BIOTECHNOLOGICAL STRATEGIES FOR IMPROVEMENT OF COCOA

S. Elain Apshara[1] and M. K. Rajesh[2]

[1]Central Plantation Crops Research Institute, Regional Station, Vittal, Karnataka

[2]Central Plantation Crops Research Institute, Kasaragod, Kerala

Introduction

Cocoa, the source of chocolates is an internationally acclaimed crop because of its value in commercial and export market. According to the World Cocoa Foundation 40 to 50 million people depend upon cocoa worldwide for their livelihood and about three million tons of cocoa is produced annually, which corresponds to a global market value of $5.1 billion. The sale of Indian chocolate industry has increased from $175 million in 2000 to $450 million in 2005, which is mainly because of the concern towards nutritional and therapeutic qualities of cocoa. Presently in the cocoa producing regions it is estimated that only 30% of the varieties grown are the result of selection programmes and the remaining are the chance introductions of plants with low disease resistance and yield potential. They were originally derived from the Amazon basin and propagated by farmers through seeds passed down through the generations (Bennett, 2003). Though large genetic variation has been identified in wild populations throughout the Amazon, this diversity has not been widely utilized in terms of practical incorporation into cultivated varieties (Bartley, 2005). This illustrates the primitive state of domestication of cocoa, but it also highlights the opportunities for future enhancement through the application of modern molecular breeding approaches.

Cocoa Molecular Genetics Research Community

It is essential that researchers in developed and developing countries establish and maintain strong working relationships, collaborative research and training programmes to maximize the potential impact of our research on the cocoa farmers and the environment. So, International Group for Genetic Improvement of Cocoa (INGENIC) was formed in 1994 by Pennsylvania State University (PSU),USA, which includes over 300 members, representing 35 developing and developed countries. The INGENIC Study Group for Molecular Biology (INGENIC-MOL-BIOL) was also chartered in 2003 to coordinate the activities of the members interested in molecular approaches (Guiltinan, 2007). An international research symposium is held by INGENIC in a developing country once in three years.

Molecular Techniques and Strategies for Cocoa Improvement

1. Molecular Markers

Germplasm Evaluation *via* DNA Fingerprinting

One of the central applications of molecular biology is the use of molecular markers in studying relationships between accessions using phylogenetic analysis. For cocoa, this has been approached with isozymes, Random Amplified Polymorphic DNA (RAPD), Restriction Fragment Length Polymorphisms (RFLP) and other types of genomic DNA markers. Most recently, microsatellite markers (SSR) have gained acceptance as the most accurate and reliable method (Crouzillat *et al.,* 2000; Gomes *et al.,* 2000; Dias 2001; Motamayor *et al.,* 2002; Kuhn *et al.,* 2003; Pugh *et al.,* 2004). These investigations elucidate the origin of different cocoa lineages spanning back to ancient Mayas and has important implications to breeders who wish to incorporate diverse sources of resistance and other traits into breeding programmes. In this way, using molecular phylogenetic analysis of chloroplast DNA sequences, *Theobroma cacao* and its related species have been reclassified from the family Sterculiaceae to the Malvaceae (Bayer *et al.,* 1999; Whitlock *et al.,* 2001). This is important to cocoa researchers who wish to extrapolate information from other related species, such as cotton, which is in the same family as cocoa and *Arabidopsis thaliana,* a model plant for molecular-genetic research worldwide that is also closely related to cocoa.

Cryer *et al.* (2005) described two methods of standardization of microsatellite allele profiles between different laboratories, which was performed on a total of 429 cocoa accessions. They fingerprinted these accessions with 15 microsatellite markers that have been agreed upon as common set for worldwide standardization. Importantly, the data was deposited in the International Cocoa Germplasm Database, which is an easily accessible and searchable database for cocoa molecular genetic data (http://www.icgd.reading.ac.uk/netscape_index.htm).

Molecular Mapping

Quantitative Trait Loci (QTL) mapping is a technique central to developing markers to assist breeders in molecular based selection schemes. QTLs for various traits in cocoa have been identified, including resistance to fungal diseases, and various yield and morphological traits of interest such as fruit size and seed size (Motilal *et al.,* 2000; Clement *et al.,* 2003a, 2003b). One landmark study was the creation of a reference molecular-genetic map of cocoa, which was developed by the Lanaud group in France (Risterucci *et al.,* 2000; Pugh *et al.,* 2004). This map currently consists of >250 SSRs and >400 RFLP, RAPD, AFLP and isozyme markers, covering ~900 cM of the 10 cocoa chromosomes with an average marker distance of ~2 cM. They also established large number of microsatellite markers for mapping in cocoa.

2. Gene Discovery

BAC Library Resources

Two Cocoa Bacterial Artificial Chromosome libraries (BAC libraries) have been constructed recently. A BAC library is useful for gene discovery as it contains large fragments of genomic DNA each cloned into a bacterial strain that is arrayed in microtiter plates, making it simple to screen and to isolate specific genomic regions. Clement *et al.* (2004) reported on the construction of a BAC library from the genotype Scavina-6, is one of the most well-known and utilized genotypes of cocoa.

EST Resources for Cocoa

Expressed Sequence Tags (ESTs) are short stretches of DNA sequences determined from collections of cDNAs synthesized from RNA, and thus represent a snapshot of the genes expressed in a plant. Jones *et al.* (2002) published the first study using an EST approach to generate large numbers of expressed sequence tags for cocoa. In this study, leaf and seed cDNA libraries were sequenced and a unigene set of 1380 sequences were assembled. Additionally, these sequences were used to create a microarray, which was used to demonstrate the specificity of tissue specific expression of a number of genes. INGENIC-MOL-BIOL established a large EST database for cocoa and this project was led by Claire Lanaud of CIRAD and implemented by the French CNS (Centre National de Sequençage).

At CPCRI we have downloaded the cocoa EST sequences available in the public domain and made into contigs. Microsatellites were located in these ESTs and contigs using five softwares (MISA, TRA, TROLL, SSRIT and SSR primer). MISA gave maximum coverage of SSRs in cocoa ESTs and contigs, although TRA was able to detect higher order (>5-mer) repeats. The frequency of SSRs was one per 26.9 kb in the known set of ESTs. One-third of the repeats in EST-contigs were found to be trimeric. A few rare repeats like 21-mer repeat were also located. A/T repeats were most abundant among the mononucleotide repeats and the AG/GA/TC/CT type was the most frequent among dimerics. Flanking primers were designed using Primer3 program and verified experimentally for PCR amplification. The results of the study are made freely available online database (http://www.riju.mybioscience.com/cocoa/). Seven primer pairs amplified genomic DNA isolated from leaves were used to screen a representative set of 12 accessions of cocoa (Riju *et al.*, 2009).

Genomic DNA extraction with young cocoa leaves by DNeasy mini kit was done. The agarose gel electrophoresis showed high quality genomic DNA in cocoa, which were intact, high in molecular weight, without protein and RNA contamination. Genomic DNA was quantified and among the EST-SSR markers amplified, seven markers (*viz.* TH 1, TH 2, TH 4, TH 8, TH 9, TH 10 and TH 11) produced

amplicons of expected size. When screened with 12 cocoa accessions for level of polymorphism, 27 polymorphic alleles were produced which ranged from two to six alleles per locus with an average of 3.85 alleles per locus. The average polymorphism information content (PIC) value was 0.57. The similarity index, based on Dice coefficient, obtained after pairwise comparison among 12 cocoa accessions showed the highest index of 0.80 in the accessions Jerangau Red Axil (JRA) and VTLC-1 and the lowest (0.111) was observed between VTLCP-22 and VTLCC-1. The dendrogram generated with cluster analysis separated the 12 cocoa accessions into two major clusters at 35% similarity level. The first major cluster had five sub-clusters and included nine accessions. Accessions JRA and VTLC-1 exhibited 80% similarity. The second major cluster had two sub-clusters. VTLCC-1 formed a distinct accession. 2-D and 3-D principal coordinate analysis also showed VTLCC-1 as a distinct accession. The accessions, in general, were scattered across the co-ordinates.

Resistance Gene Analogs

Cocoa is prone to pathogens which causes an estimated loss of 8.1 lakh tons annually (30% of world production). Using a degenerate primer-PCR based approach, Lanaud *et al.* (2004) isolated a set of defense gene analogs from cocoa including several kinases and several pathogenesis-related genes of the PR class 2 and 5 families.

Floral Development Genes

Swanson (2005) compared the developmental biology of cocoa flower with *Arabidopsis* from both the morphological and molecular levels. He isolated a series of cocoa floral specific genes using degenerate RT-PCR and used these to examine gene expression during early development using *in situ* hybridization.

3. Genetic Transformation

Using *Agrobacterium tumefaciens* based transformation of cultured somatic embryos, a transformation system for cocoa capable of producing whole plants was established recently in Guiltinan lab, PSU (Maximova *et al.,* 2003).

4. Databases and Bioinformatic tools

Databases and bioinformatic tools exist to store and manage data related to *T. cacao*

- **TropgeneDB** (http://tropgenedb.cirad.fr) is organized on a crop basis with presently nine modules (banana, cocoa, coconut, coffee, cotton, oil palm, rice, rubber tree and sugarcane). The most common data stored in TropgeneDB are genetic and physical maps, marker information, Quantitative Trait Loci (QTL's), sequence data, and molecular data on genetic resources.
- **ESTtik:** The Expressed Sequence Tag Treatment and Investigation Kit tool (EST*tik)* was initiated to analyze and store results from processing of cDNA. The EST*tik* pipeline programs are a set of Perl packages, which successively performs base calling, vector trimming, assembly and functional annotation. EST*tik* will be used to store and annotate the EST sequences obtained from deeper Criollo transcriptome sequencing.
- **CocoaGen DB,** a web portal on cocoa, that comprises molecular genetic, genomic and phenotypic data, was initiated through a collaborative project involving CIRAD, University of Reading (School of Plants Sciences, UK) and USDA/ARS (United States Department of Agriculture, USA). This combines molecular genetic information from TropgeneDB (http://tropgenedb.cirad.fr) and phenotypic data from ICGD (International Cocoa Germplasm Database). It comprised around 1500 clones with their genotypes at various markers, six genetic maps, 950 markers, 98 QTL's and 250 sequences data. CocoaGenDB is available through internet at the URL http://cocoagendb.cirad.fr.

5. Cocoa Genome Sequencing

Translational Genomics

Theobroma cacao L. is a simple diploid with ten chromosomes (2n=2x=20) and with a small genome of

size 390 Mb to 415 Mb. Cocoa is a member of the order Malvales that includes the important crop plant cotton. Both are members of the Eurosids II group of plants that contains the Brassicales, including *Arabidopsis* (Soltis *et al.,* 2002). The close evolutionary relatedness of these three species suggests that cotton and cocoa are excellent crop plants for translational research. Due to its small genome size, its ability to reproduce both vegetatively and sexually and the richness of its fruits in secondary metabolisms, cocoa represents a good model to study perennial woody fruit crops.

Marker Assisted Selection Based Breeding

In the near future, local accessions, well adapted to regional environmental and soil conditions, will be crossed with internationally tested genotypes with disease resistance, high yield and other quality traits. Molecular markers will be used to screen segregating progeny for desired traits while retaining locally desired adaptive phenotypes. Gene pyramiding will be used to enhance resistance durability. Selected progenies will be vegetatively propagated through a combination of tissue culture, grafting and budding, for distribution to farmers. The genetic diversity of germplasm can be safeguarded in large cryopreservation storehouses.

Cocoa genome was sequenced in 2010 with the following goals

- Sequencing the cocoa genome (Criollo) and perform the sequence assembly integrating all resources.
- Performing a complete and detailed annotation of the cocoa genome sequence to trigger gene discovery and facilitate map based cloning strategies.
- Establishing a performant database to manage and exploit the annotated sequences data, with links with other molecular resources as linkage maps, BAC ends and EST's.
- Integrating genetic and genomic sequences data to identify key genes involved in traits variations.
- Providing a general knowledge on the cocoa genome organization and gene content. Benefit by comparative genomics approaches with the model plants to identify and understand more easily the function of key cocoa genes.
- Providing new genomic performant tools to facilitate and stimulate germplasm characterization and allelic diversity of key genes.
- Increasing breeding efficiency by marker assisted selection based on the genomics tools.
- Participating in the global understanding of plant evolution and complex biological processes.

REFERENCES

1. Bartley, B.G.D. 2005. The genetic diversity of Cocoa and its utilization. CAB International, Wallingford, UK. 341 p.
2. Bayer, C., Fay, M.F., De Bruijn, A.Y., Savolainen, V., Morton, C.M., Kubitzki, K., Alverson, W.S. and Chase, M.W. 1999. Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: A combined analysis of plastid atpB and rbcL DNA sequences. *Bot. J. Linn.* Soc. 129: 267 - 303.
3. Bennett, A.B. 2003. Out of the Amazon: *Theobroma Cacao* enters the genomic era. *Trends Plant Sci.* 8: 561-563.
4. Clement, D., Lanaud, C., Sabau, X., Fouet, O., Le Cunff, L., Ruiz, E., Risterucci, A.M., Glaszmann, J.C. and Piffanelli, P. 2004. Creation of BAC genomic resources for cocoa for physical mapping of RGA containing BAC clones. *Theor. Appl. Genet.*108: 1627-1634.
5. Clement, D., Risterucci, A.M., Motamayor, J.C., N'Goran, J. and Lanaud, C. 2003a. Mapping QTL for yield components, vigour, and resistance to *Phytophthora palmivora* in Cocoa. *Genome* 46: 204-212.
6. Clement, D., Risterucci, A.M., Motamayor, J.C., Ngoran, J. and Lanaud, C. 2003b. Mapping quantitative trait loci for bean traits and ovule number in cocoa. *Genome* 46: 103-111.

7. Crouzillat, D., Phillips, W., Fritz, P.J. and Petiard, V. 2000. Quantitative trait loci analysis in Cocoa using molecular markers. Inheritance of polygenic resistance to *Phytophtora palmivora* in two related Cocoa populations. *Euphytica* 114: 25-36.

8. Cryer, N.C., Fenn, M.G.E., Turnbull, C.J. and Wilkinson, M.J. 2005. Allelic size standards and reference genotypes to unify international cocoa microsatellite data. *Genetic Resour Crop Evol*. 53: 1643-1652.

9. Dias, L.A.S. 2001. Genetic Improvement of Cocoa. (http://ecoport.org ep ? Search Type=earticle View & earticleId=197)

10. Gomes, L.M.C., Melo, G.R.P., Faleiro, F.G., Silva, S.D.M., Araujo, R.C., Bahia, R.C., Morales, M.G. and Ahnert, D. 2000. Genetic diversity of *Crinipellis perniciosa* from the south region of Bahia-Brazil using RAPD molecular markers. *Fitopatologia Brasileira* 25: 377.

11. Guiltinan, M. 2007. Cocoa. *In:* Pua EC, Davey MR (eds) Biotechnology in Agriculture and Forestry - Transgenic Crops VI. Springer-Verlag, Berlin Heidelbelg.

12. Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D., Retzel, E.R. and Jones, C.A. 2002. Gene discovery and microarray analysis of Cocoa varieties. *Planta* 216: 255-264.

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

14. Lanaud, C., Risterucci, A.M., Pieretti, I., N'goran, J.A.K. and Fargeas, D. 2004. Characterisation and genetic mapping of resistance and defence gene analogs in cocoa. *Mol. Breed.* 13: 211-227.

15. Maximova, S., Miller, C., Antunez de Mayolo, G., Pishak, S., Young, A. and Guiltinan, M.J. 2003. Stable transformation of *Theobroma cacao* L. and influence of matrix attachment regions on GFP expression. *Plant Cell Rep* 21: 872-883.

16. Motamayor, J.C., Risterucci, A.M., Lopez, P.A., Ortiz, C.F., Moreno, A. and Lanaud, C. 2002. Cocoa domestication I: the origin of the Cocoa cultivated by the Mayas. *Heredity* 89: 380-386.

17. Motilal, L.A., Sounigo, O., Thevenin, J.M., Howell, M.H., Pieretti, I., Risterucci, A.M., Noyer, J.L. and Lanaud, C. 2000. *Theobroma cacao* L.: Genome map and QTLs for *Phytophthora palmivora* resistance. *In: 13th Annual Cocoa Research Conference,* Kota Kinabalu, Malaysia.

18. Pugh, T., Fouet, O., Risterucci, A.M., Brottier, P., Abouladze, M., Deletrez, C., Courtois, B., Clement, D., Larmande, P. and Ngoran, J.A.K. 2004. A new Cocoa linkage map based on codominant markers: development and integration of 201 new microsatellite markers. *Theor. Appl. Genet.* 108: 1151- 1161.

19. Riju, A., Rajesh, M.K., Fasila Sherin, P.T.P., Chandrasekar, A., Elain Apshara, S. and Arunachalam, V. 2009. Mining of expressed sequence tag libraries of cacao for microsatellite markers using five computational tools. *Journal of Genetics* 88 (2): 217-225.

20. Risterucci, A.M., Grivet, L., Ngoran, J.A., Pieretti, I., Flament, M.H. and Lanaud C. 2000. A high-density linkage map of Cocoa. *Theor. Appl. Genet.* 101: 948-955.

21. Soltis, D.E., Soltis, P.S., Albert, V.A., Oppenheimer, D.G., DePamphilis, C.W., Ma, H., Frohlich, M.W. and Theissen, G.2002. Missing links: the genetic architecture of flower and floral diversification. *Trends Plant. Sci.* 7: 22-31.

22. Swanson, J.D. 2005. Flower development in Cocoa: An assessment of morphological and molecular conservation of floral development between *Arabidopsis thaliana* and *Theobroma cacao. PhD Dissertation.* Pennsylvania State University, University Park, Pennsylvania, USA.

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