Chapter 6
Genomics of *Theobroma cacao*, "the Food of the Gods"

Mark J. Guiltinan, Joseph Verica, Dapeng Zhang, and Antonio Figueira

Abstract *Theobroma cacao*, the chocolate tree, is an important tropical tree-crop that provides sustainable economic and environmental benefits to some of the poorest and most ecologically sensitive areas of the world. Recent progress in the development of genomics tools for cacao is reviewed. These include a reference molecular genetic map, simple sequence repeats and other molecular markers, two germplasm databases with microsatellite DNA fingerprints and other molecular data, many quantitative trait loci mapping projects which have identified disease resistance and yield component loci, several expressed sequence tag resources, a cacao microarray, bacterial artificial chromosome libraries, and a genetic transformation system. The evolutionary relatedness of cacao with other important crops and model plant systems positions cacao genomics to play a significant role in translational plant genomics. The future prospects for the contribution of cacao genomics to improvement of this crop for sustainable cacao production and as a tool for poverty alleviation and environmental stabilization are discussed.

6.1 Introduction

6.1.1 Economic, Agronomic, and Societal Importance of Cacao

Theobroma cacao L. (cacao; Malvaceae *sensu lato*) is a small under-story tree endemic to the lowland rainforests of the Amazon basin (Wood and Lass 1985; Bartley 2005). Today, cacao is grown throughout the humid tropics, often in agroforestry-ecosystems with other fruit and commodity crops. The annual world production of cocoa (the product obtained from dried, fermented cacao seeds) is approximately 3 million tons, with 2/3 being processed into cocoa powder and cocoa butter, and the remaining 1/3 used for cocoa liquor (the flavor and color component of chocolate) (Wood and Lass 1985). Cocoa is the major export commodity

M.J. Guiltinan
The Pennsylvania State University, Department of Horticulture, Life Sciences Building, University Park, PA 16802-5807, USA
e-mail: mjg9@psu.edu

P.H. Moore, R. Ming (eds.), *Genomics of Tropical Crop Plants*

of several countries in West Africa (68% of world production), providing major economic resources to Ivory Coast, Cameroon, Nigeria, and Ghana. Other major cocoa exporters include Ecuador, Venezuela, Brazil, Costa Rica, Malaysia, and Indonesia. Worldwide, approximately 5 to 6 million smallholder farmers grow 95% of the world's production. World cocoa export trade is $5 to $6 billion/year. In the United States, the use of cocoa and cocoa butter in chocolate manufacturing, cosmetics, and other products drives an approximately $70 billion dollar market, providing over 60,000 jobs (Morais 2005). US chocolate production also uses large amounts of sugar, nuts, and milk valued at approximately $3 billion/year in receipts to American farmers.

Cacao-growing regions are largely centered in important biodiversity hotspots, and in proximity to 13 of the world's most biologically diverse regions (Piasentin and Klare-Repnik 2004). Because cacao is a shade-grown perennial tree crop with a cropping cycle of 50+ years, cultivation provides environmental benefits such as enhancement of biodiversity in avian migratory corridors, soil and watershed conservation, and buffer zones near endangered rainforest habitats (Rice and Greenberg 2003; Ruf and Zadi 2003). Development agencies such as USAID, Conservation International, The World Wildlife Federation, and the World Cocoa Foundation are increasingly aware of the role of cacao in stabilizing local economies and environments and have stepped up their involvement with cacao farmers in these regions (Guyton et al. 2003).

Cacao diseases reduce the potential crop by an estimated 810,000 tons annually (30% of world production) and individual farm losses can approach 100% (Keane 1992; Bowers et al. 2001). For example, in Southern Bahia, Brazil, introduction of the witches' broom disease, caused by the fungus *Moniliophthora perniciosa* (Aime and Phillips-Mora 2005), resulted in a decrease in production from 300,000 tons in 1989 to 130,000 tons 10 years later, for an estimated loss of $220 million each year (Pereira et al. 1990). In addition, economic loss due to cacao disease has caused widespread social disruption among smallholder growers. Widely dispersed cacao pathogens include several species of *Phytophthora* that cause multiple diseases of economic importance including pod rot, trunk canker, and leaf-blight (Appiah et al. 2003; Chowdappa et al. 2003; Appiah et al. 2004). *Phytophthora megakarya*, the most aggressive species, is reported to have entered Ivory Coast, the world's leading cocoa producer (Nyasse et al. 2002; Opoku et al. 2002; Appiah et al. 2003; Risterucci et al. 2003; Efombagn et al. 2004). Other important diseases and pests include frosty pod in Central America caused by *Moniliophthora roreri* (Evans et al. 2003), the cocoa pod borer in Asia (Day 1984; Santoso et al. 2004), and cocoa swollen shoot virus in West Africa (Hanna 1954; Hervé et al. 1991; Muller and Sackey 2005).

6.1.2 Cacao as an Experimental Organism

Theobroma cacao is a simple diploid with ten chromosomes ($2n = 2x = 20$) and a small genome. Published genome size estimates vary from 390 Mb to 415 Mb

(Figueira et al. 1992; Couch et al. 1993); however, recent fluorescent flow-cytometer measurements estimate a genome size of 447 Mb ± 11 Mbp 2C (Arumuganathan, Carlson, and Guiltinan, unpublished). This size represents the average of 10 genotypes measured, with three replicates per genotype. This is a relatively small size for a plant genome and thus makes certain aspects of cacao genomics more feasible, such as the possibility of sequencing the entire genome.

Cacao has several limitations as an experimental organism. For example, its life cycle takes a minimum of 2–3 years from seed to seed, and progeny of crosses must be grown for many years to fully evaluate their productivity and disease resistance characteristics. Many cacao genotypes are self-incompatible, making genetic analysis and breeding strategies labor intensive. Furthermore, the plants require large areas of land and large inputs of labor to maintain and evaluate field tests. In addition, the seeds are recalcitrant, so germplasm must be conserved as living collections in the field or greenhouses. These and other factors combine to make cacao a very difficult and slow experimental system.

6.1.3 Current Status of Cacao Genetics and Breeding

Breeding programs were established in the major cacao growing countries starting in the 1920s (Toxopeus 1969; Bartley 2005). In the 1930s and 40s, valuable germplasm was collected from the Amazon regions of Brazil, Ecuador, and Peru (Pound 1940; Bartley 2005), and clonal descendants of these accessions are maintained in present day germplasm collections (Lockwood and Gyamfi 1979; Engels 1981; Iwaro et al. 2003). Large genetic variation has been identified in wild populations throughout the Amazon, but this diversity has not yet been widely incorporated into cultivated varieties (Bartley 2005). Disease resistance is currently the primary trait targeted by cacao breeders. Other important traits include yield efficiency, flavor characteristics, cocoa butter content (% seed lipid content) and quality (fatty acid saturation), tolerance to abiotic stress, and various horticultural traits such as precocity, rootstock/scion interactions, plant height, and stature.

The fields of cacao genetics, breeding, and biotechnology have been the subject of a number of relatively recent review articles (Bartley 1994; Hughes and Ollennu 1994; Bennett 2003; Silva and Figueira 2005; Guiltinan 2007, Maximova et al. 2007), two books (Dias 2001; Bartley 2005), and one conference proceedings (Eskes 2003).

6.1.4 Cacao Genomic Resources:

Today's cacao genetics research community is well organized, highly collaborative, and poised to make use of new genomics resources (reviewed by Bennett 2003). To formally foster collaboration and communication between cacao breeders and geneticists, the International Group for Genetic Improvement of Cocoa (INGENIC)

was formed in 1994 (http://ingenic.cas.psu.edu). It now includes over 300 members, representing 35 developing and developed countries around the world. To coordinate the activities of the INGENIC members interested in molecular approaches, the INGENIC Study Group for Molecular Biology (INGENIC-MOL-BIOL) was formally chartered in October of 2003 (Johnson 2003). An international research symposium is held by INGENIC in a developing country every third year.

Most of the cocoa producing-countries have research facilities funded by international and national organizations that support agricultural research, such as the Cocoa Research Institute of Ghana (CRIG), the Cocoa Research Institute of Nigeria (CRIN), Institute of Agricultural Research for Development (IRAD - Cameroon), the Commissco Executiva do Plano da Lavoura Cacaueira (CEPLAC – Brazil), and the Instituto Nacional Autonomo de Investigaciones Agropecuarias (INIAP – Ecuador). The main strengths of these organizations are the many scientists with strong experiences in cacao agriculture, as well as their extensive field sites and breeding programs. It is essential that researchers in developed and developing countries establish and maintain strong working relationships and collaborative research and training programs to maximize the potential impact of their research on the cacao farmers and the environment. The Cocoa Producers' Alliance (COPAL) also supports basic research in cacao and sponsors the International Cocoa Research Conference (http://www.copal-cpa.org/).

In the United States, the USDA-ARS has established two centers of cacao research in Beltsville, MD, and Miami, FL, which carry out a wide array of research projects with collaborating laboratories worldwide (Pugh et al. 2004). Several Universities have research programs in cacao, including The Pennsylvania State University, which is home to the American Cocoa Institute Endowed Program in the Molecular Biology of Cacao. Several centers of excellence in cacao research in Europe can be found in the Centre de Cooperation Internationale en Recherche Agronomique pour le Developpement (CIRAD), Montpellier, France, The University of Reading, UK, and at other sites. Several centers of excellence in cacao research can also be found in Central and South America; Centro Agronómico Tropical de Investigación y Enseñanza, Costa Rica (CATIE), University of São Paulo, Brazil (USP), Universidade Estadual de Santa Cruz, Brazil (UESC) and others.

6.2 Genomic Mapping and Characterization

6.2.1 Molecular Markers for the Characterization of Cacao Germplasm

In the past two decades, research on the genetics of cacao has benefited enormously from molecular markers. Significant progress has been made in molecular characterization of cacao germplasm. The key objectives of this line of work include: reducing redundancy and mislabeling in cacao gene banks, understanding

genetic diversity in *ex situ* collections and in farmer's fields, verifying genealogical information and characterizing germplasm for useful agronomic traits.

Isozymes were the first molecular markers utilized in cacao (Lanaud 1986). Although the available loci and the numbers of polymorphisms typically generated by each isozyme were low, this simple system enabled assessment of genetic diversity and mating system, assisted genotype identification, and contributed to linkage mapping (Ronning and Schnell 1994; Sounigo et al. 2005; Warren et al. 1994; Lachenaud et al. 2004). However, the isozyme markers are outdated because of their low polymorphism and the environmental effect on the "phenotype."

Commonly used DNA markers in cocoa include restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and simple sequence repeats (SSRs). These markers differ in genomic abundance, level of polymorphism detected, locus specificity, reproducibility, technical requirements, and financial cost. They have been used to answer various research questions in cacao.

RFLP is a non-polymerase chain reaction (PCR) based DNA marker that was first applied in cacao in the early 1990s (Laurent et al. 1994). The polymorphism of RFLP is moderately high in cacao. Its high reproducibility and the co-dominant allelic nature make it a suitable marker for the construction of genetic linkage maps and tagging genes and quantitative trait loci (QTL) linked to characters of agronomic importance (Lanaud et al. 1995, Risterucci et al. 2000) and assessment of genetic diversity (N'Goran et al. 1994; Motamayor et al. 2002). Because RFLP probes are invariably specific to a limited number of loci, RFLP is not a very effective tool for cacao genotype identification. Another major drawback is that RFLP is not amenable to automation and data generation is both laborious and expensive.

RAPD was the first PCR based DNA fingerprinting method to be applied for the genetic characterization of cocoa (Wilde et al. 1992). This system is technically simple to perform but has a low reproducibility between experiments and laboratories. RAPD is not a direct measure of heterozygosity which makes it less useful for actual genotyping as opposed to simply distinguishing between clones. The number of variable markers that can be generated by RAPD is small and not suitable for high-throughput application. Nevertheless, RAPD is a simple-to-use DNA marker for scoring variations between cacao clones. Russell et al. (1993) showed that a minimum of three RAPD primers were able to distinguish 25 cacao accessions according to their geographical origin. Various studies used RAPD for identification of accession mislabeling and duplication (Christopher et al. 1999; Faleiro et al. 2002; Sounigo et al. 2005). RAPD has also been widely used in analyses of genetic diversity (Figueira et al. 1994; N'Goran et al. 1994; Lerceteau et al. 1997; Whitkus et al. 1998; Marita et al. 2001; Yamada et al. 2001; Faleiro et al. 2004; Sounigo et al. 2005).

AFLP is a PCR based fingerprinting protocol that combines the strength of RAPD and RFLP (Vos et al. 1995). AFLP is highly polymorphic with considerable reproducibility within a laboratory. Similar to RAPD, AFLP does not require sequence data for primer construction. However, because of its dominant nature, AFLP is not a direct measure of heterozygosity so it has limited use in genotyping.

Perry et al. (1998) reported identification of parental plants of cacao and their hybrids using AFLP. AFLP was reliable for distinguishing closely related cocoa varieties (Saunders et al. 2001). Queiroz et al. (2003) reported the identification of a major QTL associated with resistance to witches' broom disease, based on AFLP linkage mapping. However, there have been few studies to date using AFLP in cacao germplasm management. The major disadvantages of AFLP include the dominance of alleles and the possible non-homology of co-migrating fragments belonging to different loci, which limit its wide application in cacao.

In the past decade, SSRs, also known as microsatellites, have emerged as the most widely used marker for cacao, enabling great strides to be made in the characterization of cocoa germplasm (Lanaud et al. 1999). SSRs are typically co-dominant and multiallelic, allowing precise discrimination (or matching) of individual clones based on the multi-locus fingerprints. The data analysis and interpretation of results fit the genetic model of cacao. As SSR sequence information can be easily shared between laboratories, data generated in different laboratories can be standardized through a combination of ring testing of reference genotypes, standardized protocol for allele sizing, and the development of marker-specific size ladders containing all the common alleles for a given locus (Sampaio et al. 2003; Cryer et al. 2006a, b). The standardized datasets then can be merged for joint analysis (George et al. 2004; Presson et al. 2006).

Because of these major advantages, SSR has been the system of choice for cacao in the past decade. It has been applied to various aspects of cacao germplasm management, including:

A. *Identification of mislabeled and duplicate accessions in germplasm collections:* Incorrect labeling of accessions has been a significant problem that has hindered the efficient conservation and use of cacao germplasm (Motilal and Butler 2003; Turbull et al. 2004). RAPD and AFLP have sufficient discriminatory power to separate accessions with different genotypes, but these markers are very limited in their ability to determine absolute identity between two individual trees due to artifact polymorphisms. The RAPD and AFLP based conclusion that two individuals are identical was usually only approximate, and no statistical rigor was added to this assertion (Perry et al. 1998; Christopher et al. 1999; Sounigo et al. 2005). This lack of precision limits the application of RAPD and AFLP for cacao germplasm identification.

   SSR markers, used together with high throughput genotyping facilities, enable large-scale assessment of genetic identity in cacao gene banks. In contrast to dominant markers such as AFLP and RAPD, identical cacao genotypes can have an exact match in the multi-locus SSR profiles (Zhang et al. 2006a). Thus, SSR based multi-locus genotyping has been widely applied in various national and international germplasm collections (Boccara and Zhang 2006; Cryer et al. 2006a, b; Zhang et al. 2006a; Johnson et al. 2007). SSR was also used to verify genetic identities in the breeding progenies (Takrama et al. 2005) and to monitor germplasm integrities for accessions maintained in vitro (Rodriguez et al. 2004). An international consortium was formed to identify the

cacao genotypes and describe the genetic diversity in the major international and national cocoa collections maintained in the Americas. To date, more than 4,000 cocoa accessions maintained in the two international genebanks (CATIE, Costa Rica, The International Cocoa Genebank, Trinidad (ICG, T) and in several other national collections in the Americas have been genotyped with a set of 15 standard microsatellite (SSR) loci. Based on the multi-locus SSR profiles, duplicated accessions, both within and among different collections, were unambiguously identified. The reference profiles, together with the derived information in genetic diversity, are being submitted to the international databases The International Cocoa Germplasm Database (ICGD) [http://www.icgd.rdg.ac.uk/] and CocoaGenDB).

B. *Investigation of genealogical relationships:* Cacao trees have an open-pollinated breeding system. Their outcrossing nature and insects-mediated pollen dispersal mean that the majority of the progenies are based on the random mating of several parents. Therefore, the co-dominant nature of SSR is highly suitable for the investigation of the alleged genealogical relationship. Parentage and sibling relationship analysis between individuals is based on formal estimates derived from allele frequencies. Statistical rigor can be formally computed for the estimation. Parentage analysis is extremely useful for understanding gene flow in natural populations, monitoring parentage contribution in cacao seed gardens, and breeding programs. Using SSR markers, Schnell et al. (2005) determined the parental population for a group of productive and unproductive seedlings growing in Hawai'i. Seven SSR loci were found to have alleles that were associated with productive or unproductive seedlings. Motamayor et al. (2003) reported SSR allelic evidence that the Trinitario cacao is a result of natural hybridization and subsequent introgressions between Lower Amazon Forastero and ancient Criollo varieties.

   Parentage analysis based on the multi-locus SSR data is now routinely applied to verify the recorded pedigree in cocoa germplasm. For populations known to have family structures but lacking detailed pedigree information (i.e., most of Pound's collections made in the 1930s), re-construction of genealogical relationship is being performed using the multi-locus SSR data.

C. *Assessment of on-farm diversity:* Using SSR markers, Aikpokpodion et al. (2005) assessed the genetic diversity of the cacao cultivated in different agro-ecologies in Nigeria. A low adoption rate of improved germplasm was revealed in farmers' fields through SSR analysis. They also showed that only a small fraction of genetic diversity in the national germplasm collection has been exploited (Bhattacharjee et al. 2004). Another survey conducted in southern Cameroon found that farm accessions differed by geographical origin and parentage. The progeny of some promising Upper Amazon clones (T-clones) were poorly represented in the cacao farms (Efombagn et al. 2006). An SSR based diversity survey was also carried out in Ivory Coast and Ghana (Opoku et al. 2005).

   In a study of trees found near the center of origin of cacao, Zhang et al. (2006b) observed a high allelic diversity in semi-domesticated cocoa trees from the Ucayali valley of Peru. Their results substantiate the hypothesis that the Peruvian

Amazon hosts a high level of cacao genetic diversity and that the diversity has a spatial structure, which highlights the need for additional collection and conservation measures for cacao germplasm in this region.

D. *Investigating phylogeography:* Understanding the spatial patterns of biodiversity, the processes of gene flow, and the historical and climatic effects on the distribution of genetic diversity is essential for sustainable conservation and effective use of cacao germplasm. SSR is an ideal tool for assessing intra-specific phylogeography for cacao. Motamayor et al. (2002) assessed the allelic composition in varieties from Central America (genotypes from Guiana, Amazonia, and Orinoco regions were compared using SSR). Their results support the hypothesis that cacao originated in the upper Amazon and suggest the likely dispersal route from the Amazon to Central America and Mexico. Sereno et al. (2006) compared samples from natural populations from the Brazilian Amazon (Acre, Rondonia, lower Amazon, and upper Amazon). The Brazilian upper Amazon population was found to have the largest genetic diversity and therefore was suggested to be part of the center of diversity for the species. SSR was also used to examine the origin and dispersal of "Nacional" cocoa in Ecuador. The result obtained supports the hypothesis that "National" cacao had an early establishment in the coastal region of Ecuador before other exotic germplasm was introduced into this region (R. Loor and F. Amores, personal communication).

As the sequence data of ESTs are increasingly generated and deposited in public databases, EST derived SSR markers are increasingly becoming available (Borrone et al. 2007). Database availability significantly reduces the cost of identifying and developing useful SSR markers. Moreover, a putative function may be assigned to the locus based upon the sequence homology. Another key achievement that significantly enhances the effectiveness of SSR is the newly available statistical tools that do not require prior assumptions of population boundaries, e.g., maximum likelihood and Bayesian approaches (Pritchard et al. 2000; Piry et al. 2004). These tools, together with powerful computers, greatly increase the potential for using SSR markers in studying population structure, environmental adaptation, genotype and population assignment, and kinship analysis.

RFLP, RAPD, AFLP, and SSR were the markers of choice during the last two decades. Among them, SSR markers are preferred for cacao germplasm characterization and this system will continue to play a major role in cocoa breeding and germplasm research. However, SSR uses electrophoresis based assays and is therefore time consuming and expensive. Moreover, SSR is not an ideal system for conducting association studies in plants because of the occurrence of homoplasy (Rafalski 2002). Emphasis is now shifting towards the development of molecular markers that can be detected through non-gel-based assays. One of the most popular of these is single nucleotide polymorphism (SNP), which detects sites where DNA sequences differ by a single base pair. SNPs are the most abundant class of polymorphisms in plant genomes (Buckler and Thornsberry 2002; Zhang and Hewitt 2003). Assays of SNPs do not require DNA separation by size and, therefore,

can be automated in an assay-plate format or on microchips (Rafalski 2002). When a large amount of genotyping needs to be done in a timely fashion, automation is essential to dramatically reduce the costs per data point and improve the repeatability of the result among experiments and laboratories. Another major advantage of SNPs is that they can often be identified in candidate genes with potentially important functions (Kuhn et al. 2003, 2004; Borrone et al. 2004). Therefore, it is an ideal system for the association of SNP haplotypes at candidate gene loci with phenotypes. SNP data are also unambiguous so that merge data between platforms or laboratories is straightforward. Because of these advantages, SNPs will likely become the marker-of-choice in the future. Development of SNP markers is progressing rapidly in cocoa (Cryer N, Kuhn D, and Lanaud C; personal communication). The large-scale application of SNPs in cacao will significantly increase our ability to address more complex questions in cocoa germplasm characterization.

6.2.2 Linkage Mapping

The first cacao linkage map was developed for the 'UPA402' x 'UF676' progeny, containing 193 marker loci (mainly RFLPs and RAPDs), covering 759 cM in 10 linkage groups, corresponding to the haploid chromosome number (Lanaud et al. 1995) (Table 6.1). This map was later saturated with additional markers (mainly AFLP and SSRs) to a total of 424 loci (Risterucci et al. 2000), and it became the consensus linkage map and reference for linkage group or chromosome numbering for *T. cacao* (Clément et al. 2001. The latest published version of this high-density map contained 465 markers, 268 of which were SSRs and 16 resistance gene analogs (RGA), covering 782.8 cM, and it was based on 135 individuals (Pugh et al. 2004) (Table 6.1).

Crouzillat et al. (1996) developed a second linkage map, using a backcross population of 131 plants derived from a cross between a single F1 tree from 'Catongo' x 'Pound12' to a recurrent 'Catongo'. The final backcross map contained 140 markers (RFLPs and RAPDs) and included two morphologic loci (self-compatibility and anthocyanin synthesis), and covered 944 cM (Crouzillat et al. 2000b). An additional genetic map was developed for the related F1 population from the cross 'Catongo' x 'Pound12', with 162 markers, covering 772 cM (Crouzillat et al. 2000a).

Subsequently, another eight linkage maps were developed to search for genomic regions associated with resistance to various *Phytophthora* species and strains, evaluated by diverse methods of inoculation (Table 6.1; Flament et al. 2001; Motilal et al. 2002; Clement et al. 2003b; Risterucci et al. 2003). Simultaneously, -QTL- associated with yield components and plant vigor were also assessed in some of these families.

An F2 population, derived from the 'ICS1' x 'Scavina6' cross, has been specifically developed and used to identify genomic regions associated with witches' broom disease resistance, caused by *Moniliophthora perniciosa* (Queiroz et al. 2003; Brown et al. 2005; Faleiro et al. 2006). This F2 high-density genetic map currently contains nearly 500 markers, including 270 SSR loci, and is based on 146 individuals. Recently, additional mapping populations derived from novel sources of resistance to witches' broom were developed in Brazil to identify QTLs associated

Table 6.1 Details about currently published cacao genetic maps with emphasis on identification of QTLs associated with disease resistance

Families	No. of Plants	No. of Markers	No. of Linkage Groups	Map Length (cM)	Evaluation method for resistance	Reference
					Phytophthora	
UPA 402 × UF 676	144	193	10	759	Artificial inoculation of leaf disks	Lanaud et al., 1995;
	181	473		887	and % pod rot rate under field	Risterucci et al., 2000;
	135	465		783	conditions with *P. palmivora*	Pugh et al., 2004;
T60/887 × IFC 2	(59)		11	793	Artificial inoculation of leaf disks	Flament et al., 2001;
T60/887 × IFC 5	(56)				and wounded pods, and % pod	
	112	198			rot rate under field conditions with *P. palmivora*	
Catongo × Pound12	55	162	12	772	Attached pod inoculation without wounding	Crouzillat et al., 2000b
Catongo × (Catongo × Pound12)	131	140	10	944		Crouzillat et al., 2000b
ICS 84 × UPA 134	62	224	15	548	Artificial inoculation of leaf disks	Flament, 1998
SNK 10 × UPA 134	78	151	16	617	and wounded pods, and % pod	Flament, 1998
SNK 413 × IMC 67	58	119	15	419	rot rate under field conditions with *P. megakarya*	Flament, 1998
DR 1 × Catongo	107	192	9	653	% pod rot rate under field	Clement et al., 2003b
S 52 × Catongo	101	138	11	589	conditions with *P. palmivora*	Clement et al., 2003b

Table 6.1 (continued)

Families	No. of Plants	No. of Markers	No. of Linkage Groups	Map Length (cM)	Evaluation method for resistance	Reference
IMC 78 × Catongo	128	223	10	721		Clement et al., 2003b
(Scavina 6 × H) × IFC 1	151	213	10	682	Artificial inoculation of leaf disks with two strains of *P. megakarya*, *P. palmivora*, *P capsicii*	Risterucci et al., 2003
IMC 57 × Catongo	155	235	12	1427	Artificial inoculation of leaf disks with *P. palmivora*	Motilal et al., 2002
					Moniliophthora perniciosa	
F_2 (ICS 1 × Scavina 6)	146	178	10	672	Field resistance	Brown et al., 2005;
	82	342	16	670		Faleiro et al., 2006
ICS 39 × CAB 214	116	57	9	501	Artificial inoculation under field conditions	Figueira et al., 2006
ICS 39 × CAB 208	168	57	9	541		
					Moniliophtora roreri	
Pound7 × UF273	256	180	10	878	Artificial inoculation of attached pods, scored for internal and external lesion	Brown et al., 2006

with resistance (Figueira et al. 2006). Preliminary QTLs have been associated with resistance to frosty pod disease caused by *Moniliophthora roreri*, in the 'Pound7' x 'UF273' progeny (Brown et al. 2006). Another F2 population was developed to identify QTLs associated with flavor and seed quality (Crouzillat et al. 2001).

6.2.3 Quantitative Trait Loci Mapping

Theobroma cacao genetic maps have been used to detect QTLs for various agronomically important traits, including resistance to the three major fungal diseases (Table 6.1), yield components, plant vigor, and quality traits (Lanaud et al. 1996; Motilal et al. 2002; Flament et al. 2001; Clement et al. 2003a; Clement et al. 2003b; Guimarães et al. 2003; Queiroz et al. 2003; Risterucci et al. 2003; Brown et al. 2005).

A. Disease resistance: Black pod rot, caused by various species of *Phytophthora*, is the most important disease of cacao worldwide. *Phytophthora palmivora* occurs globally, while *P. megakarya* is restricted to West Africa, and *P. capsicii* and *P. citrophthora* occur in the Americas. To identify genomic regions associated with resistance to *Phytophthora* species and isolates, 12 linkage maps have been published for populations in Ivory Coast, Costa Rica, Cameroon, Trinidad, and France (Table 6.1). Evaluation of *Phytophthora* resistance for QTL analyses have been performed by either using natural pod rot losses, by artificial inoculation of attached pods (wounded or not), or by inoculation of leaf disks (Table 6.1).

Based on natural infection rates under field conditions in Ivory Coast, QTLs for resistance against *P. palmivora* were detected for both parents of the 'UPA402' x 'UF676' progeny on chromosome 1, explaining 15% to 19% of the variation of the trait, and a minor one on chromosome 9 (Lanaud et al. 2000). When a similar evaluation was conducted for a two-year harvest period using the 'T60/887' x 'IFC2/IFC5' progenies, one major QTL was identified on chromosome 10, explaining 17% of the phenotypic variance (Flament et al. 2001). Based on data between years eight and 13 after planting, Clement et al. (2003b) detected a significant QTL for resistance against *P. palmivora* for two parents 'DR1' and 'IMC78' on the same region of chromosome 4, explaining 10.1% and 22.6%, respectively of the trait variation.

When the resistance against *P. palmivora* of the 'Catongo' x 'Pound12' F1 and the BC1 progenies was evaluated based on artificial pod inoculation under field conditions, six QTLs were detected on five linkage groups (Crouzillat et al. 2000a). Only one QTL (on chromosome 9) was common to both populations, with a major effect on the F1, explaining nearly 48% of the variance for the trait. Flament et al. (2001) also used artificial inoculation of attached pods (nonetheless wounded) in the search for resistance, identifying two QTLs on chromosomes 2 and 6 of 'T60/887'. This minor QTL identified on chromosome 2 was co-localized with one identified by Crouzillat et al. (2000a).

A screening method for resistance against *Phytophthora* based on inoculation of leaf disks has been used to identify QTLs (Lanaud et al. 2000; Flament et al. 2001; Risterucci et al. 2003). However, correlations between resistance evaluated by leaf disk inoculation with field pod rot rate or pod inoculations with *P. palmivora* have

been weak and non-significant, possibly because leaf disk inoculations are highly influenced by the environment and have low precision and reproducibility. Flament et al. (2001) identified two QTLs on chromosomes 6 and 3 that were associated with resistance to *P. palmivora* based on the leaf test. These are at distinct locations to those identified by field pod rot rate or by pod inoculation, suggesting either that each evaluation method accounted for a distinct genetic mechanism of resistance, or the size of the progeny and/or the low reproducibility of the tests were not sufficiently accurate to detect all QTLs involved. Similar results were obtained for tests with *P. megakarya* (Flament 1998) with QTLs for resistance identified on chromosome 9 (evaluated by leaf inoculation), and on chromosome 2 (by pod inoculation) on 'UPA 134'.

Based on inoculation of leaf disks of the ('Scavina6' x Hybrid) x 'IFC1' progeny, using two isolates of each of three *Phytophthora* species (*P. palmivora, P. megakarya, P. capsicii*), 13 QTLs for resistance were detected in six chromosome regions, explaining between 7.5% to 12.4% of the phenotypic variation (Risterucci et al. 2003). A region on chromosome 5 contained QTLs for resistance against five strains belonging to the three *Phytophthora* species, whereas two other regions on chromosomes 1 and 6 enclosed QTLs for resistance against two species, suggesting that some of the resistance factors against these *Phytophthora* species might be shared. A QTL for resistance against *P. palmivora* based on field pod inoculation was also identified on chromosome 5 of the BC1 progeny by Crouzillat et al. (2000a), where a cluster of RGA have been also localized (Lanaud et al. 2004). Additionally, a QTL for resistance against one strain of *P. megakarya* was detected on chromosome 3, near another one identified by leaf test against *P. palmivora* in 'T60/887' (Flament et al. 2001). Motilal et al. (2002) using leaf inoculation of 'IMC57'x 'Catongo' with *P. palmivora* identified three major QTLs on chromosomes 1, 9, and 3 or 8, that co-localized with QTLs detected in other studies.

The difficulty in identifying consistent QTLs for *Phytophthora* resistance might be in part due to lack of precision in phenotype evaluation, or a complex multigenic mechanism of resistance, under a large environmental effect, or even because of the small size of the populations segregating for the traits. Nevertheless, identification of QTLs on similar chromosomal regions offers the possibility for marker-assisted breeding for *Phytophthora* resistance, which would include selection of favorable alleles for pyramiding resistance genes in specific populations and in recurrent selection.

Witches' broom disease, caused by the fungus *Moniliophthora perniciosa*, is a severe constraint to cacao production in the Americas, where it was responsible for the collapse of the industry in Surinam, Trinidad, Ecuador, and more recently in Brazil. To identify markers associated with witches' broom resistance, an F2 population, derived from selfing 'TSH516', a selected hybrid from an 'ICS1' x 'Scavina6' cross was developed (Queiroz et al. 2003). A major QTL was identified on chromosome 9 near the SSR locus mTcCIR35, responsible for up to 51% of the phenotypic variance for resistance, while a secondary minor QTL was detected on chromosome 1, near a RGA locus (Brown et al. 2005; Faleiro et al. 2006). To enable map-based cloning of this major resistance gene, a BAC library was constructed from 'Scavina 6' (Clement et al. 2004), and a larger F2 population is currently being developed in Brazil.

Frosty-pod rot of cacao, caused by *Moniliophtora roreri*, is another extremely destructive disease restricted to the Americas where it is a serious yield-limiting factor in Central America. The population 'Pound7' x 'UF273' was used to construct a linkage map and it was evaluated for frosty-pod rot resistance by scoring attached pods for internal and external lesions after artificial inoculations with *M. roreri* (Brown et al. 2006). Three major QTLs for resistance against internal and external lesions by *M. roreri* were detected at the same region on chromosomes 2 and 8, and an additional one for external resistance was found on linkage group 7. QTLs associated with witches' broom and frosty-pod resistance are potentially useful in preventive breeding.

B. Yield components, plant vigor and quality traits: In general, the yield of dry fermented cacao seeds produced by a tree is significantly correlated with the total number of pods harvested, but usually not correlated with the other yield components, such as pod weight, seed weight, or number of seeds per pod. Seed yield tends to be correlated with mature tree vigor, as estimated by trunk girth and/or canopy size.

Dry seed production from 55 F1 individuals of the 'Catongo' x 'Pound12' collected over 15 years allowed for the detection of 10 QTLs, distributed on eight chromosomes (Crouzillat et al. 2000b). Two genomic regions, on chromosomes 4 and 5, each explaining ca. 20% of the total phenotypic variance were detected as early as four years after planting and were consistently detected for another 12 years. Similarly, six QTLs for mean seed yield based on nine years of harvest were identified in five linkage groups in families derived from three genotypes (DR1, S52, and IMC78), each crossed with 'Catongo' (Clement et al. 2003b). The yield QTL identified on chromosome 5 in the Forastero 'IMC78' was detected at a later stage (nine years after planting), but it was more repeatable than those from the Trinitarios ('DR1' and 'S52'). The two yield QTLs identified on chromosomes 4 and 5 in 'IMC78' were located close to those detected in 'Pound12', probably because both Forastero genotypes share a common genetic origin in Peru. A QTL for yield was detected around the same region of chromosome 1 for the two Trinitario 'S52' and 'UF676' (Clément et al. 2001).

A major QTL for pod weight in 'IMC78', explaining 43.5% of the phenotypic variation, was detected on chromosome 4 (Clement et al. 2003b) near a similar QTL detected in the Trinitario 'DR1' and the Forastero 'T60/887' (Clément et al. 2001). A QTL for pod index, defined as the number of pods required to produce one kilogram of dry cacao seeds (function of pod weight), was also identified on chromosome 4 of 'Pound12' (Crouzillat et al. 2000b). Similarly, QTLs for pod weight have been co-localized on chromosome 1 for the Trinitario genotypes DR1 and UF676 (Clément et al. 2001).

In terms of seed weight, a major QTL was co-localized on the same region of chromosome 4 in 'Pound12', 'S52' and 'IMC78', explaining 23.6%, 16.2%, and 13.6% of phenotypic variation, respectively (Crouzillat et al. 2000b; Clement et al. 2003a). Another example of co-localization of QTLs included the region of chromosome 9, containing a QTL for seed weight identified in 'S52' and in 'UF676' (Clement et al. 2003a).

In cacao, seed yield at maturity appears to be associated with plant vigor, usually measured by trunk diameter or canopy size. The significant QTLs identified for canopy width for 'DR1' and 'IMC78' were co-localized with yield QTLs, while QTLs for stem diameter and trunk circumference are closely located to yield QTLs in 'IMC78' (Clement et al. 2003b).

QTLs for fat content and flavor quality attributes of cacao seeds (cocoa, floral and fruity flavor; astringency; acidity) have been identified in the 'UPA402' x 'UF676' progeny (Lanaud et al. 2005). Additionally, a specific F2 population was established in Ecuador to identify QTLs responsible for superior flavor aspects (Crouzillat et al. 2001). More recently, genomic regions involved in yield components have been identified by association or admixture mapping, instead of conventional family-derived mapping (Schnell et al. 2005; Marcano et al. 2007).

Association mapping is a new approach based on the occurrence of linkage disequilibrium over extensive genetic distances on chromosome segments in a population derived from recent hybridization, with defined founding individuals. Using a low density genome-wide scan with SSR loci, it should be possible to identify significant linkage disequilibrium and test for statistical association between phenotypes and markers, without the requirement of specific populations derived from controlled crosses and segregation (Schnell et al. 2005; Marcano et al. 2007). The first application of association mapping in cacao analyzed 99 productive and 50 unproductive trees from a population at Waialua, Hawai'i, and individuals from three presumed parental populations to identify associations between markers and yield components (Schnell et al. 2005). From the 65 SSR loci analyzed, 17 displayed a significant association with yield components, whereas 13 (76.4%) were located in genomic regions previously assigned to QTLs for yield components on chromosomes 1, 2, 3, 4, and 9. Co-localization of significant associations and QTL for pod number in 'DR1' and 'IMC78' were identified on chromosomes 4 and 9 (Clement et al. 2003b).

Based on admixture mapping, Marcano et al. (2007) also identified 15 genomic regions associated with seed and fruit weight in two populations, one including 150 Criollo/Trinitario accessions from a germplasm collection in Costa Rica, and the other 291 trees from a plantation in Venezuela. Linkage disequilibrium extended to up to 25-35 cM in both populations. From the 15 identified genomic regions associated with the traits, 10 were localized near previously identified QTLs for the same traits based on four distinct families (Crouzillat et al. 2000b; Clement et al. 2003a), while five were novel regions, indicating the usefulness of this new approach.

Association mapping offers great potential for cacao breeding, because it minimizes the requisite of developing specific populations, an impediment due to long juvenile phase. This mapping approach opens new possibilities with existing breeding or commercial populations, and germplasm collections, currently under molecular characterization.

Genomic regions associated with yield components and plant vigor have been identified in some of the linkage maps, and more recently by admixture mapping. Many showed co-localization on chromosomes across different genotypes,

especially those sharing a common origin. For example, chromosome 4 displayed various major QTLs for distinct yield components for more than one genotype (Crouzillat et al. 2000b; Clement et al. 2003; Schnell et al. 2005; Marcano et al. 2007). Some of the QTLs or markers have potential for use in indirect selection, but the effective use of these markers will require new approaches in cacao breeding. Developing specific populations will be required to apply genotype-building strategies (e.g., population recurrent selection), based on combining favorable alleles at all loci with phenotype evaluation.

6.3 Genomics Resources for Cacao

6.3.1 Development of BAC Libraries: BAC Physical Mapping

Two BAC libraries for cacao have been created. The first library was constructed by CIRAD, the French federal agricultural development research agency, using the Scavina-6 genotype (Clement et al. 2004). The second library was constructed by the Clemson University Genomics Institute (CUGI) under a commission from the USDA Miami Tropical Research Station (USDA-ARS Subtropical Horticulture Research Station (SHRS), Miami, FL. The CUGI library was constructed using the LCT-EEN 37 genotype and is commercially available to the public. Each of the BAC libraries contains over 36,864 clones with average insert sizes of 120 kb, representing 10 haploid genome equivalents.

The BAC libraries serve as a vital resource for a variety of structural and evolutionary genomic studies. For example, the BAC clones can be constructed into contigs by restriction fingerprinting analysis. The resulting contigs can then be anchored onto existing genetic maps to generate a physical map of the cacao genome. The BAC clones can also be utilized as a resource for map-based cloning of agriculturally important genes, and can serve as a template for genomic sequencing assembly in the future.

6.3.2 ESTs

One of the keys to gaining a better understanding of how cacao plants grow, develop, and respond to their environment lies in the unraveling of the gene expression networks underlying these processes. Toward this end, several groups have undertaken discovery programs aimed at identifying the genes that are expressed in given tissues at specific times and in response to specific biotic and abiotic stimuli.

Several cacao EST sequencing projects have resulted in 6,569 ESTs being deposited to date in dbEST (http://www.ncbi.nlm .nih.gov/dbEST/) (Jones et al. 2002; Verica et al. 2004). The Masterfoods Company generated a collection of ESTs isolated from cacao leaves and beans. These ESTs were assembled into a unigene set consisting of 1,380 sequences. Amplified inserts for each of the unigenes were used

to construct a microarray which can be used to study the expression of these genes in different tissues and in response to a variety of biotic and abiotic stimuli (Jones et al. 2002).

A second EST collection was developed with a focus on characterizing genes expressed during defense responses in cacao (Verica et al. 2004). Suppressive-subtractive hybridization (SSH) and macroarray analysis were used to identify cacao ESTs representing genes upregulated by defense signaling molecules. Differential expression analysis using macroarrays identified 475 upregulated clones. These clones, as well as 1,639 randomly chosen cDNAs, were sequenced and assembled into contigs. A total of 1,256 unigenes was obtained, including 330 representing upregulated genes. Eight hundred sixty-five unigenes were assigned to functional classes using BLAST. Eight percent of the sequences up-regulated by the defense inducers were similar to defense proteins in other plants. cDNAs were identified with sequences similar to known defense-related genes including heat shock proteins, NPR1 (a transcriptional regulator that mediates the expression of salicylic acid (SA)- and jasmonic acid (JA)- responsive genes), and several pathogenesis related (PR) genes, including chitinases, (shown to enhance resistance against fungal pathogens). An additional 8% of the up-regulated genes are predicted to play roles in signaling, although their roles in defense are unclear. A database including the ESTs from both of these projects was compiled by The Institute for Genomic Resources (TIGR) and is available online (see section 6.3.5).

A major EST sequencing project is currently underway managed by the French cacao research lab within CIRAD (C. Lanaud personal communication). The goal of the project is to sequence and annotate the ends of 200,000 clones from 28 cDNA libraries constructed from mRNA isolated from different cacao tissues, as well as transcripts induced by a variety of biotic and abiotic stimuli. The libraries were contributed in a cooperative effort by the international cacao molecular biology community. The project is scheduled to be completed in 2007, at which time all sequence data will be made publicly available. CIRAD researchers have created an annotated database to house the data and allow text or DNA based searches within a user-friendly web-based interface. With the completion of this project, the accumulated EST resources will contain representatives of nearly all expressed cacao genes. Genes of low abundance and genes exclusive to tissues or inductive conditions not sampled will remain to be discovered.

6.3.3 Microarrays

Although the EST libraries serve as a valuable resource for understanding gene expression, their utility is limited by the fact that they cannot be used simultaneously to compare differences in expression between multiple tissues or between multiple treatments. Microarrays allow the comparative measurement of gene expression levels in thousands of genes in a single experiment. In this type of analysis, sequences corresponding to specific genes are spotted onto array, and RNA from a tissue under

study is labeled then hybridized to the array. The arrays can subsequently be used to study the transcription profiles of all the arrayed genes in a given tissue, or in response to a specific stimulus.

Through the efforts of an international cacao research consortium, a unigene dataset of all publicly available cacao DNA sequences has been compiled (M. Guiltinan, unpublished). Using these sequences, a unigene set consisting of 2,781 sequences was assembled. From these, a set of 50-mer oligonucleotides unique to each unigene sequence was designed and synthesized by the MWG company (Germany) with funding from Mars Inc. Oligonucleotide microarrays were printed with five replicate sets on each slide. The specificity of the arrays was validated by comparison of leaf vs. fruit RNA, demonstrating that the majority of genes detected at higher levels in leaves are involved in photosynthesis, as would be expected. The microarrays have been made available to cacao researchers worldwide and have been used to study pathogen and endophyte interactions by one of the authors (MJG). Although microarray experiments have proved an invaluable research tool, these analyses have not been without their problems. For example, microarray analyses have been shown to have a potentially high rate of both false positives and false negatives. In addition, similar analyses performed in different laboratories have resulted in a limited overlap in gene signatures. To confirm the validity of microarray results, it is necessary to use an independent methodology.

6.3.4 Genetic Transformation System

A genetic transformation system for cacao was developed in the Guiltinan lab (Guiltinan et al. 1998; Antúnez de Mayolo 2003; Antúnez de Mayolo 2003; Maximova et al. 2003, reviewed in Maximova et al. 2007). This system is based on somatic embryogenesis as a regenerative system. Primary somatic embryos are used as explants for co-cultivation with *Agrobacterium*-harboring Ti-plasmids containing the GFP marker gene and the kanamycin selectable marker gene. The use of a dual selection system allowed the optimization of the many variables impinging on transformation frequency and for the visible selection of high expressing embryos. Transformation of cacao, regeneration of plantlets, and their subsequent analysis is a time consuming process requiring about one year from construct to small plantlet. Moreover, although the transformation system was shown to be reproducible, the transformation frequencies are low compared to other species, further increasing the difficulty and cost to produce sufficient numbers of independent transformants for analysis.

Once developed, this system was used to study the function of the cacao chitinase A gene by over-expression of this gene in transgenic cacao plants (Maximova et al. 2005). The *chiA* gene was shown to confer enhanced resistance to a fungal pathogen in the transgenic cacao leaves demonstrating the efficacy of this system as a tool for functional genomics research in cacao. The potential application of this

technology to crop improvement in cacao is uncertain, however, as consumer and industry concerns over the GMO issue continue.

6.3.5 Informatics Databases

The molecular, morphological, and pedigree data for cacao germplasm are currently managed in several international databases. One is International Cocoa Germplasm Database (ICGD) (University of Reading/Euronext.liffe; http://www.icgd.rdg.ac.uk/), which mainly contains passport data, characterization data, and conservation information of cocoa germplasm, supplied directly by research institutes and from various publications. Another database is CocoaGenDB (http://cocoagendb.cirad.fr/), which was developed and maintained at CIRAD (in partnership with ICGD and USDA). CocoaGenDB combines the genomic data of cacao held in CIRAD's TropGENE database and the germplasm information from ICGD. In CocoaGen DB, a Java applet tool was developed for the visualization of genealogy and for alleles tracking among cocoa germplasm. The CocoaGen DB has been modeled to integrate sequences, clustering, Blast and GO annotation and libraries description of the increasing ESTs data in cocoa. Microsatellites and SNPs derived from the analysis of cDNA collection will also be added to the database. The cocoa gene index was initially developed at TIGR based on publicly available EST data. The TIGR database contains a searchable annotated EST unigene database of all cacao ESTs submitted to the GenBank database at the National Center for Biotechnology Information (NCBI). Genbank contains all publicly available cacao genomic, cDNA and EST data.

6.4 Perspectives

Cacao is a crop of major economic importance to smallholder farmers in the developing world and to the food industry worldwide. It is also of major ecological importance as it is cultivated as a sustainable shade crop inside rainforests. The development of genomics resources for cacao will facilitate the breeding of improved varieties of cacao, most urgently those with improved disease resistance. Because of the importance of cacao to the chocolate industry, its value in economic advancement of developing countries, and its potential as an ecologically sustainable crop, applied genomics to this crop has tremendous potential for exerting a wide array of major impacts. In the future, genomic sequencing of cacao will provide a comprehensive dataset from which markers and genes of interest can be chosen for guiding marker assisted selection programs to speed cacao improvement and allow the relatively rapid introgression of resistance genes, genes for yield and quality traits into varieties of cacao adapted to local conditions throughout producing countries. In this way, genomics will have a major impact on reducing poverty, stabilizing

economies, and protecting the fragile environments in the countries where cacao is a major export crop.

In addition to its intrinsic importance as the source of cocoa for the chocolate industry, cacao genomics will also contribute in a wider sense to the growing knowledge base of plant biology in general. Cacao is one of the 14 crops of major economic importance in the Eurosids II group that includes the model plant system Arabidopsis: the others being Brassica sp. (11 crops), cotton, and citrus (Fig. 6.1). There are at least 21 additional species of economic interest in this group. Cacao is particularly well positioned for genomic comparisons with Arabidopsis, cotton, and other species, and offers an important benefit in having a moderately sized, simple diploid genome (in contrast for example to Brassica, citrus, and cotton). Detailed comparative genomics among Arabidopsis and these crops will enrich our understanding of gene function, genome evolution, and developmental mechanisms in all members of the group.

Furthermore, cacao offers unique opportunities for addressing biological questions of broad interest to plant biologists, for example woody tree development, phase change (juvenile to adult growth habit), disease resistance in a tropical

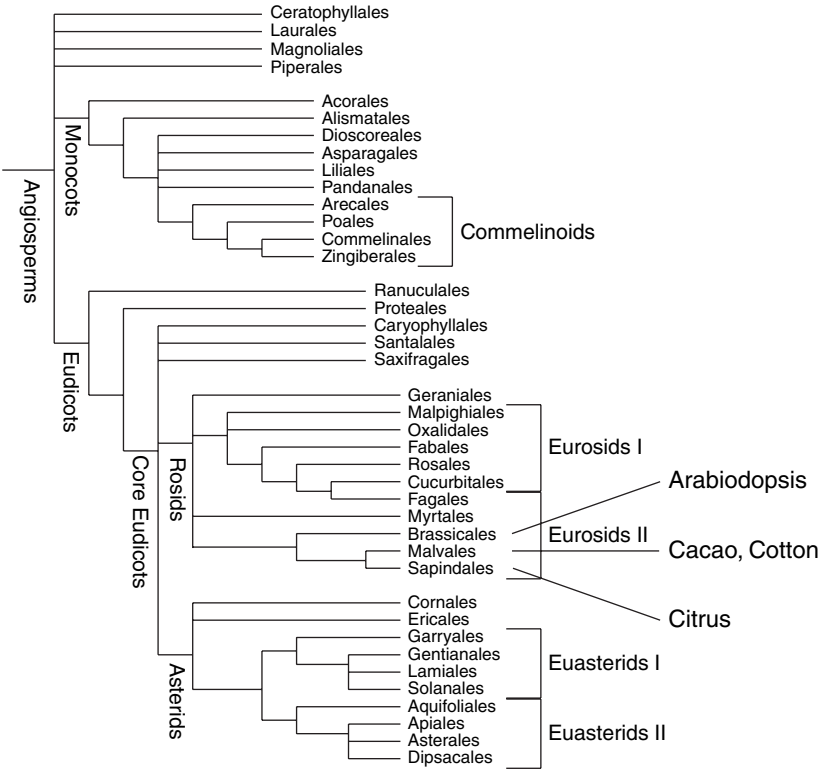

Fig. 6.1 Phylogeny of Flowering Plants Relatedness of Arabidopsis, Brassica, cacao, cotton and citrus

environment, ovarian self-incompatibility, and the phytochemical basis of flavor and aroma. Translation of genomics resources to cacao will make these questions and others accessible to plant biologists in the future. For certain, *Theobroma cacao*, the plant named the Food of the Gods, will hold many interesting and important discoveries for future generations of plant scientists to come.

References

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

Antúnez de Mayolo G (2003) Genetic Engineering Of *Theobroma Cacao* And Molecular Studies On Cacao Defense Responses, A Thesis in Integrative Biosciences. Ph.D. Thesis: The Pennsylvania State University, The Graduate School, Graduate Degree Program in Integrative Biosciences

Antunez de Mayolo G, Maximova SN, Pishak S, Guiltinan MJ (2003) Moxalactam as a counterselection antibiotic for *Agrobacterium*-mediated transformation and its positive effects on *Theobroma cacao* somatic embryogenesis. Plant Sci 164:607–615

Appiah AA, Flood J, Bridge PD, Archer SA (2003) Inter- and intraspecific morphometric variation and characterization of *Phytophthora* isolates from cocoa. Plant Pathol 52:168–180

Appiah AA, Flood J, Archer SA, Bridge PD (2004) Molecular analysis of the major *Phytophthora* species on cocoa. Plant Pathol 53:209–219

Bartley BGD (1994) A review of cacao improvement: fundamentals, methods and results. http://ingenic.cas.psu.edu/proceedings.htm. Proc Intl Workshop Cocoa Breeding Strategies. pp. 3–16

Bartley BGD (2005) The Genetic Diversity of Cacao and its Utilization. (Wallingford, UK: CABI Publishing)

Bennett AB (2003) Out of the Amazon: *Theobroma cacao* enters the genomic era. Trends Plant Sci 8:561–563

Bhattacharjee R, Aikpokpodion P, Kolesnikova-Allen M, Badaru K, Schnell RJ (2004) West African Cocoa: A pilot study on DNA fingerprinting the germplasm from Cross River State of Nigeria. INGENIC Newslett 9:15–20

Boccara M, Zhang D (2006) Progress in resolving identity issues among the Parinari accessions held in Trinidad: the contribution of the collaborative USDA/CRU project. CRU Ann Rpt 2005, Cacao Research Unit, The University of the West Indies, St. Augustine, Trinidad and Tobago

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

Borrone JW, Brown JS, Kuhn DN, Schnell RJ (2007) Microsatellite markers developed from *Theobroma cacao* L. expressed sequence tags. Mol Ecol Notes 7: 236–239

Bowers JH, Bailey BA, Hebbar PK, Sanogo S, Lumsden RD (2001) The impact of plant diseases on worldwide chocolate production (APSNet; Plant Health Progress, http://www.plantmanagementnetwork.org/pub/php /review/ cacao/)

Brown JS, Kuhn DN, Lopez U, Schnell RJ (2005) Resistance gene mapping for witches' broom disease in *Theobroma cacao* L. in an F2 population using SSR markers and candidate genes. J Am Soc Hort Sci 130:366–373

Brown J, Phillips W, Power E, Kroll C, Cervantes-Martinez C, et al. (2006) Preliminary QTL mapping for resistance to frosty pod in cacao (*Theobroma cacao* L.). Proc Intl Cocoa Res Conf 15: (in press)

Buckler ES, Thornsberry J (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5:107–111

Chowdappa P, Brayford D, Smith J, Flood J (2003) Molecular discrimination of *Phytophthora* isolates on cocoa and their relationship with coconut, black pepper and bell pepper isolates based on rDNA repeat and AFLP fingerprints. Curr Sci 84:1235–1237

Christopher Y, Mooleedhar V, Bekele F, Hosein F (1999) Verification of accession in the ICG, T using botanical descriptors and RAPD analysis. Ann Rpt 1998. St. Augustine, Trinidad and Tobago: Cocoa Research Unit, The University of the West Indies. pp. 15–18

Clément D, Risterucci AM, Lanaud C (2001) Analysis of QTL studies related to yield and vigour traits carried out with different cocoa genotypes. Proc Intl Workshop New Technol Cocoa Breeding (INGENIC). pp. 127–134

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

Clement D, Risterucci AM, Motamayor JC, N'Goran J, Lanaud C (2003b) Mapping QTL for yield components, vigor, and resistance to *Phytophthora palmivora* in *Theobroma cacao* L. Genome 46:204–212

Clement D, Lanaud C, Sabau X, Fouet O, Le Cunff L, et al. (2004) Creation of BAC genomic resources for cocoa (*Theobroma cacao* L.) for physical mapping of RGA containing BAC clones. Theor App Genet 108:1627–1634

Couch JA, Zintel HA, Fritz PJ (1993) The genome of the tropical tree *Theobroma cacao* L. Mol Gen Genetics 237:123–128

Crouzillat D, Lerceteau E, Pétiard V, Morera J, Rodriguez H, et al. (1996) *Theobroma cacao* L: A genetic linkage map and quantitative trait loci analysis. Theor Appl Genet 93:205–214

Crouzillat D, Phillips W, Fritz PJ, Petiard V (2000a) Quantitative trait loci analysis in *Theobroma cacao* using molecular markers. Inheritance of polygenic resistance to *Phytophthora palmivora* in two related cacao populations. Euphytica 114:25–36

Crouzillat D, Menard B, Mora A, Phillips W, Petiard V (2000b) Quantitative trait analysis in *Theobroma cacao* using molecular markers. Euphytica 114:13–23

Crouzillat D, Rigoreau M, Cabigliera M, Alvarez M, Bucheli P, Pétiard V (2001) QTL studies carried out for agronomic, technological and quality traits of cocoa in Ecuador. Proc Intl Workshop New Technol Cocoa Breeding (INGENIC). pp. 120–126

Cryer NC, Fenn MGE, Turnbull CJ, Wilkinson MJ (2006a) Allelic size standards and reference genotypes to unify international cocoa (*Theobroma cacao* L.) microsatellite data. Genetic Res Crop Evol 53:1643–1652

Cryer NC, Fenn MGE, Turnbull CJ, Wilkinson MJ (2006b) Allelic size standards and reference genotypes to unify international cocoa (*Theobroma cacao* L.) microsatellite data. Genetic Res Crop Evol, http://dx.doi.org/10.1007 /s10722-005-1286-9

Day RK (1984) Population dynamics of cocoa pod borer *Acrocercops cramerella*: Importance of host plant cropping cycle. Intl Conf Cocoa and Coconuts. pp. 1–9

Dias LAS (2001) Genetic Improvement of Cacao. Editora Folha de Vicosa Ltda. xii + 578 pp. Translation (2005) by Cornelia Elisabeth Abreu-Reichart, Vicosa, M.G., Brazil; aided by the Editor and Peter Griffee of FAO and supported by FAO. Web version: http://ecoport.org/ep?SeachType=earticleView &earticleId=197

Efombagn MIB, Marelli JP, Ducamp M, Cilas C, Nyasse S, et al. (2004) Effects of fruiting traits on the field resistance of cocoa (*Theobroma cacao* L.) clones to *Phytophthora megakarya*. Phytopathology 152:557–562

Efombagn MIB, Sounigo O, Nyass S, Manzanares-Dauleux M, Cilas CB, et al. (2006) Genetic diversity in cocoa germplasm of southern Cameroon revealed by simple sequences repeat (SSRs) markers. African J Biotech 5:1441–1449

Engels J (1981) Genetic resources of cacao, A catalogue of the CATIE collection. (Turrialba, Costa Rica)

Eskes A (2003) Proc Intl Workshop Cocoa Breeding for Improved Production Systems: INGENIC International Workshop on Cocoa Breeding. Bekele F, End M, Eskes A (eds) (Accra, Ghana: INGENIC and Ghana Cocoa Board 2005, http://ingenic.cas. psu.edu/proceedings.htm)

Evans HC, Holmes KA, Reid AP (2003) Phylogeny of the frosty pod rot pathogen of cocoa. Plant Pathol 52:476–485

Faleiro FG, Yamada MM, Lopes UV, Pires JL, Faleiro ASG, et al. (2002) Genetic similarity of *Theobroma cacao* L. accessions mantained in duplicates in the Cacao Research Center germplasm collection, based on RAPD markers. Crop Breed Appl Biotechnol 2:439–444

Faleiro FG, Lopes UV, Yamada MM, Melo GRP, Monteiro WR, et al. (2004) Genetic diversity of cacao accessions selected for resistance to witches' broom disease based on RAPD markers. Crop Breed Appl Biotechnol 4:12–17

Faleiro F, Queiroz V, Lopes U, Guimarães C, Pires J, et al. (2006) Mapping QTLs for witches' broom (*Crinipellis perniciosa*) resistance in cacao (*Theobroma cacao* L.). Euphytica 149:227–235

Figueira A, Janick J, Goldsbrough P (1992) Genome size and DNA polymorphism in *Theobroma cacao*. J Am Soc Hort Sci 117:673–677

Figueira A, Janick J, Levy M, Goldsbrough P (1994) Reexamining the classification of *Theobroma cacao* l. using molecular markers. J Am Soc Hort Sci 119:1073–1082

Figueira A, Albuquerque P, Leal-Jr G (2006) Genetic mapping and differential gene expression of Brazilian alternative resistance sources to witches' broom (causal agent *Crinipellis perniciosa*). Proc Intl Cocoa Res Conf 15: (in press)

Flament MH (1998) Cartographie genetique de facteurs impliques dans la resistance du cacaoyer (Theobroma cacao L.) a *Phytophthora megakarya* et a *Phytophthora palmivora* (Montpellier, France: Ecole Nationale Supérieure Agronomique de Montpellier)

Flament MH, Kebe I, Clement D, Pieretti I, Risterucci AM, et al. (2001) Genetic mapping of resistance factors to *Phytophthora palmivora* in cocoa. Genome 44:79–85

George MLC, Li W, Moju C, Dahlan M, Pabendon M, et al. (2004) Molecular characterization of Asian maize inbred lines by multiple laboratories. Theor Appl Genet. 109:80–91

Guiltinan M (2007) Cacao. In: Pua VE, Davey M (eds) Biotechnology in Agriculture and Forestry - Transgenic Crops. Berlin Heidelbelg: Springer-Verlag

Guiltinan MJ, Li Z, Traore A, Maximova S, Pishak S (1998) High efficiency somatic embryogenesis and genetic transformation of cacao. Ingenic Newsl

Guimarães CT, Queiros VT, Mota JWS, Pereira MG, Daher RF, et al. (2003) A cocoa genetic linkage map and QTL detection for witches' broom resistance. Plant Breeding 122: 268–272

Guyton B, Lumsden R, Matlick BK (2003) Strategic plan for sustainable cocoa production. Manufacturing Confectioner 83:55–60

Hanna AD (1954) Application of a systemic insecticide by trunk implantation to control a mealybug vector of the cacao swollen shoot virus. Nature 173:730–731

Hervé L, Djiekpor E, Jacquemond M (1991) Characterization of the genome of cacao swollen shoot virus. J Gen Virol 72:1735–1739

Hughes JA, Ollennu LAA (1994) Mild strain protection of cocoa in Ghana against cocoa swollen shoot virus-a review. Plant Pathol 43:442–457

Iwaro AD, Bekele FL, Butler DR (2003) Evaluation and utilization of cacao (*Theobroma cacao* L.) germplasm at the International Cocoa Genebank, Trinidad. Euphytica 130:207–221

Johnson L (2003) INGENIC Workshop, Cocoa genomics group. Gro Cocoa 4, 4–5 (http://www.cabi-commodities.org/Acc/ACCrc/PDFFiles/GROC/GROC.htm)

Johnson SE, Mora A, Schnell (2007) Field Guide efficacy in the identification of reallocated clonally propagated accessions of cacao. Genetic Res Crop Evol (in press)

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

Keane PJ (1992) Diseases and pests of cocoa: An overview. Cocoa pest and disease management in Southeast Asia and Australasia, FAO Plant Production and Protection Paper 112:1–12

Kuhn DN, Heath M, Wisser RJ, Meerow A, Brown JS, et al. (2003) Resistance gene homologues in *Theobroma cacao* as useful genetic markers. Theor Appl Genet 107:191–202

Kuhn, DN, Borone J, Meerow A, Motamayor JC, Brown JS, et al. (2004) Single strand conformation polymorphism analysis of candidate genes for reliable identification of alleles by capillary array electrophoresis. Electrophoresis 26:112–115

Lachenaud P, Sounigo O, Oliver G (2004) Genetic structure of Guianan wild cocoa (*Theobroma cacao* L.) described using isozyme electrophoresis. Plant Genetic Res Newsl 139:24–30

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 30:259–270

Lanaud C, Risterucci AM, N'Goran AKJ, Clement D, Flament MH, et al. (1995) A genetic linkage map of *Theobroma cacao* L. Theor Appl Genet 91:987–993

Lanaud C, Kebe I, Risterucci AM, Clement D, N'Goran JKA, et al. (1996) Mapping quantitative trait loci (QTL) for resistance to *Phytophthora palmivora* in *T. cacao*. Proc. Intl Cocoa Res Conf 12:99–105

Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet A, et al. (1999) Isolation and characterization of microsatellites in *Theobroma cacao* L. Mol Ecol 8:2141–2143

Lanaud C, Kébé I, Risterucci A, Clément D, N'Goran J, et al. (2000) Mapping quantitative trait loci (QTL) for resistance to *Phytophthora palmivora* in *T. cacao*. Proc Intl Cocoa Res Conf 12:99–105

Lanaud C, Risterucci AM, Pieretti I, N'goran JAK, Fargeas D (2004) Characterisation and genetic mapping of resistance and defence gene analogs in cocoa (*Theobroma cacao* L.). Mol Breeding 13:211–227

Lanaud C, Boult E, Clapperton J, N'Goran J, Cros E, et al. (2005) Identification of QTLs related to fat content, seed size and sensorial traits of *Theobroma cacao*. Proc Intl Cocoa Res Conf 13:1119–1126

Laurent V, Risterucci AM, Lanaud C (1994) Genetic diversity in cocoa revealed by cDNA probes. Theor Appl Genet 68:193–195

Lerceteau E, Robert T Pétiard V, Crouzillat D (1997) Evaluation of the extent of genetic variability among *Theobroma cacao* accessions using RAPD and RFLP. Theor Appl Genet 95:10–19

Lockwood G, Gyamfi MMO (1979) The CRIG cocoa germplasm collection with notes on codes used in the breeding programme at Tafo and elsewhere (Ghana: Cocoa Research Institute). 62 pp

Marcano M, Pugh T, Cros E, Morales S, Portillo Paez EA, et al. (2007) Adding value to cocoa (*Theobroma cacao* L.) germplasm information with domestication history and admixture mapping. Theor Appl Genet 114: 877–884

Marita JM, Nienhuis J, Pires JL, Aitken WM (2001) Analysis of genetic diversity in *Theobroma cacao* with emphasis on withes' broom disease resistance. Crop Sci 41:1305–1316

Maximova S, Miller C, Antunez de Mayolo G, Pishak S, Young A, et al. (2003) Stable transformation of *Theobroma cacao* L. and influence of matrix attachment regions on GFP expression. Plant Cell Rep 21:872–883

Maximova SN, Marelli JP, Young A, Pishak S, Verica JA, et al. (2005) Over-expression of a cacao class I chitinase gene in *Theobroma cacao* L. enhances resistance against the pathogen, *Colletotrichum gloeosporioides*. Planta 224:740–749

Maximova, SN, Tan CL, Guiltinan MJ (2007) Cocoa In: Kole C, Hall TC (eds) A Compendium of Transgenic Crop Plants. Volume 7. Blackwell Publishing, Malden, MA, USA: In Press

Morais RC (2005) The Genomes of cocoa. In Forbes,(March 14) pp. 110–112

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

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

Motilal L, Butler D (2003) Verification of identities in global cacao germplasm collections. Genetic Res Crop Evol 50:799–807

Motilal L, Sounigo O, Thévenin J, Risterucci A, Pieretti I, et al. (2002). *Theobroma cacao* L.: genome map and QTLs for *Phytophthora palmivora* resistance. Proc Intl Cocoa Res Conf 13 Kota Kinabalu, Malaysia, 2000. Cocoa Producer's Alliance, Lagos, Nigeria, pp. 111–118

Muller E, Sackey S (2005) Molecular variability analysis of five new complete cacao swollen shoot virus genomic sequences. Arch Virol 150:53–66

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

Nyasse S, Despreaux D, Cilas C (2002) Validity of a leaf inoculation test to assess the resistance to *Phytophthora megakarya* in a cocoa (*Theobroma cacao* L.) diallel mating design. Euphytica 123:395–399

Opoku IY, Akrofi AY, Appiah AA (2002) Shade trees are alternative hosts of the cocoa pathogen *Phytophthora megakarya*. Crop Protection 21:629–634

Opoku IY, Bhattacharjee R, Kolesnikova-Allen M, Enu-Kwesi L, Asante E, et al. (2005) Impact of Breeders' Collection on Cocoa Plantings of Ghana: Assessment by Molecular Marker Analysis and Farmers' Field Survey. A paper presented in First COCOBOD Conference and Promotion of Local Consumption of Cocoa Products and 24 th Biennial Conference of The Ghana Science Association. Accra Ghana . 1–4 August 2005

Pereira JL, Ram A, Defigueiredo JM, Dealmeida LCC (1990) First occurrence of witches' broom disease in the principal cocoa-growing region of Brazil. Trop Agric 67:188–189

Perry MD, Davey MR, Power JB, Lowe KC, Bligh HFJ, et al. (1998) DNA isolation and AFLP genetic fingerprinting of Theobroma cacao (L.). Plant Mol Biol Rep 16:49–59

Presson A, Sobel E, Lange K, Papp J (2006) Merging microsatellite data. J Comput Biol 13:1131–1147

Piasentin F, Klare-Repnik L (2004) Biodiversity conservation and cocoa agroforests. Gro Cocoa 5:7–8

Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, et al. (2004) Geneclass2: A Software for Genetic Assignment and First-Generation Migrant Detection. J Heredity 95:536–539

Pound FJ (1940) Witches' broom resistance in cacao. Trop Agric 17:6–8

Pritchard JK, Stephens M, Donnelly PJ (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Pugh T, Fouet O, Risterucci AM, Brottier P, Abouladze M, et al. (2004) A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. Theor Appl Genet 108:1151–1161

Queiroz VT, Guimaraes CT, Anhert D, Schuster I, Daher RT, et al. (2003) Identification of a major QTL in cocoa (*Theobroma cacao* L.) associated with resistance to witches' broom disease. Plant Breed 122:268–272

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Rice RA, Greenberg R (2003) The Chocolate Tree: Growing cacao in the forest. Natural History 112:36–43

Risterucci AM, Grivet L, Ngoran JA, Pieretti I, Flamen, MH, et al. (2000) A high-density linkage map of *Theobroma cacao* L. Theor Appl Genet 101:948–955

Risterucci AM, Paulin D, Ducamp M, N'Goran JAK, Lanaud C (2003) Identification of QTLs related to cocoa resistance to three species of *Phytophthora*. Theor Appl Genet 108:168–174

Ronning CM, Schnell RJ (1994) Allozyme Diversity in a Germplasm Collection of *Theobroma cacao* L. J. Heredity 85:291–295

Rodriguez LCM, Wetten AC, Wilkinson MJ (2004) Detection and quantification of in vitro-culture induced chimerism using simple sequence repeat (SSR) analysis in *Theobroma cacao* (L.). Theor Appl Genet 110:157–166

Ruf F, Zadi H (2003) Cocoa: From deforestation to reforestation (Migratory Bird Center: Smithsonian National Zoological Park, (http://nationalzoo.si.edu/ConservtionAndScience/MigratoryBirds/Research/Cacao/ruf.cfm))

Russell JR, Hosein F, Johnson E, Waugh R, Powell W (1993) Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis. Mol Ecol 2:89–97

Santoso D, Chaidamsari T, Wiryadiputra S, de Maagd RA (2004) Activity of *Bacillus thuringiensis* toxins against cocoa pod borer larvae. Pest Management Sci 60:735–738

Sampaio P, Gusmão L, Alves C, Pina-Vaz C, Amorim A, Pais C (2003). Highly polymorphic microsatellite for identification of *Candida albicans* strains. J Clin Microbiol 41:552–557

Saunders JA, Hemeida AA, Mischke S (2001) USDA DNA fingerprinting programme for identification of *Theobroma cacao* accessions. p. 108–114. In F Bekele et al (ed) Proc Intl Workshop New Technol Cocoa Breeding 2000. INGENIC Press, London

Sereno ML Albuquerque PSB, Vencovsky R, 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

Schnell RJ, Olano CT, Brown JS, Meerow AW, Cervantes-Martinez C, et al. (2005) Retrospective determination of the parental population of superior cacao (*Theobroma cacao* L.) seedlings and association of microsatellite alleles with productivity. J Am Soc Hort Sci 130:181–190

Silva CRS, Figueira A (2005) Phylogenetic analysis of *Theobroma* (Sterculiaceae) based on Kunitz-like trypsin inhibitor sequences. Plant Syst Evol 250:93–104

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

Takrama JF, Cervantes-Martinez CT, Phillips-Mora W, Brown JS, Motamayor JC, et al. (2005) Determination of off-types in a cacao breeding program using microsatellites. INGENIC Newsl 10:2–8

Toxopeus H (1969) Cacao (*Theobroma cacao*). pp 79–109. In: Fewerda FP (ed) Outlines of perennial crop breeding in the tropics, (Landbouwhogesschool Wageningen, The Netherlands)

Turnbull CJ, Butler DR, Cryer NC, Zhang D, Lanaud C, et al. (2004) Tackling mislabelling in cocoa germplasm collections. INGENIC Newsl 9:8–11

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

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, et al. (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

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

Whitkus R, De La Cruz M, Mota-Bravo L, Gomez-Pompa A (1998) Genetic diversity and relationships of cacao (*Theobroma cacao* L.) in southern Mexico. Theor Appl Genet 96:621–627

Wilde J, Waugh R, Powell W (1992) Genetic fingerprinting of *Theobroma* clones using randomly amplified polymorphic DNA markers. Theor Appl Genet 83:871–877

Wood GAR, Lass RA (1985) Cocoa. (New York: Longman Scientific & Technical)

Yamada MM, Faleiro FG, Lopes UV, Bahia RCS, Pires JL, et al. (2001) Genetic variability in cultivated cacao populations in Bahia, Brazil, using isozymes and RAPD markers. Crop Breed Appl Biotechnol 1:377–384

Zhang D-X, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Mol Ecol 12:563–584

Zhang D, Mischke S, Goenaga R, Hemeida AA, Saunders JA (2006a) Accuracy and reliability of high-throughput microsatellite genotyping for cacao clone identification. Crop Sci 46:2084–2092

Zhang D, Arevalo-Gardini E, Mischke S, Zúñiga-Cernades L, Barreto-Chavez A, et al. (2006b) Genetic diversity and structure of managed and semi-natural populations of cacao (*Theobroma cacao*) in the Huallaga and Ucayali valleys of Peru. Ann Bot 98:647–655