Chapter 16

Cocoa

☆ *Lambert A. Motilal, Antoinette Sankar, David Gopaulchan and Pathmanathan Umaharan*

1. Introduction

Theobroma cacao L. (chocolate tree; 2n = 2x = 20), is an outbreeding understory tree, commonly referred to as cacao or cocoa, although the latter has been traditionally referred to the processed products of the cacao tree. Cacao was used by the early indigenous peoples of South and Central America more than 3,000 years ago, including the Olmecs, Mayan and later Aztec civilizations in Central America (Young, 1994; Coe and Coe, 1996; Henderson *et al.*, 2007; Powis *et al.*, 2011). The objects of these early exploits, referred to in the literature as "Criollo" cacao, were once taxonomically regarded as a separate species, *T. cacao* ssp. *cacao* (Cuatrecasas, 1964). Although studies have confirmed the Amazon basin as the centre of diversity of cacao and home to the 'Forastero' cacao (Cheesman, 1944; Motamayor et al., 2008; Thomas et al., 2012), there has been little understanding of diversity or phylogenetic relationships between geographical variants or populations. Further, the origin of 'Trinitario' cacao, a hybrid population between Forastero and Criollo and the basis for the fine/flavour cacao industry, has been mired in mystery. The contribution of molecular profiling of cacao has been tremendous, over the past four decades, on reinforcing or refining anthropological studies, resolving taxonomic issues, defining populations, understanding the origin of 'Trinitario' cacao and contributing to the overall understanding of the evolution of cacao.

Cacao has a little understood breeding system enforced by a complex multi-allelic incompatibility system encompassing features of both the sporophytic and gametophytic systems resulting in potential cross-compatible, incompatible and self-compatible reactions depending on the maternal tree and surrounding pollen donors (Knight and Rogers, 1953, 1955; Cope, 1958, 1962; Baker *et al.*, 1997; Ford and

Wilkinson, 2012). Highly inbred landraces also exist in Central and South America, referred to as 'Criollo' or 'Amelonado' or 'Nacional' (Bartley, 2005; Motamayor *et al.*, 2008; Loor Solorzano *et al.*, 2012). With the onset of European colonisation of the Americas, cacao was transported out of the South American mainland into the Caribbean Islands, West Africa, Asia and the Pacific islands (Wood, 1991; Lockwood and End, 1993; Bartley, 2005). Many of these introductions originated from the same geographic area or were descended from seedlings or seeds of few fruits that were transported in oceanic voyages. Plant material conveyed to and established at one site, often served as the secondary focal point for distribution to other distant areas. This, and the fact that a limited number of parents have been used in breeding programmes, has resulted in cultivated cacao possessing a narrow genetic base. The exploitation of genetic diversity in plant breeding has been hindered by poor understanding of the breeding system of cacao, the absence of reliable genetic information on the inheritance of traits and limited genetic markers to support breeding.

Cacao encompasses wild, semi-cultivated and cultivated varieties and the current form of the species is relatively unchanged from the *T. cacao* in the centre of diversity. Due to its recalcitrance and its predominantly outcrossing nature arising from its complex incompatibility system, cacao genetic resources are maintained as living collections in field genebanks (Figure 16.1). Cacao germplasm has been named according to the country, farm, collecting expedition, river system and research station (Lockwood and Gyamfi, 1979; Turnbull and Hadley, 2012). Cacao genetic resources are maintained in over 50 collections worldwide (Motilal and Butler, 2003) with two Universal Collections – the Centro Agronómico Tropical de Investigación y Enseñanza in Costa Rica and the International Cocoa Genebank Trinidad (ICGT) in Trinidad and Tobago. Information on cocoa varieties, or accessions as they are called in the germplasm collections, held in the worldwide collections with accessible records are contained in the International Cocoa Germplasm Database (ICGD) maintained by the University of Reading (Turnbull and Hadley, 2012). Details on accession nomenclature can be found in Lockwood and Gyamfi (1979), Iwaro *et al.* (2003), Bekele *et al.* (2006) and Turnbull and Hadley (2012). In the ICGD, there are about 29,500 accession names of which there are 13,000 synonyms (Turnbull and Hadley, 2012) giving approximately 16,500 putatively unique accessions. Accession nomenclature takes the form of an alphanumeric system where the lettered prefix indicates the accession group and the numeric suffix indicates the fruit, budwood or seed source. Related cacao accessions may therefore be identical to another accession, be within the same family structure as full-sibs or half-sibs, belong to the same accession group or belong to the same population. Cacao accessions may carry the same name but may be genetically distinct from each other, because the same name was unknowingly applied more than once to different germplasm collected in different expeditions; seed-derived descendants from a mother plant were given the same name as the mother plant; or from errors in recording during collection expeditions and reporting the accession name in the scientific literature and germplasm documentation.

Efforts to conserve cacao genetic resources, as repositories for safeguarding the livelihoods of smallholder farmers and cacao businesses, have been hindered by a poor understanding of the diversity, genetic structure and phylogenetic relationships among geographical populations. A more comprehensive understanding of the structure and diversity of cocoa is not only critical to understanding the gaps within the two Universal Collections of cacao germplasm, but is also important to improve the effectiveness and efficiency of collections through establishment of core and minicore collections. Poor fidelity in collections, seed gardens and clonal gardens is a hindrance to the efficiency of breeding programmes and also contributes to the poor quality of planting material supplied to farmers. Molecular markers play an important role in curation, characterisation, augmentation, utilisation and distribution of cacao germplasm. Multilocus molecular profiles can be used for identity analysis, genetic relatedness, genetic diversity, pedigree analysis, phylogenetic assessment and genetic ancestry.

The chapter traces the major historical advances in molecular profiling of cocoa, the current state of the art that has led to refining taxonomical relationships, improving the understanding of the evolution of diversity, its structure and phylogenetic relationships among populations, advancing the conservation and exploitation of genetic resources in breeding, enhancing the deployment of quality planting material in farmer fields, and future possibilities.

2. History of Molecular Profiling in Cocoa

Fruit and seed traits, in conjunction with geographic distribution, were used to classify cocoa into two groups of Criollo and Forastero (van Hall, 1914; Cheesman, 1944; Cuatrecasas, 1964) which were thought to be two separate subspecies *T. cacao* and *T. sphaerocarpum* respectively (Cuatrecasas, 1964). The Forastero group was further partitioned, based on fruit dimension and basal constriction, into Angoleta, Cundeamor, Amelonado and Calabacillo forms (Toxopeus, 1985). However, the limitations and ambiguity of this system have been recognised and a more systematic description of the phenotypic states (Enriquez and Soria, 1966, 1967; Engels *et al.*, 1980; Engels, 1983a,b), as used in most other crops, was instituted using heritable phenotypic features of leaf pubescence and colour; pod shape, features and colour; seed morphology and floral morphology. Phenetic dendrograms have been developed for cacao (Bekele and Bekele, 1996; Aikpokpodion, 2010; Bekele *et al.*, 1994; Bekele *et al.*, 2008a; Bekele *et al.*, 2008b; Maharaj *et al.*, 2011). However, the collection of phenotypic data is time-consuming and burdensome, particularly for reproductive traits in tree crops, phenotypic plasticity, and inconsistent scoring primarily due to improper or insufficient training of data collectors and the influence of modifying factors on trait expression. The use of biochemical and DNA markers, therefore supplanted phenotype characterisation in understanding taxonomic relationships. These molecular markers were heritable, more numerous, less susceptible to environmental vagaries and allowed for a more consistent, reliable, repeatable and reproducible study of genetic variation within any species.

Figure 16.1: Cacao Germplasm, in the International Cocoa Genebank Trinidad, Maintained as Living Trees.

A: Propagated accession with multiple trunks. The genotyped accession (note blue label on central trunk) was sampled. If the trunks are identical to each other, fruits can be used from all trunks for this accession; B: An accession with green fruits which become yellow upon ripening; C: An accession with reddish fruits that will become orange-yellow upon ripening.

Note that the fruit forms in A, B and C are different from each other indicating that these are different accessions. DNA fingerprinting allows for the unique identification of these accessions and provides an estimate of the relatedness of the accessions (Photographs courtesy of Lambert A. Motilal).

2.1. Protein, Allozyme and Isozyme Polymorphism

Protein banding patterns (Kaosiri and Zentmyer, 1980; Erselius and De Vallavieille, 1984; Chowdappa and Chandramohanan, 1995) and isoenzymes (Lanaud and Berthaud, 1985; Atkinson *et al.*, 1986; Lanaud, 1986; Yidana *et al.*, 1987; Oudemans and Coffey, 1991 a,b,c; Oudemans *et al.*, 1994) were the first to be used in molecular profiling in cacao. Allozymes are different forms of the same enzyme that may differ in electrophoretic mobility and can be separated in starch gels, stained and visualised. Following the early work on isozymes studies by Lanaud and Berthaud (1985), Atkinson *et al.* (1986), Lanaud (1986) and Yidana *et al.* (1987), attempted to characterise the genetic diversity held at the ICGT started at the Cocoa Research Centre, then Cocoa Research Unit, in 1984, using the isoperoxidase system although other enzymes including acid phosphatase were also investigated (Yidana and Kennedy, 1984-86). Later, other enzyme systems were incorporated to characterise cacao populations at the Cocoa Research Centre (Warren, 1994). Sounigo *et al.* (2005) reported the classification of 459 cacao accessions from 26 accession groups with isoenzymes and provided evidence for the separation of wild and cultivated types. Warren *et al.* (1995) found that the acid phosphatase and isocitrate dehydrogenase loci were linked to one of the loci in the incompatibility complex of cacao. Although isozymes provided for robust markers capable of being assayed early in the life time of cacao, there were concerns about the number of available markers and their repeatability, particularly across different laboratories.

2.2. Restriction Fragment Length Polymorphism

The use of DNA fingerprinting assays to analyse a plant genome was first reported by Ryskov *et al.* (1988) using restriction fragment length polymorphism (RFLP). The procedure involved digesting the DNA with a restriction enzyme to generate fragments of different sizes, separating the fragments using gel electrophoresis and transferring to a nitrocellulose or nylon membrane. The membrane is then probed with a radioactively labelled DNA fragment and the pattern visualised through autoradiography. Kurt *et al.* (1989) later used synthetic oligonucleotides that targeted simple, repetitive DNA sequences to generate hypervariable DNA fingerprinting profiles in barley and chickpea (*Cicer arietinum*). The technique, however, needed large amounts of DNA. In cacao, RFLPs were used to study the genetic diversity (Laurent *et al.*, 1993, 1994; Figueira *et al.*, 1994; N'Goran *et al.*, 1994; Lerceteau *et al.*, 1997), and later used to create the first genetic maps (Lanaud *et al.*, 1995; Crouzillat *et al.*, 1996). There is considerable support in these studies for the presence of greater diversity in South America and the Upper Amazon Basin, in particular, and the notion of this region as the centre of diversity for cacao. The markers could also be used to separate Criollo from Forastero types.

2.3. Random Amplified Polymorphic DNA

The field of DNA fingerprinting was transformed by the emergence of polymerase chain reaction (PCR)-based technology. Arbitrary primer sequences were used to amplify anonymous segments of genomic DNA to produce polymorphic banding patterns (Williams *et al.*, 1990; Caetano-Anolles and Brant,

1991; Welsh and McClelland, 1990). The assay published by Williams *et al.* (1990). based on the amplification of random segments in the genome with single primers of arbitrary nucleotide sequence called random amplified polymorphic DNA (RAPD), has since been reported in numerous studies. Unlike RFLP, the technique was quick and simple as it did not include blotting or hybridising steps and required only small amounts of DNA. Additionally, there was no requirement for DNA probes or sequence information for primer design as the primers detected polymorphisms in the absence of specific nucleotide sequence information.

The RAPD system was quickly adopted by cacao scientists for fingerprinting purposes with Wilde *et al.* (1992) being the first to report RAPD polymorphism in cacao. These authors assessed 14 primers on 13 genotypes, inclusive of two *Herrania* and one *T. microcarpum* L. individuals. One primer was identified that separated all the individuals. The marker system was adopted by CRC in 1993 and the reproducibility of the system was investigated by Christopher (1993) who used 11 accessions in common to that of Russell *et al.* (1993). Christopher and Sounigo (1995) screened 150 decameric RAPD primers and identified 18 promising primers from which five were selected for identity and cluster analysis. These five primers yielded 17 polymorphic loci, with 1-7 loci per primer and were used to separate 27 of 47 accessions. These authors therefore recommended additional primers (and hence more loci) for identity analysis. Subsequently, with refinement of the technique, RAPDs were used for linkage maps, identity analysis and genetic diversity studies in cacao (Figueira *et al.*, 1994; N'Goran *et al.*, 1994; Lerceteau *et al.*, 1997; Whitkus *et al.*, 1998; Kasran and Subali, 2002; Lanaud *et al.*, 2004; Sounigo *et al.*, 2005; Leal *et al.*, 2008). Analysis of RAPD profiles provided support for unique diversity in southern Mexico (Whitkus *et al.*, 1998), separation of dwarf and vigorous clones (Kasran and Subali, 2002), the distinction of Criollo from Forastero (Figueira *et al.*, 1994; N'Goran *et al.*, 1994; Lerceteau *et al.*, 1997), and the separation of 22 accession groups from each other inclusive of wild and improved cultivars (Sounigo *et al.*, 2005). However, the repeatability and reproducibility of the technique was greatly affected by reaction conditions and comparison of DNA fingerprinting profiles within and among laboratories was severely affected.

2.4. Amplified Fragment Length Polymorphism

Another approach to using PCR in DNA fingerprinting was the development of the amplified fragment length polymorphism (AFLP) assay (Vos *et al.*, 1995). The procedure involved digesting the DNA with a restriction enzyme mixture, ligating oligonucleotide adapters, amplifying the restriction fragments and pre-selective amplification followed by selective-PCR. With this method, a high number of polymorphic bands could be produced and highly reproducible unique DNA fingerprinting patterns generated. Nevertheless, the assay is technically challenging and a larger quantity of DNA than RAPDs is required. The first reported AFLP study in cacao (Perry *et al.*, 1998) reported inter- and intra-varietal differences even among phenotypically similar cacao varieties. AFLPs were mainly used in cocoa to saturate linkage maps (Risterucci *et al.*, 2000; Clément *et al.*, 2003a, b; Quieroz *et al.*, 2003; Lanaud *et al.*, 2004). However, AFLPs have not been adopted for cultivar identification or study of genetic diversity at any of the cocoa genebanks worldwide.

2.5. Inter-simple Sequence Repeat Polymorphism

Inter-simple sequence repeat (ISSR) –PCR, developed by Zietkiewicz *et al.* (1994), is another PCR-based DNA profiling assay. The technique is used to detect the variation in length between two simple sequence repeats (SSRs, microsatellites) in the genome. SSRs are polymorphic loci consisting of multiple tandem repeating sequences usually two to six base pairs in length. The target of an ISSR assay is any DNA segment located within an amplifiable distance between two identical SSR regions oriented in opposite direction. The ISSR markers are mostly inherited as dominant alleles although co-dominance has been reported in maize (Gupta *et al.*, 1994) and citrus (Sankar and Moore, 2001). Although ISSRs were recommended by Charters and Wilkinson (2000)for fingerprinting cacao, its use as a tool for cacao profiling was limited because of the speed of success in the cloning and isolation of microsatellite sequences. A common disadvantage of the RAPD, AFLP and ISSR techniques is the difficulty to determine whether an amplified DNA segment is heterozygous or homozygous at a particular locus. As a result, the marker alleles generated by these assays are dominant which reduces their information content. This limitation was overcome when cloning and sequencing of microsatellite sequences led to SSR markers that were co-dominant in their inheritance.

2.6. Simple Sequence Repeat Polymorphism

Simple Sequence Repeat (SSR) assays, in contrast to the previous markers, provide a combination of locus-specific, co-dominantly inherited bands with high levels of polymorphism. The amplified regions have been termed sequence tagged microsatellite sites (a variant of a sequence tagged site) because the microsatellite is effectively tagged in the genome through the design of a forward and a reverse primer from the unique regions flanking the microsatellite. Microsatellite mutation rate in the genome has been reported to be as high as 10^{-2} per generation (Weber and Wong, 1993; Li *et al.*, 2002), often through alteration of repeat number (Li *et al.*, 2002). Repeat number changes have been theorised to occur through DNA slippage during replication or asymmetric recombination between DNA strands (Tachida and Iizuka, 1992). These polymorphisms have been shown to be co-dominantly inherited, and are therefore valuable for DNA typing and other genetic profiling applications. Another type of marker, SSR-RFLP, combined the power of RFLP with the simplicity of PCR using primers designed around microsatellite-containing regions in the genome. Attempts had been made to evaluate the potential of SSR-RFLP for use in developing country laboratories (P. Umaharan, pers comm). The concept involved the amplification of a fragment, surrounding an SSR locus, long enough to allow subsequent digestion using restriction enzymes, to reveal alleles based on variations in restriction fragment length through gel electrophoresis. However, this method has not been successfully adopted.

The SSR-PCR product size can be predicted based upon the sequence information and used to confirm successful amplification. Sizing of the amplified fragments/alleles can be done manually with size standards or using a high throughput sequencing system for automated sizing. PCR amplification of SSR length polymorphisms in plants was first reported by Akkaya *et al.* (1992) using

soybean (*Glycine max*) DNA. Primers that targeted the DNA sequences flanking hypervariable SSR regions in the genome were used in the assay. A limitation of this technique is that species-specific primer pairs are required to produce SSR markers. The development of such primer assays can be time consuming and costly, although assays developed for a species may be transferable across different taxonomic levels depending on the SSR loci and organism involved (Rosetto, 2001; Scott, 2001; Motilal, 2004a).

Cacao-based SSRs were first developed by Lanaud *et al.* (1999). Argout *et al.* (2008) later identified 2252 SSRs from 149,650 ESTs derived from a transcriptome set of 56 cDNA over different cacao organs and environmental conditions. Dimeric to hexameric pure SSRs and compound SSRs were identified with the dimeric and trimeric SSRs accounting for 88 per cent of all SSRs identified and with the poly$(AG)_n$ and poly$(AAG)_n$ motifs being most abundant in the unigenes (Argout *et al.*, 2008).

Variable numbers of SSR loci have been employed in identity and genetic diversity studies. A set of 15 SSR primer pairs (Saunders *et al.*, 2004) were recommended for resolution of identity issues. Cryer *et al.* (2006) later reported on the use of reference genotypes and allelic size standards to unify allele calls. However, ascertainment bias in sample selection affects SSR informativeness (Johnson *et al.*, 2009; Motilal *et al.*, 2009) which may limit the utility of these 15 loci. Errors in SSR typing arise from allele drop out and false alleles and these error rates were locus dependent (Zhang *et al.*, 2006b). Repeat typing was recommended to obtain reliable consensus genotypes (Zhang *et al.*, 2006b).

To date, SSR-PCR has been used for off-types and cacao clone identification (Figueira, 1998; Risterucci *et al.*, 2000, 2001; Motilal, 2004b; Schnell *et al.*, 2004; Cryer *et al.*, 2006; Motilal *et al.*, 2009, 2011; Zhang *et al.*, 2006a,b, 2007, 2009a,b). The combined probability of identity among siblings (PID_{SIB}) is the probability that two sibling individuals drawn at random from a population have identical genotypes (Evett and Weir, 1998; Waits *et al.*, 2001). The PID_{SIB} was recommended to be used in assigning identity match declarations (Zhang *et al.*, 2009a).

Microsatellite polymorphisms have also been used in linkage map generation (Lanaud *et al.*, 2004; Pugh *et al.*, 2004) and for tagging genes for quantitative trait loci towards marker assisted selection (Clement *et al.*, 2003a,b; Brown *et al.*, 2005, 2007; Schnell *et al.*, 2005, 2007a; Lanaud *et al.*, 2009; Marcano *et al.*, 2009). The utility of SSRs in parentage analysis (Schnell *et al.*, 2005) describing genetic diversity and in assessing population ancestry is well known (Lanaud *et al.*, 1999, 2001; Opoku *et al.*, 2007; Motamayor *et al.*, 2008; Marcano *et al.*, 2009; Zhang *et al.*, 2006a,b, 2007, 2009a,b, 2011; Irish *et al.*, 2010; Motilal *et al.*, 2010, 2011, 2012, 2013; Susilo *et al.*, 2011; Trognitz *et al.*, 2011, 2013; Thomas *et al.*, 2012; Dinarti *et al.*, 2015; Santos *et al.*, 2015).

2.7. Single Nucleotide Polymorphism

Single nucleotide polymorphisms (SNPs) are variations at individual nucleotides at specific positions in the genome. SNPs are reportedly the most abundant class of polymorphisms in genomes including the plant genome (Buckler and Thornsberry, 2002; Brumfield *et al.*, 2003). Density estimates have yielded one SNP per 170 bp in

the rice genome (Yu *et al.*, 2002), 1900 bp in humans (The International SNP Map Working Group, 2001), 3300 bp in *Arabidopsis thaliana* (The Arabidopsis Genome Initiative 2000), 20,500-24,600 bp in chickpea (Kujur *et al.*, 2015). SNPs can potentially exist at any position throughout the genome, including coding, non-coding and inter-genic regions and hence offers the greatest promise for molecular profiling. Besides its abundance, SNPs present the advantage of being identified without the need for electrophoresis and hence size homoplasy, platform issues and binning problems are avoided. Genotyping error rates are low (less than 1 in 2000; Ranade *et al.*, 2001) and SNP genotypes can be scored with minimal human intervention. Furthermore, similar to SSRs, SNPs are co-dominant but, unlike SSRs, are biallelic rather than multiallelic. However, triallelic states in humans (Hüebner *et al.*, 2007; Hodgkinson and Eyre-Walker, 2010; Jenkins *et al.*, 2014) or multiallelic SNPs (Jenkins *et al.*, 2014) due to copy number variation (MacConaill *et al.*, 2007) have been reported. Multiallelic SNPs have been attributed to sequencing errors (Beissinger *et al.*, 2014). The codominant biallelic state for SNPs will be discussed in this chapter. Three states – homozygous 1, homozygous 2 and the heterozygous condition can therefore be detected by a single biallelic SNP. For example the genotypes AA, GG and AG will be detected using a SNP with biallelic states A/G. Due to this level of detection, SNPs have limited resolving power per locus compared to microsatellites (Glaubitz *et al.*, 2003). However, when the number of sites is factored in, the overall resolving power of SNPs may be comparable to that of other DNA markers. SNPs have therefore emerged as the next generation of molecular markers for species identification and genetic diversity measurements.

SNPs have been used for diverse applications in crops including providing evidence for selection (Beissinger *et al.*, 2014); varietal identification (Ganal *et al.*, 2009); describing population genetic structure (Schmid *et al.*, 2003; Kujur *et al.*, 2015); tagging genes or quantitative trait loci in rice (Konishi *et al.*, 2006), maize (Buckler *et al.*, 2009) and oil palm (Pootakham *et al.*, 2015); generating linkage maps in many crops including *Arabidopsis* (Cho *et al.*, 1999), rice (Nasu *et al.*, 2002; Feltus *et al.*, 2004; Shen *et al.*, 2004) and oil palm (Pootakham *et al.*, 2015); and for genome wide association studies (GWAS) (Bélo *et al.*, 2008; Pajerowska-Mukhtar *et al.*, 2009).

The earliest SNP studies in cacao are those of Borrone *et al.* (2004) and Kuhn *et al.* (2005). These markers in cacao have been derived from single strand conformation polymorphism (Kuhn *et al.*, 2005; Livingstone *et al.*, 2011); expressed sequenced tags (Lima *et al.*, 2009; Allegre *et al.*, 2012) which could come from conserved ortholog set sequences (Kuhn *et al.*, 2012); from microsatellites (Dadzie *et al.*, 2013) using genome walking (Parker *et al.*, 1991) and from RNAseq data (Livingstone *et al.*, 2015). Argout *et al.* (2008) identified 5246 SNPs which were distributed as transitions (A/T - G/C; 54.2 per cent), transversions (32.1 per cent) and indels (13.7 per cent) in their EST database. Subsequently, two cacao genomes were mapped: one a Belizean Criollo cultivar B97-61/B2 (Argout *et al.*, 2011) and another from the Amazonian MATINA 1/6 (Motamayor *et al.*, 2013). These genomes were used to identify numerous SNPs (Allegre *et al.*, 2012). SNPs in cocoa, have been used to generate linkage maps (Allegre *et al.*, 2012; Livingstone *et al.*, 2015), tag genes (Borrone *et al.*, 2004; Lima *et al.*, 2009; Motamayor *et al.*, 2013), determine parentage (Ji *et al.*, 2013; Takrama *et al.*, 2014),

evaluate genetic diversity (Ji *et al.*, 2013) and identify hybrids and varieties (Kuhn *et al.*, 2010; Livingstone *et al.*, 2011; Takrama *et al.*, 2012, 2014; Fang *et al.*, 2014).

3. Development of a Global Reference SNP Panel for Resolving Identities in Cacao

Notwithstanding the ongoing cacao SNP studies, neither a logical methodology for selecting SNPs for resolution of identities has been developed nor has there been an attempt to use such an approach to identify a global reference SNP panel for resolution of identities. SNPs were reportedly selected based on their level of polymorphism and to represent all ten chromosomes. Polymorphism estimates, however, have an ascertainment bias as it is affected by the number of samples and the composition of the samples. For instance, even monomorphic SNPs (Kuhn *et al.*, 2010; Livingstone *et al.*, 2011; Allegre *et al.*, 2012; Ji *et al.*, 2013) can be deemed polymorphic depending on the set of cacao plants screened (Livingstone *et al.*, 2011). In this chapter, we present a global reference SNP panel selected based on screening a subset of germplasm from the ICGT for their ability to discriminate the genetic variability for cacao.

The ICGD contains SNP data on 603 accessions from 160 loci (Turnbull and Hadley, 2012) of which 54 bears the "Tc" prefix similar to that of Ji *et al.* (2013) and Takrama *et al.* (2014). These TcSNPs were derived from a set of over 2000 SNPs by CIRAD from Illumina GoldenGate Assays as described by Allegre *et al.* (2012) and prioritised by Michel Boccara at CRC, Trinidad (Michel Boccara, personal communication). Fang *et al.* (2014) employed 44 loci which had the capacity to discriminate amongst the 160 individuals studied but could not resolve the SCA/Ucayali accessions from the MO accessions into their respective population groups of Contamana and Nacional. Ji *et al.* (2013) and Livingstone *et al.* (2015) have suggested that 26 and 30 SNP loci respectively would be sufficient for SNP profiling.

The multilocus profiles of 546 SNPs from 81 cocoa DNA samples were obtained from Illumina GoldenGate Assays as described by Allegre *et al.* (2012). The data was used to obtain estimates of H_e and PIC using PowerMarker (Liu and Muse, 2005), as well as, estimates of simulated power ($simPW_R$) and informativeness for relatedness (I_r) using the program KinInfor v1 (Wang, 2006). Using these outputs and the location of SNPs on the linkage groups (http://cocoagendb.cirad.fr/), different datasets (Table 16.1) were compiled to assess the choice of SNPs on resolution of tree identities.

As was demonstrated for SSR loci, the composition of the panel of polymorphic loci is important and all the members do not need to be the most informative (Motilal *et al.*, 2009). The SNP loci should be selected as a set based on the capacity to discriminate among all individuals. Although, at least 30 loci could give at least 95 per cent resolution, the number of near matches across the different SNP combinations indicate that these combinations would inflate the relatedness of distinct accessions. Relatedness becomes overestimated as the incidence of missing data is increased. If by chance, missing data occurs at the few sites that differentiate the near misses, then even with a reasonably large number (100) of SNPs, distinct but closely related accessions would be deemed equivalent.

Table 16.1: Ascertainment of SNP Selection on Identity Resolution of 81 Cacao Accessions

SNP Selection	*Number of SNPs*	*Resolution(%)*	*Number of Sample Pairs Matching at all but n loci*	*Number of Synonymous Groups*	PID_{SIB}
Random 15	15	90.1	n = 1: 9 n = 2: 33	3 sets: 2 doublets and 1 quadruplet	2.6×10^{-3}
Random 30	30	96.3	n = 1: 6 n = 2: 4	1 triplet	1.5×10^{-5}
Random 45	45	95.1	n = 1: 0 n = 2: 4	1 quadruplet	1.0×10^{-6}
Random 60	60	96.3	n = 1: 3 n = 2: 0	1 triplet	7.6×10^{-10}
Random 75	75	100	n = 1: 4 n = 2: 4	none	1.0×10^{-9}
Random 100	100	96.3	n = 1: 0 n = 2: 3	1 triplet	9.1×10^{-13}
Chromosome 1	92	100	n = 1: 4 n = 2: 2	none	9.9×10^{-12}
Chromosomes 2 and 3	112	97.5	n = 1: 0 n = 2: 3	1 doublet	1.6×10^{-15}
Chromosomes 2 and 8	92	96.3	n = 1: 3 n = 2: 0	1 triplet	2.0×10^{-13}
Chromosomes 4 and 5	112	100	n = 1: 2 n = 2: 1	none	7.0×10^{-16}
Chromosomes 6 and 9	100	97.5	n = 1: 4 n = 2: 5	1 doublet	6.6×10^{-13}
Chromosomes 6, 7, 8, & 10	99	97.5	n = 1: 3 n = 2: 2	1 doublet	1.7×10^{-14}
Ten SNPs distributed per chromosomen = 2: 0	100	95.1	n = 1: 4	1 quadruplet	2.3×10^{-13}
SNPs at every 7.5 cM for each chromosome	100	97.5	n = 1: 0 n = 2: 5	1 doublet	1.6×10^{-13}
[1]Top 100 by H_e	100	97.5	n = 1: 3 n = 2: 3	1 doublet	1.1×10^{-22}
[1]Top 100 by PIC	100	96.3	n = 1: 0 n = 2: 3	1 triplet	2.8×10^{-23}
[2]Top 100 by $simPW_R$	100	100	n = 1: 0 n = 2: 1	None	1.4×10^{-17}
[2]Top 100 by I_r	100	96.3	n = 1: 0 n = 2: 3	1 triplet	2.8×10^{-23}
Select set of 60	60	100	n = 1: 0 n = 2: 0	None	2.1×10^{-13}

1 Choice based on SNPs ranked by expected heterozygosity (H_e), polymorphism information content (PIC) estimated in PowerMarker (Liu and Muse, 2005) from 546 SNPs/81 DNA samples.

2 Choice based on SNPs ranked by simuleted power (simPWR) or informativeness for relatedness (I_r) estimated in KinInfor v1 (Wang, 2006) from 546 SNPs/81 DNA samples.

A panel of 192 SNPs (Table 16.2) was compiled based on the above results, on accumulated SNP data at CRC and published information (Ji *et al.*, 2013; Takrama *et al.*, 2014). Eighteen of the 'Select set of 60' loci in Table 1 were retained in the CIRAD-CRC SNP panels but only one of these (TcSNP0886) was common to the set of loci employed by the previously mentioned sources. The complete panel covered 43 of the 54 TcSNP loci in the ICGD database (Turnbull and Hadley, 2012), 56 of the 70 SNP loci employed by Ji *et al.* (2013) and 45 of the 48 SNP loci employed by Fang *et al.* (2014). The complete panel is designed as two complementary 96- member subsets of which Panel A could be used as the first panel. This panel contained 36 loci in common with Fang *et al.* (2014).

Table 16.2: CIRAD-CRC Panel of SNPs for Identity Resolution in Cacao

Panel[1]	TcSNP	Chrom[2]	Panel	TcSNP	Chrom	Panel	TcSNP	Chrom
A	13	2	A	261	10	A	642	8
A	33	2	A	290	3	A	645	5
A	49	4	A	305	6	A	660	5
A	64	5	A	309	6	A	702	10
A	131	8	A	313	2	A	703	3
A	135	4	A	316	2	A	704	10
A	139	8	A	326	3	A	736	5
A	143	3	A	329	7	A	749	4
A	144	10	A	339	9	A	750	6
A	149	5	A	364	9	A	751	5
A	150	5	A	372	4	A	791	7
A	151	8	A	380	2	A	799	8
A	154	1	A	391	4	A	800	8
A	189	8	A	397	2	A	814	7
A	192	9	A	429	2	A	835	5
A	193	9	A	469	7	A	836	2
A	194	1	A	522	5	A	841	5
A	198	4	A	556	6	A	852	3
A	226	9	A	560	10	A	872	4
A	230	10	A	591	1	A	878	3
A	240	1	A	602	6	A	885	2
A	242	9	A	619	6	A	891	2
A	899	8	A	1309	8	B	214	2
A	917	10	A	1350	1	B	256	8
A	929	3	A	1362	8	B	258	7
A	945	3	A	1414	9	B	259	1
A	994	6	A	1484	6	B	280	3
A	998	5	A	1520	8	B	341	9
A	999	8	A	1522	2	B	363	3
A	1010	1	A	1527	6	B	421	3
A	1011	1	B	19	3	B	455	9

Contd...

Panel[1]	TcSNP	Chrom[2]	Panel	TcSNP	Chrom	Panel	TcSNP	Chrom
A	1028	4	B	**25**	9	B	456	4
A	**1038**	**5**	B	32	4	***B***	***461***	***5***
A	**1060**	**2**	B	**75**	**2**	B	**480**	9
A	**1062**	**3**	***B***	***97***	***2***	B	519	4
A	**1075**	**1**	B	105	8	B	531	1
A	1144	6	B	126	5	**B**	**534**	1
A	1165	2	B	141	3	B	546	10
A	1200	6	B	148	3	**B**	**577**	5
A	1201	7	B	164	1	B	589	3
A	1216	1	***B***	***173***	***1***	B	606	**7**
A	**1253**	9	B	**174**	4	B	607	1
A	**1270**	**7**	B	176	3	***B***	***636***	***8***
A	1275	1	B	186	?	B	640	2
B	644	9	B	954	1	B	1237	7
B	669	2	B	964	6	B	1266	6
B	674	10	B	1019	5	B	1293	?
B	**723**	**10**	B	1053	2	B	1302	1
B	737	9	B	1058	1	B	1308	1
B	739	3	B	**1063**	7	B	1329	1
B	***773***	***4***	B	1074	10	B	**1331**	10
B	785	1	***B***	***1093***	***4***	***B***	***1349***	***5***
B	**823**	**5**	B	1106	2	B	1401	1
B	833	9	B	1112	5	B	1404	2
B	857	9	B	1136	2	B	1408	5
B	867	9	B	1156	6	B	1416	5
B	871	5	B	1158	5	***B***	***1422***	***2***
B	881	5	B	**1159**	**2**	B	1426	8
B	***886***	***4***	***B***	***1160***	***4***	B	1432	1
B	901	6	B	1195	2	**B**	**1453**	**5**
B	***906***	***2***	B	1205	4	B	1457	3
B	933	9	B	1223	5	**B**	**1458**	**1**
B	947	9	B	1228	?	B	1483	8
B	**953**	4	B	1229	4	B	1524	2

1. Each panel consists of 96 loci.
2. Chromosome; Bold black font are loci in common with Ji *et al.* (2013). Bold, italic, blue font are loci in the 'Select set of 60' loci from Table 16.1 that was useful in discriminating among accessions. One marker (TcSNP0886) is common to both datasets.
3. SNPs with uncertain chromosome assignment are indicated by ?.

Table 16.3: Number of SNP Markers Used in Varietal Studies in Cacao

Number of SNPs	*Sample size*	*Germplasm Source*	*Genetic Groups Covered*	*Resolution*	*Reference*
6	50	Ghana (CRIG germplasm)	Iquitos, Nanay, Marañón	30 per cent; 8 synonymous groups present	Takrama *et al.* (2012).
13	186	Cameroon, Trinidad	Amelonado, Guiana, Iquitos, Nanay, Nacional	not given; SNPs identified 24.2 per cent offtypes	Livingstone *et al.* (2011).
18	171	USDA-ARS SHRS[1] (Florida, USA)	Seedlings from four mother trees representing three genotypes	Not given; apparently 100 per cent	Livingstone *et al.* (2012).
48 assessed; 44 usable	48	Brazil, Costa Rica, Ecuador, Peru, Trinidad	Amelonado, Criollo, Curaray, Iquitos, Nanay, Marañón	Full; no matching samples	Fang *et al.* (2014).
54 assessed; 53 employed	160	Ghana (CRIG germplasm)	Amelonado, Curaray, Iquitos, Nanay, Marañón	Not given; apparently 100 per cent	Takrama *et al.* (2014).
100 assessed; 70 employed	115	Costa Rica, Honduras, Nicaragua, Trinidad	Amelonado, Criollo, Contamana, Iquitos, Nacional, Nanay,	86.1 per cent; 6 synonymous groups: 5 Criollo sets and 1 Amelonado set	Ji *et al.* (2013).

1 Subtropical Horticulture Research Station.

4. Applications of Molecular Profiling

4.1. Resolving Fidelity Issues

Unambiguous identification of individuals and usage of correctly named accessions are important in managing germplasm collections, seed gardens and clonal gardens, ensuring fidelity during germplasm transfer, parental choice in breeding programmes and reducing the errors associated with experimental analyses especially when dealing with phenotypic data. The problem of mislabelling in crop genebanks is well recognised (van Hintum, 2000; Hurka *et al.*, 2004) and has been shown to be an issue in cocoa germplasm collections (Figueira, 1998; Christopher *et al.*, 1998; Risterucci *et al.*, 2001; Motilal and Butler, 2003; Sounigo *et al.*, 2006a; Motilal *et al.*, 2009, 2011, 2012). Cases where accessions have the same name, but have different genetic profiles, are termed homonymous errors. Cases where accessions have different names but the same genetic profile are termed synonymous errors.

The decreasing cost of molecular marker technologies have led to their routine application for fidelity testing in cocoa germplasm collections, clonal gardens and breeding. A standard set of 15 SSR loci were recommended for cacao identity studies (Saunders *et al.*, 2004). However, full resolution of germplasm accessions was not possible with this set of 15 and a set of nine SSR loci that performed better in separating accessions, was later recommended (Motilal *et al.*, 2009). Similarly two panels of 96 SNPs (192 loci; Section 4.3) are recommended for resolving the identity of cacao accessions. In the past, 48-100 SNPs have been used in cacao for identity studies (Table 16.3). The random match probability (Budowle *et al.*, 2000) and the estimated Amazonian population of 6.81×10^{10} cacao trees (Motilal *et al.*, 2009) was used to obtain a conservative estimate of 1.48×10^{-13} which could represent the upper boundary for identity declaration. Values for PID_{SIB} fall in the range of 10^{-10} (Motilal *et al.*, 2012; 26 SSR loci); 10^{-9} (Takrama *et al.*, 2014; 53 SNPs); 10^{-6} to 10^{-5} (Zhang *et al.*, 2006a,b, 2009a, 2011; 15 SSRs) and 10^{-5} from 26 (Ji *et al.*, 2013) or 44 SNPs (Fang *et al.*, 2014) have been reported. The lower polymorphism of SNP markers relative to SSR loci means that more SNP loci are required than SSR loci to obtain the same PID_{SIB} value. The number of SNP loci suggested to be sufficient (based on the number of, and types of accessions present in the study), to get the same order of magnitude of probability in declaring identity, is nearly twofold that of the number of SSRs previously used among cacao scientists worldwide (Livingstone *et al.*, 2011, 2015; Ji *et al.*, 2013).

The SNPs identified in Table 2 generated a combined PID_{SIB} of 2.9×10^{-31} from a set of 184 loci on 1800 samples inclusive of variable numbers of replicated trees per accession. The suitability of this panel in determining genetic identity and the identification of a best possible minimum set of SNP loci for use in varietal identification has been evaluated using cacao accessions in the ICGT.

4.2. Defining Populations and Phylogenetic Relationships

Molecular fingerprinting techniques such as isozyme electrophoresis, RFLP, RAPD and AFLP and more recently, SSR and SNP assays have allowed for more

rapid and accurate investigating of phylogenetic relationships of cacao germplasm using evolutionary models. Sounigo *et al.* (2005), using RAPDs and isozyme electrophoresis techniques, separated germplasm from the ICGT into three major clusters, with one consisting mainly of trees native to Peru, Ecuador and French Guiana; one mainly of all the cultivated Trinitario and one exclusively of trees cultivated in Ecuador. Irish *et al.* (2010) found that accessions generally grouped together according to their geographical origin and traditional genetic background. One cluster contained mainly Trinitario type accessions with varieties mostly from Mexico, Central America, and the Caribbean; while the Amelonado, SIAL and SIC accessions that originated from Brazil, were grouped in another cluster; and the Ecuadorian accessions such as EET and UF which are linked to "Nacional" ancestry, formed another cluster.

Motamayor *et al.* (2008) using 106 SSR markers on samples with provenance from the wild, partitioned cacao genetic diversity into 10 phylogenetic groups–Amelonado, Contamana, Criollo, Curaray, Guiana, Iquitos, Marañón, Nacional, Nanay and Purús. These groups with the exception of Criollo (Mesoamerican origin) are of Amazonian origin. Thomas *et al.* (2012) mapped the 10 genetic groups to distinct geographical locations, with distinct allelic compositions, suggesting that the genetic groups may represent geographically isolated populations. However, as explorations are made in South American countries with endemic cacao, the number of genetic groups presently defined is expected to increase as was recently observed with the discovery of a unique cacao population in Bolivia (Zhang *et al.*, 2011).

The phylogenetic relationships among accessions collected from different expeditions have been further dissected in other studies. Molecular profiles (Sounigo *et al.*, 2002, 2005; Boccara and Zhang, 2006, 2007, 2008; Zhang *et al.*, 2009a; Motilal *et al.*, 2009, 2012, 2013; Loor Solorzano *et al.*, 2012; Boza *et al.*, 2013) and morphological data (Bekele *et al.*, 2005, 2006), have been used to establish the proximity of IMC and NA accessions to each other. This suggests that genetic material collected from the two river systems may have some common ancestry.

The POUND accessions were reportedly collected as budwood from the same area, and supposedly from the same mother trees as the IMC, NA, SCA accessions which were collected as fruits (Pound, 1945; Bartley, 2005). The POUND accessions have been found to be distributed among the SCA, IMC and NA clusters from molecular data (Motamayor *et al.*, 2008; Zhang *et al.*, 2009a; Motilal *et al.*, 2009, 2012; Loor Solorzano *et al.*, 2012) and this was better demonstrated with RAPD data than with isoenzyme data (Sounigo *et al.*, 2005). Using morphological data, the POUND accessions were positioned within the same cluster as NA and PA accessions (Bekele *et al.*, 2006), or dispersed amongst IMC, NA and SCA accessions (Bekele *et al.*, 2005).

Refractario is a group of related accessions that were mass selected for witches' broom disease resistance under natural disease pressure in Ecuador (Pound 1938, 1943; Bartley, 2001, 2005). Phylogenetically, the Refractario accessions cluster separately as a single clade quite distinct from Amelonado, Contamana, Criollo, Guiana, Iquitos, Marañón and Nanay accessions (Motilal *et al.*, 2013). The Refractario germplasm pool has two main subclusters (Zhang *et al.*, 2007; Motilal *et al.*, 2012) and was shown to be genetically distinct from Trinitario accessions based both

on molecular profiles (Zhang *et al.*, 2007; Boccara and Zhang, 2007, 2008) and morphological variation (Bekele *et al.*, 2006).

4.3. Corroborating Family Structure

A cacao fruit has variable number of seeds ranging between 15-60 seeds. Each seed can have a potentially different pollen parent which depends on the surrounding trees, the cross-compatibility of the maternal tree and the self-compatibility of the maternal tree. Accessions derived from fruits of a tree are therefore expected to be half-sibs or full sibs with respect to each other. Contrariwise, germplasm collected as budwood from trees in a particular locale, may or may not be related to each other. Furthermore, accessions derived from bulked seed lots or mixed germplasm may have hidden family structure. Since historical records and inventory may be lacking in detail, the use of molecular markers can provide clarity. Knowledge of the family structure is important in planning breeding experiments. Heterosis can be exploited if crosses are between diverse genotypes. Germplasm curators can use family structure in describing the genetic diversity of the collection and obtaining a measure of the redundancy of the collection (Motilal *et al.*, 2013).

For instance, the SCA, IMC, PA and NA accessions were derived from 1, 2, 7-20, and 14-17 mother trees respectively according to collection records (Pound, 1938; Bekele *et al.*, 2005). Molecular evidence has been used to show that the SCA and IMC accessions came from two trees each, the PA accessions from 20 trees, and the NA accessions from 22 mother trees (Zhang *et al.*, 2009a) which is a reasonable fit to the historical literature. The MO accessions were derived from two mother trees (Zhang *et al.*, 2009a) which agreed with the suggestion of Bartley (2005). that fruits were collected from more than one tree.

Additional work in the ICGT with a higher complement of Refractario accessions, but also containing Amelonado, Criollo, GU, NA, IMC, and PA accessions, identified 56 full-sib families and 189 half-sib families each of which contained between 2-17 individuals (Motilal *et al.*, 2013). The Guiana accessions formed a family cluster that could not be linked with the other family units of the 189 half-sib families. High sibship in the Refractario accessions was expected as this germplasm was collected as fruits from about 80 trees that showed disease resistance and these trees were derived from seedlings obtained from disease-free trees (Pound, 1938, 1943). Hybrid types like the ICS and TRD accessions also exhibited family structure and these accession groups contained 12 full-sib families with 2-9 members per family (Johnson *et al.*, 2009). Likewise, up to 12 full-sib families were found in the Indonesian cacao germplasm which were likely derived from Upper Amazon Forastero, Venezuelan germplasm and Criollo germplasm (Susilo *et al.*, 2011).

4.4. Evidence for Anthropogenic Effects

Evidence for anthropogenic effects on cacao genetic diversity can also be gleaned from ancestry and phylogenetic analysis. Thomas *et al.* (2012) provided evidence to support the role of glaciation and human mediation on cocoa population differentiation. The PA accessions, form part of the Marañon group that was collected from the Upper Amazon region, and have been shown to be genetically similar to

Lower Amazon Forastero material from Brazil (Motamayor *et al.*, 2008) Bartley (2005) suggested on the basis of morphological observations, that the PA group probably descended from planting material imported from Brazil. Zhang *et al.* (2009a), using Amelonado accessions from Brazil (Lower Amazon Forastero) showed that these clustered with the PA accessions and merged into the same population group when large groups (K=3) were compiled using the software 'Structure'. These and other data from isozyme and RAPD data (Sounigo *et al.*, 2005) could be taken to support the hypothesis of Bartley (2005). However, the descent of Lower Amazon material from the Upper Amazon via a stopping point at the area where the PA accessions were collected can also be supported by the data. This alternative view is bolstered by the results of Thomas *et al.* (2012) and Sereno *et al.* (2006) who advocated that part of the Lower Amazon cacao population was derived from the Upper Amazon. The Amelonado and the Marañon-Amazon River cluster were also suggested to be derived from the same genepool that led to the differentiation of the Iquitos and Púrus clusters (Thomas *et al.*, 2012). The heterozygosity of descended populations is expected to be less than that of founder populations. Examination of the data of Thomas *et al.* (2012) show that the Marañon related clusters have higher H_e values than those of Guiana or Amelonado clusters. Motilal *et al.* (2013) found that the PA as an accession group had more microsatellite alleles and higher H_e than reference Amelonados from Brasil. The Amelonados in Brasil and the Guiana germplasm may therefore be the result of strong bottlenecks aided by human selection from the same initial pool in the centre of origin and diversity in the Western Upper Amazon that generated the PA accessions. This area includes the Amazon region of Peru and is close to the Brazilian Upper Amazon.

4.5. Evolution of the Trinitario Complex

Trinitario types from different countries have been reported (Toxopeus, 1985; Bartley, 2005) and supposedly arose in several countries as natural hybrids of Forastero and Criollo germplasm (Cheesman, 1944; Toxopeus, 1985; Wood, 1985a,b). Separation of Trinitario types from Honduras and Nicaragua have been reported using SNP data (Ji *et al.*, 2013). However, the Trinitarios from Trinidad appear to cluster with the Trinitarios from Nicaragua according to the data presented by Ji *et al.* (2013). Similarly, the Trinitarios from Trinidad clustered among the Trinitarios from Indonesia, from the figures of Susilo *et al.* (2011) and Dinarti *et al.* (2015). The Trinitarios from Trinidad were shown to have a greater contribution of Forastero ancestry than Criollo ancestry with admixture from Upper Amazon and Lower Amazon according to SSR analyses (Motilal *et al.*, 2010). This was later corroborated by Yang *et al.* (2013), using chloroplast SSRs and SNPs, to establish that three lineages (Upper Amazon Forastero, Lower Amazon Forastero and Criollo) were present in the Trinitario complex in Trinidad.

4.6. New Populations and Landraces

Genetic ancestry in cocoa has been assessed using the software 'Structure' (Pritchard *et al.*, 2000; Hubisz *et al.*, 2009) with SSR and, recently, SNP datasets in a model-based clustering method. Since Motamayor *et al.* (2008), there has been little additional insight on cocoa ancestry although Zhang *et al.* (2011) described a new

population from Bolivia. Germplasm collections need to be assessed for genetic structure after identity issues have been clarified. This enables curators to obtain empirical *a priori* proof by genetic information on the representation of the known existing cocoa populations within the germplasm collection. This aspect has been addressed in germplasm collections in Cameroun (Efombagn *et al.*, 2008), Costa Rica (Zhang *et al.*,.2009b), Dominican Republic (Boza *et al.*, 2013), Puerto Rico (Irish *et al.*, 2010), and Trinidad (Zhang *et al.*, 2007, 2009a; Johnson *et al.*, 2009; Motilal *et al.*, 2010, 2012).

Farmer varieties or landraces are also commonly studied for their ancestral contributions. Fine flavour cocoa originates from Criollo (Toxopeus, 1985) and Nacional backgrounds (Loor Solorzano *et al.*, 2012). Flavour sensory attributes that could be influenced by genetic makeup include floral and dry fruit flavour (Ziegleder, 1990; Chanliau and Cros, 1999; Lanaud *et al.*, 2005; Trognitz *et al.*, 2013). Due to the historic movement of germplasm there have been repeated introductions into many countries from few source countries, multiple subsequent transfers between countries within a region, and more recently movement of germplasm from genebanks via plant quarantine centres. These multiple introductions have resulted in diverse ancestral backgrounds and admixed individuals in countries with cultivated but not endemic cacao. The genetic ancestry therefore reflects the historical movement of cacao and can be used to retrace the history where historical records do not exist. The population ancestry of farmer varieties have been reported for Brazil (Santos *et al.*, 2015), Cameroon (Efombagn *et al.*, 2008), Dominican Republic (Boza *et al.*, 2013), Honduras and Nicaragua (Ji *et al.*, 2013), Indonesia (Dinarty *et al.*, 2015), Peru (Zhang *et al.*, 2006a, 2011) and Trinidad (Yang *et al.*, 2013).

In order to derive maximum benefit, studies on cacao ancestry need to incorporate reference profiles, either generated from within the study or obtained from open-access deposited information as in the ICGD. This will allow the proper population ancestries to be identified instead of being unknown contributions as in Trognitz *et al.* (2011, 2013). Although, SSR profiles from different platforms are usually different, the allele calls can be made consistent provided that common accessions were scored between platforms. However, this may not always be possible and the availability of SNP data in the ICGD (Turnbull and Hadley, 2012) is therefore a good alternative especially as the SNP data is expected to be more transferable due to non-reliance on size polymorphisms. Accessions allocated to genetic clusters and with SNP data in the ICGD can therefore be mined to typify their genetic ancestry.

4.7. Core Collection

A core collection is a subset of minimal size with maximum representation of the original germplasm collection. Core collections contain maximal variability in 10-15 per cent of the accessions (Frankel, 1984; Frankel and Brown, 1984; Brown, 1989a,b). Van Hintum *et al.* (2000) reported that most core collections contain 5–20 per cent of the original germplasm collection. Although core collections have been identified in various plant collections such as olive (Belaj *et al.*, 2012), pear (Miranda *et al.*, 2010) and Sea Island cotton (Mei *et al.*, 2012), a core collection for cacao is yet to

be adopted. Sounigo *et al.* (2006b) identified a set of 110 accessions in the ICGT based on RAPD and isozyme diversity, disease resistance, and seed traits of importance to the industry. Recently, a core collection of 59 accessions were identified based on the genetic diversity of 24 SSR loci on 414 accessions in the ICGT (Motilal *et al.*, 2013). Santos *et al.* (2015) reported a core collection of 27 plants based on the genetic diversity of 30 SSR loci on 279 varieties from germplasm collection and farmers' fields in Brazil. In the Brazilian core collection, seven plants possessed 70 per cent of the alleles and five of these plants originated from the farmers' fields indicative of the low diversity of the Amelonado accessions in the germplasm collection. However, core collections based on molecular genetic diversity may not adequately represent the agronomic or morphological diversity. Hence, after characterisation of a germplasm collection at the molecular and phenotypic levels, either a composite core collection based on the diversity over both levels can be identified or several core collections can be developed independently from genotypic and phenotypic data.

4.8. Evolutionary Phylogeny

Theobroma was previously placed in the Family Sterculiaceae (Purseglove, 1974), but is now reassigned to Family Malvaceae (Alverson *et al.*, 1999; Bayer *et al.*, 1999). Cuatrecasas (1964) described 22 species of *Theobroma* which are closely allied with the 17 species of *Herrania* Goudot (Whitlock and Baum, 1999). Figueira *et al.* (1994) on the basis of rDNA polymorphism found that all *Herrania* and *Theobroma* species with the exception of *T. cacao* formed a single cluster. Whitlock *et al.* (2001). reported that these two genera formed a single clade based on plastid *ndhF* sequences. Whitlock and Baum (1999), on the basis of sequenced vicilin gene, showed that *Herrania* and *Theobroma* were monophyletic groups but monophyly was more strongly supported for *Herrania* than *Theobroma*. A similar conclusion based on trypsin inhibitor gene sequences was obtained (Silva and Figueira, 2005). Both Whitlock and Baum (1999) and Silva and Figueira (2005) supported the inclusion of *T. mammosum* into the section Glossopetalum rather than as the separate section Andropetalum. In contrast, Borrone *et al.* (2007) obtained much stronger support for the monophyly of *Theobroma* from sequences of five WRKY genes. However, the separation of *Theobroma* into two main clades and the intra-cluster separation of *T. cacao* from its other sister species within the clade was consistently reported (Whitlock and Baum, 1999; Silva and Figueira, 2005; Borrone *et al.*, 2007).

More recently, there has been interest in the DNA barcoding of plants with the aim of identifying and classifying unknown plant species. The barcoding technique involves characterising species based on a DNA segment from a standard and agreed-upon position in the genome. The sequence used is usually short and can be obtained quickly and cheaply. The *trnH-psbA* and ITS spacer regions (Kress *et al.*, 2005) in conjunction with *rbcL* and *matK*, as barcode regions (Fazekas *et al.*, 2012) have been used for species identification in land plants. However, while these regions may distinguish *Theobroma* spp. from each other and from related allies like *Herrania*, *Cola* and *Guazuma*, it has not been established that these barcode regions can reveal intra-species differences. The identification of a set of SNP loci, to generate a unique haplotype per variety, is therefore recommended.

4.9. Genome-Wide Association Studies and Marker Assisted Selection

Extensive and intensive molecular profiling has generated thousands of markers which can be tested for association with a desired trait. The markers so identified have the advantage of being broad-based in application instead of being restricted to a population or populations (Yu and Buckler, 2006). Spurious or false associations are minimised by accounting for population stratification and relatedness (Aranzana *et al.*, 2005; Price *et al.*, 2006; Yu *et al.*, 2006). Association studies have been reviewed and proven useful in understanding the genetic basis for complex traits in many plants (Abdurakhmonov and Abdukarimov, 2008; Zhu *et al.*, 2008; Ingvarsson and Street, 2011; Soto-Cerda and Cloutier, 2012; Gupta *et al.*, 2014). Association mapping is an emergent field in cacao and few studies have been published. Motamayor *et al.* (2013), using association mapping among other techniques, identified a candidate gene from 71 accessions for fruit colour and a single SNP among the 168 that were studied, that affected transcript expressions of the gene.

Markers identified as being strongly associated with a trait may be employed in marker assisted selection programmes. Crouzillat *et al.* (2000b), from simulation studies, demonstrated that, in cacao, the use of molecular markers alone or in combination with phenotypic selection was more effective than phenotypic evaluation only. Reviews on marker-assisted selection have been published (Paterson *et al.*, 1991; Young, 1999; Hospital, 2003; Peleman *et al.*, 2005) and the advantages outlined (Collard *et al.*, 2005). The efficiency of marker assisted versus phenotype assisted selection is higher for traits of low heritability (Collard *et al.*, 2005) but cost effectiveness need to be evaluated for each case (Dreher *et al.*, 2003). Phenotypic evaluations may be time-consuming, difficult or costly (Dreher *et al.*, 2003; Young, 1999; Yu *et al.*, 2000). However, current trends indicate a reducing cost for SNP genotyping which should make marker assisted selection more cost-effective and therefore more favourable than phenotyping. The limited availability of land resources in terms of quantity and tenure issues may also weigh against phenotyping in cacao due to the long vegetative phase and number of years needed to obtain productivity values. Marker assisted selection programmes using quantitative trait loci in cacao have been reported (Schnell *et al.*, 2007a; Schnell *et al.*, 2007b) but those using association mapping are now underway (Motamayor *et al.*, 2010, 2013).

5. Future Prospects

5.1. Methodology

The methodologies for DNA profiling are expected to change as sequencing technologies become more powerful and cheaper, and as researchers develop faster and more efficient bioinformatics tools. On one hand platforms for DNA profiling are becoming more elaborate allowing researchers to rapidly obtain data for 1000's of SNPs across the genome for 1000's of individuals. Advances in DNA hybridisation, florescence microscopy and solid surface DNA capture have allowed for the development of large SNP arrays where 1000's SNPs can be assayed and interpreted in real-time (Ganal *et al.*, 2012, Kim and Misra, 2007). In parallel, there are technologies that seek to improve the efficiency of SNP-based DNA profiling

using nanofluidics (Fang *et al.*, 2014, Wang *et al.*, 2009) or other miniaturised PCR volume systems and using robotic platforms, which together seek to reduce the cost of DNA profiling, improve the throughput rate, while reducing the human effort and time spent. Additionally, application of isothermal DNA amplification (Li and Macdonald, 2015, Zhao *et al.*, 2015) has facilitated the development of a field-based rapid method of DNA fingerprinting of organisms. This coupled with access to on-line databases and sophisticated background data processing will allow for sampling, analysis and data processing to occur in the field.

On the other hand, with the improvement in efficiency and the consequent reduction of the cost of next generation DNA sequencing technologies genotyping-by-sequencing has become a method of choice for DNA profiling. This is particularly useful in organisms for which *a priori* genomic information is not available. This approach also allows for sequencing predetermined areas of the genome for genetic variation using various enrichment methods. The technology also allows for the detection of insertions, deletions and microsatellites in addition to SNPs and considerably reduces ascertainment bias compared to array based methods.

With the success of next generation sequencing, and the availability of accurate genome sequence information there are new possibilities for molecular fingerprinting including the use of transposable elements which are known to contribute to a large proportion of genome variation. Polymorphisms in transposable elements can arise from the potential variability in copy number between and within species (Vicient *et al.*, 1999; Pearce *et al.*, 2000; Huang *et al.*, 2008). Sveinsson *et al.* (2013) found that there was active evolution of transposable elements in cacao and suggested that transposon fingerprinting could be used in identifying and characterising cacao accessions.

As deep sequencing and resequencing technologies become cheaper and more extensively used there is likely to be convergence of methodologies of DNA profiling for various purposes, such as genomic selection, genetic diversity, phylogenetics and evolutionary studies.

5.2 Food Forensics

High quality DNA was detected in fermented and dried beans and partially degraded DNA in roasted nibs and chocolates (Crouzillat *et al.*, 2000a). These authors showed that the cacao used to make chocolate could be traced from bean to product using RAPDs, SSRs and sequencing of a fragment of a seed storage protein gene. The adulteration of fermented and dried cocoa beans as a single bean incident could be detected with 44 SNPs (Fang *et al.*, 2014). The establishment of SNP panels that can distinguish fine flavour cocoa from bulk cocoa, as well as discriminate among fine flavour cocoa from different geographic areas would be a useful certification tool. In addition, molecular profiling can be used to support geographical indicators by supporting the uniqueness of cacao genotypes by field, farm or region; establishing the similarity of genotypes to enable bulk fermentation from members of co-operatives; and to brand genetic origin chocolates.

References

Abdurakhmonov, I.Y. and Abdukarimov, A. (2008). Application of association mapping to understanding the genetic diversity of plant germplasm resources. *International Journal of Plant Genomics*, Article ID 574927, doi:10.1155/2008/574927.

Aikpokpodion, P.O. (2010). Variation in agro-morphological characteristics of cacao, *Theobroma cacao* L., in farmers' fields in Nigeria. *New Zealand Journal of Crop and Horticultural Science*, **38**: 157-170.

Akkaya, M.S., Bhagwat, A.A. and Cregan, P.B. (1992). Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics*, **132**: 1131-1139.

Allegre, M., Argout, X., Boccara, M., Fouet, O., Roguet, Y., Bérard, A., Thévenin, J.-M. *et al.* (2012). Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large scale genetic studies and breeding of *Theobroma cacao* L. *DNA Research*, **19**: 23-35. doi:10.1093/dnares/dsr039.

Alverson, W.S., Whitlock, B. A., Nyffler, R., Bayer, C. and Baum, D.A. (1999). Phylogeny of the core Malvales: evidence from *ndhF* sequence data. *American Journal of Botany*, **86**: 1474-1486.

Aranzana, M. J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., Lister, C., Molitor, J., Shindo, C., Tang, C., Toomajian, C., Traw, B., Zheng, H., Bergelson, J., Dean, C., Marjoram, P. and Nordborg, M. (2005). Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics*, 1: e60. doi:10.1371/*journal.pgen*.0010060.

Argout, X., Fouet, O., Wincker, P., Gramacho, K., Legavre, T., Sabau, X., Risterucci, A.M., *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. doi. 10.1186/1471-2164-9-512.

Argout, X., Salse, J., Aury, J.-M., Guiltinan, M.J., Droc, G., Gouzy, J., Allegre, M., *et al.* (2011). The genome of *Theobroma cacao*. *Nature Genetics*, **43**(2): 101-108.

Atkinson, M., Withers, L.A. and Simpson, M.J.A. (1986). Characterisation of cacao germplasm using isoenzyme markers. *Euphytica*, **35**: 741-750.

Baker, R.P., Hasenstein, K.H. and Zavada, M.S. (1997). Hormonal changes after compatible and incompatible pollination in *Theobroma cacao* L. *HortScience*, **32**: *1231-1997*.

Bartley, B.G.D. (2001). Refractario – an explanation of the meaning of the term and its relationship to the introductions from Ecuador in 1937. *INGENIC Newsletter*, **6**: 10-15.

Bartley, B.G.D. (2005). The genetic diversity of cacao and its utilization. UK: CABI Publishing.

Bayer, C., Fay, M.F., De Bruijn, P.Y., Savolainen, V., Morton, C.M., Kubitzki, K., Alverson, W.S. and Chase, M.W. (1999). Support for an expanded family concept of Malvaceae within a recircumscibed order Malvales: a combined analysis of plastid atpB and rbcL DNA sequences. *Botanical Journal of the Linnaean Society*, **129**: 267-303.

Beissinger, T.M., Hirsch, C.N., Vaillancourt, B., Deshpande, S., Barry, K., Buell, C.R., Kaeppler, S.M., Gianola, D. and de Leon, N. (2014). A genome-wide scan for evidence of selection in a maize population under long-term artificial selection for ear number. *Genetics*, **196**: 829-840.

Bekele, F. and Bekele, I. (1996). A sampling of the phenetic diversity of cacao in the International Cocoa Gene Bank of Trinidad. *Crop Science*, **36**: 57-64.

Bekele, F.L., Bekele, I., Butler, D.R. and Bidaisee, G.G. (2006). Patterns of morphological variation in a sample of cacao (*Theobroma cacao* L.) germplasm from the International Cocoa Genebank, Trinidad. *Genetic Resources and Crop Evolution*, **53**: 933-948.

Bekele, F.L., Bidaisee, G.G., Persad, N. and Bhola, J. (2005). Examining phenotypic relationships among Upper Amazon Forastero clones. In: Annual Report 2004, St. Augustine, Trinidad and Tobago. *Cocoa Research Unit*, The University of the West Indies. pp. 27-42.

Bekele, F.L., Butler, D.R. and Bidaisee, G.G. (2008a). Upper Amazon Forastero cacao (*Theobroma cacao* L.) 1: An assessment of phenotypic relationships in the International Cocoa Genebank, Trinidad. *Tropical Agriculture (Trinidad)*, **85**: 1-15.

Bekele, F.L., Iwaro, A.D., Butler, D.R. and Bidaisee, G.G. (2008b). Upper Amazon Forastero cacao (*Theobroma cacao* L.) 2: An overview of Parinari clones from a breeder's perspective. *Tropical Agriculture (Trinidad)*, **85**: 16-33.

Bekele, F.L., Kennedy, A.J., Mc David, C., Lauckner, F.B. and Bekele, I. (1994).) Numerical taxonomic studies on cacao (*Theobroma cacao* L.) in Trinidad. *Euphytica*, **75**: 231-240.

Belaj, A., Dominguez-García, M.C., Atienza, S.G., Urdíroz, N.M., de la Rosa, R., Satovic, Z., Martín, A. *et al.* (2012). Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. *Tree Genetics and Genomes*, **8**: 365-378.

Bélo, A. Zheng, P., Luck, S., Shen, B., Meyer, D.J., Li, B., Tingey, S. and Rafalski, A. (2008). Whole genome scans detects an allelic variant of *fad2* associated with increased oleic acid levels in maize. *Molecular Genetics and Genomics*, **279**(1): 1-10.

Boccara, M. and Zhang, D. (2006). Progress in resolving identity issues among the Parinari accessions held in Trinidad: the contribution of the collaborative USDA/CRU project. In: Annual Report 2005, St. Augustine, Trinidad. *Cocoa Research Unit*, The University of the West Indies. pp. 25-32.

Boccara, M. and Zhang, D. (2007). The contribution of the collaborative USDA/CRU project to resolve identity issues for trees in Marper Farm with missing labels. In: Annual Report 2006, St. Augustine, Trinidad. Cocoa Research Unit, The University of the West Indies. pp. 25-32.

Boccara, M. and Zhang, D. (2008). Identity assessment of Refractario origin cocoa accessions held in Trinidad: the contribution of the collaborative USDA/CRU project. In: Annual Report 2007, St. Augustine, Trinidad. Cocoa Research Unit, The University of the West Indies. pp. 30-36.

Borrone, J. W., Kuhn, D.N. and Schnell, R.J. (2004). Isolation, characterization, and development of WRKY genes as useful genetic markers in *Theobroma cacao*. *Theoretical and Applied Genetics*, **109**: 495-507.

Borrone, J.W., Meerow, A.W., Kuhn, D.N., Whitlock, B.A. and Schnell, R.J. (2007). The potential of the WRKY gene family for phylogenetic reconstruction: An example from the Malvaceae. *Molecular Phylogenetics and Evolution*, 44: 1141-1154.

Boza, E.J., Irish, B.M., Meerow, A.W., Tondo, C.L., Rodríguez, O.A., Ventura-López, M., Gómez, J.A., Moore, J.M., *et al.* (2013). Genetic diversity, conservation, and utilization of *Theobroma cacao* L.: Genetic resources in the Dominican Republic. *Genetic Resources and Crop Evolution*, **60**: 605-619.

Brown, A.H.D. (1989a). A case for core collections. In: Brown, A.H.D, Frankel, O.H., Marshall, D.R. and Williams, J.T. (Eds.), *The Use of Plant Genetic Resources*. Cambridge University Press, Cambridge. pp. 136–156.

Brown, A.H.D. (1989b). Core collections: A practical approach to genetic resources management. *Genome*, **31**: 818-824.

Brown, J.S., Phillips-Mora, W., Power, E.J., Krol, C., Cervantes-Martinez, C., Motamayor, J.C. and Schnell, R.J. (2007). Mapping QTLs for resistance to frosty pod and black pod diseases and horticultural traits in *Theobroma cacao* L. *Crop Science*, **47**: 1851-1858.

Brown, J.S., Schnell, R.J., Motamayor, J.C., Lopes, U., Kuhn, D.N. and Borrone, J.W. (2005). Resistance gene mapping for witches' broom disease in *Theobroma cacao* L. in a F2 population using SSR markers and candidate genes. *Journal of the American Society for Horticultural Science*, **130**: 366-373.

Brumfield, R.T., Beerli, P., Nickerson, D.A. and Edwards, S.V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution*, **18**: 249-256.

Buckler, E.S., Holland, J.B., Bradbury, P.J., Acharya, C.B., Brown, P.J., Browne, C., Ersoz, E., Flint-Garcia, S., *et al.* (2009). The genetic architecture of maize flowering time. *Science*, **325**(5941): 714-718.

Buckler, E.S. and Thornsberry, J. (2002). Plant molecular diversity and applications to genomics. *Current Opinions in Plant Biology*, **5**: 107-111.

Budowle, B., Chakraborty, R., Carmody, G. and Monson, K.L. (2000). Source attribution of a forensic DNA profile. *Forensic Science Communications*, **2**(3): 6 pp. http://www.fbi.gov/hq/lab/fsc/backissu/july2000/source.htm (accessed July 07, 2008).

Caetano-Anolles, G. and Brant, B. (1991). DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Nature Biotechnology*, **9**: 553-557.

Chanliau, S. and Cros, E. (1999). Influence du traitement post-récolte et de la torréfaction sue le développement de l'arôme cacao. In: Proceedings of the 12th *International Cocoa Research Conference*, 17-23 November, 1996, Salvador, Bahia, Brazil. COPAL, Malaysia. pp. 959-964.

Charters, M.Y. and Wilkinson, M.J. (2000). The use of self-pollinated progenies as "in-groups" for the genetic characterization of cocoa germplasm. *Theoretical and Applied Genetics*, **100**: 160-166.

Cheesman, E.E. (1944). Notes on the nomenclature, classification and possible relationships of cacao populations. *Tropical Agriculture (Trinidad)*, **21**: 144-159.

Cho, R.J., Mindrinos, M., Richards, D.R., Sapolsky, R.J., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T.L., Stammers, M., Federspiel, N., *et al.* (1999). Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nature Genetics*, **23**: 203-207.

Chowdappa, P. and Chandramohanan, R. (1995). Electrophoretic protein patterns of three species of *Phytophthora* associated with black pod disease of cocoa (*Theobroma cacao* L.). *Journal of Biosciences*, **20**: 637-644.

Christopher, Y. (1993). The reproducibility of DNA polymorphisms in *Theobroma cacao* L. as exhibited using RAPD analysis. In: Report for 1993. *Cocoa Research Unit*, The University of the West Indies. pp. 26-31.

Christopher, Y., Mooleedhar, V., Bekele, F. and Hosein, F. (1998). Verification of accession in the ICG,T using botanical descriptors and RAPD analysis. In: Report for 1998. *Cocoa Research Unit*, The University of the West Indies. pp. 15-18

Christopher, Y. and Sounigo, O. (1995). The use of RAPD for characterization and genetic assessment of cacao. In: Report for 1995. Cocoa Research Unit, The University of the West Indies. pp. 38-51

Clément, D., Risterucci, A.M., Motamayor, J.C., N'Goran, J. and Lanaud, C. (2003a). Mapping QTL for yield components, vigor and resistance to *Phytophthora palmivora* in *Theobroma cacao* L. Genome, **46**: 204-212.

Clément, D., Risterucci, A.M., Motamayor, J.C., N'Goran, J. and Lanaud, C. (2003b). Mapping quantitative trait loci for bean traits and ovule number in *Theobroma cacao* L. *Genome*, **46**: 103-111.

Coe, S. D. and Coe, M. D. (1996). The true history of chocolate. London: Thames and Hudson, Ltd.

Collard, B.C.Y., Jahufer, M.Z.Z., Brouwer, J.B. and Pang, E.C.K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, **142**: 169-196.

Cope, F.W. (1958). Incompatibility in *Theobroma cacao*. *Nature*, **181**: 279.

Cope, F.W. (1962). The mechanism of pollen incompatibility in *Theobroma cacao* L. *Heredity*, **17**: 157-182.

Crouzillat, D., Bellanger, L., Rigoreau, M., Bucheli, P. and Pétiard, V. (2000a). Use of DNA markers for determination of the genetic origin of cacao in fermented or roasted beans and chocolate. In: *Proceedings of the 13th International Cocoa Research Conference*, 9-14 October 2000, Kota Kinabalu, Malaysia. COPAL, Lagos, Nigeria. pp. 71-78.

Crouzillat, D., Menard, B., Mora, A., Phillips, W. and Petiard, V. (2000b). Quantitative trait analysis in *Theobroma cacao* using molecular markers. Yield QTL detection and stability over 15 years. *Euphytica*, **114**: 13-23.

Crouzillat, D., Lerceteau, E., Petiard, V., Morera, J., Rodriguez, H., Walker, D., Phillips, W., Ronning, C., Schnell, R., Osei, J. and Fritz, P. (1996). *Theobroma cacao* L.: A genetic linkage map and quantitative trait loci analysis. *Theoretical and Applied Genetics*, **93**: 205-214.

Cryer, N.C., Fenn, M.G.E., Turnbull, C.J. and Wilkinson, M.J. (2006). Allelic size standards and reference genotypes to unify international cocoa (*Theobroma cacao* L.) microsatellite data. *Genetic Resources and Crop Evolution*, **53**: 1643-1652.

Cuatrecasas, J. (1964). Cacao and its allies. A taxonomic revision of the genus *Theobroma*. In Contributions to the U.S. National Herbarium, **35**(6): 375-614. Washington, DC: Smithsonian Institution.

Dadzie, A.M., Livingstone III, D.S., Opoku, S.Y., Takrama, J., Padi, F., Offei, S.K., Danquah, E. Y., Motamayor, J.C., Schnell, R.J. and Kuhn, D.N. (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.

Dinarti, D., Susilo, A.W., Meinhardt, L.W., Ji, K., Motilal, L.A., Mischke S. and Zhang D. (2015). Genetic diversity and parentage in farmer selections of cacao from Southern Sulawesi, Indonesia revealed by microsatellite markers. *Breeding Science*, **65**: 438–446. doi:10.1270/jsbbs.65.438

Dreher, K., Khairallah, M., Ribaut, J. and Morris, M. (2003). Money matters (I): Costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT. *Molecular Breeding*, 11: 221-234.

Efombagn, I.B.M., Motamayor, J.C., Sounigo, O., Eskes, A.B., Nyassé, S., Cilas, C., Schnell, R., Manzanares-Dauleux, M.J. and Kolesnikova-Allen, M. (2008). Genetic diversity and structure of farm and GenBank accessions of cacao (*Theobroma cacao* L.) in Cameroon revealed by microsatellite markers. *Tree Genetics and Genomes*, **4**: 821-823.

Engels, J.M.M. (1983a). A systematic description of cacao clones I. The discriminative value of quantitative characteristics. *Euphytica*, **32**: 377-385.

Engels, J.M.M. (1983b). A systematic description of cacao clones II. The discriminative value of qualitative characteristics and the practical compatibility of the discriminative value of quantitative and qualitative descriptors. *Euphytica*, **32**: 387-396.

Engels, J.M.M., Bartley, B.G.D. and Enriquez, G.A. (1980). Cacao descriptors, their states and modus operandi. *Turrialba*, **30**: 209-218.

Enríquez, C.G. and Soria, V.J. (1966) Estudio de la variabilidad de varias características de las mazorcas de cacao (*Theobroma cacao* L.). *Fitotecnica Latinoamericana*, **3**: 99-118.

Enríquez, C.G. and Soria, V.J. (1967) Selección y estudio de los carácteres útiles de la flor para la identificación y descripción de cultivares de cacao. *Cacao (Costa Rica)*, **12**: 8-16.

Erselius, L.J. and De Vallavieille, C. (1984). Variation in protein profiles of *Phytophthora*: comparison of six species. *Transactions of the British Mycological Society*, **83**: 463-472.

Evett, I.W. and Weir, B.S. (1998). Interpreting DNA evidence: statistical genetics for forensic scientists. Sunderland, MA: Sinauer.

Fang, W., Meinhardt, L.W., Mischke, S., Bellato, C.M., Motilal, L. and Zhang, D. (2014). Accurate determination of genetic identity for a single cacao bean, using molecular markers with a nanofluidic system, ensures cocoa authentication. *Journal of Agricultural and Food Chemistry*, **62**: 481-487 doi: 10.1021/jf404402v, Web: 19 Dec 2013

Fazekas, A.J., Kuzmina, M.L., Newmaster, S.G. and Hollingsworth, P.M. (2012). DNA barcoding methods for land plants. In W. J. Kress and D. L. Erickson (eds.), DNA Barcodes: *Methods and Protocols, Methods in Molecular Biology*, **858**, doi. 10.1007/978-1-61779-591-6_11.

Feltus, F.A., Wan, J., Schulze, S.R., Estill, J.C., Jiang, N. and Paterson, A.H. (2004). An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. *Genome Research*, **14**: 1812-1819.

Figueira, A. (1998). Homonymous genotypes and misidentification in germplasm collections of Brazil and Malaysia. INGENIC Newsletter, **4**: 4-8. http://ingenic.cas.psu.edu/newsletters.htm (accessed July 07, 2008).

Figueira, A., Janick, J., Levy, M. and Goldsbrough, P. (1994).) Reexamining the classification of *Theobroma cacao* L. using molecular markers. *Journal of the American Society for Horticultural Science*, **119**: 1073-1082.

Ford, C.S. and Wilkinson, M.J. (2012). Confocal observations of late-acting self-incompatibility in *Theobroma cacao* L. *Sexual Plant Reproduction*, **25**: 169-183. doi:10.1007/s00497-012-0188-1.

Frankel, O. H. (1984). Genetic perspectives of germplasm conservation. In Genetic manipulation: impact on man and society Part III, Paper No. 15, edited by W. K. Arber, K. Llimensee, W. J. Peacock and P. Starlinger, 161-170. Cambridge: Cambridge University Press.

Frankel, O.H. and Brown, A.H.D. (1984). Plant genetic resources today: a critical appraisal. In Crop Genetic Resources: Conservation and Evaluation, edited by J. H. W. Holden and J. T. Williams, 249–257. London: George Allen and Unwin.

Ganal, M.W., Altmann, T. and Röder, M.S. (2009). SNP identification in crop plants. *Current Opinion in Plant Biology*, **12**: 211-217.

Ganal, M.W., Polley, A., Graner, E.-M., Plieske, J., Wieseke, R., Luerssen, H. and Durstewitz, G. (2012). Large SNP arrays for genotyping in crop plants. *Journal of Biosciences*, **37**: 821-828.

Glaubitz, J.C., Rhodes, O.E. and DeWoody, J.A. (2003). Prospects for inferring pairwise relationships within single nucleotide polymorphisms. *Molecular Ecology*, **12**: 1039-1047.

Gupta, P.K., Kulwal, P.L. and Jaiswal, V. (2014). Association mapping in crop plants: Opportunities and challenges. *Advances in Genetics*, **85**: 109-147.

Henderson, J.S., Joyce, R.A., Hall, G.R., Hurst, W.J. and McGovern, P.E. (2007). Chemical and archaeological evidence for the earliest cacao beverages. *Proceedings of the National Academy of Sciences USA*, **104**: 18937-18940.

Hodgkinson, A. and Eyre-Walker, A. (2010). Human triallelic sites: Evidence for a new mutational mechanism? *Genetics*, **184**: 233-241. doi: 10.1534/genetics.109.110510

Hospital, F. (2003). Marker-assisted breeding. In: *Plant Molecular Breeding*, edited by H. J. Newbury, 30-59. Oxford, UK: CRC Press, Blackwell.

Huang, X., Lu, G., Zhao, Q., Liu, X. and Han, B. (2008). Genome-wide analysis of transposon insertion polymorphisms reveals intraspecific variation in cultivated rice. *Plant Physiology*, **148**: 25-40.

Hubisz, M.J., Falush, D., Stephens, M. and Pritchard, J.K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**: 1322-1332.

Hüebner, C., Petermann, I., Browning, B.L., Shelling, A.N. and Ferguson, L.R. (2007). Triallelic single nucleotide polymorphisms and genotyping error in genetic epidemiology studies: MDR1 (ABCB1) G2677/T/A as an example. *Cancer Epidemiology, Biomarkers and Prevention*, **16**(6): 1185-1192. doi: 10.1158/1055-9965.EPI-06-0759.

Hurka, H., Neuffer, B. and Friesen, N. (2004). Plant genetic resources in botanical gardens. In: Proceedings of the 21st International Symposium on Classical vs. Molecular Breeding of Ornamentals, Part II 25-29 August, 2003, edited by G. Forkmann and S. Michaelis. *Acta Horticulturae*, **651**: 35-44.

Ingvarsson, P.K. and Street, N.R. (2011). Association genetics of complex traits in plants. New Phytologist, **189**: 909–922. doi: 10.1111/j.1469-8137.2010.03593.x.

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

Iwaro, A.D., Bekele, F.L. and Butler, D.R. (2003). Evaluation and utilisation of cacao (*Theobroma cacao* L.) germplasm at the International Cocoa Genebank, Trinidad. *Euphytica*, **130**: 207-221.

Jenkins, P.A., Mueller, J.W. and Song, Y.S. (2014). General triallelic frequency spectrum under demographic models with variable population size. *Genetics*, **196**: 295-311.

Ji, K., Zhang, D., Motilal, L.A., Boccara, M., Lachenaud, P. and Meinhardt, L.W. (2013). Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. *Genetic Resources and Crop Evolution*, **60** (2): 441-453.

Johnson, E.S., Bekele, F.L., Brown, S.J., Song, Q., Zhang, D., Meinhardt, L.W. and Schnell, R.J. (2009). Population structure and genetic diversity of the Trinitario cacao (*Theobroma cacao* L.) from Trinidad and Tobago. *Crop Science*, **49**: 564-572. doi:10.2135/cropsci2008.03.0128.

Kaosiri, G. and Zentmyer, G.A. (1980). Protein, esterase and peroxidase patterns in the *Phytophthora palmivora* complex from cacao. *Mycologia*, **72**: 998-1000.

Kasran, R. and Subali, S. (2002). Classification of *Theobroma cacao* L. germplasm using random amplified polymorphic DNA. In: *Proceedings of the 13th International Cocoa Research Conference*, 9-14 October 2000, Kota Kinabalu, Malaysia. COPAL, Lagos, Nigeria. pp. 229-235.

Kim, S. and Misra, A. (2007). SNP genotyping: technologies and biomedical applications. *Annual Review of Biomedical Engineering*, **9**: 289-320.

Knight, R. and Rogers, H.H. (1953). Sterility in *Theobroma cacao* L. *Nature*, **172**: 164.

Knight, R. and Rogers, H.H. (1955). Incompatibility in *Theobroma cacao* L., **9**: 69-77.

Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T. and Yano, M. (2006). An SNP caused loss of seed shattering during rice domestication. *Science*, **312**(5778): 1392-1396.

Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. and Janzen, D.H. (2005). Use of DNA barcodes to identify flowering plants. *Proceedings of The National Academy of Sciences (USA)*, **102**(23): 8389-8374.

Kuhn, D.N., Borrone, J., Meerow, A.W., Motamayor, J.C., Brown, J.S. and Schnell, R.J. (2005). Single-strand conformation polymorphism analysis of candidate genes for reliable identification of alleles by capillary array electrophoresis. *Electrophoresis*, **26**: 112-125.

Kuhn, D.N., Figueira, A., Lopes, U., Motamayor, J.C., Meerow, A.W., Cariaga, K., Freeman, B., Livingstone III, D.S. and Schnell, R.J. (2010). Evaluating *Theobroma grandiflorum* for comparative genomic studies with *Theobroma cacao*. *Tree Genetics and Genomes*, **6**: 783-792.

Kuhn, D.N., Livingstone III, D., Main, D., Zheng, P., Saski, C., Feltus, F.A., Mockaitis, K., Farmer, A.D., May, G.D., Schnell, R.J. and Motamayor, J.C. (2012). Identification and mapping of conserved ortholog set (COS) II sequences of cacao and their conversion to SNP markers for marker-assisted selection in *Theobroma cacao* and comparative genomics studies. *Tree Genetics and Genomes*, **8**: 97-111.

Kujur, A., Bajaj, D., Upadhyaya, H.D., Das, S., Ranjan, R., Shree, T., Saxena, M.S., Badoni, S., Kumar, V., Tripathi, S., Gowda, C.L.L., Sharma, S., Singh, S., Tyagi, A.K. and Parida, S.K. (2015). Employing genome-wide SNP discovery and genotyping strategy to extrapolate the natural allelic diversity and domestication patterns in chickpea. *Frontiers in Plant Science*, 6 article 162. doi: 10.3389/fpls.2015.00162.

Kumar, L.S. (1999). DNA markers in plant improvement: An overview. *Biotechnology Advances*, **17**: 143-182.

Kurt, W., Weigand, F., Driesel, A.J., Kahl, G., Zischler, H. and Epplen, J.T. (1989). Polymorphic simple GATA/GACA repeats in plant genomes. *Nucleic Acids Research*, **17**: 10128-10128.

Lanaud, C. (1986). Genetic studies of *Theobroma cacao* L. with the help of enzymatic markers. I. Genetic control and linkage of nine enzymatic markers. *Café Cacao Thé*, **30**(4): 259-270.

Lanaud, C. and Berthaud J. (1985). Mise en evidence de nouveaux marqueurs genetiques chez *Theobroma cacao* L. Par les techniques d'electrophese. In: *Proceedings of the 9th International Cocoa Research Conference*, 12-18 February, 1984, Lomé, Togo. COPAL, Lagos Nigeria. pp. 249-253.

Lanaud, C., Boult, E., Clapperton, J., N'Goran, J.N.K., Cros, E., Chapelin, M., Risterucci, A.M., Allaway, D., Gillmore, M., Cattaruzza, A., Fouet, O., Clement, D. and Petithuguenin, P. (2005). Identification of QTLs related to fat content, seed size and sensorial traits in *Theobroma cacao* L. In: Proceedings of the 14th International Cocoa Research Conference, 13-18 October, 2003, *Cocoa Producers' Alliance*, Malaysia. pp. 1120-1126.

Lanaud, C., Clément, D., Flament, M.H., Risterucci, A.M., Kébé, I., Nyassé, S., Sounigo, O., Motilal, L.A., Thévenin, J.M., Paulin, D., Ducamp, M., N'Goran, J.K.A., Fargeas, D., Cilas, C. (2004). Genetic mapping of quantitative trait loci for black pod resistance in cocoa. In: Cilas,C. and Despréaux, D. (Eds.), *Improvement of cocoa tree resistance to Phytophthora diseases*. CIRAD, France. pp. 147-164.

Lanaud, C., Fouet, O., Clement, D., Boccara, M., Risterucci, A. M., Surujdeo-Maharaj, S., Legavre, T. and Argout, X. (2009). A meta-QTL analysis of disease resistance traits of *Theobroma cacao* L. *Molecular Breeding*, **24**: 361-374.

Lanaud, C., Motamayor, J.C. and Risterucci, A.-M. (2001). Implications of new insight into the genetic structure of *Theobroma cacao* L. for breeding strategies. In *Proceedings of the International Workshop on New Technologies for Cocoa Breeding*, 89: 107, Kota Kinabalu, Malaysia: London: Ingenic Press

Lanaud, C., Risterucci, A.M., N'Goran, J.A.K., Clement, D., Flament, M.H., Laurent, V. and Falque, M. (1995). A genetic linkage map of *Theobroma cacao* L. *Theoretical and Applied Genetics*, **91**: 987-993.

Lanaud, C., Risterucci, A-M., Piretti, I., Falque, M., Bouet, A. and Lagoda, P.J.L. (1999). Isolation and characterization of microsatellites in *Theobroma cacao* L. *Molecular Ecology*, **8**: 2141-2143.

Laurent, V., Risterucci, A.M. and Lanaud, C. (1993). Variability for nuclear ribosomal genes within *Theobroma cacao*. *Heredity*, **71**: 96-103.

Laurent, V., Risterucci, A.M. and Lanaud, C. (1994). Genetic diversity in cacao revealed by cDNA probes. *Theoretical and Applied Genetics*, **88**: 193-198.

Leal, J. B., dos Santos, L. M., dos Santos, C. A. P., Pires, J. L., Ahnert, D. and Correa, R. X. (2008). Genetic diversity among farm and germplasm bank accessions of cacao in Bahia, Brazil. Diversidade genetica entre acessos de cacau de fazendas e de banco de germoplasma na Bahia. *Pesquisa Agropecuaria Brasileira*, **43**(7): 851-858.

Lerceteau, E., Robert, T., Pétiard, V. and Crouzillat, D. (1997). Evaluation of the extent of genetic variability among *Theobroma cacao* accessions using RAPD and RFLP markers. *Theoretical and Applied Genetics*, **95**: 10-19.

Li, J. and Macdonald, J. (2015). Advances in isothermal amplification: novel strategies inspired by biological processes. *Biosensors and Bioelectronics*, **64**: 196-211.

Li, Y.-C., Korol, A. B., Fahima, T., Beiles, A. and Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, **11**(12): 2453-2465.

Lima, L. S., Gramacho, K. P., Carels, N., Novais, R., Gaiotto, F. A., Lopes, U. V., Gesteira, A. S., Zaidan, H. A., Cascardo, J. C. M., Pires, J. L. and Micheli, F. (2009). Single nucleotide polymorphisms from *Theobroma cacao* expressed sequence tags associated with witches' broom disease in cacao. *Genetics and Molecular Research*, **8**(3): 799-808.

Liu, K. and Muse, S. V. (2005). PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics, **21**: 2128-2129. V 3.25 http://www.powermarker.net (accessed November 07, 2012).

Livingstone III, D.S., Freeman, B., Motamayor, J. C., Schnell, R. J., Royaert, S., Takrama, J., Meerow, A. W. and Kuhn, D.N. (2012). Optimizing of a SNP assay for genotyping *Theobroma cacao* under field conditions. *Molecular Breeding*, 30: 33-52.

Livingstone, D. S., Motamayor, J. C., Schnell, R. J., Cariaga, K., Freeman, B., Meerow, A. W., Brown, J. S. and Kuhn, D. N. (2011). Development of single nucleotide polymorphism markers in *Theobroma cacao* and comparison to simple sequence repeat markers for genotyping of Cameroon clones. *Molecular Breeding*, **27**: 93-106.

Livingstone, D., Royaert, S., Stack, C., Mockaitis, K., May, G., Farmer, A., Saski, C., Schnell, R., Kuhn, D. and Motamayor, J. C. (2015). Making a chocolate chip: development and evaluation of a 6K SNP array for *Theobroma cacao*. *DNA Research*, **22**(4): 279-291.

Lockwood, G. and End, M. (1993). History, technique and future needs for cacao collection. In: Proceedings of International Workshop on Conservation, Characterization and Utilization of Cocoa Genetic Resources in the 21st Century, 13-17 September 1992, Port-of-Spain, Trinidad. *The Cocoa Research Unit*, Trinidad. pp 1-14.

Lockwood, G. and Gyamfi M. M. O. (1978).The CRIG cocoa germplasm collection with notes on the codes used in the breeding programme at Tafo and elsewhere. Technical Bulletin No. 10: Tafo: *Cocoa Research Institute*.

Loor Solorzano, R. G., Fouet, O., Lemainque, A., Pavek, S., Boccara, M., Argout, X., Amores, F., Courtois, B., Risterucci, A. M. and Lanaud, C. (2012). Insight into the wild origin, migration and domestication history of the fine flavour Nacional *Theobroma cacao* L. variety from Ecuador. *PLoS ONE*, 7(11): e48438. doi:10.1371/*journal.pone*.0048438.

MacConaill, L. E., Aldred, M. A., Lu, X. and LaFramboise, T. (2007). Toward high-throughput SNP genotyping in the presence of inherited copy number variation. *BMC Genomics*, **8**: 211. doi: 10.1186/1471-2164/8/211

Maharaj, K., Maharaj, P., Bekele, F. L., Ramnath, D., Bidaisee, G. G., Bekele, I., Persad, C., Jennings, K. and Sankar, R. (2011). Trinidad selected hybrids: An investigation of the phenotypic and agro-economic traits of 20 selected cacao cultivars. *Tropical Agriculture (Trinidad)*, **88**: 175-185.

Marcano, M., Morales, S., Hoyer, M. T., Courtois, B., Risterucci, A. M., Fouet, O. and Pugh, T. (2009). A genome-wide admixture mapping study for yield factors and morphological traits in a cultivated cocoa (*Theobroma cacao* L.) population. Tree Genetics and Genomes, **5**: 329–337. Mei, Y., Zhou, J., Xu, H. and Zhu, S. (2012). Development of Sea Island cotton (*Gossypium barbadense* L.) core collection using genotypic values. *Australian Journal Crop Science*, **6**: 673-680.

Miranda, C., Urrestarazu, J., Santesteban, L. G., Royo, J. B. and Uribina, V. (2010). Genetic diversity and structure in a collection of ancient Spanish pear cultivars assessed by microsatellite markers. *Journal of the American Society Horticultural Science*, **135**: 428-437.

Motamayor, J., Brown, J., Kuhn, D. and Schnell, R. (2010). Linkage disequilibrium in *Theobroma cacao* L. populations. *Abstract submitted to Plant and Animal Genome Conference*, San Diego, California January 2010. Retrieved 17 December 2012 from http://www.sparc.ars.usda.gov/research/publications/publications.htm?seq_no_115=259537 and pf=1

Motamayor, J. C., Lachneaud, P., da Silva e Mota, J. W., Loor, R., Kuhn, D. N., Brown, J. S. and Schnell, R. J. (2008). *Geographic and genetic population differentiation of the Amazonian chocolate tree* (*Theobroma cacao* L). *PLoS ONE*, **3**(10): e3311. doi:10.1371/journal.pone.0003311.

Motamayor, J. C., Mockaitis, K., Schmutz, J., Haiminen, N., Livingstone III, D., Cornejo, O., Findley, S. D., Zheng, P., *et al.* (2013). The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod colour. *Genome Biology*, **14**: r53 doi:10.1186/gb-2013-14-6-r53.

Motilal, L.A. (2004a) Potential of cacao microsatellites amplification across diverse plant taxa. In: Thangadurai, D., Pullaiah, T. and Balatti, P. A. (Eds.), *Genetic Resources and Biotechnology* (*Vol. 2*). Regency Publications, New Delhi. pp. 24-49.

Motilal, L.A. (2004b). Tree identification by simple sequence repeats: a synopsis for 2000-2003. In Annual Report 2003, 13-21. St. Augustine, Trinidad and Tobago: Cocoa Research Unit, The University of the West Indies.

Motilal, L. and Butler, D. (2003). Verification of identities in global cacao germplasm collections. *Genetic Resources and Crop Evolution*, **50**: 799-807.

Motilal, L.A., Zhang, D., Mischke, S., Meinhardt, L.W. and Umaharan, P. (2013). Microsatellite-aided detection of genetic redundancy improves management of the International Cocoa Genebank, Trinidad. *Tree Genetics and Genomes*, **9**: 1395-1411.

Motilal, L.A., Zhang, D., Umaharan, P., Boccara, M., Mischke, S., Sankar, A. and Meinhardt, L. W. (2012). Elucidation of genetic identity and population structure of cacao germplasm within an international cacao genebank. Plant Genetic Resources: Characterization and Utilization, **10**: 232-241.

Motilal, L.A., Zhang, D., Umaharan, P., Mischke, S., Boccara, M. and Pinney, S. (2009). Increasing accuracy and throughput in large-scale microsatellite fingerprinting of cacao field germplasm collections. *Tropical Plant Biology*, **2**: 23-37.

Motilal, L.A., Zhang, D., Umaharan, P., Mischke, S., Mooleedhar, V. and Meinhardt, L.W. (2010). The relic Criollo cacao in Belize – genetic diversity and relationship with Trinitario and other cacao clones held in the International Cocoa Genebank Trinidad. *Plant Genetic Resources: Characterization and Utilization*, **8**(2): 106-115.

Motilal, L.A., Zhang, D., Umaharan, P., Mischke, S., Pinney, S. and Meinhardt, L.W. (2011). Microsatellite fingerprinting in the International Cocoa Genebank Trinidad: accession and plot homogeneity information for germplasm management. *Plant Genetic Resources: Characterization and Utilization*, **9**: 430-438. doi:10.1017//S1479/26211100058X

Nasu, S., Sukuzi, J., Ohta, R., Hasegawa, K., Yui, R., Kitazawa, N., Monna, L. and Minobe, Y. (2002). Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa, Oryza rufipogon*) and establishment of SNP markers. *DNA Research*, **9**: 163-171.

N'Goran, J.A.K., Laurent, V., Risterucci, A.M. and Lanaud, C. (1994).) Comparative genetic diversity studies of *Theobroma cacao* L. using RFLP and RAPD markers. *Heredity*, **73**: 589-597.

Opoku, S.Y., Bhattacharjee, R., Kolesnikova-Allen, M., Motamayor, J.C., Schnell, R., Ingelbrecht, I., Enu-Kwesi, L. and Adu-Ampomah, Y. (2007). Genetic diversity in cocoa (*Theobroma cacao* L.) germplasm collection from Ghana. *Journal of Crop Improvement*, **20**: 73-87.

Oudemans, P. and Coffey, M.D. (1991a) Isozyme comparison within and among worldwide sources of three morphologically distinct species of *Phytophthora*. *Mycological Research*, **95** (1): 19-30.

Oudemans, P. and Coffey, M. D. (1991b) A revised systematics of twelve papillate *Phytophthora* species based on isozyme analysis. *Mycological Research*, **95**(9): 1025-1046.

Oudemans, P. and Coffey, M.D. (1991c) Relationships between *Phytophthora* species: evidence from isozyme analysis. In: Lucas, A., Shattock, R.C., Shaw D.S. and Cooke, L.R. (Eds.), *Phytophthora* (Eds.), Cambridge University Press, United Kingdom. pp. 184-203

Oudemans, P., Forster, H. and Coffey, M.D. (1994).) Evidence for distinct isozyme subgroups within *Phytophthora citricola* and close relationships with *P. capsici* and *P. citrophthora*. *Mycological Research*, **98**: 189–199.

Pajerowska-Mukhtar, K., Stich, B., Achenbach, U., Ballvora, A., Lübeck, J., Strahwald, J., Tacke, E., Hoffbert, H.-R., Ilarionova, E., Bellin,D., Walkemeier, B., Basekow, R., Kersten, B. and Gebhardt, C. (2009). Single nucleotide polymorphisms in the Allene Oxide Synthase 2 gene are associated with field resistance to late blight in populations of tetraploid potato cultivars. *Genetics*, **181**(3): 1115-1127.

Parker, J.D., Rabinovitch, P.S. and Burmer, G.C. (1991). Targeted gene walking polymerase chain reaction. *Nucleic Acids Research*, **19**(11): 3055-3060.

Paterson, A., Tanksley, S. and Sorrels, M.E. (1991). DNA markers in plant improvement. *Advances in Agronomy*, **44**: 39-90.

Pearce, S.R., Knox, M., Ellis, T.H., Flavell, A.J. and Kumar, A. (2000). Pea Ty1-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. *Molecular and General Genetics*, **263**: 898-907.

Peleman, J.D., Sørensen, A.P. and van der Voort, J.R. (2005). Breeding by design: exploiting genetic maps and molecular markers through marker-assisted selection. In: Meksem, K. and Kahl, G. (Eds.), *The Handbook of Plant Genome Mapping: Genetic and Physical Mapping*. Weinheim, WILEY-VCH Verlag GmbH and Co. KGaA. pp. 109-129

Perry, M.D., Davey, M.R., Power, J.B., Lowe, K.C., Bligh, F.J., Roach, P.S. and Jones, C. (1998). DNA isolation and AFLP™ genetic fingerprinting of *Theobroma cacao* L. *Plant Molecular Biology Reporter*, **16**(1): 49-59.

Pootakham, W., Jomchai, N., Ruang-areerate, P., Shearman, J. R., Sonthirod, C., Sangsrakru, D., Tragoonrung, S. and Tangphatsornruang, S. (2015). Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics*, **105**: 288-295.

Pound, F.J. (1938) Cacao and witches' broom disease (*Marasmius perniciosus*) of South America with notes on other species of *Theobroma*. Port-of-Spain: Yuille's Printery.

Pound, F.J. (1943) Cacao and witches' broom disease (*Marasmius perniciosa*). Report on a recent visit to the Amazon territory of Peru, September 1942–February 1943. Port of Spain: Yuille's Printery.

Pound, F.J. (1945) A note on the cocoa population of South America. In: Report and *Proceedings of the 1945 Cocoa Conference*. London, pp. 131-133.

Powis, T.G., Cyphers, A., Gaikwad, N.W., Grivetti, L. and Cheong, K. (2011). Cacao use and the San Lorenzo Olmec. *Proceedings of the National Academy of Sciences USA*, **108**: 8595-8600.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, R. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, **38**: 904-909.

Pugh, T., Fouet, O., Risterucci, A.M., Brottier, P., Abouladze, M., Deletrez, C., Courtois, B., Clement, D., Larmande, P., N'Goran, J.A.K. and Lanaud, C. (2004). A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. *Theoretical and Applied Genetics*, **108**: 1151-1161

Purseglove, J.W. (1974). Tropical Crops – Dicotyledons, pp. 543-586. Essex: Longman.

Pritchard, J.K., Stephens, M. and Donnelly, P. (2000). Inference of population structure from multilocus genotype data. *Genetics*, **155**: 945-959.

Quieroz, V.T., Guimarães, C.T., Ahnert, D., Schuster, I., Daher, R. T., Pereira, M.G., Miranda, V.R.M., Loguércio, L.L., Barros, *E.G.* and Moreira, M.A. (2003). Identification of a major QTL in cocoa (*Theobroma cacao* L.) associated with resistance to witches' broom disease. *Plant Breeding*, **122**: 268-272.

Ranade, K., Chang, M.-S., Ting, C.-T., Pei, D., Hsiao C.-F., Olivier, M., Pesich, R., Herbert, J., Chen, Y.-D. I, Dzau, V.J., Curb, D., Olshen, R., Risch, N., Cox, D.R. and Botstein, D. (2001). High-throughput genotyping with single nucleotide polymorphisms. *Genome Research*, 11: 1262-1268.

Risterucci, A.M., Eskes, B., Fargeas, D., Motamayor, J.C. and Lanaud, C. (2001). Use of microsatellite markers for germplasm identity analysis in cocoa. In: *Proceedings of the Third International Group for Genetic Improvement of Cocoa (INGENIC) International Workshop on the New Technologies and Cocoa Breeding*, 16-17 October, 2000, 25-33. Kota Kinabalu, Malaysia.

Risterucci, A.M., Grivet, L., N'Goran, J.A.K., Pieretti, I., Flament, M.H. and Lanaud C. (2000). A high density linkage map of *Theobroma cacao*. *Theoretical and Applied Genetics*, **101**: 948-955.

Rossetto, M. (2001). Sourcing of SSR markers from related plant species. In: *Plant genotyping: the DNA fingerprinting of plants*, Henry, R. J. (ed.), CAB International: UK, pp. 211-224.

Russell, J.R., Hosein, F., Johnson, E., Waugh, R. and Powell, W. (1993). Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis. *Molecular Ecology*, **2**: 89-97.

Ryskov, A.P., Jincharadze, A.G., Prosnyak, M.I., Ivanov, P.L. and Limborska, S.A. (1988). M13 phage DNA as a universal marker for DNA fingerprinting of animals, plants and microorganisms. *FEBS Letters*, **233**: 388-392.

Sankar, A.A. and Moore, G.A. (2001). Evaluation of inter-simple sequence repeat analysis for mapping in *Citrus* and extension of genetic linkage map. *Theoretical and Applied Genetics*, **102**: 206–214.

Santos, E.S.L, Cerqueira-Silva, C B.M., Mori, G.M., Ahnert, D., Mello, D.L.N., Pires, J.L., Corrêa, R.X. and de Souza, A.P. (2015). Genetic structure and molecular diversity of cacao plants established as local varieties for more than two

centuries: *the genetic history of cacao plantations in Bahia*, Brazil. PLoS ONE, 10(12): e0145276. doi:10.1371/journal.pone.0145276

Saunders, J.A., Mischke, S., Leamy, E.A. and Hemeida, A.A. (2004). Selection of international molecular standard for DNA fingerprinting of *Theobroma cacao*. *Theoretical and Applied Genetics*, 110: 41-47.

Schmid, K.J., Sörensen, T.R., Stracke, R., Törjék, O., Altmann, T., Mitchell-Olds, T. and Weisshaar, B. (2003). Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. *Genome Research*, 13: 1250-1257.

Schnell, R.J., Brown, J.S., Olano, C.T., Power, E.J., Krol, C.A., Kuhn, D.N. and Motamayor, J.C. (2004). Frequency of off-type progeny among original ICS 1 x SCA 6 reciprocal families and parental clones used for disease resistance selection in Trinidad. *INGENIC Newsletter* 9: 34-39.

Schnell, R.J., Brown, J.S., Kuhn, D.N., Cervantes-Martinez, C., Borrone, J.W., Olano, C.T., Motamayor J.C., *et al.* (2007a). Current challenges of tropical tree crop improvement: Integrating genomics into an applied cacao breeding program. *Acta Horticulturae*, 738: 129-144.

Schnell, R.J., Kuhn, D.N., Brown, J.S., Olano, C.T., Phillips-Mora, W., Amores, F.M. and Motamayor, J.C. (2007b). Development of a marker assisted selection program for cacao. *Phytopathology*, 97: 1664-1669.

Schnell, R.J., Olano, C.T., Brown, J.S., Meerow, A.W., Cervantes-Martinez, C., Nagai, C. and Motamayor, J.C. (2005). Retrospective determination of the parental population of superior cacao (*Theobroma cacao* L.) seedlings and association of microsatellite alleles with productivity. *Journal of the American Society of Horticultural Science*, 130: 181-190

Scott, K.D. (2001). Microsatellites derived from ESTs, and their comparison with those derived from other methods. In: *Plant genotyping: the DNA fingerprinting of plants*, Henry, R. J. (ed.), CAB International: UK, pp. 225-237.

Sereno, M.L., Albuquerque, P.S.B., Vencovsky, R. and Figueira, A. (2006). Genetic diversity and natural population structure of cacao (*Theobroma cacao* L.) from the Brazilian Amazon evaluated by microsatellite markers. *Conservation Genetics*, 7: 13-24.

Shen, Y.J., Jiang, H., Jin, J.P., Zhang, Z.B., Xi, B., He, Y.Y., Wang, G., Wang, C., Qian, L., Li, X., Yu, Q.B., Liu, H.J., Chen, D.H., Gao, J.H., Huang, H., Shi, T.L. and Yang, Z.N. (2004). Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiology*, 135: 1198-1205.

Silva, C.R.S. and Figueira, A. (2005). Phylogenetic analysis of *Theobroma* (Sterculiaceae) based on Kunitz-like trypsin inhibitor sequences. *Plant Systematics and Evolution*, 250: 93-104. doi 10.1007/s00606-004-0223-2.

Soto-Cerda, B.J. and Cloutier, S. (2012). Association mapping in plant genomes, genetic diversity in plants, edited by Mahmut Caliskan, InTech. http://www.

intechopen.com/books/genetic-diversity-in-plants/association-mapping-in-plant-genomes (accessed February 13, 2013)

Sounigo, O., Ramdahin, S. and Christopher, Y. (2002). Etude de la diversite genetique du cacaoyer a l'aide des RAPD. In: *Proceedings of the 13th International Cocoa Research Conference*, 9-14 October 2000, Kota Kinabalu, Malaysia. COPAL, Lagos, Nigeria. pp. 119-125.

Sounigo, O., Risterucci, A-M., Clément, D., Fouet, O. and Lanaud, C. (2006a). Identification of off-types of clones used in the International Clone Trial using DNA analyses. In: *Eskes, A.B. and Efron, Y. Global Approaches to Cocoa Germplasm Utilization and Conservation*. Final report of the CFC/ICCO/IPGRI project on "Cocoa Germplasm Utilization and Conservation: a Global Approach" (1998-2004), edited by, Amsterdam, The Netherlands: CFC/London, UK: ICCO/ Rome, Italy: IPGRI. http://www.bioversityinternational.org/publications/pubfile.asp?ID_PUB=1172 (accessed July 07, 2008). pp. 82-86.

Sounigo, O., Bekele, F.L., Iwaro, A.D., Thévenin, J-M., Bidaisee, G., Umaharan, R., Sankar, A., Sukha, D., *et al.* (2006b). Description of cocoa clones proposed for the "CFC/ICCO/IPGRI Project Collection. In: Eskes, A.B. and Efron, Y. (Eds.), *Global Approaches to Cocoa Germplasm Utilization and Conservation*. Final report of the CFC/ICCO/IPGRI project on "*Cocoa Germplasm Utilization and Conservation*: a Global Approach" (1998-2004). Amsterdam, The Netherlands: CFC/London, UK: ICCO,/Rome, Italy: IPGRI. pp. 67-81.

Sounigo, O., Umaharan, R., Christopher, Y., Sankar, A. and Ramdahin, S. (2005). Assessing the genetic diversity in the International Cocoa Genebank, Trinidad (ICG,T) using isozyme electrophoresis and RAPD. *Genetic Resources and Crop Evolution*, **52**: 1111-1120.

Susilo, A.W., Zhang, D., Motilal, L.A., Mischke, S. and Meinhardt, L.W. (2011). Assessing genetic diversity in Java fine-flavour cocoa (*Theobroma cacao* L.) germplasm by using simple sequence repeat (SSR) markers. *Tropical Agriculture and Development*, **55**: 84-92.

Sveinsson, S., Gill, N., Kane, N.C. and Cronk, Q. (2013). Transposon fingerprinting using low coverage whole genome shotgun sequencing in cacao (*Theobroma cacao* L.) and related species. *BMC Genomics*, **14**: 502. http://www.biomedcentral.com/1471-2164/14/502

Tachida, H. and Iizuka, M. (1992). Persistence of repeated sequences that evolve by replication slippage. *Genetics*, **131**: 471-478.

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

Takrama, J., Kun, Ji., Meinhardt, L., Mischke, S., Opoku, S.Y., Padi, F.K. and Zhang, D. (2014). Verification of genetic identity of introduced cacao germplasm in Ghana using single nucleotide polymorphism (SNP) markers. *African Journal of Biotechnology*, **13**(21): 2127-2136.

The *Arabidopsis* Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**: 796-815.

The International SNP Map Working Group (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, **409**: 928-933.

Toxopeus, H. (1985). Botany, types and populations. In: Wood, G.A.R. and Lass, R.A. (Eds.), *Cocoa*, 4th Edition. Longman, London. pp. 11-37

Thomas, E., van Zonneveld, M., Loo, J., Hodgkin, T., Galluzzi, G. and van Etten, J. (2012). Present spatial diversity patterns of *Theobroma cacao* L. in the *Neotropics reflect genetic differentiation in Pleistocene refugia followed by human-influenced dispersal*. PLoS ONE, **7** (10): e47676. doi:10.1371/journal.pone.0047676.

Trognitz, B., Cros, E., Assemat, S., Davrieux, F., Forestier-Chiron, N., Ayestas, E., Kuant, A., Scheldeman, X. and Hermann, M. (2013). Diversity of cacao trees in Waslala, Nicaragua: associations between genotype spectra, *product quality and yield potential*. *PLoS ONE*, **8**(1): e54079. doi 10.1371/journal.pone.0054079

Trognitz, B., Scheldeman, X., Hansel-Hohl, K., Kuant, A., Grebe, H. and Hermann, M. (2011). Genetic population structure of cacao plantings within a young production area in Nicaragua. *PLoS ONE*, **6**(1): e16056. doi: 10.1371/journal.pone.0016056

Turnbull, C.J. and Hadley, P. (2012). ICGD 2012: International Cocoa Germplasm Database CD-ROM. CRA Ltd./NYSE Liffe/University of Reading, UK.

Van Hall, C.J.J. (1914) *Cocoa*. Macmillan: London.

Van Hintum, T.J.L. (2000). Duplication within and between germplasm collections. III. A quantitative model. *Genetic Resources and Crop Evolution*, **47**: 507-513.

Van Hintum, Th. J. L., Brown, A.H.D., Spillane, C. and Hodgkin, T. (2000). *Core Collections of Plant Genetic Resources*. IPGRI Technical Bulletin, **3**.

Vicient, C.M., Suoniemi, A., Anamthawat-Jónsson, K., Tanskanen, J., Beharav, A., Nevo, E. and Schulman, A.H. (1999). Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell Online*, **11**: 1769-1784.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Friters, A., Pot, J., Paleman, J. and Kuiper, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**: 4407-4414.

Waits, L.P., Luikart, G. and Taberlet, P. (2001). Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology*, **10**: 249-256.

Wang, J. (2006). Informativeness of genetic markers for pairwise relationship and relatedness inference. *Theoretical and Population Biology*, **70**(3): 300-321. doi:10.1016/j.tpb.2005.11.003

Wang, J., Lin, M., Crenshaw, A., Hutchinson, A., Hicks, B., Yeager, M., Berndt, S., Huang, W.-Y., Hayes, R.B. and Chanock, S.J. (2009). High-throughput single nucleotide polymorphism genotyping using nanofluidic Dynamic Arrays. BMC *Genomics*, **10**: 1. doi: 10.1186/1471-2164-10-561.

Warren, J.M. (1994).) Isozyme variation in a number of populations of *Theobroma cacao* obtained through various sampling regimes. *Euphytica*, **72**: 121-126.

Warren, J., Misir, S. and Kalai (1995). Isozyme markers for self-compatibility and yield in *Theobroma cacao* (cacao). *Heredity*, **74**: 354-356.

Weber, J.L. and Wong, C. (1993). Mutation of human short tandem repeats. *Human Molecular Genetics*, **2**: 1123-1128.

Welsh, J. and McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research*, **18**: 7213-7218.

Whitkus, R., de la Cruz, M., Mota-Bravo, L. and Gómez-Pampa, A. (1998). Genetic diversity and relationships of cacao (*Theobroma cacao* L.) in southern Mexico. *Theoretical and Applied Genetics*, **96**: 621-627.

Whitlock, B.A. and Baum, D.A. (1999). Phylogenetic relationships of *Theobroma* and *Herrania* (Sterculiaceae) based on sequences of the nuclear gene *Vicilin*. *Systematic Botany*, **24**: 128-138.

Whitlock, B.A., Bayer, C. and Baum, D.A. (2001). Phylogenetic relationships and floral evolution of the Byttnerioideae ("Sterculiaceae" or Malvaceae s.l.) based on sequences of the chloroplast gene, *ndhF*. *Systematic Botany*, **26**: 420-437.

Wilde, J., Waugh, R. and Powell, W. (1992). Genetic fingerprinting of *Theobroma* clones using randomly amplified polymorphic DNA markers. *Theoretical and Applied Genetics*, **83**: 871-877.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, **18**: 6531-6535

Wood, G.A.R. (1985a). History and development. In Cocoa, Wood, G.A.R and Lass, R.A. (Eds.) 4th edition, Chapter 1, Longman, London. pp.1-10.

Wood, G.A.R. (1985b). Quality and inspection. In Cocoa, Wood, G.A.R and Lass, R.A. (Eds.) 4th edition, Chapter 14, Longman, London. pp. 505-527.

Wood, G.A.R. (1991). A history of early cocoa introductions. *Cocoa Growers Bulletin*, **44**: 7-12.

Yang, J.Y., Scascitelli, M., Motilal, L.A., Sveinsson, S., Engels, J.M.M., Kane, N., Dempewolf, H., Zhang, D., Maharaj, K. and Cronk, Q.C.B. (2013). Complex origin of Trinitario-type *Theobroma cacao* (Malvaceae) from Trinidad and Tobago revealed using plastid genomics. *Tree Genetics and Genomes*, **9**: 829-840. doi 10.1007/s11295-013-0601-4.

Yidana, J.A. and Kennedy, A.J. (1984-86). Isozyme variation in cocoa. In: Annual Report 1984-86, Cocoa Research Unit: The University of the West Indies. pp. 6-11.

Yidana, J.A., Kennedy, A.J. and Withers, L.A. (1987). Variation in peroxidase isozymes of cocoa (*Theobroma cacao* L.). In: *Proceedings of the 10th International Cocoa Research Conference*, 17-23 May 1987, St. Domingue, Dominican Republic. COPAL, Lagos, Nigeria. pp. 719-722.

Young, A.M. (1994).) The chocolate tree: A natural history of cacao. Washington DC, USA: Smithsonian Nature Books.

Young, N.D. (1999). A cautiously optimistic vision for marker-assisted breeding. *Molecular Breeding*, **5**: 505-510.

Yu, J. and Buckler, E.S. (2006). Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*, **17**: 155-160. doi:10.1016/j.copbio.2006.02.003.

Yu, J., Hu, S., Wang, J., Wong, G. K.-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., *et al.* (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, **296**: 79-92. doi 10.1126/science.1068037.

Yu, K., Park, S. and Poysa, V. (2000). Marker-assisted selection of common beans for resistance to common bacterial blight: Efficacy and economics. *Plant Breeding*, **119**: 411-415.

Yu, J., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., *et al.* (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, **38**: 203-208. doi:10.1038/ng1702

Zhang, D., Arevalo-Gardini, E., Mischke, S., Zúñiga-Cernades, L., Barreto-Chavez, A., Adriazola del Aguila, J. (2006a). Genetic diversity and structure of managed and semi-natural populations of cocoa (*Theobroma cacao*) in the Huallaga and Ucayali Valleys of Peru. *Annals of Botany*, **98**: 647-655.

Zhang, D., Mischke, S., Goenaga, R., Hemeida, A.A. and Saunders, J.A. (2006b). Accuracy and reliability of high-throughput microsatellite genotyping for cacao clone identification. *Crop Science*, **46**: 2084-2092.

Zhang, D., Boccara, M., Motilal, L., Butler, D.R., Umaharan, P., Mischke, S. and Meinhardt, L. (2007). *Microsatellite variation and population structure in the "Refractario" cacao of Ecuador. Conservation Genetics*, **9**(2): 327-337. doi 10.1007/s10592-007-9345-8

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

Zhang, D., Mischke, S., Johnson, E.S., Phillips-Mora, W. and Meinhardt, L. (2009b). Molecular characterization of an international cacao collection using microsatellite markers. *Tree Genetics and Genomes*, **5**: 1-10.

Zhang, D., Martínez, W.J., Johnson, E.S., Somarriba, E., Phillips-Mora, W., Astorga, C., Mischke, S. and Meinhardt, L.W. (2011). Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia. *Genetic Resources and Crop Evolution*, **59**(2): 239-252.

Zhao, Y., Chen, F., Li, Q., Wang, L. and Fan, C. (2015). Isothermal amplification of nucleic acids. *Chemical Reviews*, **115**: 12491-12545.

Zhou, H. and Steffenson, B. (2013). Genome-wide association mapping reveals genetic architecture of durable spot blotch resistance in US barley breeding germplasm. *Molecular Breeding*, **32**: 139-154. doi:10.1007/s11032-013-9858-4.

Zhu, C., Gore, M., Buckler, E. S. and Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome*, **1**: 5-20 doi:10.3835/plantgenome2008.02.0089.

Ziegleder, G. (1990). Linalool contents as characteristics of some flavour grade cocoas. *Zeitschrift für Lebensmittel-Untersuchung und Forschung*, **191**: 306-309.

Zietkiewicz, E., Rafalski, A. and Labuda, D. (1994).) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics*, **20**: 176-183.