Current Challenges of Tropical Tree Crop Improvement: Integrating Genomics into an Applied Cacao Breeding Program

R.J. Schnell, J.S. Brown, D.N. Kuhn, C. Cervantes-Martinez, J.W. Borrone and C.T. Olano
USDA, Miami, FL
USA

J.C. Motamayor
Mars Inc., Miami, FL
USA

U. Lopes
Comissão Executiva do Plano da Lavoura Cacaueira (CEPLAC)
Ilheus, Bahia
Brazil

W. Phillips, E. Johnson and E.J. Monteverde-Penso
Centro Agronómico Tropical de Investigación y Ensenañza (CATIE)
Turrialba
Costa Rica

F. Amores
Estación Experimental Tropical Pichilingue (EET)
Quevedo
Ecuador

Keywords: *Theobroma cacao*, molecular markers, MAS, QTL, linkage mapping

Abstract

Theobroma cacao **L. is an understory tree from the Amazon basin that can be cultivated in a sustainable agro-forestry system, providing income to small farmers while maintaining biodiversity. Four main genetic groups of cacao are traditionally described: Criollo, Trinitario, and lower and upper Amazon Forastero. During the 17th and 18th centuries, plants derived from a small number of parents were distributed to many tropical regions of the world, resulting in commercial plantings with a narrow genetic base. Production of cacao in tropical America has been severely affected by two fungal pathogens causing diseases known as witches' broom (WB) and frosty pod (FP). These, along with another pan-tropical fungal disease, black pod (BP), were responsible for over 700 million USD in losses in 2001. Currently, WB and FP are confined to tropical America; however, commercial populations in West Africa and South Asia are highly susceptible to both diseases. Traditional cacao breeding programs have only been marginally successful in producing resistant material with suitable commercial characteristics. In 1999, the USDA-ARS in collaboration with Mars Inc. initiated a project to apply modern molecular genetic techniques to cacao breeding. The objectives were to develop an international Marker Assisted Selection (MAS) breeding program focusing on disease resistance, to identify new sources of resistance in unexploited germplasm, and to identify the genes involved with disease resistance. Over 320 microsatellite and 50 candidate gene markers are being used to map families segregating for resistance to WB, FP, and BP diseases. Quantitative Trait Loci (QTL) have been identified for resistance to WB and FP and these are being employed in MAS. The utility of Association Mapping for productivity traits has been demonstrated providing an alternative method to traditional mapping. Microsatellite and candidate gene markers have also been used to estimate the genetic diversity in over 1,300 individuals representing 70 different domesticated and semi-domesticated/wild cacao populations. Genetic diversity has been found to be much higher in the semi-domesticated/wild populations from the upper Amazon which may contain new sources of disease resistance. For the gene discovery effort, a Bacterial Artificial Chromosome (BAC) library has been produced from 'LCTEEN37', resistant to WB, and work is underway to identify and sequence gene(s) responsible for a major WB QTL. Large evaluation trials, developed using MAS, are located in Costa Rica, Ecuador, Brazil, and Papua New Guinea with additional QTL evaluation studies in Ghana, Nigeria, Costa Rica, and Ecuador. All these projects are collaborations with**

Proc. IS on Biotechnol. Temp. Fruit Crops & Trop. Species
Eds. R.E. Litz and R. Scorza
Acta Hort. 738, ISHS 2007

national agricultural institutes in the respective countries. The international MAS project is expected to produce new disease resistant cultivars by 2012. Genetic stocks developed in this project will be distributed to areas currently free of WB and FP in anticipation of the arrival of these diseases. International collaboration and sharing of genetic resources will ensure that crop losses due to these pathogens are manageable and will contribute to stability in the supply of cocoa beans.

INTRODUCTION

Theobroma cacao L. is an understory tree from the Amazon basin and the seeds are used as raw material for chocolate containing products. Cacao is an out-crossing diploid ($n=x=10$) with a relatively small genome (447 Mb) (Guiltinan, pers. commun.). Much genetic diversity exists within *T. cacao*; however, current commercial cultivars have a very narrow genetic base (Warren and Kennedy, 1991). The primary types of cacao grown have traditionally been referred to as Forastero, Trinitario, and Criollo, with most of the world's production based on Forastero-derived types. The Forastero material is usually further divided into upper Amazon and lower Amazon types. The Trinitario are hybrids between Forastero and Criollo types (Cheeseman, 1944; Motamayor et al., 2002, 2003).

Cacao was domesticated in northern South America; however, cultivation and the cultural use of cacao were much more extensive in Mesoamerica. According to the Maya and Aztec religions, cacao was a gift from the gods and was used in many ritual ceremonies, as a medicine, and by the nobility-specifically priests, government officials, and military officers. Cacao also has a long history in Ecuador where a special flavor, high value cacao known as Nacional is produced. Wild Criollo cacao trees existed near Lake Maracaibo in Venezuela. These were used by the Indians for ritual ceremonies and were the source of the seed for the first plantations in Venezuela. In Brazil extensive areas of wild Forastero types existed along the banks of the Amazon from where the first Amelonado was collected in 1639. The Amelonado type was taken to Bahia in 1750 and this became the major production area in Brazil. The Spanish were the first Europeans to encounter cacao when they conquered the Aztecs in 1519 and it was first prepared as a beverage for the Spanish Court in 1544. The drink became popular in Europe causing the French to establish cacao plantations in the Caribbean while the Portuguese, Spanish, and English established plantations in colonies in West Africa and Asia. The latter remain major production areas today (Dillinger et al., 2000).

Trinidad holds a special place in the genetics and breeding of cacao, and currently houses the most extensive international gene bank for cacao in the world. The first Criollo plantings were in 1525 by the Spanish, followed by the development of the Trinitaro types in the 18th century. The first scientific breeding effort was established in the early 1930s and led to the selection of the first 100 superior clones named the Imperial College Selections (ICS; Pound, 1943). A number of breeding programs have been developed in Central and South America, West Africa, and South Asia; however most have only been marginally successful in genetically improving cacao (Lockwood, 2003).

Cacao cultivation and production of cacao beans for the American chocolate industry is a multibillion-dollar effort centered largely in Africa, Asia, and South America. Significant amounts of USA-produced agricultural commodities, including milk, sugar, almonds, peanuts, and corn syrup sweeteners, are used in the chocolate and confectionery industry (http://www.chocolateusa.org/about-choc/facts/trivia/economic. asp). Cacao production is plagued by very serious losses globally from pests and diseases (Table 1). For the long term wellbeing of the industry, there is a need to foster international collaboration among producers, and to develop superior trees that can stand up to disease and still produce high quality cocoa beans. In 1999, the USDA-ARS in collaboration with Mars Inc. initiated a project to apply modern molecular genetic techniques to cacao breeding. The primary goal of this project is to develop a Marker Assisted Selection (MAS) program and to disseminate new, productive, disease resistant cultivars of cacao.

In order to facilitate the development of a MAS program, several steps needed to be implemented. 1) Cooperative projects had to be developed with national agricultural

research services and international organizations actively working on disease resistance and increased productivity. 2) Molecular markers had to be developed for DNA fingerprinting, genetic diversity analysis, and the production of linkage maps. 3) Mapping populations had to be identified that could be used to locate quantitative trait loci (QTL) for disease resistance, quality, and productivity. 4) Molecular techniques needed to be applied to identify the genes involved in the disease resistance mechanism.

SPECIFIC COOPERATIVE AGREEMENTS WITH NARS AND INTERNATIONAL INSTITUTES

Specific Cooperative Agreements have secured cooperation with national and international research organizations in Central and South America, West Africa, and South Asia. The South and Central American institutes are the Tropical Agricultural Research and Education Center (CATIE) in Turrialba, Costa Rica, the Instituto Nacional de Investigaciones Agricolas y Pecuarias (INIAP) Estacion Experimental Pichilingue (EET Pichilingue) in Quevedo, Ecuador, and the Mars Almirante Experimental Station in Bahia, Brazil. Collaboration with West African institutes is through the International Institute for Tropical Agriculture (IITA) in Nigeria and the Cocoa Research Institute of Ghana (CRIG). In Asia, collaboration is with the Coconut and Cacao Institute (CCI) in Papua New Guinea. The number of genotypes currently under evaluation in the breeding program is approximately 23,000 (Table 2). Formal, non-funded agreements are also maintained with Centre de Cooperation Internationale en Recherche Agronomique pour le Developpement (CIRAD) in France, Comissão Executiva do Plano da Lavoura Cacaueira (CEPLAC) in Ilheus and Belem in Brazil, the University of Reading in the UK, the University of Florida, Florida International University, and Pennsylvania State University in the USA.

MARKER DEVELOPMENT

Microsatellite Markers

Approximately 320 microsatellite markers have been developed by CIRAD and USDA-Miami (Lanaud et al., 1999; Brown et al., 2005; Pugh et al., 2004). Polymerase Chain Reaction (PCR) amplifications are carried out according to the protocol developed by Schnell et al. (2004). Microsatellite alleles differ in length and can be assayed by Capillary Electrophoresis (CE). CE is performed on an ABI Prism 3100 or an ABI Prism 3730 Genetic Analyzer (Applied Biosystems, Inc.). Resulting data are analyzed with GeneMapper 3.0 (Applied Biosystems, Inc.) for internal standard and fragment size determination and for allelic designations. The microsatellite names, primer sequences, annealing temperatures, and repeat type are listed at http://www.SHRS.cacao.primers.

Identification of Candidate Gene Markers Using Degenerate PCR Primers

Resistance Gene Homologues (RGH) of the Nucleotide Binding Site/Leucine Rich Repeat (NBS/LRR) class and WRKY transcription factor genes were isolated from cacao by using degenerate PCR primers for a highly conserved region of these candidate genes (Kuhn et al., 2003; Borrone et al., 2004). In the initial RGH studies, we did not identify any Toll Interleukin Receptor (TIR) NBS/LRR genes, even though these genes should occur as frequently as the NBS/LRR genes. Using a novel algorithm to design degenerate primers, we amplified RGH that contain the TIR motif in addition to the NBS and LRR domains, and identified TIR NBS/LRR genes from cacao. Despite >90% nucleotide sequence identity, the sequences did not represent a single cluster in the cacao genome. Three of the TIR loci were mapped and two were found to be on different chromosomes (Kuhn et al., 2006).

Validation of the Single Strand Conformation Polymorphism (SSCP) Assay for Candidate Genes

Single-nucleotide polymorphisms (SNPs), both as single base pair substitutions and single base pair insertions/deletions (indels) are the most common sequence differences found between alleles. Methods have been developed for high-throughput detection of SNPs, but these methods require a priori knowledge of the SNP being assayed or sequence information surrounding the SNP. A method to detect novel polymorphisms without a priori knowledge is SSCP. SSCP is a sensitive, economical procedure that indirectly detects SNPs in amplified DNA fragments of the same length. Polymorphisms are detected as alterations of mobility induced by nucleotide differences that cause stable changes in conformation of single strand DNA. Initially developed for polyacrylamide gel electrophoresis, SSCP has been adapted to CE.

We have validated the CE-SSCP method and its usefulness for estimating genetic diversity in a high-throughput analysis (Kuhn et al., 2005; Kuhn and Schnell, 2005). We were able to distinguish novel alleles of a candidate gene by mobility alone if fragments of 150-250 nucleotides were used. In that size range, CE-SSCP could reliably distinguish alleles that differed by a single nucleotide in sequence using a commercially available molecular weight standard. This information has provided the technical foundation for the development of genetic markers from candidate genes by two strategies: using degenerate PCR primers to regions of sequence that are highly conserved across many species and mining cacao Expressed Sequence Tag (ESTs) available from GenBank.

Development of Candidate and Microsatellite Markers from the EST Database

Candidate gene markers developed from EST have several advantages. 1) They can be mined from available EST data. 2) They can be identified by comparing the translated EST sequence to GenBank database sequences. 3) By using the chromosomal position of the candidate gene in *Arabidopsis* to identify other genes that may be near it (synteny), we may identify other candidate genes closer to the QTLs for witches' broom (WB) and frosty pod (FP) resistance. To test the feasibility of developing markers from ESTs, 39 different ESTs were chosen and SSCP primers produced (labeled on both strands). Eight of these ESTs were chosen because they were almost identical at the nucleotide level with cotton and *Arabidopsis* ESTs. The other ESTs were chosen because they coded for proteins involved in disease or stress resistance (two putative NBS/LLR genes, hypersensitive-induced protein, SA induced protein, O-succinyl homoserine lyase, putative mitogen activated protein kinase, chitinase, cellulase, elicitor induced protein, drought induced protein, water stress protein, and dehydration induced protein), fatty acid or lipid synthesis (stearoyl ACP desaturase, oleoyl ACP thioesterase, 3-oxoacyl-ACP reductase, 3-ketoacyl CoA thiolase, diacyl glycerol kinase, omega-6 desaturase, malonyl CoA-ACP transacylase), or developmental regulatory genes (MADS-box genes CONSTANS-like and SQUAMOSA, and an AP2/EREBP ovule development protein). To determine if they were polymorphic and if they could be mapped, the primers were tested against 95 different cacao cultivars, which included the parents of all cacao mapping populations, in addition to cultivars known to be heterozygous and genetically diverse.

For the 39 primer pairs tested, 6 of 39 (15%) did not amplify even after being tested with temperature gradient; 6 of the 39 (15%) gave products too large or amplification was too low to be analyzed by SSCP. Of the 27 primer pairs that gave sufficient amplification of fragments between 150 and 250 nucleotides in length, 12 (44%) can be mapped. CA795469 (a putative kinase and one of the ESTs that shared high nucleotide identity with cotton and *Arabidopsis*) mapped to the very end of linkage group 1, extending the map in that region by 4 cM.

We have also developed microsatellite markers from these ESTs. ESTs present in GenBank May 2005 were converted into a set of 2336 unigenes. Two hundred and ten unigenes were identified as containing 293 SSRs, with AG-rich repeats dominating all repeat size ranges. Primers were designed for 43 SSR-containing unigenes; 29 primer pairs gave PCR products of the expected size. PCR products from 26 primer pairs were

polymorphic in length, and 28 loci could potentially be mapped in at least one segregating population, with two primer pairs each amplifying an additional product. The overall 60% success rate suggests that approximately 120 useful EST-SSR markers could be developed from the current *T. cacao* unigene set (Borrone et al., 2006).

Use of Markers in the Breeding Program

1. Parentage Analysis in Full-Sib Families and Determination of Offtypes in Germplasm Collections. A major problem in cacao breeding is the inconsistency in the performance of known clones when used as parents in field trials, in part due to incorrect genetic identification. It has been estimated that the misidentification of cacao accessions could be as high as 20 to 30% in some major germplasm collections (Motilal and Butler, 2003; Saunders et al., 2004). The utility of microsatellites as a tool for the identification of mislabeled accessions in field genebanks has been demonstrated by several researchers in the cacao community (Motilal and Boccara, 2004; Turnbull et al., 2004). Schnell et al. (2004) demonstrated with microsatellite markers that many seedlings from the important cross of 'SCA6' x 'ICS1' made in Trinidad to select for WB resistance were misidentified. Parentage analysis was performed using the program CERVUS (Marshal et al., 1998; Slate et al., 2000). This software uses a simulation program to generate log-likelihood scores and provides a confidence statistic for assigning paternity. Parentage analysis of the 186 trees sampled, assigned 134 as true hybrids (72.0%) with 80% confidence. In cases where 'ICS1' or 'SCA6' was not identified as the most likely parent, from 3 to 17 mismatching loci occurred. The remaining 52 offtype trees could be categorized as selfs of 'ICS1' (6 plants), 'ICS1' x unknown (13 plants), 'SCA6' x unknown (2 plants), and 31 plants that had neither 'ICS1' nor 'SCA6' as a parent.

Making controlled crosses for genetic studies is essential for cacao improvement, and the above results demonstrate the need to confirm the integrity of parents and progeny with molecular markers, until the crossing technique used by any given research group is routinely accurate. Obviously, strict attention must be paid to avoid pollen contamination, as a high number of off-type plants renders a population unsuitable for scientific studies, and can hinder breeding progress when an incorrect parent is identified as a progenitor of a productive seedling.

Correct identification of full-sib families is important in the estimation of heritabilities, predictions of genetic gain, and identification of superior parents. As demonstrated in this study, pollen contamination and mislabeling of plant families has been a serious problem in cacao improvement programs and in germplasm management. Other microsatellite analyses on segregating populations carried out in our lab have shown even higher proportions of offtype progeny. Microsatellite analyses are continuously being used to check pollinations before establishing family or QTL trials, because the benefits of their use offset their expense and laboriousness.

2. Genetic Diversity Analysis. Genetic diversity among 1300 individuals from >70 localities has been analyzed using one hundred microsatellite markers. To avoid the influence of mislabeled samples, Bayesian statistics and computer simulations were used to infer genetic groups for population genetic analyses using the program Structure (Prichard et al., 2000). Structure does not require prior information on the population of origin and we were able to exclude mislabeled individuals identified using this approach. At least thirteen major genetic groups were delineated based on this analysis, instead of the traditional genetic groups Criollo, Trinitario, and upper and lower Amazon Forrastero. A genetic distance matrix between individuals was estimated using Shared Allele Distance coefficient and a Multidimensional Scaling Plot generated using SAS version 9.1 (SAS Institute Inc., Cary, NC). Figure 1 illustrates the genetic distances among the individuals from the 13 groups. Individuals of the same population cluster together except those from the Solimoes group which spread from the left to the right half of the plane.

We compared the diversity of germplasm traditionally used in breeding and the unexploited germplasm. Much higher diversity values were estimated for the unexploited germplasm than for the traditionally used parental cultivars. One parameter is of

particular interest: the number of alleles exclusive to each group. The germplasm traditionally used only has 44 exclusive or "private alleles", whereas the unexploited germplasm contained 471 such alleles (Table 3). Based on our results, most of the recently collected germplasm has not been used in breeding programs.

LINKAGE MAPPING AND QTL DISCOVERY

The First F_2 Map in Cacao Made from Microsatellite Markers, and the Identification of QTLs for WB Resistance

This cross was produced, grown and evaluated for WB resistance in CEPLACI. Tissue samples were collected at CEPLAC and sent to the SHRS for DNA extraction and fragment analysis. A genetic linkage map was created from 146 cacao trees using this F_2 population produced by selfing an F_1 progeny ('TSH516') of the cross 'ICS1' x 'SCA6' (Brown et al., 2005; Faleiro et al., 2006). Microsatellite markers (170) were used principally for this map, with 12 candidate genes (eight RGH and four stress related WRKY genes), for a total of 182 markers. Joinmap® software (Van Ooijen and Voorrips, 2001) was used to create the map, and 10 linkage groups were obtained, corresponding to the 10 known chromosomes of cacao. Our map encompassed 671.9 cM, approximately 100 cM less than most previously reported cacao maps, and 213.5 cM less than the one reported high-density map (Pugh et al., 2004). Approximately 27% of the markers showed significant segregation distortion, mapping together in six genomic areas. Four of these markers also showed distortion in other cacao maps. As the map comes from an F_2 population, in which there are fewer segregation types than in a cross of two heterozygous F_1 clones, it is more robust. There was remarkably good correlation between marker order and relative distances between markers when compared with the map of Pugh et al. (2004), produced from a greater number of codominant markers. Only four pair-wise inversions of single marker positions differed between the two maps, and the mean Pearson correlation coefficients across linkage groups for positions of markers in common was 0.9927. Two QTL for resistance to witches' broom disease were found, one producing a major effect and one a minor effect, both showing important dominance effects (Fig. 2). One RGH flanked the minor QTL for witches' broom resistance, implying possible association. QTLs mapped in F_2 populations produce estimates of additive and dominance effects, not obtainable in F_1 crosses. As dominance was clearly shown in the QTL found in this study, this population merits further study for evaluation of dominance effects for other traits.

A Second Cacao Map from the 'Pound7' x 'UF273' to Identify QTLs for Frosty Pod Resistance

This cross was produced, grown, and agronomically evaluated at CATIE. Tissue samples were collected and brought to the SHRS for DNA extraction and fragment analysis. One hundred eighty loci were analyzed for the entire population, although gametes from one parent, 'UF273', had come from two different trees, resulting in two different populations, one with 185 trees, and one with 71 trees, differing at 22 loci. We were able to use all 256 trees by entering missing data for the 71 trees at the loci where they differed from the larger population of 185 trees. Some robustness was lost due to the 22 missing loci; however, the net gain in map precision was larger when the two populations were joined in this manner. Eleven linkage groups were found when the map was composed using recombination distances no larger than 0.5 and LOD scores no larger than 1.0 (Fig. 2). In the map composed from the 185 tree population, there was very little segregation distortion; however, in the combined map, 33 loci showed highly significant segregation distortion, some of which was likely due either to missing data or from combining the two populations. The extra (eleventh) linkage group is actually a piece of linkage group 3. This map covered a total of 916.5 cM with all 11 linkage groups, or 889.8 cM with just the ten linkage groups that map correctly. QTLs for jorquette height and three FP resistance QTLs have been mapped using MapQTL version 4.0 (Van Ooijen, 2000; Fig. 2).

Creation of the First Combined Map over Populations

These first two cacao maps were combined using all linkage groups except the 11th linkage group in the 'Pound7' x 'UF273' F_1 cross. This map (Fig. 2) contained 10 linkage groups, named as in the convention established by Lanaud et al. (1995). Chromosome lengths varied from 51.36 cM to 114.19 cM, with a total genomic length of 885.4 cM. Average distance between markers was 3.43 cM, from a total of 254 markers. This map compares well with the high-density map of Pugh et al. (2004) both in length of individual linkage groups (68.4 cM to 111.3 cM) and in total genomic length (885.4 cM), based on 424 markers with an average distance of 2.1 cM between markers.

Mapping QTLs for Disease Resistance in Cacao Using a Haplotypic Approach

Developing large cacao mapping populations has been a restrictive task given the morphological and biological characteristics of the species. The integration of QTL analysis into breeding strategies, rather than being considered as a separate process, has been proposed to increase the power and accuracy of QTL detection. We developed a method that groups F_1 breeding populations by common founders and statistically associates founder-origin probabilities that trace the common founder genome (haplotypes) in a given region of the progeny genome with the phenotypic expression, using a linear mixed model approach (Cervantes-Martinez and Brown, 2004). Crosses among selected clones were made in the partial full-sib mating, and subsequently evaluated in a field experiment over a five-year period beginning two years after planting at CATIE the cacao breeding program for yield and disease resistance. We studied the combining abilities of the clones and the predominant gene action for disease resistance, yield and agronomic traits in this set of crosses. The results demonstrated that healthy pods and percent pods with frosty pod have additive gene action, while total number of pods and trunk diameter have both additive and non-additive gene action. Also the crosses of the two parental founders 'UF712' and 'UF273 type I' were determined to be suitable for QTL mapping by population pooling methods (Cervantes-Martinez et al., 2006). We are currently conducting research to construct a combined molecular linkage map, by pooling the F_1 crosses that have 'UF712' as a common parent, and developing user friendly software macros in SAS (version 9.1; SAS Institute Inc.) to estimate QTLs for resistance to frosty pod, yield, and agronomic traits.

Association Mapping in Cacao

The development and evaluation of mapping populations for QTL identification is very time consuming and current QTL identification for productivity traits and disease resistance is based on a few F_1 families and a single F_2 population. An attractive alternative approach for perennial tree crops like cacao is association mapping. Association mapping was developed for medical genetics, but successfully applied to plants where a QTL for flowering time was associated with a DNA polymorphism in a candidate gene locus among inbred lines of maize (Thornsberry et al., 2001). Association analysis or linkage disequilibrium (LD) mapping in plants provides an alternative method for evaluating the basis of quantitative variation in perennial fruit crops where family designs are difficult and recombinant inbred lines impossible to produce. Utilizing 99 selected productive plants and 50 selected unproductive plants growing at the Dole Food Co. Hawaii fields on Oahu, we demonstrated that associations could be detected between microsatellite alleles and the productive vs. unproductive seedlings. In addition, microsatellite loci associated with productivity co-located near previously mapped QTLs for productivity traits (Schnell et al., 2005).

Seventeen of 65 loci were identified as having a significant linear (additive) trend with an α level of 0.05 or lower. Of these, 13 (76.4%) were located in areas of the chromosomes previously associated with QTLs for productivity. No significant association of microsatellite alleles with productive seedlings were detected for markers on linkage groups 5, 6, 7, or 8. Of the 17 loci with alleles associated with the productive seedlings, one was located on linkage group 1, five on linkage group 2, two on linkage

group 3, five on linkage group 4, three on linkage group 9, and one on linkage group 10. An example of loci with a significantly associated allele is provided in Figure 3 where Cir 60 and Cir 152 co-locate with QTLs for bean length and pod weight and Cir 100 and Cir 240 co-locate with a QTL for wet bean weight on linkage group 2. On linkage group 9 Cir 205 and Cir 212 co-located with a QTL for pod number and wet bean weight. The number of alleles for the 17 significant loci associated with productivity ranged from 4 to 13 and a single allele could be associated with the productive seedlings at four loci, while two alleles were associated for 11 loci, and three alleles at two loci (Schnell et al., 2005).

Association mapping is an alternative approach to the traditional family-based mapping (F_1, F_2, backcross). Although significant correlations of particular alleles for 17 loci and productivity variation were found, the inference is conditional on the precision achieved by discrete data collected for two years. Additional quantitative trait data that includes yield and yield components collected for several years will be required to corroborate these associations. If these associations are repeatable, the procedure will have utility in screening germplasm collections and influence MAS in cacao.

GENE DISCOVERY

Development of a Subtracted cDNA Library for Frosty Pod from Susceptible and Resistant Cacao Cultivars

Hand-pollinated pods, both inoculated and not inoculated with frosty pod, were produced at CATIE and freeze dried. mRNA was isolated from these pods and four subtracted libraries were produced. We are currently screening approximately 4000 colonies (~1000 per library). Preliminary sequencing of 100 clones per library has identified the forward subtracted library of the resistant cultivars as the most complex. Comparisons of the 100 sequences from each library identified 42 sequences unique to the forward subtracted library. This represents genes whose transcription is up-regulated. Primers have been produced for amplifying fragments between 150 and 250 nucleotides in genomic DNA for these unique ESTs. The next step is to determine if they are polymorphic in parents of mapping populations, which would allow mapping onto the cacao genome.

Analysis of QTL-Containing Regions from a BAC Library

A BAC library has been constructed from 'LCTEEN37'. The BAC library contains 36,864 clones (an 11-fold coverage of the *T. cacao* genome) with an average insert size of 120 kb, and is of good quality. Hybridization with chloroplast and mitochondrial specific probes has determined that only 2.5% and 1.5% of the clones in the BAC library are from chloroplast and mitochondrial DNA, respectively. Analysis of the Brazilian F_2 population segregating for resistance to WB has identified a QTL within a 20-25 cM region of linkage group 9 that contributes up to 20-25% of the variance. The goal is to utilize the BAC library to further define this region, and ultimately identify the gene(s) and the alleles involved in conferring increased disease resistance. DNA probes have been developed from an initial set of markers mapped within the QTL identified on linkage group 9. Hybridization of these probes has identified an appropriate number of clones (between 6-14 clones per probe) from the BAC library. The positive clones are currently being "fingerprinted" to determine whether they indeed represent overlapping fragments. DNA probes are currently being developed for additional markers mapped within the QTL, as well as end sequences of the BACs already identified. Clones will be identified and "fingerprinted" until a complete physical map of the QTL on linkage group 9 is obtained.

CONCLUSIONS

Over 23,000 genotypes are currently under evaluation. Many of these families have been verified using a subset of the microsatellite markers, alerting us to identity problems and mislabeled parents. Insuring the integrity of our breeding populations has

significantly reduced errors in our phenotypic data allowing us to discard plants resulting from self-pollination, pollen contamination, and mislabeling in the nursery. This approach has allowed obtaining better genetic parameter estimates in our analyses of segregating populations (Cervantes et al., 2006).

Currently over 100 QTLs have been identified and are listed in the CocoaGenDB (http://cocoagendb.cirad.fr). Many of these QTLs are not stable over genetic backgrounds and most have not been tested in multiple environments over multiple years. Based on the LOD scores, number of years tested, and number of individuals in the families, we believe that approximately 40 of these QTLs could be stable and therefore useful over genetic backgrounds and years, although this is speculative. We are in the process of confirming the utility of these QTLs in families for clone selection. One of the major objectives of the project over the next few years will be the development and confirmation of the stability of QTLs for productivity, quality, and disease resistance.

The major and minor QTLs for WB resistance are being used for selection. The major QTL on linkage group 9 has four loci with alleles significantly associated with resistance and the minor QTL on linkage group 1 has three loci with alleles significantly associated with resistance. In each case the 'SCA6' parent of 'TSH516' contributed the desirable allele to the plants with the resistant phenotypes (Table 4). 'TSH516' was self-pollinated and the mapping was accomplished in an F_2 population as previously described. To confirm the stability of these markers we are testing the association of the resistant haplotype with field resistance in families with a 'SCA6' or 'SCA12' parent or grandparent in their background in our disease field site at INIAP.

As we continue to identify and confirm the stability and utility of QTLs for disease resistance, productivity, and quality traits, they will be incorporated into the MAS program. The goal is to greatly reduce the number of seedlings planted in field evaluations using the QTL analysis. The outline of this Accelerated Cacao MAS program now under development, involves two major six year phases. Phase I starts with producing full-sib families from selected parents. Seedlings resulting from these hybridizations are screened with a number of QTL-associated markers and only those with the desirable haplotype retained for further field testing. The retained seedlings are cloned and four to 12 plants of each genotype are planted for preliminary phenotypic evaluation. Phenotypic evaluation begins in year 3 and is carried out for an additional three years. The best 20% of the seedlings would be advanced to the next phase. Phase II involves placing these selected clones into a larger replicated field trial for another six year period. Phenotypic data would be collected starting in year 2 and continuing through year 6 when the final selections are made and released as new varieties. A complete cycle of selection takes 12 years using this Accelerated Cacao MAS breeding program.

The current breeding project has been underway for five years and significant progress has been made in laying the foundation for a MAS program. The development of molecular markers, production of linkage maps, identification of QTLs for traits of interest, and the development of new mapping and selection methodologies have been necessary to implement a MAS program. If progress continues, we should be able to fully execute an Accelerated Cacao MAS within the next four years. The time and expense associated with the development of the tools and genetic knowledge for the MAS have been significant. However, based on preliminary cost estimates, a savings of approximately 50% per cycle of selection can be realized using the Accelerated Cacao MAS. This shorter time line and marker enhanced selection for field testing is far superior to traditional cacao selection techniques currently in use. The expected outcomes from the Accelerated Cacao MAS are new, precocious, high yielding, disease resistant, high quality cultivars that increase income for farmers while having the quality attributes required by the confectionary industry.

Literature Cited

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*. Theor. Appl. Genet. 109:495-507.

Borrone, J.W., Brown, J.S., Kuhn, D.N., Motamayor, J.C. and Schnell, R.J. 2006. Simple sequence repeat markers developed from *Theobroma cacao* expressed sequence tags. International Plant and Animal Genome Conference XIV, San Diego, CA.

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 an F_2 population using SSR markers and candidate genes. J. Amer. Soc. Hort. Sci. 130:366-373.

Cervantes-Martinez, C. and Brown, S.J. 2004. A haplotype-based method for QTL mapping of F1 populations in outbred plant species. Crop Sci. 44:1572-1583.

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

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

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

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

Dillinger, T.L., Barriga, P., Escarcega, S., Jimenez, M., Salazar-Lowe, D. and Grivetti, L. 2005. Food of the gods: cure for humanity? A cultural history of the medicinal and ritual use of chocolate. J. Nutrition 130:2057-2072.

Faleiro, F.G., Quieroz, V.T., Lopez, U.V., Guimarães, C.T., Pires, J.L., Yamada, M.M., Araújo, I.S., Pereira, M.G., Schnell, R., Souza Filho, G.A., Ferreira, C.F., Barros, E.G. and Moreira, M.A. 2006. Mapping QTLs for witches' broom (*Crinipellis perniciosa*) resistance in cacao (*Theobroma cacao* L.). Euphytica 147 (in press).

Kuhn, D.N., Heath, M., Wisser, R.J., Meerow, A., Brown, J.S., Lopes, U. and Schnell, R.J. 2003. Resistance gene homologues in *Theobroma cacao* as useful genetic markers. Theor. Appl. Genet. 107:191-202.

Kuhn, D., Borrone, J., Meerow, A., 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. and Schnell, R.J. 2005. Use of capillary array electrophoresis single strand conformational polymorphism analysis to estimate genetic diversity of candidate genes in germplasm collections. Meth. Enzymol. 395:238-258.

Kuhn, D.N., Narasimham, G., Nakamura, K., Brown, J.S., Schnell, R.J. and Meerow, A.W. 2006. Identification of cacao TIR-NBS-LRR resistance gene homologues and their use as genetic markers. J. Amer. Soc. Hort. Sci. (in press).

Lanaud, C., Risterucci, A.M., N'Goran, J.A.K., Clement, D., Flament, M.H., Laurent, V. and Flaque, M. 1995. A genetic linkage map of *Theobroma cacao* L. Theor. Appl. Genet. 91:987-993.

Lanaud, C., Risterucci, A.M., Piepertti, I., Falque, M., Bouet, A. and Lagoda, P.J.L. 1999. Isolation and characterization of microsatellites in *Theobroma cacao* L. Mol. Ecol. 8:2141-2152.

Lockwood, R. 2003. Who needs clothing? INGENIC Newsletter 8:2-5.

Marshall, T., Slate, C.J., Kruuk, L. and Pemberton, J.M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7:639-655.

Motamayor, J.C., Risterucci, A.M., Lopez, P.A., Ortiz, C.F., Moreno, A. and Lanaud, C. 2002. Cacao domestication I: the origin of the cacao cultivated by the Mayas.

Heredity 89:380-386.
Motamayor, J.C., Risterucci, A.M., Heath, M. and Lanaud, C. 2003 Cacao domestication II: progenitor germplasm of the Trinitario cacao cultivar. Heredity 91:322-330.
Motilal, L. and Butler, D. 2003. Verification in global cacao germplasm collections. Genet. Res. Crop Evol. 50:799-807.
Motilal, L.A. and Boccara, M. 2004. Screening and evaluation of SSR primers in gel systems for the detection of off-types in cocoa field. INGENIC Newsletter 9:21-24.
Pound, F.J. 1943. Cocoa and witches' broom disease (*Marasmius perniciosus*) of South America with notes on other species of *Theobroma*. Yuille's printery, Port of Spain, Trinidad and Tobago.
Pritchard, J.K., Stephens, M. and Gonnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.
Pugh, T., Fount, 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. Theor. Appl. Genet. 108:1151-1161.
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* L. Theor. Appl. Genet. 101:948-855.
Saunders, J.A., Mischke, S. Leamy, E.A. and Hameida, A.A. 2004 Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao*. Theor. Appl. Genet. 110:41-47.
Schnell, R.J., Heath, M.A., Johnson, E.S., Brown, J.S., Olano, C.T. and Motamayor, J.C. 2004. Frequency of off-type progeny among the original ICS1 x SCA6 reciprocal families made for selection for disease resistance in Trinidad. INGENIC newsletter 9:34-39.
Schnell, R.J., Olano, C.T., Brown, J.S., Meerow, A.M., 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 2 microsatellite alleles with productivity. J. Amer. Soc. Hort. Sci. 130:181-190.
Slate, J., Marshall, T. and Pemberton, J.M. 2000. A retrospective assessment of the accuracy of the paternity inference program Cervus. Mol. Ecol. 9:801-808.
Thornsberry, J., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D. and Buckler, E.S. 2001. Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics 28:286-289.
Turnbull, C.J., Butler, D.R., Cryer, N.C., Zhang, D., Lanaud, C., Daymond, A.J., Ford, C.S., Wilkinson, M.J. and Hadley, P. 2004 Tackling mislabeling in cocoa germplasm collections. INGENIC 9: 8-11.
Warren, J.M. and Kennedy, A.J. 1991. Cocoa breeding revisited. Cocoa Growers' Bulletin 44:18-24.
Van Ooijen, J.W. 2000. MapQTL ® version 4.0 Userfriendly Power in QTL Mapping; Addendum to the Manual of Version 3.0. Plant Research International, Wageningen, The Netherlands.
Van Oooijen, J.W. and Voorrips, R.E. 2001. JoinMap Version 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, Netherlands.

Tables

Table 1. Disease losses from the three major pathogens of *Theobroma cacao.*

Diseases	Pathogen	Region	Reduced production	
			(tons x 1000)[1]	($ million)[2]
Black Pod	*Phytophthora* spp.	Africa/Brazil/Asia	450	485
Witches' Broom	*Moniliophthora perniciosa*	Latin America	250	270
Frosty Pod Rot	*Moniliophthora roreri*	Latin America	30	32

[1]CABI Commodities (http://www.dropdata.net/cocoa/cocoa_prob.htm)
[2]January, 2001: value = $1,078.00/metric ton.
ICCO Annual Averages (http://www.icco.org/prices/pricesave.htm)

Table 2. Genotypes under evaluation in the USDA-ARS international breeding program.

Location	Population	Number of genotypes	
CATIE in Costa Rica	F1 mapping population 'Pound7' x 'UF273'	260	
	The 34-Hybrid Family Trial	2,720	
	Two segregating populations for QTL identification	800	
	Parents of the 34-Hybrid Families Trial	68	
	Selected clone trail	160	
	Sub-total:		4,008
INIAP in Ecuador	Witches' broom disease plot	7,320	
	Old populations under selection	4,800	
	Chalmers and Allen Collection	300	
	Sub-total:		12,420
CCI in PNG	VSD and BP mapping populations		1,600
Brazil	Almirante (59 families)	4,113	
	Germplasm collection	379	
	Sub-total:		4,492
USA	Miami (quarantine and nursery)	250	
	Mayaguez (germplasm and nursery)	300	
	Sub-total:		550
All locations	Total:		23,070

Table 3. Comparison of genetic diversity of cacao germplasm traditionally used in cacao breeding and unexploited germplasm. Hn.b., Unbiased gene diversity; Hobs, Observed heterozygosity.

Germplasm type	Hn.b.	Hobs	Mean number of alleles	Number of private alleles
Traditionally used germplasm (Criollo, Amelonado, Nacional, Pound Collection) (n = 454)	0.65 (0.17)	0.32 (0.12)	9.62	44
Unexploited germplasm (n = 495)	0.74 (0.14)	0.39 (0.11)	14.02	471

Table 4. Use of QTL information in Marker Aided Selection. 'TSH516' is an F_1 seedling of the cross 'ICS1' x 'SCA6'. For each of the seven loci linked to the WB QTLs the favorable allele (indicated in bold) is from the 'SCA6' parent.

QTL	Marker	'ICS1'	'SCA6'	'TSH516'	Resistance associated allele
Major QTL on linkage group 9 flanking markers	mTcCir 24	184, 184	184, **192**	184, **192**	**192**
	mTcCir 35	235, 235	**230, 230**	**230,** 235	**230**
	mTcCir 157	149, 149	149, **153**	149, **153**	**153**
	mTcCir 160	288, 288	288, **293**	288, **293**	**293**
Minor QTL on linkage group 1 flanking markers	RGH11	G626, G626	G626, **G649**	G626, **G649**	**G649**
	mTcCir 22	296, 286	278, **282**	**282,** 286	**282**
	mTcCir 264	196, 196	208, **210**	196, **210**	**210**

Figures

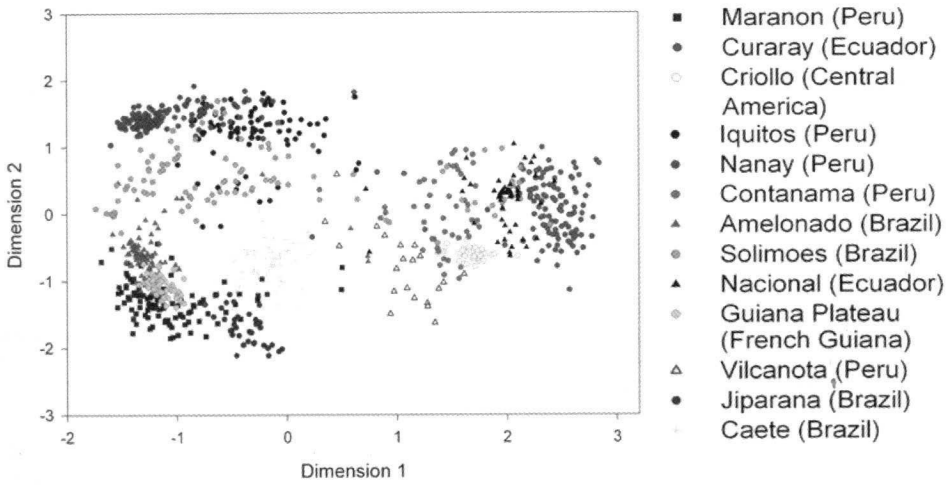

Fig. 1. Multidimensional Scaling Plot of 949 genotypes based on a 97 microsatellite Shared Allele Distance matrix using SAS (version 9.1; SAS Institute Inc., Cary, NC).

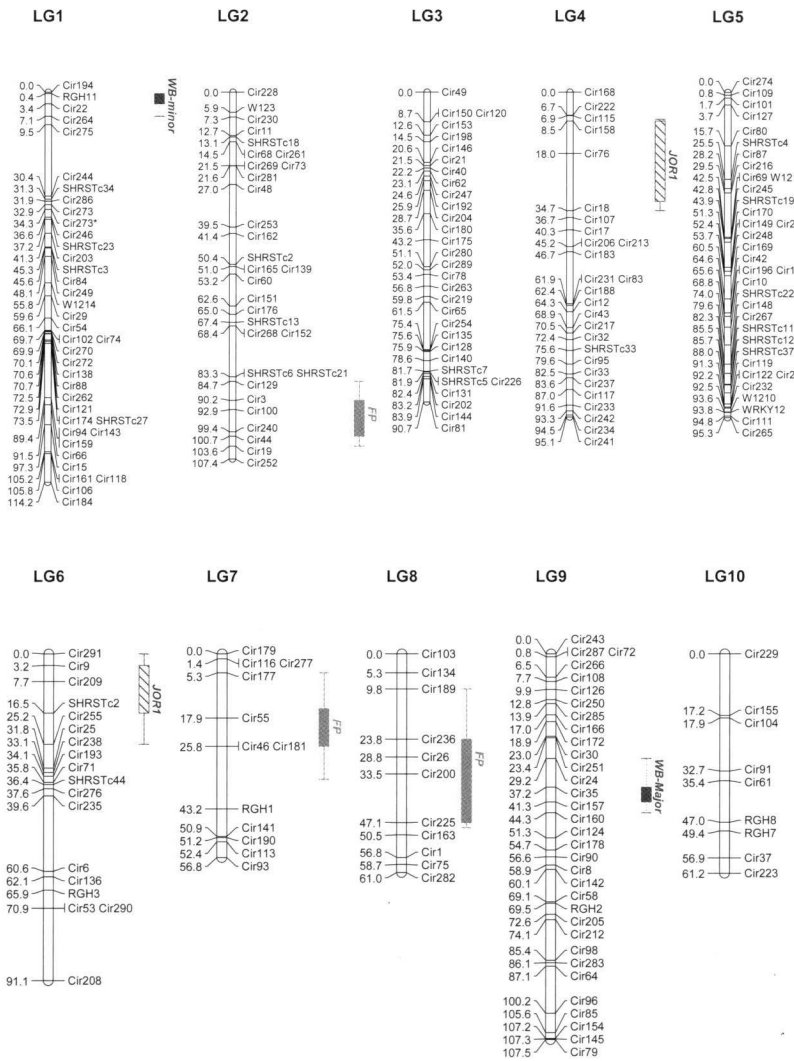

Fig. 2. The combined map from the F_2 population created by selfing 'TSH516' and from the F_1 population from crossing 'Pound7' x 'UF273'. Disease resistance and productivity QTL locations indicated by vertical bars on the left of the linkage groups are as follows: Frosty pod resistance (*FP*), Witches' broom resistance - major QTL (*WB-Major*), Witches' broom resistance - minor QTL (*WB-minor*), jorquette height (JOR1). Markers designated Cir were developed at CIRAD. Markers developed at the USDA-ARS are designated RGH and SHRSTc for resistance gene homologue and microsatellite markers, respectively.

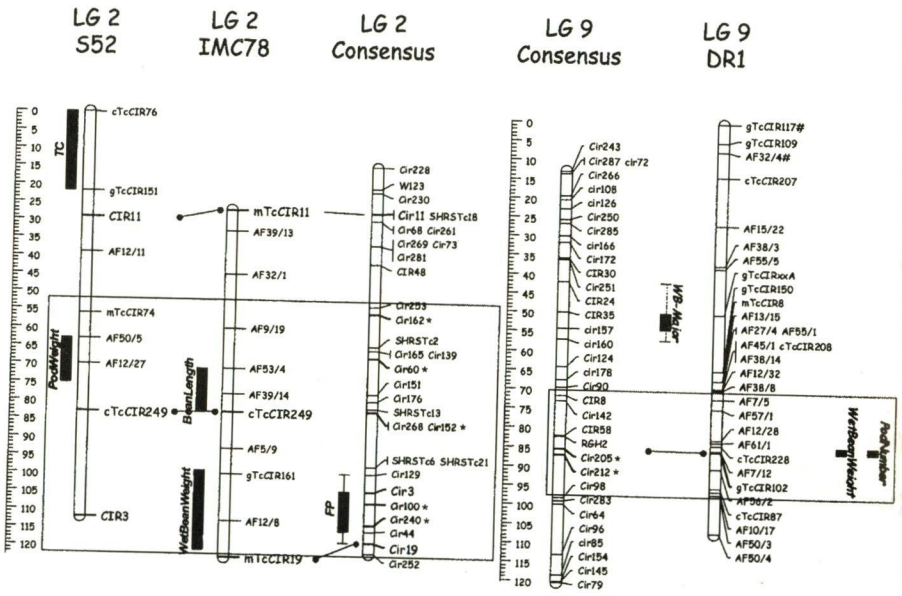

Fig. 3. Microsatellite markers and productivity QTL locations for linkage groups 2 and 9. Microsatellite markers found to be associated with productivity traits by Armitage's trend test are indicated with a *. Linkage map comparisons by linkage group (LG) indicate location of RFLP, AFLP, and microsatellite markers in progeny of IMC78 x Catongo (IMC78); S52 x Catongo (S52); and DR1 x Catongo (DR1) (Clement et al., 2003a,b). Disease resistance and productivity QTL locations indicated by vertical bars on the side of the linkage groups are as follows: Frosty pod resistance (*FP*), Witches' broom resistance - major QTL (*WB-Major*), bean length (*beanlength*), trunk circumference (*TC*), pod weight (*PodWeight*), wet bean weight (*WetBeanWeight*), and pod number (*PodNumber*).