Chapter 33

Cocoa

☆ *V. Arunachalam*

1. Introduction

Cocoa (*Theobroma cacao* L.) is an important perennial tree of the tropics and forms a major ingredient of chocolates and confectionery recipes. It is a shade loving tree commonly grown as inter/mixed crop in orchards and plantation crops. Cocoa is a diploid plant (2n = 2x = 20) and belongs to family *Malvaceae* with an estimated genome size of 380 Mbp (Figueira *et al.*, 1992). Genetic improvement of cocoa is aimed at improving traits such as yield, bean quality and resistance to diseases. Three groups of cacao cultivars are known traditionally worldwide: Criollo, Forastero and the hybrid between the two, Trinitario. Criollo is a high quality variety, but with poor yield and susceptible to fungal diseases. Currently, microsatellite markers have greatly aided in grouping the cacao cultivars into ten major groups (Motamayor *et al.*, 2008). Cocoa is vegetatively propagated by grafting method- hence desirable single plants can be multiplied clonally. Clonal selection among progenies helps in the further improvement in the traits of interest. Recent developments in genomics and transcriptomics research have paved the way for understanding of genes, enzymes, and pathways of important traits of cocoa crop. Genomic science has provided novel molecular markers hence quick strategies are available to cocoa breeders. Readers are advised to refer review articles on cocoa agronomy and general aspects (Badrie *et al.*, 2015); genomics (Bennett, 2003; Guiltinan *et al.*, 2008), linkage analysis and QTL mapping (Lanuad *et al.*, 2009), Witches' broom disease interactions (Teixeira *et al.*, 2015) for further details. The objective of the current chapter is to highlight the major developments in the areas of cocoa genomics and transcriptomics research.

2. Cocoa Genomics

The cacao plant is characterised by small chromosomes, single secondary constriction, and lack of C-banding due to small size (2.01×10^8) of its genome (Couch *et al.*, 1993). Besides these physical attributes of chromosomes, extent of methylation in the cacao genome has also been studied in detail. A massive transcriptome resource was generated by sequencing 56 cDNA libraries representing various tissues, genotypes and environments, which yielded 149,650 sequences linked to genes governing important traits in cocoa (Argout *et al.*, 2008). Bacterial Artificial Chromosome (BAC) resources were developed for the first time by Clement *et al.* (2004) to aid physical mapping of resistance genes. Generated from the genotype 'SCA6', the BAC resource contains approximately 11 genome equivalents, with an average insert size of 120 kb. Soon, a second BAC library was created; the library obtained from 'LCT-EEN37' genotype, collected in Ecuador, represents approximately 11 genome equivalents with an average insert size of 120 kb. BAC libraries were also generated from Criollo genotype 'B97-61', collected from Belize and a Forastero genotype, 'Matina 1-6', collected from lower Amazon. The main aim of construction of these BAC resources was to support genome sequencing projects. BAC-based approach was also followed to sequence and assemble a QTL-rich region, of ~3 Mbp cacao genome which revealed several important genes governing the traits such as black pod disease resistance, bean shape index, and pod weight (Feltus *et al.*, 2011).

The International Cacao Genome Sequencing (ICGS) consortium was formally organized in 2006 at the International Cocoa Research Conference held in Costa Rica. Two years later, in 2008, the Cocoa Genome Consortium (CGS)—an industry-funded partnership—was formed. The objectives of these two consortiums were to undertake sequencing of the genomes of two quite distinct genetic groups of cacao: Criollo and Amelonado, respectively. The rationale of sequencing two different genetic types of cocoa was that the comparison of these two quite distinct types would provide a far deeper understanding of the structure and function of the cocoa genome compared to a single genome (Guiltinan and Maximova, 2015). Both these types were chosen because of their highly homozygous genomes which would immensely facilitate final genome assemblies.

The ICGS sequenced the genome of a Criollo type (B97-61/B2) collected from Maya mountains of Belize (Argout *et al.*, 2011), whereas the CGS sequenced Matina 1-6, a Costa Rican variety from the Matina river valley (Motamayor *et al.*, 2013). The sequencing data revealed a difference of about 3.4% (430 Mbp and 445 Mbp for B97-61/B2 and Matina 1-6, respectively) between the genomes of these two sequenced types of cacao; most of the difference was accounted for by the increased amount of repetitive DNA and transposons found in the Matina 1-6 genome (Motamayor *et al.*, 2013). Approximately 29, 000 genes have been predicted in cacao genome, similar to number of genes found in the model plant *Arabidopsis*. In addition, over 700 novel genes were found; these genes have been suggested to be involved in specializations within the cacao lineage (family Malvaceae).The global organization of the genomes of the two cocoa types was found to be quite similar; however, 12 relatively small regions were found to be located on different chromosomes in the

two types, which could have possibly have arisen as a result of transposon activity (Motamayor *et al.*, 2013). The release of these two genomes is a major milestone in the genomics of cocoa as it offered interesting conclusions and useful information for cocoa breeders. Argout *et al.* (2011) have also listed the genes responsible for fungal and oomycetes resistance and flavour quality of cocoa.

3. Omics of Fungal Disease Resistance

3.1. Witches' Broom Disease

Witches' broom disease (WBD), caused by the pathogenic hemibiotrophic fungus *Moniliophthora perniciosa*, is a major disease of *Theobroma cacao* and can cause up to 90% yield loss. The disease has severely affected the cocoa industry in Brazil and is a major factor of yield loss in cacao (Teixeira *et al.*, 2014). Gesteira *et al.* (2007) generated cDNA libraries from meristem of resistant cacao genotype, TSH1188, and the susceptible genotype, Catongo, after inoculation with *M. perniciosa*. From these two libraries, a total of 6884 ESTs could be obtained, which corresponded to 2926 non-redundant sequences (2585 singletons plus 341 contigs). Putative functional categories could be assigned to 54% of these sequences. Even though the overall distribution of sequences in functional categories between the two libraries was quite similar, differences could be observed with respect to genes encoding PR proteins in TSH188, the resistant genotype, and genes involved in programmed cell death (PCD) in Catongo, the susceptible genotype.

Two SSH (subtractive suppressive hybridization) libraries were constructed from meristems, collected from the resistant genotype CAB 214 and the susceptible genotype ICS39, after inoculation with *M. perniciosa* by Leal *et al.* (2007), subtracting common transcripts in both directions. A total of 104 and 187 unique sequences were obtained respectively, from each of these two libraries. Out of the 23 genes evaluated by RT-qPCR, only 16 were induced in the susceptible genotype, while 21 were induced in the resistant genotype.

Gene expression analyses in the disease resistant variety 'TSH1188' and the susceptible variety 'Catongo', revealed the production of ROS (reactive oxygen species) and elicitor molecules during infection followed by detoxification of ROS in the resistant variety (da Hora Junior *et al.*, 2012). The report also identified 154 and 227 genes from TSH118 and Catongo, respectively, which were differentially expressed during fungal infection. One hundred and fifty-three genes, potentially related to plant-pathogen interaction, were identified from infected plant tissue libraries along with 71 putative SNPs (Lima *et al.*, 2009).

Lopes *et al.* (2010) focused on cacao transcription factors (TFs) by developing a macroarray with 88 TF cDNA from interaction libraries (Gesteira *et al.*, 2007). Seventy-two TFs were found differentially expressed between the susceptible (Catongo) and resistant (TSH1188) genotypes and/or during the disease time course—from 24 to 30 days after infection. Most of the TFs differentially expressed belonged to bZIP, MYB and WRKY families, and presented opposite expression patterns in susceptible and resistant cacao–*M. perniciosa* interactions. The results of the macroarray were confirmed by RT-qPCR for bZIP and WRKY TFs (Lopes *et al.*,

2010). On the other hand, SVP (short vegetative phase), which shared similarity with a *Populus tomentosa* MAD-Box transcription factor, was up-regulated in resistant CAB plants (Leal *et al.*, 2007). Dual RNA-seq analysis was performed to characterize transcriptional changes of both cacao and *M. perniciosa* during infection of pathogen in biotrophic stage. The normal transcriptional machinery of cacao was found to be heavily disturbed during the infection of *M. perniciosa* leading to hormonal imbalances in host (Teixera *et al.*, 2014). In addition, infection due to *M.perniciosa* causes carbon deprivation status in the cacao plants leading to premature senescence. Thus the phenotypic symptoms associated with fungal infection and transition from biotropic to necrotic stage has been correlated with transcriptional changes of cacao plants. Transcriptomics studies of fungus in necrotrophic phase revealed upregulation of the secreted proteins that are actively involved in pathogenesis by degradation of plant cell wall (Meinhardt *et al.*, 2014).

Analysis of segregating mapping population of cocoa (derived from a cross between the resistant 'TSH 1188' and the tolerant 'CCN 51') for Witches' broom disease (WBD) identified seven QTLs spread over five chromosomes that confer resistant to WBD. The study not only identified potential candidate disease resistance genes in the QTL regions, but a few SNP-based molecular markers were also proposed to aid breeding for resistance against WBD (Royaert *et al.*, 2016). Ultimately, greater understanding of molecular mechanism underlying WBD has led to development of WBD Transcriptome Atlas initiative (http://www.lge.ibi.unicamp.br/wbdatlas), which is a repository of all sequence libraries generated from cacao-*M. perniciosa* interactions.

3.2. Black Pod /Pod Rot Disease

Black pod, caused by *Phytophthora megakarya*, is a major disease of cocoa in West African and Asian countries. The genetic resistance of cacao to three pathogens of pod rot (*Phytophthora palmivora*, *P. megakarya* and *P. capsici*) were studied by linkage analysis. A total of 13 QTLs, spread over six chromosomes, were identified. Among these, a major one was qPsp-5 governing resistance to five races and three species of pathogens (Risterucci *et al.*, 2003). Three QTLs for black pod resistance were found on LG 4, 8, and 10, with the most favourable alleles coming from the cacao genotype Pound 7 (Brown *et al.*, 2007). A meta-QTL study on cocoa combining QTL and linkage mapping experiments revealed only 13 major and consensus QTLs (Clement *et al.*, 2003; Brown *et al.*, 2007) from the linkage studies (Review: Lanuad *et al.*, 2009). The study lists unique and common QTLs for pod rot and disease resistance. *In silico* analysis of transcriptomics data from cocoa infected by black pod disease, found 272 enzymes corresponding to 114 metabolic pathways. The annotated enzymes from the study were involved in amino acid biosynthesis and phenylpropanoid biosynthesis. The study has implications in understanding the biotic stress response pathway in cacao (Naganeeswaran *et al.*, 2012). A major QTL region spanning 3 Mb size on LG5 covering the resistance to black pod disease was sequenced from Matina 1–6 cacao clone to identify SNP markers on COS (conserved orthologous sequence) genes (Kuhn *et al.*, 2012). Unravelling molecular mechanism underlying resistance to pod rot pathogen (*Phytophthora tropicalis*) in two cacao genotypes with contrasting disease resistance traits Scania6 (Sca6-resistant) and Imperial College Selection 1

(ICS1-susceptible), revealed that salicylic acid treatment enhanced production of reactive oxygen species (ROS) through upregulation of organeller genes involved in ROS production in Sca6 whereas ICS1 produced pathogenesis related proteins (Fister et al., 2015). Furthermore, it was also deduced that transient overexpression of *TcNPR1*- a transcriptional regulator involved in salicylic acid dependent immune system also- enhanced resistance to pod rot pathogen (Fister *et al.*, 2015).

3.3. Ceratocystis Wilt Resistance

Ceratocystis wilt is a lethal wilt disease of cacao present in Caribbean and Central and South America. A mapping population involving Scavina 6 (Sca 6; resistant genotype) × Imperial College Selections 1 (ICS 1; susceptible genotype) and a set of EST-SSR markers was employed by Santos *et al.* (2013) to identify additional SSR markers (CEPEC13, CEPEC 14, CEPEC 28 and CEPEC 17) and to tag resistance genes of the disease.

4. Genomics of Flooding Stress Tolerance

Studies conducted regarding flooding stress tolerance (40-day flooding) on 35 elite cacao genotypes, also identified polymorphism for 248 alleles of 18 microsatellite loci (Bertolde *et al.*, 2010). Gene expression pattern in cocoa plants suffering from soil anoxia caused by flooding found three major genes (alcohol / lactate hydrogenases and pyruvate carboxylase in leaves and roots. Activity of these enzymes also differed during flooding in TSA-792 and TSH-774 the stress tolerant and susceptible genotypes respectively (Bertolde *et al.*, 2014).

5. Omics of Cacao *in vitro* Culture

Deciphering molecular mechanisms underlying cocoa somatic embryogenesis (SE) would help devise better tools for regeneration. A gene governing somatic embryogenesis in cocoa called as leafy cotyledon1 (*TcLEC1*) was characterised by Alemanno *et al.* (2008). Later, a candidate gene *TcBBM*, known as baby boom transcription factor orthologous to that of *Arabidopsis*, was identified as a biomarker for embryogenesis in cacao tissue (Florez *et al.*, 2015) and was confirmed by over-expression in *Arabidopsis* and cocoa transgenic systems. Transient expression studies of *TcLEC1* and *TcBBM* and enhanced somatic embryogenesis observed in cacao underlined the importance of transcriptional factors (TFs) induced embryogenesis and identification of functional biomarkers associated with SE in cacao (Zhang *et al.*, 2014; Florez *et al.*, 2015). Most importantly, studies on cacao TFs have opened new avenues for development of efficient regeneration system thereby making cocoa transgenics a possibility.

6. Genomics of Pod Colour

The genome sequence of Matina 1-6 was analysed *via* haplotype, association mapping and gene expression studies to identify candidate gene(s) that governs cacao pod colour (Motomayor *et al.*, 2013). The R2R3 MYB class transcription factor *TcMYB113* has been found to be involved in red colour pigmentation of cocoa pods. Furthermore, SNP identified in the *TcMYB113* has been shown to affect the activity

of *trans*-acting siRNAs (siRNAs) targeting *TcMYB113* and hence cause pod colour variation (Motomayor et al., 2013).

7. Omics of Cacao Flavour

Flavour of cacao is an important criterion for chocolate industry and it has also been demonstrated that flavour is a genetically controlled character besides flavour components depend on post-harvest processing conditions (Clapperton *et al.*, 1994). Relative expression levels of *TcLIS* (cacao linalool synthase) from cotyledons during fermentation were found to increase in 'ICS1' and 'Nacional' seeds (Sabau *et al.*, 2006). Similarly, to decipher other quality parameters, cDNA macroarray based expression analysis was carried out to delineate genes involved in terpenes and polyphenol biosynthesis pathways (Sabau *et al.*, 2012).

8. Bioinformatics and Databases

Storage and retrieval of the genomics data on user-friendly online databases and web servers is important so as to use the information for crop breeding. Table 33.1 lists GenBank accessions that house major genomic and transcriptomic resources of cacao. Cacao genome database (http:// www.cacaogenomedb.org) provides several services of visualization and browsing of the DNA sequence, genes on all the 10 chromosomes of cocoa plant (Zheng, 2012). TropGeneDB (http://tropgenedb.cirad.fr) holds updated information regarding cocoa linkage maps, QTLs, molecular markers and genes (Hamelin *et al.*, 2013).

Table 33.1: Major Genomic and Transcriptomic Resources of Cacao

GenBank Id(s)	*Type*	*Materials*	*Reference*
CU469588 to CU633156	56 EST libraries	Genotypes differing self compatibility, disease resistance, flavour / quality of bean and corresponding tissues	Argout *et al.* (2008)
CACC01000001–CACC01025912	Whole genome sequence of 10 chromosomes	Criollo cultivar	Argout *et al.* (2011)
ALXC01000000	Genome assembly of whole genome	Matina 1-6 clone	Motamayor *et al.* (2013)

With the advent of cost and time effective sequencing technologies, use of computational approaches becomes essential for analysis of voluminous data from genomics and transcriptomics projects (review by Arunachalam, 2014). Gene index provides curated sequence information and details of the known genes in an organism. TIGR gene index was developed for cocoa plant (Quackenbush *et al.*, 2001). Few attempts were made using bioinformatics tools to mine the SSR (Riju *et al.*, 2009) and SNP (Lima *et al.*, 2009; Riju and Arunachalam, 2010) markers. The SNPs mined from cocoa ESTs are available in the online database (http://www.riju.byethost31.com/cocoa/ccsnp.html). Of the 6578 EST sequences from seven tissues/libraries, a density of one SNP/166 bp and one Indel/360 bp were found in cocoa. Similarly, Cacao EST sequences are mined for simple sequence repeats by

performing computational analysis . Identified SSRs and primers designed were made available as online database as http://riju.byethost31.com/cocoa/. Putative function of the simple sequence repeat containing sequences were analysed *in silico* and found to be PGK (phosphoglycerate kinase) and gibberellin 20-oxidase 1, among others (Riju *et al.*, 2009).

A bioinformatics tool has been developed to locate RAPD/ISSR primers and design *in silico* SCAR primers (PremKrishnan and Arunachalam, 2012). The software was used to mine the complete genome, EST and core nucleotide sequences of cacao for RAPD/ISSR priming sites and iSCAR (*in silico* Sequence Characterised Amplified Region) Markers (http://www.bioinfoindia.org/fv-iscardb/) have been designed. The database current holds 9123 predicted SCAR markers for use in cacao breeding programs (PremKrishnan and Arunachalam, in press).

Conclusion

The availability of the cacao genome sequence will accelerate the discovery of candidate genes underlying important QTLs identified in genetic studies and in relation to functional genomics. The genome sequencing of both contrasting Criollo and Forastero genotypes, originated from distinct genetic groups, will allow to produce a wide SNP resource useful for all genetic and genomics studies. Developments in the field of cacao 'omics' especially in transcriptomics have unraveled role of molecular components that play significant role in somatic embryogenesis, conferring resistance to diseases such as WBD, pod rot, identification of biomarkers associated with flooding tolerance etc would aid in effective designing future crop improvement programmes.

References

Alemanno, L., Devic, M., Niemenak, N., Sanier, C., Guilleminot, J., Rio, M. and Montoro, P. (2008). Characterization of leafy cotyledon1-like during embryogenesis in *Theobroma cacao* L. *Planta*, **227(4)**: 853-866.

Argout, X., Fouet, O., Wincker, P., Gramacho, K., Legavre, T., Sabau, X., Risterucci, A.M, Da Silva, C., Cascardo, J., Allegre, M. and Kuhn, D. (2008). Towards the understanding of the cocoa transcriptome: Production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* L. generated from various tissues and under various conditions. *BMC Genomics*, **9(1)**: 512.

Argout, X., Salse, J., Aury, J. M., Guiltinan, M. J., Droc, G., Gouzy, J. Allegre, M., Chaparro, C., Legavre, T., Maximova, S.N. and Abrouk, M. (2011). The genome of *Theobroma cacao*. *Nature Genetics*, **43**: 101-108.

Arunachalam, V. (2014). Computational genomics of plants. CAB Reviews. Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources.*CAB Reviews*: (CABI Press): **9(036)**: 1-9.

Badrie, N., Bekele, F., Sikora, E. and Sikora, M. (2015). Cocoa agronomy, quality, nutritional, and health aspects. *Critical Reviews in Food science and Nutrition*, **55**(5): 620-659.

Bennett, A. B. (2003). Out of the Amazon: *Theobroma cacao* enters the genomic era. *Trends in Plant Sciences*, **8(12)**: 561-563.

Bertolde, F.Z., Almeida, A.A.F. and Pirovani, C.P. (2014). Analysis of gene expression and proteomic profiles of clonal genotypes from *Theobroma cacao* subjected to soil flooding. *PLoS ONE*, **9**(10): e108705. doi: 10.1371/journal.pone.0108705.

Bertolde, F.Z., Almeida, A-AF, Correa, R.X., Gomes, F.P., Gaiotto, F.A., Baligar, V.C. and Loguercio, L. L. (2010). Molecular, physiological and morphological analysis of water logging tolerance in clonal genotypes of *Theobroma cacao* L. *Tree Physiology*, **30**: 56–67.

Brown, J. S., Phillips-Mora, W., Power, E. J., Krol, C., Cervantes-Martinez, C., Motamayor, J. C. and Schnell, R. J. (2007). Mapping QTLs for resistance to frosty pod and black pod diseases and horticultural traits in *Theobroma cacao* L. *Crop Science*, **47**(5): 1851-1858.

Clapperton, J. F., Yow, S. T. K., Chan, J. and Lim, D. H. K. (1994). Effects of planting materials on flavour. *Cocoa Growers' Bulletin*, 48: 47–59.

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 (*Theobroma cacao* L.) for physical mapping of RGA containing BAC clones. *Theoretical and Applied Genetics*, **108**(8): 1627-1634.

Clement, D., Risterucci, A. M., Motamayor, J. C., N'goran, J. and Lanaud, C. (2003). Mapping QTL for yield components, vigour and resistance to *Phytophthora palmivora* in *Theobroma cacao* L. *Genome*, **46(2)**: 204-212.

Couch, J. A., Zintel, H. A. and Fritz, P. J. (1993). The genome of the tropical tree *Theobroma cacao* L. *Molecular and General Genetics*, **237(1-2)**: 123-128.

Feltus, F.A., Saski, C.A., Mockaitis, K., Haiminen, N., Parida, L., Smith, Z., Ford, J.,Staton, M.E., Ficklin, S.P., Blackmon, B.P., Cheng, C.H. (2011). Sequencing of a QTL-rich region of the Theobroma cacao genome using pooled BACs and the identification of trait specific candidate genes. *BMC Genomics*, **27**(12): 379.

Figueira, A., Janik, J. and Goldsbrough, P. (1992). Genome size and DNA polymorphism in *Theobroma cacao*. *Journal of the American Society for Horticultural Science*, **117:** 673-677.

Fister, A.S., O'Neil, S.T., Shi, Z., Zhang, Y., Tyler, B.M., Guiltinan, M.J. and Maximova, S.N. (2015). Two *Theobroma cacao* genotypes with contrasting pathogen tolerance show aberrant transcriptional and ROS responses after salicylic acid treatment. *Journal of Experimental Botany*, doi:10.1093/jxb/erv334.

Florez, S. L., Erwin, R. L., Maximova, S. N., Guiltinan, M. J. and Curtis, W. R. (2015). Enhanced somatic embryogenesis in *Theobroma cacao* using the homologous BABYBOOM transcription factor. *BMC Plant biology*, **15**(1): 121.

Gesteira, A.S., Micheli, F., Carels, N., da Silva, A.C., Gramacho, K.P., Schuster, I., Macedo, J., Pereira, G.A.G. and Cascardo, J.C.M., 2007. Comparative analysis of expressed genes from cacao meristems infected by *Moniliophthora perniciosa*. *Annals of Botany*, **100**: 129–140

Guiltinan, M. J., Verica, J., Zhang, D. and Figueira, A. (2008). Genomics of Theobroma cacao, "The Food of the Gods". In: *Genomics of Tropical Crop Plants*. Springer, New York. pp. 145-170.

Hamelin, C., Sempere, G., Jouffe, V. and Ruiz, M. (2013), TropGeneDB, the multi-tropical crop information system updated andextended. *Nucleic Acids Research*, **41**: D1172–1175.

Kuhn, D. N., Livingstone III, D., Main, D., Zheng, P., Saski, C., Feltus, F. A., Mockaitis, K., Farmer, A.D., May, G.D., Schnell, R.J. and Motamayor, J.C. (2012). Identification and mapping of conserved ortholog set (COS) II sequences of cacao and their conversion to SNP markers for marker-assisted selection in Theobroma cacao and comparative genomics studies. Tree Genetics & Genomes, **8(1):** 97-111.

Lanaud, C., Fouet, O., Clément, D., Boccara, M., Risterucci, A. M., Surujdeo-Maharaj, S. and Argout, X. (2009). A meta–QTL analysis of disease resistance traits of *Theobroma cacao* L. *Molecular Breeding*, **24**(4): 361-374.

Leal, Jr, G.A., Albuquerque, P.S.B., Figueira A. (2007). Genes differentially expressed in Theobroma cacao associated with resistance to witches' broom disease caused by *Crinipellis perniciosa*. *Molecular Plant Pathology*, **8**: 279-292.

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

López, C.M.R., Wetten, A.C. and Wilkinson, M.J. (2010). Progressive erosion of genetic and epigenetic variation in callus-derived cocoa (*Theobroma cacao*) plants. *New Phytologist*, **186**: 856-868.

Meinhardt, L.W., Costa, G.G., Thomazella, D.P., Teixeira, P.J., Carazzolle, M.F., Schuster, S.C., Carlson, J.E., Guiltinan, M.J., Mieczkowski, P., Farmer, A. and Ramaraj, T. (2014). Genome and secretome analysis of the hemibiotrophic fungal pathogen,Moniliophthororaroreri, which causes frosty pod rot disease of cacao: mechanisms of the biotrophic and necrotrophic phases. *BMC Genomics*, **27(15)**:164.

Motamayor, J. C., Lachenaud, P., Mota, J. W. D. S., Loor, R., Kuhn, D. N., Brown, J. S. and Schnell, R. J. (2008), Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). *PLoS ONE*, **3**(10): e3311.

Motamayor, J.C., Mockaitis, K., Schmutz, J., Haiminen, N., Livingstone, D. III, Cornejo, O., Findley, S.D., Zheng, P., Utro, F., Royaert, S., Saski, C., Jenkins, J., Podicheti, R., Zhao, M., Scheffler, B.E., Stack, J.C., Feltus, F.A., Mustiga, G.M., Amores, F., Phillips, W., Marelli, J.P., May, G.D., Shapiro, H., Ma, J., Bustamante, C.D., Schnell, R.J., Main, D., Gilbert, D., Parida, L. and Kuhn, D.N. (2013). The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod colour. *Genome Biology*, **14**: 1–24.

Naganeeswaran, S.A., Subbian, E.A. and Ramaswamy, M. (2012). Analysis of expressed sequence tags (ESTs) from cocoa (*Theobroma cacao* L.) upon infection with *Phytophthora megacarya*. **8**: 65-69.

Premkrishnan, B.V. and Arunachalam, V. (2012). *In Silico* RAPD priming sites in expressed sequences and iSCAR markers for oil palm. *Comparative and Functional Genomics*, doi: 10.1155/2012/913709.

Premkrishnan, B.V. and Arunachalam, V. (In Press). Database of predicted SCAR markers in five fruit and three vegetable crops. *J Genet*. Unedited article online http://www.ias.ac.in/public/Resources/General/jgen/jgen-14-703-ue.pdf.

Quackenbush J., Cho D., Lee F. L., Hott I., Karamychera S. And Parizi B. (2001). The *TIGR* gene indices: Analysis of gene transient sequences in highly sample eukaryotic species. *Nucleic Acids Research*, **29**: 159–164.

Riju, A. and Arunachalam, V. (2010). Electronic sorting of SNP/Indel Sites in expressed sequence tag libraries of cocoa (*Theobroma cacao* L.). *Genes, Genomes and Genomics*, **4**: 37-40.

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

Risterucci, A. M., Paulin, D., Ducamp, M., N'goran, J. A. K. and Lanaud, C. (2003). Identification of QTLs related to cocoa resistance to three species of *Phytophthora*. *Theoretical and Applied Genetics*, **108(1)**: 168-174.

Royaert, S., Jansen, J., da Silva, D.V., de Jesus Branco, S.M., Livingstone, D.S., Mustiga, G., Marelli, J.P., Araújo, I.S., Corrêa, R.X. and Motamayor, J.C. (2016) .Identification of candidate genes involved in Witches' broom disease resistance in a segregating mapping population of *Theobroma cacao* L. in Brazil. *BMC Genomics*, **17(1)**: 1.

Sabau, X., LoorSolorzano, R. G., Boccara, M., Fouet, O., Jeanneau, M., Argout, X., Legavre, T., Risterucci, A. M., Vincker, P., Da Silva, C., Jimenez, J. C. and Cros, E. (2006). Preliminary results on linalool synthase expression during seed development and fermentation of nacional and trinitario clones. *15th International Cacao Research Conference*, San Jose, Costa Rica.

Santos L.L.L., Legravre, T., Lanaud, C. Schnell, R.J., Pires, J.L., Lopes, U.V., Micheli. F. and Gramacho, K.P. (2013). Identification, characterization and mapping of EST-derived SSRs from the cacao–*Ceratocystis cacaofunesta* interaction *Tree Genetics & Genomes*, **9**: 117–127.

Teixeira, P.J., Thomazella, D.P., Reis, O., do Prado, P.F., do Rio, M.C., Fiorin, G.L., José , J., Costa, G.G., Negri, V.A., Mondego, J.M. and Mieczkowski P. (2014). High-resolution transcript profiling of the atypical biotrophic interaction between *Theobroma cacao* and the fungal pathogen *Moniliophthora perniciosa*. *Plant Cell*, **26(11)**: 4245-4269.

Teixeira, P.J.P.L., Thomazella, D.P.dT. and Pereira, G.A.G. (2015). Time for Chocolate: Current understanding and new perspectives on cacao witches' broom disease research. *PLoS Pathology*, **11**(10): e1005130. doi:10.1371/journal.ppat.1005130.

Zhang, Y., Clemens, A., Maximova, S.N. and Guiltinan, M.J. (2014). The *Theobroma cacao* B3 domain transcription factor *TcLEC2* plays a duel role in control of embryo development and maturation. *BMC Plant Biology*, **14**: 106. doi: 10.1186/1471-2229-14-106

Zheng, P. (2012). Cacao Genome Database. In: *Plant and Animal Genome XX Conference*, January 14-18, 2012.