTcCIR37 genotype	mTcCIR37 s112	mTcCIR211 genotype	mTcCIR211 s1036	mTcCIR222 genotype	mTcCIR222 s296	mTcCIR222 s316
Parents							
'Pound7' (female)	139/146	C/G	182/182	A/T	218/220	T	T
	139 allele	C	182 allele	T	218 allele	T	T
	146 allele	G	182 allele	A	220 allele	T	T
'P30' (male)	164/164	G	182/182	A	216/216	T	G
Progeny							
TC11535	146/164	G	182/182	A/T	216/220	T	G/T
TC11536	146/164	G	182/182	A/T	216/220	T	G/T
TC11541	139/164	C/G	182/182	A	216/220	T	G/T
TC11554	146/164	G	182/182	A/T	216/220	T	G/T
TC11555	146/164	G	182/182	A/T	216/220	T	G/T
TC11556	146/164	G	182/182	A/T	216/220	T	G/T
TC11557	146/164	G	182/182	A/T	216/218	T	G/T
TC11558	139/164	C/G	182/182	A	216/220	T	G/T
TC11559	146/164	G	182/182	A/T	216/220	T	G/T
TC11560	146/164	G	182/182	A/T	216/220	T	G/T
TC11561	146/164	G	182/182	A/T	216/218	T	G/T

Single letters represent a homozygous allele for the nucleotide listed. Nucleotide pairs divided by a slash (/) represent heterozygous alleles for the nucleotides listed. Microsatellite genotypes represented by fragment length of each allele in nucleotides are separated by a forward slash (/).

DISCUSSION

In cacao, a number of QTL have been identified using microsatellite markers that could be used for MAS (Clement et al. 2003a, b; Risterucci et al. 2003; Brown et al. 2007; Schnell et al. 2007; Brown et al. 2008; Lanaud et al. 2009). The assay of microsatellite markers in cocoa-producing countries such as Ghana has been difficult because of problems in obtaining access to high-resolution microsatellite assay platforms and maintaining a reliable supply of reagents and supplies for those platforms. Previously these problems were circumvented through collaborations with laboratories in Europe or the United States for genotyping large populations (Pugh et al. 2004; Takrama et al. 2005). Despite the success of these collaborations, accessibility to and firsthand knowledge of the germplasm material at a breeding site would be beneficial to a MAS program. To this end, the development of a genotyping system in cacao-breeding regions is preferred over international collaborations where samples may change hands and labels several times. To overcome the difficulties associated with genotyping in cacao-growing regions, a SNP assay based on the 5' nuclease TaqMan assay was developed (Livingstone et al. 2012) and is now being used at CRIG to determine off-types in clonal germplasm collections (Takrama et al. 2012). The SNP markers have the following advantages over microsatellite markers: They can be assayed with great specificity on a variety of platforms that are less expensive than those used to assay microsatellites. They are unambiguous (the outcome is either A, C, G, T, not an amplified fragment length) on any assay platform unlike microsatellite genotypes, which differ in reported length from platform to platform (platform-dependent). Unambiguous SNP genotypes can be shared between labs and in international databases, unlike platform-dependent microsatellite genotypes. However, to be able to take advantage of the QTL identified with microsatellite markers, a method to convert them to SNP markers became necessary. The GenomeWalker protocol provided the requisite platform to facilitate this conversion. In this study, eight microsatellite markers that flank important QTL that could not be easily assayed in Ghana were successfully converted to SNP markers for the TaqMan assay. The rate of marker conversion is increased by the availability of longer regions of flanking sequence surrounding the microsatellites. These longer flanking sequences allow for better primer design for both genome walking and sequencing across a diversity panel. With the recent release of the cacao genome by an international consortium (Argout et al. 2011) and a collaboration of USDA-ARS, Mars Inc., and IBM (http://www.cacaogenomedb.org), the need for genome walking becomes less as sequencing primers can now be designed to regions near known microsatellites. However, for those working on crops without a known genome sequence, genome walking out from known markers provides a simple and successful way to convert microsatellites to SNPs.

Additionally, because the microsatellite markers converted were associated with QTL of desired traits, these SNP markers may be useful for MAS. While the association of the SNP markers with QTL traits is currently undergoing evaluation, the ability to assay markers within QTL at cacao breeding sites will be an advantage to cacao MAS programs worldwide. An added benefit of marker conversion is the ability to evaluate loci in populations that were not previously informative. For example, although associated with a black-pod QTL (Brown 2005), mTcCIR211 was monomorphic, and thus uninformative in both of the populations utilized in this study. However, the SNP marker identified adjacent to this SSR locus (mTcCIR211s1036) was polymorphic in the 'Pound 7' × 'P 30' population, thereby providing the cacao breeder with an informative marker that is linked to the black-pod QTL where previously none existed.

The determination of off-types in cacao-breeding programs remains a serious problem. In this study, some of the individuals tested were detected as off-types using microsatellite and SNP markers. Genetic fingerprinting using six highly polymorphic microsatellite markers revealed that five individuals (5.9%) were off-types in the 'Pound 7' × 'P 30' cross and nine (8.6%) in the 'PA 7' × 'P 30' cross. When using the converted SNP markers, the same off-types were also identified with at least one SNP marker. The biallelic nature of SNP markers may require use of more markers to obtain the same discrimination power as microsatellites, and it is highly likely that five SNP markers are not sufficient to differentiate all off-types. Because a number of the off-types were only identified by a single SNP, more SNP markers should be employed to increase the reliability of off-typing with SNPs. However, because the microsatellites used for off-typing were selected for their high polymorphism across many cultivars (Saunders et al. 2004), it is of interest to note that the five SNP markers were able to identify all the off-types identified by the more polymorphic microsatellites. With the availability of the complete cacao genome and more highly saturated genetic maps, core subsets of SNPs are being developed to provide a greater number of SNP markers that are evenly distributed across the cacao genome and can be used for both off-typing and MAS.

Among the off-types identified using microsatellite markers, seven possessed at least one maternal allele, which suggests that pollen from different paternal parents might have been used to pollinate the maternal tree instead of 'P 30' pollen, or the pollen used was contaminated with pollen from a different source. It was also observed that six other off-types detected possessed only the paternal allele, suggesting that the male pollen was used to pollinate entirely different maternal trees probably intended for a different cross, but that had been added to the seedlings in the nursery. The remaining off-type (Tc11672) did not have any allele in common with either parent. This observation seems to confirm an earlier claim of seedling mix-up in the nursery, because all pollinated pods were raised in a central nursery

before transplanting. These events point out the critical need for a convenient genotyping assay, such as SNPs, that can be performed locally and prior to planting of seedlings in the field.

One off-type (Tc11581) may be the result of a maternal selfing, as all the alleles recorded were common to the mother (Table 1 and 7). This is interesting because Upper Amazon collections, such as Pound 7, are known to be self-incompatible and, in the tree nursery, the self-incompatible genotype is always used as the maternal parent to avoid the possibility of self-pollination. A further compatibility study on this individual will be very useful in explaining this observation.

The detection of off-types in the progeny may explain observations made by Adomako (2006), who established that progeny performance for resistance to black pod disease could not be reliably predicted from the parental disease reaction. Presence of off-types could have masked the true potentials of the progeny when their performance was compared with a different set of progeny performance. Crouizillat et al. (2000) and Cervantes-Martinez et al. (2006) could reliably predict progeny performance from parental disease reaction. It is possible that they were dealing with fairly homozygous populations. It is, therefore, recommended that seedlings should be fingerprinted before planting out in experimental trials.

Another possibility that would explain the inconsistent association of particular microsatellite alleles with favorable traits is that, for some loci, microsatellite alleles are not identical by descent. Although we only analyzed three of the converted microsatellite loci with our SNP and sequence data, we were able to demonstrate that, for two of them (mTcCIR37 and mTcCIR211), alleles that were the same length did not have identical sequences in the flanking regions. With the availability of the complete genome sequence for at least two cacao individuals, more microsatellite loci can be studied. It was surprising that two of the three loci showed size homoplasy of alleles, which is not commonly observed at that rate (Estoup et al. 2002; Schroeder et al. 2009). Thus, even though SNPs are only biallelic, they can provide a greater amount of information about the source of the favorable allele (i.e., from which lineage) than the more polymorphic microsatellite markers.

This study has shown the successful application of genome walking to convert microsatellite markers into SNP markers. This is especially useful in orphan crops where little genome sequence is available. The conversion to SNP markers not only supplies the cacao breeder with additional markers linked to QTL, but also allows for the ability to use these markers at breeding centers, reducing the need to outsource genotyping. This will give the breeder more control over the evaluation of a given cross and reduce the time needed to apply the results of those evaluations as samples that will not need to be shipped internationally. The ability to convert well-established microsatellite markers into SNPs should help serve cacao MAS breeding efforts worldwide. The findings have also shown that with a simple

fluorescence microplate reader, SNP markers have provided the needed tool for quick screening of genotypes in breeding programs in a cocoa-producing country. Scientists at CRIG currently run these SNP assays to confirm clonal materials in the gene bank.

REFERENCES

Adomako, B. 2006. Combining ability analysis of blackpod disease incidence in cocoa genotypes in Ghana. *Trop. Sci*. 46 (4): 201–204. DOI:10.1002/ts.176.

Argout, X., J. Salse, J. M. Aury, M. J. Guiltinan, G. Droc, J. Gouzy, M. Allegre, et al. 2011. The genome of *Theobroma cacao*. *Nat. Genet*. 43(2): 101–108. DOI:10.1038/ng.736.

Brown, J. S., W. Phillips-Mora, E. J. Power, C. Krol, C. Cervantes-Martinez, J. C. Motamayor, and R. J. Schnell. 2007. Mapping QTLs for resistance to frosty pod and black pod diseases and horticultural traits in *Theobroma cacao* L. *Crop Sci*. 47(5): 1,851–1,858. DOI:10.2135/cropsci2006.11.0753.

Brown, J. S., R. Sautter, C. Olano, J. Borrone, D. Kuhn, J. Motamayor, and R. Schnell. 2008. A composite linkage map from three crosses between commercial clones of cacao, *Theobroma cacao* L. *Trop. Plant Biol*. 1(2): 120–130.

Cervantes-Martinez, C., J. S. Brown, R. J. Schnell, W. Phillips-Mora, J. F. Takrama, and J. C. Motamayor. 2006. Combining ability for disease resistance, yield, and horticultural traits of cacao (*Theobroma cacao* L.) clones. *J. Am. Soc. Hort. Sci*. 131(2): 231–241.

Cheesman, E. 1944. Notes on the nomenclature, classification and possible relationships of cacao populations. *Trop. Agric*. 21:144–159.

Clement, D., A. M. Risterucci, J. C. Motamayor, J. N'goran, and C. Lanaud. 2003a. Mapping QTL for yield components, vigor, and resistance to *Phytophthora palmivora* in *Theobroma cacao* L. *Genome* 46(2): 204–212. DOI:10.1139/g02–125.

Clement, D., A. M. Risterucci, J. C. Motamayor, J. N'goran, and C. Lanaud. 2003b. Mapping quantitative trait loci for bean traits and ovule number in *Theobroma cacao* L. *Genome* 46(1): 103–111. DOI:10.1139/g02–118

Collard, B.C., and D. J. Mackill. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. Royal Soc. London Ser. B*. 363(1,491): 557–572. DOI:10.1098/rstb.2007.2170

Crouzillat, D., W. Phillips, P. Fritz, and V. Pétiard. 2000. Quantitative trait loci analysis in *Theobroma cacao* using molecular markers. Inheritance of polygenic resistance to *Phytophthora palmivora* in two related cacao populations. *Euphytica* 114(1): 25–36. DOI:10.1023/a:1003994212394

Estoup, A., P. Jarne, and J. M. Cornuet. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol. Ecol*. 11(9): 1,591–1,604. DOI:10.1046/j.1365-294X.2002.01576.x.

Ewing, B., and P. Green. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*. 8(3): 186–194.

Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res*. 8(3): 175–185.

Gordon, D., C. Abajian, and P. Green. 1998. Consed: A graphical tool for sequence finishing. *Genome Res*. 8(3): 195–202.

International Cocoa Organization. 2010. 2009/2010 Annual Report. http://www.icco.org/about-us/icco-annual-report.

Irish, B. M., R. Goenaga, D. P. Zhang, R. Schnell, J. S. Brown, and J. C. Motamayor. 2010. Microsatellite fingerprinting of the USDA-ARS Tropical Agriculture Research Station cacao (*Theobroma cacao* L.) germplasm collection. *Crop Sci*. 50(2): 656–667. DOI:10.2135/cropsci2009.06.0299.

Lanaud, C., O. Fouet, D. Clement, M. Boccara, A. M. Risterucci, and S. Surujdeo-Maharaj, T. Legavre, and X. Argout. 2009. A meta-QTL analysis of disease resistance traits of *Theobroma cacao* L. *Mol Breed*. 24(4): 361–374. DOI:10.1007/s11032-009-9297–4.

Lanaud, C., A. M. Risterucci, I. Pieretti, M. Falque, A. Bouet, and P. J. L. Lagoda. 1999. Isolation and characterization of microsatellites in *Theobroma cacao* L. *Mol. Ecol*. 8(12): 2,141–2,143. DOI:10.1046/j.1365-294x.1999.00802.x.

Livingstone, D. S., B. Freeman, J. C. Motamayor, R. J. Schnell, S. Royaert, J. Takrama, A. W. Meerow, and D. N. Kuhn. 2012. Optimization of a SNP assay for genotyping *Theobroma cacao* under field conditions. *Mol. Breed*. 30(1): 33–52. DOI:10.1007/s11032-011-9596–4.

Livingstone, D. S., B. Freeman, C. L. Tondo, K. A. Cariaga, N. H. Oleas, A. W. Meerow, R. J. Schnell, and D. N. Kuhn. 2009. Improvement of high-throughput genotype analysis after implementation of a dual-curve sybr green I-based quantification and normalization procedure. *Hortsci*. 44(5): 1,228–1,232.

Livingstone, D. S., J. C. Motamayor, R. J. Schnell, K. Cariaga, B. Freeman, A. W. Meerow, J. S. Brown, and D. N. Kuhn. 2011. Development of single nucleotide polymorphism markers in *Theobroma cacao* and comparison to simple sequence repeat markers for genotyping of Cameroon clones. *Mol. Breed*. 27(1):93–106. DOI:10.1007/s11032-010-9416–2.

Motamayor, J. C., P. Lachenaud, J. Mota, R. Loor, D. N. Kuhn, J. S. Brown, and R. J. Schnell. 2008. Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). *Plos One* 3(10). DOI:10.1371/journal.pone.0003311.

Motilal, L., and D. Butler. 2003. Verification of identities in global cacao germplasm collections. *Gen. Res. Crop Evol*. 50(8): 799–807. DOI:10.1023/a:1025950902827.

Opoku, S. Y., R. Bhattacharjee, M. Kolesnikova-Allen, J. C. Motamayor, R. Schnell, I. Ingelbrecht, L. Enu-Kwesi, and Y. Adu-Ampomah. 2007. Assessment of genetic diversity and population structure in West African cocoa: A case study on collections from Ghana. *J. Crop Improv*. 20(1–2): 73–87.

Parker, J. D., P. S. Rabinovitch, and G. C. Burmer. 1991. Targeted gene walking polymerase chain reaction. *Nucleic Acids Res*. 19(11): 3,055–3,060. DOI:10.1093/nar/19.11.3055

Ploetz, R. C. 2007. Cacao diseases: Important threats to chocolate production worldwide. *Phytopathol*. 97(12): 1,634–1,639. DOI:10.1094/phyto-97-12–1634.

Pugh, T., O. Fouet, A. M. Risterucci, P. Brottier, M. Abouladze, C. Deletrez, B. Courtois, D. Clement, P. Larmande, J. A. K. N'goran, and C. Lanaud. 2004. A new

cacao linkage map based on codominant markers: Development and integration of 201 new microsatellite markers. *Theor. Appl. Genet*. 108(6): 1,151–1,161. DOI:10.1007/s00122-003-1533–4.

Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol*. 5(2): 94–100. DOI:10.1016/s1369-5266(02)00240–6.

Risterucci, A. M., D. Paulin, M. Ducamp, J. A. N'goran, and C. Lanaud. 2003. Identification of QTLs related to cocoa resistance to three species of *Phytophthora*. *Theor. Appl. Genet*. 108(1): 168–174. DOI:10.1007/s00122-003-1408–8.

Rosen, S., and H. J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics methods and protocols: Methods in molecular biology,* edited by S. Krawetz and S. Misener, 365–386. Totowa, NJ: Humana Press.

Saunders, J. A., S. Mischke, E. A. Leamy, and A. A. Hemeida. 2004. Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao*. *Theor. Appl. Genet*. 110(1): 41–47. DOI:10.1007/s00122-004-1762–1.

Schnell, R. J., D. N. Kuhn, J. S. Brown, C. T. Olano, W. Phillips-Mora, F. M. Amores, and J. C. Motamayor. 2007. Development of a marker assisted selection program for cacao. *Phytopathol*. 97(12): 1,664–1,669.

Schnell, R .J., C. T. Olano, J. S. Brown, A. W. Meerow, C. Cervantes-Martinez, C. Nagai, and J. C. Motamayor. 2005. Retrospective determination of the parental population of superior cacao (*Theobroma cacao* L.) seedlings and association of microsatellite alleles with productivity. *J. Am. Soc. Hort. Sci*. 130(2): 181–190.

Schroeder, K. B., M. Jakobsson, M. H. Crawford, T. G. Schurr, S. M. Boca, D. F. Conrad, R. Y. Tito, et al. 2009. Haplotypic background of a private allele at high frequency in the Americas. *Mol. Biol. Evol*. 26(5): 995–1,016. DOI:10.1093/molbev/msp024.

Stephens, M., J. S. Sloan, P. D. Robertson, P. Scheet, and D. A. Nickerson. 2006. Automating sequence-based detection and genotyping of SNPs from diploid samples. *Nat. Genet*. 38(3): 375–381. DOI:10.1038/ng1746.

Takrama, J. F., C. Cervantes-Martinez, W. Phillips-Mora, J. S. Brown, J. C. Motamayor, and R. J. Schnell. 2005. Determination of off-types in a cacao breeding programme using microsatellites. *INGENIC Newsl*. 10:2–8.

Takrama, J. F., A. M. Dadzie, S. Y. Opoku, F. K. Padi, B. Adomako, Y. Adu-Ampomah, D. S. Livingstone Iii, J. C. Motamayor, R. J. Schnell, and D. N. Kuhn. 2012. Applying SNP marker technology in the cacao breeding program in Ghana. *African Crop Sci. J*. 20(1): 67–75.

Zhang, D. P., M. Boccara, L. Motilal, S. Mischke, E. S. Johnson, D. R. Butler, B. Bailey, and L. Meinhardt. 2009. Molecular characterization of an earliest cacao (*Theobroma cacao* L.) collection from Upper Amazon using microsatellite DNA markers. *Tree Genet. Genomes* 5(4): 595–607. DOI:10.1007/s11295-009-0212–2.

Zhang, D. P., W. J. Martinez, E. S. Johnson, E. Somarriba, W. Phillips-Mora, C. Astorga, S. Mischke, and L. W. Meinhardt. 2012. Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia. *Genet. Res. Crop Evol*. 59(2): 239–252. DOI:10.1007/s10722-011-9680–y.