

# Out of the Amazon: *Theobroma cacao* enters the genomic era

Alan B. Bennett

Mann Laboratory, Department of Vegetable Crops, University of California, One Shields Avenue, Davis, CA 95616, USA

***Theobroma cacao* has long been a crop of near-mystical human importance among indigenous Mesoamerican cultures and its importance in modern culture is growing with new realizations of the potential health benefits of cocoa polyphenolic compounds. However, cultivated *T. cacao* is vulnerable to emerging disease pressures because it has a narrow genetic base, and systematic genetic improvement of the crop is imperative. A wide range of genomic tools and resources has been developed and are providing the basis for genome-based breeding and gene discovery.**

*Theobroma cacao* has a rich history of human use, beginning with the Olmec and Mayan peoples of the eastern Mexican gulf who believed that cacao was delivered to them directly by God [1]. From its divine origins in Mesoamerica, cacao was introduced to the Spanish royal court in the mid-1550s and products derived from cacao continue to hold a special appeal in modern culture. There is also extensive documentation of medicinal uses of cacao that can be traced to ancient Aztec documents. A surviving document, the Florentine Codex, was compiled in 1590 by a Spanish priest, which extensively documents cacao-based preparations and the illnesses that they were used to prevent or cure [1]. Over the following centuries, the medicinal uses of cacao have diminished but recent research has demonstrated an unanticipated role of cacao in promoting cardiovascular health [2]. The basis of the health-promoting effects of cacao is that certain polyphenols (catechins and flavanols) are potent antioxidants that can attenuate inflammatory processes in atherosclerosis and block expression of cellular adhesion molecules. Some cacao products contain high enough levels of biologically active polyphenols to exert both acute and chronic antioxidant-associated health benefits [2].

In spite of its rich history and continuing significant role in human consumption and potentially in human health, the *T. cacao* tree has received relatively little attention as a crop species. Bertus Eskes [3] estimated that only ~30% of *T. cacao* cultivated today are of selected varieties. The other 70% of trees are traditional populations, and these are grown increasingly from seeds collected by farmers from preferred trees and likely to be inbred. A recent analysis of the 'best' *T. cacao* clones identified by two experienced breeders indicated that only 20% originated from any form of breeding program and that less than 1% of the clones originated in the past 20 years [4]. This

situation is surprising because *T. cacao* germplasm was extensively collected in the 1930s. These collections have been continuously maintained as an important genetic resource but have only received limited systematic use for breeding. A consequence of the narrow genetic base of cultivated cacao resulted in the devastation of cacao production in the state of Bahia, Brazil during the 1990s. In 1989, *Crinipellis pernicioso*, a fungus that causes witches' broom disease (Figure 1), was first identified in Bahia and within a decade cacao production fell from 383 000 tons to an estimated 125 000 tons today. This genetic vulnerability to a single disease introduction has raised the specter that similar crop failures could occur in other major cacao-producing regions of the world with devastating effects on the global supply of the crop. In Bahia, the infrastructure has been developed to produce cloned *T. cacao* plants in sufficient quantities to replant the



**Figure 1.** (a) Large-scale commercial production of rooted cuttings is being undertaken to replant large areas of *Theobroma cacao* lost to a single disease, witches' broom (*Crinipellis pernicioso*) in the state of Bahia, Brazil. (b) Basidiocarps of *Crinipellis pernicioso*. (Photograph courtesy of Robert Lumsden, USDA.)

entire Brazilian crop (Figure 1). However, the lack of systematic evaluation of the genetic structure and variation of the new *T. cacao* clones raises questions as to whether these efforts can outpace the witches' broom disease or have the genetic resilience in the face of other potential new diseases. There is now broad recognition that systematic genetic improvement of *T. cacao* is imperative and coincides with the development of a suite of genomic tools for the species that are accelerating this process.

### Genomic tools for germplasm evaluation and breeding

The development of a high-density molecular map of the *T. cacao* genome has had rapid and pronounced implications for clarifying the natural history of cacao, in guiding genomic-based breeding strategies and in germplasm conservation. Molecular mapping of the cacao genome was initiated in 1995 [5] and recently resulted in the development of a high-density linkage map comprising 424 markers in ten linkage groups with an average spacing of 2.1 cM [6]. The availability of a molecular map and of characterized molecular markers has contributed to rapid advances in several areas. For example, the origin of *T. cacao*, a controversial topic for several decades, has been recently clarified. The earliest studies suggested that *T. cacao* originated in the Upper Amazon near the eastern edges of the Andes, and that Central American populations had been introduced from this center of origin [7]. Subsequently, many authors proposed that the Central American 'Criollo' types and South American 'Forestero' types of *T. cacao* had separate origins and evolved independently [8,9]. Recent analysis using more-detailed molecular markers has provided strong support for a single origin in the Upper Amazon, with the Central American 'Criollo' group probably originating from a few individuals transported by humans from South America and spread throughout Central America [10].

The primary efforts and advances based on the *T. cacao* molecular map have been in the identification of quantitative trait loci (QTLs) for major traits including resistance to pod rot, *Phytophthora palmivora*, and resistance to witches' broom, *C. perniciosus* [11–13]. In one study, pod rot resistance assayed by pod inoculations identified a single QTL in two separate populations, explaining ~47% of the phenotypic variance in resistance [11]. In a second study, pod rot resistance was assayed using field tests, as well as artificial inoculations of leaves and pods. Five QTLs for resistance were identified but none was common between the three assayed traits, suggesting that there could be multiple mechanisms contributing to a complex resistance phenotype [12].

QTLs contributing to witches' broom resistance were evaluated in an F<sub>2</sub> population derived from a cross that included the *T. cacao* cultivar 'Scavina-6', which is well known to exhibit resistance to witches' broom. 'Scavina-6' germplasm is apparently randomly represented in hybrids grown throughout Bahia, Brazil and has been long-suspected to be the major contributor to farmer-identified trees resistant to the disease. In this study, a single QTL was identified explaining ~35% of the phenotypic variance for witches' broom resistance [13]. The identification

of this molecular marker is a crucial step in determining the molecular basis of witches' broom resistance in farmer selections in Bahia, Brazil and to determine if this is the single source of resistance or if other sources exist that can be combined for increased durability of the trait.

Perhaps most importantly, the molecular map of the *T. cacao* genome is providing rational direction for systematic breeding programs that are seeking to broaden the genetic base of cultivated trees as well as to characterize the existing *T. cacao* germplasm collections. For example, it has been reported that of the genetically diverse 'Forastero' types collected in the 1930s, a sample of pods from nine trees were sent to Ghana and their progeny have formed the basis of production and breeding populations in West Africa and Malaysia [14]. Molecular characterization of diverse germplasm is providing the basis for rationally increasing the genetic base for cacao breeding. The results to date also suggest that the previously collected *T. cacao* germplasm was narrow and that there is likely to be substantial unsampled genetic diversity in wild populations that are still to be collected.

To assess the scope and quality of preserved *T. cacao* germplasm, a program is underway to genetically fingerprint the international *T. cacao* germplasm collection [15]. Because *T. cacao* seeds are short-lived, germplasm collections are maintained as living trees and it is estimated that 20–30% of the collective global collection are mislabeled and the degree of redundancy between collections is unknown. A panel of microsatellite single sequence repeats has been identified that exhibit high reproducibility and a high degree of polymorphism, which allow the differentiation of divergent cultivars. These microsatellite markers have been proposed as the international molecular standards for DNA fingerprinting of *T. cacao* and the development of a unified database of *T. cacao* germplasm is now underway [15].

### Genomic tools for gene discovery

Although relatively less attention has been directed towards gene discovery and expression analysis in *T. cacao*, significant advances have been made in this arena as well [16]. A major study sequenced >5500 expressed sequence tags (ESTs) and identified a unigene set of 1380 unique gene sequences derived from bean and leaf cDNA libraries, setting the stage to identify bean and leaf differentially expressed sequences. Microarrays were constructed using sequences from the *T. cacao* unigene set and used in demonstration experiments to evaluate genotype and tissue specificity of gene expression and validated by real-time PCR. The results demonstrated that the EST and microarray genome resource performed as anticipated and is likely to be useful in combination with the developing germplasm resources to evaluate the underlying molecular basis of disease responses as well as bean quality traits that contribute to bean yield, composition and flavor.

### Conclusion

The cocoa crop is at a scientific crossroads. *T. cacao* is one of a few important crops that has received little systematic genetic improvement. Its economic importance and genetic vulnerability, as illustrated in Bahia, Brazil over the past

decade, has mobilized a partnership of private and public research groups to focus on the development of new technologies to support *T. cacao* genetic improvement ([http://www.ars-grin.gov/ars/SoAtlantic/Miami/ngr/cacao\\_genetics\\_meeting\\_summary.pdf](http://www.ars-grin.gov/ars/SoAtlantic/Miami/ngr/cacao_genetics_meeting_summary.pdf)) with a degree of coordination by the International Group for Genetic Improvement of Cocoa (INGENIC), which promotes the exchange of information and international collaboration on topics related to cocoa genetics. This initiative comes at a time of unprecedented new genomic tools that dramatically increase the genetic knowledge to inform breeding strategies and to accelerate notoriously slow tree breeding programs. Over the next decade, *T. cacao* is likely to become a model of an essentially wild crop being transformed through genome-based breeding. It will be interesting to see whether these new tools can provide the basis simultaneously to address the improvement of cocoa productivity while maintaining the quality characteristics of the bean that are crucial for chocolate production.

#### References

- Dillinger, T.L. *et al.* (2000) Foods of the gods: cure for humanity? A cultural history of the medicinal and ritual uses of chocolate. *J. Nutr.* 130, 2057S–2072S
- Kris-Etherton, P.M. and Keen, C.L. (2002) Evidence that the antioxidant flavonoids in tea and cocoa are beneficial for cardiovascular health. *Curr. Opin. Lipidology* 13, 41–49
- Eskes, B. (2001) Introductory notes. *Proceedings of the International Workshop on New Technologies for Cocoa Breeding, Kota Kinabalu, Malaysia*, INGENIC, London, UK, pp. 8–11
- Lockwood, R. (2003) Who needs clothing? *INGENIC Newsl.* 8, 2–5
- Lanaud, C. *et al.* (1995) A genetic linkage map of *Theobroma cacao* L. *Theor. Appl. Genet.* 91, 987–993
- Risterucci, A.M. *et al.* (2000) A high-density linkage map of *Theobroma cacao* L. *Theor. Appl. Genet.* 101, 948–955
- Cheesman, E.E. (1944) Notes on the nomenclature, classification and possible relationships of cocoa populations. *Trop. Agri.* 21, 144–159
- Cuatrecasas, J. (1964) Cacao and its allies: a taxonomic revision of the genus *Theobroma*. *Contrib. US Herb.* 35, 379–614
- De la Cruz, M. *et al.* (1995) Origins of cacao cultivation. *Nature* 375, 542–543
- Motamayor, J.C. *et al.* (2002) Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* 89, 380–386
- Crouzillat, D. *et al.* (2000) Quantitative trait loci analysis in *Theobroma cacao* using molecular markers: inheritance of polygenic resistance to *Phytophthora palmivora* in two related cacao populations. *Euphytica* 114, 25–36
- Flament, M-H. *et al.* (2001) Genetic mapping of resistance factors to *Phytophthora palmivora* in cocoa. *Genome* 44, 79–85
- Queiroz, V.T. *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
- Lanaud, C. *et al.* (2001) Implications of new insight into the genetic structure of *Theobroma cacao* L. for breeding strategies. *Proceedings of the International Workshop on New Technologies for Cocoa Breeding, Kota Kinabalu, Malaysia*, INGENIC, London, UK, pp. 89–107
- Saunders, J.A. *et al.* (2001) USDA DNA fingerprinting program for identification of *Theobroma cacao* accessions. *Proceedings of the International Workshop on New Technologies for Cocoa Breeding, Kota Kinabalu, Malaysia*, INGENIC, London, UK, pp. 108–114
- Jones, P.G. *et al.* (2002) Gene discovery and microarray analysis of cacao (*Theobroma cacao* L.) varieties. *Planta* 216, 255–264

1360-1385/\$ - see front matter © 2003 Elsevier Ltd. All rights reserved.  
doi:10.1016/j.tplants.2003.10.004

## Why are there so many carbohydrate-active enzyme-related genes in plants?

Pedro M. Coutinho, Mark Stam, Eric Blanc and Bernard Henrissat

Architecture et Fonction des Macromolécules Biologiques, UMR 6098, Centre National de la Recherche Scientifique, Universités Aix-Marseille I and II, 31 Chemin Joseph Aiguier, 13402 Marseille cedex 20, France

**Plants contain far more carbohydrate-active enzyme-encoding genes than any other organism sequenced to date. The extremely large number of glycosidase and glycosyltransferase-related genes in plant genomes can be explained by the complex structure of the plant cell wall, by ancient genome duplication and by recent local duplications, but also by the recent emergence of novel and unrelated protein functions based on widely available pre-existing scaffolds.**

In plants, carbohydrates in the form of glycosides are central to many biological pathways, from cell wall structure to energy, signalling and defence. Glycosides are made from activated sugars by glycosyltransferases

and are degraded by glycoside hydrolases (glycosidases). Genes encoding glycosidases and glycosyltransferases in all organisms (all data available from the Carbohydrate-Active enZymes server at <http://afmb.cnrs-mrs.fr/CAZY/>) are currently analysed and listed based on the classifications of glycosidases [1] and glycosyltransferases [2] in sequence- and structure-based families. Now that the genomes of many organisms have been completely sequenced, it is possible to analyse the content of genomes from a global glycobiochemical perspective.

#### Content of genomes in carbohydrate-active enzyme-related genes

One of the intriguing features of the analysis of the sequenced genomes is that for bacteria and for eukaryotes there appears to be a global correlation between the number of glycosidase and glycosyltransferase-related genes and the

Corresponding author: Bernard Henrissat (bernard.henrissat@afmb.cnrs-mrs.fr).