Single Nucleotide Polymorphisms (SNPs) and haplotype structuring in the latex biosynthesis genes of *Hevea brasiliensis*

Thesis submitted to
The University of Kerala
for the award of the Degree of
Doctor of Philosophy
in Biotechnology

THOMAS K U

THE RUBBER RESEARCH INSTITUTE OF INDIA
RUBBER BOARD
(Ministry of Commerce & Industry, Govt. of India)
Kottayam – 686009, Kerala, INDIA

August 2016

DECLARATION

I hereby declare that this Ph.D thesis entitled **"Single Nucleotide Polymorphisms (SNPs) and haplotype structuring in the latex biosynthesis genes of *Hevea brasiliensis*"** is an independent work carried out by me under the supervision of Dr. T. Saha, Principal Scientist, Genome Analysis laboratory, at the Rubber Research Institute of India and it has not been submitted anywhere else for any other degree, diploma or title.

Kottayam

01/08/2016

Thomas K.U

भारतीय रबड गवेषण संस्थान
THE RUBBER RESEARCH INSTITUTE OF INDIA
(Ministry of Commerce and Industry, Government of India)
(वाणिज्य मन्त्रालय, भारत सरकार)

Phone : 91-481-2353311 (10 lines)
Fax : 91-481-2353327
Email : saha@rubberboard.org.in
Web : http://www.rubberboard.org.in

रबड़ बोर्ड
RUBBER BOARD
कोट्टयम - 9
KOTTAYAM 686 009

Certificate

This is to certify that the research work embodied in the thesis entitled, **"Single nucleotide polymorphisms (SNPs) and haplotype structuring in the latex biosynthesis genes of *Hevea brasiliensis*"** is an authentic record of original research work carried out by Mr. Thomas KU at the Rubber Research Institute of India, Kottayam - 686 009 under my supervision for the award of the degree of Doctor of Philosophy in Biotechnology under the Faculty of Applied Sciences, University of Kerala, Thiruvananthapuram, Kerala. It is also certified that the work presented in this thesis has not been submitted earlier for any other degree or diploma elsewhere.

01 August 2016
Kottayam

Dr. T. Saha
Principal Scientist
Genome Analysis Laboratory

Acknowledgement

First and foremost I look up with gratitude to God Almighty for the wisdom, health and strength bestowed on me to undertake this research work and enable its completion.

I express my reverence and gratitude to my supervisor Dr. T Saha for his invaluable support and counsel throughout this project. Being the pioneer in SNP studies in rubber this work could not have been possible without his vision and advice. I express my sincere thanks to Dr. James Jacob, Director of Research, RRII, Kottayam for granting me the permission for conducting my doctoral study at the Genome Analysis Lab of RRII and also for extending the necessary infrastructure for the successful completion of my PhD programme.

I wish to extend my special thanks to Smt. Minimol Ravindran, Sr. R Anantharamanan and Dr. S Pramod for their immense support , care and assistance provided during various stages of my work at RRII. I gratefully acknowledge the help and support extended to me by Dr. K Bini for the successful completion of this work. Thanks are due to Dr. Bindu Roy, Dr. Shaji Philip and Dr. Molly Thomas for providing me with consumable and reagents required for the work when there was a shortage. Special thanks are due to Sri. Madhusudhanan for helping me with the art work in the thesis.

I thank all my colleagues and students in RRII and Rubber Board, who have in some way or other helped me to complete my thesis work.

I am greatly indebted to my father Sri. Baby K Thomas whose love, prayers and encouragement was a constant source of energy for me to complete my thesis work. I thank my wife, Deepa for her whole hearted support and my son Daniel, to bear with my temper, short-comings and time restraints to take care of his needs during this period. I also wish to place on record my gratitude to my sister Bindu , my brother George and their respective families for the encouragement and prayers.

This thesis is dedicated to the memories of my dear mother Smt. Annie George who always wanted to see me earn my doctoral degree but did not stayed back to see it full filled. I consider this thesis as a reward to her endless prayers.

THOMAS K U

Contents

Page No.

Chapter 2

Materials and Methods

Chapter 3

Results

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AFLP	-	Amplified fragment length polymorphism
APS	-	Ammonium per sulphate
AS-PCR	-	Allele specific polymerase chain reaction
BD	-	Bodjong Datar
CTAB	-	Cetyl trimethyl ammonium bromide
dNTP	-	Deoxynucleotide triphosphate
EDTA	-	Ethylene diamine tetra acetic acid
Gl	-	Glenshiel
GT	-	Godang Tapen
HRM	-	High resolution meltcurve
LINE	-	Long interspersed nuclear element
LTR	-	Long terminal repeats
NCBI	-	National centre for biotechnology information
PB	-	Prang Besar
PCR	-	Polymerase chain reaction
PR	-	Proefstation Voor Rubbe
QTL	-	Quantitative trait loci
RAPD	-	Random amplified polymorphic DNA
RFLP	-	Restriction fragment length polymorphism
RNA	-	Ribonucleic Acid
RRIC	-	Rubber Research Institute of Ceylon
RRII	-	Rubber Research Institute of India

RRIM	-	Rubber Research Institute of Malaysia
SCAR	-	Sequence-characterized amplified region
SCATC	-	South China Academy of Tropical Crops
SINE	-	Short interspersed nuclear element
SSR	-	Simple sequence repeats
SNP	-	Single nucleotide polymorphism
TAE	-	Tris acetate EDTA buffer
TE	-	Tris-EDTA
TBE	-	Tris borate EDTA buffer
TEMED	-	Tetramethyl ethylene diamine
Tjir	-	Tjirandji
qRT-PCR	-	Quantitative real-time PCR
µL	-	microlitre

ABSTRACT

Single nucleotide polymorphisms (SNPs) and haplotype structuring in the latex biosynthesis genes of *Hevea brasiliensis*

Hevea brasiliensis (Para rubber tree), a member of the family *Euphorbiaceae*, produces about 98% of the world's natural rubber. Conventional genetic analysis in *Hevea* is difficult because of its perennial nature, long breeding and selection cycles, and difficulties in raising F_2 progenies. The development of molecular markers for genetic improvement is comparatively difficult in rubber due to their narrow genetic base, polygenic nature of traits like yield, disease resistance etc. In this context single nucleotide polymorphisms (SNPs) and indels are having relevance due to their abundance in the genome and wide distribution pattern even in the vicinity of virtually every gene. SNPs provide an important source of molecular markers that are useful in high-resolution genetic linkage mapping, map-based positional cloning, detection of marker-trait associations, phylogenetic analysis etc.

The present work was taken up with the objective of identification and characterisation of SNPs from the complete genomic sequence of major genes involved in the rubber biosynthesis pathway, so that the SNPs associated with functionality of a gene can ultimately be used as the marker in early selection of potentially high-yielding genotypes in *Hevea* breeding programs. Eight major rubber biosynthesis genes involved in the synthesis of key enzymes catalysing crucial steps in the upper mevalonate pathway leading to IPP monomer synthesis (*HMGS, HMGR, MVK and PMVK*), the chain initiation process (*FDPS, GGDPS*) and the final elongation steps (*CPT* and *REF*) leading to natural rubber production, were sequenced from five genetically divergent popular cultivated clones originated from Southeast Asian countries including India. Altogether, 172 SNPs were identified and characterised based on their position in the above genes. Overall SNP frequency data from the eight genes ranged from 1 SNP in every 53 bp to 410 bp, which is relatively high among crop plants indicating the rich allelic diversity prevailing in popular *Hevea* clones. From the allelic status of SNPs in latex biosynthesis genes it is assumed that the high homozygosity in the initiator molecule synthesising genes like *FDPS* and *GGDPS* observed in RRII 105 and RRIM 600 may have influence in their productivity. Haplotype reconstruction and its association with gene expression was also attempted in order to characterise the rubber clones and to understand the effect of haplotypic variation on respective gene expression. Putative intragenic recombination sites were predicted by statistical analysis to understand the inheritance pattern of SNP and to establish clonal identity. Recombinant sites were

identified even within the coding regions and the number of recombination sites was independent of the size of the gene.

Our results also revealed the existence of retrotransposon activity in the intronic region of a functional gene in *Hevea* like *FDPS* and the structural modifications induced by them in regulatory motif and splice sites, which may have functional role in gene expression. Comparative analysis of the protein sequences encoded by the genes involved in MVA pathway in related plant species revealed that some of the non-synonymous SNPs are highly conserved across species and genera, whereas some are unique to *Hevea* shedding light on the evolution of rubber biosynthesis genes in *Hevea brasiliensis*. *In silico* 3D protein structure modelling studies- a novel approach to identify the impact of non-synonymous SNPs on the protein structure of HMGS and GGDPS revealed that the non-synonymous SNPs did not result in any major change in the active binding sites of the protein.

Selected SNPs identified from the eight genes were converted to SNP markers using the High Resolution Melt Curve (HRM) based genotyping strategy. Segregation studies in a progeny population using these markers helped to map the latex biosynthesis genes in an existing linkage map of rubber. In order to evaluate the impact of identified haplotypes on the expression of respective gene, transcript abundance of each gene in latex from the five clones were estimated using real time gene expression study. qRT PCR studies indicated the possible association of specific haplotype combinations with gene expression, which may have a downstream impact on the level of rubber biosynthesis.

The results from this extensive sequence structure analysis carried out in major rubber biosynthesis genes in *Hevea* indicated that despite the universal presence of mevalonate pathway genes in all living organisms, the sequence structure varies depending on the functional property of each organism and the sequence variation in rubber indicated their species-specific evolutionary patterns, which might be responsible for latex producing capacity of this wonder tree. The Information generated is important for the development of high yielding rubber clones using marker assisted selection (MAS) by considerably reducing the time required for initial selection and screening. Moreover the best haplotype combination identified in these key biosynthesis genes can be transferred to the existing cultivated clones for genetic enhancement of latex production in *Hevea* through cisgenic approach.

Chapter 1

INTRODUCTION

1.1 An overview on Natural Rubber

Natural rubber (NR) is a high molecular weight polymeric substance with viscoelastic properties ideal for various dynamic or static industrial and medical applications. Rubber is synthesized in more than three thousand plant species, confined to approximately three hundred genera of seven families which mainly include Euphorbiaceae, Apocynaceae, Asclepiadaceae, Asteraceae, Moraceae, Papaveraceae, and Sapotaceae (Lewinsohn, 1991; Cornish *et al.*, 1993a). Though rubber containing latex is produced from all these plants, latex of high quality in sufficient quantity in an economically sustainable way is produced mainly from the rubber tree *(Hevea brasiliensis)*. The uniqueness of natural rubber lies in its physical properties of extensibility and toughness. Due to the crystalline arrangement of its molecules (*cis* 1-4 polyisoprene), the strength of the natural rubber is enhanced several folds and therefore it is considered to be "self-reinforcing." Despite all these favourable properties natural rubber is susceptible to temperature variations and will lose all the above mentioned properties under low (−25 °C) and high (50 °C) temperatures. Moreover it's cross reactivity to hydrocarbon oils as well as with oxygen and ozone in the atmosphere leads to the breakage of carbon-carbon double bonds resulting in the softening and cracking of the material over time (Encyclopædia Britannica). These disadvantages are overcome to a great extent by cross-linking the polymer chains through the process known as vulcanization.

The main use of natural rubber is in automobile sector, where major portion of the NR used is for heavy duty tyres. In addition to tyres, a modern automobile has more than 300 components made out of rubber and many of these are processed from NR. Uses of NR in hoses, footwear, battery boxes, foam mattresses, balloons, toys etc., are

well known. In addition to this, NR now finds extensive use in soil stabilization, vibration absorption and in road making. Isoprene rubber (IR) or synthetic rubber, a substitute of NR up to some extent is manufactured by the polymerization of synthetic isoprene, which is obtained from the thermal cracking of the naphtha fraction of petroleum. The product is at most 98 percent *cis*-1,4 polyisoprene, but it does not crystallize as readily as the natural material, and is not as strong or as tacky in the raw (unvulcanized) state. (Encyclopædia Britannica). Therefore NR is still inevitable for the manufacturing of many products which plays critical role in various sectors like, health, industry, defence and space research.

1.2. Para rubber tree- *Hevea brasiliensis*

The principal source of natural rubber is *Hevea brasiliensis* (Willd.ex Adr. De Juss) Muell. Arg. a tropical tree species belonging to the *Euphorbiaceae* family. It is native to the rainforests in the Amazon region of South America, including Brazil, Venezuela, Ecuador, Colombia, Peru and Bolivia. These trees are generally found in low-altitude moist forests. *Hevea brasiliensis* is a deciduous tropical tree with a cylindrical trunk which attains a height of 25-35 m with much-branched leafy canopy. It is monoecious with both male and female flowers found in the same inflorescence with a high degree of cross pollination. The fruit is an exploding 3-lobed, 3-seeded ellipsoidal capsule. The economic life span of the tree is around 32 years and the plants start its yield from the sixth year onwards. Natural rubber is present in colloidal form in the latex which is the cytoplasm of laticiferous cells present in the bark of rubber trees. The chemical composition of latex varies depending on the clone, season, soil condition, tapping frequency etc., but in general it contains approximately 30-40% rubber, 55-65 % water, and negligible quantities of resins, protein, sugar and ash (Majumder *et al.*, 2014). Latex is extracted by wounding the bark of rubber trees where laticifers are present abundantly by a process known as "tapping" and is collected in containers attached to the bark. Around eighty percentage of the crop from plantations is in the form of latex (Figure.1).

Figure 1. The source of natural rubber - *Hevea brasiliensis*.

Main photo: Rubber planted in the low lying areas of southern Indian without disturbing the natural habitat. Inset- (top) A tapped rubber tree- latex oozing from the bark of a rubber tree to the collecting cup attached to the stem. (bottom) Transporting of collected latex to processing centre over a swinging bridge in the forests of southern India constructed by the British in 1937 (Photograph courtesy- Dr.Vinoth Thomas).

Fresh Latex is slightly alkaline in nature which becomes acidic rapidly due to bacterial action which will lead to coagulation of the latex. Therefore before processing, anticoagulants like formalin, ammonia etc. are added to it at specified concentration. The coagulated latex can be used for making sheets, which can be dried and stored for various applications. Alternatively, preserved latex concentrates are also prepared and used for the manufacturing of various products.

Hevea was introduced to the modern world as a source of rubber by Sir Hentry Wickham by making a collection of rubber seeds from the confluence of river Tapajos and Amazon river in 1876 (Schultes, 1977). Seeds were sent to Kew Botanic Gardens in London and from there it was distributed to different Southeast Asian countries. The

seedling stock survived at Malaysia was known as the "Wickham base" (Simmonds, 1989) and it became the source of planting material to all Southeast Asian countries, Sri Lanka and India.

1.2.1. Genome content of *H. brasiliensis*

Though *H. brasiliensis* is generally considered as diploid, there were disputes regarding the ploidy level of its genome as it was initially considered as amphidiploids (2n=36, x=9) (Ong *et al.*, 1976). Molecular studies suggested that it might be allotetraploid in nature (Leitch *et al.*, 1998). However, karyotyping studies by Saraswathyamma *et al.* (1984) and the recent molecular genetic studies have shown its diploid nature with 2n=2x=36 chromosomes (Clement-Demange *et al.*, 2000). Low and Bonner (1985) reported that *Hevea* nuclear genome contains 48% of slowly annealing DNA (putative single copy) and 32% middle repetitive sequences with remaining highly repetitive or palindromic DNA. Uncertainty about the genome size was due to the ploidy assignment to the rubber genome (4C) against respective nuclear DNA content of the cells (4C = 8.3 pg) determined by flow cytometry (Bennett and Leitch, 1995). A statement made in the same report on the average DNA content of each chromosome (~0.2 pg) helped us to estimate further the genome size based on the total DNA content of *Hevea brasiliensis* bearing 36 chromosomes (7.2 pg). Therefore, haploid genome size will be 3.6 pg, equivalent to ~3.5 Gb. The recent whole genome sequencing studies indicate that the haploid genome size of *Hevea* is in the range of 1.5 to 2.0 Gb (Rahman *et al.*, 2013; Tang *et al.*, 2016), which appeared to be half of the estimated genome size. However, all the above studies suggest that *Hevea* has a complex genome with the presence of abundant repetitive regions and the genome size yet to be estimated correctly.

1.2.2. Rubber biosynthesis in *Hevea*

Rubber biosynthesis in *Hevea* laticifer cells has become a major field of the research and researchers all around the world are working to fully understand the mechanism by which this wonder product is synthesised in plants. Rubber molecules (*1,4*

cis-polyisoprene) are formed from polymerisation of molecules with 5 carbons isopentenyl diphosphate (IPP) and aggregated as rubber particles packaged within a membrane that protects them from oxidation in latex vessels. The general metabolic pathway of rubber biosynthesis is as follows: Sucrose from photosynthesis is actively transported into lateciferous cells through the plasmalemmic membrane, and is then hydrolyzed into glucose and fructose by invertase. These sugars are then converted into acetyl-CoA through glycolysis. Three molecules of acetyl-CoA are condensed into mevalonic acid and then to IPP. Polymerization of thousands of IPP leads to dimethylallelyl pyrophosphate (DMAPP), Farnesyl diphosphate (FDP) and geranylgeranyl diphosphate (GGPP). The action of the enzyme *cis*-prenyltransferase associated with REF (a molecule fixed on the rubber particles membrane) on the above complex will ultimately lead to natural rubber.

1.2.3. *Hevea* breeding

Breeding in *Hevea* is aimed with the objective of developing high yielding clones combined with desirable secondary characters like disease resistance and timber quality. It is evident that the breeding efforts were fruitful with the development of high yielding clones with a production potential of 3500 kg/ ha/year from the gene pool introduced to Southeast Asia by Sir Henry Wickham in 1876 with an average yield of just 200 to 300 kg/ha/year (Licy *et al.*, 1997; Panikkar *et al.*, 1980). Being a perennial tree crop, there are lots of breeding constraints in *Hevea brasiliensis* like the narrow genetic base, non-synchronous flowering, low fruit set, long gestation period, heterozygous nature and absence of fully reliable early selection parameters. The common procedures employed for crop improvement involves introduction, plus tree/ mother tree/ ortet selection, hybridization followed by selection. Introduction involves introducing new already established clones from one country or region to another country for direct planting after evaluation under local agroclimatic conditions and as base material for genetic improvement using selected clones among them. Apart from introduction the most important procedure followed for breeding in rubber is clonal selection. They can be either primary clones or hybrid clones, the former developed

through ortet or mother tree selection and the latter through controlled hybridization and selection programs. Usually selective hybridisation of promising parents is done among themselves and also with wild germplasm lines. The progenies are directly selected from seedling nurseries and cloned for further evaluation. Also, natural seedling population or half-sib population are screened for desirable characters including resistance to fungal pathogens during which susceptible and poor performing clones were discarded. Polycross gardens comprising of pre-potent clones are also utilised and the selection is generally exercised in the polyclonal seedling orchards, even at the stage of maturity. In addition to introduction, selection and hybridization methods in breeds, techniques like polyploidy, mutation and *in vitro* culture also have been attempted in *Hevea.*

1.2.3.1. Challenges in *Hevea* breeding and genetics

Crop improvement through breeding in rubber faces a number of challenges. The main constraints are: (1) the long time necessary for accurate selection, (2) the low female fertility which limits recombination and (3) the difficulties in the exchange of planting material. As there is less possibility to overcome these constraints, rubber breeding has to adapt in-depth analytical studies applied to small sets of genotypes as well as to develop simple, direct and robust approaches for selection directly from the wild germplasm. Among all, improving the latex yield potential is the major challenge faced by breeders due to the ever increasing demand for natural rubber. The current maximum level of latex yield performance can be estimated to around 30% more than RRII 105. So the target will be to produce a clone having higher productivity with the available genetic variability in the Wickham population using the available methods and tools.

1.2.3.2. Popular cultivated *Hevea* clones

Breeding in *Hevea* was always aimed with increasing the productivity and the popular clones currently available are the result of extensive breeding process using conventional breeding practices. Even though the initially developed clones were

mostly primary clones evolved by ortet selection, several hybridisation programmes were initiated during the first half of 20th century mainly by the Southeast Asian countries like Malaysia and Indonesia resulting in the development of high yielding varieties. Later on, countries like Sri Lanka, India, China and Thailand started research on rubber leading to the development of large number of clones with various desirable traits from their hybridisation experiments. In the earlier days rubber clones are denominated with a first part in letters (abbreviation of the origin) and a second part in figures. Examples for the denominations with their emblematic clones during the first half of the 20th century are AVROS (AV49, AV255, AV352, AV2037), Bodjong Datar (BD5, BD10), Djasinga (Djas1), Glenshiel (Gl1), Gondang Tapen (GT1), Kali Djeroek (KD1), Landb. Mij. "Oud Djember" (LMOD53), Lands Caoutch. Bedrijf (LCB1320), Pataroeman (Pat190), Pilmoor (Pil D65), Prang Besar (PB186), Proefstation voor Rubber (PR107 = LCB510), Tjirandji (Tjir1, Tjir16), Waringiana (War4), etc. A denomination was established to describe the different types of seeds or the genetic origin of clones issued from recombination: "illegitimate seedling families" (ill.) are issued from commercial plantings, with no known genetic origin; when only the mother parent of one clone is known (i.e. AV163), the origin of the clone is denominated as AV163 ill. "Legitimate" full-sib seedling families issued from hand pollination are indicated by mentioning first the female (seed) and then the male (pollen) parent (i.e. PB186 × Tjir16). In the recent times clones became representatives of the research centres from where they were developed with the abbreviation of the institute followed by the number like RRII 105 (Rubber Research Institute of India 105), RRIM 600 (Rubber Resarch Institute of Malayasia 600), RRIC 100 (Rubber Research Institute of Ceylon 100) etc.

1.2.3.3. Biotechnological interventions in genetic improvement of *Hevea*

Being a perennial outbred species with low fruit set and long breeding cycle, quick development of new high yielding clones is not easy in *Hevea* like other crops. In this context, the emergence of biotechnology, a combination of cellular and recombinant DNA techniques, has increased the availability of molecular tools and

methodologies for *Hevea* improvement. In addition, the understanding of *Hevea* genome and genetics has been significantly improved by the recent developments in the area of functional genomics and genotyping technologies. *In vitro* approaches applied to rubber tree like tissue culture, haplogenesis, somatic embryogenesis, protoplast culture, germination of immature embryos, cultivation of laticiferous tissue has been reviewed by Carron *et al.* (1989, 2001, 2005). Somatic embryogenesis was studied in *Hevea* to achieve rapid clonal propagation as an alternative to the use of grafting with mature trees. Moreover somatic embryogenesis as a means of regeneration opens up the possibilities for transgene technology in rubber. Somatic embryogenesis in rubber is becoming standardized in different laboratories worldwide as an efficient system for plant regeneration from cells (Sushamakumari *et al.*, 2000). Biotechnological tools have been used in a focussed manner to transfer various agronomically important genes associated with tolerance to abiotic stress, production of recombinant proteins, increased rubber biosynthesis etc through the *in vitro* channel (Jayashree *et al.*, 2003; Sobha *et al.*, 2003; Thlaseedharan *et al.*, 2008; Rekha *et al.*, 2014). At the same time, efforts have been made to transform *Hevea* cells *in vitro* in order to increase genetic variation in a targeted way and complement plant breeding efforts with the possibility of modifying cultivated clones with specific genes. However, in the short term, genetic transformation is becoming a powerful tool for investigating the rubber genome functioning with the assistance of targeted mutations. Altogether biotechnological interventions like *in-vitro* propagation and application of molecular markers hold promise as effective shortcut to achieve the difficult objectives in rubber breeding like shortening gestation period and selection of potential varieties in the juvenile stage itself.

1.3. Molecular markers in breeding

Marker-assisted selection offers prospects of accelerating the process of long term breeding objectives offered by the conventional approaches. Modern plant breeding methods are based on scientific principles of genetics, which began only with the rediscovery of Mendel's paper that was originally published in 1866 (Mendel,

1866). Science of genetics has progressed at a rapid pace and has laid the foundation for molecular plant breeding with the integration of new advances in biotechnology, genomics and molecular marker applications. The advent of molecular marker technology has revolutionized the entire scenario of biological sciences and widened the existing knowledge on polymorphism between individuals – 'from morphological to molecular level'. The recent PCR based approach, gel free visualization of PCR products and automation at various steps are the boons to the molecular marker approaches adopted for genome mapping and genetic diversity analysis in any organism. Molecular markers are based on naturally occurring polymorphisms in the DNA sequence, *ie.*, base pair deletions, substitutions or additions (Gupta *et al.*, 1999). They are considered as landmarks on chromosomes, which are essential in finding out where the genes are placed in a genetic map. Molecular markers can identify a particular aspect of phenotype or genotype or both, whose inheritance can be easily followed from generation to generation. Isozymes were the first developed markers before the advent of DNA based molecular markers (Markert and Moller, 1959). Molecular markers are widely used for diversity analysis of exotic germplasm and genotyping of cultivars, phylogenetic and evolutionary studies, construction of genetic linkage maps and gene tagging, mapping of genes of interest followed by map based cloning and finally for marker assisted selection (MAS).

1.4. Desirable properties of an ideal marker

The following attributes are expected to be present in a useful genetic marker:

- Highly polymorphic nature
- Co-dominance (ability to distinguish homozygous and heterozygous states)
- Genomic abundance
- Neutral behaviour to environmental or management practices
- Locus specificity
- Small quantities of template DNA required for genotyping
- High reproducibility

- Technically less demanding
- Ease of access
- Easy and fast assay
- Amenability to automation
- Easy exchange of data between laboratories
- Less operational and developmental costs

1.5. Classification of molecular markers

Molecular markers are broadly classified into three types based on the molecules and identification technology used; a) biochemical, b) hybridization based, and c) polymerase chain reaction (PCR) based. Biochemical markers like isozymes were the first to evolve followed by DNA based markers. Based on chronological evolution, DNA based markers which became available during the last two decades, can be broadly classified into three types (a) the first generation molecular markers including Restriction Fragment Length Polymorphisms (RFLPs), Random Amplified Polymorphic DNAs (RAPDs) and their modifications, (b) the second generation molecular markers including Simple Sequence Repeats (SSRs), Amplified Fragment Length Polymorphisms (AFLPs) and their modification forms, and (c) the third generation molecular markers including Expressed Sequence Tags (ESTs) and Single Nucleotide Polymorphisms (SNPs). Among these markers SNPs are considered as the new generation markers which are currently being used widely than any other marker.

1.5.1. Biochemical markers (Isozyme)

Isozyme markers have been used for over 60 years for various research purposes in biology. They are electrophoretically separable variants of an enzyme with qualitatively the same catalytic function (Bergmann *et al.*, 1989). Isozymes have been proven to be reliable genetic markers in breeding and genetic studies of plant species and are successfully used in several crop improvement programmes (Baes and Custsem, 1993). In the 1980s, isozymes were used in rubber as proteic genetic markers for

cultivar identification, genetic diversity analysis, control of progenies issued from hand pollination and reproductive biology (Chevallier, 1988; Paiva *et al.*, 1994; Sunderasan *et al.*, 1994) The main drawback of isozyme is their relatively low abundance and low level of polymorphism.

1.5.2. Hybridization based molecular markers

Restriction fragment length polymorphisms (RFLPs) were the most widely used hybridization based marker until the discovery of PCR based markers. The first DNA marker using RFLP technique in the construction of genetic maps was reported by Botstein *et al.* (1980) in human and thereafter used in plant research (Weber and Helentjaris (1989). These markers are co-dominant in nature and provide extremely reliable method for DNA typing, determining genetic relationships and genetic linkage mapping studies. But there are certain disadvantages like, the requirement of large quantity of DNA (50-200 μg), low levels of polymorphism exhibited and the lengthy procedure associated with it. Minisatellites/variable number of tandem repeats (VNTR) is similar to RFLPs but due to the high mutation rate of minisatellites, the level of polymorphism is substantial resulting in unique multilocus profiles for individuals.

1.5.3. PCR-based markers

The PCR based markers are mainly of two types, arbitrarily-primed markers and sequence tagged site (STS) markers from a known sequence. Arbitrarily-primed markers include randomly amplified polymorphic DNAs (RAPDs) (Welsh *et al.*, 1990) amplified fragment length polymorphisms (AFLPs) (Vos *et al.*, 1995) and inter-simple sequence repeats (ISSRs) (Zietkiewicz *et al.*, 1994). Though these technologies could generate large number of markers in a single reaction, the lack of reproducibility especially in the case of RAPDs was a major disadvantage. AFLP technique is an intermediate between RFLP and PCR, and is based on the detection of genomic restriction fragment by PCR amplification and can be used for DNA of any origin or complexity. Sequence tagged sites (STS) are single locus markers with known sequences (Olsen *et al.*, 1989). They include microsatellite (SSR) markers consisting of

tandemly repeating mono, di, tri, tetra or penta nucleotide distributed throughout the genomes which are technically difficult to develop but very reliable. They can be easily integrated into plant breeding programme for marker assisted selection of the trait of interest. Microsatellites are co-dominant, highly abundant in genomic region of eukaryotes and randomly distributed throughout the genome with preferential association in low copy regions (Morgante *et al.*, 2002).

Single Nucleotide Polymorphism (SNP) is a variation in a single nucleotide that occurs at a specific position in the genome, at which different sequence alternatives (alleles) exist in a population and each variation is present to some appreciable degree within a population (e.g. >1%). SNP and insertions/deletions (indels) are highly abundant and distributed throughout the genome in plants (Garg *et al.,* 1999). Automated SNP genotyping make it a very effective technique for diversity studies and segregation analysis of large populations. In plants, SNP seems to be more abundant than those in human genome as evidenced by the existence of one SNP per 20 bp in wheat (Picoult *et al.*, 1999) and one SNP per 70 bp in maize genome (Bhattramakki *et al.*, 2002). SNP studies provide a frame work for examining how population history, breeding system and selection affect genetic loci and delineate the mechanisms that lead to evolutionary diversification of genomes (Kota *et al.*, 2003).

1.6. SNP identification and genotyping techniques

Re-sequencing is considered as the best method for the discovery of SNPs and is still the gold standard in SNP identification. Sequencing can be either direct amplicon sequencing or cloned amplicon sequencing where heterozygous SNPs may be unambiguously determined in primary analysis which could be used for determining the haplotype structure. Another method which is widely practiced after the advent of next generation sequencing (NGS) technologies is the *in silico* SNP mining approach, which makes use of the transcriptome and whole genome sequencing data from various studies. In the eighties and nineties, other techniques like denaturing/temperature gradient gel electrophoresis (D/TGGE) and single strand conformational polymorphism

(SSCP) (Orita *et al.*, 1989) were widely used to detect SNP polymorphisms in forest and horticultural trees (Plomion *et al.*, 1999; Etienne *et al.*, 2002).

High-throughput, high density SNP genotyping has become an essential tool for QTL mapping, association genetics, gene discovery etc. in several crops (Zhu *et al.*, 2001; Rafalski, 2002). Because of the growing need of high-throughput genotyping due to the ever increasing genome sequencing information from the various transcriptome and whole genome assemblies studies, SNP markers are the marker of choice as it can be easily automated for quicker analysis. For the same reason, electrophoresis based genotyping is avoided and direct sequencing based (NGS) and fluorescence or luminescence signal detection methods are favoured. The basic principles that are used currently to distinguish between the possible nucleotides at the SNP sites at high-throughput are based on

- Allele specific PCR (Probe based and gel based - Includes Allele-specific oligo nucleotide hybridization, DNA chips, Bead based techniques etc.
- Template dependent DNA strand elongation by a DNA polymerase (Primer extension, Pyrosequencing etc)
- Double strand dependent ligation (Oligonucleotide ligation assay (OLA) coupled with DNA chips or Bead based techniques)
- Mismatch detection (Denaturing high performance liquid chromatography (DHPLC), High Resolution Meltcurve (HRM) etc.)

Allele specific oligonucleotide hybridisation related techniques (ASOH)

Allele specific PCR: The basic idea of this technique is that PCR amplification is dramatically reduced if there is a mismatch at the last 3' base of one of the primers. Primers are designed in such a way that the last 3' base of one of the primers will be allele specific depending on the SNP and the other primer in a conserved sequence not too far from the first one. Genotyping is performed by scoring the presence absence of the amplicon which can be performed directly on the total genomic DNA.

LNA based allelic specific primers: Locked Nucleic Acid (LNA™) nucleosides are a class of nucleic acid analogues in which the ribose ring is "locked" by a methylene bridge connecting the 2'-O atom and the 4'-C atom (Exiqon A/S, Denmark). In LNA technology, the modified nucleoside incorporated into a DNA or RNA oligonucleotide makes the pairing with a complementary nucleotide strand more rapidly with higher stability, which is demonstrated by an increase in the duplex melting temperature (Tm) of 2-8 °C per LNA™ monomer. The LNA based allele specific PCR method makes use of primers containing a single 3'-terminal Locked Nucleic Acid (LNA) base where the higher effective annealing temperatures of LNA-containing primers is increased 2-8 °C as compared to normal primers making allele discrimination more easier and specific (Lattora *et al.*, 2003).

5' nuclease assay (TaqMan™ assay): This technique uses the real time quantitative PCR method based on 5' exonuclease activity of the Taq DNA polymerase (Heid *et al.*, 1996). An oligonucleotide probe coupled to a reporter dye and a quencher dye at both ends is designed which hybridises with the SNP containing region. SNPs are detected as fluorescence intensity variations based on the release of reporter dye. Despite of its simplicity and high throughput, the high cost of the machine as well as the probes limits its usage widely.

DNA chip based techniques also makes use of allele specific oligonucleotide hybridisation, where hybridization is performed simultaneously on different oligonucleotides corresponding to each allele of SNP, and the oligonucleotides are identified by their position on the array. Similarly, bead based approaches are based on simultaneous hybridizations on different oligonucleotides, but each type of oligonucleotide is identified by being bound to fluorescent micro beads that can emit fluorescence when exited by laser.

Template dependent DNA strand elongation methods

Primer extension is a method of genotyping based on template dependent DNA strand elongation by a DNA polymerase which makes use of MALDI-TOF detecting the primer extension of a single allele specific oligonucleotide based on the molecular

weight of the extended product (Sauer *et al.*, 2000). Pyrosequencing is another reliable method of SNP genotyping based on template dependent DNA strand elongation amenable to high-throughput. It is a real-time DNA sequencing method based on the transformation of pyrophosphates released during DNA elongation by DNA polymerase into measurable light. During DNA elongation, a single pyrophosphate molecule is released following the incorporation of a single nucleotide which is detected by the laser. Since the dNTPs are sequentially added one by one to the reaction mix, it is possible to identify the correct complimenting base being incorporated.

Double strand dependent ligation

Oligonucleotide ligation assays (OLA): OLAs are based on the property of ligases that, two nucleotides cannot be efficiently ligated if one of them makes a mismatch with its complimentary strand. This method can be used in association with DNA chips or bead based techniques instead of ASOH.

Mismatch detection related techniques

Denaturing high performance liquid chromatography (DHPLC): In this technique, PCR products of wild-type and mutant alleles, differing by as little as a single base pair, are denatured by heating at 95°C and re-annealed by slow cooling. The resultant wild-type and mutant homoduplexes melt at higher temperatures than the mismatch containing wild-type/mutant heteroduplexes. Therefore hetroduplexes will usually elute before the homoduplexes. The difference in melting temperature between homo and heteroduplexes having one or more mismatches is measured using ion-pair high performance liquid chromatography (HPLC) at a precisely control temperature.

High Resolution Melting analysis (HRM): High Resolution Melting (HRM) is a novel, homogeneous, close-tube, post-PCR method, based on mismatch detection enabling genomic researchers to analyze genetic variations (SNPs, mutations, methylations) in PCR amplicons. It goes beyond the power of classical melting curve analysis by allowing studying the thermal denaturation of a double-stranded DNA in

much more detail and with much higher information yield than ever before. Melting curve analysis using fluorescent dyes has proven to be a highly sensitive method for mutation discovery and SNP genotyping (Liew *et al.*, 2004; Cho *et al.*, 2008; Smith *et al.*, 2008). In this method, PCR of a small amplicon consisting of the SNP is performed in the presence of a saturating intercalating dye such as EvaGreen®, which binds to double-stranded but not single-stranded DNA. When the PCR products were subjected to melt curve analysis, genetic variants with differences in base composition result in differences in melting temperature, which are detected by monitoring fluorescence intensity changes (Zhou *et al.*, 2004). In HRM assay method, the analysis can be performed in less than 10 min after PCR. Compared to probe based and sequencing based methods, this method is very cheap and does not require post-PCR sample processing. . Additionally, prior knowledge of the SNP variation present in a genotype is not required. Thus HRM analysis offers unique opportunities for genotyping species with limited genomic resources (Koeyer *et al.*, 2010). HRM has been used in genotyping in many plant species including apple (Chagné *et al.*, 2008), almond (Wu *et al.*, 2009), potato (Koeyer *et al.*, 2010) and alfalfa (Han *et al.*, 2012).

1.7. Molecular markers analysis in *Hevea*

The molecular markers have a great impact on various aspects of crop improvement from diversity studies to genome mapping, gene tagging, map based cloning and physical mapping of genes. The technique of molecular biology provides path-breaking information regarding time scale on which closely related species have diverged and types of genetic variations are associated with species formation (Ratnaparkhe *et al.*, 1998). They are also valuable in the analysis of gene pool variation of crops during the process of cultivar development and classification of germplasm. Molecular markers were being used in rubber extensively for the last three decades for diversity studies, characterisation of clones, linkage map construction and QTL analysis. In rubber, isozymes were utilized for clonal identification in right from the eighties. (Chevallier, 1988). Tools like minisatellites (Besse *et al.*, 1993a), RFLPs (Besse *et al.*, 1994), RAPDs (Varghese *et al.*, 1997; Venkatachalam *et al.*, 2001, 2002),

AFLPs (Lespinasse *et al.*, 2000); and SSRs (Atan *et al.*, 1996; Low *et al.*, 1996; Roy *et al.*, 2004, Saha *et al.*, 2005, Mantello *et al.*, 2012) were developed in *Hevea brasiliensis*.

Diversity analyses using molecular markers can effectively classify the existing biodiversity among plant species, which can be further useful in wild gene introgression programmes (Brown and Kresovich, 1996). STMS, ISSR, AFLP markers etc. can provide useful criteria for evaluating the gene pool of crop plants. In *Hevea*, identification of 73 Wickham clones was carried out with 13 probes associated with restriction enzyme *Eco* RI (Besse *et al.*, 1993b). Similarly, RFLPs were also used for identification of progenies with two common parents (Low *et al*, 1996). The genetic diversity of Amazonian and Wickham clones were also assessed by Besse *et al.* (1994) based on RFLP profiles and concluded that Amazonian population brought genetic enrichment to *Hevea* germplasm. Ribosomal DNA variations were also studied using this technique (Besse *et al.*, 1993a?). Genetic diversity analysis using microsatellite markers confirmed that wild accessions are more polymorphic than cultivated Wickham clones (Lekawipat *et al.*, 2003). RAPD markers were successfully used to estimate genetic distances and relationships in cultivated *Hevea* clones and to distinguish them based on their origin (Varghese *et al.*,1997; Venkatachalam *et al.*, 2002). Additionally, mitochondrial DNA (mtDNA) was analyzed over 395 genotypes of rubber with heterologous probes from broad bean by Luo and Boutry (1995). They reported high mt DNA polymorphism in Amazonian accessions than most of the Wickham clones except GT1 which appeared little different from the other Wickham clones analyzed.

Development of genetic maps is one of the important applications of molecular markers in any species. Genetic maps indicate the position and relative distances between markers along chromosomes . A linkage map allows the selection of markers which are evenly distributed over the genome, thus enhancing the probability of finding markers linked to quantitative trait loci (QTL). Molecular markers that are linked to these QTL will co-segregate with the genes involved in desirable traits and could be

used efficiently to follow introgressions and accumulation of favourable traits during recombination eventss. In order to facilitate quantitative trait loci (QTL) mapping in rubber, several research groups have constructed genetic linkage maps in rubber from both intraspecific and interspecific crosses. Lespinasse *et al.,* (2000) developed the first rubber tree genetic map from an inter-specific cross between *H. brasiliensis* and *H. benthamiana.* Another map based on SSR markers was constructed by Feng *et al,* (2010). A linkage map exclusively based on RAPD markers was constructed by Novalina and Sagala (2013). A linkage map making use of RAPD markers, AFLP markers, SSR markers and SNP markers was constructed by Bini (2013). Another map was constructed by Souza *et al.* (2013) using SSR markers to identify QTLs associated with growth-related traits in a full-sib family. To achieve high-density linkage maps, researchers have shifted from anonymous markers such as AFLPs and microsatellites to direct analyses of sequence variations. Advances in high-throughput next generation sequencing technologies have enabled large scale SNP discovery in rubber tree and aided in the development of high density linkage maps (Mantello *et al.*, 2014; Pootakham *et al.*, 2011, 2015 Shearman *et al.*, 2015). The above studies on sequence diversities have shown that SNP frequencies in rubber are comparable to that in other crops. Molecular markers also aid in locating genes governing agronomically important characters through such genetic linkage maps. Major genes can be mapped by establishing association between molecular markers and inherited traits. Mapping of genes can also be done by anchoring large contigs and scaffolds obtained from RNA sequencing experiments in to linkage groups. Adopting this strategy Shearman *et al.* (2015) mapped 20,143 predicted genes from rubber tree which will be useful for further mapping studies.

Recent molecular genetic investigations in rubber indicate that SNPs including insertion/deletions (indels) can provide a rich source of useful molecular markers in genetic analysis in *Hevea.* However, there are only limited reports regarding the identification and utilization of SNP markers in rubber. First report on large scale SNP identification and validation of SNP in *Hevea* was reported by Pootakham *et al.* (2011).

They identified 5883 biallelic SNPs by transcriptome sequencing and 50 of them were validated in *Hevea* clones. Later on, Mantello *et al.* (2014) reported 404,114 SNPs by *de novo* assembly of tanscriptome data of rubber tree generated using NGS. They also validated 78 SNPs in 36 genotypes. By an allele-specific amplification technology Leonardo *et al.* (2014) characterized 23 *Hevea* genotypes using 191 SNPs screened out of 2,191 ingle Nucleotide Variation (SNVs) discovered by RNA sequencing method. In the following year, Shearman *et al.* (2015) identified more than three lakh SNPs from six popular clones and used these SNPs to construct a linkage map for rubber. They reported a frequency of approximately 1 SNP in every 270 nucleotides in the *Hevea* genome. Recently SNP/Indels were also used to investigate yield heterosis of rubber tree (Li *et al.*, 2016).

1.8. SNP marker analysis based on candidate genes (CG)

Marker assisted selection (MAS) facilitate selection of plants carrying genomic regions that are involved in the expression of traits of interests through molecular markers. For successful adoption of a marker aided selection, a tight linkage between marker and the gene of interest is essential. In the context of MAS, DNA markers can trace favourable alleles, which are both dominant and recessive across generations to identify most suitable individual among the segregating progeny based on allelic composition. Molecular markers were employed for marker assisted selection in plants right from the late eighty's itself (Melchinger, 1990; Hospital and Charosset, 1997). The first substantive article on the application of MAS in plant breeding using DNA markers was for soybean cyst nematode (*Heterodera glycines*) resistance (Concibido *et al.*, 1996). Though indirect associations like the above examples are available for linkage between marker and trait of interest, the possibility of association will be much higher if the marker is from the CG itself. The primary intention of identification of markers from CGs is their utilisation for marker assisted selection (MAS) in the concerned crop as it is the most rapid and sensitive technique to screen genotypes having potential desirable traits. Therefore SNP markers from candidate genes are considered as the marker of choice for MAS in any crop plant.

In terms of monogenic traits, development of efficient markers from CG was always easy as there is only one gene associated with the trait. But most often, characters like crop yield, tolerance to abiotic stress etc. are controlled by a set of genes and their collective interactive expression network will yield the desired traits which are known as quantitative traits. In quantitative traits also, when biological function of the gene or genes could be associated with the trait of interest, candidate gene (CG) approach emerge as a viable alternative strategy for effective marker assisted selection (MAS) (Pflieger *et al.,* 2001).The CG approach has been used widely for QTL identification and MAS with success in several crop plants like maize (Byrne *et al.*, 1996), tomatos (Boylan and Quail, 1991), pepper (Lefebvre *et al.*, 1998), Potato (Leister *et al.*, 1998) etc. In tree crops, CG approach was used to identify markers from drought-adaptation and photosynthesis-related gene loci in poplar (Chu *et al.*, 2009), candidate gene for lignin composition in Eucalyptus (Mandrou *et al.*, 2012), resistance gene in Aprocot (Mariette *et al.*, 2015) etc.

In the case of *Hevea,* numerous studies pertaining to the biochemical synthesis of natural rubber have shown that natural rubber synthesis in *Hevea* is a product of the Mevalonate pathway. It is generally believed that the genes involved in this pathway may be associated with the quantity and quality of latex produced. The results from the following studies in *Hevea* strongly support the above argument (Dennis, 1989; Priya, 2005; Cornish and Blakeslee, 2011; Liu *et al.*, 2016; Tang *et al.*, 2013, 2016). Moreover it is anticipated that the functional polymorphisms arising out of structural variability of these genes may influence the quantity of latex produced from rubber (Uthup *et al.*, 2013, 2016). Consequently SNP markers linked to the genes involved in MVA pathway may be developed as efficient molecular markers to screen the best phenotype at very early stages of selection itself.

Until now, very few CGs have been validated for identification of polymorphisms in rubber. Since traits like rubber yield which is of agronomic

importance is polygenic in nature, MAS programs targeting the selection of best clones should not interfere with the breeding efficiency as well. Validated CGs for QTLs will provide very efficient molecular markers since recombinations between markers and QTLs would be absent. In the coming years, it is likely that validated CG will ease the creation of new commercial varieties in rubber. In conclusion, the CG approach is a powerful strategy for identifying and isolating genes controlling rubber biosynthesis in *Hevea* and to establish their association with desirable traits like higher yield and disease tolerance in rubber.

1.9. Yield related candidate genes selected for the present study

Pioneering work on cholesterol biosynthesis by the laboratories of Lynen (Lynen *et al.*, 1958) and Bloch (Bloch *et al.*, 1959; Chaykin *et al.*, 1958) identified the essential enzymes responsible for the synthesis of isopentenyl diphosphate (IPP), the indispensable building block and universal precursor of all isoprenoids. Since IPP is mainly synthesised *via* mevalonate in the cytosol, this pathway was named as the mevalonate pathway. Plant isoprenoids are the most diverse class of natural compounds which comprise around 23,000 substances such as gibberellins, carotenoids, chlorophyll side chains, plastochinone side chains, sesquiterpenes, sterols, brassinosteroids, dolichol and mitochondrial ubiquinone side chains (Nagata *et al.*, 2002; Kasahara *et al.*, 2004). A vast array of key physiological functions is built on the diversity of isoprenoid compounds representing the largest natural product class simply synthesized *via* the isoprene units that are generated from basic molecules generated in glycolysis. Though a large majority of their functions still remains elusive, it is evident that isoprenoids and the distinct pathways they are derived from play multi-faceted, but vital roles in plants. Briefly, they are involved in electron transport and redox chemistry of photosynthetic light harvesting (ubiquinones, plastoquinones; chlorophylls), photo-oxidative protection (carotenoids), protein prenylation, developmental control (phytohormones), as well as interspecies defense. Moreover, novel drug active compound classes of ginkgolides, sesquiterpene lactones (Artemisinin: anti-infective agent in treating malaria), and taxoids (Paclitaxel: cancer treatment) are of isoprenoid origin (Bhangu-Uhlmann, 2011).

Thus, isoprenoids are not only crucial for the plant, but their research is also of eminent biotechnological and economic interest.

In plants, isoprenoid metabolism is a complicated process because the isoprene building units IPP and DMAPP is synthesised in plastids also by an alternate pathway namely the 2-C-methyl-D-erythritol 4-phosphate or MEP pathway (Arigoni *et al.*,1997). Although expression of an enzyme of the DOXP/MEP pathway could be proven in laticifers of *H. brasiliensis*, radiolabeling studies using intermediates of the cytosolic MVA pathway provided evidence that synthesized IPP from MVA pathway is preferentially incorporated into natural rubber (Skilleter and Kekwick, 1971) whereas experiments using [1-13C] 1-deoxy-D-xylulose triacetate, an intermediate of the MEP pathway could not detect any rubber molecules that carry an isotope label (Ko *et al.*, 2003; Sando *et al.*, 2009). Furthermore experimental evidence by Sando *et al.*, (2008a, 2008b) showed that at least one gene for each step in MVA pathway shows abundant expression in latex, which further corroborates the proposition that it is the MVA rather than the MEP pathway which is involved in rubber biosynthesis. Based on the above studies depicting the major role played by the MVA pathway in rubber biosynthesis, our study was focussed on the genes encoding the enzymes involved in MVA pathway rather than MEP pathway. Figure 1.2 shows the schematic representation describing metabolic map leading to natural rubber biosynthesis in plants.

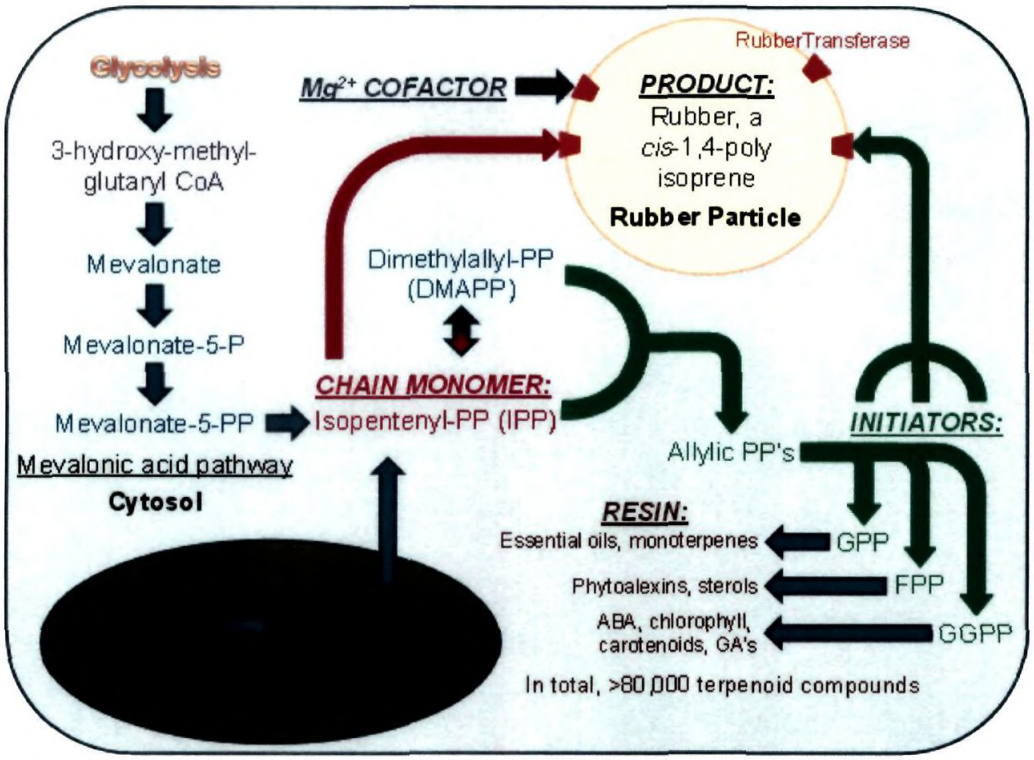


Figure 1.2. Metabolic map of rubber synthesis in plant cells.

IPP (isopentenyl pyrophosphate), the rubber chain monomer essential for chain elongation is synthesized *via* either the mevalonic acid pathway in the cytosol or the MEP pathway in the plastid (rate-limiting metabolic intermediates are shown in purple). IPP and its stereoisomer DMAPP are combined to make several allylic pyrophosphates. GPP (C-10), FPP (C-15) and GGPP (C-30) are able to serve as rubber chain initiators (pathway in green) in various species, and are also the building blocks for several other compounds found in the resin of rubber producing cells (pathway in brown). Rubber synthesis occurs in the unilamellar rubber particle (yellow), requires a Mg^{2+} ion cofactor (grey), and is mediated by the rubber transferase protein() (orange). (Adapted from Cornish and Blakeslee, 2011)

The entire rubber biosynthesis metabolic pathway can be broadly classified in to three stages. Stage A: Upper isoprenoid pathway which includes enzymes involved in the synthesis of IPP from AcetylCoA molecules. Stage B: Middle isoprenoid pathway with enzymes involved in initiation of the polymerisation of IPP units (initiator molecules). Stage C: Lower isoprenoid pathway having enzymes responsible for the chain elongation of the IPP monomer units (elongator molecules).

In the present study eight important genes representing all the three stages of rubbe biosynthesis in *Hevea* were selected as candidate genes for SNP identification Schematic representation of the above classification is shown in **Figure 1.3.**

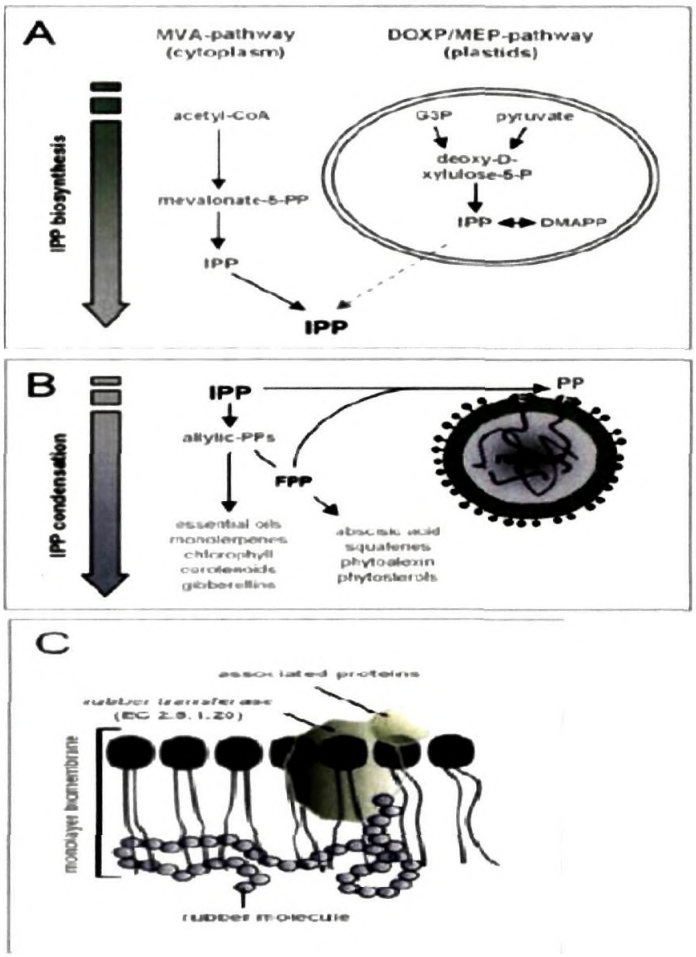


Figure 1.3. Schematic representation of the various stages of the biosynthesis o poly(cis-1,4-isoprene). (A) The monomeric subunit of natural rubber IPP is synthesize by the MVA pathway and the DOXP/MEP pathway in higher plants from acetyl-CoA o glyceraldehyde-3-phosphate and pyruvate, respectively, (B) Chain elongation of IPP i initiated by FPP (FDP) and GGDP for the synthesis of isoprenoids such as allyli diphosphates, side chains of chlorophylls, and natural rubber, (C) Natural rubber i synthesized, regulated and elongated by the activity of rubber particle-associate proteins such as a rubber transferase (*cis*-prenyltransferase, REF and othe proteinaceous factors at the monolayer biomembrane surface of rubber particles (Adapted from Gronover *et al.*, 2011)

1.9.1. Stage A: Upper rubber biosynthesis pathway genes

The major intermediatory molecules synthesised are acetoacetyl CoA, 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA), mevalonate, mevalonate phosphate and mevalonate diphosphate. In the first biochemical reaction of the pathway three Ac-CoA molecules undergo sequential Claison-type and aldol condensations to generate 3-hydroxy-3-methylglutaryl CoA (HMG-CoA), catalyzed by the enzymes acetoacetyl-CoA thiolase (AACT) and hydroxymethyl-glutaryl-CoA synthase (HMGS). The conversion of HMG-CoA into MVA occurs in two successive NADPH-dependent reduction steps via an enzyme-bound intermediate, catalyzed by HMG-CoA reductase (HMGR). Mevalonic acid then undergoes two sequential ATP-dependent phosphorylations catalyzed by mevalonate kinase (MVK) and mevalonate-5-phosphate kinase or phosphomevalonate kinase (PMVK) resulting in the formation of mevalonate-5-diphosphate (MVADP). Decarboxylation of MVADP then generates the C5 building block IPP through the action of mevalonate 5-diphosphate decarboxylase (MDD). Finally, isopentenyl diphosphate isomerase (IDI) is responsible for isomerization of IPP into its allylic isomer DMAPP and vice versa (Okada *et al.*, 2008). Out of the above enzymes, four enzymes (HMGS, HMGR, MVK and PMVK) involved in the synthesis of key intermediates in the pathway was selected for the study anticipating that the genes encoding these proteins may have impact on their expression which ultimately influence the latex yield in rubber. Figure 1.4. represents the steps and enzymes involved in the upper mevalonate pathway.

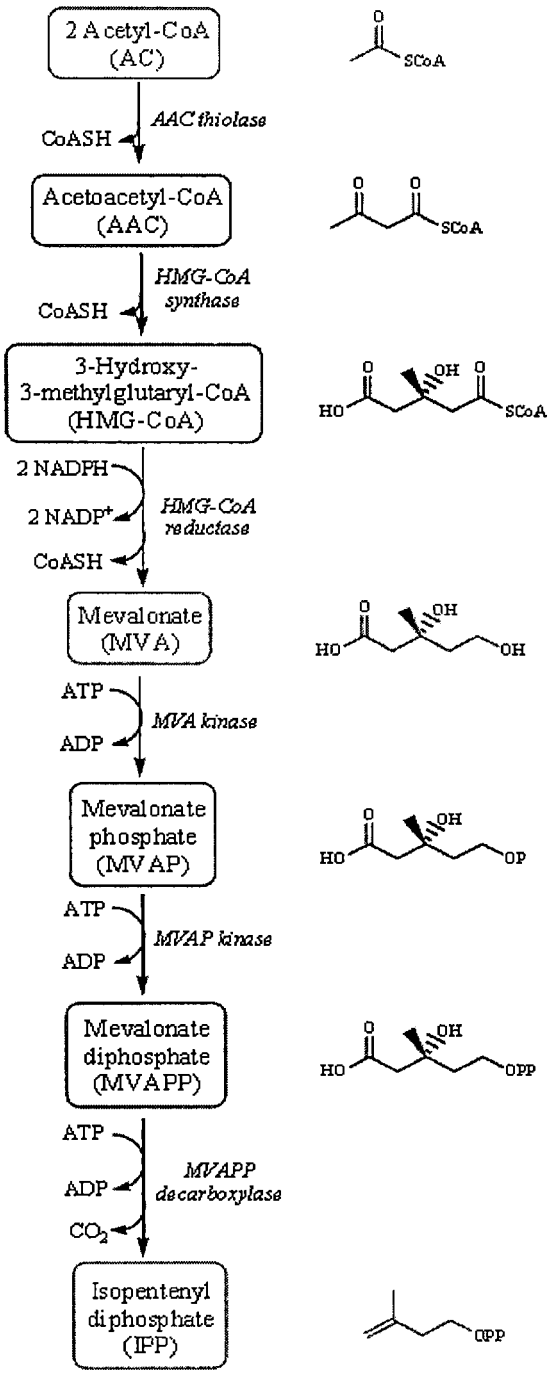


Figure 1.4. Stage A: Upper isoprenoid pathway leading to the biosynthesis of IPP units. Major enzymes like HMGS, HMGR, MVK are shown to involve in this pathway

HMGS: HMGS catalyse the main condensation step of converting Acetoacetyl CoA to HMG-CoA. It has a homodimeric quaternary structure with the overall αβαβα core, conserved among all condensing enzymes and thiolases, reflecting their common evolutionary lineage. All members of this 'condensing-enzyme' super family use a catalytically essential cysteine derived thiolate nucleophile (Kumari *et al.*, 2013). HMGS is one of the two enzymes in the entire pathway for which 3-dimensional structure data is available in plants (Pojer *et al.,* 2006).

HMGR: The first dedicated step in the MVA pathway is the formation of mevalonate which is catalysed by the rate-limiting enzyme HMGR. Due to its highly regulatory and complex nature it is one of the most extensively studied enzymes of the MVA pathway (Hemmerlin *et al*, 2012). The plant HMGRs are smaller than animals and yeast, and they are the only membrane bound enzyme in the path way containing two trans-membrane domains which enables its attachment onto the ER (Vollack *et al.*, 1994).

MVK and PMVK: The crucial final reactions that metabolize Mevalonate to IPP are mainly catalysed by the enzymes MVK and PPVK. They belonging to the GHMP kinase super-family and are believed to have evolved *via* multiple gene duplications followed by neo-functionalization of a common ancestor (Andreassi *et al.*, 2009; Romanowski *et al.*, 2002; Kumari *et al.*, 2013). An important characteristic of MVK enzyme is its sensitivity to feedback inhibition by mevalonate diphosphate and other branch point intermediatory initiator molecules such as Geranyl diphosphate (GDP), Geranylgeranyl diphosphate (GGDP), and Farnesyl diphosphate (FDP) (Andreassi *et al.*, 2004; Gray and Kekwick ,1972). In the case of PMVK, the role of divalent cations like $Mg2^{+}$ in the activity of kinase family members was established by Schulte *et al.* (1999) through their kinetic analysis studies in *Catharanthus roseus.* Furthermore the NMR-based dynamics and chemical shift perturbation measurement studies have shown that PMVK undergoes substantial substrate-induced changes and that the high density of positively charged side chains in the protein plays a role in the charge neutralization

and binding. Sequence analysis of PMVK from various sources reveals that though divergent plant and bacterial sequences shows some similarity, animal proteins exhibits significant divergence in protein folding characteristics (Herdendorf and Miziorko, 2006). Despite this difference, all PMVKs including plant proteins have a high number of arginine and lysine residues and they occur in various spatial clusters all over the surface of the protein (Kumari *et al.*, 2013).

1.9.2. Stage B: Middle rubber biosynthesis pathway genes (initiator molecules)

The second stage in isoprenoid biosynthesis involves the assembly of the C5 building blocks into isoprenoid precursors or prenyl diphosphate homologs having an allylic double bond. The final product of Stage-A is the production of isoprene in the form of two isomers, IPP & DMAPP. The initial condensation of DMAPP, the first allylic diphosphate with one or more IPP molecules in a 'head-to-tail' fashion results in the formation of C10 geranyl diphosphate (GPP or GDP), C15 farnesyl diphosphate (FPP or FDP) and C20 geranylgeranyl diphosphate (GGPP or GGDP), and in each case, the resulting prenyl pyrophosphate product possesses the allylic arrangement required for addition of another IPP molecule, such that, in principle, this repetitive cycle of IPP addition can go on indefinitely (Liang *et al.,* 2002; Kloer *et al*, 2006). Allylic diphosphates such as FDP serve as outlet points for a variety of products. Thus, Stage-B essentially encompasses 'prenylation'or the sequential condensations of IPP with allylic diphosphates to generate isoprenoid carbon skeletons with widely variable chain lengths, from GDP (C10) to natural rubber (C[10,000). GGDPS and FDPS are two major enzymes responsible for these reactions. It could be shown that initiation of rubber biosynthesis is most efficient with FDP in *H. brasiliensis*, *Ficus elastica* and *P. argentatum* in *in-vitro* experiments (Xie *et al.*, 2008). Moreover the sequential condensation of the non-allylic IPP in *cis*-configuration proceeds during rubber biosynthesis at the priming allylic substrate FDP (Cornish, 2001). These observations combined with the finding that FDP is synthesized in the cytosol of *H. brasiliensis* indicate that it is the most likely initiator molecule for rubber biosynthesis (da Costa *et al.*, 2006).

1.9.3. Stage C: Lower rubber biosynthesis pathway genes (elongator- molecules)

cis-prenyltransferase: cis-prenyltransferases (CPTs), are the key downstream enzymes involved in the synthesis of natural rubber which are crucial in determining the molecular weight of the rubber. They elongate a short all '*trans*' precursor, oligoprenyl diphosphate, by sequential addition of the desired number of IPP molecules, which, in turn, results in the formation of a long hydrocarbon skeleton of '*cis*' units (Cornish, 1993a). *cis*-prenyl transferases or rubber transferases have substrate binding constants distinct from other soluble and membrane-bound CPTs and are able to synthesize varying molecular weight *cis*-1,4-polyisoprene. add Rubber transferase is localized to the surface of cytosolic vesicles known as rubber particles, and biosynthesis is initiated through the binding of an allylic pyrophosphate primer (initiators).

REF: Rubber elongation factor (REF) is a 14.6 kDa primary protein, which actually appears to have a role in latex coagulation (a defence mechanism). It is associated with the rubber particle in a ratio of one REF to one rubber molecule (Dennis *et al.*, 1989). Therefore it is expected that REF play a functional role at the interface between the hydrophobic, rubber particle bound, high molecular mass (1000 kDa) *cis*-polyisoprenyl pyrophosphate and the prenyltransferase in the C-serum. Since it is found in a tightly bound state to serum-free rubber particles purified from *H. brasiliensis* latex, REF in coordination with prenyltransferases add multiple cis-isoprene units to rubber molecules. Positive correlation of REF expression with the rubber yield was established by Oh *et al.*, (1999) and Priya *et al*, (2007) . Additional evidence for the abundance of REF transcripts in latex was provided by Chow *et al*, (2007).

Despite of all the information available so far on *Hevea* genetics, genome and rubber biosynthesis with the help of existing tools and methodologies, an efficient MAS strategy for crop improvement in rubber could not be achieved so far. Even though the lack of complete information on the exact mechanism towards the final steps in rubber biosynthesis process in *Hevea* can be partially responsible for this shortcoming, insufficiency of focussed information on the sequence structure of the known genes

involved in rubber biosynthesis in *Hevea* is a major hindrance in attaining the above goal. Therefore a concentrated study on the concerned genes was designed to address the above issue so that it would equip the molecular breeders working on this important tree crop.

1.10. Objectives of the study

An important part of rubber research is concerned with creation and selection of new *Hevea* clones to produce high yield, which are adapted to different agro-climatic conditions. Conventional genetic analysis in *Hevea* is difficult because of its perennial nature, long breeding and selection cycles, and difficulties in raising F_2 progenies. Since the cultivated rubber is having narrow genetic base, development of informative molecular markers for genetic analysis and MAS is comparatively difficult. Genetic mapping of candidate genes is very difficult unless the markers are generated from the gene itself or the marker is in close association with the gene. Therefore detection of SNPs in candidate genes for yield trait will make it an ideal marker for genetic analysis and association studies related to yield and rubber quality. Single nucleotide variations in the coding regions of rubber biosynthesis genes may have a role in the expression characteristics of concerned protein/enzyme also. Therefore, identification of these SNPs, followed by marker generation will be very useful in the characterisation of existing cultivated clones as well as in the screening of wild accessions on the basis of latex yield. Above all the SNPs associated with functionality of a gene can ultimately be used as markers in early selection of potentially high-yielding clones/genotypes in *Hevea* breeding programs.

In the above context the present work was carried out with the following major objectives:

- To identify single nucleotide polymorphisms (SNPs) in rubber biosynthesis genes
- To develop SNP marker from candidate genes for marker assisted selection in rubber

- To understand the inheritance of SNPs and haplotypes in a mapping population for genetic mapping
- To understand the evolutionary relationship of SNPs in mevalonate pathway genes in plants
- To identify putative association between specific rubber-biosynthesis gene haplotypes with their expression which may ultimately lead to the development of SNP markers linked to higher yield in rubber
- To understand the functional significance of non-synonymous SNPs in rubber

Chapter 2

MATERIALS AND METHODS

2.1. Plant materials

The plant materials used for the study includes popular clones of *Hevea brasiliensis* widely cultivated in Southeast Asian rubber growing countries, *Hevea* wild germplasm accessions collected by IRRDB from the Amazonian forests, and other *Hevea* species.

2.1.1. Popular clones used for SNP identification

Five clones popularly cultivated in the Asia pacific region were selected for the SNP identification of rubber biosynthesis genes. They are: RRII 105, RRII 118, RRIM 600, RRIC 52 and GT1 The clones were all good yielders with maximum genetic diversity between them. Data from genetic diversity analysis using molecular markers carried out in the Genome Analysis laboratory of RRII was also considered for the selection of the five clones for SNP identification. Based on their origin they represent India, Sri Lanka, Malaysia and Indonesia. Detailed description of the characteristics of each clone selected for the study is given below.

RRII 105: RRII 105 is a high yielding popular clone developed by the Rubber Research Institute of India (RRII). It is a hybrid of two primary clones Tjir1 and Gl1. The clone is sturdy with average vigour and good branching habit (Nair and George,1969). Bark contains a high number of latex vessel rows. The estimated yield is 2490 kg per ha per annum (Nazeer *et al.*, 1987). The clone is highly susceptible to pink disease and tapping panel dryness (TPD). Fair degree of tolerance to abnormal leaf fall disease was reported under normal conditions (George, 1980).

RRII 118: RRII 118 is a vigorously growing clone developed by the Rubber Research Institute of India by crossing two Sri Lankan primary clones Mil3/2 and Hil 28. It has tall straight trunks with branched and moderately dense canopy. Bark contains medium

number of latex vessel rings. The estimated yield is 1845 kg per ha per annum. The clone shows moderate tolerance to all major diseases (Saraswathyamma *et al.*, 2000).

RRIM 600: RRIM 600 is another high yielding clone evolved by Rubber Research Institute of Malayasia (RRIM) which is grown extensively in all rubber growing countries. Its parents are Tjir 1 and PB 86. They are tall with straight stems and fairly heavy branches. Vigour before tapping is low but girth increment upon tapping is good. Average yield of 2200 kg per ha per annum was observed during large scale trials in India (Marattukalam *et al.*, 1992). The clone exhibits a rising yield trend. Susceptibility to powdery mildew, TPD and wind damage is low but the clone is highly susceptible to Phythothora and Pink disease.

RRIC 52: RRIC 52 is a primary clone evolved by the Rubber Research Institute of Ceylon (RRIC). It is a moderate yielder with yield dependent on water availability though the plants shows better draught tolerance property. Leaves are glossy. It is resistant to Oidium having only 20% defoliation in the epidemic years with an additional advantage of better inherited vigour than the other Sri Lankan clones (Fernando, 1973).

GT1: GT1 is an extensively cultivated primary clone developed at Gondang Tapen Estate in Indonesia. It is one of the oldest primary clone evolved by ortet selection and is still widely cultivated as one of the clone with good timber quality with moderate latex yield. It has an upright stem which is slightly twisted. Estimated yield from commercial holdings in India is around 1400 kg per ha per annum. This clone also shows a rising yield trend with tapping. It shows moderate incidence to pink and ALF disease while TPD and wind damage is mild. The clone is moderately susceptible to powdery mildew. Genetically it is considered to be much diverse from the other popular clones.

In addition to the above five clones another 35 popular clones developed from various rubber growing countries in the world was used for screening in connection with the analysis of selected genes in the study. These clones were not considered for

SNP discovery, but used for genotyping and confirmation of the inheritance pattern of certain SNPs and indels.

2.1.2. Wild accessions and other *Hevea* species

Wild accessions used in the study comprised of 60 representatives of the *Hevea* wild germplasm collection made by International Rubber Research and Development Board (IRRDB) in 1983 from the three provinces of Brazil namely Acre, Mato Grosso and Rondonia. These plants were maintained as germplasm collection at Central Experimental Station, Chethakal farm of the RRII by the Germplasm division. The individual accessions include 24 plants collected from Acre, 18 from Mato Grosso and 18 from Rondonia. Apart from *H. brasiliensis* plants, the following *Hevea* species were also used for screening of the intragenic retro-element in *FDPS* gene *viz. H. benthamiana, H. spruceana, H. nitida, H. camargoana* and *H. pauciflora.* The details of popular clones used in the study is listed in Table 2.1.

2.1.3. Progeny population used for segregation analysis

The mapping population used for the segregation analysis in the present study comprised of 46 full-sib F_1 progenies developed from a controlled interspecific cross between the *Hevea* clones RRII 105 (female parent) and RRII 118 (male parent). The hybridization program was conducted in 1995 as part of the *Hevea* breeding program of RRII (Kavitha K Mydin, personal communication).

2.2. Collection of leaf samples

Tender healthy green leaves of the *Hevea* clones were collected from germplasm division of RRII campus and immediately kept in ice and brought to the laboratory for DNA isolation. Leaves were washed with sterile distilled water supplemented with a few drops of Tween-20 for 2- 3 minutes. Then washed in sterile water for few minutes and blotted dry on Whatman filter paper. The margins of the leaves were removed and 0.5 gram samples were weighed, excluding the midrib and disease spots if any on the lamina.

Table 2.1. List of popular clones used in the study with their parentage and country of origin.

Sl. No.	Country of origin	Clone	Parentage
1	India	RRII 105	Tjir1 x GL1
2	Malaysia	GL 1	primary clone
3	Indonesia	PB5/51	PB 56 x PB 24
4	Indonesia	PB 86	Primary clone
5	Indonesia	PB 217	PB 5/51 x PB 6/9
6	Indonesia	PB 235	PB 5/51 x PB S/78
7	Indonesia	PB 314	RRIM 600 x PB 235
8	Indonesia	PB 28/59	Primary clone
9	Indonesia	PR 107	Primary clone
10	Malaysia	RRIM 600	Tjir1 x PB 86
11	Malaysia	RRIM 703	RRIM 600 x RRIM 500
12	Malaysia	RRIM 701	44/553 x RRIM 501
13	India	RRII 118	Mil3/2 x Hil 28
14	India	RRII 33	Primary clone
15	Sri Lanka	RRIC 52	Primary clone
16	Sri Lanka	RRIC 104	RRIC 52 x Tjir 1
17	Sri Lanka	RRIC 100	RRIC 52 x PB 83
18	Indonesia	Tjir 16	Primary clone
19	Indonesia	BD 10	Primary clone
20	Indonesia	GT1	Primary clone
21	Thailand	KRS 25	Primary clone
22	India	RRII 5	Primary clone
23	Indonesia	PR 255	Tjir 1 x PR 107
24	Indonesia	PR 261	Tjir 1 x PR 107
25	China	SCATC 88/13	RRIM 600 x Pil B 84
26	China	SCATC 93/114	TR 31 45 x HK 3 11
27	China	Haiken 1	Primary clone
28	Indonesia	Tjir 1	Primary clone
29	India	RRII 203	PB 86 x Mil3/2
30	India	RRII 414	RRII 105 x RRIC 100
31	India	RRII 417	RRII 105 x RRIC 100
32	India	RRII 422	RRII 105 x RRIC 100
33	India	RRII 429	RRII 105 x RRIC 100
34	India	RRII 430	RRII 105 x RRIC 100
35	Malaysia	PB 280	Primary clone
36	Malaysia	PB 28/83	Primary clone
37	Malaysia	PB 330	PB 5/51xPB 32/36
38	Malaysia	PB 255	PB 5/51xPB 32/36
39	Malaysia	PB 260	PB 5/51xPB 49
40	Malaysia	PB 311	RRIM 600x PB 235

2.3. Genomic DNA extraction and purification

Extraction and purification of the total genomic DNA were carried out following a modified CTAB method of Doyle and Doyle (1990). Leaf samples (0.5 g) were ground to fine powder in the presence of liquid nitrogen using a pre-chilled mortar and pestle. The powder was homogenized in 5 ml of 2X CTAB buffer preheated to 60 ^{0}C [2% (w/v) CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris–HCl (pH 8)] β-mercaptoethanol [0.3% (v/v)] was added just before use. The entire content was transferred to polypropylene tubes and was incubated in a water bath at 65°C for 45 minutes with slow intermittent mixing by swirling at an interval of 10 minutes. After cooling, equal volume of 24:1 chloroform:isoamyl alcohol was added. Mixed well by gently inverting the tubes to and fro for 10 minutes. Centrifuged at 10,000 rpm for 10 minutes at room temperature and supernatant was transferred to fresh tubes without disturbing the interphase. It was treated with 25 µl RNase (10 mg/ml) and incubated at 37°C for one hour. An equal volume 24:1 chloroform: isoamyl alcohol was added and mixed well by slow inversion for 10 minutes. Spinned at 10,000 rpm for 10 minutes, at room temperature. The aqueous phase was transferred to a fresh tube and extracted with 0.6 volume of ice cold isopropyl alcohol and mixed slowly. The precipitated DNA was spooled and transferred to a new 1.5 ml eppendorf tube containing 1 ml 75% ethanol with a pipette tip. Spinned for 10 minutes at 4 °C at 10,000 rpm. Supernatant was removed and added 1 ml 75% ethanol again. Spinned as above. The supernatant was discarded and the pellet was kept for air drying. The pellet was dissolved in 500 µl sterile water.

2.4. Quantifying DNA concentration and purity

The DNA concentration and the purity were determined by NanoDrop ND-2000C spectrophotometer (Nanodrop Technologies) and stored at 4 °C.

2.5. Gel electrophoresis and photography

The DNA was electrophoresed on a 0.7 % (w/v) agarose gel at 60 V for about 1-2 hours. Lambda DNA/*Eco*R I+*Hin*d III double digest was used as molecular weight

marker. The gels were visualized on UV transilluminator and photographed with Stratagene Eagle Eye II digital gel documentation system.

2.6. Retrieving rubber biosynthesis gene sequences from NCBI database and primer designing

Selection of gene sequences for primer designing in the current study was done based on the previous reports on the expression characteristics of rubber biosynthesis by Sando *et al.*, 2008 and Chow *et al.*, 2012). Only selected sequences of rubber biosynthesis genes with higher expression in the latex of *H.brasiliensis* and/or sequences having high similarity to them were considered for the analysis.

Hydroxy methylglutrayl Co-A synthase gene (*HMGS*): Full length cDNA of *HMGS* with Accession No AB294688.1 was retrieved from NCBI Genbank database. The sequence of 1.8 kb consisting of 5' and 3' UTR regions were used to design three sets of over-lapping primers initially expecting the presence of introns in genomic sequence. Since genomic sequence information was not available then, PCR amplification was initially performed using these primers based on cDNA sequence. Additional extension primers were designed at a later stage after initial sequencing to fill the gaps.

Hydroxy Methylglutrayl Co-A reductase gene (*HMGR*):Full length *HMGR* genomic sequence with the Accession No.X54659 belonging to the clone RRIM 600 was retrieved from the NCBI Genbank database. The sequence was 2.41 kb in length consisting of three introns. Four pairs of overlapping primers were designed for PCR amplification and sequencing.

Mevalonate kinase gene (*MVK*):Complete cds of *MVK* gene from RRIM 600 with Accession No AF429384 was downloaded from NCBI Genbank. The sequence was 1.5 kb consisting of 5' and 3' UTR regions. Three pairs of overlapping primers were designed and synthesised covering the entire gene.

Phospho mevalonate kinase gene (*PMVK*): Complete cds of *PMVK* from RRIM 600 with Accession No AF429385 was downloaded from NCBI Genbank. The sequence was 1.78 kb consisting of 5' and 3' UTR regions. Four pairs of overlapping primers

were designed initially and got synthesised covering the entire gene. Additional primers for sequencing were designed at later stage.

Farnasyl diphosphae synthase gene (*FDPS*): The entire *FDPS* genomic sequence of 4.9 kb was downloaded from NCBI Genbank database with the Accession No.EF593108.1 and eight sets of overlapping primers were designed and synthesised spanning the entire gene

Geranylgeranyldiphosphate synthase (*GGDPS*): The entire *GGDPS* sequence of 1.5 kb including 5' and 3' UTR regions belonging to RRIM 600 were retrieved from NCBI Genbank database with the Accession No. AB294713. Three pairs of overlapping primers were designed and synthesised covering the entire gene.

cis-prenyltransferase gene (*CPT*): Complete cds of *cis*-prenyltransferase from RRIM 600 with Accession No.AB061234 was downloaded from NCBI Genbank database. The sequence was around 1.28 kb consisting of 5' and 3' UTR regions. Three pairs of overlapping primers were designed and synthesised covering the entire gene.

Rubber elongation factor gene (*REF*): Full length genomic region of *REF* gene consisting of 5' and 3' UTR regions belonging to clone RRII 105 was obtained by combining the REF promoter region and partial cds (Accession No.AY712939.1) and the complete cds (Accession No.AY299405). The total sequence region analysed was around 1.73 kb. Three pairs of overlapping primers were designed and got synthesised covering the entire gene.

A list of primers designed for amplification and sequencing of the eight rubber biosynthesis genes was mentioned in Table 2.2.

Table 2.2. List of primers designed for amplification and sequencing of the eight rubber biosynthesis genes

Gene	Primers sequence in 5' to 3' direction
HMGS	HbHMGS-1F: TCATCCTTCTCAGTCTCTCTTTTTC HbHMGS-1R: CATCCCATGAACTGCTCTCA HbHMGS-2F: CACTGACATTGAAGGCGTTG HbHMGS-2R: GCTGGCATTCCTCACAAAGT HbHMGS-3F: TTTGAAGGCAAGCAATTCTCT HbHMGS-3R: GCCTTCAACTTTCCTGCAAC HbHMGS-F1a-F: CAGGGACCCCTCTCCTTAAT HbHMGS-F1a-R: ACCATCATGAGCCTCCAGTG HbHMGS-F1b-F: GCTCAATTTCCACCTGCATC HbHMGS-F1b-R: TGGCAAAGTTACAGCCTCGT HbHMGS-F1c-F: TGCGCTAAATCAGCTTGTTC HbHMGS-F1c-R: CAAGTCCATAGCGTCCATCC HbHMGS2a-R: CATATATTTGGTTGGCAGCTTG
HMGR	HbHMGR:1F: GCTCATCATTCCCTCTTCCTC HbHMGR:1R: CATTTCGTCTTCCTCCGAGA HbHMGR:2F: TACCATTATTGCCGCACCTA HbHMGR:2R: CAGTGCTGCAGCTGAATCTT HbHMGR:3F: ATTTGCGAGGCTCCAAGG HbHMGR:3R: CCTTCTTCACCACCTCTTCCT HbHMGR:4F: AGCCTGCTGCTGTAAATTGG HbHMGR:4R: CTGCAATGGCAGACATCAAG
MVK	HbMVK1-F:GCCAAAGCTGAGTGTTCATTTC HbMVK1-R: ACTACGGTGGCAGGCTTAAAT HbMVK2-F: GCACTCACTTCTGGAGTGTCAG HbMVK2-R: AGGTGACTGGATGATGTTAGCC HbMVK3-F: GCTGGTGTTTCAGAGAGAACCT HbMVK3-R: CCCCAGAAACTTCAAAGAAAGT
PMVK	HbPMVK1-F: CAGACCAAAGAGACACAAGTGC HbPMVK1-R: TCAAGACCTTGCAAGAGTAGCTT HbPMVK2-F: ACAAGCAGTGCAATTTGCTGTA HbPMVK2-R: CTGAGCAGAGGAAAGCACTTCT HbPMVK3-F: CGGCAGTGGATTTGATGTTAG HbPMVK3-R: GATTCAGGCTCTATCGGAACAC HbPMVK4-F: TAGGATCAAGAAATGCCATGC HbPMVK4-R: CCAGTCTTCTTTGTCTGATGATCT HbPMVK-3a-F: GCTGCTTTACTTCATCACCTTG HbPMVK-3a-R: AGCCATTTTGGATTGAGTGG HbPMVK-3b-F:TCCTCGTACACCCACTAGCTCT HbPMVK-3b-R:GTCAGGAATTACTCAGTGGTGAGA HbPMVK-3c-F: GAGTAGTGTCCCCATCATACCA HbPMVK-3c-R: TATACGCGTCCCAATGTTCTTC HbPMVK-3d-F: ACATGGAGAAAGTTGTCAGAGG HbPMVK-3d-R: AACTGAACATACCGGAACACCT continued to next page..

Gene	Primers sequence in 5' to 3' direction
FDPS	HbFDP1-F:ACAACGCCAAGAACAGGAAT HbFDP1-R: CGCAACTCTTATCTTGCGACT HbFDP2-F:CGTATACACATGTTTGTGGGTGT HbFDP2-R: TGCCAAGAAGTTAAAGGATAACAAA HbFDP3-F: GCCTAAACAAGCAAATTCCCTA HbFDP3-R: TTCCTACCTCCAGGCACATT HbFDP4-F:CTGACTCTGGCATGTGCTCT HbFDP4-R: GAATCCTGGGAATGTGATTG HbFDP5-F: TTGAAGGTTGGTCTGATTGC HbFDP5-R: TTGGTGCCATACACATTGCT HbFDP6-F:TGGTTTTCATTTCAAACTGCAT HbFDP6-R :ACCCTGTTTGGATCTGTTCG HbFDP7-F:CCTCAACTTTTCCTGGATTTTC HbFDP7-R:GCTTTCCCATAGTGCTCCTG HbFDP8-F:TGTTGTCTTTGGTCCTTCTTG HbFDP8-R:TGCTACAAGCAGAATAAGCAATC
GGDPS	HbGGP1-F: CACAGGGACGCTATTCAATCC HbGGP1-R: GGCAGCTTCTAAAGCTTGGTTA HbGGP2-F: CAGAATAATCCACCACCCTCTT HbGGP2-R: ATCAACTACTTGGCCAGCAACT HbGGP3-F: CACTCCTAGCATTTGCTTTTGA HbGGP3-R: TGCATTAGTTTTGCCTGTAAGC
Cis pre	HbCis-Pre1-F: CCCCACATCATCTTCTCTTCTC HbCis-Pre1-R: CTCATGAGGTTTCCTTCGAAAA HbCis-Pre2-F: GGCTGGATTTTTAGCTCTTCTG HbCis-Pre2-R: CGAATCAGAACATCAGGATAAGG HbCis-Pre3-F: GAGAACATGGAGTCGTATTCTGG HbCis-Pre3-R: GAGCTCATCTACAGAGAGAGACCA
REF	HbREF1-F: AATCCCCATTCTAAATCGAC HbREF1-R: CTCCCCCTGCTGTTAATTTG HbREF2-F: GCAAACCAAGTTCCTCGTTATC HbREF2-R: GCCCCTATCATCAACCGTATAA HbREF3-F: GCTTCATTTTCCCATTTGAG HbREF3-R: TCAATTGGCCCTTTATTCTC

2.7. PCR amplification and processing of amplified products for sequencing

The PCR reaction was carried out in a 50 µl final volume containing 100 ng of genomic DNA, 0.2 µM each of the forward and reverse primers, 200 µM dNTPs and 1X

Taq DNA polymerase (Advantage Taq, Clontech, USA) along with the buffer supplied. The temperature cycle profiles were as follows. An initial denaturation of 95 ^{0}C for 2 min was followed by 95 ^{0}C for 30 sec, 54 ^{0}C for 30 sec, 68 ^{0}C for 40 sec for a total of 35 cycles. Final extension was given for 4 min at 68 ^{0}C, and hold at 4 ^{0}C. The PCR products were run on 1% agarose gel and the gels were stained with ethidium bromide and viewed using a gel documentation system (Eagle Eye II, Stratagene). The successfully amplified loci were processed for further sequencing reactions (direct PCR product sequencing as well as sequencing after cloning). Multiple PCR reactions were set for each gene to get sufficient products for sequencing. PCR products were purified using Illustra PCR DNA and Gel Band purification Kit (GE Health care). Concentration of PCR products was adjusted to 50 ng/µl using MilliQ water for sequencing. Each locus was sequenced in both directions in each individual to verify sequence variants.

2.8. Cloning and recombinant plasmid isolation for haplotype detection

The gel-purified PCR product was ligated to pGEM–T Vector (Promega, USA). Ligation mix was prepared by taking 2X rapid ligation buffer, pGEM-T Vector (50 ng/µl), T4 DNA ligase (3 U/µl) and purified DNA. All the reagents ware mixed gently and kept at 4^{0}C overnight. *E. coli* DH5α cells were transformed with ligated products and plated on to LB medium supplemented with ampicillin (100 µg/ml) as well as IPTG – X gal for blue - white selection of transformed colonies. Plates were incubated overnight at 37^{0}C. Colony PCR was performed using vector directed primers (T_7 and SP_6) in a 10 µl final volume containing 0.2 µM each of T_7 and SP_6 primers, 200 µM dNTPs and 0.8 units of *Taq* DNA polymerase (GE Healthcare), prescribed PCR buffer and a single white colony.

For plasmid isolation, single positive colony was streaked on LB- ampicillin plate. A single colony from the plate was grown overnight in LB broth supplemented with ampicillin at 37^{0}C. Plasmid was isolated using the kit (Illustra Plasmidprep Mini Spin kit, GE Healthcare). The concentration of isolated plasmid DNA was checked and adjusted to 50 ng/µl for sequencing. Schematic representation of the experimental scheme from sample collection to DNA sequencing is depicted in Figure 2.1.

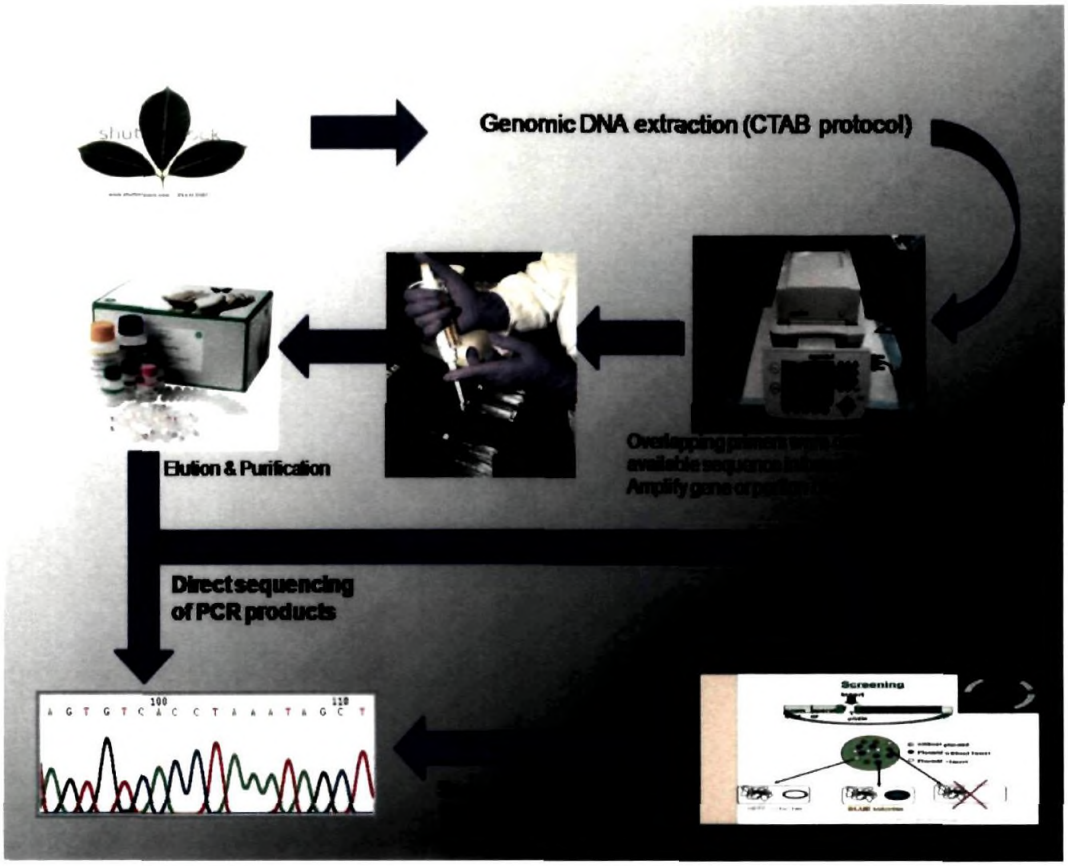


Figure 2.1. Schematic representation of the experimental scheme from sample processing to DNA sequencing.

2.9. Sequence analysis and identification of SNPs

All sequencing reactions were performed using ABI Prism Big Dye Chemistry and were read on ABI 3700 automated sequencer (Applied Biosystems) at the sequencing facility of Macrogen Inc., Republic of Korea. The sequences from the five clones for each respective gene fragment were edited and aligned using DNASISMAX software (Hitachi Solutions America, Ltd). Initial screening for the presence of SNPs and indels was done by comparing direct PCR product sequences and cloned fragment

sequences. Poor quality sequences from the ends of the chromatogram were not considered for SNP detection. Heterozygous loci were identified by locating double peaks in the PCR product sequence chromatogram followed by confirmatory crosschecking with the cloned fragment sequence. Simultaneously each of the chromatograms was visually screened for the identifications of SNPs which were not noticed during the multiple sequence alignment. The SNPs were provided numbers depending on their position from the first base of the primer binding site at 5' end and were named as per the following format. Gene name followed by the base position and alleles [eg. HbGene0000AT].

2.10. SNP genotyping

SNP genotyping was performed by the following techniques: Allele-specific PCR (AS-PCR), Sequence Tagged Site (STS) and HRM. The genotyping techniques were initially standardized with two parental genotypes (RRII 105 and RRII 118) of the progeny population whose alleles were identified by sequencing.

2.10.1. SNP genotyping of popular clones using Locked Nucleic Acid (LNA) primers

AS-PCR was used to genotype the forty popular clones. The allele specific primers were designed manually based on the SNP information from *HMGR* gene. The SNP HbHMGR924AT was selected for the analysis. Two forward primers (LnHMGR-924A-F:5'-GGCATGACAAGAGCGCCTGTTGT+A-3' and LnHMGR-924T-F: 5'-GGCATGACAAGAGCGCCTGTTGT+T-3') were designed in such a way that the last 3' base consisted of each alleles of this SNP (ie., A and T). The last allele specific base was modified in to an LNA base so that the specificity of the reaction was increased several folds. A common reverse primer was used (HbHMGR-2R: 5'-CAGTGCTGCAGCTGAATCTT-3'). Two PCR reactions were performed for each clone (using both forward allele specific primers). Scoring was done in the following manner. If amplification was observed only for LnHbHMGR924A-F , the allele status is "A/A" (homozygous) and if amplicon was observed only for LnHbHMGR924T-F the allele status is "T/T" (homozygous) and if amplification was seen in both the reactions,

the genotype is "A/T" (heterozygous). This simple presence/ absence PCR amplification was performed directly on total genomic DNA. Initial optimization was performed with the five clones where allele status is known from the sequence analysis. In the final screening also these five clones were included as controls to verify the accuracy of the method. For achieving the allele specificity, primers were tested initially with varying annealing temperatures and $MgCl_2$ concentrations. The PCR reactions were performed in a total volume of 20 µl with 0.2 µ m of the allele-specific and common gene specific reverse primer, 10 mM of dNTPs, 0.1 µl of Advantage ® 2 polymerase mix (Clonetech), along with its buffer (10X Advantage 2SA) and 40 ng of genomic DNA. The temperature cycling parameters followed were an initial denaturation at 95^0C for one minute followed by a touch down PCR programme of 10 cycles : 95^0C for 30 seconds, 70^0C for 30 seconds, 72^0C for 45 seconds with, $\Delta\downarrow$ 0.3^0C for 10 cycles. This was followed by a normal cycling of 95^0C for 30 seconds, 67^0C for 30 seconds, 72^0 C for 45 seconds for 35 cycles followed by a final extension at 72^0C for 5 minutes. The PCR products were electrophoresed on a 1.4% agarose gel for 3 hours in 1x TAE buffer at 70 V, stained with ethidium bromide (0.5 µg/ml) and visualized on a UV transilluminator.

2.10.2. SNP genotyping of progeny population using Sequence Tagged Site (STS) marker

An STS marker was developed from the indels existing in the first intron of *FDPS* gene. This marker was used to genotype all the sixty one individuals from the progeny population with the intention of assessing the segregation pattern of intron1 alleles. Genotyping was performed using the primer combination HbFDP2-F: CGTATACACATGTTTGTGGGTGT and HbFDP2-R:TGCCAAGAAG TTAAAGGATAACAAA. The PCR products were checked on 2.5% agarose gel for allele identification and scoring. The allele and genotype frequency was estimated and the contribution to chi-square was calculated from the values obtained. Similarly another STS marker (HbPMVKindel) was developed based on the large indel (1.2 kb) identified from the 9^{th} intron of *PMVK* gene. Like the *FDPS* gene marker, this marker also was screened in the progeny. Genotyping was performed using the primer

combination HbPMVK4-F:AGGATCAAGAAATGCCATGC and HbPMVK-4a-R: CTGTAAGCAGCGAAACATCTTG.

2.10.3. SNP genotyping of progeny population using HRM technique

Genotyping was performed by post PCR analysis method ie., High Resolution Melt (HRM) analysis that measures the dissociation of double-stranded DNA at high temperature, and permits the analysis of genetic variations (SNPs, mutations, methylation) in PCR amplicons (Figure 2.2). The progeny population of 46 individuals were genotyped using SNP markers developed from six latex biosynthesis genes for segregation analysis to assess the mode of inheritance of the haploblock it resides. Since polymorphisms were not detected for the genes *HMGR* and *REF* in the parents (RRII 105 & RRII 118) SNP markers from these genes could not be developed for genotyping the above progeny. Details of the SNPs used in segregation analysis is shown in Table 2.3.

The HRM analysis was performed using Type-it® HRM™ PCR kit (Qiagen, Crawley, UK) following the manufacturer's instructions; 25 ng of genomic DNA of the parents and progenies was added as template to the HRM PCR reaction mix comprising 7.5 µl of 2X HRM PCR Master Mix (including EvaGreen fluorescent dye), 0.7 µM of forward and reverse primers and sterile nuclease free water to 15 µl. HRM analysis was performed using LightCycler 480 (Roche) with PCR steps comprised of an initial PCR activation step at 95°C (5 min), followed by 45 cycles of 95°C (10 sec), 55°C annealing (30 sec) and 72°C extension (10 sec). Both the parents were considered as melting standard for the data analysis by included them in the reaction in triplicate. List of gene-specific primers used for HRM genotyping is shown in Table 2.4.

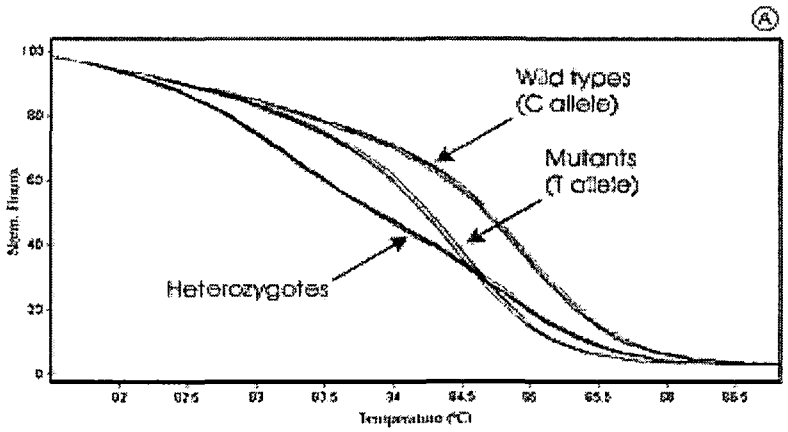


Figure 2.2. The temperature shift pattern of the two SNP alleles (C& T) plotted against the fluorescence intensities. All the three genotypes (C/C homozygous, T/T homozygous, C/T heterozygous) have different melting temperature in their respective duplexes which is reflected in the graph.

Table 2.3. Details of SNP markers with their respective genes and segregation type codes as per JoinMap program.

Sl No	Gene	SNP marker	SNP Alleles	Parental genotypes RRII105 X RRII118	Join Map Segregation type
1	HMGS	HbHMGS3059AG	A/G	AG X GG	<lm x ll>
2	FDPS	HbFDPS1380CT	C/T	CC X CT	<nnxnp>
3	cis-prenyltransferase	Hbcis-pre-1438CT	C/T	CT X TT	<lm x ll>
4	GGDPS	HbGGDP741CT	C/T	CC X CT	<nnxnp>
5	MVK	HbMVK2628AG	A/G	AG X GG	<lm x ll>
6	PMVK	HbPMVK102CT	C/T	CTX TT	<lm x ll>
7	PMVK	HbPMVK1786CT	C/T	CT X CC	<lm x ll>

Table 2.4. List of gene-specific primers used for HRM genotyping

Gene	Locus /position	SNP alieles	Primers in 5' to 3' direction	Amplicon size (bp)
HMGS	3059	A/G	hrm HMGS-A3059G-F:TTAGGTTGTGGACGGCAAACTTTCT hrm HMGS-A3059G-R:AGAGTAACATAGTAAGCATCTGCAC	205
FDPS	1380	C/T	hrmFDP- C1380T-F:ATATGGATCGAGGATCTGCTGG hrmFDP- C1380T-R:AAGAATACTGATTATGCACGGCA	160
cis-pre	1438	C/T	hrmcis-pre-C1438T-F:CCTCTTTGCTAGCAGGCTGTTC hrmcis-pre-C1438T-R:GAGCTCATCTACAGAGAGAGACCA	156
GGDPS	741	C/T	hrmGGP-C741T-F:GATTCATACTATGTCTCTCATCCA hrmGGP-C741T-R:AGCAAATGCTAGGAGTGCGTCA	220
MVK	2628	A/G	hrm-MVK-A2628G-F:CAAATGGGCTCTCGAAGGTGAAA Hrm-MVK-A2628G-R:GCAGAGATAAAGCAGGTGGAGTACA	166
PMVK	102	C/T	hrm-PMVK-C102T-F CTTGCCTGTAACTCTCTGCCG hrm-PMVK-C102T-R:GATACTGGAAAAGAAGATTTGAGAGAG	170
PMVK	1786	C/T	hrm-PMVK-C1786T-F:CTTTCAACTTCGTGTGAGTGGATG hrm-PMVK-C1786T-R:GAATGTCTCGAAGACCACCTAATTGA	140

After the PCR step, the amplicons were subjected to HRM step where the melt temperature range was set at 65°C-95.4°C with continuous fluorescence data acquisition mode (25 acquisition /sec with a ramp rate of 0.02°C/s. This step was followed by a cooling step of 40 °C. The Light Scanner Data Analysis software (Version 2.0, Idaho Technology) was used to analyse the data and to produce normalised dissociation curves and difference plots. Genotype data is obtained from the result output section which provides allelic status of each locus.

2.11. Phylogenetic analysis

Phylogenetic trees were constructed using conceptually translated protein sequences downloaded from NCBI data base to understand the sequence diversity and evolutionary relationship of respective genes. One or two representative sequences from each organism belonging to major domains of life other than plants, like, archae, eubacteria, fungi, arthropods, fishes, birds and mammals etc were taken for the analysis. In the case of plants, HMGS sequences from related as well as distant

genera/family/species were considered for a relationship study with *Hevea brasiliensis* sequences. The sequences were aligned and a tree was constructed using ClustalW.

2.12. Haplotype reconstruction and analysis

The SNP data generated from each of the eight genes from five clones was analysed using DNA polymorphism analysis softwares like DnaSP, Haploview, SNiPLAY and DARwin v.5.

2.12.1. Haplotype prediction by DnaSP

DnaSP, DNA sequence polymorphism analysis tool, is an interactive computer program for the analysis of DNA polymorphism from nucleotide sequence data (Librado *et al.*, 2009). The program, addressed to molecular population geneticists, calculates several measures of DNA sequence variation within and between populations (with or without the sliding window method) in noncoding, synonymous or nonsynonymous sites. Linkage disequilibrium, recombination, gene flow and gene conversion parameters and some neutrality tests, Fu and Li's, Hudson, Kreitman and Aguadé's, McDonald and Kreitman and Tajima's tests could be carried out using this software.

After identification and confirmation of the SNPs, they were analysed using the DNA sequence analysis module of DnaSP for haplotype identification. The aligned sequences containing the SNPs were formatted as per the software requirement. The SNPs and their allele status was marked and read as unphased (or genotype) data files (diploid individuals) in FASTA format. The IUPAC nucleotide ambiguity codes were used to represent heterozygous sites by DnaSP. The haplotype phases from unphased data were reconstructed by DnaSP using the algorithms provided by PHASE, fastPHASE and HAPAR modules. A coalescent-based Bayesian method was used by the software to infer the haplotypes. Haplotypes were generated clicking the "generate haplotype data file" link. A pure parsimony approach was used to estimate the haplotypes and the optimal solution which requires less haplotypes to resolve the

genotypes was selected. It was also used to estimate the recombination rate and prediction of recombination hot spots within each gene sequence.

2.12.2. Haplotype analysis by Haploview

Haploview is bioinformatics software which is designed to analyze and visualize patterns of linkage disequilibrium (LD) in genetic data (Barrett et al., 2005). Haploview can also perform association studies, choosing tag SNPs and estimating haplotype frequencies. It was used to construct haplotype blocks from the SNP data and to reconfirm the haplotype sequences obtained form DnaSP.

2.12.3. Haplotype analysis by SNiPLAY v.2

SNiPlay is a web based tool for SNP analysis like SNP detection, haplotype reconstruction, diversity analysis and linkage disequilibrium (Dereeper *et al.*, 2011). This software was also used to cross check the results from DnaSP and to plot Venn diagrams to understand the distribution status of the identified SNPs from each gene.

2.12.4. Haplotype analysis by DARwin v.5

The software DARwin v.5 (Perrier *et al.*, 2003) was used to calculate dissimilarity from the haplotype data and to build a phylogenetic tree applying the Unweighted Neighbour-Joining method to visualize genetic relationships among the detected haplotyes. The significance of each node was evaluated by bootstrapping data for 1000 replications of the original matrix.

2.13. Integration of latex biosynthesis genes in linkage map using JoinMap

Mapping data were obtained by visual scoring of the banding pattern segregation by the STS marker. Genotyping data from the HRM analysis was directly taken from the gene scanning software output file. Segregation ratio was calculated based on the number of parental alleles for both the markers. All scorings were performed according to the standard coding systems described in JoinMap 4.0. Scored data available from the

earlier mapping study (Bini, 2013) for all the old markers following JoinMap format was used. As per this format the scoring lm x ll was done for the heterozygous marker present only in the female parent and nn x np for the heterozygous marker present only in the male parent. In the case of newly developed markers based on the current SNP data, only 1:1 segregation ratio was expected as per mendelian inheritance mode as the markers were either heterozygous in the female parent or in the male parent. Details of the HRM based SNPs markers used for progeny genotyping, the parental allelic combination and JoinMap scoring format are shown in Table 2.3. Simple text files were used for the JoinMap data file preparation called as 'locus genotype files'. Separate data sets based on the heterozygosity of the markers were organized for each parent: one for the female parent RRII 105 and another for the male parent RRII 118 . A total of 169 markers (164 old + 6 new markers) were utilized for the linkage map construction in RRII 105 and 147 markers (145 old +2 new markers) were utilized for genetic map construction in RRII 118. For RRII 105 out of the five new markers, four were SNP markers and one STS marker (HbPMVK indel). Individual genotype frequency was calculated for all individuals with the available markers., The similarity of loci was determined using the similarity of loci tab-sheet, where we could exclude one of the loci, showing similarity value 1, since identical loci will map exactly at the same position. Kosambi mapping function was used for the calculation of genetic distances in centimorgan (cM). Marker groups were determined using a minimum LOD score of 3.0 and a recombination frequency of 0.4 and the other parameters kept as default.

2.14. Gene specific analysis

Additional screening and analysis was carried out for the genes *HMGS* and *FDPS* in order to understand certain gene specific features observed in them like the presence of a partial retrotransposon sequence in *FDPS* and a putative hypervariable region in the *HMGS* gene.

2.14.1. Sequencing of the putative variable region of *HMGS* gene

The presence of six heterozygous SNPs in a 758 bp region in between the SNPs at base positions 3059 and 3817 in the five clones made it an attractive target to study variability among other *Hevea* clones as well. Since the five genotypes could easily be differentiated by the allele status at just three loci in this region, the possibility of developing this region as a marker for differentiating popular clones was explored by screening additional 14 popular clones. Moreover the results may also help us to understand the distribution of these SNPs and their inheritance pattern. The study also intended to verify whether the SNPs occurring in this region were inherited by reciprocal recombination (crossing over) or by gene conversion (non- reciprocal) which does not follow Mendelian inheritance mode. The region of interest was amplified using the flanking primers HbHMGS-Var-F:5'-CTACCTCATGGCTCTTGATTCC-3' andHbHMGS-Var-R:5'-AGGACTAAGCCCTTATGTTGCAT-3'. The products were eluted separately from the gel and subjected to direct sequencing as well as cloned plasmid sequencing. The 19 clones including the earlier five were broadly classified based on their geographic origin as Indian, Sri-Lankan and Southeast Asian (Malayasia and Indoneasia). DnaSP 4.0 as well as SNiPlay software was used to reconstruct the haplotypes. The software DARwin v.5 was used to calculate dissimilarity from the haplotype data and to build a phylogenetic tree applying the Unweighted Neighbour-Joining method to visualize genetic relationships among the detected haplotyes. The list of 19 clones included in this study is given in Table 2.5.

2.14.2. Screening of popular clones, wild accessions and other *Hevea* species for identifying intron1 alleles of *FDPS* gene

Since hyper variable *FDPS* gene intron1 alleles were detected from the five popular clones analysed initially, the study was extended to additional thirty five popular clones and 60 wild germplasm accessions to locate all the existing allele variants in *Hevea*. *FDPS* intron1 of other *Hevea* species like *H. benthamiana, H. spruceana, H. nitida, H. camargoana* and *H. pauciflora* were also explored. The following primers flanking the hyper-variable region of intron1 were used for

genotyping using the PCR conditions mentioned above (HbFDP2-F: 5'-CGTATACACATGTTTGTGGGTGT-3' and HbFDP2-R: 5'-TGCCAAGAAGT TAAAGGATAACAAA-3'. PCR products were checked on 2.5% agarose gel for discriminating the alleles and their scoring. The sample details and the type of analysis performed is summarised in Table 2.6.

Table 2.5. List of clones used for studying the *HMGS* variable region with their parental details

No	Clone name	Country of Origin	Parentage/Ortet selection
1	RRII 105	Indian	Tjir 1x GI 1
2	Tjir 1	Indonesia	Primary clone
3	GI 1	Malaysia	Primary clone
4	RRII 203	Indian	PB86 x MIL3/2
5	PB 86	Malaysia	Primary clone
6	MIL3/2	Sri Lanka	Primary clone
7	RRIM 600	Malaysia	Tjir 1x PB86
8	RRII 414	Indian	RRII 105 X RRIC 100
9	RRII 422	Indian	RRII 105 X RRIC 100
10	RRII 429	Indian	RRII 105 X RRIC 100
11	RRII 430	Indian	RRII 105 X RRIC 100
12	RRIC 100	Sri Lanka	RRIC 52 X PB83
13	RRIC 104	Sri Lanka	RRIC 52 X Tjir 1
14	PB 314	Malaysia	RRIM 600 X PB 235
15	PB 235	Malaysia	PB 5/51 X PB S/78
16	RRIC 52	Sri Lanka	Primary clone
17	RRII 118	Indian	Mil 3/2 X Hil 28
18	PR 107	Indonesia	Primary clone
19	GT 1	Indonesia	Primary clone

2.14.2.1. Sequence analysis of *FDPS* intron1 alleles

The three different intron1 allele sequences obtained from the initial five clones (FINT1-A & FINT1-C from RRII 118 and FINT1-B from the other four clones) were aligned using DNASIS MAX. Gaps as well as SNPs were identified from the aligned sequences which were later confirmed by checking the chromatograms of respective sequences. Based on the initial screening done in forty popular clones, 60 wild accessions and five *Hevea* species, cloning and sequencing of *FDPS* intron1 region

was carried out from selected clones/accessions. The selection was made based on fragment size variation observed during electrophoresis at the time of initial screening. The additional plants include one popular clone (RRII 5), four wild accessions and four *Hevea* species. *FDPS* hyper-variable intron1 region of all these sequences were compared with the earlier five sequences to detect intron1 specific SNPs, repeats and indels. Phylogenetic analysis of these sequences was done using ClustalW (http://www.genome.jp/tools/clustalw/) to understand their evolutionary relationship. Blastx analysis with NCBI as well as Maize transposable element (TE) database (http://maizetedb.org/~maize/) was performed to detect the presence of transposable elements within the sequences. The first intronic region of the gene from the above plants was also analysed using plantCARE software online (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to identify the putative *cis* regulatory elements.

2.15. Comparative distribution study of non-synonymous SNPs

In-order to identify the differentiation and conserved nature of non-synonymous SNPs in due course of evolution of rubber biosynthesis genes, a comparative study with respect to *Hevea* haplotypes was undertaken using the DNA and protein sequence of representative members of various plant species. The haplotypes were selected based on their unique combination of coding SNPs and were aligned with the conceptually translated amino acid (aa) sequence of plant species included in the phylogenetic analysis to identify the aa changes resulting from the non-synonymous SNPs. The resulting aa change for each coding SNP was traced in other plant species. The aa changes resulting from nucleotide changes other than that present in the *Hevea* haplotypes were not considered for the comparative study

2.16. Latex sample collection for RNA isolation

Fifteen to 20 year old mature tapping tress of the clones RRII 105, RRII 118, RRIM 600 RRIC 52 and GT1 were selected for the study. Latex was collected from two

trees of each clone from CES Chetakal and Malankara estate. The trees were tapped freshly and twenty ml of fresh latex was collected to 50 ml corning tubes containing 20 ml of RNA extraction buffer making up the total volume to 40 ml. The latex was mixed well with the buffer and kept in ice and transported immediately to lab for processing.

From the 40ml, 20 ml was transferred to fresh tube for processing and the rest was preserved in -80°C freezer. 2% beta-ME (beta-mercaptoethanol) was added to the sample extraction buffer mix and mixed well. The samples were centrifuged at 6000 rpm for 10 min to settle the solid particles. After spinning the upper clear phase was transferred to another fresh tube and was extracted two times with equal volume of chloroform isoamyl alcohol (24:1) followed by centrifugation at 7000 rpm for 5 min at RT. The upper phase was recovered and 0.3 w/v of 8M LiCl was added followed by mixing by inverting the tubes several times. The tubes were stored for precipitation at 4°C overnight. The solution was centrifuged next day at 7000 rpm for 20 min at 4°C and the supernatant was discarded and pellet was washed with 2M LiCl followed by centrifugation again at 7000 rpm for 20 min at 4°C. Supernatant was discarded and the pellet was solubilised in 0.5 ml of SSTE buffer. Chloroform isoamyl alcohol extraction was performed once again and the upper phase was transferred to fresh1.5 ml tube. The total RNA was precipitated by adding double the volume of absolute alcohol and kept at -80°C for 30 min to 1 hr. The RNA was pelleted by spinning at 12000 rpm for 20 min and the supernatant was discarded. The pellet was vacuum dried and resuspended in 50-100 µl DEPC treated water. The quality of RNA isolated was checked on 1% agarose gel followed by quantification using NanoDrop spectrophotometer.

2.17. Relative gene expression studies

Gene expression patterns of all the eight rubber biosynthesis genes were studied in five popular clones to identify association between the identified haplotypes of each gene and their expression. Two microgram of total RNA extracted from the latex of two plants of each clone was subjected to DNase treatment using TURBO DNA-free kit (Ambion, USA) and cDNA was synthesized using Superscript III reverse transcriptase (Invitrogen) following the manufacturer's instructions. Relative

quantitative gene expression analysis was carried out using specific primer pairs for each gene on Light Cycler 480 II (Roche Real Time PCR System). List of qPCR primers is provided in Table 2.7. qPCR was performed in a 15 µl reaction mixture. Each PCR reaction was performed in triplicate with no template controls (NTC). Actin was used as endogenous control for the qPCR analysis. The relative quantification (RQ) values were analyzed using Light Cycler 480 Software and the expression rate of these genes is represented as the fold change in transcript level normalized to the actin gene, relative to that in RRII 105 plants.

2.18. *In silico* protein 3D structural modelling analysis

The major aim of the homology modelling study was to verify whether the non-synonymous SNPs identified in *HMGS* and *GGDPS* genes from *Hevea* genotypes leads to any structural modifications in their active binding sites. Since no crystal protein structure was available for both the genes from *Hevea*, molecular modelling approach was adopted. SWISS-MODEL - a fully automated protein structure homology-modelling server was used for constructing the structure for *HMGS* and *GGDPS* (Schwede and Kopp, 2003; Arnold *et al*., 2006; Biasini *et al*., 2014). In SWISS-MODEL the initial template was searched against the swiss-model template library (SMTL - version 2015-02-11) using blast and HHblits methods (Remmert *et al*., 2011). In the case of HMG-CoA synthase, the protein sequence from *Hevea* was extracted with the length of 465 amino acids. Six non-synonymous SNPs were identified in HMG-CoA Synthase gene at position 172 (N -> D), 176 (V -> A), 276 (A -> V), 308 (P -> L), 316 (P -> S) and 383 (T -> A). Protein with these amino acid changes caused by SNP variations at all combinations observed in different haplotypes was subject to molecular modelling in order to identify whether these SNPs forms part of ligand binding region in HMGS. Our protein sequence was modelled using the crystal structure (PDB id: 2FA3) of HMGS from B*rassica juncea* based on template similarity. After successful selection of template, the protein models were built based on the target-template alignment using Promod-II (Guex and Peitsch, 1997). Since the structure complexes with acetyl-CoA, the impact of SNPs on this binding site was examined in detail. The

hydrogen bond distance between acetyl-CoA and binding region was calculated. The structural changes of the modelled structure were analysed using Deep View-spdbv 3.7 (Guex, and Peitsch, 1996) and chimera-1.10.1 (Pettersen *et al.*, 2004). In the case of GGDPS enzyme, protein crystal structure 2J1P (Geranylgeranyl diphosphate synthase from *Sinapis alba*) was selected as the best template for modelling. Following the same approach as mentioned above protein structure of *Hevea* GGPS was modelled and the five non-synonymous SNPs (14 Y, 100 L, 200 A, 287 D, 338 K) were located and highlighted in the modelled structure.

Table 2.6. Samples and the type of analysis performed for *FDPS* gene

Whole gene sequencing (Both alleles)	Intron1 sequencing	Retro mapping
RII 105 RII 118 RIM 600 RIC 52 T1 — popular clones	RRII 105 RRII 118 (both alleles) RRIM 600 RRIC 52 GT1 RRII 5 — popular clones	Forty popular clones [Includes RRII 5 and the five clones shown in column one].
	Acre-9 Acre-19 Rondonia-6 Rondonia-10 — wild accessions	Sixty wild accessions and five *Hevea* species [Includes those given in column two].
	H.benthamiana *H.spruciana* *H.pauciflora* *H.nitida* — *Hevea species*	Progeny population [Sixty one plants]

Table 2.7. Genes specific real time PCR primers used for the relative quantification of eight genes involved in rubber biosynthesis.

Gene	qPCR primer sequence in 5' to 3' direction
HMGS	QHMGS1-F:TCTATGCCCAGAAGGCTGTTG QHMGS1-R:TCCTGGCATGCTACATGACTTCC
HMGR	QHMGR1-F: AACAAAGAGTCGCCAGGATC QHMGR1-R:GCAATGGCAGACATCAAGGA'
MVK	QMVK-F:ATGCCGTTGATTCTATCAGTAATG QMVK-R:CTTCTCAGTTATGGACACATCATCTG
PMVK	QPMVK-F:AGGCATATCATTCAGTCGTAGTTGAT QPMVK-R:TGATTGTCAATTGAGAGTTCCTTTG
FDPS	QFDP-F: TGCAAGCAGTGTTGAAGTCCTT QFDP-R: CGTCATCCAGTCTTTGTCCATGT
GGDPS	QGGP-F:GACCTCCCTTGCATGGATAAC QGGP-R:CACGTCCTCTCCAAACACGAT
CPT	QCis-Prenyl-F:CACGTGGTGTGGGCAGTAATTA QCis-Prenyl-R:TCCGAGTTTATTCCACAAGGATATG
REF	QREF:F:TGTCAAGGACGCATCTATCCAA QREF:R:GTCTGCCCAGGCAAAGAAGA

Chapter 3

RESULTS

3.1. Hydroxy methylglutaryl-CoA synthase (*HMGS*) gene

HMGS catalyse the main condensation step of converting Aceto acetyl CoA to HMGCoA. The HMG-CoA synthase in *H. brasiliensis* has been reported to be present mainly in the C-serum. Previous studies have shown that there exist reasonable correlation between the total activity of HMGS and dried rubber content from each tapping like in the case of HMGR activity. Due to the above correlation it is assumed that that the rubber biosynthetic pathway in *Hevea* is co-ordinately regulated by the activity of both HMGS and HMGR. Due to the significant role of HMGS in the pathway, the sequence structure of the corresponding gene was analysed in detail.

3.1.1. Phylogenetic analysis of HMGS

A phylogenetic tree was constructed using the available HMGS protein sequences of *Hevea brasiliensis*, selected plant species, representative organisms from major domains of life like animals, fishes, birds, bacteria, fungi etc. The phylogenetic tree constructed based on the amino acid sequences clearly depicted the hierarchical linkage of this gene across major kingdoms and divisions Figure 3.1. Clear differentiation based on sequence structure was observed for organisms from different stratas of life. *H. brasiliensis HMGS* gene sequence was clustered along with its orthologs in species like *Ricinus communis, Populus trichocarpa* etc., which formed a sub-group within the major cluster of plant kingdom. Cereals formed another sub-group within this cluster. Bacterial species including archae, proteo and acinetobacter formed separate group clearly depicting the hierarchy of evolution. Fishes, birds and animals along with humans formed another major cluster. Interestingly fungi were placed in the animals group whereas nematodes formed a separate line.

3.1.2. SNP identification in *HMGS* gene

Multiple sequence alignment of the coding regions of *HMGS* sequences from our study with the previously available cDNA sequences of *HMGS1* as well as *HMGS2* showed no major variations except SNPs existing in this region. Twenty bi-allelic SNPs were detected from the 5437 bp sequenced genomic region of *HMGS* from five popular clones resulting in an average of 1 SNP in every 272 bp. Characteristics of the 20 SNPs across the studied popular clones are listed in Table 3.1.

Of the 20 SNPs, 13 were from the non-coding regions and seven were located in coding region. A putative variable region of 754 bp with six highly heterozygous SNPs differentiating all the five clones were identified from base position 3059 to 3817. One SNP from the coding region (HbHMGS4345CT) was synonymous in nature and the remaining six were non-synonymous. The non-synonymous SNPs and the corresponding amino acid changes are given below.

HbHMGS 2167 (GAT Asp acid ⟷ AAT Asparagine)
HbHMGS 2180 (GCT Alanine ⟷ GTT Valine)
HbHMGS 3513 (GCT Alanine ⟷ GTT Valine)
HbHMGS 4052 (CCG Proline ⟷ CTG Leucine)
HbHMGS 4075 (CCT Proline ⟷ TCT Serine)
HbHMGS 4427 (GCA Alanine ⟷ ACA Threonine)

Fourteen SNPs were transition events and and six SNPs were transversions. Nucleotide diversity (Pi) of the studied gene was 0.00119. Out of the 20 SNP loci, 11 were single variable sites while the remaining nine were informative sites.

Based on the geographic region of origin, the clones where grouped in to three as Indian (RRII 105 & RRII 118), Southeast Asian (RRIM 600 & GT1) and Sri Lankan (RRIC 52) to understand the regional distribution of SNPs in *HMGS*. The distribution of 20 SNPs in the five clones is depicted in the Venn diagram (Figure 3.2).

The SNPs were also analysed using the online SNP analysis software SNiPlay v.2 (http://sniplay.southgreen.fr/cgi-bin/analysis.cgi) to characterise individual SNPs and their allelic status. The output is given in Table 3.2.

3.1.3. Haplotype structuring of full-length *HMGS* gene

Haplotype re-construction using DnaSP yielded eight haplotypes. The haplotypes with their allele frequency in each clone is shown in Table 3.3

Three intragenic recombinations were identified between base positions: (2010- 2704) (2704-3059) and (3766-3817).

3.1.4. Sequence analysis of the variable region in *HMGS* gene

The 754 bp highly variable region was comprised of four small exons. The six SNPs falling in this region were HbHMGS3059AG, HbHMGS3513CT, HbHMGS3598CT, HbHMGS3734GT, HbHMGS3766AG and HbHMGS3817GT. Except the non-synonymous SNP HbHMGS3513CT, all these SNPs were existing in the intronic region. No new SNPs were identified from the additional 14 clones sequenced. The allelic status of the six SNPs in 19 popular clones with parental details is shown in the Table 3.4.

As expected the non-synonymous SNP HbHMGS3513CT had the lowest frequency with 17 clones having the dominant allele "C" in homozygous state resulting in the codon GCT (Alanine) instead of GTT (Valine). The SNP data from six loci in these clones were further verified with their parental loci information and all of them showed possible combinations of their parental alleles in all the six loci. Figure 3.3 depicts the chromatogram with the allelic status of the locus HbHMGS3817GT in 400 series hybrid clones developed from the cross between RRII 105 and RRIC 100.

3.1.5. Haplotype structuring of the *HMGS* variable region

The estimated Haplotype diversity and nucleotide diversity using DnaSP was Hd: 0.677 and Pi: 0.00236 respectively. Haplotype re-construction using the PHASE algorithm of DnaSP resulted in 7 types. Table 3.5 shows the grouping of seven haplotypes with the genotypes representing them. Hap_1_Var was observed in 10 clones belonging to all the three geographic regions with a frequency of 26.3%. Hap_2_Var was found to be the most prevalent haplotype which had a frequency of 52.6% (20/38). All the Indian clones had either Hap_1_Var or Hap_2_Var except RRII 118. Hap_4_Var and Hap_5_Var were exclusively observed in three Sri Lankan clones with Hap_2_Var as the common alternate haplotype. Hap_4_Var, Hap_6_Var and

Hap_7_Var present in RRIC 100, RRII 118 and GT1 were noted as the rare haplotypes prevailing among the popular clones. They also share Hap_2_Var as the alternate haplotype. Altogether Sri Lankan clones appeared to be the most diverse clone with the presence of four haplotypes in just four clones. Though nine clones were of Southeast Asian origin, they harboured only four different haplotypes. Interestingly Hap_3_Var was noted exclusively in three clones of Southeast Asian origin and they had either Hap _1_Var or Hap_2_Var as the alternate haplotype.

In the phylogenetic tree (Figure 3.4) constructed based on the seven haplotypes, in the variable region, Hap_3_Var formed a separate cluster evolutionarily distant from the other two clusters. Another cluster comprised of haplotypes: Hap_1_Var, Hap_5_Var and Hap_7_Var and a third cluster had rest of the haplotypes (Hap_2_Var, Hap_4_Var and Hap_6_Var). Similar results were obtained by haplotype analysis using SNiPLAY also. The distribution of the six SNPs in 19 clones based on their geographic origin is shown in Figure 3.5.

3.1.6. Comparative DNA and protein sequence analysis of *HMGS*

Comparative analysis of HMGS protein sequences in diverse plant species with that of *Hevea* revealed that the region consisted of non-synonymous SNPs HbHMGS2167AG and HbHMGS2180CT were highly conserved across different phylum and class. Conversely the remaining four SNPs were harboured in partially conserved regions. Interestingly for the SNP HbHMGS2167AG, the "G" allele resulting in Aspartic acid (Asp) was prevailing in the majority of higher plants whereas the "A" allele leading to Asparagine (Asn) was seen among the members of grass family. For the SNPs HbHMGS2180CT and HbHMGS3513CT the "C" allele resulting in Alanine (Ala) was prevalent than Valine (Val) resulting from the "T" allele. Interestingly the aa Threonine (Thr) due to the presence of allele "A" in SNP HbHMGS4427AG was noted only in *Pinus* apart from *Hevea*. On the contrary, SNPs HbHMGS4052CT (Proline/Leucine) and HbHMGSCT4075 (Proline/Serine) were found to be unique to *Hevea* since these aa changes were not observed in any other species analysed. Additionally, SNP annotation using the inbuilt database of SNiPlay detected the presence of HbHMGS3513CT, HbHMGS4052CT and HbHMGS4427AG

in the *HMGS* gene within chromosome number two of *Vitis vinefera*. HbHMGS3513CT was also identified within the *HMGS* gene in a whole genome scaffold of *Citrus clementae*. The amino acid changes resulting from the six non-synonymous SNPs in *Hevea* haplotypes and selected plant species are schematically represented in Figure 3.6.

3.1.7. *In silico* protein 3D structural modelling of HMGS protein

HMGS from *Brassica juncea* was identified as the best available template for the modelling study. 2FA3 is a crystallized protein structure of HMGS from *Brassica juncea* in complex with acetyl-CoA. Interestingly this structure shared over 85% sequence homology with HMGS from *Hevea*. The acetyl-CoA binding site in 2FA3 protein was identified along with their hydrogen bond distance. The binding site analysis revealed that the amino acids Ser 31, Lys 32, Pro 155, Tyr 151, His 247, Lys 256 and Arg 296 are part of binding cavity. Comparison of each SNP in *HMGS* with respect to its location in the protein revealed that they do not have any direct interaction with the acetyl-CoA binding site and none of them were found inside or near the protein-small molecule interaction interface of HMGS. The 3D protein structure showing the estimated location of six non-synonymous SNPs is depicted in the Figure 3.7.

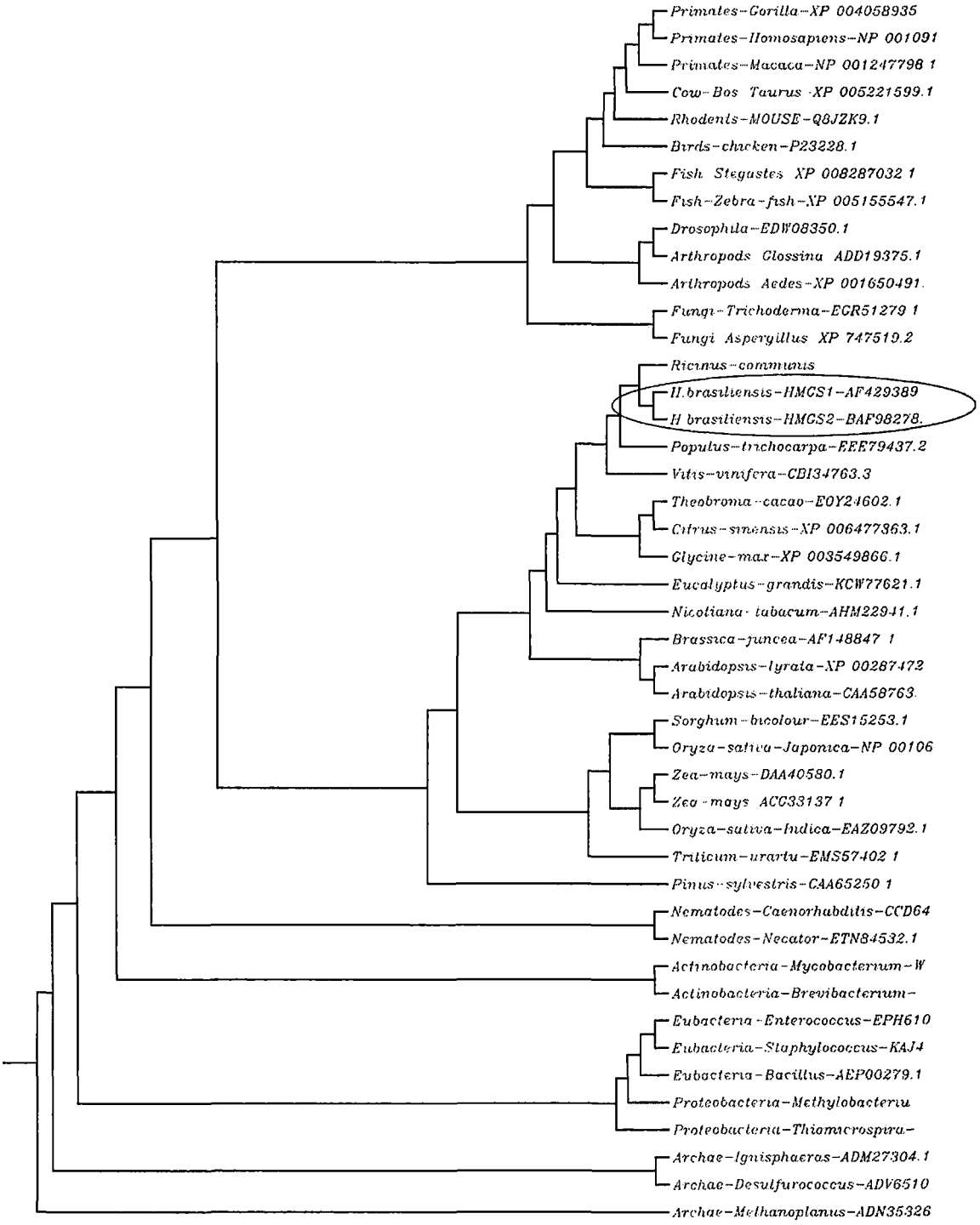


Figure 3.1. HMGS phylogenetic tree

Phylogenetic tree constructed using conceptually translated protein sequences of *HMGS* from selected plant species and representative organisms belonging to major domains of life. The tree depicts the evolutionary relationship of *HMGS* gene across different organisms and the clusters shows the sequence structure variation existing in the *HMGS* gene.

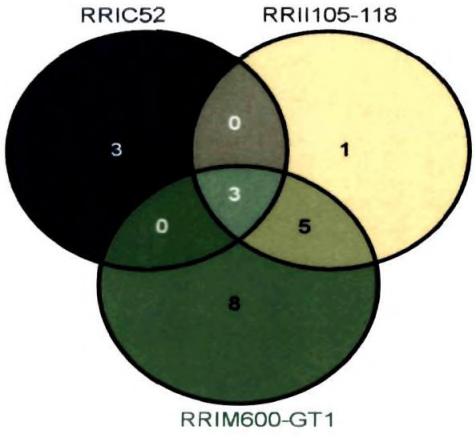


Figure 3.2. Distribution of *HMGS* SNPs across the three groups *viz.* Indian (RRII 105 & RRII 118) Sri Lankan (RRIC 52) and Southeast Asian (RRIM 600 and GT1)

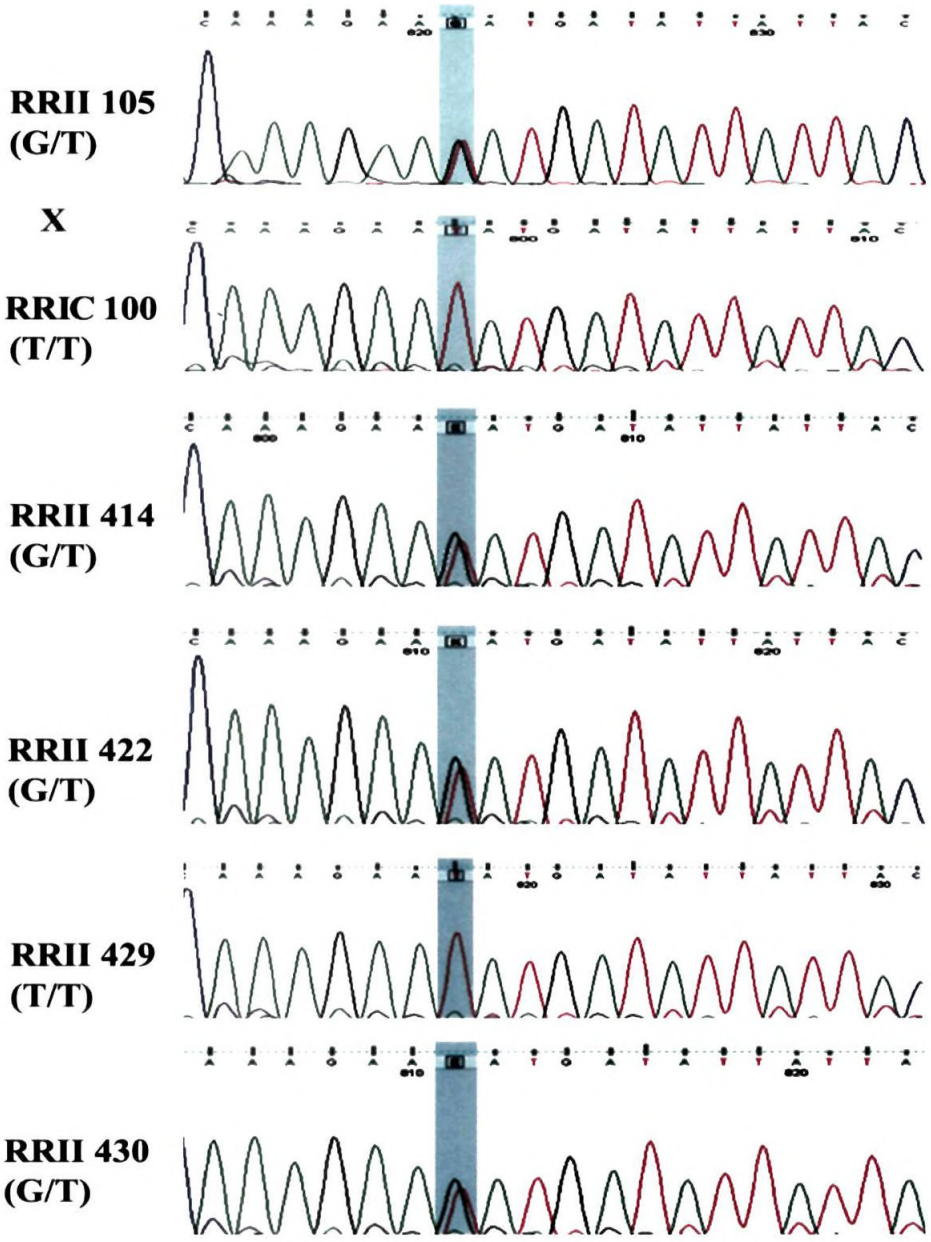


Figure 3.3. Sequence chromatogram depicting the allele status of a parent- hybrid combination: Allelic status of the loci HbHMGS3817(G/T) in the parents RRII 105 x RRIC 100 and their hybrids RRII 414, RRII 422, RRII 429 and RRII 430 are highlighted in the chromatogram. The crossing of GT (RRII 105) x TT (RRIC 100) type resulted in the segregation of alleles to G/T (RRII 414, RRII 422, RRII 430) and T/T (RRII 429) types respectively. Similarly all the six loci were cross verified in all the combinations studied

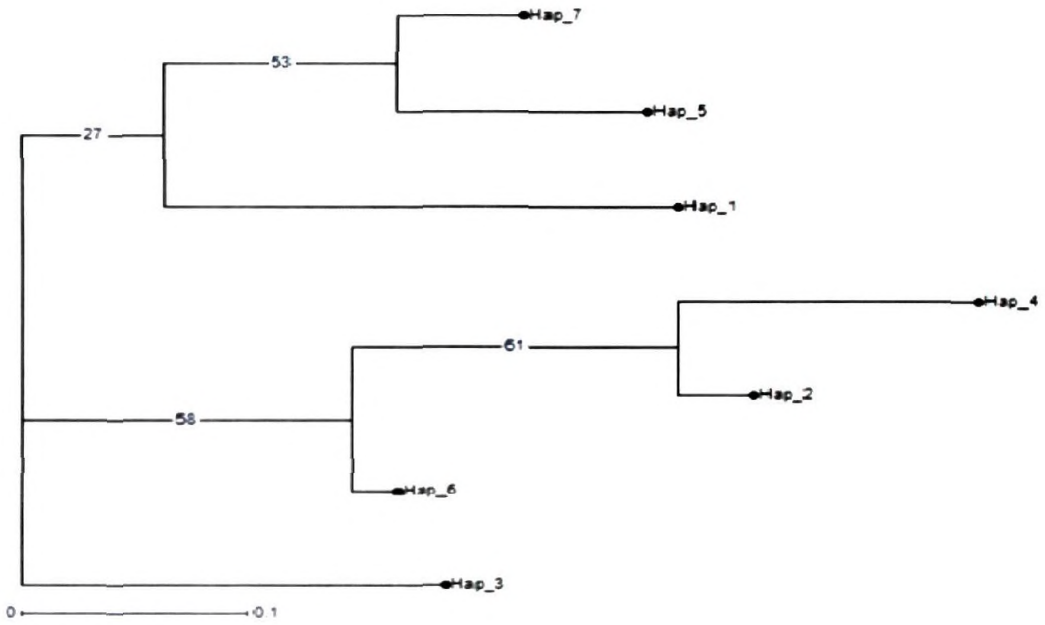


Figure 3.4. Phylogram based on the haplotypes generated exclusively using the six SNPs in the variable region of *HMGS* gene.

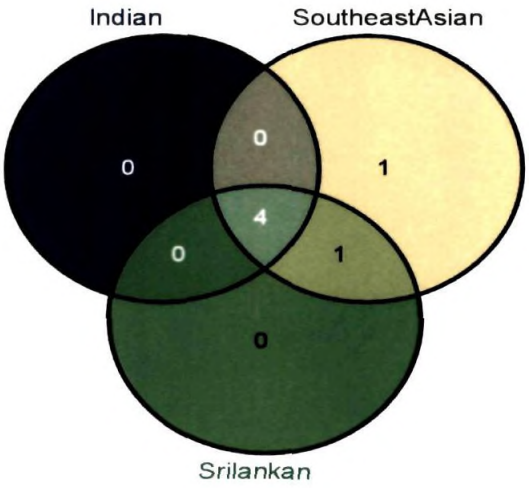


Figure 3.5. Distribution of six SNPs in the variable region of *HMGS* gene. The 19 popular clones were grouped in to three based on their geographic origin as Indian, Sri Lankan and Southeast Asian. The Venn diagram shows the unique as well as common SNPs among these regions.

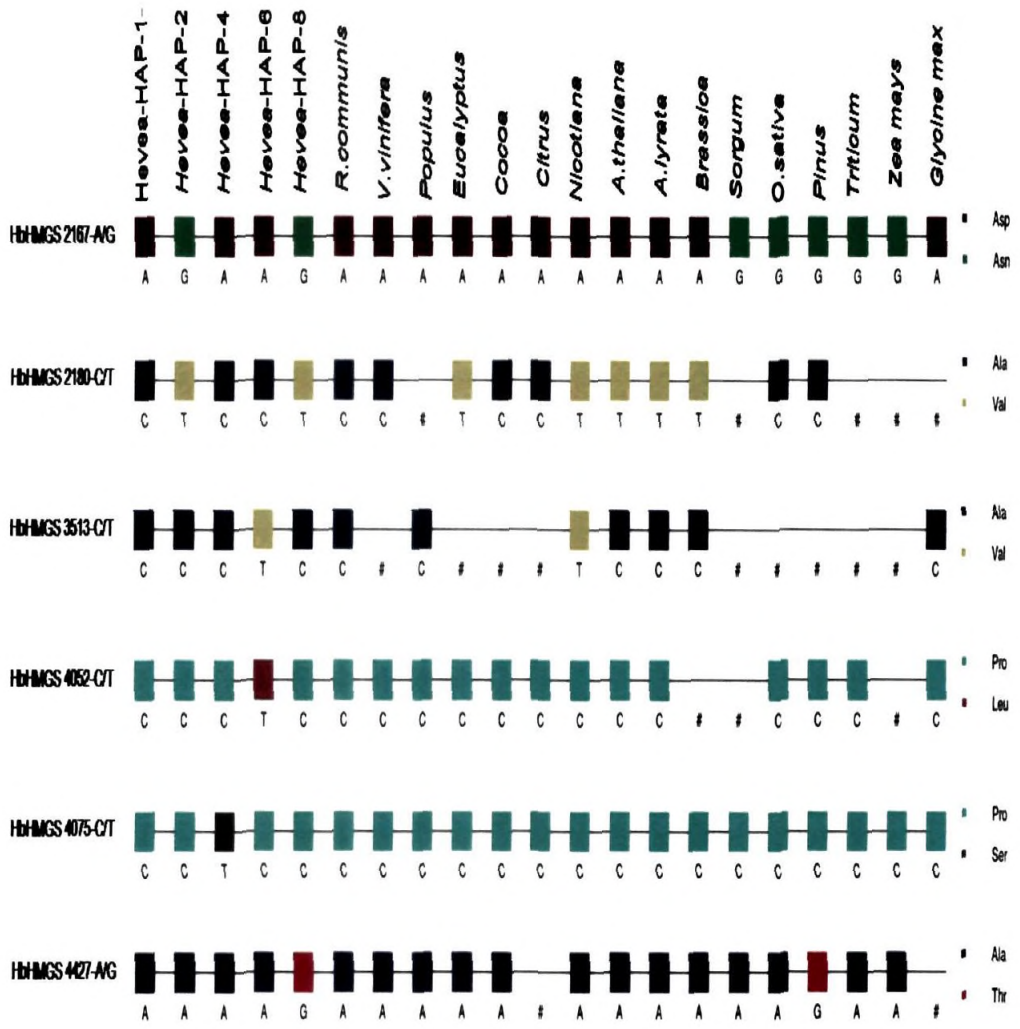


Figure 3.6. Comparative amino acid distribution study based on the non-synonymous SNPs in *HMGS* gene. The amino acid variations due to the codon change induced by the six non-synonymous SNPs in *Hevea* were traced in the *HMGS* gene of selected plant species having similarity to *Hevea* sequences. Changes induced by SNPs not characterised in *Hevea* were excluded from the comparative analysis. The non-synonymous SNP coding regions were traced only in conserved regions. # non-conserved portions and non-synonymous substitutions other than those characterised in *Hevea*. [Amino acids are represented in rectangular coloured boxes for each SNP and their corresponding non-synonymous SNP alleles are given right below the boxes].

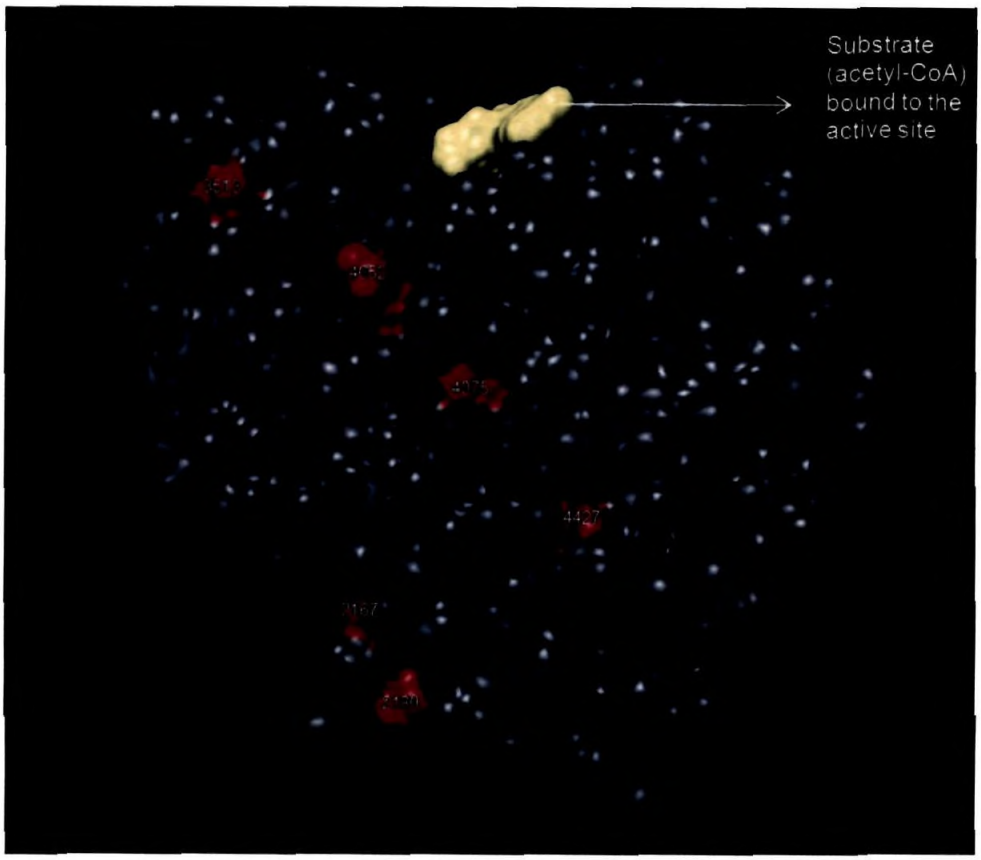


Figure 3.7. Predicted 3D structure of *Hevea* HMG-CoA synthase.
The 3D structure of *Hevea* HMG-CoA synthase bound to its substrate (acetyl-CoA). Amino acid positions which were associated with the non-synonymous substitutions are highlighted. Numbers refer to position of substitutions in the corresponding nucleotide sequence

Table 3.1. SNPs obtained from the full length gene sequence of *HMGS*

Genotype	Respective base position of each SNP locus																		
	2010	2089	2167	2180	2280	2705	2907	2921	2935	3059	3513	3598	3734	3766	3817	4052	4075	4345	4427	4616
RRII105	G/C	T/T	A/G	C/T	C/T	C/C	T/T	A/A	T/T	A/G	C/C	T/T	T/G	A/G	T/G	C/C	C/C	C/C	G/G	A/A
RRII118	G/G	T/T	A/G	C/T	C/T	C/C	T/T	A/G	T/T	G/G	C/C	T/T	G/G	A/A	T/G	C/C	C/C	C/C	G/G	A/A
RRIM600	G/C	T/T	G/G	C/C	T/T	A/C	T/T	A/A	T/T	A/A	C/C	C/T	T/G	A/G	G/G	C/C	C/T	C/T	G/G	A/A
RRIC52	G/G	A/T	G/G	C/C	T/T	A/A	T/T	A/A	T/T	A/G	C/T	T/T	G/G	A/G	T/G	C/T	C/C	C/C	G/G	A/A
GT1	G/G	T/T	A/G	C/T	C/T	C/C	C/T	A/A	A/T	A/G	C/C	T/T	G/G	A/A	T/G	C/C	C/C	C/C	GA	A/T

SNP No	Position consensus	Variation	SNP frequency	Nb readable individual	Major allele	Minor allele	Homo-zygotes major allele	Homo-zygotes minor allele	Hetero-zygotes
1	2010	[C/G]	20.0 %	5	G	C	3	0	2
2	2089	[A/T]	10.0 %	5	T	A	4	0	1
3	2167	[A/G]	30.0 %	5	G	A	2	0	3
4	2180	[T/C]	30.0 %	5	C	T	2	0	3
5	2280	[T/C]	30.0 %	5	T	C	2	0	3
6	2704	[A/C]	30.0 %	5	C	A	3	1	1
7	2906	[T/C]	10.0 %	5	T	C	4	0	1
8	2922	[A/G]	10.0 %	5	A	G	4	0	1
9	2935	[A/T]	10.0 %	5	T	A	4	0	1
10	3059	[A/G]	50.0 %	5	A	G	1	1	3
11	3512	[T/C]	10.0 %	5	C	T	4	0	1
12	3606	[T/C]	10.0 %	5	T	C	4	0	1
13	3733	[T/G]	20.0 %	5	G	T	3	0	2
14	3766	[A/G]	30.0 %	5	A	G	2	0	3
15	3817	[T/G]	40.0 %	5	G	T	1	0	4
16	4050	[T/C]	10.0 %	5	C	T	4	0	1
17	4073	[T/C]	10.0 %	5	C	T	4	0	1
18	4343	[T/C]	10.0 %	5	C	T	4	0	1
19	4425	[A/G]	10.0 %	5	G	A	4	0	1
20	4614	[A/T]	10.0 %	5	A	T	4	0	1

Table 3.3. *HMGS* gene haplotypes with their respective allele frequency in five genotypes

Sl No	Haplotype	Sequence 5' to 3' direction	Haplotype Frequency (%)	Haplotype Frequency (genotypes)
1	HAP_1	CTGCTCTATACTTGGCCCGA	10	[RRII 105-a]
2	HAP_2	GTATCCTATGCTGATCCCGA	20	[RRII 105-b RRII 118-a]
3	HAP_3	GTGCTCTGTGCTGAGCCCGA	10	[RRII 118-b]
4	HAP_4	CTGCTATATACCTGGCTTGA	10	[RRIM 600-a]
5	HAP_5	GTGCTCTATACTGAGCCCGA	20	[RRIM 600-b GT1-b]
6	HAP_6	GTGCTATATATTGGTTCCGA	10	[RRIC 52_a]
7	HAP_7	GAGCTATATGCTGAGCCCGA	10	[RRIC 52_b]
8	HAP_8	GTATCCCAAGCTGATCCCAT	10	[GT1_a]

*"a" and "b" represent the two homologus chromosomes of the diploid *Hevea brasiliensis*

Table 3.4. Allelic status of six SNPs in the putative variable region of *HMGS* gene in 19 popular clones

Genotype No	Parents	3059	3513	3598	3734	3766	3817
RRII 105	Tjir1x GI1	A/G	C/C	T/T	G/T	A/G	G/T
Tjir1	Primary clone	A/G	C/C	T/T	G/T	A/G	T/T
GI1	Primary clone	A/G	C/C	T/T	G/T	A/G	G/T
RRII 203	PB 86 x MIL3/2	A/G	C/C	T/T	G/T	A/G	G/T
PB 86	Primary clone	A/G	C/C	C/T	G/G	A/A	G/T
Mil 3/2	Primary clone	A/G	C/C	T/T	G/T	A/G	G/T
RRIM 600	Tjir1x PB 86	A/A	C/C	C/T	G/T	A/G	G/G
RRII 414	RRII 105 X RRIC 100	A/G	C/C	T/T	G/T	A/G	G/T
RRII 422	RRII 105 X RRIC 100	A/G	C/C	T/T	G/T	A/G	G/T
RRII 429	RRII 105 X RRIC 100	G/G	C/C	T/T	G/G	A/A	T/T
RRII 430	RRII 105 X RRIC 100	A/G	C/C	T/T	G/T	A/G	G/T
RRIC 100	Primary clone	G/G	C/C	T/T	G/T	A/A	T/T
RRIC 104	Tjir1x RRIC 52	A/G	C/T	T/T	G/G	A/G	G/T
PB 314	RRIM 600 X PB 235	A/G	C/C	T/T	G/G	A/A	G/T
PB 235	Primary clone	G/G	C/C	T/T	G/G	A/A	T/T
RRIC 52	Primary clone	A/G	C/T	T/T	G/G	A/G	T/G
RRII 118	Primary clone	G/G	C/C	T/T	G/G	A/A	T/G
PR 107	Primary clone	A/G	C/C	T/T	G/T	A/G	G/T
GT1	Primary clone	A/G	C/T	T/T	G/G	A/A	T/G

Allelic status of six SNPs in the selected clones confirming their pedigree. Since the clones could not be differentiated based on six loci information, the region cannot be developed as a suitable marker region for clone identification.

Table 3.5. Haplotypes re-constructed using the SNPs in the variable region of *HMGS* gene with their frequencies in 19 popular clones.

HAP_1_Var ACTTGG	HAP_2_Var GCTGAT	HAP-3_Var ACCGAG	HAP_4_Var GCTTAT	HAP_5_Var ATTGGG	HAP_6_Var GCTGAG	HAP_7_Var ATTGAG
RRII 105-a RRII 203-a Tjir1-a GL1-a Mil-a RRII 414-a RRII 422-a RRII 430-a RRIM 600-a PR 107-a	RRII 105-b RRII 203-b Tjir1-b GL1-b Mil-b RRII 414-b RRII 422-b RRII 430-b RRII 429-a RRII 429-b PR 107-b RRIC 100-a RRIC 104-a PB 86-a PB 314-a PB 235-a PB 235-b RRIC 52-a RRII 118-a GT1-a	PB 86-b RRIM 600-b PB 314-b	RRIC 100-b	RRIC 104-b RRIC 52-b	RRII 118-b	GT1-b

*"a" and "b" represent the two homologus chromosomes of the diploid *Hevea brasiliensis*

3.2. Hydroxy methylglutaryl-CoA reductase (*HMGR*) gene

HMGR is considered as a rate-limiting enzyme in the MVA pathway due to its involvement in dedicated step of catalysing the formation of mevalonate from HMG Co-A. Due to its highly regulatory and complex nature it is one of the most extensively studied enzymes of the MVA pathway. Because of the significant role played by HMGR in MVA pathway, in-depth analysis of its gene sequence structure in prevailing popular clones is essential to better understand its functional role in NR production in *Hevea*.

3.2.1. Phylogenetic analysis of HMGR

A phylogenetic tree was constructed using the available conceptually translated HMGR protein sequences of Hevea *brasiliensis*, selected plant species and representative organisms from major domains of life like animals, fishes, birds, bacteria, fungi etc. The phylogenetic tree constructed based on the protein sequences clearly depicted the hierarchical linkage of this major gene across kingdoms and divisions (Figure 3.8). Clear differentiation based on sequence structure was observed for organisms from different stratas of life. *H. brasiliensis HMGR* haplotypes were clustered together showing close similarity with their orthologs in species like *Jatropha curcuras, Ricinus communis*, etc., which formed a sub-group within the major cluster of plant kingdom. Cereals formed another sub-group within this cluster. The fossil plant *Ginko biloba* was not clustered along with any of the plants but formed a distinct line within the plant group. *Archae* bacteria, *eubacteria*, arthropods and fungi formed separate clusters clearly depicting the hierarchy of evolution. Fishes, birds and animals along with human formed another major cluster.

3.2.2. SNP identification in *HMGR* gene

Complete genomic sequence of 2560 bp was sequenced from the five *H. brasiliensis* clones. Both the 5'UTR (57 bp) and 3'UTR (170 bp) were included in sequence analysis. The sequenced regions consisted of three introns and four exons. Fourteen bi-allelic SNPs including two indels were identified from the gene resulting in an average of 1 SNP in every 182 bp. Base position of the identified SNPs and their allele status are shown in the Table 3.6. RRII 105, RRII 118 and RRIC 52 showed complete

homozygosity for the SNPs identified. Moreover RRII 105 and RRII 118 shared the same alleles while RRIC 52 alleles were different. RRIM 600 and GT1 were found to be highly heterozygous for *HMGR* gene. Out of the 14 SNPs including two indels, six were from exonic region and eight from intronic region. Furthermore, three SNPs were of transition events and nine were transversions. Nucleotide diversity (Pi) of the studied gene was 0.00268. All the six exonic SNPs (HbHMGR786GC, HbHMGR859GC, HbHMGR888AT, HbHMGR924AT, HbHMGR1149CT, and HbHMGR1320AC) were synonymous in nature. Characteristics of the 14 SNPs obtained from *HMGR* gene is shown in Table 3.7.

3.2.3. Haplotype structuring of full-length *HMGR* gene

Haplotype analysis using all the SNP data by DnaSP confirmed the presence of five haplotypes. Haplotype diversity (He) was: 0.84444 . The haplotypes with their allele frequency in each clone is shown in Table 3.8.

Out of the 14 SNP loci, eight were single variable sites and remaining six were informative sites. Two recombination sites were detected between the base positions (924-1129) (1149-1588).

A phylogram was constructed based on the haplotypes reconstructed using the 14 SNPs in the *HMGR* gene from five genotypes (Figure 3.9). Relationship of the identified haplotypes based on the phylogram generated shows the presence of three distinct groups. Hap_2 and Hap_4 formed one group whereas Hap_1 and Hap_6 formed another. Hap_3 seems to be unique and distinct.

3.2.4. Development of LNA based SNP markers and screening of popular clones

Allele-specific LNA primers designed for differentiating the "A" and "T" alleles at base position 924 of *HMGR* gene was optimized and genotyping of 40 popular clones were carried out. For SNP (A/T) at position 924, 21 popular clones showed homozygous state for allele "A". Only RRIC 52 had homozygous "T" allele, whereas the remaining 18 clones were heterozygous. Figure 3.10 shows the genotyping results of the 40 popular clones using allele specific LNA based primers based on the SNP HbHMGR924AT.

3.2.5. Comparative study of distribution of coding SNPs in *HMGR* gene

Since all the coding SNPs identified in *HMGR* were synonymous in nature, the comparative study was limited to nucleotide sequence alone instead of protein sequence. The six synonymous SNPs in *Hevea* were traced in *HMGR* gene orthologs in selected plant species. SNPs identified only in *Hevea* were considered for the comparative study. The synonymous SNP containing regions were mostly conserved. The result of the comparative DNA sequence study is shown in Table 3.9. From the table it is evident that the SNPs HbHMGR786GC, HbHMGR858GC, HbHMGR924AT and HbHMGR1320AC were widely distributed, whereas the SNP HbHMGR1149CT was found unique to *Hevea* only. Despite the overall conserved nature of this region, two entirely different alleles "A" and "G" were observed in other plant species. Aligned sequences highlighting this portion are shown in Figure 3.11. Similar trend was noted for the SNP HMGR924AT where the allele "A" was observed only in *Hevea* haplotypes.

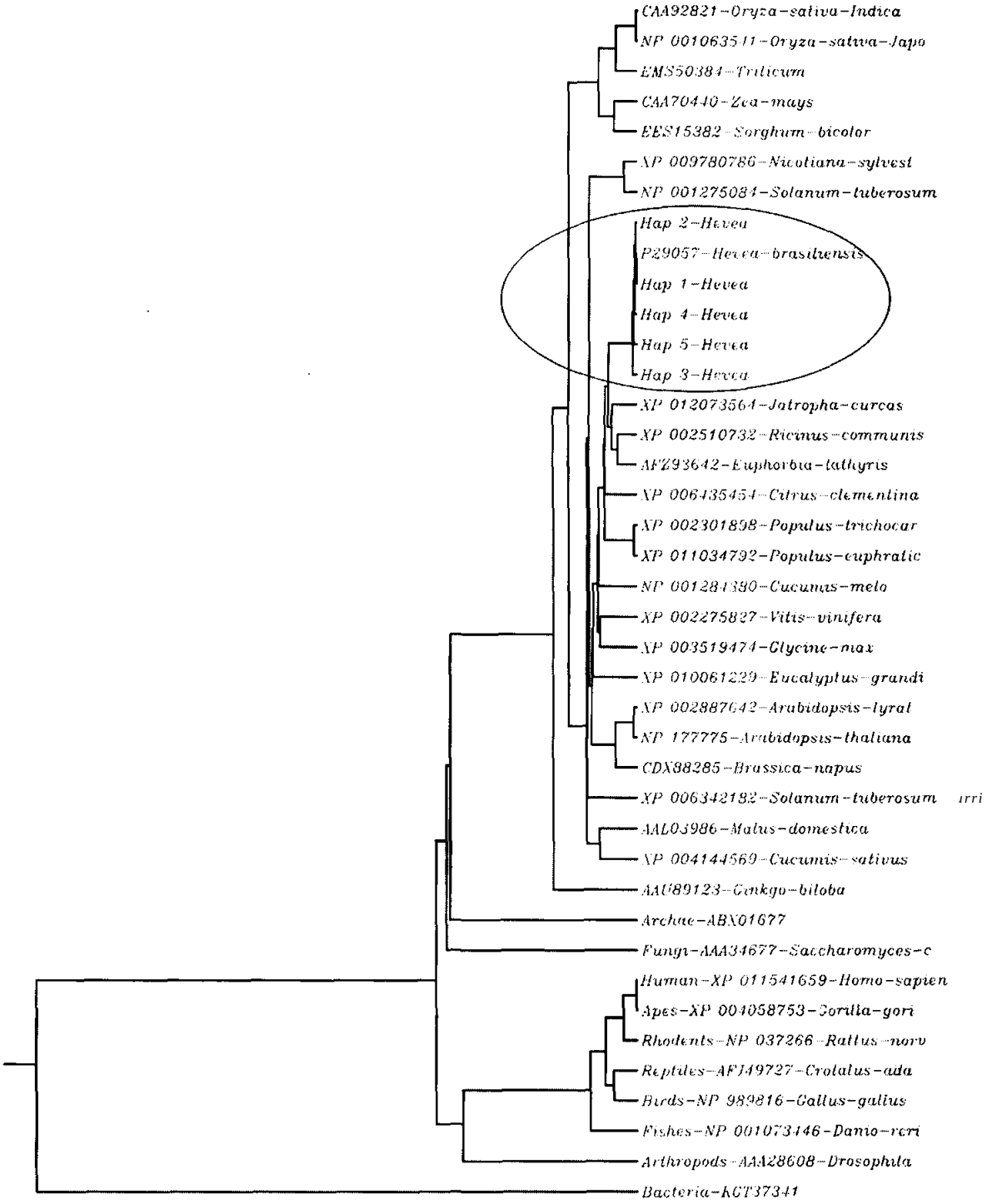


Figure 3.8. Phylogenetic tree of HMGR
Phylogenetic tree constructed using the encoded protein sequences of *HMGR* from selected plant species and organisms belonging to major domains of life. The tree depicts the evolutionary relationship of *HMGR* gene across different organisms.

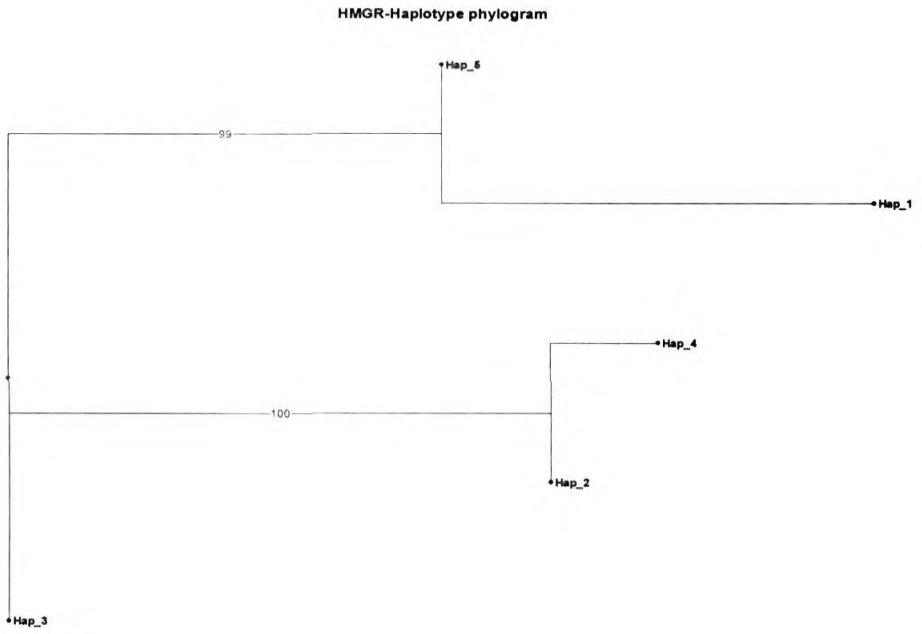


Figure 3.9 Phylogram generated based on the similarity of haplotypes constructed using the fourteen SNPs in the *HMGR* gene.

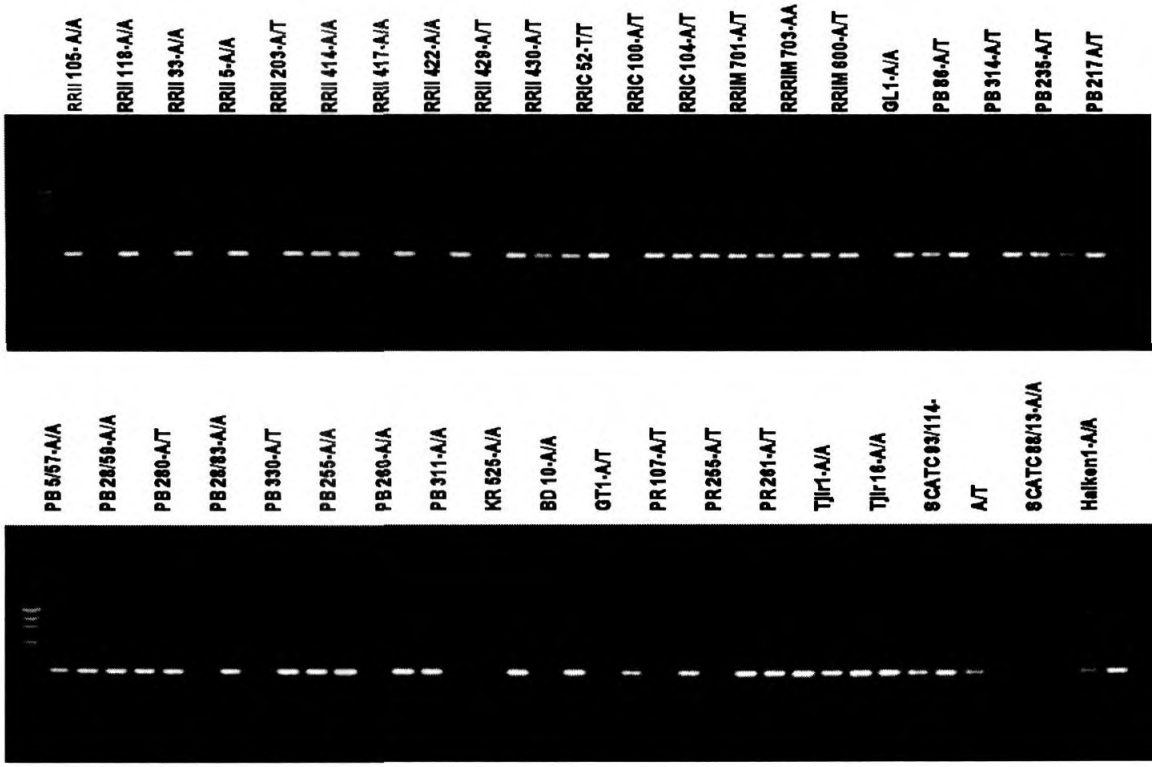


Figure 3.10. Genotyping of 40 popular clones using allele specific LNA based primers based on the SNP HbHMGR924AT. PCR products of reactions amplified using respective A & T allele specific primers were loaded side by side. Name and allelic status of each clone is indicated top of respective lanes.

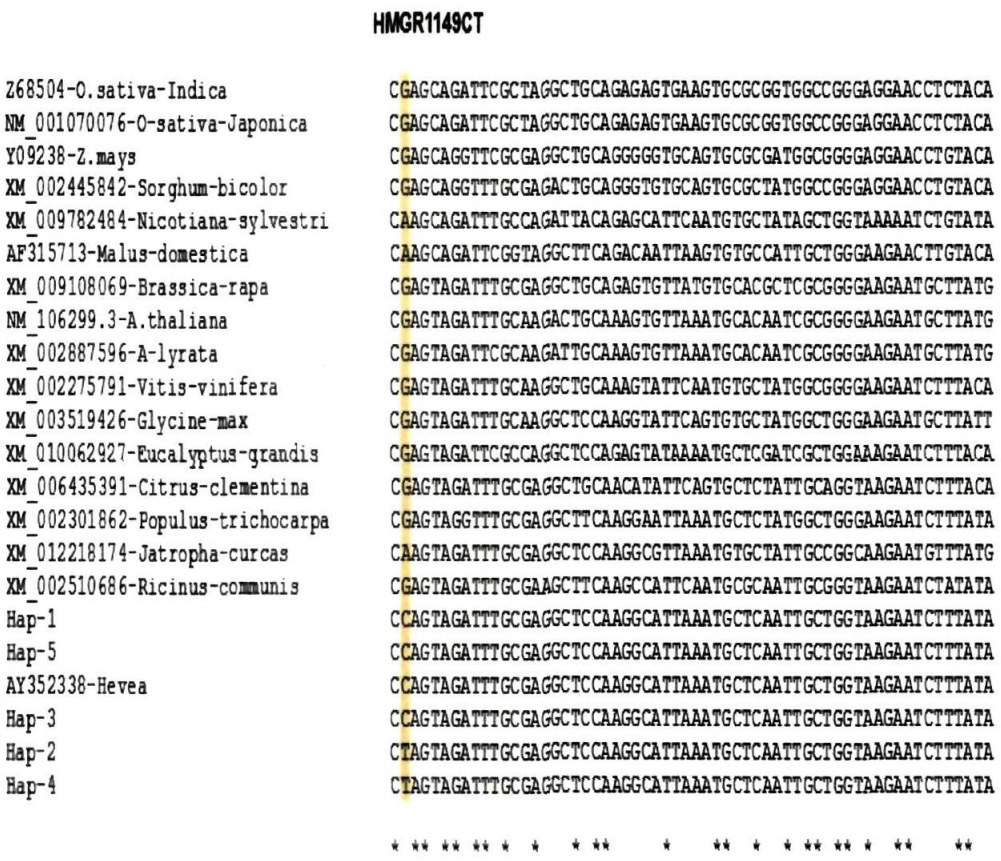


Figure 3.11. Portion of the multiple sequence alignment of *HMGR* gene from *Hevea* haplotypes and related species. The aligned sequences shows the distribution of the SNP HbHMGR1149CT in *Hevea* and other plant species (highlighted).

Table 3.6. SNPs obtained with their allelic status from the full-length gene sequence of *HMGR* gene

Genotype	Respective base position of each SNP locus													
	786	859	888	924	1129	1149	1320	1341	1368	1451	1587	1636	1666	1693
RRII105	G/G	G/G	C/C	A/A	C/C	C/C	A/A	T/T	G/G	T/T	A/A	-/-	T/T	G/G
RRII118	G/G	G/G	C/C	A/A	C/C	C/C	A/A	T/T	G/G	T/T	A/A	-/-	T/T	G/G
RRIM600	G/G	CG	TC	TA	G/C	TC	C/C	-/-	T/T	A/A	T/T	T/T	C/C	T/T
RRIC52	C/C	C/C	T/T	T/T	G/G	T/T	C/C	-/-	T/T	A/A	T/T	T/T	C/C	T/T
GT1	CG	CG	TC	TA	G/C	TC	CA	-/T	TG	AT	T/T	T/T	C/C	T/T

[--] deletion)

Table 3.7. Characteristics of the 14 SNPs in *HMGR* gene

SNP No	Position consensus	Variation	SNP freq	Nb readable individ	Major allele	Minor allele	Homo-zygotes major allele	Homo-zygote minor allele	Hetero-zygote
1	786	[C/G]	30.0 %	5	G	C	3	1	1
2	858	[C/G]	40.0 %	5	G	C	2	1	2
3	888	[T/C]	40.0 %	5	C	T	2	1	2
4	924	[A/T]	40.0 %	5	A	T	2	1	2
5	1129	[C/G]	20.0 %	5	C	G	3	0	2
6	1149	[T/C]	20.0 %	5	C	T	3	0	2
7	1320	[A/C]	30.0 %	5	A	C	3	1	1
8	1341	indel	40.0 %	5	insert	del	3	2	0
9	1369	[T/G]	30.0 %	5	G	T	3	1	1
10	1451	[A/T]	30.0 %	5	T	A	3	1	1
11	1588	[A/T]	40.0 %	5	A	T	3	2	0
12	1638	indel	40.0 %	5	insert	del	3	2	0
13	1665	[T/C]	40.0 %	5	T	C	3	2	0
14	1692	[T/G]	40.0 %	5	G	T	3	2	0

Table 3.8. Haplotypes re-constructed using the SNPs in *HMGR* gene with their respective frequency in five genotypes

Sl No	Haplotype	Sequence 5' to 3' direction	Haplotype Frequency (%)	Haplotype Frequency (genotypes)
1	HAP_1	GGCACCATGTA-TG	40	[RRII 105_a, RRII 105_b, RRII 118_a RRII 118_b
2	HAP_2	GCTTGTC-TATTCT	10	[RRIM 600_a]
3	HAP_3	GGCACCC-TATTCT	10	[RRIM 600-b]
4	HAP_4	CCTTGTC-TATTCT	30	[RRIC 52_a, RRIC 52_b, GT1_a]
5	HAP_5	GGCACCATGTTTCT	10	[GT1-b]

*"a" and "b" represent the two homologus chromosomes of the diploid *Hevea brasiliensis*

Table 3.9. Comparative study of distribution of coding SNPs in *HMGR* gene in rubber and other plant species

Coding SNPs	*Hevea*-HAP-1	*Hevea*-HAP-2	*Hevea*-HAP-3	*Hevea*-HAP-4	*Hevea*-HAP-5	*R.communis*	*V.vinifera*	*Populus tricocarpa*	*Eucalyptus*	*Jatropha curcas*	*Citrus clementia*	*Nicotiana tobaccum*	*A.thaliana*	*A.lyrata*	*Brassica juncea*	*Malus-domestica*	*Glycine max*	*Sorgum*	*O.sativa (indica)*	*O.sativa(Japonica)*	*Zea mays*
HMGR786GC	G	G	G	C	G	G	G	G	G	#	G	G	#	#	#	C	G	C	C	C	C
HMGR858GC	G	C	G	C	G	G	G	#	G	G	#	C	#	C	#	#	C	C	C	C	C
HMGR888CT	C	T	C	T	C	#	#	T	C	#	#	#	C	#	T	#	#	C	C	C	C
HMGR924AT	A	T	A	T	A	T	#	T	#	T	T	#	T	T	#	#	T	T	#	#	#
HMGR1149CT	C	T	C	T	C	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#
HMGR1320AC	A	C	C	C	A	C	C	C	C	C	A	C	A	A	A	A	C	C	C	C	C

The six coding *HMGR* SNPs identified in *Hevea* clones were traced in selected plant species. SNPs identified only in *Hevea* were considered for the comparative study.
non-conserved regions and nucleotide changes other than those detected in *Hevea*

3.3. Mevalonate Kinase (*MVK*) gene

MVK is one of the important enzyme which catalyse the crucial final reactions of converting mevalonate to mevalonate diphosphate. An important characteristic of MVK enzyme is its sensitivity to feedback inhibition by mevalonate diphosphate and other branch point intermediatory initiator molecules such as Geranyl diphosphate (GDP), Geranylgeranyl diphosphate (GGDP), and Farnesyl diphosphate (FDP). Due to its nature as an intermediary enzyme, it is responsible for the synthesis of various secondary metabolites in plants involving complex networking. Therefore it is imperative to understand the DNA sequence structure of the gene responsible for the synthesis of MVK in *Hevea*.

3.3.1. Phylogenetic analysis of MVK

The phylogenetic tree constructed based on the protein sequences clearly depicted the hierarchical linkage of this gene across major kingdoms and divisions Figure 3.12. Clear differentiation based on sequence structure was observed for organisms from different stratas of life similar to what observed in the case of *HMGS* and *HMGR*. The *H. brasiliensis* MVK protein sequence was clustered along with its orthologs in species like *Jatropha curcuras, Ricinus communis*, etc., which formed a sub-group within the major cluster of plant kingdom. Members of grass family like rice , mays etc., formed another sub-group within this cluster. Archae bacteria, eubacteria, arthropods and fungi formed separate clusters clearly depicting the hierarchy of evolution. Fishes, birds and animals along with human formed another major cluster.

3.3.2. SNP identification in *MVK* gene

Twenty six bi-allelic SNPs were detected from the 4.1 kb sequenced genomic region of *MVK* gene from five popular clones resulting in an average of 1 SNP in every 158 bp. Sequence analysis revealed that there are five exons and four introns in the sequenced region. The SNPs, their position and allelic status are showed in Table 3.10.

Majority of SNPs (> 80 %) of *MVK* gene showed homozygosity in three clones: RRII 118, RRIM 600 and GT1. Surprisingly RRII 105 which showed high homozygosity for most of the rubber biosynthesis genes appeared highly heterozygous for *MVK* gene (92 %) (Table 3.11).

Out of the 26 SNPs, eight were from the coding region and the remaining 18 from non-coding region. Eleven SNPs were transversion events and 15 were transitions. Nucleotide diversity (Pi) of the studied gene was 0.00315. Of the eight coding SNPs four were non synonymous in nature and the remaining four synonymous.

The non-synonymous SNPs and the corresponding amino acid changes observed in MVK gene are given below

HbMVK2490(A/G) (GAT Asp acid ⟷ AAT Asparagine)
HbMVK 2518 (A/T) (CAG Glutamine ⟷ CTG Leucine)
HbMVK 2628 (A/G) (ACT Threonine ⟷ GCT Alanine)
HbMVK 2907(A/C) (ATG Methionine ⟷ CTG Leucine)

Interestingly the SNP locus HbMVK2490AG was heterozygous in all the five clones. Rest of the four SNPs in coding region : HbMVK2387AG, HbMVK2420AG, HbMVK2912AT and HbMVK3218AT were synonymous in nature. The distribution of 26 SNPs in the five clones was depicted as Venn diagram (Figure 3.13). Unique SNPs were observed in the Sri Lankan clone RRIC 52 and the Southeast Asian clones RRIM 600 and GT1.

3.3.3. Haplotype structuring of full-length *MVK* gene

Eight haplotypes were determined using DnaSP. The haplotypes with their allele frequency in each clone are shown in Table 3.12. Eight intragenic recombination events were detected. Recombination has been detected between base positions: (71-84), (440-1658), (1774-1857), (1857-2022), (2320-2379), (2379-2387), (2420-2490) and (2490-2518).

A phylogram was generated based on the similarity of haplotypes reconstructed using the 26 SNPs in the *MVK* gene (Figure 3.14). Relationship of the haplotypes shows the presence of three clusters. Haplotypes Hap_2, Hap_4 and Hap_8 formed one cluster whereas Hap_1, Hap_5, Hap_6 and Hap_7 formed another. Hap_3 haplotype appeared to be distinct from the rest forming a separate group.

3.3.4. Comparative study of distribution of non-synonymous SNPs in *MVK*

Comparative analysis of MVK protein sequences from diverse plant species including haplotypes of *Hevea* clones revealed that the region containing non-synonymous SNPs, HbMVK2628AG and HbMVK2907AC were highly conserved across different phylum and classes whereas the SNP locus HbMVK2490AG containing region was partially conserved. Interestingly for the SNP HbMVK2628AG the allele "G" resulting in the amino acid Alanine was observed in members of the family *Fabaceae* (*Glycine max* and *Medicago truncatula*) and *Brassicasia (Arabidopsis lyrata, Arabidopsis thaliana* and *Brassica rapa*) in addition to *Euphorbeceae* family member *Hevea* (Hap_2 and Hap_3).

Similarly for the SNP HbMVK2907AC the allele "A" resulting in the amino acid Methionine was observed only in two *Citrus* species in addition to *Hevea* haplotypes. Same trend was observed in the case of the partially conserved SNP HbMVK2490AG also. Additionally, SNP annotation using the inbuilt database of SNiPlay detected the presence of HbMVK2379AG [intronic], HbMVK2387AG [coding], HbMVK2420AG [coding], HbMVK2490AG [coding], HbMVK2628AG [coding], HbHMGS3218 AT [coding] , and HbMVK3297CT [intronic] in the *MVK* gene in chromosome 14 of *Vitis vinefera.* The amino acid changes resulting from the four non-synonymous SNPs in *Hevea* haplotypes and selected plant species are shown in Table 3.13. The multiple sequence aligned portion of MVK protein sequence showing the amino acid substitution [Threonine (ACA) to Alanine (GCA)] resulting from the non-synonymous SNP HbMVK2628AG is shown in Figure 3.15.

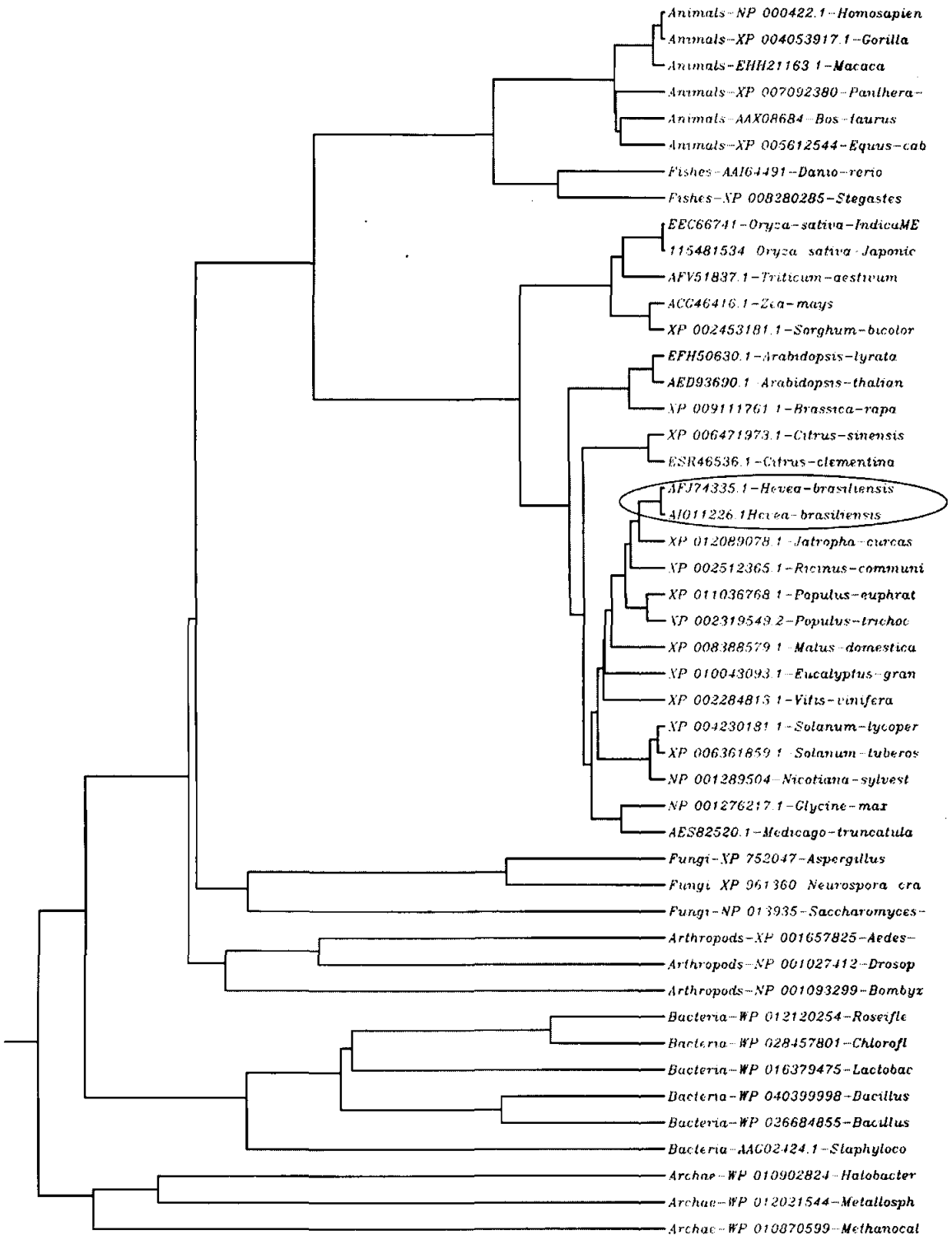


Figure 3.12. Phylogenetic tree of MVK
Phylogenetic tree constructed using the encoded amino acid sequences of *MVK* from selected plant species and representative organisms belonging to major domains of life. The tree depicts the evolutionary relationship of *MVK* gene across different organisms

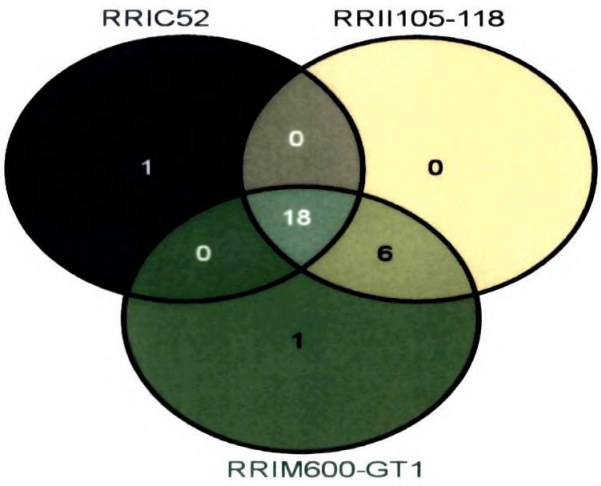


Figure 3.13. Distribution of SNPs in *MVK* gene across the three groups *viz.* Indian (RRII 105 & RRII 118) Sri Lankan (RRIC 52) and Southeast Asian (RRIM 600 and GT1)

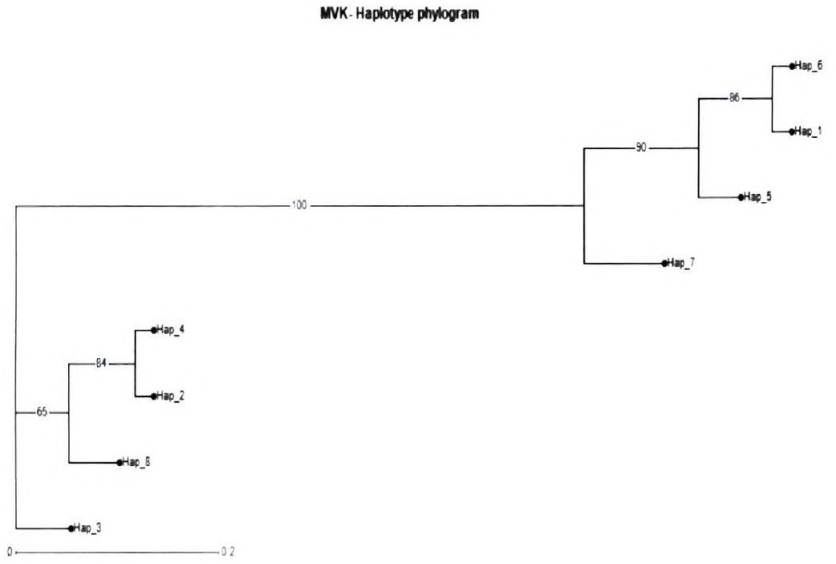


Figure 3.14. Phylogram generated based on the haplotypes consisted of 26 SNPs in *MVK* gene.

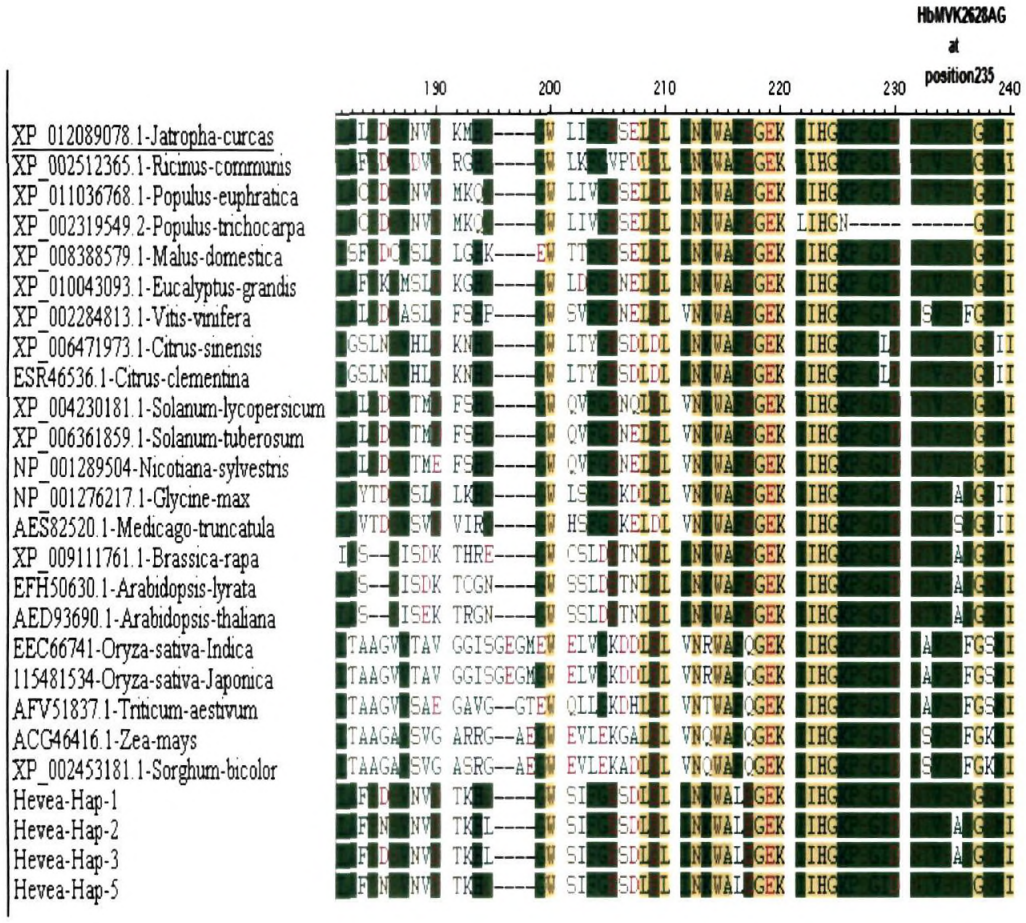


Figure 3.15. Portion of the multiple alignment of MVK protein sequence from *Hevea* haplotypes and related species. The distribution of amino acid substitutions resulting from the non-synonymous SNP HbMVK2628AG in selected plant species is denoted by the position number (position 235).

Table 3.10. SNPs obtained with their allelic status from the full length gene sequence of *MVK* gene

Genotype	Respective base position of each SNP locus																	
	71	84	119	349	441	470	1658	1774	1857	1908	2022	2320	2338	2379	2387	2420	2490	2518
RRII105	C/C	A/G	GT	A/G	GT	A/T	C/T	C/T	C/T	A/G	G/C	A/T	C/C	A/G	A/G	A/G	A/G	A/T
RRII118	C/C	G/G	G/G	G/G	T/T	A/A	C/T	C/T	C/T	A/A	G/G	T/T	C/C	A/G	G/G	A/A	A/G	T/T
RRIM600	A/A	A/A	T/T	A/A	G/G	T/T	C/C	C/C	T/T	G/G	C/C	A/A	C/C	A/G	A/A	G/G	A/G	A/A
RRIC52	A/A	G/G	G/G	G/G	G/G	A/A	C/T	C/T	C/C	A/A	G/C	A/T	C/T	A/G	A/G	A/G	A/G	A/T
GT1	C/C	G/G	G/G	G/G	T/T	A/A	C/T	C/T	C/C	A/A	G/G	T/T	C/C	A/G	G/G	A/A	A/G	T/T

Genotype	Respctive base position of each SNP locus							
	2628	2874	2907	2912	3218	3297	3375	3432
RRII105	A/G	A/C	A/C	A/T	A/T	C/T	C/T	A/G
RRII118	G/G	A/A	C/C	T/T	T/T	C/C	T/T	A/A
RRIM600	A/A	C/C	A/A	A/A	A/A	T/T	C/C	A/G
RRIC52	A/G	A/C	A/C	A/T	A/T	C/T	C/T	A/A
GT1	G/G	A/A	C/C	T/T	T/T	C/C	T/T	A/A

Table 3.11. Characteristics of the 26 SNPs in *MVK* gene.

SNP	Position consensus	Variation	SNP freq	Nb readable individu	Major allele	Minor allele	Homo-zygotes major allele	Homo-zygote minor allele	Hetero-zygote
1	71	[A/C]	40.0 %	5	C	A	3	2	0
2	84	[A/G]	40.0 %	5	G	A	2	1	2
3	119	[T/G]	40.0 %	5	G	T	2	1	2
4	349	[A/G]	30.0 %	5	G	A	3	1	1
5	440	[T/G]	50.0 %	5	T	G	2	2	1
6	469	[A/T]	30.0 %	5	A	T	3	1	1
7	1658	[T/C]	40.0 %	5	C	T	1	0	4
8	1774	[T/C]	40.0 %	5	C	T	1	0	4
9	1857	[T/C]	40.0 %	5	C	T	2	1	2
10	1909	[A/G]	30.0 %	5	A	G	3	1	1
11	2022	[C/G]	40.0 %	5	G	C	2	1	2
12	2320	[A/T]	40.0 %	5	T	A	2	1	2
13	2338	[T/C]	10.0 %	5	C	T	4	0	1
14	2379	[A/G]	50.0 %	5	A	G	0	0	5
15	2387	[A/G]	40.0 %	5	G	A	2	1	2
16	2420	[A/G]	40.0 %	5	A	G	2	1	2
17	2490	[A/G]	50.0 %	5	A	G	0	0	5
18	2518	[A/T]	40.0 %	5	T	A	2	1	2
19	2628	[A/G]	40.0 %	5	G	A	2	1	2
20	2874	[A/C]	40.0 %	5	A	C	2	1	2
21	2907	[A/C]	40.0 %	5	C	A	2	1	2
22	2912	[A/T]	40.0 %	5	T	A	2	1	2
23	3218	[A/T]	40.0 %	5	T	A	2	1	2
24	3297	[T/C]	40.0 %	5	C	T	2	1	2
25	3375	[T/C]	40.0 %	5	T	C	2	1	2
26	3432	[A/G]	20.0 %	5	A	G	3	0	2

Table 3.12. Haplotypes reconstructed using the SNPs in *MVK* gene with their respective allele frequency in each clone.

Sl:No	Haplotype	Seq 5' to 3' direction	Haplotype Frequency (%)	Haplotype frequency genotypes
1	Hap_1	CATAGTCCTGCACGAGGAACAAATCG	10	[RRII 105-a]
2	Hap_2	CGGGTATTCAGTCAGAATGACTTCTA	20	[RRII 105-b GT1-b]
3	Hap_3	CGGGTACCCAGTCGGAGTGACTTCTA	20	[RRII 118-a GT1-a]
4	Hap_4	CGGGTATTTAGTCAGAATGACTTCTA	10	[RRII 118-b]
5	Hap_5	AATAGTCCTGCACAAGAAACAAATCA	10	[RRIM 600-a]
6	Hap_6	AATAGTCCTGCACGAGGAACAAATCG	10	[RRIM 600-b]
7	Hap_7	AATGGACCCACATAAGGAACAAATCA	10	[RRIC 52-a]
8	Hap_8	AGGGGATTCAGTCGGAATGACTTCTA	10	[RRIC 52-b]

*"a" and "b" represent the two homologus chromosomes of the diploid *Hevea brasiliensis*

Table 3.13. Comparative study of distribution ofnon-synonymous SNPs and resulting amino acids of *MVK* gene in rubber and selected plant species

No	Organism	HbMVK2490AG Aspartic acid(GAC)/ Asparagine (AAC)	2518A/T Not conserved	HbMVK2628AG Threonine(ACA)/ Alanine (GCA)	HbMVK2907AC Methionine (ATG) Leucine (CTG)
1	*Jatropha curcas*	Asp	#	Threo	Leucine
2	*Ricinus communis*	Asp	#	Threo	Leucine
3	*Populus euphratica*	Asp	#	Threo	Leucine
4	*Populus trichocarpa*	Asp	#	#	Leucine
5	*Malus domestica*	Asp	#	Threo	Leucine
6	*Eucalyptus grandis*	#	#	Threo	Leucine
7	*Vitis vinifera*	Asp	#	Threo	Leucine
8	*Citrus sinensis*	Asparagine	#	Threo	Met
9	*Citrus clementina*	Asparagine	#	Threo	Met
10	*Solanum lycopersicum*	Asp	#	Threo	Leucine
11	*Solanum tuberosum*	Asp	#	Threo	Leucine
12	*Nicotiana sylvestris*	Asp	#	Threo	Leucine
13	*Glycine max*	Asp	#	Alanine (Fabaceae family)	Leucine
14	*Medicago truncatula*	Asp	#	Alanine (Fabaceae family)	Leucine
15	*Brassica rapa*	#	#	Alanine (brassicasia family)	#
16	*Arabidopsis lyrata*	#	#	Alanine(brassicasia family)	#
17	*Arabidopsis thaliana*	#	#	Alanine(brassicasia family)	#
18	*Oryza sativa Indica*	#	#	Threo	Leucine
19	*Oryza sativa Japonica*	#	#	Threo	Leucine
20	*Triticum aestivum*	#	#	Threo	Leucine
21	*Zea mays*	#	#	Threo	Leucine
22	*Sorghum bicolor*	#	#	Threo	Leucine
23	Hevea-Hap-1	Asp	#	Threo	Met
24	Hevea-Hap-2	Asparagine	#	Alanine	Leucine
25	Hevea-Hap-3	Asp	#	Alanine	Leucine
26	Hevea-Hap-5	Asparagine	#	Threo	Met

The four non-synonymous *MVK* SNPs identified in *Hevea* clones and the resulting amino acid changes were traced in selected plant species. SNPs identified only in *Hevea* were considered for the comparative study. The non-synonymous SNP containing regions were mostly conserved. # non-conserved portions and nucleotide changes other than those detected in *Hevea*

3.4. Phospho Mevalonate Kinase (*PMVK*) gene

PMVK is an enzyme involved in the final step of MVA pathway where mevalonate diphosphate is converted to IPP monomer. Like MVK, it belongs to the GHMP kinase superfamily and are believed to have evolved *via* multiple gene duplications followed by neo-functionalization, of a common ancestor. Due to their dynamic nature in response to various substrates and the presence of positively charged side chains involved in the charge neutralization and binding, they are considered to play a crucial role in rubber biosynthesis in *Hevea*. Therefore sequence structure analysis of the gene responsible for the synthesis of PMVK was carriedout.

3.4.1. Phylogenetic analysis of PMVK

The phylogenetic tree constructed based on the encoded amino acid sequence of the gene *PMVK* clearly depicted the hierarchical linkage of this major enzyme across kingdoms and divisions. Differentiation based on sequence structure observed for organisms from different stratas of life was clearly shown in Figure 3.16. Protein sequences based on the *H. brasiliensis PMVK* gene haplotypes (based on non-synonymous SNPs) were clustered together with available *Hevea* sequences in the database. The sequences showed close similarity with their orthologs in species like *Jatropha curcuras, Ricinus communis*, etc., which formed a sub-group within the major cluster of plant kingdom. Cereals formed another sub-group within this cluster. Archae bacteria, eubacteria, arthropods and fungi formed separate clusters clearly depicting the hierarchy of evolution. Fishes, birds and animals along with human formed another major cluster. Viral sequences appeared to be the most distant one from all other sequences.

3.4.2. SNP identification in *PMVK* gene

The entire genomic region of *PMVK* gene was found to be approximately 9 kb in size with 10 exons. A total of 22 SNPs were identified from the analyzed sequences. SNP frequency of 1 in every 410 bp was observed, which was the lowest among all the rubber biosynthesis related genes analyzed so far. The 22 SNPs with complete phase information is showed in Table 3.14. Out of six SNPs identified in the coding region, four were synonymous and two were non synonymous in nature. Fourteen SNPs were

of transition events and eight SNPs were transversions. Nucleotide diversity (Pi) of the studied gene was 0.00104. Of the 6 coding SNPs, two were non-synonymous in nature. They are

HbPMVK5114GC (CTA Leucine ⟷ GTA-Valine)
HbPMVK8395AG (ATG-Methionine ⟷ GTG-Valine)

RRII 118 was found to be completely homozygous for all the SNPs identified in *PMVK* gene. RRIM 600 and GT1 also showed homozygosity in all the loci except for the SNPs at base position 297 and 3227. On the contrary RRII 105 and RRIC 52 showed more than 70% heterozygosity. Distribution of the 22 SNPs across the three groups based on the geographic region of origin showed that only two SNPs were shared among the Indian, Sri Lankan and Southeast Asian clones. Surprisingly unique SNPs were observed only in the Indian clones (Figure 3.17).

3.4.3. Fragment size variation at the 3' end of *PMVK* gene

PCR amplification of the last 3 kb portion of the *PMVK* gene resulted in fragments of different size from different clones. While RRII 118, RRIM 600 and GT1 had a bigger fragment of approximately 2 kb, RRII 105, RRIC 52 and *Hevea benthamiana* had fragment size of 0.9 kb (Figure 3.18). Sequencing of both the fragments confirmed that they are part of *PMVK* gene itself. It was identified that the smaller fragment occurred due to a deletion of around 1.2 kb in the last intron (10^{th}). Repeated PCR experiments using different primer combinations with additional clones like RRIC 100 revealed that both bands were present in RRII 105, RRIC 52 and RRIC 100 (Figure 3.19). It was noted that the upper band was faint in these clones compared to the lower band probably due to the preferential amplification of the smaller fragment. In order to confirm that the different sized fragments were not isoforms and to prove that they are alleles, both the fragments were separately isolated, cloned and sequenced. Moreover their segregation was also checked in a progeny of 46 individuals from a cross by the parents RRII 105 x RRII 118. Since RRII 105 was heterozygous (having the upper and lower bands) and RRII 118 was homozygous (having only the upper band) the progenies were expected to have both the types. Out of the 46 progenies,

twelve were homozygous for the upper allele like RRII 118 and the remaining 34 were heterozygous as in the case of RRII 105. Figure 3.20 shows the segregation of the two alleles in this progeny. Though the ratio is skewed showing significant deviation from the Mendelian inheritance mode, the resulting segregation ratio matched with the results of the segregation study conducted using two different SNP markers from the same gene using HRM technique.

3.4.4. Haplotype structuring of full-length *PMVK* gene

Haplotype reconstruction using DnaSP yielded eight haplotypes. The haplotypes with their allele frequency in each clone is showed in Table 3.16. Haplotype diversity "He" was found to be 0.844 . Two intrageneic recombination events were detected between base positions (1443-3227) and (3227-6584).

A phylogram was constructed based on the similarity of haplotypes constituted of 22 SNPs in the *PMVK* gene from five genotypes (Figure 3.21). The phylogram generated from the six haplotypes shows that the three haplotypes: Hap_1, Hap_2 and Hap_6 are related whereas Hap_3, Hap_4 and Hap_6 are clustered separately.

3.4.5. Comparative study of distribution of non-synonymous SNPs in *PMVK* gene

Comparative analysis of the encoded PMVK protein sequences from diverse plant species with that of *H.brasiliensis* revealed that the region harbouring the non-synonymous SNP HbPMVK5114GC and HbPMVK8395AG were conserved across different phylum and classes included in the study (Table 3.17). Interestingly the SNP HbMVK8395AG responsible for the amino acid change from Valine to Methionine was observed only in *Hevea.* On the contrary, amino acid change from Leucine to Valine resulting from the SNP HbPMVK5114GC was observed in other crops also. Figure 3.22 shows the multiple sequence aligned portion of PMVK protein sequence highlighting the amino acid change from Methionine to Valine resulting from the non-synonymous SNP HbPMVK8395AG.

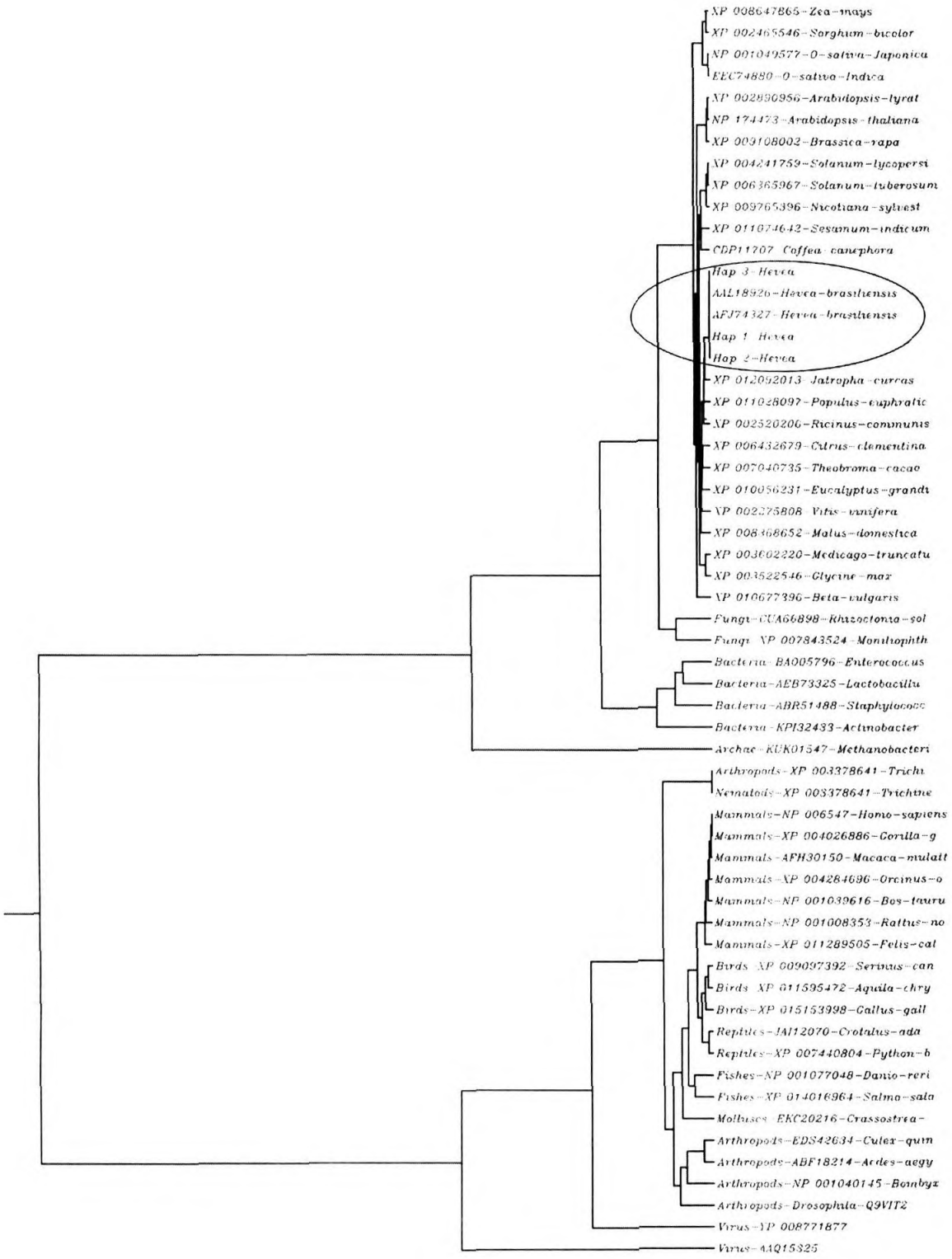


Figure 3.16. Phylogenetic tree of PMVK
Phylogenetic tree constructed using the encoded protein sequences of *PMVK* from selected plant species and representative organisms belonging to major domains of life. The tree depicts the evolutionary relationship of *PMVK* across different organisms.

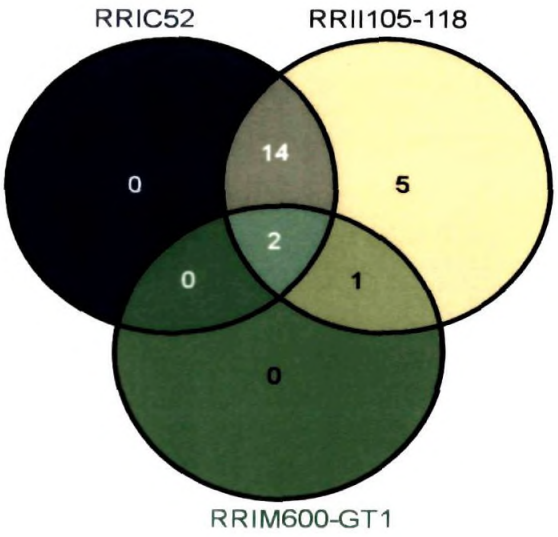


Figure 3.17. Distribution of SNPs in *PMVK* gene across the three groups *viz.* Indian (RRII 105 & RRII 118) Sri Lankan (RRIC 52) and Southeast Asian clones (RRIM 600 and GT1)

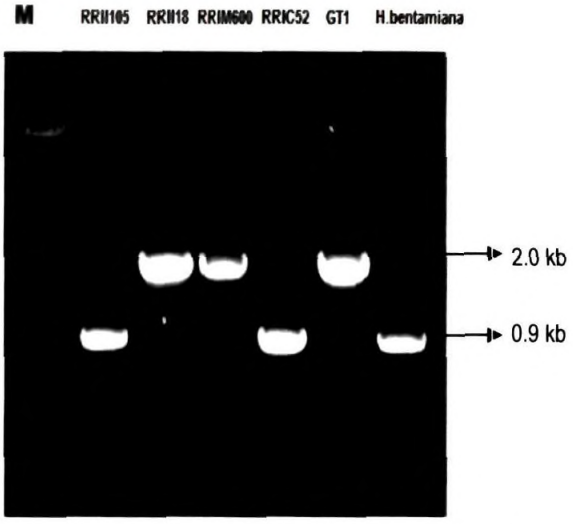


Figure 3.18. Gel picture showing the PCR product amplified from the 3' end of *PMVK* gene from different clones and the species *H. benthamiana* showing fragment size variation

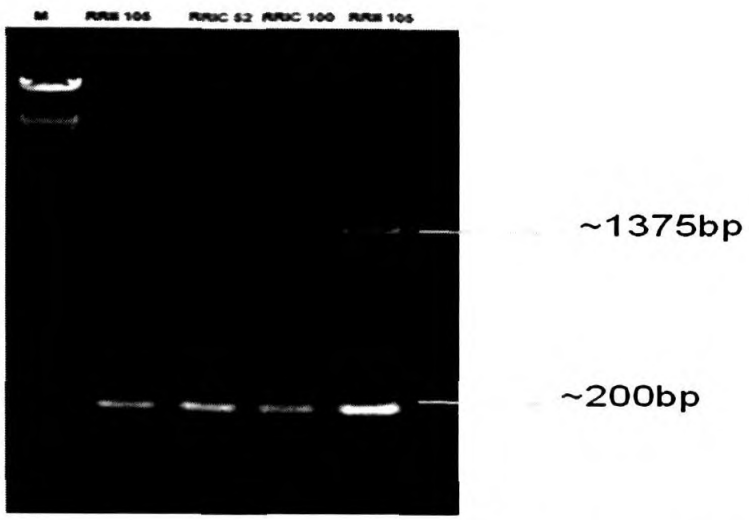


Figure 3.19. Gel picture showing the amplification of the large PMVK gene indel using different flanking primers in selected clones. A faint upper band was seen in all case and the size difference was approximately the same (1.2 kb) as noted in Figure 3.18.

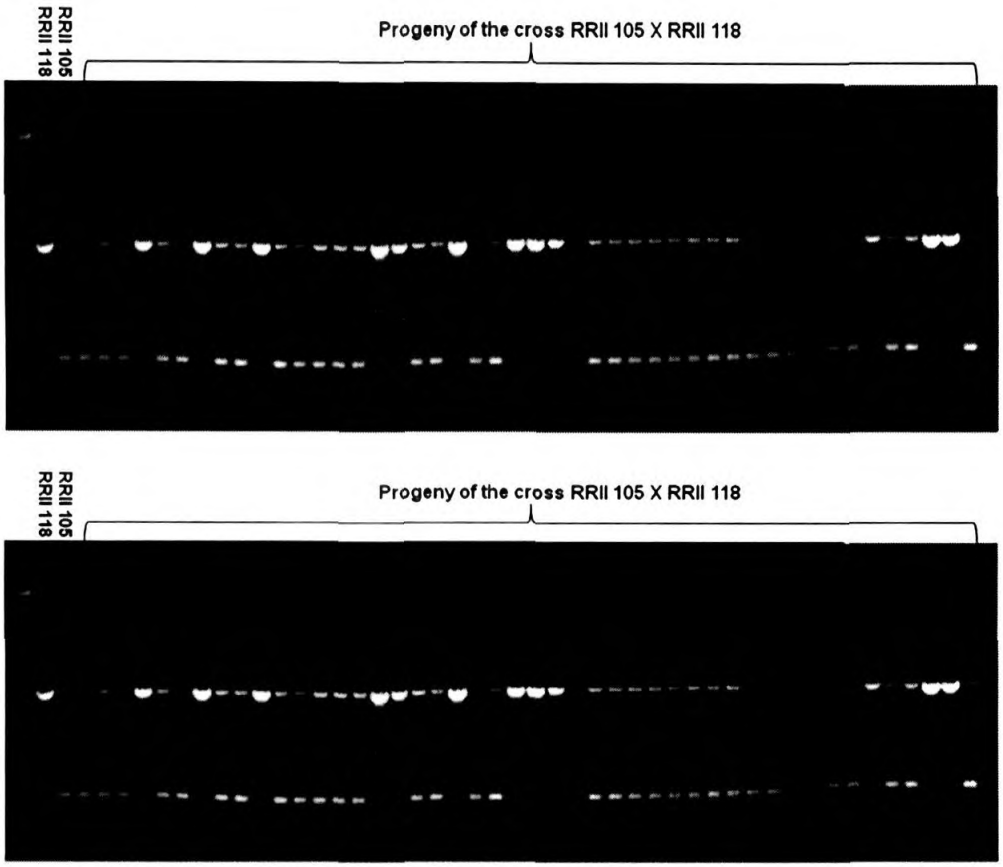


Figure 3.20. Gel picture showing the segregation pattern of upper and lower alleles (indels) of *PMVK* gene in a progeny population derived from the cross between RRII 105 and RRII 118.

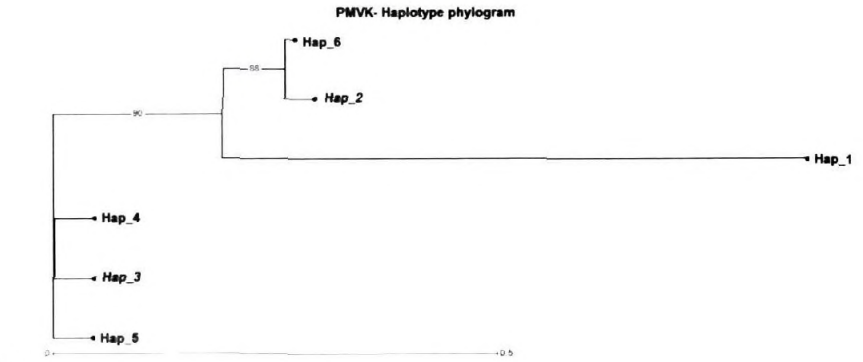


Figure 3.21. Phylogram generated based on the haplotypes reconstructed using the 22 SNPs in the *PMVK* gene

```
MUSCLE (3.8) multiple sequence alignment

                                                                    HbPMVK 8395 (A/G)
NP_001049577-O-sativa-Japonica          -MEVVASAPGKVLVAGGYLVLERPNAGLVLSTSARFYAIVRPLRDELSHDSWAWAWTDVK
EEC74880-O-sativa-Indica                -MEVVASAPGKVLVAGGYLVLERPNAGLVLSTSARFYAIVRPLRDELSHDSWAWAWTDVK
XP_008647865-Zea-mays                   -MEVVASAPGKVLIAGGYLVLERPNAGLVLSTTARFYAVVRPLRDSLPADAWAWAWTDVK
XP_002465546-Sorghum-bicolor            -MEVVASAPGKVLIAGGYLVLERPNAGLVLSTTARFYAVVRPIRDSLPADSWAWAWTDVK
XP_010677396-Beta-vulgaris              -MEIVASAPGKVLMTGGYLVLERPNAGLVLSTNARFYAIVKPFHDDVTSGSWAWSWMDVK
XP_009108002-Brassica-rapa              -MAVVASAPGKVLMTGGYLILERPNAGLVLSTNARFYAIVKPINDEVKPESWAWKWTDVK
XP_002890956-Arabidopsis-lyrata         -MAVVASAPGKVLMTGGYLVLDKPNAGLVLSTNARFYAIVKPINEDVKPESWAWKWTDVK
NP_174473-Arabidopsis-thaliana          -MAVVASAPGKVLMTGGYLVLEKPNAGLVLSTNARFYAIVKPINEEVKPESWAWKWTDVK
XP_011074642-Sesamum-indicum            -MAVVASAPGKVLMTGGYLILERPNAGIVLSTNARFYAIVKPLYEETKPESWAWAWTDVK
CDP11707-Coffea-canephora               ------------MTGGYLILERPNAGIVLSTNARFYAIVKPLYEEIKPESWAWAWTDVK
XP_009765396-Nicotiana-sylvestri        -MAVVASAPGKVLMTGGYLVLERPNAGIVLSTNARFYAIVKPLYEEIKPESWAWAWADVK
XP_004241759-Solanum-lycopersicu        -MAVVASAPGKVLMTGGYLVLERPNAGIVLSTNARFYAIVKPLHEEIIPESWAWAWTDVK
XP_006365967-Solanum-tuberosum          -MAVVASAPGKVLMTGGYLVLERPNAGIVLSTNARFYAIVKPLHEEIIPESWAWGWADVK
XP_003602220-Medicago-truncatula        MAVVVASAPGKVLMTGGYLVLERPNAGLVLSTNARFYAIVKPIYPQTKPDSWAWAWSDVR
XP_003522546-Glycine-max                MGEVVSSAPGKVLMTGGYLILERPNAGLVLSTNARFYAIVKPIHHEIKPDSWAWAWTDVR
XP_008368652-Malus-domestica            MEAVVASAPGKVLLTGGYLILERPNAGLVLSTNARFFAIVKPXYQPLKPBSXASGCTDVK
XP_002275808-Vitis-vinifera             -MAVVASAPGKVLMTGGYLILERPNAGIVLSTNARFYAIVKPLREDIEPDSWAWAWTDVK
XP_010056231-Eucalyptus-grandis         -MAVVASAPGKVLMTGGYLVLERPNAGIVLSTNARFYAIVKPICEEIKPDSWAWVWTDVK
XP_007040735-Theobroma-cacao            -MEVVASAPGKVLMTGGYLILERPNAGIVLSTNARFYAIVKPIHEDVKPESWAWAWTDVK
XP_006432679-Citrus-clementina          -MAVVASAPGKVLITGGYLILERPNAGIVLSTNARFYAIVKPLYEQVKPDSWSWGWTDVK
XP_011028097-Populus-euphratica         -MAVVASAPGKVLMTGGYLILERPNAGIVLSTNARFYAIVKPLYEEIKPDSWAWAWTDVR
XP_002520206-Ricinus-communis           -MAVVASAPGKVLMTGGYLILERPNAGIVLSTNARFYAIVKPLYDEIKPDSWSWAWTDVK
XP_012092013-Jatropha-curcas            -MAVVASAPGKVLMTGGYLILERPNAGIVLSTNARFYAIVKPLYDEIKPDSWAWAWTDVK
Hap_2-Hevea                             -MAVVASAPGKVLMTGGYLILERPNAGIVLSTNARFYAIVKPIYDEIKPDSWAWAWTDVK
AFJ74327-Hevea-brasiliensis             -MAVVASAPGKVLMTGGYLILERPNAGIVLSTNARFYAIVKPIYDEIKPDSWAWAWTDVK
Hap_3-Hevea                             -MAVVASAPGKVLMTGGYLILERPNAGIVLSTNARFYAIVKPIYDEIKPDSWAWAWTDVK
Hap_1-Hevea                             -MAVVASAPGKVLMTGGYLILERPNAGIVLSTNARFYAIVKPIYDEIKPDSWAWAWTDMK
                                                 ::****:*:.****:****.***:*:*.*       : :    *:.
```

Figure 3.22. Multiple sequence alignment of PMVK protein from *Hevea* haplotypes and other plants species with the amino acid change introduced by the SNP HbPMVK8395 (highlighted)

Table 3.14. SNPs obtained with their allelic status from the full-length gene sequence of *PMVK*

Genotype	Respective base position of each SNP locus																				
	102	297	1443	1787	2046	2281	2425	2428	2596	2669	2787	3104	3141	3227	4705	4864	5114	6584	6602	7850	8309	8395
RRII 105	C/T	A/G	A/T	C/T	C/T	A/G	A/T	C/T	A/G	A/T	GT	A/G	GT	C/T	GC	C/T	GC	C/C	C/C	C/C	T/T	G/G
RRII 118	T/T	G/G	A/A	C/C	T/T	G/G	A/A	C/C	G/G	A/A	T/T	G/G	G/G	C/T	G/G	T/T	C/C	T/T	T/T	T/T	G/G	A/A
RRIM 600	T/T	A/G	A/T	C/C	T/T	G/G	A/A	C/C	G/G	A/A	T/T	G/G	G/G	C/T	G/G	T/T	C/C	T/T	T/T	T/T	G/G	A/A
RRIC 52	C/T	A/A	A/T	C/T	C/T	A/G	A/T	C/T	A/G	A/T	GT	A/G	GT	C/T	GC	C/T	GC	C/C	C/C	C/C	T/T	G/G
GT1	T/T	A/G	A/A	C/C	T/T	G/G	A/A	C/C	G/G	A/A	T/T	G/G	G/G	C/T	G/G	T/T	C/C	T/T	T/T	T/T	G/G	A/A

Table 3.15. Characteristics of the 22 SNPs in *PMVK* gene

SNP	Position consensus	Variation	SNP freq	Nb readable individus	Major allele	Minor allele	Homo-zygotes major allele	Homo-zygotes minor allele	Hetero-zygotes
1	102	[T/C]	20.0 %	5	T	C	3	0	2
2	297	[A/G]	50.0 %	5	A	G	1	1	3
3	1443	[A/T]	30.0 %	5	A	T	2	0	3
4	1787	[T/C]	20.0 %	5	C	T	3	0	2
5	2046	[T/C]	20.0 %	5	T	C	3	0	2
6	2281	[A/G]	20.0 %	5	G	A	3	0	2
7	2425	[A/T]	20.0 %	5	A	T	3	0	2
8	2428	[T/C]	20.0 %	5	C	T	3	0	2
9	2596	[A/G]	20.0 %	5	G	A	3	0	2
10	2669	[A/T]	20.0 %	5	A	T	3	0	2
11	2787	[T/G]	20.0 %	5	T	G	3	0	2
12	3104	[A/G]	20.0 %	5	G	A	3	0	2
13	3141	[T/G]	20.0 %	5	G	T	3	0	2
14	3227	[T/C]	30.0 %	5	C	T	2	0	3
15	4705	[C/G]	20.0 %	5	G	C	3	0	2
16	4864	[T/C]	20.0 %	5	T	C	3	0	2
17	5114	[C/G]	20.0 %	5	C	G	3	0	2
18	6584	[T/C]	40.0 %	5	T	C	3	2	0
19	6602	[T/C]	40.0 %	5	T	C	3	2	0
20	7850	[T/C]	40.0 %	5	T	C	3	2	0
21	8309	[T/G]	40.0 %	5	G	T	3	2	0
22	8395	[A/G]	40.0 %	5	A	G	3	2	0

Table 3.16. Haplotypes generated using the SNPs in *PMVK* gene with their respective allele frequency in each clone

Sl No	Haplotype	Seq	Haplotype frequency (%)	Haplotype frequency genotypes
1	Hap_1	CATTCATTATGATTCCCCCCTG	20	[RRII 105-a RRIC52-a]
2	Hap_2	TGACTGACGATGGCGTGCCCTG	10	[RRII 105-b]
3	Hap_3	TGACTGACGATGGCGTCTTTGA	40	[RRII 118-a RRII 118-b RRIM 600-b GT1-b]
4	Hap_4	TATCTGACGATGGCGTCTTTGA	10	[RRIM 600-a]
5	Hap_5	TAACTGACGATGGCGTGCCCTG	10	[RRIC 52-b]
6	Hap_6	TAACTGACGATGGTGTCTTTGA	10	[GT1-a]

*"a" and "b" represent the two homologus chromosomes of the diploid *Hevea brasiliensis*

Table 3.17. Comparative study of distribution of *PMVK* non-synonymous SNPs and resulting amino acids in rubber and selected plant species

Sl No	Organism	HbPMVK 5114 (G/C) Leucine(CTA)/ Valine (GTA)	HbPMVK 8395 (A/G) Methionine (ATG)/Valine (GTG)
1	*Jatropha curcas*	#	Valine
2	*Ricinus communis*	#	Valine
3	*Populus euphratica*	#	Valine
4	*Sesamum indicum*	#	Valine
5	*Malus domestica*	Valine	Valine
6	*Theobroma cacao*	Leucine	Valine
7	*Eucalyptus grandis*	#	Valine
8	*Vitis vinifera*	Leucine	Valine
9	*Coffea canephora*	Leucine	Valine
10	*Citrus clementina*	#	Valine
11	*Beta vulgaris*	#	Valine
12	*Solanum lycopersicum*	#	Valine
13	*Nicotiana sylvestris*	#	Valine
14	*Glycine max*	#	Valine
15	*Medicago truncatula*	#	Valine
16	*Brassica rapa*	#	Valine
17	*Arabidopsis lyrata*	#	Valine
18	*Arabidopsis thaliana*	#	Valine
19	*Oryza sativa Indica*	Valine	Valine
21	*Oryza sativa Japonica*	Valine	Valine
22	*Zea mays*	#	Valine
23	*Sorghum bicolor*	#	Valine
24	Hevea-Hap-1	Leucine	Valine
25	Hevea-Hap-2	Valine	Valine
26	Hevea-Hap-3	Leucine	Methionine

The two non-synonymous *PMVK* SNPs identified in *Hevea* clones and the resulting amino acid changes were traced in selected plant species. # non-conserved portions and nucleotide changes other than those detected in *Hevea*

3.5. Farnesyl diphosphatase synthase (*FDPS*) gene

FDPS is a key initiator molecule responsible for the initiation of the chain elongation of IPP monomer units to form poly isoprene units. It is mainly involved in the sequential condensations of IPP with allylic diphosphates to generate isoprenoid carbon skeletons with widely ranging chain lengths. Previous studies in several latex producing plants and the finding that FDP is synthesized in the cytosol of *H. brasiliensis* indicate that it is the most likely initiator molecule for rubber biosynthesis in *Hevea*. Moreover the concentration of FDPS plays a significant role in determining the rate of the chain elongation which ultimately affects the rate of production of rubber in *Hevea*. Due to their crucial and species specific role in rubber biosynthesis in *Hevea* the sequence structure of the gene responsible for the synthesis of this enzyme was analysed in detail.

3.5.1. Phylogenetic analysis of FDPS

The phylogenetic tree constructed based on the encoded amino acid sequences of *FDPS* from different organisms clearly depicted the hierarchical linkage of this gene across major kingdoms and divisions. *FDPS* from *Hevea brasiliensis* was clustered along with the plant species like *Euphorbia pekinensis*, *Populus trichocarpa* etc., which formed a sub-group within the major cluster of plant kingdom. Cereals formed another sub-group within this cluster. As expected, the living fossil plant *Ginkgo biloba*, was placed in a separate branch away from all other plant species. The cluster of bacterial species including archae, proteo and acinetobacter were placed much away from the plant group. Fishes, birds and animals along with human formed another major cluster whereas fungi and arthropods formed two separate lines (Figure 3.23).

3.5.2. PCR amplification and primary sequence analysis of entire *FDPS* gene from five popular clones

PCR amplification and sequence analysis of the entire genomic portion of *FDPS* gene from five diverse popular *Hevea brasiliensis* clones was carried out with the intention of discovering SNPs from this gene. Amplification of the first intronic region using the primer combination HbFDP2-F & HbFDP2-R from RRII 118 produced two bands of different size based on the electrophoretic mobility whereas a single band of

intermediate size was observed in the other four clones (Figure 3.24). Primary analysis of this region using the cloned fragment sequence data showed the presence of major indels within the first intronic region responsible for the size variation of amplicons. Moreover, sequence analysis lead to the inference that the three different bands observed were actually three different alleles of the same *FDPS* gene. Interestingly, the rest of the gene from position 1379 bp to the 3' end of the gene appeared to be conserved apart from the occurance of SNPs which are normally seen in a population. Due to the aforementioned discrepancy in fragment size and sequence variation of the first intronic region, the first 1379 bp including 5'UTR, first exon and intron were analysed in detail separately from the rest of the gene. The analysis revealed the presence of eleven introns and twelve exons from the entire *FDPS* gene. Sequence structure of the entire *FDPS* gene showing 5' & 3' UTRs, large indels, introns and exons is schematically depicted in Figure 3.25.

3.5.3. Analysis of intron1 region of *FDPS* gene from five popular clones

PCR amplification of the intron1 region using primers flanking the variable region (HbFDP2-F & HbFDP2-R) in the five selected clones yielded two unique alleles of approximately 760 bp and 675 bp in RRII 118 and a single allele of around 690 bp in the other four clones (RRII 105, RRIM 600, RRIC 52, GT1). The upper and lower alleles observed in RRII 118 were designated as FINT1-A and FINT1-C whereas the single common allele of intermediate size observed in the other four clones was named as FINT1-B (Figure 3.24). This disparity prompted us to analyse Mil 3/2, one of the parents of RRII 118. The presence of same allelic combination in Mil 3/2 confirmed the existence and origin of the above alleles in RRII 118. Sequence analysis of intron1 alleles from the five popular clones using ClustalX as well as DNASIS revealed that the two alleles in RRII 118 varied significantly in terms of size and sequence from the common allele present in the other four clones. The presence of a hyper-variable region in the first 792 bp of the first intron including a major insertion of 64 bp in the FINT1-A allele and a deletion of 22 bp in the FINT1-C allele were found to be responsible for the size difference (Figure 3.26). Interestingly, the 64 bp insert in the FINT1-A allele appeared to be part of a 96 bp repeat region comprising of three 32 bp repeats with almost 95% accuracy. Repetition of this 32 base pair string was not

observed in FINT1-C and FINT1-B alleles. The 22 bp deletion event was observed only in the FINT1-C allele and no repeats were found associated with this site. Sequence variations within the major indels are summarised in Table 3.18. Exclusive sequence analysis of the 5' UTR region from the five clones revealed that at 155 bp (-11 bp upstream of the start codon), a single nucleotide variation (SNV) from C to T occurred in the plus strand of RRII 118 where, the "T" allele resulted in the introduction of a gibberellin responsive motif (GARE- TCTGTTG) in the sequence (Figure 3.27).

3.5.4. SNPs other than the major intronic indels in *FDPS* gene

Apart from the above large indels, thirty five SNPs including three single nucleotide indels were identified from the first intronic region. Sequence comparison of the hyper-variable region of the FINT1-A allele displayed only 91.4% similarity to the common FINT1-B allele whereas FINT1-C showed 95.5% similarity. It should be noted that 23 out of 35 single nucleotide variations observed in RRII 118 were in heterozygous state. A triallelic loci having alleles "A", "G" and "C" was observed at base position 819 of FINT1-A, FINT1-C and FINT1-B respectively. Another multiallelic locus was identified at base position 1065 (within the 22 bp indel) where FINT1-B had "T", FINT1-A had "G" and FINT1-C had "-" (the locus was within the 22 bp deleted region in FINT1-C). In addition to the above SNVs and indels, a nine base pair fragment was found repeated two times and a three base pair fragment (CTA) at base position 701 was found to be absent in FINT1-B allele (Figure .3.26).

Blastx analysis of the first intronic hyper-variable region revealed its homology to a non-LTR retrotransposon of the RTE super family. The FINT1-A allele showed more homology to the retroelement than the other two alleles. Blastx analysis of this region with the maize transposable element database showed its homology to a partial RIT_ jare like LINE retroelement which is a non-LTR retrotransposon (Table 3.19). The retroelement harbouring first intron also had a terminal polymorphic CT repeat having $(CT)^{13}$ in RRII 118 and $(CT)^{9}$ in other genotypes.

3.5.5. Estimation of *FDPS* intron1 allele status in popular clones, wild accessions, and *Hevea* species

In order to assess the distribution and status of the retrotransposon harbouring *FDPS* gene, 40 popular clones (including the initial five clones), 60 wild germplasm accessions and five *Hevea* species were analysed (Figure 3.28). Out of the 40 popular clones, 30 were found to be homozygous with 29 clones having FINT1-B allele and one (RRII 5) having FINT1-D allele (a new allele even smaller than FINT1-C). The smaller size was due to the deletion of a 184 bp region of the retroelement. The heterozygous clones had the combinations, FINT1-A/ FINT1-B (3 clones), FINT1-B/FINT1-D (3 clones), FINT1-C/ FINT1-D (3 clones) and FINT1-A/FINT1-C (1 clone) respectively. FINT1-A and FINT1-C in homozygous state were not observed in any of the clones. Screening of wild accessions and *Hevea* species revealed that the intron1 locus was homozygous in 48% wild accessions and all the five species analysed. Homozygous FINT1-B allele had the highest representation with their presence in 40% of the wild accessions. Apart from the aforementioned four major alleles, novel alleles were identified exclusively from the Rondonia accessions (Figure 3.29).

3.5.6. Sequence analysis of *FDPS* intron1 alleles from selected plants

Intron1 alleles from selected plants from each group (popular clones, wild accessions and *Hevea* species) which appeared to be unique based on electrophoretic mobility were amplified separately, cloned and sequenced. Multiple alignment and phylogenetic analysis of these intron1 sequences with the earlier allele sequences revealed that the most represented allele in entire population is the highly conserved FINT1-B allele. Popular clones except RRII 118 and RRII 5 were clustered together (Figure 3.30). Interestingly, the morphologically much different *H. pauciflora* was also found in the same group due to their 100% sequence similarity. In RRII 118, the upper allele (FINT1-A) was grouped along with Acre as well as Rondonia plants (Acre-19 and Ron-10) whereas the lower allele (FINT1-C) was restricted to Acre plants alone. The rarest and smallest of all the allele (Ron-6) was observed only in three Rondonia plants. Surprisingly, this unique allele which had a major deletion of 337 bp had maximum homology to the transposable element (Table 3.19).

3.5.7. Prediction of regulatory elements and splice sites from the first intron of *FDPS* gene

Search for *cis*-regulatory elements in the first intron comprising of 1379 bp from *Hevea* clones/accessions revealed the presence of 19 major regulatory motifs. The *cis*-regulatory motifs present in the *FDPS* intronel alleles of eleven representative plants are listed in Table 3.20. *H. benthamiana* had the maximum number with 17 motifs while Ron-6 had the least with twelve. Some of the major motifs like 5' UTR Py-rich stretch, ACE motif, ARE motif, "CAAT" motif, P-box, Skn-1_motif and TATA-box were present in all the alleles. A "CGTCA" motif involved in MeJA-responsiveness was found to be absent in both FINT1-A and FINT1-C alleles as well as in *H. nitida*. Similarly a "TATC-box" involved in gibberellin-responsiveness was absent in FINT1-C and Ron-6 allele. The *cis*-acting TCA element involved in salicylic acid responsiveness was found to be present only in FINT1-C. A "TC" element involved in defence and stress responsiveness was found unique to the allele in *H.benthamiana* and an F-box binding domain with putative signal transduction role was identified in FINT1-A allele of three Rondonia accessions.

Sequence analysis for splice site prediction revealed the existence of consensus donor site (agcggGTacttc) and acceptor site (gttgcAGatgtt) typical to intron-exon junctions. However, several additional acceptor and donor splice sites having scores above threshold were predicted by the FSPLICE module of softberry (www.softberry.com). For example, the donor site "ggaggGTgagtc" was found in all the alleles except in FINT1-A due to a single nucleotide change of "G" to "T" at position 688 (Figure 3.26). Similarly, two unique acceptor sites due to an indel at position 977were detected in FINT1-D allele. In the FINT1-C allele sequence, one acceptor site was converted to another acceptor site by a point mutation at base position 1109. The intronl allele sequences reported in this thesis were submitted to GenBank under the accession no KC886384 to KC886399.

3.5.8. Segregation analysis of *FDPS* intronl alleles in a progeny population

In order to confirm that the alleles observed were of the same gene (not the paralogs) and to ascertain their mode of inheritance, allele segregation was tested in a

progeny of full-sib family of parents RRII 105 and RRII 118. The female parent RRII 105 was homozygous (FINT1-B/ FINT1-B) while the male parent RRII 118 was heterozygous (FINT1-A/FINT1-C). As per Mendelian mode of inheritance the expected allele frequency for the alleles FINT1-A: FINT1-B: FINT1-C in the progeny was 1:2:1 and the expected allele segregation pattern is 50% FINT1-B/FINT1-A and 50% FINT1-B/FINT1-C (1:1). Out of the 60 progenies, 32 had FINT1-B/FINT1-C and 26 had FINT1-B/ FINT1-A allelic combinations which followed expected ratio of 1:1. Data for two plants were missing (Figure 3.31). Allele and genotype frequencies of the inron1 polymorphic loci of *FDPS* gene is shown in Table 3.21

3.5.9. Sequence analysis of *FDPS* gene excluding 5' end and first intron

Altogether 25 SNPs were identified from the 3598 bp region starting from base position 1380 onwards. RRII 105 was found to be completely homozygous in this region whereas RRIM 600 showed 88% homozygosity. RRII 118 had 60% homozygosity while both RRIC 52 and GT1 were found to be highly heterozygous with only five homozygous SNP loci. The entire region had only two exonic SNPs (HbFDPS4804 and HbFDPS4813) and both of them were synonymous in nature. The distribution of 25 SNPs in the five clones based on their region of origin is depicted in the Venn diagram (Figure 3.32). More number of unique SNPs was seen among the Indian clones than the others. Furthermore more number of SNPs were shared between the Sri Lankan clone RRIC 52 and the Indian clones.

3.5.10. Haplotype structuring of *FDPS* gene excluding the 5' end and first intron

PHASE analysis as well as haplotype identification was carried out on the unphased genotypic data using the Arlequin module of DnaSPv5 software. The software detected 25 polymorphic sites of which the SNPs at base position 2127 (G/C), 2755 (A/G), 3475 (A/T), 4489 (C/T) and 4707 (C/T) were variable single haplotypes. There were 20 informative sites with two variants. Seven haplotypes were identified with a haplotype diversity Hd: 0.867 (Table 3.23).

The minimum number of recombination events was calculated based on four-gamete test and one event was observed between base position 1380 and 1585. Hap_1

was the only haplotype seen in completely homozygous state (RRII 105) which occurred with a total frequency of four in the five clones analysed (in both alleles of RRII 105, one allele each in RRIM 600 and GT1). The haplotype, Hap_2 was constructed from the upper allele of RRII 118 and Hap_3 from the lower allele. Allelic variations in RRII 118 were observed at nine positions out of the 25 sites. The other four haplotypes were constructed from RRIM 600, RRIC 52 and GT1 respectively. Haplotypes Hap_4 and Hap_5 had minor variations from Hap_1 while haplotypes Hap_6 and Hap_7 displayed a single nucleotide variation from Hap_3.

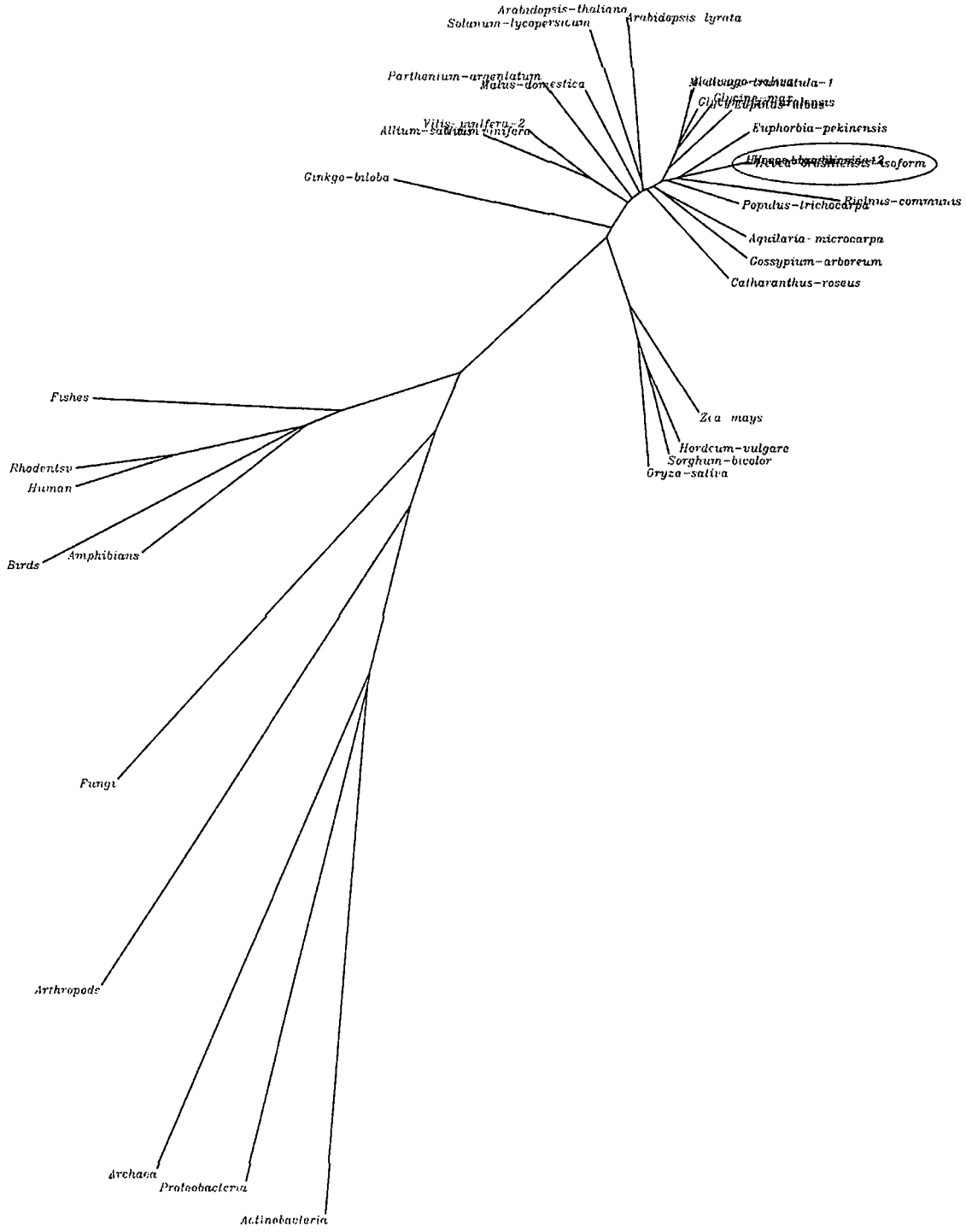


Fig.3.23. Phylogenetic tree of FDPS
Phylogenetic tree constructed using the protein sequences encoded by *FDPS* from selected plant species and representative organisms belonging to major domains of life. The tree depicts the evolutionary relationship of *FDPS* gene across different organisms.

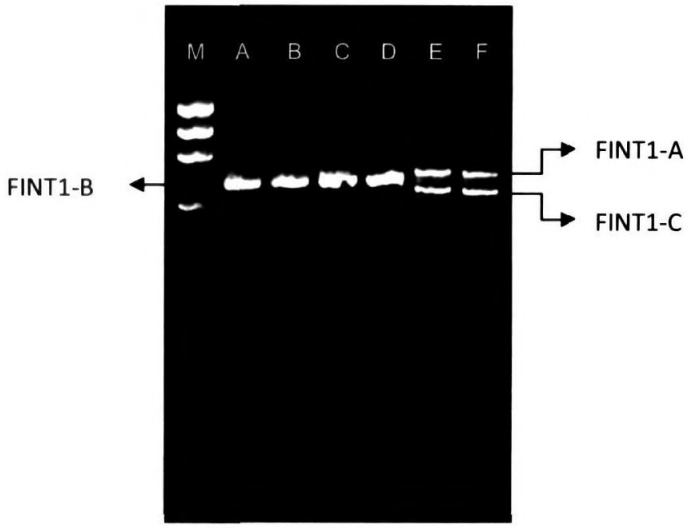


Figure 3.24. Gel picture showing different alleles of *FDPS* gene first intron

A- RRII 105, B- RRIM 600, C- RRIC 52, D- GT1, E- RRII 118, F- Mil 3/2 (one of the parents of RRII 118), M – Marker. RRII 105, RRIM 600, RRIC 52 and GT1 shared the same allele (middle band- FINT1-B). RRII 118 had two unique alleles (upper & lower bands -FINT1-A & FINT1-C). Mil 3/2 had the same allelic combination as that of RRII 118.

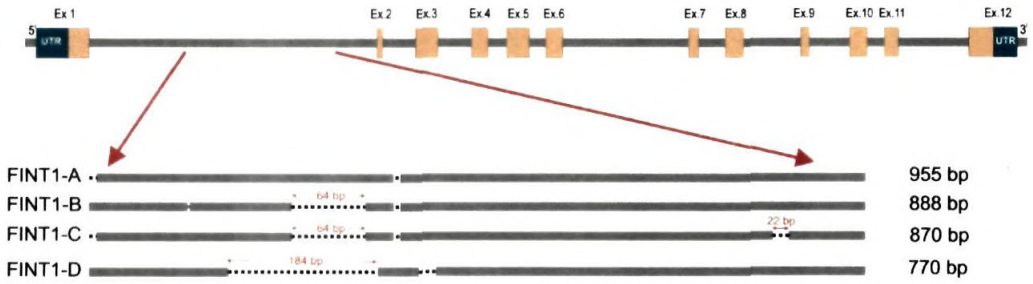


Figure 3.25. Schematic representation of *FDPS* gene showing 5' & 3' UTRs, indels, introns and exons: The four major intron1 alleles (FINT1-A, FINT1-B, FINT1-C, FINT1-D), with their size and location in large first intron are highlighted. The 64 bp indel in FINT1-B & FINT1-C, the 22 bp deletion in FINT1-C and the 184 bp deletion in FINT1-D are denoted by dots. The size of the alleles in base pairs obtained by amplifying the variable region using common conserved flanking primers are indicated in the right side. Ex= exon.

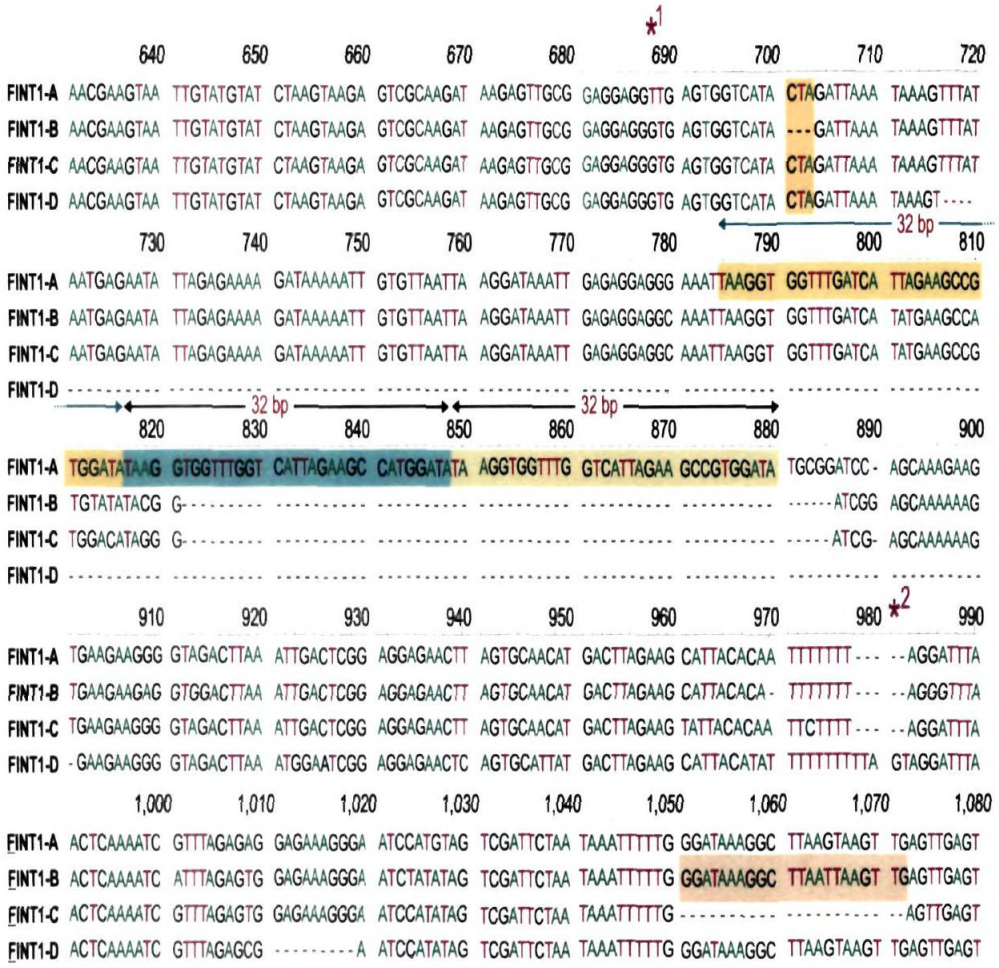


Figure 3.26. Multiple aligned nucleotide sequences of the *FDPS* intron1 alleles highlighting large indels and repeats.
The hyper-variable region harboured four major first intronic alleles (FINT1-A, FINT1-B, FINT1-C and FINT1-D). A major indel of 64 bp comprising of 32 bp repeats in FINT1-A starting from position 785 is highlighted. Another 22 bp deletion which was observed only in FINT1-C starts from position 1050 (highlighted in different colour). The core reverse transcriptase retro-sequence is absent in FINT1-D due to a large deletion starting from position 717. *1 denotes the deletion of the splice site "ggaggGTgagtc" in FINT1-A due to a single nucleotide variation (SNV) from G to T at position 688. *2 denote a five base addition in FINT1-D from 977 bp resulting in two novel acceptor sites.

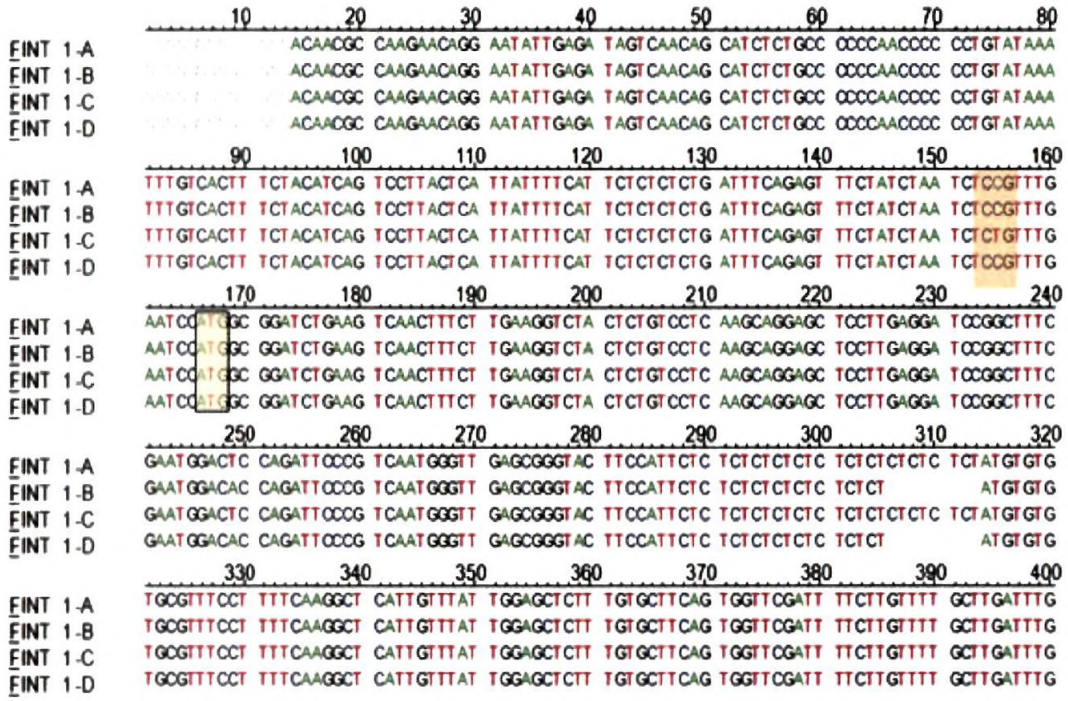


Figure 3.27. Multiple sequence alignment of the four major *FDPS* alleles highlighting the C/T single nucleotide variation at position 155 resulting in the induction of a gibberellins responsive motif (GARE) TCTGTTG in the FINT1-C allele of RRII 118. The start codon (ATG) is also highlighted.

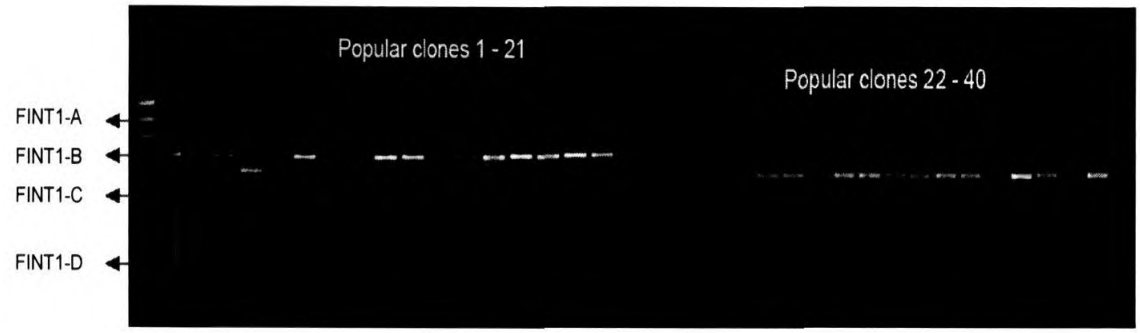


Figure 3.28. Gel picture showing the *FDPS* intron1 allelic status in forty popular clones including the initial five clones. All the four major alleles (FINT1-A, FINT1-B, FINT1-C, and FINT1-D) in various combinations can be observed. FINT1-D in homozygous state was seen in RRII 5 (5^{th} lane from the left).

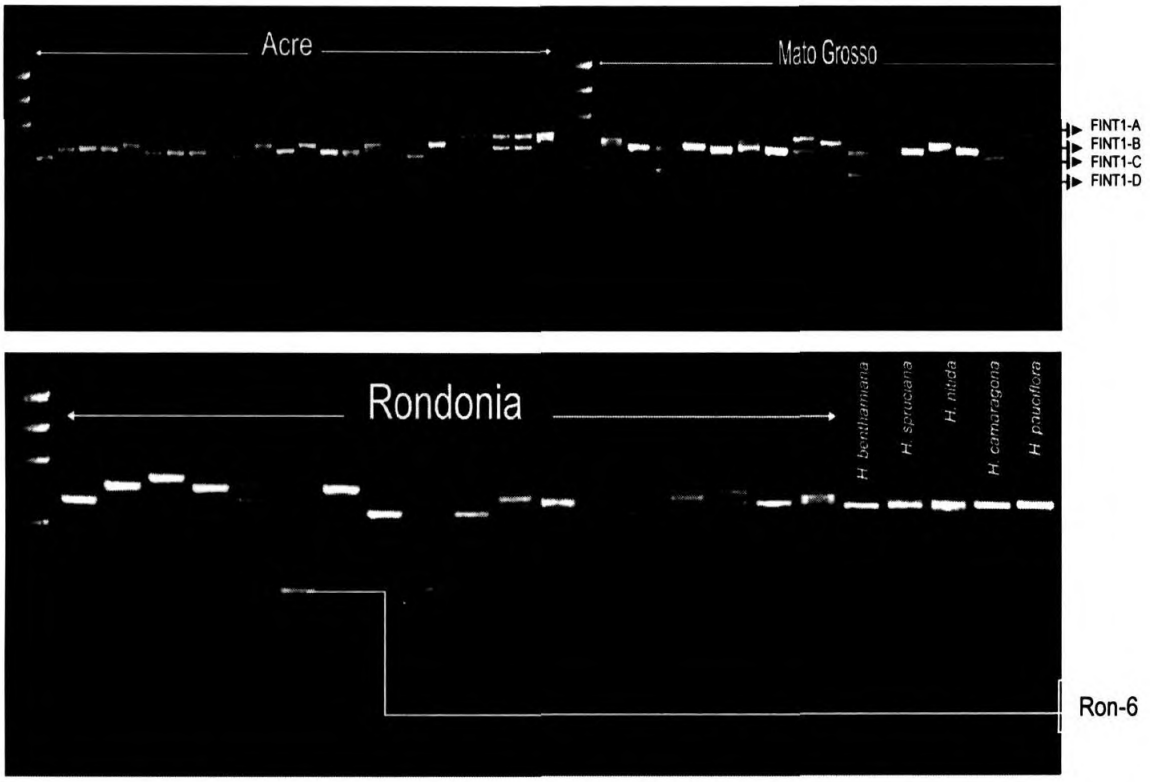


Figure 3.29. Gel picture showing the *FDPS* intron1 allelic status in sixty wild accessions and five *Hevea* species

24 Acre accessions, 18 Mato Grosso accessions, 18 Rondonia accessions and five *Hevea* species were loaded serially. All the four alleles (FINT1-A, FINT1-B, FINT1-C, FINT1-D) were noted in Acre and Mato Grosso accessions. Rondonia accessions have unique alleles not present in the other two groups. The rarest and smallest of all the alleles was observed only in three Rondonia plants of which Ron-6 was homozygous and the other two in combination with FINT1-A and FINT1-C respectively

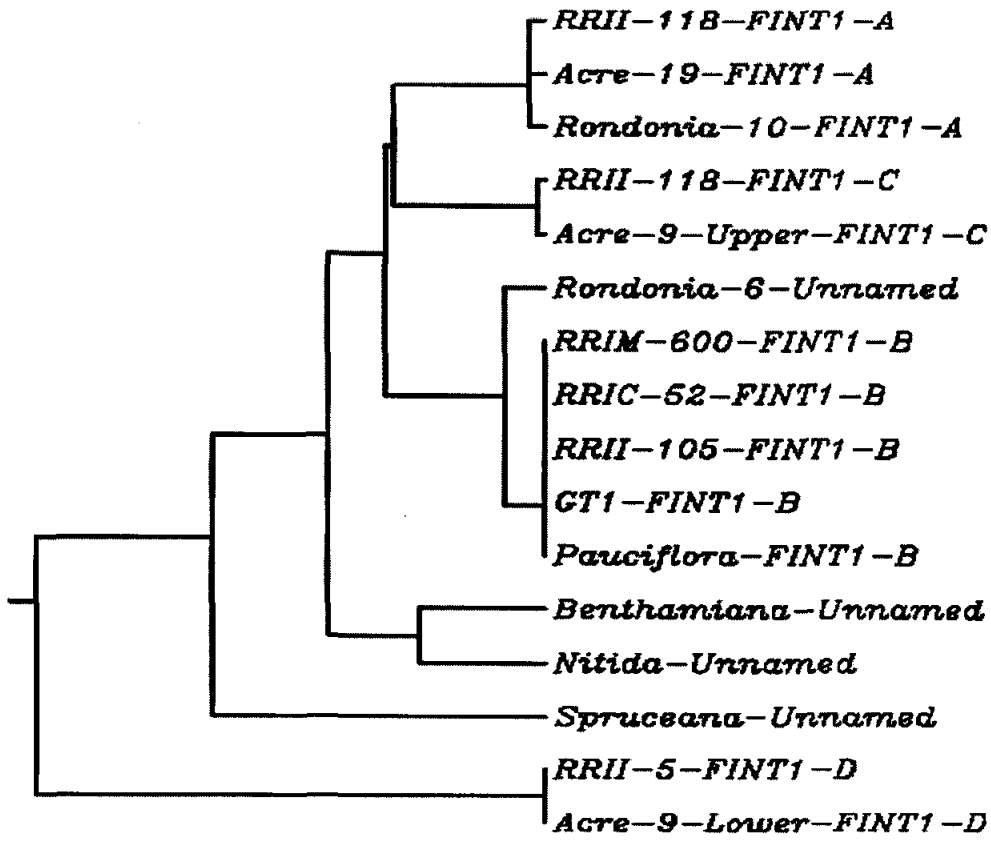


Figure 3.30. Phylogenetic tree constructed using the *FDPS* intron1 allele sequences from representative plants of popular rubber clones, wild accessions and *Hevea* species.

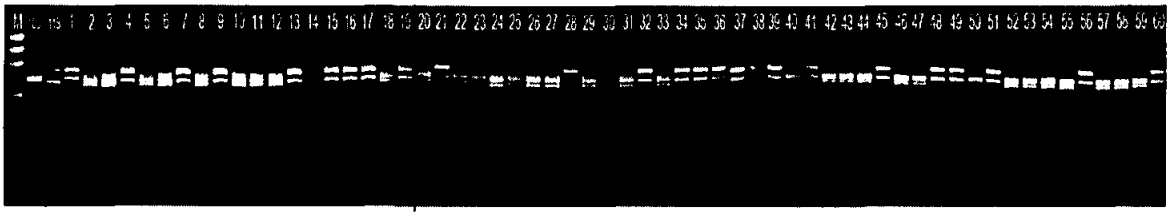


Figure 3.31. Gel picture showing *FDPS* intron1 allele segregation in a progeny population

The *FDPS* intron1 alleles, FINT1-A, FINT1-B and FINT1-C are segregating as per the Mendelian mode of inheritance in a full-sib progeny of the parents RRII 105 FINT1-B/ FINT1-B (female) and RRII 118 FINT1-A/ FINT1-C (male). Out of the sixty progenies, 32 had FINT1-B/ FINT1-C combination and 26 had FINT1-B/ FINT1-A combinations. Data for two plants were missing.

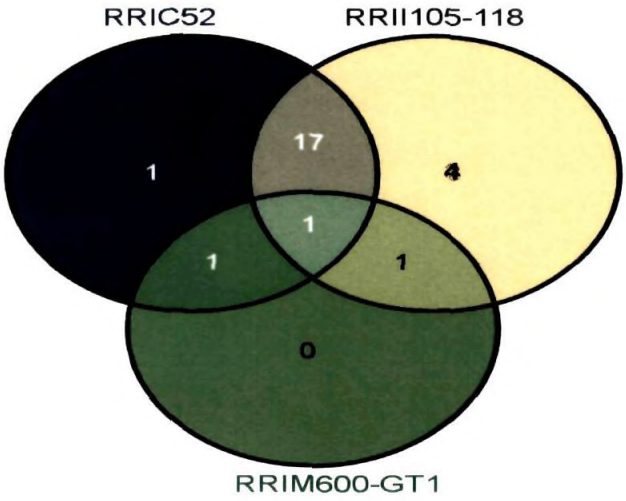


Figure 3.32. Distribution of SNPs in the *FDPS* gene across the three groups *viz.* Indian (RRII 105 & RRII 118) Sri Lankan (RRIC 52) and Southeast Asian (RRIM 600 and GT1)

Table 3.18. Sequence variations among intron1 major indels in *FDPS* gene

Alleles	FINT1-A			FINT1-B	FINT1-C
Status of 32 bp repeat in each allele	I	II	III	I	I
13th base position of the 32bp repeat	A	G	G	G	G
18th base position of the 32bp repeat	T	T	T	A	A
19th base position of the 32bp repeat	A	A	A	T	T
26th base position of the 32bp repeat	G	A	G	A	G
29th base position of the 32bp repeat	G	G	G	T	G
Status of 22 bp indel in each allele	present			present	absent
1065th base position (within the 22bp indel)	G			T	-

In FINT1-A allele, the indel site consists of three 32 bp repeats (I, II & III) with minor variations whereas the 32 bp is not repeated in the other two alleles resulting in 64 bp sequence variation between FINT1-A and FINT1-B/FINT1-C alleles. The 22 bp deletion was observed only in FINT1-C allele. These sequence variations were assumed to be the outcome of repeated copy paste events mediated by the retro elements. Deletion of 22bp starting from position 1051 was observed only in FINT1-C allele.

Table 3.19. Blastx result of the *FDPS* intron1 alleles

No	Plant & Allele name	Size bp	NCBI Blastx result	Acc No	Max score	Coverage	E value	Maize TE DB tblastx result	Score	E value
1	**Ron-6** (Unnamed)	637	Putative reverse transcriptase, [Solanum demissum]	AAT40500.1	64.3	29%	7e-10	RIT_jare_AC204843-0	71	2e-13
2	**Ron-10 [FINT1-A]**	955	Putative reverse transcriptase, [Solanum demissum]	AAT40500.1	57.8	19%	3e-07	RIT_jare_AC204843-0	63	8e-11
3	**RRII 118 [FINT1-A]**	955	Putative reverse transcriptase, [Solanum demissum]	AAT40500.1	57.8	19%	6e-07	RIT_jare_AC204843-0	**63**	9e-11
4	**Acre-19 [FINT1-A]**	955	Putative reverse transcriptase, [Solanum demissum]	AAT40500.1	57.8	19%	6e-07	RIT_jare_AC204843-0	63	9e-11
5	***H.spruciana*** (Unnamed)	888	Polyprotein, putative [Solanum demissum]	AAT40504.2	55.8	29%	2e-06	RIT_jare_AC204843-0	57	1e-10
6	***H.bentamiana*** (Unnamed)	879	Putative reverse transcriptase, [Solanum demissum]	AAT40500.1	51.6	21%	5e-05	RIT_jare_AC204843-0	46	1e-05
7	**RRII-118 [FINT1-C]**	870	Putative reverse transcriptase, *Solanum demissum*]]	AAT40500.1	48.9	21%	5e-04	RIT_jare_AC204843-0	44	7e-05
8	**Acre-9 [FINT1-C]**	870	Putative reverse transcriptase, [*Solanum demissum*]	AAT40500.1	48.9	21%	5e-04	RIT_jare_AC204843-0	44	7e-05
9	***H.pauciflora*** **[FINT1-B]**	888	Putative reverse transcriptase, [*Solanum demissum*]	AAT40500.1	45.1	20%	0.011	RIT_jare_AC204843-0	42	2e-04
10	**RRII 105 [FINT1-B]**	888	Putative reverse transcriptase, [*Solanum demissum*]	AAT40500.1	42.7	11%	0.065	RIT_jare_AC204843-0	42	2e-04
11	**RRII-5 [FINT1-D]**	770	No similarity	--	--	--	--	DTT_ZM00002_consensus	29	1.1
12	**Acre-9 [FINT1-D]**	770	No similarity	--	--	--	--	DTT_ZM00002_consensus	29	1.1
13	***H.nitida*** (Unnamed)	899	Polyprotein, putative [*Solanum demissum*]	AAT40504.2	32.0	16%	1.1	DTT_ZM00002_consensus	29	1.3

Table 3.20. The lists of cis-regulatory motifs present in the *FDPS* intronel alleles of eleven representative plants

Sl No	Motif name	Sequence	Putative Function	Ron-6 (Un named)	RRII 105 [B]	*H.pauciflora* [B]	Ron-10 [A]	RRII118 [A]	Acre-19 [A]	RRII-118 [C]	Acre-9 [C]	Acre-9 [D]	RRII5 [D]	*Hevea spruciana* (Un named)	*Hevea bentamiana* (Un-named)	*Hevea nitida* (Un-named)
1	5UTR Py-rich stretch	TTTCTCTCTCTC TC	cis-acting element conferring high transcription levels	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
2	ACE	ACTACGTTGG	cis-acting element involved in light responsiveness	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
3	ARE		cis-acting regulatory element essential for the anaerobic induction	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
4	CAAT-box	CAAT	common cis-acting element in promoter and enhancer regions	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
5	CGTCA-motif	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness	Y	Y	Y	N	N	N	N	N	Y	Y	Y	Y	N
6	GATA-motif	AAGATAAGATT	part of a light responsive element	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
7	I-box	aAGATAAGA	part of a light responsive element	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y
8	P-box	CCTTTTG	gibberellin-responsive element	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
9	Skn-1_motif	GTCAT	cis-acting regulatory element required for endosperm expression	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
10	TATA-box	TATA	core promoter element around -30 of transcription start	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
11	TATC-box	TATCCCA	cis-acting element involved in gibberellin-responsiveness	N	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y
12	TCA-element	TCAGAAGAGG	cis-acting element involved in salicylic acid responsiveness	N	N	N	N	N	N	Y	Y	N	N	N	N	N
13	TC-rich repeats	ATTCTCTAAC	cis-acting element involved in defense and stress responsiveness	N	N	N	N	N	N	N	N	N	N	N	**Y**	N
15	TGACG-motif	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness	Y	Y	Y	N	N	N	N	N	Y	Y	Y	Y	N
16	Unnamed__6	GCATTTTTATCA	SEF4 factor binding site	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
17	Circadian	CAANNNNATC	cis-acting regulatory element involved in circadian control	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
18	F-box	CTATTCTCATT	Signal transduction and cell cycle regulation	N	N	N	**Y**	**Y**	**Y**	N	N	N	N	N	N	N
19	O2-site	GATGATATGG	cis-acting regulatory element involved in zein metabolism regulation	N	N	N	N	N	N	N	N	N	N	N	Y	N

List of cis-regulatory motifs identified from the first intronic region of *FDPS* gene from representative plants of popular clones, wild germplasm accessions and *Hevea* species. The regulatory motif sequence, putative function and status in each plant are listed. (Y= present, N= absent)

Table 3.21. Segregation ratios of three *FDPS* intron1 alleles in a progeny population

Parents

RRII-105 (FINT1-B/ FINT1-B) x RRII-118 (FINT1-A/FINT1-C)

Allele frequency

Allele	Expected frequency	Observed frequency	Contribution to *chi*- square
FINT1-A	30(50%)	26 (43.3%)	0.53
FINT1-B	60 (100%	58 (96.6%)	0.06
FINT1-C	30 (50%)	32 (53.3%)	0.13
Total			0.72 (not significant)

Genotype frequency

Alleles	Expected frequency	Observed frequency	Contribution to *chi*-square
FINT1-B/ A	30 (50%)	26 (43.3%)	0.53
FINT1-B/ C	30 (50%)	32 (53%)	0.13
Total			0.66 (not significant)

Allele and genotype frequencies of the inron1 polymorphic loci of *FDPS* gene in a full sib family progeny. The alleles FINT1-A, FINT1-B and FINT1-C show no statistically significant deviation from Mendelian inheritance mode.

Table 3.22. SNP data of five popular clones from the twenty five polymorphic sites starting from 1380 bp onwards in *FDPS* gene

Genotype	Respective base position of each SNP locus																							
	1380	1585	1627	1918	1934	2013	2105	2127	2521	2755	2860	3356	3475	3743	3764	3800	3851	4234	4420	4489	4659	4707	4804	4813	4929
RRII 105	C/C	C/C	A/A	C/C	C/C	A/A	A/A	G/G	G/G	A/A	A/A	A/A	A/A	G/G	T/T	A/A	T/T	T/T	A/A	C/C	G/G	T/T	G/G	A/A	G/G
RRII 118	C/T	T/T	G/G	T/T	G/G	A/C	T/T	G/G	C/C	G/A	T/T	G/G	T/A	G/G	A/A	A/T	T/A	C/C	T/T	C/C	A/A	C/T	G/A	G/G	G/A
RRIM 600	C/T	C/C	A/A	C/T	C/C	A/A	A/A	G/G	G/G	A/A	A/A	A/A	A/A	G/T	T/T	A/A	T/T	T/T	A/A	C/C	G/G	T/T	G/G	A/A	G/G
RRIC 52	C/T	C/T	A/G	T/T	CG	A/C	A/T	G/C	G/C	A/A	A/T	A/G	A/A	TG	T/A	A/T	T/A	TC	A/T	C/C	G/A	T/T	G/A	A/G	G/A
GT 1	C/T	C/T	A/G	C/T	CG	A/C	A/T	G/G	G/C	A/A	A/T	A/G	A/A	G/G	T/A	A/T	T/A	TC	A/T	C/T	G/A	T/T	G/A	A/G	G/A

Table 3.23. Haplotypes constructed using the 25 SNPs in *FDPS* gene (excluding the first 1379 bp)

Sl No	Haplotype	Seq	Haplotype Frequency (%)	Haplotype frequency genotypes
1	Hap_1	CCACCAAGGAAAAGTATTACGTGAG	40	[RRII 105-a RRII 105-b RRIM 600-a GT1-a]
2	Hap_2	CTGTGATGCGTGTGAATCTCACGGG	10	[RRII 118-a]
3	Hap_3	TTGTGCTGCATGAGATACTCATAGA	10	[RRII 118-b]
4	Hap_4	TCATCAAGGAAAATTATTACGTGAG	10	[RRIM 600-b]
5	Hap_5	CCATCAAGGAAAATTATTACGTGAG	10	[RRIC 52-a]
6	Hap_6	TTGTGCTCCATGAGATACTCATAGA	10	[RRIC 52-b]
7	Hap_7	TTGTGCTGCATGAGATACTTATAGA	10	[GT1-b]

*"a" and "b" represent the two homologus chromosomes of the diploid *Hevea brasiliensis*

3.6. Geranylgeranyl diphosphate synthase (*GGDPS*) gene

GGDP Synthase catalyzes the condensation of IPP with an allylic prenyl diphosphate FDP to give GGDP, which is an essential precursor in the biosynthesis of several isoprenoids necessary for plant growth and development such as carotenoids, gibberellins, prenyl quinones, chlorophylls and formation of geranylgeranylated proteins. In *Hevea* it is considered as a possible initiator molecule in rubber biosynthesis process like FDPS. Therefore the gene sequence of *GGDPS* in popular *Hevea* clones was analysed in detail.

3.6.1. Phylogenetic analysis of GGDPS

A phylogenetic tree was constructed using conceptually translated GGDPS protein sequences which showed similar trend as that of the other rubber biosynthesis genes. *H. brasiliensis* showed maximum similarity with *Jatropha curcuras*. Clear differentiation based on sequence structure was observed for organisms from different stratas of life. As expected all the grass family members formed a separate sub-group. The living fossil *Ginko biloba* was the most divergent plant species placed away from other plants. Bacteria and animals formed separate groups as expected (Figure 3.33).

3.6.2. SNP identification in *GGDPS* gene

Fifteen bi-allelic SNPs were detected in 1500 bp sequenced genomic region of *GGDPS* gene from five popular clones resulting in an average of 1 SNP in every 100 bp. Sequence analysis revealed that the gene harbours no introns. The SNPs, their position and allelic status is shown in Table 3.24. Both RRII 105 and RRIM 600 were found to be highly homozygous for all the SNP loci except one heterozygous locus at base position 194. GT1 was highly heterozygous with 14 SNP loci. The distribution of 15 SNPs in the five clones based on their geographical origin is depicted in the Venn diagram (Figure 3.34). Only one unique SNP was observed in the Sri Lankan clone RRIC 52. No SNPs were shared between all the three groups. Except the two SNPs identified in the 5'UTR region, all were existing in the coding region. Eleven SNPs were the result of transition and four SNPs, were transversions. Nucleotide diversity of the studied gene (Pi) was 0.004. The characteristics of the identified SNPs were shown in Table 3.25

Of the 13 coding SNPs, five were non-synonymous in nature. The non-synonymous SNPs and the resulting amino acid changes are given below:

HbGGPS278AT (TAT- Tyrosine ⟷ TTT- Phenaylalanine)
HbGGPS535AC (ATC- Leucine ⟷ CTC- Isoleucine)
HbGGPS835AG (GCA- Alanine ⟷ ACA- Threonine)
HbGGPS1097AG (GGC- Glycine ⟷ GAC- Aspartic acid)
HbGGPS1249AC(AAG- Lysine ⟷ CAG- Glutamine)

3.6.3. Haplotype structuring of full-length *GGDPS* gene

Haplotype re-construction using DnaSP yielded 5 haplotypes. The estimated haplotype diversity, Hd was: 0.822. The haplotypes with their allele frequency in each clone is showed in Table 3.26. Interestingly the *GGDPS* gene sequence was devoid of any potential intragenic recombination sites. A phylogram was constructed based on the haplotypes consisting of 15 SNPs identified in the *GGDPS* gene (Figure 3.35). As evident from the figure two haplotypes: Hap_4 and Hap_5 evolved separately, whereas Hap_1 and Hap_2 clustered together.

3.6.4. Comparative study of distribution of non-synonymous SNPs in *GGDPS* gene

In the case of non-synonymous SNP HbGGDPS835AG, the aa Threonine was noted only in *Hevea* Hap_4. Therefore this seems to be a unique *Hevea* specific mutation. Similarly the SNP HbGGDPS1249AC resulting in the amino acid substitution from Lysine to Glutamine can be considered as a rare mutation which was noted only in *Vitis vinefera* in addition to *Hevea.* Interestingly the amino acid Aspartic acid resulting from the SNP mutation HbGGDPS1097AG seems to be specific to *Euphorbiaceae* family as it was seen only in members of this family like, *Ricinus communis, Jatropha curcas* and *Hevea brasiliensis.*

3.6.5. *In silico* protein 3D structural modelling analysis of GGDPS

Homology modelling study was carried out for *Hevea* GGDPS protein using SWISS-MODEL tool. Protein structure templates were searched using blast and HHblits methods available in SWISS-MODEL. Protein crystal structure 2J1P (Geranylgeranyl diphosphate synthase from *Sinapis alba)* was selected as the best

template for modelling. It was noted that 2J1P shared over 78% sequence identity with *Hevea GGDPS* gene. Five non-synonymous SNPs in modelled protein structure of *Hevea* GGDPS (14 Y, 100 L, 200 A, 287 D, 338 K) were highlighted using Chimera 1.10.2. Comparison of each SNP in *GGDPS* with respect to its location in the protein revealed that they do not have any direct interaction with the putative binding site and none of them were found inside or near the protein-small molecule interaction interface of GGDPS. The 3D protein structure showing the estimated location of five non-synonymous SNPs in *GGDPS* gene is depicted in Figure 3.37.

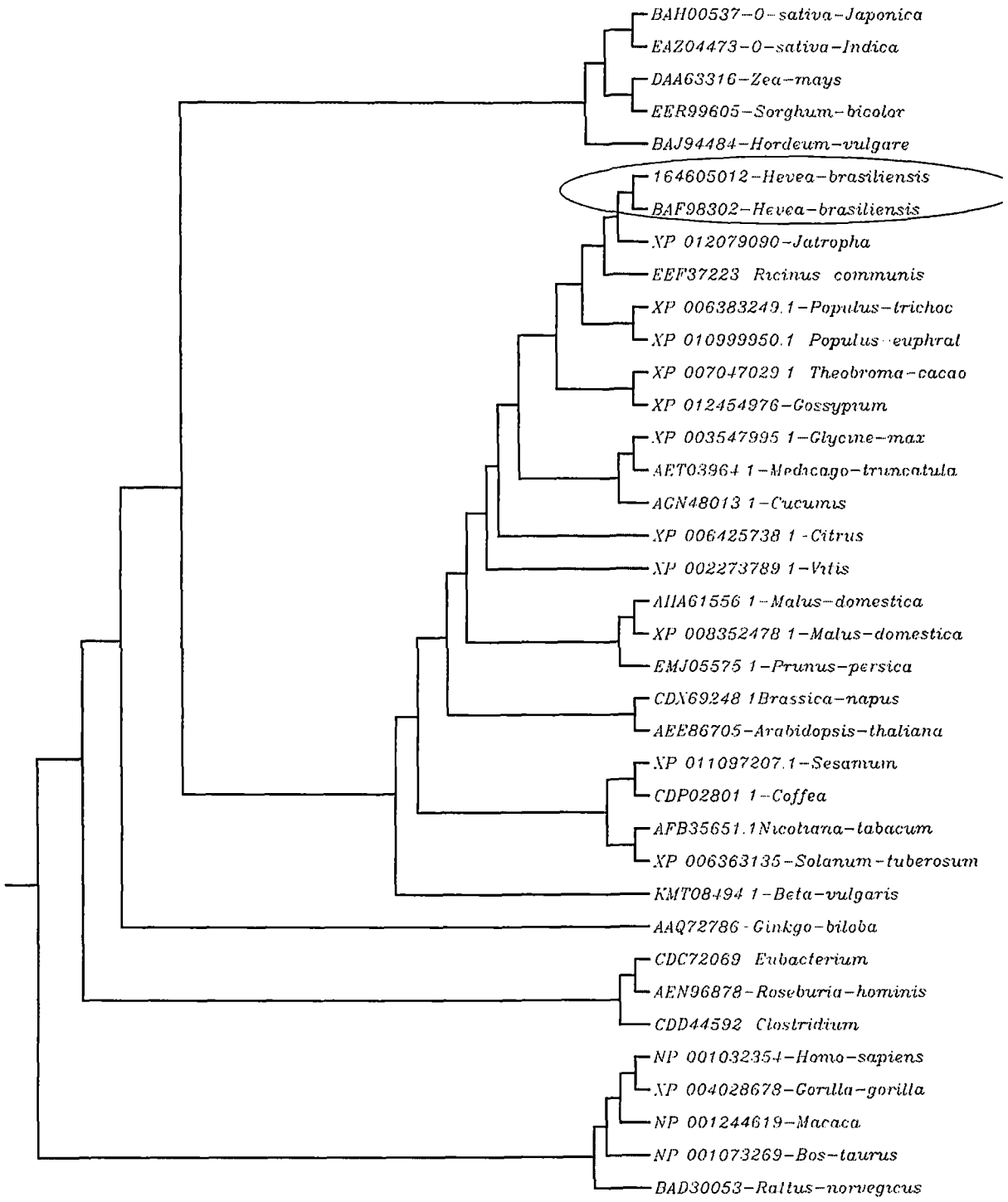


Figure 3.33. Phylogenetic tree of GGDPS
Phylogenetic tree constructed using encoded protein sequences of *GGDPS* from selected plant species and representative organisms belonging to major domains of life. The tree depicts the evolutionary relationship of *GGDPS* gene across different organisms

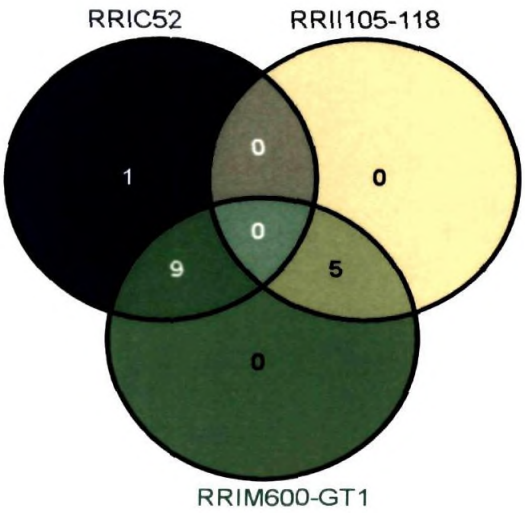


Figure 3.34. Distribution of SNPs in the *GGDPS* gene across the three groups *viz.* Indian (RRII 105 & RRII 118) Sri Lankan (RRIC 52) and Southeast Asian (RRIM 600 and GT1).

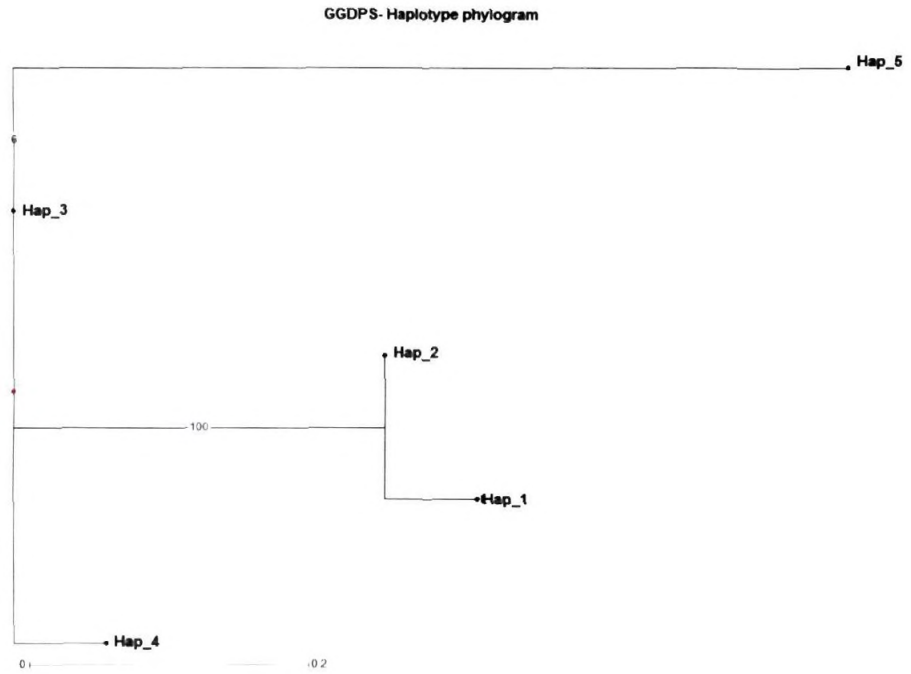


Figure 3.35. Phylogram based on the haplotypes generated using the 15 SNPs in *GGDPS* gene

```
                                    HbGGDPS 1249
AAQ72786-Ginkgo-biloba              EFAVELVNKAKEELSYFDPVKAAPLLGLADYIAFRQN
BAH00537-O-sativa-Japonica          EFAEKLLSDAREQLSGFDQETAAPLLHLANYIAYRQN
EAZ04473-O-sativa-Indica            EFAEKLLSDAREQLSGFDQETAAPLLHLANYIAYRQN
DAA63316-Zea-mays                   EFAEKLVSDATEQLACFDEEKAAPLLHLANYIAHRQN
EER99605-Sorghum-bicolor            EFAEKLLSDAIEQLDCFDKEKAAPLLHLANYIAHRQN
KMT08494.1-Beta-vulgaris            EYAKRLNKEAKEQLSGFDLEKAKPLIALADYIAYRQN
AFB35651.1Nicotiana-tabacum         DFAEELNKEAQAQLAGFDQEKAAPLIALANYIAYREN
XP_006363135-Solanum-tuberosum      EFAEELNKEAKAQLVGFDQEKAAPLFALADYIAYREN
CDX69248.1Brassica-napus            EFAEKLNREARDQLLGFDSDKVAPLLALANYIANRQN
AEE86705-Arabidopsis-thaliana       EFAEKLNREARDQLLGFDSDKVAPLLALANYIAYRQN
XP_002273789.1-Vitis                ELAEQLNKDAKDQLSGFDPDKAAPLIALSNYIAYRQN
XP_006425738.1-Citrus               KLADKLNKDAQQQLSEFDQEKAVPLIALANYIAYRQN
XP_006383249.1-Populus-trichocar    EFAEKLLNEARELLAGFDQEKAAPLIALANYIAYRQN
XP_010999950.1-Populus-euphratic    EFAEKLLNEARELLAGFDQEKAAPLIALANYIAYRQN
EEF37223-Ricinus-communis           EFAEKLNKEAQEQLAGFDPEKAAPLIALANYIAYRQN
XP_012079090-Jatropha               EFAQKLNKEAQEQLAGFDHEKAAPLIALANYIAYRQN
GGDPS-Hap-5                         EFAEQLNKEAQEQLAGFDPEKAAPLIALANYIAYRQN
GGDPS-Hap-4                         EFAEKLNKEAQEQLAGFDPEKAAPLIALANYIAYRQN
GGDPS-Hap-3                         EFAEKLNKEAQEQLAGFDPEKAAPLIALANYIAYRQN
164605012-Hevea-brasiliensis        EFAEKLNKEAQEQLAGFDPEKAAPLIALANYIAYRQN
BAF98302-Hevea-brasiliensis         EFAEKLNKEAQEQLAGFDPEKAAPLIALANYIAYRQN
GGDPS-Hap-1                         EFAEKLNKEAQEQLAGFDPEKAAPLIALANYIAYRQN
XP_003547995.1-Glycine-max          VFAAKLNKDAQDQLVGFDPVKAAPLIALANYIAYRQN
AET03964.1-Medicago-truncatula      EFAEKLNRDAQEQLSGFDLNKSAPLIALANYIAYRQN
XP_007047029.1-Theobroma-cacao      EFAEKLKSDALELLQGFDPEKAAPLVALANYIAYRQN
EMJ05575.1-Prunus-persica           EFAEKLIKDAKEQLSGFDLEKAAPLIALANYIAYREN
AHA61556.1-Malus-domestica          EFAEKLNKDAKEQLAGFDPEKAAPLIALANYIAYRQN
XP_008352478.1-Malus-domestica      EFAEKLNKDAKEQLAGFDPEKAARLIALANYIAYRQN
                                      *  *  .*   *  *:  .   *. *::*** *:*
```

Figure 3.36. Portion of the multiple aligned amino acid sequences of GGDPS from *Hevea* haplotypes and various plants species highlighting the amino acid change introduced by the non –synonymous SNP HbGGDPS1249AC

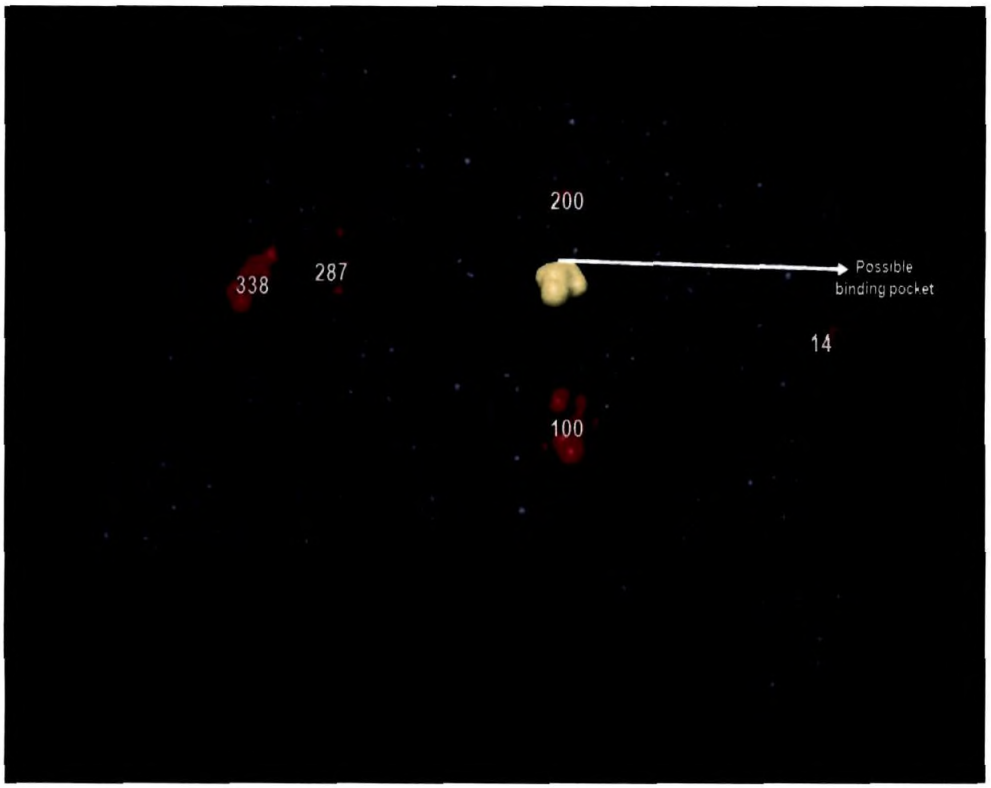


Figure 3.37. Predicted 3D structure of *Hevea* GGDP synthase. Amino acid positions which are associated with the non-synonymous substitutions are highlighted. Numbers refer to position of substitutions in the corresponding nucleotide sequence.

Table 3.24. SNPs with their allelic status identified from the full-length genomic sequence of *GGDPS* gene.

Genotype	Respective base position of each SNP locus														
	87	194	258	278	303	338	361	519	535	741	835	1097	1116	1173	1249
RRII 105	G/G	A/C	C/C	A/A	C/C	A/A	T/T	A/A	C/C	C/C	G/G	A/A	A/A	A/A	A/A
RRII 118	G/G	A/C	C/C	A/A	C/C	A/A	T/T	A/A	C/C	C/T	G/G	AG	AG	AG	A/A
RRIM 600	G/G	A/C	C/C	A/A	C/C	A/A	T/T	A/A	C/C	C/C	G/G	A/A	A/A	A/A	A/A
RRIC 52	AG	C/C	C/T	A/T	C/T	AG	C/T	AG	A/C	T/T	AG	G/G	G/G	G/G	A/C
GT1	AG	A/C	C/T	A/T	C/T	AG	C/T	AG	A/C	C/T	G/G	AG	AG	AG	A/C

Table 3.25. Characteristics of 15 SNPs in *GGDPS* gene

SNP	Position consensus	Variation	SNP freq %	Nb readable individus	Major allele	Minor allele	Homo major allele	Homo minor allele	Hetero
1	87	[A/G]	20.0 %	5	G	A	3	0	2
2	194	[A/C]	40.0 %	5	C	A	1	0	4
3	258	[T/C]	20.0 %	5	C	T	3	0	2
4	278	[A/T]	20.0 %	5	A	T	3	0	2
5	303	[T/C]	20.0 %	5	C	T	3	0	2
6	338	[A/G]	20.0 %	5	A	G	3	0	2
7	361	[T/C]	20.0 %	5	T	C	3	0	2
8	519	[A/G]	20.0 %	5	A	G	3	0	2
9	535	[A/C]	20.0 %	5	C	A	3	0	2
10	741	[T/C]	40.0 %	5	C	T	2	1	2
11	835	[A/G]	10.0 %	5	G	A	4	0	1
12	1097	[A/G]	40.0 %	5	A	G	2	1	2
13	1116	[A/G]	40.0 %	5	A	G	2	1	2
14	1173	[A/G]	40.0 %	5	A	G	2	1	2
15	1249	[A/C]	20.0 %	5	A	C	3	0	2

Table 3.26. Haplotypes generated using SNPs in *GGDPS* gene with their respective allele frequency in each rubber clone

*"a" and "b" represent the two homologus chromosomes of the diploid *Hevea brasiliensis*

Sl No	Haplotype	Seq	Haplotype Frequency (%)	Haplotype frequency genotypes
1	Hap_1	GACACATACCGAAAA	40	[RRII 105-a RRII 118-a, RRIM 600-a GT1-a]
2	Hap_2	GCCACATACCGAAAA	20	[RRII 105-b RRIM 600-b]
3	Hap_3	GCCACATACTGGGGA	10	[RRII 118-b]
4	Hap_4	GCCACATACTAGGGA	10	[RRIC 52-a]
5	Hap_5	ACTTTGCGATGGGGC	20	[RRIC 52-b GT1-b]

Table 3.27. Comparative study of distribution of *GGDPS* gene non-synonymous SNPs and resulting amino acids in rubber and selected plant species

Sl No	Organism	HbGGPS278 Not conserved	HbGGPS535 Leucine(CTC) Isoleucine (ATC)	HbGGPS835 Alanine (GCT) Threonine(ACT)	HbGGPS1097 Glycine(GGC) Aspartic acid (GAC)	HbGGPS1249 Glutamine acid (GAC) Lysine (AAG)
1	*Jatropha curcas*	#	#	Alanine	Aspartic acid	Lysine
2	*Ricinus communis*	#	Isoleucine	Alanine	Aspartic acid	Lysine
3	*Populus euphratica*	#	Leucine	Alanine	#	Lysine
4	*Populus trichocarpa*	#	Leucine	Alanine	#	Lysine
5	*Malus domestica*	#	Leucine	#	#	Lysine
6	*Beta vulgaris*	#	Leucine	Alanine	Glycine	#
7	*Vitis vinifera*	#	Isoleucine	Alanine	#	Glutamine
8	*Prunus persica*	#	Leucine	Alanine	#	Lysine
9	*Citrus clementina*	#	Leucine	Alanlne	#	Lysine
10	*Theobroma cacao*	#	Leucine	Alanine	#	Lysine
11	*Solanum tuberosum*	#	Leucine	Alanine	#	#
12	*Nicotiana tabacum*	#	Isoleucine	Alanine	#	#
13	*Glycine max*	#	Leucine	Alanine	#	Lysine
14	*Medicago truncatula*	#	Leucine	Alanine	#	Lysine
15	*Brassica napus*	#	Leucine	Alanine	#	Lysine
16	*Ginkgo biloba*	#	Leucine	Alanine	#	#
17	*Arabidopsis thaliana*	#	Leucine	Alanine	#	Lysine
18	*Oryza sativa Indica*	#	#	Alanine	#	Lysine
19	*Oryza sativa Japonica*	#	#	Alanine	#	Lysine
20	*Triticum aestivum*	#	#	Alanine	#	Lysine
21	*Zea mays*	#	#	Alanine	#	Lysine
22	*Sorghum bicolor*	#	#	Alanine	#	Lysine
23	*Hevea*-Hap-1	#	Leucine	Alanine	Aspartic acid	Lysine
24	*Hevea*-Hap-3	#	Leucine	Alanine	Glycine	Lysine
25	*Hevea*-Hap-4	#	Leucine	Threonine	Aspartic acid	Lysine
26	*Hevea*-Hap-5	#	Isoleucine	Alanine	Glycine	Glutamine

The amino acid changes resulting from the five non-synonymous SNPs in *Hevea GGDPS* gene was traced in selected plant species. SNPs identified only in *Hevea* were considered for the comparative study. The non-synonymous SNP containing regions were mostly conserved. # non-conserved portions and nucleotide changes other than those detected in *Hevea*

3.7. Rubber elongation factor (*REF*) gene

Rubber elongation factor (REF) is a major protein located on the surface of large rubber particles in latex and is involved in rubber biosynthesis in *H. brasiliensis*. *Hevea* REF promotes rubber biosynthesis *in vitro* (Dennis and Light, 1989), and its expression level in various *H. brasiliensis* accessions correlates with their rubber yields (Priya *et al.*, 2006). Due to the above crucial role played by REF in NR biosynthesis, its gene sequence from different rubber clones was studied in details.

3.7.1. Phylogenetic analysis of REF

Contrary to what observed in the case of other rubber biosynthesis genes, the conceptually translated REF protein sequences appear to be unique in *Hevea* (Figure 3.38). Protein-protein sequence blast revealed that *Hevea* REF is much different from the REFs present in other plant species. The closely related protein sequence from other species was *Jatropha* which had only 52% homology with *Hevea* REF. As the gene sequences were less in number, phylogenetic analysis was limited to few higher plants. The related genes were annotated as small rubber particle protein (*SRPP*) as well as *REF*. Figure 3.38 shows the phylogenetic tree constructed using the amino acid sequences of REF and SRPP from different plant species.

From the phylognetic tree it is evident that *Hevea REF* is a class of protein unique to *Hevea brasiliensis*. It is also evident that they share approximately 50% homology with another latex related protein called small rubber particle protein (SRPP). Though several available *Hevea* sequences were included in the study only one *Hevea* protein (AGO95096) designated as SRPP was clustered along with the proteins designated as REF in other plant species. All the other *Hevea* REF sequences formed a separate cluster demonstrating their uniqueness.

3.7.2. SNP identification in *REF* gene

A total of 32 SNPs were identified from the genomic region of 1.739 kb from each of the five rubber clones. The analysed region consisted of three exons and two introns. An SNP density of 1 in every 53 bases was obtained which was the highest among all the rubber biosynthesis genes analysed so far. RRII 105 and RRII 118 were found to be completely homozygous, whereas RRIM 600, RRIC 52 and GT1 were

highly heterozygous. In majority of the loci, the allelic status of RRIM 600 and RRIC 52 were found to be similar. The distribution of 32 SNPs in the five clones based on their region of origin is depicted as Venn diagram (Figure 3.39). Since the Indian clones were homozygous for all the loci, maximum number of SNPs were shared between the Southeast Asian clones and the Sri Lankan clone RRIC 52. Unique SNPs were noted only in Southeast Asian clones.

A total of 17 coding region SNPs were identified out of which 15 were non-synonymous. The non-synonymous SNPs and the corresponding amino acid changes are shown below.

HbREFAG752 (ACT- Threonine ⟷ GCT- Alanine)
HbREF 758GT (GCT -Alanine ⟷ TCT- Serine)
HbREF1382AT (GTC- Valine ⟷ GAC- Asp acid)
HbREF 1439AT (ATC- Isoleucine ⟷ AAC- Asparagine)
HbREF 1454CT (GCC- Alanine ⟷ GTC- Valine)
HbREF 1456AG (ATT- Isoleucine ⟷ GTT- Valine)
HbREF1460AT (CGA- Arginine ⟷ GGA- Glycine)
HbREF 1495CT (TCT- Serine ⟷ CCT- Proline)
HbREF 1505AG (GGG- Glycine ⟷ GAG- Glut acid)
HbREF 1508AG (CAG- Glutamine ⟷ CGG- Arginine)
HbREF 1520CT (CTT- Leucine ⟷ CCT- Proline)
HbREF 1522GT (GCT- Alanine ⟷ TCT- Serine)
HbREF 1533GT (TTT- Phenyl alanine ⟷ TTG- Leucine)
HbREF 1542GC (GAG- Glutamic acid ⟷ GAC- Aspartic acid)

3.7.3. Haplotype structuring of full-length *REF* gene

Haplotype re-construction using DnaSP yielded 5 haplotypes. The haplotypes with their allele frequency in each clone is shown in Table 3.30. Hap_1 occurred with the highest frequency (6/10). Intragenic recombination site were detected between positions 758 & 781 and between 781 & 1230 bp. A phylogram (Neighbouring joining tree) was constructed based on the haplotypes reconstructed using the 32 SNPs distributed in the *REF* gene (Figure 3.40). As evident from the figure, Hap_1 and

Hap_3 formed one cluster showing their relationship whereas Hap_2 and Hap_4 formed another cluster indicating their co-evolvement. Hap_5 appeared to be distinct from the other types.

Comparative study of distribution of non-synonymous SNPs was not carried out for *REF* gene due to the high sequence dissimilarity of *Hevea REF* with that of other plant species.

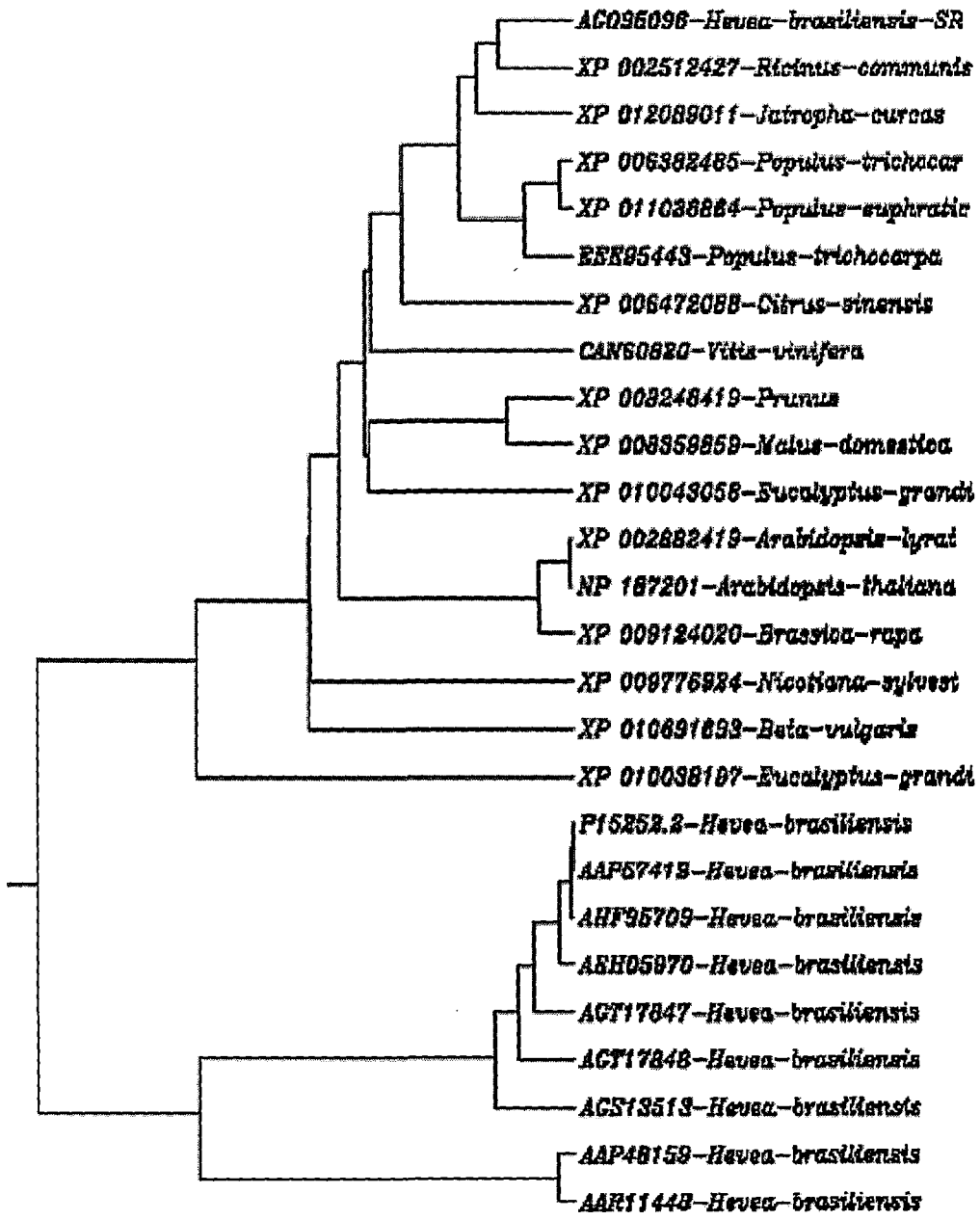


Figure 3.38. Phylogenetic tree of REF

Phylogenetic tree constructed using the encoded protein sequences of *REF* and or *SRPP* from related plant species. The two major clusters were observed, with one cluster comprised exclusively of *Hevea* REFs and the other with SRPPs from *Hevea* as well as other plant species.

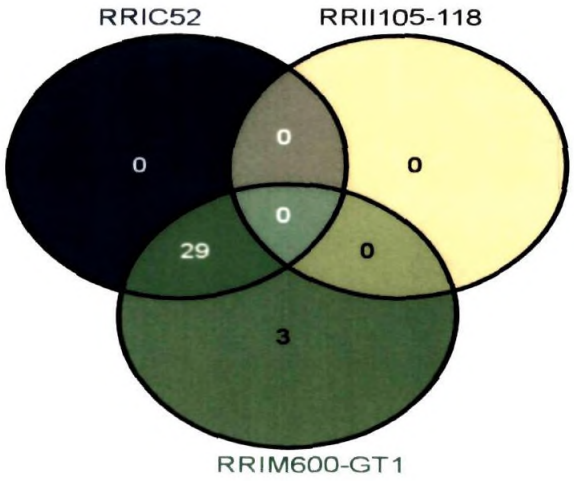


Figure 3.39. Distribution of *REF* gene SNPs across the three groups *viz.* Indian (RRII 105 & RRII 118) Sri Lankan (RRIC 52) and Southeast Asian (RRIM 600 and GT1)

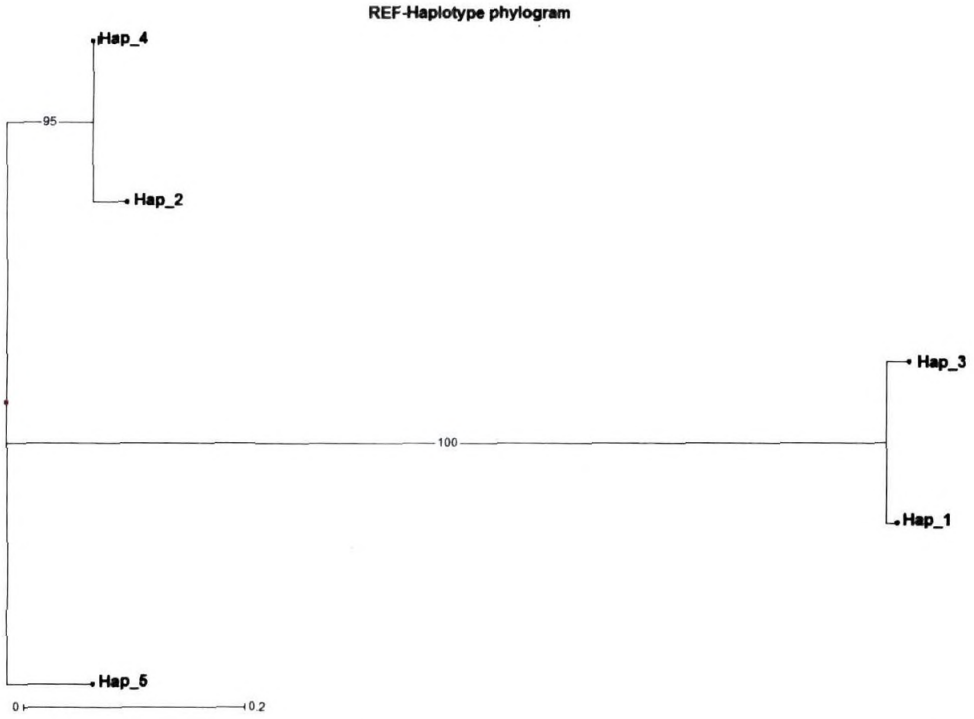


Figure 3.40 Phylogram based on the similarity of haplotypes reconstructed using the 32 SNPs in *REF* gene.

Table 3.28. SNPs with their allelic status identified from the full-length genomic sequence of *REF* gene

Genotype	Respective positions of each SNP locus																
	120	130	134	140	164	367	376	429	474	482	752	758	781	1230	1334	1359	1382
RRII 105	C/C	A/A	A/A	C/C	A/A	A/A	A/A	T/T	T/T	G/G	A/A	G/G	C/C	C/C	A/A	T/T	T/T
RRII 118	C/C	A/A	A/A	C/C	A/A	A/A	A/A	T/T	T/T	G/G	A/A	G/G	C/C	C/C	A/A	T/T	T/T
RRIM 600	C/T	A/G	A/G	C/G	A/G	A/C	A/T	C/T	G/T	A/G	A/G	G/T	C/C	C/T	A/G	C/T	A/T
RRIC 52	C/T	A/G	A/G	C/G	A/G	A/C	A/T	C/T	G/T	A/G	A/G	G/T	C/T	C/T	A/G	C/T	A/T
G/T 1	C/T	A/G	A/G	C/C	A/G	A/C	A/T	C/T	G/T	A/G	A/G	G/T	C/T	C/T	G/G	C/T	A/T

Genotype	Respective positions of each SNP locus													
	1439	1452	1454	1456	1457	1460	1495	1505	1508	1520	1522	1533	1542	1552	1585
RRII 105	T/T	A/A	C/C	A/A	T/T	G/G	T/T	G/G	A/A	T/T	T/T	T/T	G/G	C/C	T/T
RRII 118	T/T	A/A	C/C	A/A	T/T	G/G	T/T	G/G	A/A	T/T	T/T	T/T	G/G	C/C	T/T
RRIM 600	A/T	C/A	C/T	A/G	A/T	G/G	C/T	A/G	A/A	C/T	G/T	G/T	C/G	C/A	A/T
RRIC 52	A/T	C/A	C/T	A/G	A/T	G/G	C/T	A/G	A/A	C/T	G/T	G/T	C/G	C/A	A/T
G/T 1	A/T	C/A	C/T	A/G	A/T	C/G	C/T	A/G	A/G	C/T	G/T	G/T	C/G	C/A	A/A

Table 3.29. Characteristics of the 32 *REF* gene SNPs

SNP	Position consensus	Variation	SNP frequency	Nb readable Individus	Major allele	Minor allele	Homo-zygotes major allele	Homo-zygotes minor allele	Hetero-zygotes
1	120	[T/C]	10.0 %	5	C	T	4	0	1
2	130	[A/G]	30.0 %	5	A	G	2	0	3
3	134	[A/G]	30.0 %	5	A	G	2	0	3
4	140	[C/G]	20.0 %	5	C	G	3	0	2
5	164	[A/G]	30.0 %	5	A	G	2	0	3
6	367	[A/C]	30.0 %	5	A	C	2	0	3
7	376	[A/T]	30.0 %	5	A	T	2	0	3
8	429	[T/C]	30.0 %	5	T	C	2	0	3
9	474	[T/G]	30.0 %	5	T	G	2	0	3
10	482	[A/G]	30.0 %	5	G	A	2	0	3
11	752	[A/G]	30.0 %	5	A	G	2	0	3
12	758	[T/G]	30.0 %	5	G	T	2	0	3
13	781	[T/C]	20.0 %	5	C	T	3	0	2
14	1230	[T/C]	30.0 %	5	C	T	2	0	3
15	1334	[A/G]	30.0 %	5	A	G	2	0	3
16	1359	[T/C]	30.0 %	5	T	C	2	0	3
17	1382	[A/T]	30.0 %	5	T	A	2	0	3
18	1439	[A/T]	30.0 %	5	T	A	2	0	3
19	1452	[A/C]	30.0 %	5	A	C	2	0	3
20	1454	[T/C]	30.0 %	5	C	T	2	0	3
21	1456	[A/G]	30.0 %	5	A	G	2	0	3
22	1457	[A/T]	30.0 %	5	T	A	2	0	3
23	1460	[C/G]	10.0 %	5	G	C	4	0	1
24	1495	[T/C]	30.0 %	5	T	C	2	0	3
25	1505	[A/G]	30.0 %	5	G	A	2	0	3
26	1508	[A/G]	10.0 %	5	A	G	4	0	1
27	1520	[T/C]	30.0 %	5	T	C	2	0	3
28	1522	[T/G]	30.0 %	5	G	T	2	0	3
29	1533	[T/G]	30.0 %	5	T	G	2	0	3
30	1542	[C/G]	30.0 %	5	G	C	2	0	3
31	1552	[A/C]	30.0 %	5	C	A	2	0	3
32	1585	[A/T]	20.0 %	5	T	A	3	0	2

Table 3.30. *REF* gene haplotypes with their allele frequencies in each clone

Sl No	Haplotype	Seq 5' to 3' direction	Haplotype Frequency (%)	Haplotype frequency genotypes
1	Hap_1	CAACAAATTGAGCCATTTACATGTGATGTGCT	60	[RRII 105-a RRII 105-b RRII 118-a RRII 118-b RRIM 600-a GT1-a]
2	Hap_2	TGGGGCTCGAGTCTGCAACTGAGCAACTGCAA	10	[RRIM 600-b]
3	Hap_3	CAACAAATTGAGTCATTTACATGTGATGTGCT	10	[RRIC 52-a]
4	Hap_4	CGGGGCTCGAGTCTGCAACTGAGCAACTGCAA	10	[RRIC 52-b]
5	Hap_5	CGGCGCTCGAGTTTGCAACTGACCAGCTGCAT	10	[GT1 b]

*"a" and "b" represent the two homologus chromosomes of the diploid *Hevea brasiliensis*

3.8. *cis*-prenyltransferase (*CPT*) gene

cis-prenyltransferase: *cis*-prenyltransferases (CPTs), are the key downstream enzymes involved in the synthesis of natural rubber which are crucial in determining the molecular weight of the rubber. They elongate a short all-trans precursor, oligoprenyl diphosphate, by sequential addition of the desired number of IPP molecules, which, in turn, results in formation of a long hydrocarbon skeleton of *cis* units (NR). Due to their species specific function in *Hevea* rubber biosynthesis their gene sequence structure was analysed in detail.

3.8.1. Phylogenetic analysis of CPT

As in the case of other rubber biosynthesis genes, clear demarcation in terms of sequence structure was observed for the different organisms included in the study based on their sequence divergence (Figure 3.41). Since prenyltransferases comprise of a large gene family, extensive sequence variations were noted in this gene in various organisms. Various isoforms of this gene were also noted in several organisms. Due to the above reasons sequence information used for the current analysis was limited to selected plants, animals and fungi. One notable observation from the study was the divergence of *Hevea CPTs* from that of other plant species as observed in the case of *REF* gene. While the sequences from other plants came under a single cluster, *Hevea CPTs* formed a separate cluster which is clearly different from the rest demonstrating their uniqueness. As expected, organisms like mammals, fishes, arthropods and fungi formed their own distinct grouping.

3.8.2. Classification of *Hevea CPTs* based on DNA sequence

Being a major gene involved in the chain elongation of isoprene units, *cis*-prenyltransferase sequences were reported by various groups and named them without following any conventional standard nomenclature. In order to broadly classify them based on their sequence similarity, all the available *cis*-prenyltransferase cDNA sequences were downloaded from the Genebank and a phylogenetic tree was constructed by including those sequences obtained from the five clones so as to group them based on their sequence similarity. Earlier reported *CPT* sequences were reported mainly from three clones viz; RRII 105, RRIM 600 and PR255. When all the *CPT*

DNA sequences including those from the current study were analysed together using clustalW three distinct groupings were observed. Surprisingly all the five *CPT* sequences from this study formed a single distinct cluster (Figure 3.42). From the data it is clear that there exists mainly three isoforms of this important gene of which the sequences from the current study formed a new group different from the earlier forms even though primers for PCR amplification was designed based on the available sequences then. All the three isoforms were observed in RRII 105 whereas two forms each were noted in RRIM 600 and PR 255.

3.8.3. SNP identification in *CPT* gene

Twenty-one bi-allelic SNPs were identified from the cloned fragment (1.463 kb) of *cis*-prenyltransferase gene from five *Hevea* clones resulting in an average of 1 SNP in every 70 bp. Sequence analysis revealed that the ORF was of 891 bp in size and the gene is devoid of any introns. The SNPs, their position and allelic status is shown in Table 3.31. More heterozygous SNP loci were identified in RRII 105 than any other clone followed by GT1 whereas RRII 118 and RRIC 52 were completely homozygous in all the loci analysed. Out of 21 SNPs, sixteen were from the coding regions and the remaining five from 5' and 3' UTRs respectively. Nine SNPs were categorised as transversions and 12 were transitions. Nucleotide diversity of the studied gene (Pi) was 0.00611. The distribution of 21 SNPs in the five clones based on their geographic region of origin is depicted as Venn diagram (Figure 3.43). Maximum number of SNPs was shared between Indian clones and Southeast Asian clones. Of the sixteen SNPs from coding region, eight were non-synonymous in nature.

The non-synonymous SNPs and the corresponding amino acid changes are given below.

Hbcispre703AT (GAT Aspartic acid ⟷ GAA Glutamic acid)
Hbcispre 714AT (ATG Methionine ⟷ AAG Lycine)
Hbcispre 851AC (CTC Leucine ⟷ ATC Isoleucine)
Hbcispre 971AG (AGT Serine ⟷ GGT Glycine)
Hbcispre 987CT (ACT Threonine ⟷ ATT Isoleucine)
Hbcispre 1145CT (TAC Tyrosine ⟷ CAC Histidine)
Hbcispre 1187GT (TCC Serine ⟷ GCC Alanine)
Hbcispre1200GT (TTC Phenyl alanine ⟷ TGC Cystene)

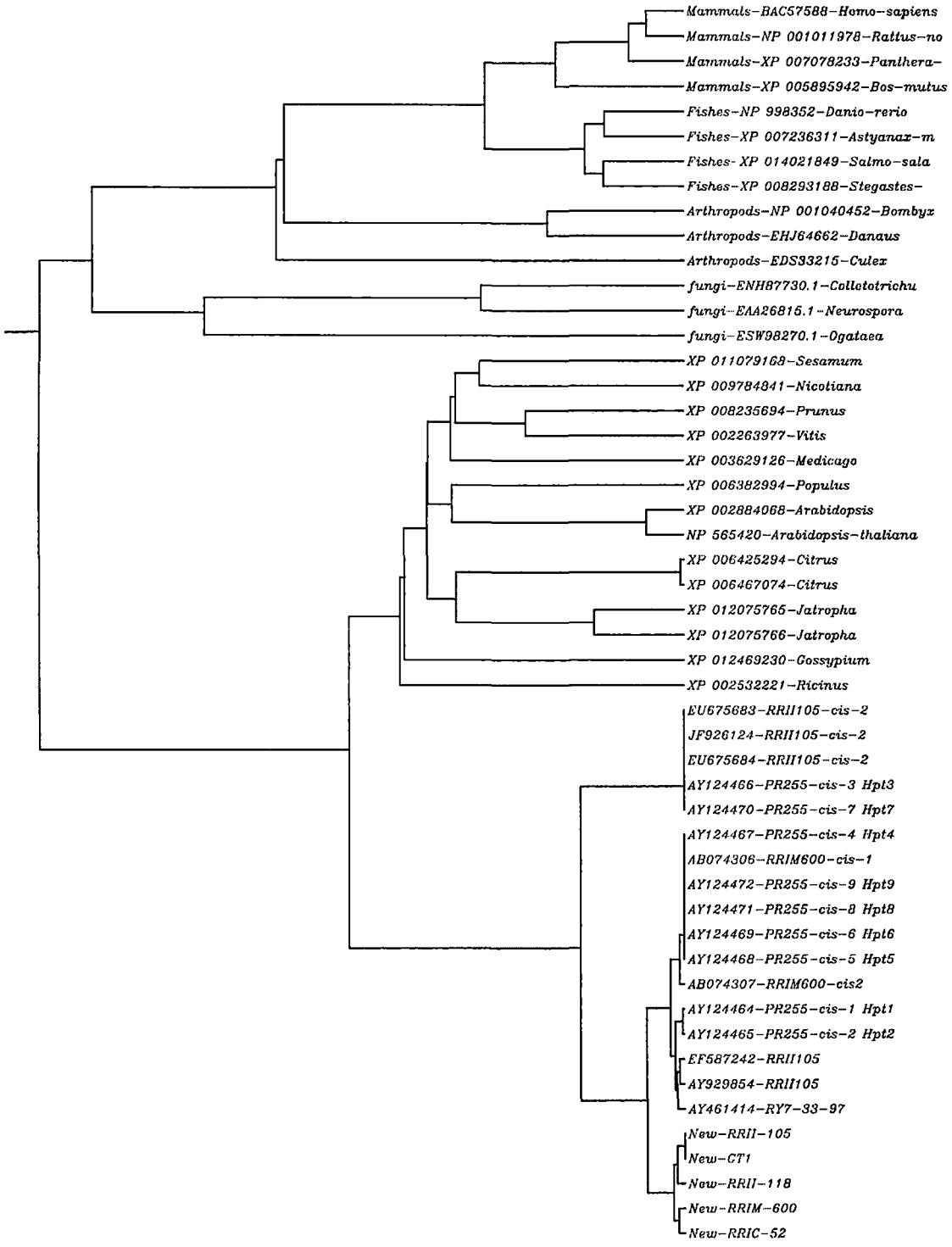


Figure 3.41 Phylogenetic tree based on cis-prenyltransferase sequence.
Phylogenetic tree constructed using the protein sequences of cis-prenyltransferase from selected plant species and representative organisms belonging to major domains of life.

3.8.4. Haplotype structuring of full-length *CPT* gene

Haplotype analysis using 17 SNPs with complete phase information yielded eight haplotypes from the five genotypes with a haplotype diversity, Hd: 0.9333. The eight haplotypes and their respective frequencies in the five clones are shown in Table 3.33. The following eight intragenic recombination events were also predicted between the base positions: (631-703) (714-736) (736-760) (760-811) (811-851) (859-971) (1145-1187) and (1200-1312).

A phylogram was constructed based on the haplotypes reconstructed using the the 17 SNPs within the *cis*-prenyltransferase gene (Figure 3.42). Hap_5 was distinct from the others whereas Hap_2, Hap_3, Hap_4 and Hap_8 were related and formed a single cluster. Similarly Hap_1, Hap_6 and Hap_7 appeared to be related and seems to evolve together.

Comparative distribution study of non-synonymous SNPs was not carried out for this gene due to the high sequence dissimilarity of *Hevea cis*-prenyltranferases with that of other plant species.

All the sequences from the eight rubber biosynthesis genes from five clones were submitted to NCBI Genbank database. Details of the submitted sequences are provided under Database submissions head at the end of the thesis

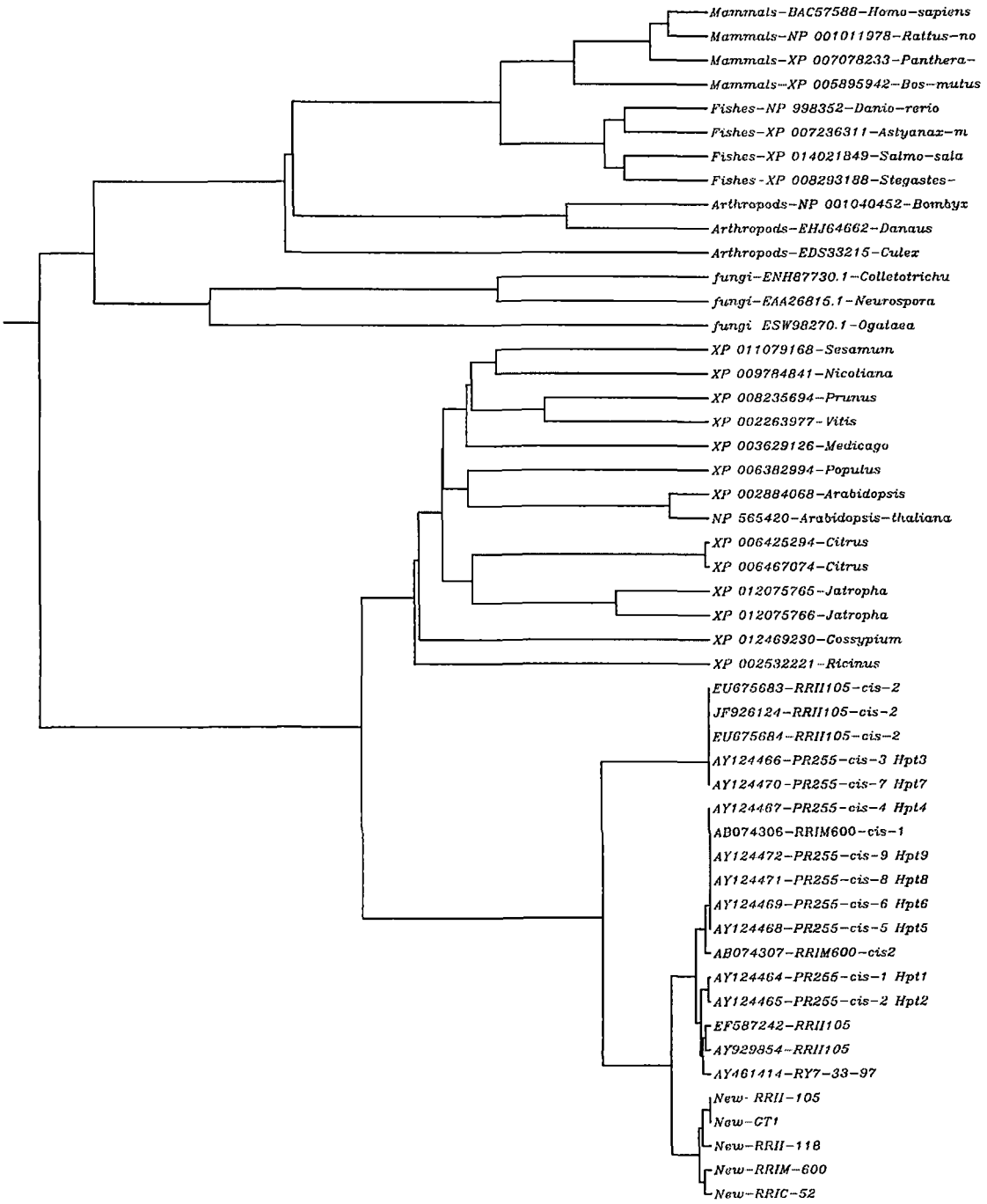


Figure 3.41 Phylogenetic tree based on cis-prenyltransferase sequence.
Phylogenetic tree constructed using the protein sequences of cis-prenyltransferase from selected plant species and representative organisms belonging to major domains of life.

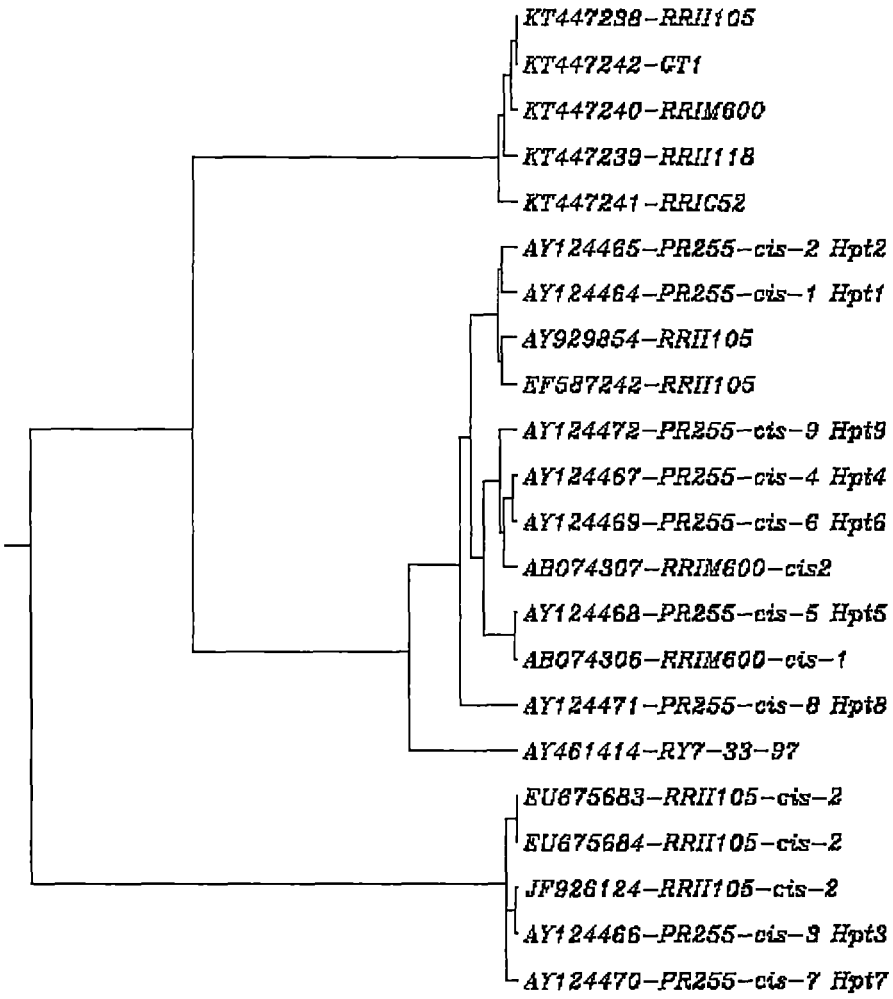


Figure 3.42. **Phylogenetic tree based on available *Hevea* cis-prenyltransferase DNA sequences. DNA sequences of *CPTs* available in public databases and haplotypes identified from the current study were used for the tree construction.**

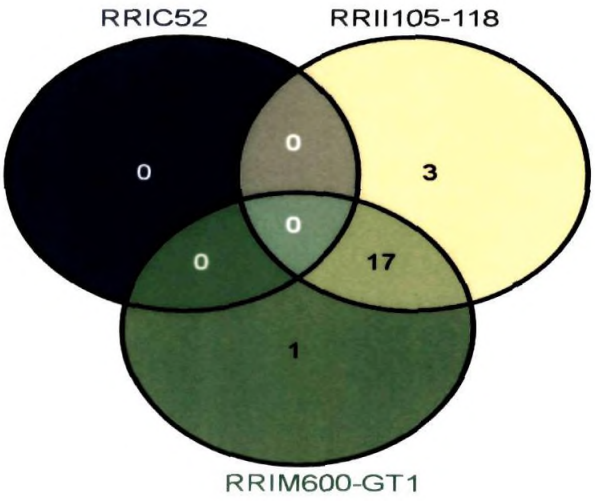


Figure 3.43. Distribution of *cis*-prenyltransferase SNPs across the three groups *viz.* Indian (RRII 105 & RRII 118) Sri Lankan (RRIC 52) and Southeast Asian (RRIM 600 and GT1)

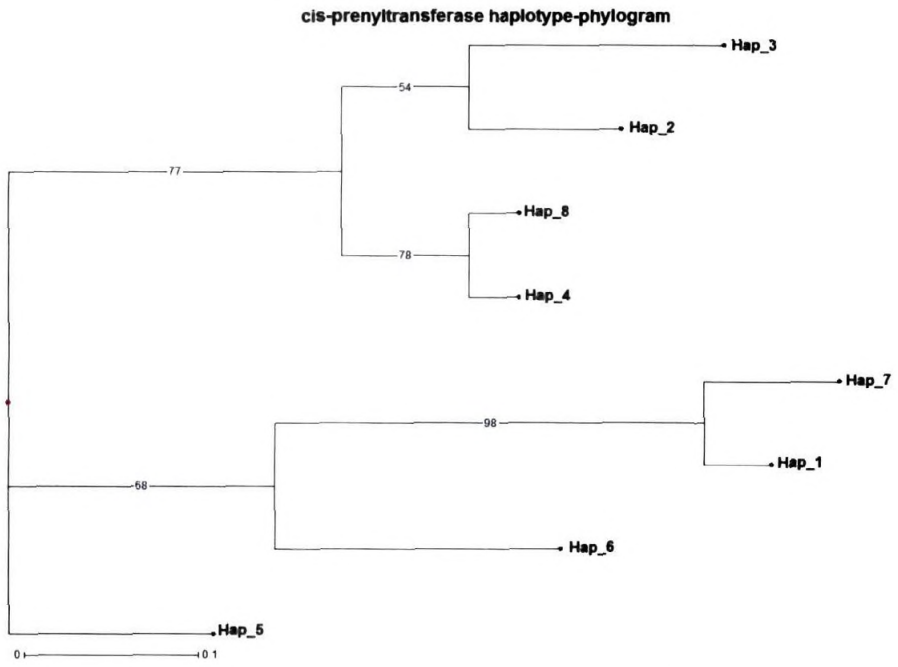


Figure 3.44. Phylogram (Neighbour joining tree) based on the haplotypes reconstructed using the 17 SNPs in *CPT* gene.

Table 3.31. SNPs obtained with their allelic status from the full-length genomic sequence of *CPT* gene.

Genotype	Respective base position of each locus																			
	169	550	619	631	703	714	736	760	811	851	859	971	987	1036	1145	1187	1200	1312	1349	1366	1438
RRII105	A/A	T/T	T/T	C/C	A/T	A/T	A/T	A/G	A/A	AC	A/G	A/G	C/T	AC	C/T	G/T	T/T	C/T	GC	C/T	C/T
RRII118	G/G	C/C	C/C	T/T	T/T	A/A	T/T	G/G	T/T	C/C	G/G	A/A	C/C	A/A	T/T	T/T	T/T	T/T	G/G	C/C	T/T
RRIM600	G/G	T/T	T/T	T/T	A/A	T/T	A/A	A/A	T/T	A/A	A/A	A/G	C/T	A/A	C/T	G/T	G/T	T/T	G/G	C/C	T/T
RRIC52	G/G	C/C	T/T	T/T	A/A	T/T	A/A	G/G	T/T	A/A	A/A	G/G	T/T	A/A	C/C	G/G	G/G	C/C	C/C	T/T	C/C
GT1	A/A	T/T	T/T	C/C	A/T	A/T	A/T	A/A	A/A	AC	A/G	A/G	C/T	AC	C/T	G/T	T/T	C/T	GC	C/T	C/T

Table 3.32. Characteristics of the 21 SNPs in *CPT* gene

SNP	Position consensus	Variation	SNP frequency	Nb readable individus	Major allele	Minor allele	Homo-zygotes major allele	Homo-zygotes minor allele	Hetero-zygotes
1	169	[A/G]	40.0 %	5	G	A	NA	NA	NA
2	550	[T/C]	40.0 %	5	T	C	NA	NA	NA
3	619	[T/C]	20.0 %	5	T	C	NA	NA	NA
4	631	[T/C]	40.0 %	5	T	C	NA	NA	NA
5	703	[A/T]	40.0 %	5	A	T	2	1	2
6	714	[A/T]	40.0 %	5	T	A	2	1	2
7	736	[A/T]	40.0 %	5	A	T	2	1	2
8	760	[A/G]	50.0 %	5	A	G	2	2	1
9	811	[A/T]	40.0 %	5	T	A	3	2	0
10	851	[A/C]	40.0 %	5	A	C	2	1	2
11	859	[A/G]	40.0 %	5	A	G	2	1	2
12	971	[A/G]	50.0 %	5	A	G	1	1	3
13	987	[T/C]	50.0 %	5	T	C	1	1	3
14	1036	[A/C]	20.0 %	5	A	C	3	0	2
15	1145	[T/C]	50.0 %	5	T	C	1	1	3
16	1187	[T/G]	50.0 %	5	T	G	1	1	3
17	1200	[T/G]	30.0 %	5	T	G	3	1	1
18	1312	[T/C]	40.0 %	5	T	C	2	1	2
19	1349	[C/G]	40.0 %	5	G	C	2	1	2
20	1367	[T/C]	40.0 %	5	C	T	2	1	2
21	1439	[T/C]	40.0 %	5	T	C	2	1	2

Table 3.33. *CPT* gene haplotypes with their respective allele frequency in each clone.

Sl No	Haplotype	Seq	Haplotype Frequency (%)	Haplotype frequency genotypes
1	Hap_1	TATAAAAGTCCGTCCTC	10	[RRII 105-a]
2	Hap_2	ATAGACGACATTTTGCT	10	[RRII 105-b]
3	Hap_3	TATGTCGACATTTTGCT	20	[RRII 118-a RRII 118-b]
4	Hap_4	ATAATAAACATTTTGCT	10	[RRIM 600-a]
5	Hap_5	ATAATAAGTACGGTGCT	10	[RRIM 600-b]
6	Hap_6	ATAGTAAGTACGGCCTC	20	[RRIC 52-a RRIC 52-b]
7	Hap_7	TATAACGGTCCGTCCTC	10	[GT1-a]
8	Hap_8	ATAAAAAACATTTTGCT	10	[GT1-b]

*"a" and "b" represent the two homologus chromosomes of the diploid *Hevea brasiliensis*

3.9. Comparative analysis of SNP number and frequency in the eight rubber biosynthesis genes

A comparison of the SNP number and frequency in the eight rubber biosynthesis genes revealed that the upstream genes like *HMGS, HMGR. MVK* and *PMVK* are highly conserved. The genes which are functionally more specific to *Hevea* like *REF* and *CPT* were the least conserved genes with very high SNP frequency rates. *REF* had the highest number of SNPs, highest SNP frequency as well as highest number of non-synonymous SNPs. The initiator molecule synthesising genes like FDPS and GGDPS were moderately conserved. Interestingly *FDPS* and *HMGR* were the only two genes were no non synonymous SNPs were identified. Additionally *FDPS* also had the highest number of SNPs from non-coding regions. This is in addition to the 35 SNPs reported from *FDPS* first intron which also harbours big indels associated with retro-elements. Table 3.34 shows the gene wise comparison of SNPs identified.

3.10. Clone wise comparative analysis of allele homozygosity in eight rubber biosynthesis genes

A comparative analysis of the Homozygosity/Heterozygosity of all SNP loci under study in each of the eight genes in five clones was done to identify the existence of possible correlation between Homozygosity/Heterozygosity of each gene and its gene expression or latex yield. The Table 3.35 shows the percentage homozygosity of each gene in the five clones and its average. From the data it is evident that overall homozygosity level was more in the clone RRII 118 followed by RRIM 600. Out of the eight genes, four genes (two upper mevalonate and two lower mevalonate genes) showed 100% homozygosity in RRII 118. Though RRIM 600 was the second highest, cent percent homozygosity was not seen in any gene. Gene wise figures indicate that *HMGR* is more homozygous than the other seven genes (fully homozygous in three clone). Higher homozygosity in the genes responsible for the synthesis of the most crucial initiator molecules (*FDPS & GGDPS*) was noted in RRII 105 followed by RRIM 600, the comparatively high yielders among the five clones under study. Interestingly, RRII105 which had 100% homozygous *FDPS* gene loci was also the clone with highest *FDPS* gene expression levels.

Table 3.34. Gene wise comparison of SNP frequency and type in eight rubber-biosynthesis genes

Sl No	Gene	Total No of SNPs	No of Non-coding/ Intronic SNPs	No of Coding /exonic SNPs	No of synonymous SNPs	No of non-synonymous SNPs	SNP frequency	SNP Frequency per 1000bp
1	HMGS	20	13	7	1	6	1/272	3.7
2	HMGR	14	8	6	6	0	1/182	5.5
3	MVK	26	18	8	4	4	1/158	6.3
4	PMVK	22	16	6	4	2	1/410	2.4
5	FDPS	25	23	2	2	0	1/145	6.9
6	GGDPS	15	2	13	8	5	1/100	10
7	CISPRE	21	5	16	8	8	1/70	14.3
8	REF	32	15	17	2	15	1/53	18.9

Table 3.35. The % age homozygosity of each gene in the five

Sl NO	GENE	RRII 105 Homo-zygosity %	RRII 118 Homo-zygosity %	RRIM 600 Homo-zygosity %	RRIC 52 Homo-zygosity %	GT1 Homo-zygosity %
1	HMGS	60	78	65	70	55
2	HMGR	100	100	64	100	29
3	MVK	8	80	88	38	85
4	PMVK	30	100	91	27	91
5	FDPS	100	60	88	20	20
6	GGDPS	93	67	93	33	7
7	REF	100	100	9	6	9
8	CPT	29	100	76	100	33
10	Average Homo-zygosity %	65	86	71	49	41

3.11. SNP marker segregation analysis and linkage map integration of the rubber biosynthesis genes

SNP genotyping was done using SNP based HRM technique. One SNP marker each from *HMGS, FDPS, MVK, GGDPS, CPT* and two SNP markers from *PMVK* gene was used for genotyping a progeny population of 46 individuals derived from a cross between RRII 105 x RRII 118. Since no SNP polymorphisms were detected for the genes *HMGR* and *REF* between the parents RRII 105 and RRII 118, these two genes could not be considered for linkage map incorporation. Markers from the genes: *HMGS* (HMGS3059AG) , *MVK* (MVK2628AG), *CPT* (Hb cis-pre-1438CT) and *PMVK* (PMVK-indel, PMVK102CT, PMVK 1786CT) were specific for the female parent RRII 105 and were scored as "lm x ll" type. Simultaneously the SNP markers from the genes *GGDPS* (GGP741CT) and *FDPS* (FDP1380CT) were specific to the male parent RRII 118 and was scored as "nn x np" type. The segregation ratio of 1:1 was followed by all the allele specific markers except that from *PMVK* gene which showed significant deviation from the Mendelian inheritance pattern. The SNP marker used, with the alleles, genotypes of the parents and the Join Map scoring type is shown in the Table 3.34.

3.11.1. Segregation pattern of *HMGS* marker

The segregation analysis of the locus HbHMGS3059AG clearly showed the presence of two genotypes in the ratio 1:1 as expected. Based on the melting curve analysis, clear differentiation was obtained between the A/G and G/G genotypes Figure 3.45. It was evident from the figure that all the 46 progenies were grouped in to either A/G or G/G type along with the parental controls included in the analysis. Twenty four progenies were genotyped as A/G and the remaining 22 as G/G. As per Mendelian mode of inheritance, the expected allele frequency for the alleles G:A in the progeny was 3:1 and the expected genotype frequency at one loci with two alleles in the progeny of parental genotypes as described above, is 50% A/G and 50% G/G (1:1). Out of the 46 progenies, 24 had A/G and 22 had G/G allelic combinations. The expected and observed frequencies for the alleles and the genotypes along with the *chi*-square values are shown in Table 3.35.

3.11.2. Segregation pattern of *MVK* gene marker

The segregation analysis of the locus HbMVK2628AG clearly showed the presence of two genotypes in the ratio 1:1 as expected. Based on the melting curve analysis, clear differentiation was obtained between the A/G and G/G genotypes (Figure 3.46). It was evident from the figure that all the 46 progenies were grouped in to either A/G or G/G type along with the parental controls included in the analysis (RRII 105 (A/G) X RRII 118 (G/G). Twenty progenies were genotyped as A/G and the remaining 26 as G/G. As per Mendelian mode of inheritance, the expected frequency for the alleles A:G in the progeny was 3:1 and the expected genotype frequency at one loci with two alleles in the progeny of parental genotypes as described above, is 50% A/G and 50% G/G (1:1). Out of the 46 progenies, 20 had A/G and 26 had G/G allelic combinations. The expected and observed frequencies for the alleles and the genotypes along with the chi-square values are shown in Table 3.36.

3.11.3. Segregation pattern of *PMVK* gene markers

The segregation analysis of the locus HbPMVK102CT showed the presence of two genotypes (C/T & T/T). Based on the melting curve analysis, clear differentiation was obtained between the C/T and T/T genotypes (Figure 3.47) based on the differentiation of the parental controls included in the analysis (RRII 105 (C/T) X RRII 118 (T/T). As per Mendelian mode of inheritance, the expected frequency for the alleles C:T in the progeny was 3:1 and the expected genotype frequency at one loci with two alleles in the progeny of parental genotypes as described above, is 50% A/G and 50% G/G (1:1). But surprisingly the expected segregation ratio of 1:1 was not observed. Thirty four progenies were genotyped as C/T and the remaining 12 as T/T which is a significant deviation from the Mendelian segregation pattern. The expected and observed frequencies for the alleles and the genotypes along with the *chi*-square values are shown in Table 3.37. Since this locus showed significant deviation from the cut off value of 3.84, another locus HbPMVK1786 CT which is around 1.5 kb downstream was tested. For this markers also clear differentiation was obtained between the C/T and C/C genotypes based on the differentiation of the parental controls included in the analysis (RRII 105 (C/T) X RRII 118 (C/C) (Figure 3.48).

Like the previous locus, thirty four progenies were genotyped as C/T and the remaining 12 as C/C which is a significant deviation from the Mendelian segregation pattern. The expected and observed frequencies for the alleles and the genotypes along with the *chi*-square values are shown in Table 3.38.

3.11.4. Segregation pattern of *CPT* gene marker

The segregation analysis of the locus HbCispre1438CT clearly showed the presence of two genotypes in the ratio 1:1 as expected. Based on the melting curve analysis, clear differentiation was obtained between the C/T and T/T genotypes (Figure 3.49). It was evident from the figure that all the 46 progenies were grouped in to respective C/T or T/T type along with the parental controls included in the analysis (RRII 105 (C/T) X RRII 118 (/TT). Twenty six progenies were genotyped as C/T and the remaining 20 as T/T. As per Mendelian mode of inheritance, the expected frequency for the alleles C:T in the progeny was 3:1 and the expected genotype frequency at one loci with two alleles in the progeny of parental genotypes as described above, is 50% C/T and 50% T/T (1:1). The expected and observed frequencies for the alleles and the genotypes along with the *chi*-square values are shown in Table 3.39.

3.11.5. Segregation pattern of *FDPS* gene marker

The segregation analysis of the loci HbFDPS1380CT clearly shows the presence of two genotypes in the ratio 1:1 as expected. Based on the melting curve analysis, clear differentiation was obtained between the C/T and C/C genotypes (Figure 3.50).

As evident from the figure all the 46 progenies were grouped in to respective C/T or C/C type along with the parental controls included in the analysis (RRII 105 (C/C) X RRII 118 (C/T). Twenty five progenies were genotyped as C/T and the remaining 21 as C/C. As per Mendelian mode of inheritance, the expected frequency for the alleles C:T in the progeny was 3:1 and the expected genotype frequency at one loci with two alleles in the progeny of parental genotypes as described above, is 50% C/T and 50% C/C (1:1). The expected and observed frequencies for the alleles and the genotypes along with the *chi*-square values are shown in Table 3.40.

3.11.6. Segregation pattern of *GGDPS* gene marker

The segregation analysis of the loci HbGGDPS741CT clearly shows the presence of two genotypes in the ratio 1:1 as expected. Based on the melting curve analysis, clear differentiation was obtained between the C/T and C/C genotypes (Figure 3.51). As evident from the figure all the 46 progenies were grouped in to respective C/T or C/C type along with the parental controls included in the analysis (RRII 105 (C/C) X RRII 118 (C/T). Twenty six progenies were genotyped as C/T and the remaining 20 as C/C. As per Mendelian mode of inheritance, the expected frequency for the alleles C:T in the progeny was 3:1 and the expected genotype frequency at one loci with two alleles in the progeny of parental genotypes as described above, is 50% C/T and 50% C/C (1:1). The expected and observed frequencies for the alleles and the genotypes along with the chi-square values are shown in Table 3.41.

3.11.7. Integration of SNP markers in to an available linkage map in rubber using JoinMap

In the analysis with JoinMap, all the markers with different segregation ratios from the earlier study reported by Bini (2003) were taken into consideration using the Join Map software. Separate linkage maps were constructed for each parent following the pseudo-test cross model analysis (Grattapaglia and Sederoff, 1994). A total of 169 markers (164 old + 6 new markers) were utilized for the linkage map construction in RRII 105 and 147 markers (145 old +2 new markers) were utilized for genetic map construction in RRII 118. For RRII 105 out of the five new markers, four were SNP markers and one STS marker (HbPMVK indel). Since the previous map was constructed using segregation data from 61 individuals the current analysis was modified in such a way that the marker data from the extra 15 individuals were considered as missing data.

In the linkage map of RRII 105, HbMVK2628AG and HbPMVK102CT formed linkage with the existing markers. The HbMVK2628AG marker got integrated to the LG-18 linkage group of the earlier reported Map (Bini., 2013). As expected the marker HbMVK2628AG got tightly linked to the earlier reported SNP marker MKF197C from the same gene. Similarly the SSR markers which were linked in the earlier map

remained linked in the same group. On the contrary the marker HbPMVK102CT and the large indel marker (HbPMVKIndel) from the same gene formed a new linkage group along with an earlier unlinked AFLP marker.

In the linkage map of RRII 118, HbFDPS1380CT from FDPS gene and HbGGDP741CT from GGDPS gene formed linkage groups with the existing markers. FDPS marker got integrated to the LG-3 linkage group of the earlier reported Map of RRII 118 (Bini., 2013). As expected all the markers which were present in the previous LG-3 group were present in the new group also. Moreover the marker HbFDPS1380CT was found linked to the intron1 STS marker from the same gene. Surprisingly six new markers (four RAPD markers, one AFLP marker and one SSR marker) which were not linked earlier got integrated to the new map. On the contrary the marker HbGGDPS741CT formed a new linkage group along with an earlier reported SNP marker from the same gene (GGDPF88). It is interesting to note that the GGDPF88 marker was unlinked in the earlier map and with the addition of the new marker HbGGDPS741CT, a new group was created by integrating one SSR marker and one RAPD marker which were not linked earlier. Figure 3.52 shows the linkage groups with the above markers integrated. The markers HbHMGS3059AG for HMGS, Hbcis-pre-1438CT for cis prenyltransferase and HbPMVK1786CT remained unlinked.

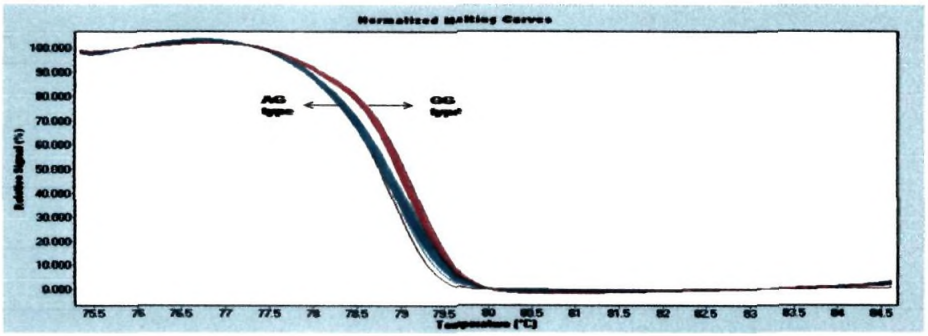


Figure 3.45. Melting curve variation shown by A/G and G/G genotypes based on HbHMGS3059AG marker. Separation of the two possible genotypes based on the melting curve analysis of a segregated progeny population derived from the cross RRII 105 (A/G) x RRII 118(G/G)

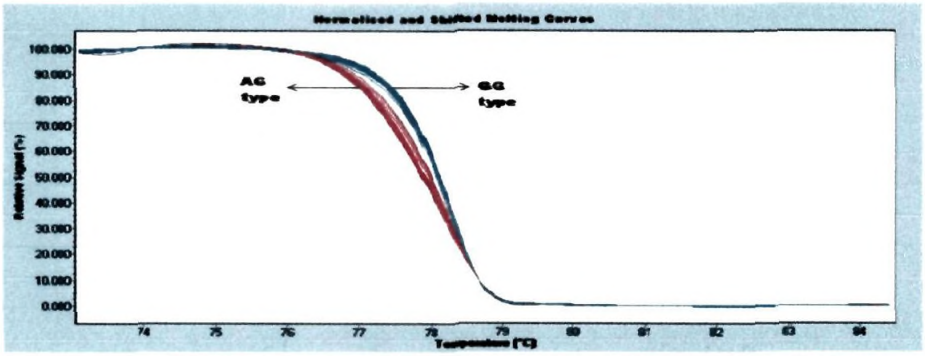


Figure 3.46. Melting curve variation of A/G and G/G genotypes based on HbMVK2628AG marker. Separation of two possible genotypes based on the melting curve analysis of a segregated progeny population derived from the cross RRII 105 (A/G) x RRII 118(G/G)

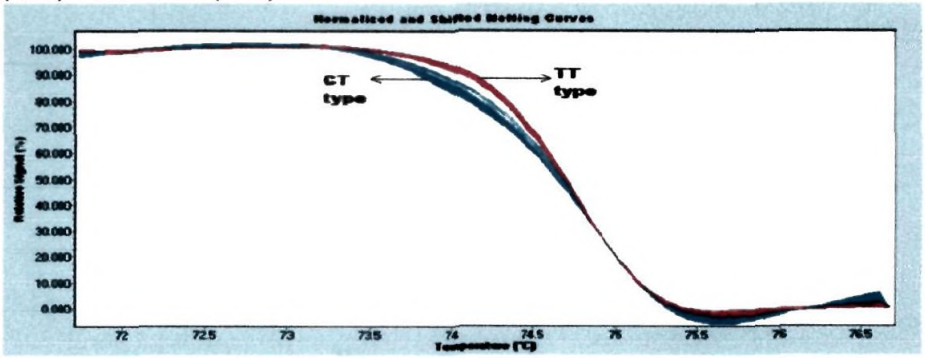


Figure 3.47. Melting curve variation of C/T and T/T genotypes based on HbPMVK102CT marker. Separation of two possible genotypes based on the melting curve analysis of a segregated progeny population derived from the cross RRII 105 (C/T) x RRII 118(T/T)

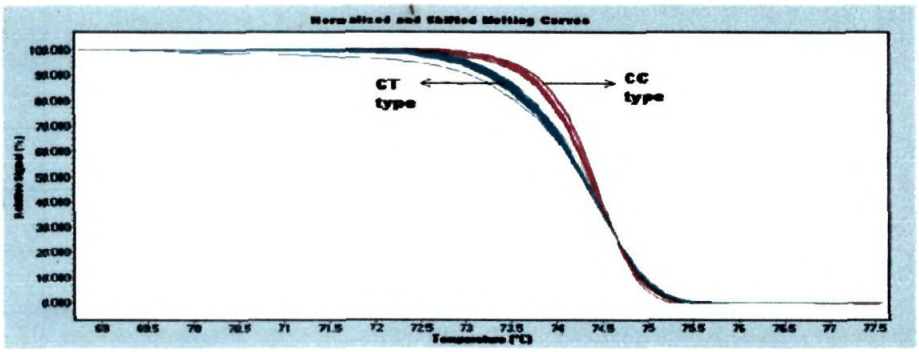


Figure 3.48. Melting curve variation of C/T and C/C genotypes based on HbPMVK1786CT marker. Separation of two possible genotypes based on the melting curve analysis of a segregated progeny population derived from the cross RRII 105 *(C/T)* x *RRII 118(C/C)*

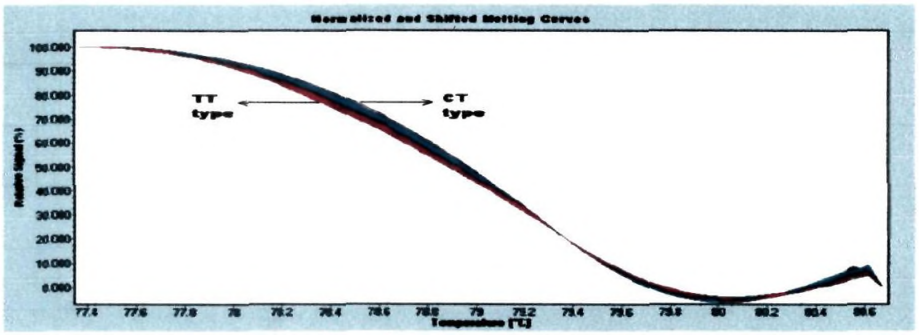


Figure 3.49. Melting curve variation of C/T and T/T genotypes based on HbCispre1438CT marker. Separation of two possible genotypes based on the melting curve analysis of a segregated progeny population derived from the cross RRII 105 (C/T) x RRII 118(T/T)

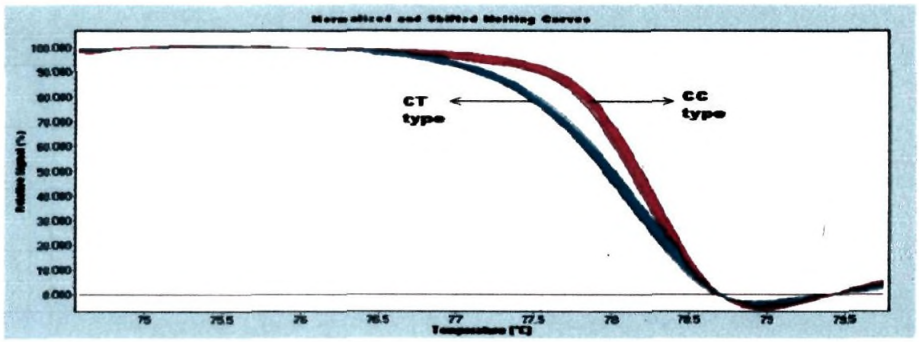


Figure 3.50. Melting curve variation of C/T and C/C genotypes based on HbFDPS1380CT marker. Separation of two possible genotypes based on the melting curve analysis of a segregated progeny population derived from the cross RRII 105 (C/C) x RRII 118(C/T)

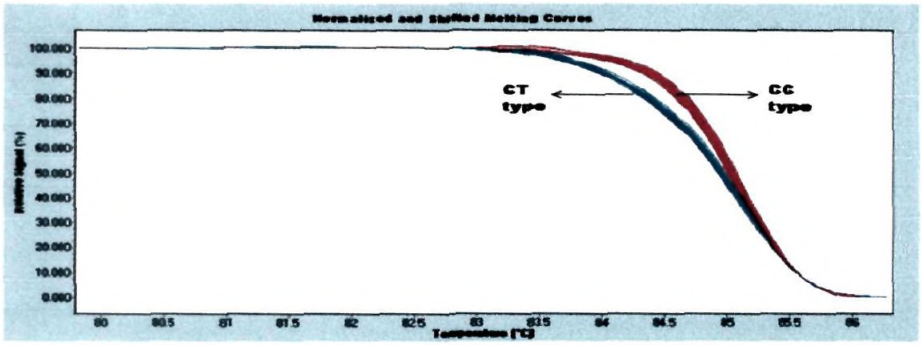


Figure 3.51. Melting curve variation of CT and CC genotypes based on HbGGDSP741CT marker. Separation of two possible based on the melting curve analysis of a segregated progeny population derived from the cross RRII 105 (C/C) x *RRII 118(C/T)*

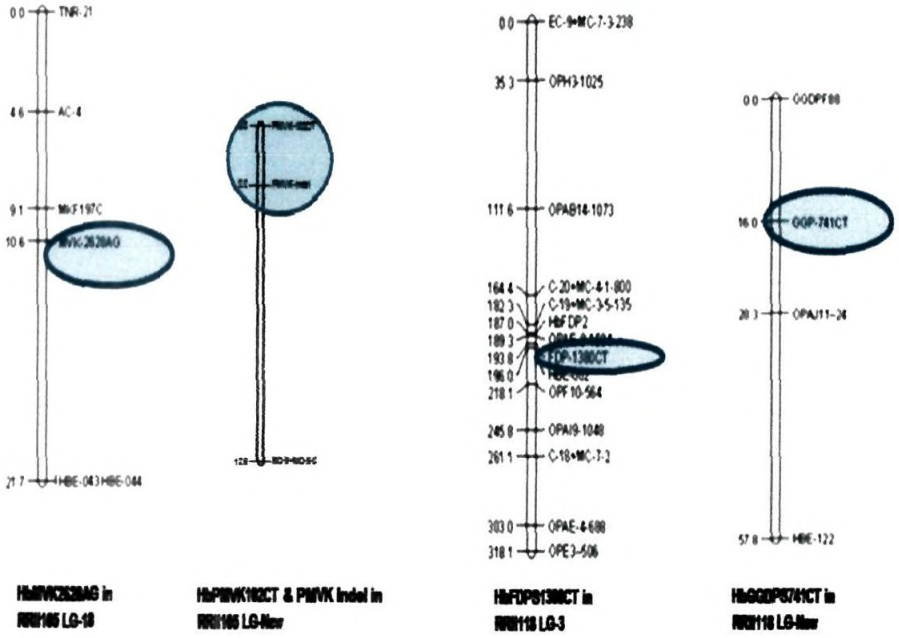


Figure 3.52. Linkage groups of RRII 105 and RRII 118 with the SNP markers from *MVK, PMVK, FDPS* and *GGDPS* genes. LG-number indicates the respective linkage groups of the linkage map generated by Bini (2013).

Table 3.36. SNP markers used for segregation analysis with their respective alleles and scoring format

Sl No	Gene	SNP name	SNP Alleles	Parental genotypes RRII105 X RRII118	Join Map Scoring type
1	HMGS	HMGS3059A/G	A/G	A/G X G/G	<lm x ll>
2	FDPS	FDPS1380C/T	C/T	C/C X C/T	<nnxnp>
3	CPT	cis-pre-1438C/T	C/T	C/T X T/T	<lm x ll>
4	GGDPS	GGDP741C/T	C/T	C/C X C/T	<nnxnp>
5	MVK	MVK2628A/G	A/G	A/G X G/G	<lm x ll>
6	PMVK	PMVK102C/T	C/T	C/TX T/T	<lm x ll>
7	PMVK	PMVK1786C/T	C/T	C/T X C/C	<lm x ll>

Table 3.37. Allele and genotype frequencies of *HMGS* marker HbHMGS3059AG in a progeny population

Parents (RRII-105 (A/ G) x RRII-118 (G/G)

Genotype frequency

Genotypes	Expected frequency	Observed frequency	Contribution to chi- square
AG	23(50%)	24 (52.1%)	0.04
GG	23 (50%	22 (47.8%)	0.05
Total			0.09 (not significant)

Allele frequency

Alleles	Expected frequency	Observed frequency	Contribution to chi-square
A	23 (25%)	24 (26.1%)	0.04
G	69 (75%)	68 (73.9%)	0.01
Total			0.05 (not significant)

Allele and genotype frequencies of SNP HbHMGS3059AG in a full sib family progeny of size 46. The calculated chisqaure value is 0.09 which is less than 3.84, the chisquare distribution table value. Therefore the genotypes "A/G" and "G/G" showed no statistically significant deviation from Mendels law of inheritance and the segregation follows the 1:1 ratio.

Table 3.38. Allele and genotype frequencies of *MVK* marker HbMVK2628AG n a progeny population

Parents (RRII-105 (A/ G) x RRII-118 (G/G)

Genotype frequency

Genotypes	Expected frequency	Observed frequency	Contribution to chi- square
AG	23(50%)	20 (43.4%)	0.39
GG	23 (50%	26 (56.6%)	0.39
Total			
			0.78
			(not significant)

Allele frequency

Alleles	Expected frequency	Observed frequency	Contribution to chi-square
A	23 (25%)	20 (21.8%)	0.39
G	69 (75%	72 (78.2%)	0.13
Total			0.52
			(not significant)

Allele and genotype frequencies of SNP HbMVK2628(A/G)) in a full sib family progeny of size 46. The calculated chisqaure value at p(>0.05) and df=1, is 0.78 which is less than 3.84 the chisquare distribution table value. Therefore the genotypes "AG" and "GG" showed no statistically significant deviation from Mendels law of inheritance and the segregation follows the 1:1 ratio.

Table 3.39. Allele and genotype frequencies of *PMVK* marker HbPMVK102CT and in a progeny population.

Parents (RRII-105 (C/ T) x RRII-118 (T/T)

Genotype frequency

Genotypes	Expected frequency	Observed frequency	Contribution to chi- square
CT	23 (50%)	34 (73.9%)	5.26
TT	23 (50%)	12 (26%)	5.26
Total			10.52 (Significant)

Allele frequency

Alleles	Expected frequency	Observed frequency	Contribution to chi-square
C	23(25%)	34 (36.95%)	5.26
T	69 (75%	58 (63.05%)	1.75
Total			7.01 (Significant)

Allele and genotype frequencies of SNP HbPMVK102CT in a full sib family progeny of size 46. The calculated chisqaure value at p(>0.05) and df=1, is 10.52 which is higher than 3.84 the chisquare distribution table value. Therefore the genotypes "CT" and "TT" showed statistically significant deviation from Mendels law of inheritance and the segregation does not follows the 1:1 ratio

Table 3.40. Allele and genotype frequencies of *PMVK* marker HbPMVK1786CT in a progeny population

Parents (RRII-105 (C/ T) x RRII-118 (C/C)

Genotype frequency

Genotypes	Expected frequency	Observed frequency	Contribution to chi- square
CT	23 (50%)	34 (73.9%)	5.26
CC	23 (50%)	12 (26%)	5.26
Total			10.52 (Significant)

Allele frequency

Alleles	Expected frequency	Observed frequency	Contribution to chi-square
T	23(25%)	12 (13.00%)	5.26
C	69 (75%)	80 (87.0%)	1.75
Total			7.01 (Significant)

Allele and genotype frequencies of SNP HbPMVK1786 CT in a full sib family progeny of size 46. The calculated chisqaure value at p(>0.05) and df=1, is 10.52 which is higher than 3.84 the chisquare distribution table value. Therefore the genotypes "CT" and "CC" showed statistically significant deviation from Mendels law of inheritance and the segregation does not follows the 1:1 ratio.

Table 3.41. Allele and genotype frequencies of *CPT* marker HbCispre1438CT in a progeny population

Parents (RRII-105 (C/ T) x RRII-118 (T/T)

Genotype frequency

Genotype	Expected frequency	Observed frequency	Contribution to chi- square
CT	23 (50%)	26 (56.6%)	0.39
TT	23 (50%)	20 (43.4%)	0.39
Total			0.78 (not significant)

Allele frequency

Allele	Expected frequency	Observed frequency	Contribution to chi-square
C	23(25%)	26 (28.3%)	0.39
T	69 (75%	66 (71.7%)	0.13
Total			0.52 (not significant)

Allele and genotype frequencies of SNP HbCispre1438CT in a full sib family progeny of size 46. The calculated chisqaure value at p(>0.05) and df=1, is 0.78 which is less than 3.84 which is the chisquare distribution table value. Therefore the genotypes "CT" and "TT" showed no statistically significant deviation from Mendels law of inheritance and the segregation follows the 1:1 ratio

Table 3.42. Allele and genotype frequencies of *FDPS* marker HbFDPS1380CT in a progeny population.

Parents (RRII-105 (C/ C) x RRII-118 (C/T)

Genotype frequency

Genotypes	Expected frequency	Observed frequency	Contribution to chi- square
CT	23 (50%)	25 (54.3%)	0.17
CC	23 (50%)	21 (45.6%)	0.17
Total			0.34 (not significant)

Allele frequency

Allele	Expected frequency	Observed frequency	Contribution to chi-square
T	23(25%)	25 (27.1%)	0.17
C	69 (75%	67 (72.8%)	0.05
Total			0.22 (not significant)

Allele and genotype frequencies of SNP HbFDPS1380CT in a full sib family progeny of size 46. The calculated chisqaure value at p(>0.05) and df=1, is 0.34 which is less than 3.84 which is the chisquare distribution table value.. Therefore the genotypes "CT" and "CC" showed no statistically significant deviation from Mendels law of inheritance and the segregation follows the 1:1 ratio.

Table 3.43. Allele and genotype frequencies of *GGDPS* marker HbGGDPS741CT in a progeny population

Parents (RRII-105 (C/ C) x RRII-118 (C/T)

Genotype frequency

Genotype	Expected frequency	Observed frequency	Contribution to chi- square
CT	23 (50%)	26 (56.6%)	0.39
CC	23 (50%)	20 (43.4%)	0.39
Total			0.78 (not significant)

Allele frequency

Alleles	Expected frequency	Observed frequency	Contribution to chi-square
T	23(25%)	26 (28.3%)	0.39
C	69 (75%	66 (71.7%)	0.13
Total			0.52 (not significant)

Allele and genotype frequencies of SNP HbGGDPS741CT in a full sib family progeny of size 46. The calculated chisqaure value at p(>0.05) and df=1, is 0.78 which is less than 3.84 which is the chisquare distribution table value. Therefore the genotypes "CT" and "CC" showed no statistically significant deviation from Mendels law of inheritance and the segregation follows the 1:1 ratio.

3.12. Relative quantification of Gene Expression

qPCR studies of rubber biosynthesis genes was performed to identify possible association between SNP haplotypes and gene expression in terms of transcript abundance in the latex of five clones, RRII 105, RRII 118, RRIM 600, RRIC 52 and GT1. Relative gene expression of the eight rubber biosynthesis genes in five clones is depicted in the Figures 3.53 to 3.60.

HMGS gene: Gene expression was significantly higher in RRII 105 than the other clones. RRII 105 had a unique combination of Hap_1 & Hap_2 which no other clone possessed. Moreover Hap_1 was unique to RRII 105. Since RRII 118 also had Hap_2 it was evident that Hap_2 alone did not have any influence on *HMGS* expression.

HMGR gene: RRII 105, RRII 118 and RRIC 52 had comparable levels of gene expression. Both RRII 105 and RRII 118 had the same haplotype (Hap_1) in both alleles whereas RRIC 52 has Hap_4 in both alleles. RRIM 600 showed comparatively lower expression levels (less than half that of the above three). GT1 showed significantly lower expression and harboured Hap_4 and Hap_5.

MVK gene: Both RRII 105 and RRII 118 had comparable expression rates and they were significantly higher than the expression in the other three clones. GT1 showed the least expression among all the five. The combinations Hap_1 & Hap_2 present in RRII 105 and Hap_3 & Hap_4 present in RRII 118 may have some influence on expression. But the Haplotypes Hap_2 & Hap_3 alone or in any other combination as seen in GT1 (Hap_2 & Hap_3) may have a negative effect on expression.

PMVK gene: Highest gene expression was observed in RRII 118 followed by RRII 105 whereas the expression levels in the other three clones were significantly low. GT1 expression was the lowest. RRII 118 had Hap_2 in homozygous state while RRII 105 had Hap_1 & Hap_2. The lowest expressing clone GT1 had the combination of Hap_2 & Hap_5.

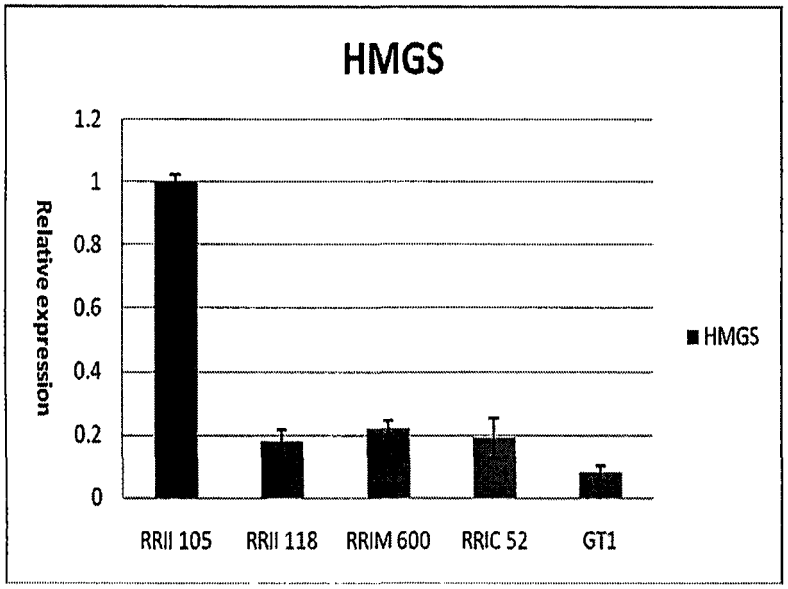
FDPS gene: Gene expression was significantly higher in RRII 105 than the other clones. Significantly lower expression (Less than half that of RRII 105) was noted in RRIM 600, RRIC 52 and GT1. RRII 105 had Hap_1 in homozygous state whereas

RRIM 600 and GT1 had Hap_1 in one of their chromosome. RRII 118 and RRIC 52 had entirely different haplotypes in their chromosomes.

GGDPS gene : RRII 105 had comparatively higher gene expression rate than the other clones. RRIM 600 had less than half fold expression than that in RRII 105. As in the case of other genes, GT1 had the lowest expression profile. Despite having the same allelic combination (Hap_1 & Hap_2) there was significance difference in GGDPS expression in RRII 105 and RRIM 600.

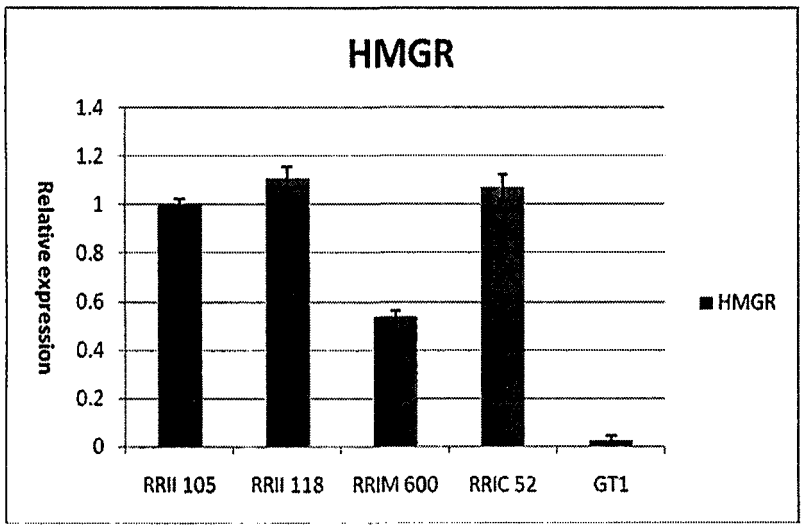
REF gene: RRII 118 had significantly higher expression than the others. Compared to RRII 118, expression level was much less in RRII 105 though they shared the same haplotype (Hap_1) in both their chromosomes. Hap_1 was also identified in the least expressing clone GT1.

CPT gene: RRII 105 had significantly higher cis-prenyltransferase gene expression than the other clones. As in the case of the above seven genes, GT1 had the least expression for CPT also. RRII 105 had a unique set of haplotypes (Hap_1 and Hap_2). Interestingly all the clones possessed their own unique haplotypes/haplotype combinations.



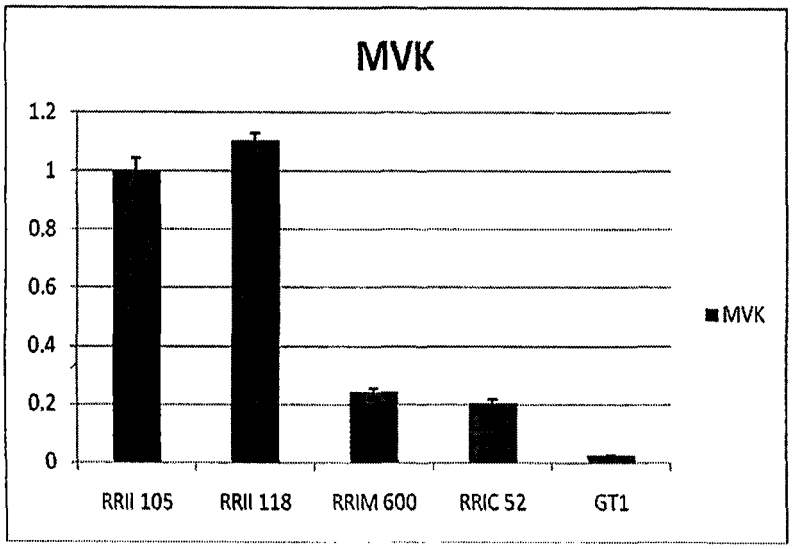
Clone	*HMGS* Haplotypes
RRII 105	Hap_1 & Hap_2
RRII 118	Hap_2 & Hap_3
RRIM 600	Hap_4 & Hap_5
RRIC 52	Hap_6 & Hap_7
GT1	Hap_5 & Hap_8

Figure 3.53. Graph showing the expression pattern of *HMGS* gene in the five clones. The expression rate was represented as the fold change in transcript level normalized to the Actin gene, relative to that in RRII 105 plants. Respective haplotypes identified in each clones is also shown.



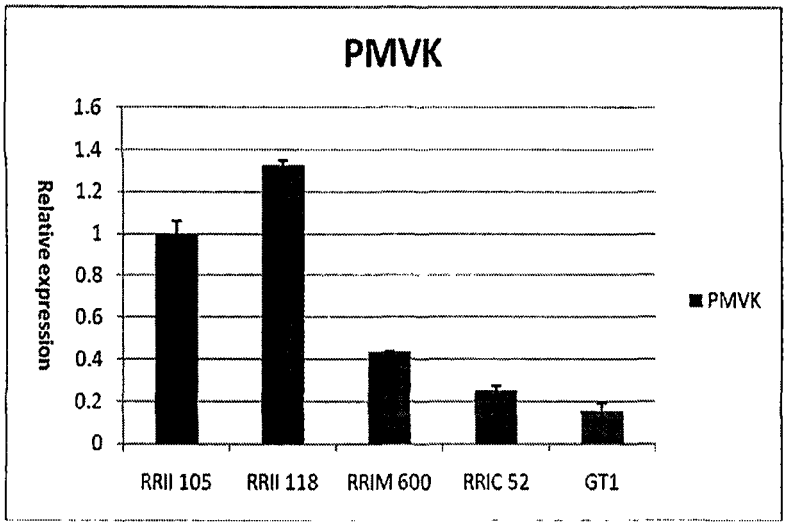
Clone	*HMGR* Haplotypes
RRII 105	Hap_1
RRII 118	Hap_1
RRIM 600	Hap_2 & Hap_3
RRIC 52	Hap_4
GT1	Hap_4 & Hap_5

Figure 3.54. Graph showing the expression pattern of *HMGR* gene in the five clones. The expression rate was represented as the fold change in transcript level normalized to the Actin gene, relative to that in RRII 105 plants. Respective haplotypes identified in each clones is also shown.



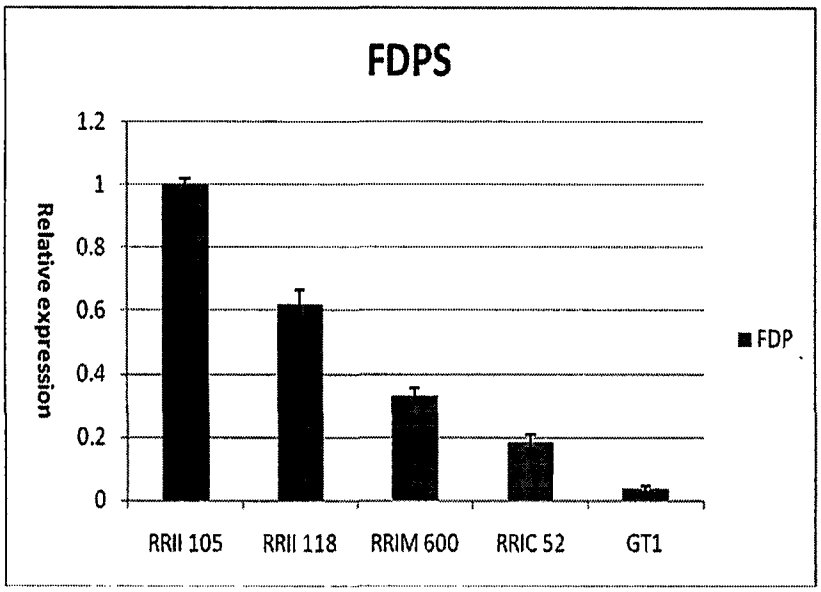
Clone	*MVK* Haplotypes
RRII 105	Hap_1 & Hap_2
RRII 118	Hap_3 & Hap_4
RRIM 600	Hap_5 & Hap_6
RRIC 52	Hap_7 & Hap_8
GT1	Hap_2 & Hap_3

Figure 3.55. Graph showing the expression pattern of *MVK* gene in the five clones. The expression rate was represented as the fold change in transcript level normalized to the Actin gene, relative to that in RRII 105 plants. Respective haplotypes identified in each clones is also shown.



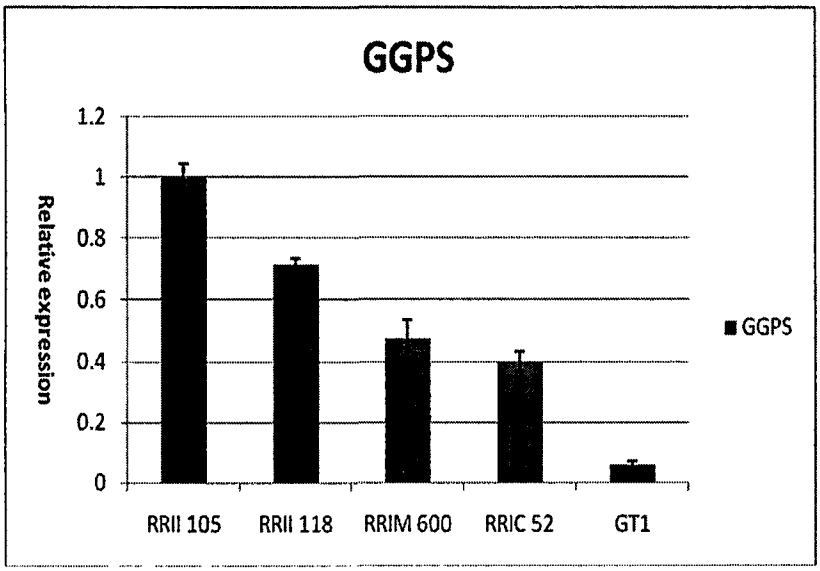
Clone	*PMVK* Haplotypes
RRII 105	Hap_1 & Hap_2
RRII 118	Hap_2
RRIM 600	Hap_2 & Hap_3
RRIC 52	Hap_1 & Hap_4
GT1	Hap_2 & Hap_5

Figure 3.56. Graph showing the expression pattern of *PMVK* gene in the five clones. The expression rate was represented as the fold change in transcript level normalized to the Actin gene, relative to that in RRII 105 plants. Respective haplotypes identified in each clones is also shown.



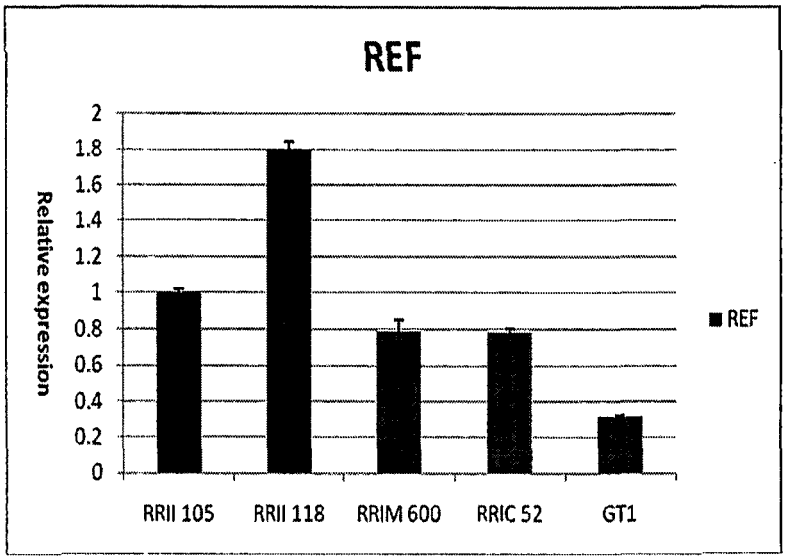
Clone	*FDPS* Haplotypes
RRII 105	Hap_1
RRII 118	Hap_2 & Hap_3
RRIM 600	Hap_1 & Hap_4
RRIC 52	Hap_5 & Hap_6
GT1	Hap_1 & Hap_7

Figure 3.57. Graph showing the expression pattern of *FDPS* gene in the five clones. The expression rate was represented as the fold change in transcript level normalized to the Actin gene, relative to that in RRII 105 plants. Respective haplotypes identified in each clones is also shown.



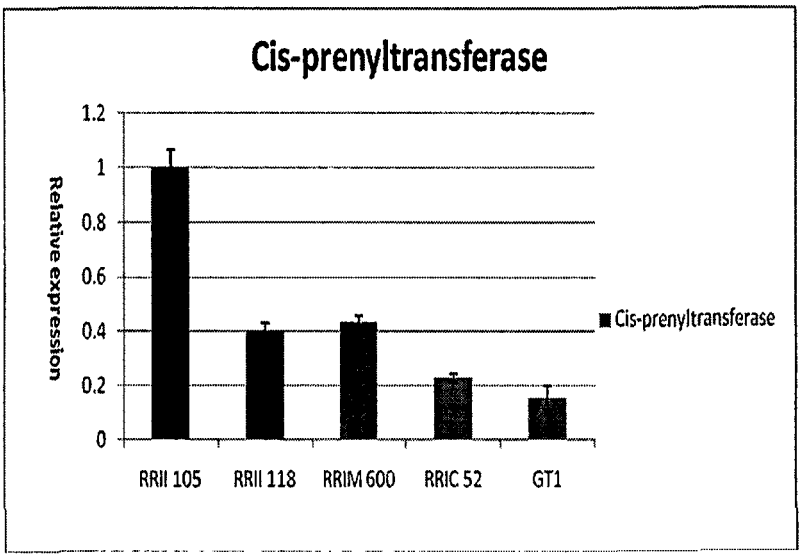
Clone	*GGDPS* Haplotypes
RRII 105	Hap_1 & Hap_2
RRII 118	Hap_1 & Hap_3
RRIM 600	Hap_1 & Hap_2
RRIC 52	Hap_4 & Hap_5
GT1	Hap_1 & Hap_5

Figure 3.58. Graph showing the expression pattern of *GGDPS* gene in the five clones. The expression rate was represented as the fold change in transcript level normalized to the Actin gene, relative to that in RRII 105 plants. Respective haplotypes identified in each clones is also shown.



Clone	*REF* Haplotypes
RRII 105	Hap_1
RRII 118	Hap_1
RRIM 600	Hap_1 & Hap_2
RRIC 52	Hap_3 & Hap_4
GT1	Hap_1 & Hap_5

Figure 3.59. Graph showing the expression pattern of *REF* gene in the five clones. The expression rate was represented as the fold change in transcript level normalized to the Actin gene, relative to that in RRII 105 plants. Respective haplotypes identified in each clones is also shown.



Clone	*CPT* Haplotypes
RRII 105	Hap_1 & Hap_2
RRII 118	Hap_3
RRIM 600	Hap_4 & Hap_5
RRIC 52	Hap_6
GT1	Hap_7 & Hap_8

Figure 3.60. Graph showing the expression pattern of *CPT* gene in the five clones. The expression rate was represented as the fold change in transcript level normalized to the Actin gene, relative to that in RRII 105 plants. Respective haplotypes identified in each clones is also shown.

Chapter 4

DISCUSSION

An important part of rubber research is concerned with creation and selection of new *Hevea* clones having higher productivity, better adaptability to different agro-climatic conditions, disease resistance etc. Conventional genetic analysis in *Hevea* is difficult because of its perennial nature, long breeding and selection cycles, and difficulties in raising F_2 progenies. Though several molecular markers like isozymes, RAPDs, RFLPs, AFLPs and SSRs have been used for studying the genetic variability in rubber the development of molecular markers for genetic improvement is comparatively difficult in rubber due to their narrow genetic base, polygenic nature of traits like yield, disease resistance etc. Therefore genetic markers from the genes responsible for the trait itself or in close association with them are required for genetic mapping and the development of effective markers for marker assisted selection (MAS). In this context Single nucleotide polymorphisms (SNPs) and indels assumes relevance due to their abundance in the genome and wide distribution pattern even in the vicinity of virtually every gene. SNP is an individual nucleotide base difference between two DNA sequences which can have a major impact on how the organism develops and responds to the environment. Furthermore they are evolutionarily stable, not changing significantly from generation to generation. SNPs provide an important source of molecular markers that are useful in high-resolution genetic mapping, map-based positional cloning, detection of marker-trait gene associations, phylogenetic analysis etc. Since mevalonate pathway is the key metabolic pathway associated with the synthesis of natural rubber, the major genes involved in this pathway in rubber were selected as the candidates for the identification of SNPs in the current study. Due to their close involvement in rubber biosynthesis it is anticipated that their gene sequence structure will have influence on their expression which may be linked to the quality and quantity of latex synthesised. Therefore detection of SNPs in the above candidate genes will make it an ideal marker for genetic analysis and association studies related to yield and rubber quality in *Hevea*.

4.1. Phylogenetic analysis

Since mevalonate pathway genes are universally present in almost all organisms, it is assumed that their occurrence and evolution over millions of years has led to several modifications in their sequence structure depending on the complexity of the organism (Sacchettini and Poulter, 1997). Being an important metabolic pathway leading to rubber biosynthesis in *Hevea brasiliensis,* attempts were made in the present study to understand the evolutionary origin and status of major MVA pathway genes in rubber using phylogenetic analysis based on available sequence databases.

Due to the highly dynamic nature of plant nuclear genomes, they vary considerably in structure and size leading to the evolution of new alleles in the micro level and new species at the macro level. Though universally present, the occurrence and evolution of mevalonate pathway genes over millions of years has led to several modifications in their sequence structure depending on the complexity of the organism. It is primarily these modifications which enabled the enzymes in this pathway to be involved in diverse vital functional activities in various organisms. For example, in *Arabidopsis thaliana* Hydroxymethyl-glutaryl-CoA synthase (HMGS) is involved in the development of reproductive organelles, fertility of pollen grains as well as in the production of phytohormones, whereas in *Brassica juncea HMGS* is involved in the development of flowers, seeds and seedlings, stress response etc. (Alex *et al.*, 2000; Nagegowda *et al.*, 2005; Ishiguro *et al.*, 2010). In Norway spruce and *Pinus* species, HMGS is mainly involved in the biosynthesis of terpenoid-based resins (Phillips *et al.*, 1999; Martin *et al.*, 2002). Therefore phylogenetic trees constructed using MVA pathway gene protein sequences from various organisms can be considered an indirect representation of their diverse functions, because the changes at genome level which occurred in the organisms over the years may be responsible for the functional differentiation of this pathway genes. The clustering, distribution and branching of each domain and the distance between them clearly depict the evolutionary ladder of life from lower simple organisms to higher complex forms. The clustering of similar type of organisms like members of eubacteria, primates, cereals etc., clearly states the relevance of these genes in phylogenetic studies and suggested the existence of a single common ancestor for each group.

Based on previous studies It is assumed that compartment-specific isoforms as in the case of mitochondrial and cytosolic isoforms of *HMGS* among eukaryotes have arisen relatively recently through gene duplication (Lange *et al.*, 2000). Focussed analysis of *Hevea HMGS* revealed that there exist several isoforms of *HMGS* with minor sequence variations (Sirinupong *et al.*, 2005). In the case of HMG-CoA reductase (*HMGR*) also clear compartmentalisation was observed between various groups of organisms (Li *et al.*, 2014). This may be attributed to the highly divergent architecture of the N-terminal region of this enzyme. On the contrary, the interlinking between the various groups may be due to the extensive sequence identity of the carboxyl-terminal region of this enzyme which contains the active site. The presence of multiple copies especially in plants also might have influenced the differentiation of species (Hampton *et al.*, 1996). The phylogenetic analysis by Li *et al.* (2014) revealed that the plant *HMGR*s were derived from one ancestor gene and finally developed into distinct groups of monocot and dicot plants. The presence of a unique cluster for the grass family members is a clear indication of the above finding. The minor variatios among the monocots may be due to species-specific gene duplications, caused mainly by segmental duplication. The *HMGR* family in flowering plants underwent species-specific expansions after the species diverged due to the requirement of various terpene compounds in their respective development. The same explanation may be applied to the variations observed among higher plant like *Ricinus communis, Vitis vinefera, Malus domestica* etc. The unique characteristics of the fossil plant *Ginko biloba* is also evident from the distance at which it was placed in the phylogenetic tree. Prior studies indicate that MVK and PMVK which catalyzed the last steps of the upper MVA pathway is poorly conserved across genomes (Wilding *et al.*, 2000). Despite this variation, the phylogenetic tree structure of both MVK and PMVK proteins was almost identical to that of the other upstream enzymes in the pathway.

The relationship of the organisms studied based on FDPS and GGDPS protein sequences suggests that there exist universally conserved as well as variable amino acid sequences of these two proteins in the organisms studied. The variations responsible for similar groupings are often known to contribute to the evolutionary mechanism resulting in the generation of new alleles. Such variation occurs mainly due to localized,

intra-chromosomal scrambling fragmentation which was proved to be a general mechanism for generating diversity in taxa other than genetic recombination (Kloeckner-Gruissem and Freeling, 1995). In the case of the *FDPS* gene in *H.brasiliensis,* point mutations like SNPs, major additions and deletions might have enhanced microevolution within the species enabling the selection of better adaptable clones. Similarly phylogenetic studies of GGDPS by Chen *et al.*, (1994) indicated that these enzymes have originated from a common progenitor (Ohnuma *et al.*, 1994). Moreover the enzymatic characters of archaeal GGDP synthase have suggested that archaeal GGDPS is near the origin (Chen and Poulter, 1994). However, the amino acid variations responsible for the product chain length difference between FDPS and GGDPS is still unknown.

Phylogenetic tree construction based on DNA and protein sequences of *cis*-prenyltransferase (*CPT*) and Rubber elongation factor (*REF*) gene revealed many interesting details which were not known earlier. Based on our preliminary analysis of *CPT* gene and protein sequences from public databases we arrived at the conclusion that there exist multiple forms of this gene coming under the gene family *cis*-prenyltransferases with various functions associated with them. Since the *CPTs* identified in bacterial species were so large and divergent in nature, a comparative analysis with *Hevea* sequences was beyond the ambit of the current study. Because huge diversity exists even among major groups within the plant kingdom as seen in monocots (grass family members), they were also excluded from the comparative analysis even though they were included in the analysis for upstream enzymes like HMGS and HMGR. Among the plant species used for comparison, a wide distribution of CPT was noticed. For instance, in *Arabidopsis thaliana*, a CPT gene family with nine members has been reported (Surmacz and Swiezewska, 2011). One member of this group, ACPT/DPS, contributed toward the biosynthesis of long-chain polyisoprenoids (Cunillera *et al.*, 2000; Oh *et al.*, 2000), while another member, AtHEPS, was recently identified as a heptaprenyl diphosphate synthase (Kera *et al.*, 2012). However, functions have not been assigned for the remaining members of the *Arabidopsis* CPT family. Although the function of the vast majority of plant *CPT* genes whose sequences are found in various databases is not well established, their homology to dehydrodolichyl

diphosphate synthases from non-plant species suggests that they are involved in the synthesis of long-chain polyisoprenoid diphosphates. Other reports of plant CPTs include several genes from *Hevea brasiliensis* and *Taraxacum brevicorniculatum* that are associated with synthesis of natural rubber particles (Asawatreratanakul *et al.*, 2003; Schmidt *et al.*, 2010). From the above studies it is evident that *CPT*s are more or less specific to each species depending on their function and in *Hevea* they are specifically designed for the synthesis of a very high molecular weight poly-isoprene (natural rubber) compared to the isoprenoids of low molecular weight synthesised in other species. This uniqueness is evident in the protein as well as gene based phylogenetic analysis. The analysis also reveals the presence of multiple isoforms of this gene in rubber similar to that observed in other species but the sequences appeared to be evolved from a single *Hevea* specific *CPT* probably by gene duplication. Our study also identified the flawed nomenclature of the *Hevea CPTs* from various studies reported as the similar sequences were named in a different way by various groups working on it. The discovery of a new putative isoform of *CPT* further suggests that there exists several highly conserved and divergent sequence regions among the different isoforms identified. Similarly REF, another terminal protein involved in the elongation of isoprene was also found to be unique in *Hevea* like its function whereas its paralog SRPP was widely conserved across species. This is evident from the grouping where *Hevea* SRPP was the only sequence clustered along with the sequences designated as SRPP and REF in other plant species including closely related *Ricinus communis* and *Jatropha curcas.* The separate grouping of *Hevea* REF suggests that SRPPs are clearly distinct from REF but was wrongly annotated as REF in other species probably due to their sequence similarity at certain motifs. The difference between REF and SRPP in rubber has been established by Tang *et al.* (2016) where they state that SRPP isoform is an ancestral protein in the *REF/SRPP* gene family and a mutation event occurred relatively recently, truncated the *SRPP* resulting in REF. Furthermore this truncation event which was unique to *Hevea* was considered to be the reason for the enhanced capacity of rubber production. The presence of a carboxy (C) terminus found in SRPP and its absence in REF further supports this assessment (Oh *et al.*, 1999; Berthelot *et al.*, 2012). From a comparative analysis of the phylogenetic trees from the present study and that from Tang *et al.* (2016) it may be assumed that

functional stress related genes responsible for stress adaption in plants may evolve independently and may not always show similarity to the most closely related species. Moreover this may be apparent in multigene families where the interlinking complex network of genes results in the desired trait. In the case of *Hevea*, even though single-gene families showed high homology to the sequences from *Manihot exculanta*, the multi gene families involved in rubber biosynthesis were much divergent and demonstrated more similarity to those sequences from *Ricinus communis* and *Jatropha*. Ultimately based on the available data it is assumed that the two terminal enzymes; CPT and REF bestowed *H. brasiliensis* its unique characteristic of high molecular weight rubber biosynthesis.

Even though the phylograms generated in the current study are simple trees based on a single protein sequence, the clustering, distribution and branching of each domain and the distance between them looks very interesting due to its concordance with the evolution of life from lower simple organisms to higher complex forms. The divergent clusters generated across the domains imply that each domain of life possesses a unique isoprenoid biosynthesis pathway whereas the convergence of *Hevea* and related plant species support the assumption that there existed a last common ancestor of all plant species. The relationship of the organisms studied suggests that there exist universally conserved as well as variable domains for each group and the MVA pathway is likely an ancestral metabolic route in all the three domains of life, Archaea, Bacteria and Eukaryotes.

4.2. SNP identification and distribution

The SNP data available from selected rubber biosynthesis genes suggest that there exists abundant genetic diversity among cultivated popular *Hevea brasiliensis* clones despite their origin from a few selected parents belonging to the Wickham collections. This is well apparent from the high level of heterozygosity at the analysed loci. This phenomenon may be attributed to the diversity originally present in the Wickham selection as well as those raised out of genetic recombination resulting from the selective hybridisation. Reduced genetic diversity is generally expected in cultivated plants than their wild counterparts due to the breeding practices followed during

domestication (Li *et al.*, 2011). Studies in Sunflower (Liu and Burke, 2006) and Apricot (Bourguiba *et al.*, 2012) strongly support this assumption. Interestingly significant genetic diversity in cultivated rubber clones was reported by several groups based on molecular genetic studies (Le Guen *et al.*, 2011; Perseguin *et al.*, 2012; Roy *et al.*, 2012). Based on recent findings by Shearman *et al.* (2015) the frequency of single nucleotide substitutions found in the rubber tree transcriptome was approximately 1 in every 270 nucleotides, comparable with the similar trend in other plant species like maize: 1 SNP/204 bp (Gore *et al.*, 2009), soybean: 1 SNP/490 bp (Choi *et al.*, 2007) and pea: 1SNP/540 bp (Leonforte *et al.*, 2013). In the present study, SNP density of 1 in every 53 bp to 410 bp in the mevalonate pathway genes with total number of 172 SNPs identified is congruent with the above reports. This trend is even evident in the coding regions also where non-synonymous SNPs resulting in major amino acid (aa) changes were noted.

In the case of *HMGS* gene, the rare alleles were observed only in heterozygous condition in all the five genotypes. Therefore, they were considered as non functional alleles apparently having less downstream impact on the gene expression. This phenomenon is supported by earlier reports by Koehn and Eanes, (1976) on the existence of strong correlation between rare alleles and their heterozygous status. Studies by Chen and Sun (2013) and Toka *et al.* (2013) established that rare mutations usually demonstrate low allele frequencies in a population. These studies substantiate the low frequency observed (7.5%) for the rare allele "T" at base position 3513 during the screened of this SNP in the *HMGS* gene in 19 clones of rubber.

A comparison of the number of coding region SNPs shows that *REF* have maximum number of non-synonymous SNPs whereas the rate limiting gene *HMGR* has none. Total number of SNPs as well as frequency was also high in *REF* closely followed by *CPT*, the two downstream genes which are more specific to *Hevea brasiliensis*. As evident from the SNP frequency estimation studies from various genes, the SNP frequency per 1,000 bp for the genes like *HMGR, HMGS, MVK, PMVK, FDPS* and *GGDPS* were comparatively much lesser than the downstream more specific genes like *REF* and *CPT*. Similar results have been reported in other plants also (Novaes *et al.*, 2008, Ingvarsson, 2008). A contradictory result was reported by Tang *et*

al. (2016) where they claimed that *Hevea* cultivars exhibit a low SNP frequency at the rate of approximately <1 SNP per kilobase which is comparatively very low with that observed in rice and date palm based on bimodal curve study using sequence data. Furthermore they claim that SNP deserts account for 42% of the *Hevea* genome, a value significantly higher than that reported in rice (8%) or date palm (18%). According to them, *REF*1, the single most highly expressed gene in latex resides in the core SNP desert indicating their highly conserved nature in *H. brasiliensis*. On the contrary our results clearly indicate that *REF* is the most unconserved gene among all the rubber biosynthesis gens with a high SNP frequency of 1 in every 53 bases.

Evolutionary variation of genes involved in plant metabolic pathways has been well established by various studies. Studies on synonymous and non-synonymous SNPs in structural genes in the anthocyanin pathway within the genus *Ipomoea* as well as for species from monocots and dicots representing different divergent times of evolution showed that downstream genes exhibited statistically significant greater divergence rates than upstream genes (Rausher *et al.*, 1999; Lu and Rausher, 2003). Similar patterns of variation were found in the carotenoid biosynthetic pathway genes from another six species (Livingstone and Anderson, 2009). Significant positive rank correlation was also found between positions in the pathway and non-synonymous substitution rates. Because upstream enzymes are intermediary of multiple end products, even slightly detrimental amino acid changes in these enzymes could have major deleterious fitness consequences. Greater constraint in upstream genes may also be explained because they are associated with pathway branches and may control pathway flux (Crabtree and Newsholme, 1987). In fact, a protein interaction network analysis showed that proteins with more interactions tend to evolve more slowly, because a greater portion of the protein is directly involved in its function, and connectivity is positively associated with pleiotropic effects on cellular function (Hahn and Kern, 2005; Rausher *et al.*, 2008). Therefore the higher nucleotide and non-synonymous variation found in the lower isoprenoid biosynthetic genes could be related with stronger purifying selection in upstream genes like *HMGS, HMGR, MVK* etc. that provide precursors for various groups of end products (triterpenes, sterols, brassinosteroids, isoprenes). SNP data analysis in forest trees from the earlier studies

suggest that 15-20% of genes are under some kind of selection pressure (Külheim *et al.*, 2009). The SNP information obtained from the rubber biosynthesis genes are in agreement with the previous studies which states that genes at a bottleneck position of a pathway may be under stronger selective pressure and may therefore have fewer SNPs than will genes at other positions. Furthermore, the number of genes with the same function may determine the SNP frequency at a particular gene. Therefore it is assumed that genes with single or very few copies within the genome are under higher selective pressure and therefore we are expected to observe less SNPs in those genes than genes that occur in multi-gene families.

Regarding greater diversity found in introns of the gene analysed compared with their exons, contrasting effects of selection within a gene could explain differential levels of nucleotide diversity on introns (Hare and Palumbi, 2003). In *HMGR, PMVK* and *FDPS* genes, the greater sequence variation in intron was caused not only by the number of SNPs but also by the greater frequency of indels. The lack of regulatory elements in the introns and major constrains on exon region due to pleiotropic effect of these primary metabolism genes could explain greater polymorphism on introns. Contradictory observations in the intronic region of *FDPS* gene despite the presence of intronic regulatory motifs can be attributed to the activity of the retro elements identified within this gene.

Even though introns are removed during the mRNA maturation process, they are known to enhance or regulate gene expression in numerous ways (Morita *et al.*, 2012). Recently several groups reported the presence of regulatory elements having a significant role in post translational modification and gene expression from the first intronic region of genes from higher plants (Gowik *et al.*, 2004; Morello *et al.*, 2008; Parra *et al.*, 2011). Because these regulatory first introns are proximal to the 5' region of the gene, they are known as *cis*-acting and can act upstream or downstream from the gene irrespective of their orientation (Clancy and Hannah, 2002; Rose, 2002). Potential mechanisms for such intron-mediated enhancement include increased transcription, splicing facilitated transcript maturation, stabilization or export and targeting of spliced transcripts for protein synthesis (Parra *et al.*, 2011). Despite these important functional

roles, introns are still structurally not as stable as are coding regions due to their susceptibility to mutations induced by various internal and external factors. One of the major factors responsible for this instability is transposable elements due to their notorious nature of high self establishment in the genic as well as non-genic regions. This may often result in the reshuffling of targeted sites due to repeated copy paste events induced by them (Gao *et al.,* 2012). In plants, retrotransposons often concentrate numerous sequence variations in to short stretches of non-coding DNA by the induction of several SNPs and SSRs as observed in *FDPS* intron1 with high SNP density in a hyper variable region of *H. brasiliensis*. The similarity of this region to a partial reverse transcriptase (RT) gene sequence identified during blastx analysis further strengthens the above assumption.

Even though characteristic features of a non-LTR retrotransposon, like the presence of direct repeats at base positions 582 (9 bp repeat), 785 (32bp repeat) and the CT repeat at position 288 as revealed by the *FDPS* intron1 sequence analysis, the RT gene seems to be incomplete or appears to be highly truncated. As per previous reports on truncations associated with non-LTR integration, this may be a consequence of the integration of prematurely terminated RT transcripts initiating at the 3'end of the RNA (Luan *et al.*, 1993). The absence of a set of uniform structural features of a TE which can sometimes only be accomplished by an examination of its reverse transcriptase (RT) encoding domain (Han 2010) is an indication of the presence of a non-LTR retrotransposable element as in the case of the *FDPS* intron. Moreover, truncated copies will, with time accumulate mutations eliminating their ORFs either fully or partially even though they may be active *via* the *trans* acting interspersed retroelements at another site. Mutations like the 35 single nucleotide variations resulted from the copy paste events induced by the retroelement on the nearby sequences provide additional evidence for the above argument. However the possibility of a non-LTR retrotransposon remaining unidentified or being misidentified as SINE insertions also cannot be ruled out. Though transposable elements are known to multiply once integrated in to a genomic region, under certain circumstances for maintaining gene integrity, they may get eliminated by self excision mechanisms or *via* recombination between 10 to 20-bp target site duplications (TSDs) flanking them (García-Pérez, 2010;

Kejnovsky *et al.*, 2012). Evidence for this possibility is obvious from the target site duplications present in the *FDPS* intronice sequences. Therefore the conserved nature as well as the high frequency of FINT1-B allele contrary to the rarer and polymorphic alleles like FINT1-A, and FINT1-C in the analysed populations may not be just coincidence, but rather an indication of its stability due to the removal of retroelement sequences. Nevertheless, the probability of such changes fuelling micro-evolution of new favourable alleles also cannot be ruled out. For example, sequence analysis indicates the possibility that FINT1-A may be the more recently evolved allele than the other two..

The discovery of new alleles in addition to the three major alleles discussed earlier from other popular *Hevea* clones, wild accessions as well as other *Hevea* species substantiate the earlier assumptions about the occurrence of TE associated turbulences in the *FDPS* gene structure in *Hevea*. Sequence data from popular clones reveals that the intron1 region up to 1379 bp is partially conserved. FINT1-B is the frequent allele as 72% of clones had only this allele which is considered to be more preferable than the others. This can be attributed to natural selection because of the similar trend observed in wild accessions and *Hevea* species where FINT1-B is the prominent allele. Same reason may apply to the occurrence of the smaller FINT1-D allele with moderate frequency (20%) in popular clones. Genotyping data of popular clone further supports this hypothesis because, both FINT1-B and FINT1-D alleles existed in homozygous state as well as in heterozygous state whereas the FINT1-A and FINT1-C alleles having similarity to the retroelement were not found in homozygous state and rarely found together with exceptions like in the case of RRII 118. In general, from the genotypic and sequence data, it can be inferred that FINT1-A, FINT1-B and FINT1-C alleles were shared among the wild accessions from all the three provinces and in popular clones whereas the FINT1-D allele though shared among popular clones in low frequency was completely absent from Rondonia accessions. Therefore it can be assumed that the original Wickham collection may be more of a representation of Acre and Mato Grosso province rather than Rondonia. Nevertheless, the presence of several uncharacterised distinct *FDP* alleles in Rondonia accessions like the smallest allele with

the largest deletion having maximum similarity to retroelement suggest the higher prevalence of intron1 structural variations among Rondonia plants.

Though several isoforms of *FDPS* gene from *Hevea* has been reported by earlier studies, all the alleles mentioned above are supposed to be of the same gene because, analysis of popular clones, wild accessions and *Hevea* species showed the existence of only one (homozygote state) or two variants (heterozygote state) of the *FDPS* intron1 loci. If these were paralogs, one would expect to see more than two variants in one individual itself. Furthermore, the observed allele frequency and the genotype frequency of the three alleles; FINT1-A, FINT1-B and FINT1-C in a full-sib progeny showed no statistically significant deviation from the expected segregation ratio of genotypes Mendels law of inheritance for both the alleles and genotypes ($P > 0.10, 0.05$) and followed a Mendelian pattern of inheritance. All the above results are well confirmatory for the assumption that the alleles observed were of the same gene.

The presence of indels and SNPs in the regulatory region of a gene may have a significant role in gene regulation due to the disruption of their recognition motifs. Since first introns are known to regulate gene expression using the *cis*-regulatory elements they carry, indels and SNVs within them are expected to have substantial impact on gene regulation (Schauer *et al.*, 2009). As Jasmonates are known signalling compounds for the production of plant defence related compounds like terpenoids, glycosteroids and alkaloids by the co-regulation of genes involved using their CGTCA element, the presence and absence of this motif in the regulatory regions of the *FDPS* gene indicate the possibility of a jasmonate-mediated regulation of isoprenoid biosynthesis (Rouster *et al.*, 1997; He and Gan, 2001). Furthermore the association of a retroelement with this motif lead us to assume that an epigenetic mode of gene regulatory mechanism by transposable elements exists in *Hevea* as in the case of other previously reported systems (Whitelaw and Martin, 2001; Kashkus *et al.*, 2002). Therefore, the absence of CGTCA-motif due to an SNP in the *FDPS* promoter region as well as in both the intron1 alleles as seen in RRII 118 may be a retrotransposon induced negative MeJa responsive regulatory mechanism in *Hevea*. Similarly, the absence of a gibberellin responsive element (TATC-box) in the FINT1-C allele of RRII

118 and the introduction of an alternate gibberellin responsive element (GARE) upstream in the promoter sequence of the same strand introduced by a single mutation may be part of an alternative response mechanism for the regulation of *FDPS* itself. The presence of an intact 5'UTR Py-rich stretch (conferring high transcription levels) and the signal transduction response domain F-box in all the plants may be associated with the retroelement function rather than with *FDPS* transcription since the selected plants consisted of different species having varying rubber biosynthesis capacity. Furthermore, the presence of an additional TATA box as well as a CAAT box in intron1 are also assumed to represent the regulatory sequences of the retroelement although, experimental evidence lacks presently. On the other hand the presence of defence and stress responsiveness element like "TC" only in *H. benthamiana* is justified as this species is considered to be comparatively more disease- tolerant than *H. brasiliensis* (Jain and Priyadarshan, 2009). Besides the disruption of regulatory motifs, insertion of a transposable element in intronic regions can lead to exonisation as well as to new patterns of pre-mRNA processing, resulting in novel responses to developmental and environmental stimuli (Weil and Wessler, 1990; Sela *et al.*, 2010). This can happen *via* the induction of new alternative splice sites or by the modification of existing canonical protosplice sites where transposons preferentially get inserted (Dibb and Newman, 1989). In the case of the *FDPS* first intron, the splice site was GG/GT and AG/AT at the donor and acceptor sites respectively. If these potential donor and acceptor splice sites are utilized efficiently by the spliceosome, transposons may get inserted between the sequences without altering the genes coding sequence. This may be the reason for the *FDPS* genes highly conserved coding region despite of having a polymorphic variable first intronic region. Alternatively, transposons themselves may harbour strong donor and acceptor splice sites near their boundaries or activate nearby latent splice sites, enabling its precise, or nearly precise, excision by the spliceosome (Yenerall and Zhou, 2012). Transposable elements are also known to eliminate or create canonical splice sites in regulatory region *via* single nucleotide variation (Huang *et al.*, 2008; Barbaglia *et al.*, 2012). A single nucleotide change at position 688 (G/C) resulting in the absence of a donor site "ggaggGTgagtc" in FINT1-A is a clear evidence for such a changes in *Hevea FDPS* gene. The five base deletion in FINT1-D from 977 bp resulting in two novel acceptor sites can also be attributed to the same reason.

Another instance is the acceptor site "attacAGaggtg" present only in FINT1-C and the site "tacacAGgtgag" absent only in FINT1-C which are basically two splice site variants generated from the same region due to a single nucleotide change at position 1109. However, transcript level studies are required to prove the functional relevance of these changes.

Structural rearrangements are the result of different kinds of mutations like sequence inversions or insertions/deletions in coding and non coding regions of the DNA. Though thousands of indels are reported from various plant species from genic as well as inter genic regions, majority of them are less than 50 bp in size. For example, whole genome analysis of *A. thaliana* revealed that only 5% of the indels are more than 50 bp. Moreover they observed that majority of the indels greater than 250 bp are associated with transposon-related proteins, indicating that a substantial proportion of the large indels are the result of transposon insertion or excision (The *Arabidopsis* Genome Initiative, 2000). The intronic indel identified in *PMVK* gene was the biggest indel reported so far from rubber biosynthesis genes. Moreover *PMVK* was the biggest mevalonate pathway gene identified so far spanning up to 9 kb in length with ten introns of which the biggest intron was the 7^{th} intron of approximately 2.0 kb followed by the 10^{th} intron of size 1.8 kb which harboured the large indel of 1.2 kb size. Since sequence analysis revealed the absence of any traces of transposable elements in this region, the large indel may be caused by segmental duplication. In this context the large indel observed in *PMVK* can be considered as a rare event which may have some influence on the gene structure and function as well (Cong *et al.*, 2008). Other than the above mentioned large indels from *FDPS* and *PMVK* genes, indels were rarely observed except the two 'one nucleotide indels' in *HMGR* gene. Interestingly the indel at base position 1341 of *HMGR* was noted in all combinations in the clones analysed *ie.,* as homozygous (deletion) in RRIM 600 and RRIC 52, homozygous (addition) in the Indian clones RRII 105 and RRII 118 and heterozygous (insertion/deletion) in GT1, whereas the indel at position 1636 was seen only in homozygous state (addition in RRIM 600, RRIC 52 , GT1 and deletion in RRII 105, RRII 118). This trend indicates that the indel at 1341 is primitive whereas the latter was more of recent origin.

The intragenic recombination sites predicted within the above genes also might have accounted for the high variability since intragenic recombination is considered as a key player in the generation of novel variants like mutations (Watt, 1972). The contribution of intragenic recombination towards allele diversity and gene evolution has been well studied in R genes of *Arabidopsis* and self incompatibility locus of *Petunia inflates* (McDowell *et al.*, 1998; Wang *et al.*, 2001). This stands true in the case of *HMGS, HMGR, MVK* and *PMVK* genes also because of the presence of several intragenic recombination sites. Recent studies on different patterns of polymorphism between different functional gene classes suggest that genes interacting with the external environment have often high levels of SNP diversity (Clark *et al.*, 2007). This stands true in the present study since *REF* and *CPT* are known for their interaction with several forms of biotic and abiotic stress (Kim *et al.*, 2004; Ponciano *et al.*, 2012). For example, stress related REF proteins from hot pepper and sweet potato are induced by abiotic stress and can improve stress tolerance when over expressed in *Arabidopsis* or tobacco (Kim *et al.*, 2010; Seo *et al.*, 2010).

Analysis of the distribution pattern of SNPs from the five clones based on the geographic region of their origin shows that majority of SNPs were not restricted to any clones of any particular region. This trend is clear evidence that the clones are all related to each other to some extent. At the same time the Sri Lankan clone RRIC 52 seems to be unique in terms of SNP as well as haplotype diversity. Uniqueness of Sri Lankan clones was also reported by Roy *et al.,* (2012). In the case of *HMGS* gene the distribution pattern of the 20 SNPs in five genotypes shows that the Southeast Asian clones possess the maximum number of unique SNPs compared to Indian and Sri Lankan clones. At the same time it is also interesting to note that the single Sri Lankan clone (RRIC 52) possessed three unique SNPs in comparison to a single unique SNP from two Indian clones. Similar trend was noted in the case of *MVK, FDPS* and *GGDPS* genes respectively. This may be attributed to the fact that RRIC 52 is a primary clone evolved by ortet selection from seedlings which might have been retained in Sri Lanka while serving as a disseminating point for the Wickham seedlings from Kew gardens to different countries like India. Overall comparison of SNP distribution from the rubber

biosynthesis genes reveals that the number of SNP alleles shared by clones from all the three regions is very less, which further highlight the existence of genetic diversity within the existing Wickham population.

Intronlessness is as a characteristic feature of prokaryotic genome which is rarely observed in eukaryotes. Though extensive studies on intronless genes has not been done in plants, several studies pertaining to the occurrence and evolution of inton less genes in plants like rice, maize, Arabidopsis etc. have been reported (Yan *et al.*, 2014; Jain *et al.*, 2008). In *Hevea*, intronless versions of *CPT*, Chitinase, *FDPS* and Hevin were reported by various groups (Saleena *et al.*, 2006; Saleena *et al.*, 2010); Bokma *et al.*, 2002; Venkatachalam *et al.*, 2016). When the available *Hevea cis*-prenyltransferase sequences in *Hevea* were investigated, it was noted that the gene was devoid of introns in all of them. Moreover the gene was observed to be universally intronless witht few exceptions like some isoforms in tomato (Akhtar *et al.*, 2013). Similarly, majority of *GGDPS* sequences reported were also intronless but intron containing isoforms were found in some plants and algae (Okada *et al.*, 2000; Liang *et al.*, 2015). On the contrary *FDPS* sequences from plants were mostly containing intron including that of rubber from the current as well as previous reports. Interestingly an intronless variant of this gene was reported very recently by Venkatachalam *et al.* (2016). Despite repeated amplification and sequencing of *FDPS* gene from five clones using several sets of primers designed from various regions of the cDNA, intronless variants were not obtained from the present study from any of the *Hevea* clones including RRII 105. Therefore we assume that the intron less variant obtained by Venkatachalam *et al.*, (2016) may be a very rare isoform. Earlier studies in different plant species imply that intronless genes in a species which does not have homologs in other species in the database, have often crucial role in performing and expressing species-specific traits or phenotypes (Jain *et al.*, 2008). This stands true in the case of CPT gene. Altogether, our findings suggest that *Hevea* harbours intronless genes in their genome involved in the rubber biosynthesis process.

Depending on the location of amino acid substitution in the protein resulting from non-synonymous SNPs, they may have impact on the structure as well as function

of the encoded protein. A major substitution resulting in a change from acidic polar Aspartic acid to neutral polar Asparagine was noted in position 2490 of *MVK* gene. Since the locus was observed to be heterozygous in all the five clones it is assumed that the mutation may be an undesired one where the recessive allele was incorporated. However its impact is nullified by the positive effect of the dominant allele. Since protein crystal structure of MVK from plants is not yet reported, *in silico* predictions at this stage could not be carried out. On the contrary, a similar change within *GGDPS* gene resulting in the substitution of neutral non-polar Glycine to acidic polar Aspartic acid (HbGGPS1097AG) may not be deleterious to the structure or function of the enzyme despite the presence of both the alleles in homozygous state in different clones. This argument is based on the *in silico* structural modelling analysis which shows that putative GGDPS binding pocket is uninterrupted by the above substitution. Presence of both the alleles in homozygous state in different clones was also observed in the case of *CPT* gene for the SNPs Hbcispre714AT and Hbcispre1145CT where substitutions resulted in change from neutral to basic amino acids. Since RRII 118 and RRIC 52 had minor allele in homozygous condition for the above SNPs at both the loci, it is assumed that they originated from the Sri Lankan clones. Correspondingly the amino acid substitutions in *REF* gene due to the non-synonymous SNPs HbREF1382AT and HbREF1505AG resulted in neutral non-polar to acidic polar amino acid change. Since one of the allele (minor allele) was not seen in homozygous condition in both the loci, it is assumed that the minor allele is the recessive one which is not favoured or expressed. Similarly the non-synonymous SNPs HbREF1460AT and HbREF1508AG resulting in basic to neutral amino acid substitutions were observed only in GT1 under heterozygous state. The very low frequency of the minor alleles in both the loci clearly states their negative selection characteristic. At the same time, the divergent nature of the clone GT1 is also expressed from the above data. Furthermore, rare SNPs like the above in GT1 may be of interest due to their potential role in explaining quantitative trait variation (Gibson, 2012).

4.3. Haplotype structuring

The term haplotype refers to a set of single nucleotide polymorphisms (SNPs) existing on one of the chromosome of a homologus pair arranged in the linear order,

whereas haplotype block refers to those neighbouring SNPs that have the tendency to occur together due to tight linkage. Usually the SNPs in a haploblock are under linkage disequilibrium where the mendelian inheritance pattern is not followed. *i.e.,* most haplotypes fall into a few classes with little evidence of recombination. This phenomenon was first observed in dense SNP data from the SNP Consortium Allele Frequency Project (Gabriel *et al.*, 2002). This grouping of SNPs into haplotype blocks has led to significant improvement in the power and accuracy of association mapping studies where a particular haplotype is associated with a trait of interest (Zhao *et al.*, 2003).

Associations between phenotypes and haplotypes, rather than single SNPs, have been extensively studied in many plants like rice, barley etc. (Begum *et al.*, 2015; Lorenz *et al.*, 2010). These studies are based on the underlying principle that density of markers in a gene have significant impact on the linkage disequilibrium structure of the SNPs which may ultimately affect the phenotype. For example, the possibility of epistasis among SNPs is much greater when SNP density is very high, and is particularly relevant to studies where many common SNPs have been typed across a candidate gene (Xu *et al.*, 2015). SNP-SNP interactions within genes and gene clusters was proved by Zhang *et al.* (2015) through their epistasis studies on the genetic architecture of sudden death syndrome resistance in soybean. In this context, the identification of haplotypes from important rubber biosynthesis genes in *Hevea* is essential to establish their probable association with a complex trait like rubber yield. Haplotype analysis in *HMGS* revealed that the incidence of any one haplotype in more than one clone was noted only in the case of Hap_2 and Hap_5 where they were distributed among clones belonging to the same geographic region like Hap_2 being present in Indian clones RRII 105 and RRII 118 while Hap_5 among the Southeast Asian clones RRIM 600 and GT1. Interestingly, despite the significant differences in parentage of RRII 105 and RRII 118 they shared one haplotype indicating their common Wickham origin probably collected from Mato Grosso province in the Amazon forests (Jones and Allen, 1992). The haplotype distribution in *HMGR* also supports this statement since both RRII 105 and RRII 118 had the same haplotype Hap_1 in homozygous state. On the contrary, RRIM 600 seems to have some unique

characteristics in their rubber biosynthesis genes which are evident from the distinct haplotypes observed in *HMGR* and *MVK* genes. The size of a haplotype block largely depends on the presence of recombination hot spots (Kauppi *et al.*, 2007). Since three possible recombination sites were predicted within the gene, there may be a maximum of four haplotype blocks within *HMGS*. In the case of *MVK*, the occurrence of eight haplotypes with low frequencies suggests that several intragenic recombinations happened in the past in the popular clones due to their high out crossing nature (Pawsoi *et al.*, 2013). The prediction of eight recombination hot spots within the gene supports this assumption.

As far as gene size is concerned, *PMVK* is the most conserved rubber biosynthesis gene among the eight genes analysed (excluding the large indel) which is evident from their comparatively low SNP frequency of 1 SNP in every 410 bp. It is interesting to note that despite possessing nine introns including a large indel the sequence structure was relatively conserved among the five popular clones analysed. This is evident from the less number of haplotypes and recombination sites from the entire 9 kb genomic region. As mentioned elsewhere, PMVKs positioning as a branch point enzyme may be the reason for its conserved nature. Additionally, the six haplotypes clearly demonstrate the diversity as well as relationship of the five clones. For example Hap_3 which had the highest frequency of 40% was shared among the Indian clone RRII 118 of Sri Lankan parents, Malaysian clone and Indonesian clone. At the same time, region specific haplotypes like Hap_2, Hap_4, Hap_5 and Hap_6 were observed in the clones from India, Malaysia, Sri Lanka and Indonesia. Similar patterns were observed for *GGDPS* haplotypes where haplotype based differentiation of clone origin could be observed. For example, Hap_1 and Hap_2 appeared to be specific to clones of Indian and Malaysian origin whereas Hap_3 and Hap_4 were highly specific to Sri Lanka (RRII 118 was developed as an Indian clone but both of its parents (Hil & Mil) are Sri Lankan primary clones evolved by ortet selection). Even though at the micro level with very limited number of samples, the above findings assumes relevance because the existence of correlation of geographic origin to genetic structure is imperative to understand germplasm diversity and for identifying traits underpinning adaptive variation in rubber (Blackmore *et al.*, 2015). The ongoing studies on the

distribution of SNP loci in glutathione peroxidise gene based on the region of origin of *H. brasiliensis* wild accessions by Saha *et al.* (unpublished) is another supportive evidence pointing the above aspect.

In the case of *FDPS* gene the conserved nature of the region starting from base position 1379 is evident from the phased SNP data which emphasize mainly three haplotypes (Hap_1,Hap_2 and Hap_3). The remaining four haplotypes identified are minor modified versions of the above three with maximum variation of just three out of twenty five SNPs (12%). The low exonic SNP frequency of 8% further ascertains their minimal contribution towards the overall *FDPS* gene diversity. From the haplotype data, it can be inferred that Hap_4 evolved from Hap_5 and Hap_5 from the most prevalent Hap_1 since, Hap_5 varied from Hap_1 at two loci and Hap_4 varied from Hap_5 at just one locus. Hap_1 was the only haplotype observed in fully homozygous state (in RRII 105). In the case of RRII 105, the close relationship between its Malaysian and Indonesian parents (Tjir1 & Gl1) may be the reason for the high homozygosity. Sequence analysis of other major rubber biosynthesis genes supports this statement (Uthup *et al.* unpublished) Similarly the close genetic relationship of the RRII 105 and RRIM 600 (Same female parent (Tjir1) of Indonesian origin; Different male parents (Gl-1 & PB-86) of common Malaysian origin) is evident from the high homozygosity and the presence of Hap_1 in both clones. The entirely different lineage of RRII 118 (hybrid of Sri Lankan primary clones Mil 3/2 x Hil 28) may be the reason for the presence of its two different haplotypes (Hap_2 &Hap_3) with a high rate of heterozygosity (40%). Though the rest of the gene posses a comparatively stable sequence structure, unlike the first 1379 bp region, we could not provide a clear explanation for the presence of SNPs and the resulting seven haplotypes from this region. The probability of recombinant events inducing SNPs is sparse in this region as the 'D' calculated between any of the two SNPs between 2nd and 25th SNP is one, indicating some lack of recombination in the last 3344 base pairs. The possible impact of the first intronic retroelement on this region also can be ruled out since the SNPs are placed randomly at distant locations from the transposable element (TE) in all the five clones. However, the one recombinant event predicted between the 1st and 2nd SNP in this region seems to be correct based on four-gamete test as the 'T' allele at position

1380 of RRIM-600 appeared to be the result of a recombination event resulting in Hap_4 from Hap_5. All the above results suggest that the last 3598 bp (three fourth) of the *FDPS* gene is more or less structurally exempt of any influence from the TE present in the first intron.

SNP frequency as well as number of non-synonymous SNPs was observed to be highest in *REF* gene than any other rubber biosynthesis gene analysed so far. Despite this high SNP frequency, only five haplotypes were identified for *REF* gene from the five genotypes. Major reason is the surprisingly similar sequence structure of this gene in RRII 105 and RRII 118 with 100% homozygosity in all the loci. The highest frequency observed for any single haplotype (60%) was Hap_1 which showed its presence in all clones except RRIC 52. Though two potential intragenic recombination spots were predicted from *REF*, the high frequency of this haplotype suggest that they segregate as a single block. Interpretations of human SNP data by Wall and Pritchard (2003a, 2003b) showing the delimitation of haplotypes by recombination hot spots supports this argument. Similar trend was noted in *HMGR* gene also where polymorphisms could not be detected between RRII 105 and RRII 118 which is partially responsible for the low haplotype diversity and recombination sites in this gene. In both the genes heterozygosity was high in the other three clones. Due to their entirely unrelated parentage, no possible explanation could be deduced for their sequence similarity in these two genes and the only possibility owes to their common Wickham origin. In *CPT* gene where high recombination frequencies were observed despite it being an intronless gene. As expected, the high recombination frequencies subsequently resulted in high haplotype diversity. Additionally, some of the minor alleles in this gene also might have contributed to the diversity since minor SNP allele frequency is a major determining factor of haplotype frequency (Micheal and Stumpf, 2004).

Reconstructions of the genealogical relationships between genes through the use of haplotype information can be used to analyze evolutionary processes at the time of lineage separation. Haplotypes are effective tools to study the evolutionary relationships of genotypes within a species as well as between different species within the same

genus (Schaal *et al.*, 2003). The simple phylogenetic trees constructed for each gene in the present study revealed interesting information on the evolution of *Hevea* rubber biosynthesis genes in popular clones (Uthup *et al.*, 2016). In the case of *HMGR* and *CPT* genes, RRIM 600 harboured unique haplotypes which were very distinct from the others. It is assumed that the unique allele might have been acquired from its male parent PB86 because it is a primary clone evolved by ortet selection in Malaysia prior to 1956, whereas the female parent Tjir1 is common for RRIM 600 as well as RRII 105. Ancestral haplotypes usually remain within the species and will be passed onto future generations at very low frequencies (Sicard *et al.*, 2015). Though less in frequency, distinct primitive characteristics of the haplotypes Hap_7 and Hap_8 of *MVK* gene can be seen in RRIC 52 where these haplotypes formed the root from which two different clusters of haplotypes are evolved. The distinctive nature of RRIC 52 as mentioned elsewhere is also evident from the distance at which Hap_5 of *PMVK* and Hap_4 of *GGDPS* were placed in their respective haplotype phylograms.

Based on the entire SNP and haplotype information from five genotypes it is evident that despite their common Wickham base origin, there exists abundant genetic diversity in the rubber biosynthesis genes among them. At the same time the SNP data also explained the common origin of the rubber biosynthesis pathway genes and its evolution in rubber. The abundance of non-synonymous SNPs indicates that most of the SNP substitutions were not of recent origin and it was naturally present in the seeds randomly collected by Wickham form the Amazon base. Analysis of downstream genes like *REF* and *CPT* revealed information on the sequence specificity of these genes in rubber and provide clues to assume their crucial role in the synthesis of high-molecular weight natural rubber in *Hevea*.

4.4. Comparative SNP distribution studies

The relationship of the plants species studied based on conceptually translated protein sequence suggests that there exist universally conserved as well as variable domains for each group. Proper assessment of this diversity at genome level is possible only by comparative genetic mapping and synteny analysis which reveal conservation of gene content, order and function among closely related taxa as well as distantly

related organisms (Bennetzen 2000; McCouch, 2001). Recently microsynteny (small-scale of synteny) has been investigated across several plant species using whole-genome sequences and selected discrete sequences to infer shared ancestry (Dohm *et al.*, 2009; Rajesh *et al.*, 2008). Though not as extensive as the above studies, comparative DNA or protein sequence analysis is an efficient way to understand the functional significance, evolution and distribution of nucleotide variations at the gene level (Tompa, 2001). This is accomplished by comparing orthologous sequences from different species and identifying the SNPs in those regions that are most well-conserved. For example, Jung *et al.* (2012) have reported the existence of conserved as well as phylum or group-specific single-nucleotide polymorphisms (SNPs) in gene sequences among ciliates. Though not specific to any phyla, the wide distribution of the coding SNPs noted in *Hevea* rubber biosynthesis genes also provides valuable information on their origin and evolution in plants. For example, the presence of SNP HbHMGS2167AG in *HMGS* gene across plant species belonging to different phylum and class is a good indication of their lineage from a common ancestor. Interestingly the presence of "Asn" instead of "Asp" in the members of grass family due to the above single nucleotide variation indicates its evolvement and speciation as a separate line from the other higher plants (Bennetzen *et al.*, 1998). It is assumed that "Asp" inducing mutation occurred in higher plants after the differentiation of cereals. In the case of Pinus, "Asn" producing allele might have been acquired from a common ancestor of angiosperms apparently originated from a progymnospermous ancestor or delineated from gymnosperms itself before their differentiation into the recent phyla (Semikhov *et al.*, 2004). The same explanation can be used to justify the presence of Valine in the case of SNPs HbHMGS2180CT and HbHMGS3513CT. With regard to the SNP HbHMGS4427AG (Alanine <>Threonine), "Thr" seems to be the primitive allele which is still retained at low frequency levels as seen in Hap_8 of *Hevea* and in *Pinus*. On the contrary, HbHMGS4052CT (Proline <> Leucine) and SNP HbHMGS4075CT (Proline <>Serine) were assumed to be evolved after the speciation of *Hevea* since they were found to be unique to *Hevea.* It is these SNPs which may aid in differentiating *Hevea* clones whereas the rest hold information on the history of gene evolvement. Similar trend can be observed in the distribution of *Hevea* specific synonymous SNPs in *HMGR* and non-synonymous SNPs in genes like *MVK, PMVK, GGDPS* etc.

Species-specific SNP markers are extensively used for differentiating land plants including trees species (Kersten *et al.*, 2015). Therefore such species-specific markers can be used for the development of DNA barcodes for differentiating *Hevea* species. SNP markers like HbMVK2628AG in *MVK* gene resulting in amino acid substitutions specific for selected plant families were also identified in *Hevea* which further shows the evolutionary relationship of the members of the families like *Fabaceae* and *Brassicasia* (*Arabidopsis lyrata, Arabidopsis thaliana and Brassica rapa*) to *Hevea.* Interestingly these amino acid substitutions were not observed in the *Euphorbiaceae* family members like *Ricinus communis* and *Jatropha curcas.* This indicate that not all genotypic variations in a species are strictly confined to the members of the same family but may extend beyond those boundaries showing close relation with members of other families which are otherwise distantly related. Being a branch point upstream enzymes associated with the synthesis of several secondary metabolites, *MVK* is presumed to be conserved as stated elsewhere. Surprisingly this nature is even evident in the case of non-coding regions of *Hevea MVK* gene like the presence of non-coding SNPs HbMVK2379AG and HbMVK3297CT in *Vitis vinefera.* On the contrary, putative family specific mutation as seen in the case of HbGGPS1097AG in *GGDPS* gene ascertains the robustness of the conventional taxonomical classification because similar amino acid substitutions were noted for this SNP in three family members of *Euphorbiaceae* included in the analysis. Comparative whole genome analysis has revealed that lineage-specific genes which have no detectable sequence similarity to genes from other lineages are present in plants (Rensink *et al.*, 2005; Graham *et al.*, 2004). Being generally conserved among taxa, they provide insight into the species specific evolutionary processes and biological functions. The sequence specificity of the downstream genes like *REF* and *CPT* can be considered as good examples for such genes in *Hevea.* Earlier modelling studies proposes that lineage-specific genes may be generated by gene duplication followed by rapid sequence divergence (Domazet-Loso and Tautz, 2003). Phylogenetic studies focussing *Hevea REF* genes agree with these findings (Tang *et al.,* 2016). It is anticipated that such genes in *Hevea* involved in the rubber biosynthesis process may be responsible for the unique rubber producing trait.

Based on the analysis of the distribution pattern of non-synonymous SNPs among diverse plant species, it is inferred that the coding region of important rubber biosynthesis genes in the upper mevalonate pathway are generally conserved despite the presence of several non-synonymous SNPs whereas the downstream genes like *REF* and *CPT* are less conserved, more species-specific in nature and therefore plays functionary role in determining the quantity and quality of latex produced. Furthermore, the data suggests that family specific as well as genus specific single nucleotide substitutions exist in plants which act as signature markers for the concerned group.

4.5. SNP marker development, segregation analysis and map integration

SNP markers are widely being used to construct high density linkage maps in various plant species with the advent of high throughput genotyping technologies (Wang *et al.*, 2016). HRM genotyping was used as an efficient genotyping strategy to map SNP anchored genes in crop plants and to incorporate them in existing linkage maps (Wu *et al.*, 2009). By applying the candidate gene strategy which has been frequently applied to identify genes controlling agronomically important traits in crop plants, we extended our study to map putative genes anchored by the previously and newly developed SNPs in a *Hevea* mapping population of 46 F_1 individuals obtained from the cross between two cultivated clones RRII 105 and RRII 118.

HRM based SNP markers were developed based on the SNP information generated from eight rubber biosynthesis genes which may have significant impact on the important trait of latex yield in rubber so that the genes associated with the markers can be mapped in a linkage map of rubber constructed earlier Bini, (2013). A total of seven SNP markers satisfying all the criteria required for HRM based genotyping were selected from six rubber biosynthesis genes. Since no SNP polymorphisms were detected for the genes *HMGR* and *REF* between the two parents of the progeny population, SNPs from these genes could not be considered as markers for the above population. SNPs are classified in to four groups based on the homoduplexes and heteroduplexes produced after amplification of a heterozygote and the predicted number of distinct nearest-neighbour thermodynamic duplexes (Tms) (Liew *et al.*, 2004). The

classification : class 1 (C/T and G/A), class 2 (C/A and G/T) , class 3 (C/G) and class 4 (A/T) was made based on the decreasing value in Tm difference. All the selected SNPs for HRM genotyping of *Hevea* population were Class I SNPs having higher Tm shift than other SNP types. Since the present study forms the first attempt to perform HRM genotyping in rubber, its efficiency was tested initially using the five clones of interest for which sequence information was already available. High sensitivity and specificity was obtained for amplicons of <400 bp) for SNP differentiation than higher amplicon size. This is in agreement with previous studies done to determine HRM efficiency by Garritano *et al.* (2009). Since most of the rubber biosynthesis genes were having high SNP density, the size of amplicon was even reduced to 150 bp in order to ward off the interference of neighbouring SNPs on the amplicon for melt curve analysis.

In rubber pseudo-test cross strategy was adopted for linkage map construction. The name "pseudo-test cross" was applied for this strategy because the testcross mating configuration of the markers is not known *a priori*, as in a conventional testcross where the tester is homozygous recessive for the locus of interest. In out-crossing tree crops, individual plants are highly heterozygous. Hence the F_1 mapping population had been widely adopted in trees due to the high heterozygosity (Grattapaglia and Sederoff, 1994; Zhang *et al.*, 2015). Lespinasse *et al.* (2000) reported the adoption of same strategy for the first time in rubber for linkage map construction using F_1 progeny populations obtained from PB 260 x RO 38 cross combination. In the reports of Feng *et al.* (2010), F_1 segregating population derived from the cross between Reyan 88-13 x IAN 873 was considered for mapping. Another genetic linkage map was constructed in *Hevea* using the full-sib progeny derived from a controlled cross between the cultivars PB 217 and PR 255 (Souza *et al.*, 2011). An integrated parental linkage map based on F_1 mapping population of RRIM 600 x RRII 105 was constructed by Triwitayakorn, (2011). The same mapping strategy was adopted in the present study using mapping data generated by Bini, (2013) in addition to the segregation data from the newly developed SNP markers. Since the allelic status of all the seven markers were homozygous in one parent and heterozygous in the other, the expected allele and genotype frequency will be always 1:1 as per Mendelian inheritance mode. The segregation analysis of all the loci except the two from *PMVK* gene in the full-sib progeny showed no statistically

significant deviation from the expected segregation ratio of genotypes ($P > 0.10, 0.05$) in the mapping population. The results further suggest the role of intragenic recombination in the induction of allele richness rather than somatic mutations which doesn't follow Mendelian inheritance.

Linkage analysis of SNPs in latex biosynthesis genes: Linkage analysis using JoinMap was performed as it was done earlier and the groups were analysed separately. The SNPs anchoring MVK locus, was mapped on LG-18 (18^{th} linkage group) of the female parent RRII 105. The tight linkage of this marker with an already reported SNP marker MKF197C from the same gene suggests that despite the high SNP frequency and presence of several recombination hot spots they segregate together as a single block. In the map of the male parent RRII 118, *FDPS* marker got integrated to the LG-3 (3^{rd} linkage group) of the earlier reported map of RRII 118 and as expected, all the markers which were present in the previous LG-3 group were present in the new group also confirming their mapping position. Since six new markers (four RAPD markers, one AFLP marker and one SSR marker) which were not linked earlier got integrated to the new map, it is assumed that the new marker have fitted in to a gap position which interlinked the previously unlinked markers to form a new group. The positioning of this marker is further confirmed by its proximity to the earlier reported STS marker developed from the same gene (*FDPS*). Though the old marker GGDPF88 from the gene *GGDPS* remained unlinked in the previous map, its interlinking with a new marker from the same gene (HbGGDPS741CT) and two additional markers yielded a new linkage group showing the significance of the new marker in further improving the old map. The markers HbHMGS3059AG and Hbcis-pre-1438CT from the genes, *HMGS* and *CPT* genes respectively remained unlinked in the present analysis. Even though these findings could not be utilized for any direct marker trait association study at this moment due to its low saturation level, incorporation of additional markers in future may render it a very useful linkage map for QTL analysis of yield related traits.

Segregation distortion is referred to as a deviation from an expected Mendelian ratio for a given genotype within a segregating population (Dufour *et al.*, 2001). Distorted segregation is a common feature noticed among different crops. In the present

study, the segregation distortion ($P \leq 0.05$) was detected only with three markers (HbPMVK1786CT, HbPMVK102CT and HbPMVKindel) from the *PMVK* gene. Lorieux *et al.* (1995) reported that segregation distortion had less impact on the recombination frequencies of co-dominant markers compared to the dominant markers. Since recombination breaks down the segregation distortion gene complex (Campos et al., 2014) the prediction of just two recombination sites from the entire 9 kb genomic region of *PMVK* suggests that *PMVK* gene is less prone to recombination resulting in distorted segregation of the entire gene. Segregation distortion of the entire *PMVK* gene is evident from the distorted ratios of the three widely spaced markers from different regions of the gene. Segregation distortion may occur due to several reasons such as deleterious genes segregating in the mapping population, heterogeneity within the parents (Grandillo and Tanksley, 1996), chromosomal rearrangements (Tanksley, 1984), reproductive differences between the two parents (Foolad *et al.*, 1995) and pre or post zygotic selection of genes flanking the distorted loci (Joobeur *et al.*, 2000). The genetic interactions among loci were also cited as a major factor inducing segregation distortions (McDaniel *et al.*, 2007). Since the distortion was observed only in the case of *PMVK* gene and all loci exhibited an excess of the maternal allele, we assume that the inter locus interactions involving nuclear–cytoplasmic factors may be the contributing agent for distortion rather than the other reasons cited above. Another possibility for the gene specific segregation distortion in *Hevea* is the selection pressure on a particular gamete and/or zygote whereby certain gametes or zygotic combinations have a reduced chance of survival resulting in skewed progeny distributions as reported by Lambrides *et al.* (2004). Since the distorted markers are excellent tools to gauge the level of segregation distortion, these markers will aid in understanding the behavioural pattern of hybrids developed for the crop improvement in *Hevea* species (Lambrides *et al.*, 2004)

Though all the three PMVK markers showed skewed segregation ratio, two of the marker got into a linkage group whereas the other did not. Other than the *PMVK* gene markers only one marker was present in the concerned linkage group within a distance of 12.6cM. Oliveira *et al.*, (2004) reported that the location of skewed markers within linkage groups influence the genetic distances of maps if they are placed in the

middle of a linkage group. Moreover such markers are located at the end of the linkage groups, sometimes forming entire linkage groups, without causing significant distance modifications. This finding favours the linkage group positioning of *PMVK* markers in the present map which formed a linkage group with the existing marker which was not linked in the previous map.

The present study forms the first report of mapping the genes *GGDPS* and *PMVK* in a linkage map in rubber. Like the segregation information obtained from the current study, SNP variants analysis and scoring of alleles by HRM technique can be extended to any gene of interest (candidate genes) and can be mapped in existing linkage maps in rubber. This approach is particularly useful for species like rubber where perfectly annotated whole genome sequences are not available but abundant transcript data from various genotypes are accessible.

4.6. Relative quantification of rubber biosynthesis genes and their association with haplotypes

Specific characteristics of plants like the level of synthesis of proteins and trait phenotypes are usually coordinated by genetic and environmental factors through the regulation of the concerned genes expression (Xu *et al.*, 2016). Therefore, expression level of candidate genes can be considered an intermediate phenotype from which conclusions about the emergence of trait of interest can be drawn (Ackermann *et al* ., 2013; Dimas *et al.*, 2009). The identification of functional SNPs in genes and analysis of their effects on phenotype may lead to better understanding of their impact on gene function for varietal improvement (Huq *et al.*, 2016). There are many reports about the effect of SNP on gene function in different crop plants (Schreiber *et al.*, 2014; Kharabian-Masouleh *et al.*, 2012; Vidal *et al.*, 2012; Xia *et al.*, 2013; Hirakawa *et al.*, 2013) All these studies shows that SNP plays an important role in determining traits of interest in crop plants by influencing the expression of concerned genes.

Since environment and haplotype were the two leading factors impacting gene expression, we explored the possibility of association of specific haplotypes of the

rubber biosynthesis genes in popular clones with their expression. Though single SNPs may have sometimes major impact on the phenotype of a plant, the rationale for testing for associations between phenotypes and haplotypes, rather than single SNPs, is due to the epistastic interactions between SNPs at a locus captured by the haplotypes (Bardel *et al.*, 2005). This is because the possibility of epistasis among SNPs is much greater when SNP density is very high, which is usually the case of SNPs in candidate genes (Clark, 2004). Using genotype and gene expression data from HapMap populations, Dimas *et al.* (2008) have proposed that interactions between protein-coding and regulatory SNPs may be common, and there is some evidence for SNP-SNP interactions within genes and gene clusters. At the same time there is no evidence to suggest that epistasis occurs frequently between randomly chosen SNPs hundreds of kilo bases apart (Hamon *et al.*, 2006). The ability to predict the downstream effects of genetic variation is critically important for understanding both the evolution of the genome and the molecular basis of expression of characters (Nackley *et al.*, 2006).

Association of different haplotypes with the expression of concerned gene is well established by various studies in the past (Kelly *et al.*, 2013). For example, relative gene expression studies on different haplotypes of the *TaGW2* gene promoter in wheat responsible for wheat grain width and kernel weight shows better expression in a particular haplotype than the rest (Qin *et al.*, 2014). Similarly He *et al.* (2006) detected haplotype specific expression variation in the gene cluster associated with a quantitative trait locus for improved yield in rice. Several such studies established that association based on haplotypes is more effective than single marker in dissection of complex traits. In the above context the possibility of association between a haplotype rather than a single SNP on the expression and finally on the yield trait in rubber is high. Therefore qPCR studies of rubber biosynthesis genes was performed to identify putative association between SNP haplotypes identified in the present study and gene expression in the clones RRII 105, RRI I118, RRIM 600, RRIC 52 and GT1. Since rubber biosynthesis related genes were reported to show higher expression in the latex compared to bark and leaf, transcript levels of these genes were estimated in the latex

collected from the five clones (Kush *et al.*, 1990; Suwanmanee *et al.*, 2002; Ko *et al.*, 2003; Asawatreratanakul *et al.*, 2003)

From the gene expression data it is evident that RRII 105 has consistently higher expression of all the eight genes followed by RRII 118. Studies on the gene expression profiles of latex biosynthesis genes support this finding (Priya, 2005; Uthup *et al.*, 2016). For *HMGS* gene if there exist any correlation between gene expression and haplotype, the best candidate will be the combination of Hap_1 and Hap_2 due to the significantly higher expression levels noted in RRII 105 (have both the haplotypes). The effect of Hap_2 alone on expression was ruled out as the same is present in RRII 118 without any significant impact, whereas the possibility of Hap_1 alone could not be neglected since it is unique to RRII 105. Similar trend was also noted in the case of *CPT* where RRII 105 has the unique combination of the haplotypes Hap_1 and Hap_2. Therefore there is a possibility that Hap_1 and Hap_2 may be associated with higher CPT expression. Contrary to the possible impact of heterozygous haplotype combinations in *HMGS*, the three clones which showed high expression rates of *HMGR* gene had homozygous haplotype status. Hap_1 as well as Hap_4 under homozygous status may have putative association with gene expression. Similar trend was noted in *FDPS* gene also where significantly lower expression was noted for RRIM 600, RRIC 52 and GT1 compared to RRII 105 despite the presence of Hap_1 in heterozygous state in RRIM 600 and GT1. Therefore it is assumed that Hap_1 may be contributing towards expression only when present in homozygous state as in the case of RRII 105. This assumption is supported by studies in loblolly pine where plants homozygous for the rarer alleles in genes involved in wood development had several fold higher expression than heterozygous plants and those homozygous plants for the common alleles (Palle *et al.*, 2013).

In the case of *MVK* gene, both RRII 105 and RRII 118 had comparable expression rates which were significantly higher than the expression in the other three clones. From the haplotype information available it may be assumed that the combinations Hap_1 & Hap_2 (RRII 105) and Hap_3 & Hap_4 (RRII 118) may have some influence on gene expression. But the haplotypes Hap_2 & Hap_3 alone or in any

other combination as seen in GT1 (Hap_2 and Hap_3) may have a negative effect on expression. As in the case of *HMGR* and *FDPS* genes, Hap_2 present in homozygous state in *PMVK* gene might have influenced the expression. Hap_2 in combination with Hap_1 also gave better expression as seen in RRII 105 whereas Hap_1 or Hap_2 in combination with other haplotypes have no positive effect towards expression. The expression profiles of *GGDPS* and *REF* genes did not showed any indications towards the existence of association between their haplotypes in the five clones. In these two genes despite having the same allelic combinations, the clones exhibited significantly divergent expression rates yielding no meaningful conclusion from the analysis. However the significance of *REF* gene structure on its expression cannot be ruled out based on the presently available data due to its crucial role in latex production. Previous studies suggest that *REF* might contribute to rubber biosynthesis by facilitating the biogenesis of large rubber particles (LRPs) or the 'growing' of small rubber particles (SRPs) to LRPs which makes up 93% of the rubber by volume in the latex (Yeang *et al.*, 1995; Dennis *et al.*, 1989; Priya *et al.*, 2005).

As mentioned in the introduction section, *GGDPS* and *FDPS* are two major enzymes responsible for the 'prenylation', or the sequential condensations of IPP with allylic diphosphates to generate isoprenoid carbon skeletons leading to the synthesis of natural rubber and their concentrations plays a significant role in determining the rate of the chain elongation which ultimately affects the rate of rubber production in *Hevea*. A notable observation made regarding the allele homozygosity and gene expression level in each clone is the correlation between higher homozygosity in the SNP loci of these two genes with their higher gene expression rate in RRII 105. Therefore it is assumed that if both the alleles of these two genes have an additive effect leading to a higher expression, it may lead to the enhanced synthesis of their products (FDP and GGDP), which may ultimately lead to more natural rubber. Since RRII 105 is considered to be the best clone in terms of *FDPS* gene expression and rubber productivity, the above data makes sense as an indirect correlation can be observed between the higher allele homozygosity of *FDPS* (100%) and *GGDPS* (93%) genes and their expression. Further extrapolation of this assumption may end up in the conclusion that the high homozygosity of these two genes may be acting as contributing factors for the high

productivity of this clone. The above assumptions and statements are contradictory to the general hypothesis that heterosis at the genome level is beneficial for the survival of the plant as well as for the expression of certain beneficial traits in crop plants. But significant increase in expression rate of specific genes were reported in several transgenic lines which are homozygous only at the loci of interest leading to higher expression of specific phenotypes (James *et al.*, 2002; Bourdon *et al.*, 2002). Simultaneously this benefit may not always be passed on to the next generation because recombination may disrupt the existing allelic combinations. This disadvantage can be overruled in the case of rubber because of its vegetative mode of propagation.

A comparative analysis of the homozygosity/heterozygosity of individual SNP loci under study in each of the eight gene in five clones suggest that overall homozygosity level was significantly more in the clone RRII 118 than the rest. The high homozygosity suggests that the parents of this clone are genetically more identical than the parents of the other clones even though much information is not available about its parents Mil 3/2 and Hil 28. Though the clones RRII 105 and RRIM 600 have comparable overall average homozygosity levels, none of the eight genes were 100% homozygous in the latter. Since they share the same female parent Tjir1, based on the available data it may be inferred that the male parent of RRIM 600 was more divergent than that of RRII 105. As mentioned elsewhere, functional SNPs are mostly non-synonymous SNPs and the presence of the favourable allele or dominant allele of selected genes in homozygous state may have an additive impact on the gene expression as well as on the phenotype. This may be true in the case of intronic as well as SNPs existing in the 5'UTR also as many such SNPs are considered to have significant functional role in the expression of the concerned gene (Palle *et al.*, 2013; Guyon-Debast *et al.*, 2010). This is relevant in the case of *Hevea* since more than 65% of the SNPs from rubber biosynthesis genes were of intronic origin. Intronic variants might affect alternative splicing of the mRNA by changing sequences involved in splicing like donor or acceptor sites like the changes induced in *FDPS* gene (Uthup *et al.*, 2013). Intron-variants could also act as enhancers resulting in the enhancement of expression of many genes. Syed *et al.* (2012) suggest that intronic SNPs may not be regulatory always but can be correlated with another SNP that actually has an impact on the

regulation of the gene. Moreover introns of a gene may also contain downstream regulatory elements known as DREs and their modifications may ultimately affect the gene regulation and signalling including post transcriptional modifications, and in this way the overall expression level as well as pattern of expression in a tissue type get affected (Greenwood and Kelsoe, 2003). Based on the above observations it may be assumed that homozygous intronic SNPs in *Hevea* may have an equally important role as that of non-synonymous SNPs.

4.7. *In silico* modelling studies on protein 3D structure

As mentioned earlier, the downstream effect of genetic variations occurring in the coding regions may ultimately leads to structural and functional changes at the protein level. Protein structural changes associated with non-synonymous SNPs are known to alter agronomic traits in plants (Weng *et al.*, 2013). In this context, *in silico* prediction of the impact of six non-synonymous SNPs in *HMGS* on the protein structural conformation variations is essential for establishing association with its catalytic activity. Putative structure of *Hevea* HMGS was predicted first *via in silico* molecular modelling approach by Sirinupong *et al.* (2005) using the crystal structure of ACP synthase III from *Mycobacterium tuberculosis* (PDB id: 1HZP) as the template citing the distant relation of HMGS to beta-ketoacyl acyl carrier protein synthase III (ACP synthase III). Unfortunately, sequence similarity between the two proteins was around 25% resulting in a model of inferior quality having less reliability. The model generated in the present study claims higher reliability because HMGS from *Brassica Juncea* (PDB id: 2FA3) which shared over 85% sequence identity with *Hevea,* was used as the template. Since none of the non-synonymous SNPs overlapped the acetyl-CoA binding region in 2FA3, we assume that the non-synonymous SNPs identified in the study may not influence the catalytic function of the protein and may result only in minor structural variations in the N and C terminals. Moreover we could also ascertain that the non-synonymous SNPs are directly not responsible for any alteration in the functional properties of the enzyme leading to changes in the quantity or quality of latex produced. The present study forms the first structural modelling study carried out on HMGS from *Hevea* with plant protein as the template and it may be considered as a reliable theoretical model for any prospective HMGS structural analysis studies in

Hevea. Similar results were also obtained for the *GGDPS* gene where protein crystal structure 2J1P (Geranylgeranyl diphosphate synthase from *Sinapis alba*) was selected as best template for modelling which shared over 78% sequence identity with *Hevea GGDPS* gene. In both the cases the non-synonymous SNPs are not interfering with the functional properties of the enzyme directly. But the actual scenario will be revealed only after the analysis of the protein crystal structure of *H.brasiliensis* itself since proteins are known to incorporate species specific modifications to suit its own special requirements (Pan *et al.*, 2016). The present study forms the first attempt to predict *in silico*, the possible impact of non-synonymous SNPs on the three dimensional structure of any protein in *Hevea.*

4.8. Summery and conclusions

A total of 172 SNPs were identified from eight major rubber biosynthesis genes in rubber showing their rich allelic diversity. Haplotype structuring was completed for these genes and potential recombination hot spots were identified shedding light on their sequence structure. Analysis of the allelic status of each SNP loci in the initiator molecule synthesising genes (*FDPS* and *GGDPS*) indicated that the high allele homozygosity observed in RRII 105 and RRIM 600 may have directly or indirectly influenced their productivity. Segregation analysis and computational prediction of recombination hot spots indicated that majority of SNPs and haplotypes in popular rubber clones are inherited through genetic recombination rather than induction through somatic mutations. Segregation studies in a progeny population using markers developed from the SNPs in eight rubber biosynthesis genes helped in the integration of these genes into an existing linkage map in rubber. The study also revealed the existence of retrotransposon activity in the intronic region of functional genes in *Hevea* and their role in the creation of indels and SNPs in intragenic regions. Furthermore the structural modifications induced by the retro-element eventually resulted in regulatory motif and splice site modifications in the *FDPS* gene of *Hevea*, which may have functional role in gene expression. Comparative protein sequence studies in several plant species revealed that some of the SNPs identified in *Hevea* are highly conserved across species and genera whereas some are unique to *Hevea* shedding light on the evolution of rubber biosynthesis genes in *Hevea brasiliensis. In silico* 3D structural

conformational studies of HMGS and GGDPS protein revealed that the non-synonymous SNPs in their gene sequence did not interfere with their active binding sites. Expression studies of latex biosynthesis genes in different genotypes indicated the possible association of specific haplotype/haplotype-combinations with gene expression and ultimately to latex yield. Above all the high allele diversity suggests that Wickham clones still hold rich genetic diversity within them which may be exploited further. Altogether the study is the first of its kind where extensive sequence structure analysis was carried out in any of the rubber biosynthesis genes in *Hevea*. The results from this learning indicate that despite the universal presence of mevalonate pathway genes in all living organisms, their sequence structure varies depending on the functional requirement of each organism. Moreover genes sequence information in *Hevea* clones indicates their species specific evolutionary characteristics which may be responsible for the high latex producing capacity of this wonder tree.

The information generated from this work may be used for diversity analysis as well as in the development of marker assisted selection (MAS) techniques for *Hevea* breeding. Prospectively, the information generated on the frequency of SNPs (nucleotide diversity) and the haplotypic diversity (heterozygosity) among individuals or within a population will be useful in the estimation of genetic diversity in rubber. SNPs and haplotypes which showed correlation with gene expression will be explored, in a large population for validation followed by their use as the marker of choice for early selection of potentially high-yielding genotypes in *Hevea* breeding programs. SNPs and haplotypes having association with latex biosynthesis in rubber can be transferred to the existing cultivated clones for genetic enhancement of latex production in *Hevea* through cisgenic approach. Additionally genetic mapping of latex biosynthesis genes on a linkage group will assist in understanding gene architecture and its interaction with neighbouring genes on a particular chromosome. The protein 3D models generated in the current study will be useful in screening out the functionally relevant mutations yet to be discovered in the future which may have significant impact on the activity of the concerned enzyme.

PUBLICATIONS AND PRESENTATIONS

Publications in internationally peer reviewed journals

Molecular evolution and functional characterisation of haplotypes of an important rubber biosynthesis gene in *Hevea brasiliensis* (2016) Thomas KU, Rajamani A, Ravindran M and Saha T. Plant Biology. doi:10.1111/plb.12433.

Impact of an intragenic retrotransposon on the structural integrity and evolution of a major isoprenoid biosynthesis pathway gene in *Hevea brasiliensis*. (2013) Thomas KU, Saha T, Ravindran M, Bini.K. Plant Physiology and Biochemistry 73: 176-188.

Presentations in conferences

Sequence structure analysis and haplotyping of HMGCo-A synthase gene in popular *H.brasiliensis* clones. Thomas K U, Anantharamanan R, Ravindran M and Saha T. Molecules in Living Cells:Mechanistic basis of function, Annual Meeting of The Indian Biophysical Society, during 8–10 February, 2016, Indian Institute of Science, Bangalore.

Structural variations associated with transposable element in an isoprenoid biosynthesis pathway gene in rubber (*Hevea brasiliensis*). K.U.Thomas, M.Ravindran, B.Karthika and T.Saha. First international and third national conference on Biotechnology, Bioinformatics and Bioengineering, 28th to 29th of June 2013, Tirupati, Andhra Pradesh. **(Won the best paper award)**

Sequences submitted to NCBI Genbank database

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447233- *Hevea brasiliensis* isolate RRII-105 geranylgeranyl diphosphate synthase (GGPS) gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447234- *Hevea brasiliensis* isolate RRII-118 geranylgeranyl diphosphate synthase (GGPS) gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447235- *Hevea brasiliensis* isolate RRIM-600 geranylgeranyl diphosphate synthase (GGPS) gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447236- *Hevea brasiliensis* isolate RRIC-52 geranylgeranyl diphosphate synthase (GGPS) gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447237- *Hevea brasiliensis* isolate GT1 geranylgeranyl diphosphate synthase (GGPS) gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447238- Hevea brasiliensis isolate RRII-105 cis-prenyltransferase gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447239- *Hevea brasiliensis* isolate RRII-118 cis-prenyltransferase gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447240- *Hevea brasiliensis* isolate RRIM-600 cis-prenyltransferase gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447241- *Hevea brasiliensis* isolate RRIC-52 cis-prenyltransferase gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447242- *Hevea brasiliensis* isolate GT1 cis-prenyltransferase gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447243- *Hevea brasiliensis* isolate RRII-105 rubber elongation factor (REF) gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447244- *Hevea brasiliensis* isolate RRII-118 rubber elongation factor (REF) gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447245- *Hevea brasiliensis* isolate RRIM-600 rubber elongation factor (REF) gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447246- *Hevea brasiliensis* isolate RRIC-52 rubber elongation factor (REF) gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447247- *Hevea brasiliensis* isolate GT1 rubber elongation factor (REF) gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT749568 - *Hevea brasiliensis* isolate RRII-105 farnesyl diphosphate synthase (FDPS) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT749569 - *Hevea brasiliensis* isolate RRIC-52 farnesyl diphosphate synthase (FDPS) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT749570 - *Hevea brasiliensis* isolate GT1 farnesyl diphosphate synthase (FDPS) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2014)
KM272630- *Hevea brasiliensis* mevalonate kinase (MVK) gene, complete cds

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447220 - *Hevea brasiliensis* isolate RRII-118 mevalonate kinase (MVK) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447221 - Hevea brasiliensis isolate RRIM-600 mevalonate kinase (MVK) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447222 -*Hevea brasiliensis* isolate RRIC-52 mevalonate kinase (MVK) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447223 - *Hevea brasiliensis* isolate GT1 mevalonate kinase (MVK) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KM272629 - *Hevea brasiliensis* hydroxymethylglutaryl coenzyme A synthase isoform (HMGS-2) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447224 -*Hevea brasiliensis* isolate RRII-118 hydroxymethylglutaryl-CoA synthase (HMGS) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447225-*Hevea brasiliensis* isolate RRIM-600 hydroxymethylglutaryl-CoA synthase (HMGS) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447226 - *Hevea brasiliensis* isolate RRIC-52 hydroxymethylglutaryl-CoA synthase (HMGS) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447227 - *Hevea brasiliensis* isolate GT1 hydroxymethylglutaryl-CoA synthase (HMGS) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447228 - *Hevea brasiliensis* isolate RRII-105 hydroxymethylglutaryl-CoA reductase (HMGR) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447229 - *Hevea brasiliensis* isolate RRII-118 hydroxymethylglutaryl-CoA reductase (HMGR) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447230 -*Hevea brasiliensis* isolate RRIM-600 hydroxymethylglutaryl-CoA reductase (HMGR) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447231-*Hevea brasiliensis* isolate RRIC-52 hydroxymethylglutaryl-CoA reductase (HMGR) gene, complete cds.

Uthup,T.K., Ravindran,M. and Saha,T (2015)
KT447232-*Hevea brasiliensis* isolate GT1 hydroxymethylglutaryl-CoA reductase (HMGR) gene, complete cds.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886384 - *Hevea brasiliensis* isolate RRII-118 farnesyl diphosphate synthase (FINT1) gene, FINT1-A allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886385 - *Hevea brasiliensis* isolate RRII-118 farnesyl diphosphate synthase (FINT1) gene, FINT1-C allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886386 - *Hevea brasiliensis* isolate RRII-105 farnesyl diphosphate synthase (FINT1) gene, FINT1-B allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886387 -*Hevea brasiliensis* isolate RRIM-600 farnesyl diphosphate synthase (FINT1) gene, FINT1-B allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886388 -*Hevea brasiliensis* isolate RRIC-52 farnesyl diphosphate synthase (FINT1) gene, FINT1-B allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886389 - *Hevea brasiliensis* isolate GT1 farnesyl diphosphate synthase (FINT1) gene, FINT1-B allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886390 - *Hevea brasiliensis* isolate RRII-5 farnesyl diphosphate synthase (FINT1) gene, FINT1-D allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886391 - *Hevea brasiliensis* isolate Rondonia-6 farnesyl diphosphate synthase (FINT1) gene, FINT1-unnamed-1 allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886392 - *Hevea brasiliensis* isolate Acre-9 farnesyl diphosphate synthase (FINT1) gene, FINT1-D allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886393 -Hevea *brasiliensis* isolate Acre-19 farnesyl diphosphate synthase (FINT1) gene, FINT1-A allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886394 - *Hevea brasiliensis* isolate Rondonia-10 farnesyl diphosphate synthase (FINT1) gene, FINT1-unnamed-2 allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886395 - *Hevea brasiliensis* isolate Acre-9 farnesyl diphosphate synthase (FINT1) gene, FINT1-C allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886396 - *Hevea pauciflora* isolate pau-1 farnesyl diphosphate synthase (FINT1) gene, FINT1-B allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886397 - *Hevea benthamiana* isolate ben-1 farnesyl diphosphate synthase (FINT1) gene, FINT1-unnamed-3 allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886398 - *Hevea nitida* isolate nit-1 farnesyl diphosphate synthase (FINT1) gene, FINT1-unnamed-4 allele, intron 1.

Uthup,T.K., Saha,T. and Ravindran,M(2013)
KC886399 -*Hevea spruceana* isolate spr-1 farnesyl diphosphate synthase (FINT1) gene, FINT1-unnamed-5 allele, intron 1.

SNPs submitted to the dbSNP database of NCBI with Sequence Submission No:

SNP Name	NCBI_SS No.	SNP Name	NCBI_SS No
HbHMGR786GC	1839574836	Hbcis-Pre851AC	1839574877
HbHMGR858GC	1839574837	Hbcis-Pre859AG	1839574878
HbHMGR888CT	1839574838	Hbcis-Pre971AG	1839574879
HbHMGR924AT	1839574839	Hbcis-Pre987CT	1839574880
HbHMGR1129CG	1839574840	Hbcis-Pre1036AC	1839574881
HbHMGR1149CT	1839574841	Hbcis-Pre1145CT	1839574882
HbHMGR1320AC	1839574842	Hbcis-Pre1187GT	1839574883
HbHMGR1368GT	1839574843	Hbcis-Pre1200TG	1839574884
HbHMGR1451AT	1839574844	Hbcis-Pre1312TC	1839574885
HbHMGR1588AT	1839574845	Hbcis-Pre1349GC	1839574886
HbHMGR1666CT	1839574846	Hbcis-Pre1366CT	1839574887
HbHMGR1693TG	1839574847	Hbcis-Pre1438TC	1839574888
HbHMGS2010GC	1839574848	HbGGPS87GA	1839574889
HbHMGS2089TA	1839574849	HbGGPS194CA	1839574890
HbHMGS2167AG	1839574850	HbGGPS258CT	1839574891
HbHMGS2180TC	1839574851	HbGGPS278AT	1839574892
HbHMGS2280TC	1839574852	HbGGPS303CT	1839574893
HbHMGS2705CA	1839574853	HbGGPS338AG	1839574894
HbHMGS2907TC	1839574854	HbGGPS361TC	1839574895
HbHMGS2921AG	1839574855	HbGGPS519AG	1839574896
HbHMGS2935TA	1839574856	HbGGPS535CA	1839574897
HbHMGS3059AG	1839574857	HbGGPS741CT	1839574898
HbHMGS3513CT	1839574858	HbGGPS835GA	1839574899
HbHMGS3598TC	1839574859	HbGGPS1097AG	1839574900
HbHMGS3734GT	1839574860	HbGGPS1116AG	1839574901
HbHMGS3766AG	1839574861	HbGGPS1173AG	1839574902
HbHMGS3817GT	1839574862	HbGGPS1249AC	1839574903
HbHMGS4052CT	1839574863	HbREF120CT	1839574904
HbHMGS4075CT	1839574864	HbREF130AG	1839574905
HbHMGS4345CT	1839574865	HbREF134AG	1839574906
HbHMGS4427GA	1839574866	HbREF140CG	1839574907
HbHMGS4616AT	1839574867	HbREF164AG	1839574908
Hbcis-Pre169GA	1839574868	HbREF367AC	1839574909
Hbcis-Pre550TC	1839574869	HbREF376AT	1839574910
Hbcis-Pre619TC	1839574870	HbREF429TC	1839574911
Hbcis-Pre631TC	1839574871	HbREF474TG	1839574912
Hbcis-Pre703AT	1839574872	HbREF482GA	1839574913
Hbcis-Pre714TA	1839574873	HbREF752AG	1839574914
Hbcis-Pre736AT	1839574874	HbREF781CT	1839574915
Hbcis-Pre760AG	1839574875	HbREF1230CT	1839574916
Hbcis-Pre811TA	1839574876	HbREF1334AG	1839574917

HbREF1359TC	1839574918
HbREF1382TA	1839574919
HbREF1439TA	1839574920
HbREF1452AC	1839574921
HbREF1454CT	1839574922
HbREF1456AG	1839574923
HbREF1457TA	1839574924
HbREF1460GC	1839574925
HbREF1495TC	1839574926
HbREF1505GA	1839574927
HbREF1508AG	1839574928
HbREF1520TC	1839574929
HbREF1522TG	1839574930
HbREF1533TG	1839574931
HbREF1542GC	1839574932
HbREF1552CA	1839574933
HbREF1585TA	1839574934
HbMVK71CA	1839574935
HbMVK84AG	1839574936
HbMVK119TG	1839574937
HbMVK349AG	1839574938
HbMVK441GT	1839574939
HbMVK470TA	1839574940
HbMVK1658CT	1839574941
HbMVK1774CT	1839574942
HbMVK1857TC	1839574943
HbMVK1908GA	1839574944
HbMVK2022CG	1839574945
HbMVK2320AT	1839574946
HbMVK2338CT	1839574947
HbMVK2379AG	1839574948
HbMVK2387AG	1839574949
HbMVK2420GA	1839574950
HbMVK2490GA	1839574951
HbMVK2518AT	1839574952
HbMVK2628AG	1839574953
HbMVK2874CA	1839574954
HbMVK2907AC	1839574955
HbMVK2912AT	1839574956
HbMVK3218AT	1839574957
HbMVK3297TC	1839574958
HbMVK3375CT	1839574959
HbMVK3432AG	1839574960
HbFDPS1380CT	1839574961
HbFDPS1585CT	1839574962
HbFDPS1627AG	1839574963
HbFDPS1918CT	1839574964
HbFDPS1934CG	1839574965
HbFDPS2013AC	1839574966
HbFDPS2105AT	1839574967
HbFDPS2127GC	1839574968
HbFDPS2521GC	1839574969
HbFDPS2755AG	1839574970
HbFDPS2860AT	1839574971
HbFDPS3356AG	1839574972
HbFDPS3475AT	1839574973
HbFDPS3743GT	1839574974
HbFDPS3764TA	1839574975
HbFDPS3800AT	1839574976
HbFDPS3850TA	1839574977
HbFDPS4234TC	1839574978
HbFDPS4420AT	1839574979
HbFDPS4489CT	1839574980
HbFDPS4659GA	1839574981
HbFDPS4707TC	1839574982
HbFDPS4804GA	1839574983
HbFDPS4813AG	1839574984
HbFDPS4929GA	1839574985
HbPMVK102CT	1839574986
HbPMVK297AG	1839574987
HbPMVK1443AT	1839574988
HbPMVK1787CT	1839574989
HbPMVK2046CT	1839574990
HbPMVK2307AG	1839574991
HbPMVK2485AT	1839574992
HbPMVK2488CT	1839574993
HbPMVK2596AG	1839574994
HbPMVK2669AT	1839574995
HbPMVK2787GT	1839574996
HbPMVK3104AG	1839574997
HbPMVK3141GT	1839574998
HbPMVK3227CT	1839574999

plant biology

Plant Biology ISSN 1435-8603

RESEARCH PAPER

Molecular evolution and functional characterisation of haplotypes of an important rubber biosynthesis gene in *Hevea brasiliensis*

T. K. Uthup, A. Rajamani, M. Ravindran & T. Saha

Genome Analysis Laboratory, Rubber Research Institute of India, Kottayam, Kerala, India

Keywords
Haplotype structure; *Hevea brasiliensis*; HMGCoA synthase; isoprenoid biosynthesis; SNPs.

Correspondence
Thomas K. Uthup, Genome Analysis Laboratory, Rubber Research Institute of India, Rubber Board PO, Kottayam 686009, Kerala, India.
E-mail: thomasku79@gmail.com

Editor
U. Wittstock

Received: 26 August 2015; Accepted: 12 January 2016

doi:10.1111/plb.12433

ABSTRACT

Hydroxy-methylglutaryl coenzyme-A synthase (HMGS) is a rate-limiting enzyme in the cytoplasmic isoprenoid biosynthesis pathway leading to natural rubber production in *Hevea brasiliensis* (rubber). Analysis of the structural variants of this gene is imperative to understand their functional significance in rubber biosynthesis so that they can be properly utilised for ongoing crop improvement programmes in *Hevea*. We report here allele richness and diversity of the HMGS gene in selected popular rubber clones. Haplotypes consisting of single nucleotide polymorphisms (SNPs) from the coding and non-coding regions with a high degree of heterozygosity were identified. Segregation and linkage disequilibrium analysis confirmed that recombination is the major contributor to the generation of allelic diversity, rather than point mutations. The evolutionarily conserved nature of some SNPs was identified by comparative DNA sequence analysis of HMGS orthologues from diverse taxa, demonstrating the molecular evolution of rubber biosynthesis genes in general. *In silico* three-dimensional structural studies highlighting the structural positioning of non-synonymous SNPs from different HMGS haplotypes revealed that the ligand-binding site on the enzyme remains impervious to the reported sequence variations. In contrast, gene expression results indicated the possibility of association between specific haplotypes and HMGS expression in *Hevea* clones, which may have a downstream impact up to the level of rubber production. Moreover, haplotype diversity of the HMGS gene and its putative association with gene expression can be the basis for further genetic association studies in rubber. Furthermore, the data also show the role of SNPs in the evolution of candidate genes coding for functional traits in plants.

INTRODUCTION

Hevea brasiliensis (Euphorbiaceae) is the chief source of natural rubber, a critically strategic industrial raw material that is indispensable for the production of a wide range of products having diverse applications, ranging from healthcare to space science. Natural rubber is practically pure poly-cis-1,4 isoprene. It is synthesised in the latex vessels of rubber trees by the mevalonate or 2-C-methyl-D-erythritol 4-phosphate pathway (Chow *et al.* 2012). The limitations in conventional breeding practices to increase rubber productivity prompted investigations into the underlying molecular mechanisms of natural rubber biosynthesis, leading to identification of several important genes and enzymes involved (Tang *et al.* 2013). 3-Hydroxy-3-methylglutaryl coenzyme-A synthase (HMGS) is an important gene in the mevalonate pathway, which catalyses the condensation of acetyl-CoA and acetoacetyl-CoA to form HMG-CoA. This HMG-CoA acts as the substrate for HMG-CoA reductase (HMGR) to yield mevalonate, which is further converted to different isoprenoid compounds, e.g. growth regulators, chlorophyll, phytoalexins, natural rubber, etc. (Suwanmaneea *et al.* 2002). The involvement of HMGS and HMGR in the early steps of rubber biosynthesis and the positive correlation between their activity and the dry rubber content of the latex from rubber trees is well established (Wititsuwannakul 1986; Suvachittanont & Wititsuwannakul 1995). Even though clear-cut evidence at the gene level is not available so far to link expression with the quality and quantity of latex produced, the rate-limiting role of these two enzymes in the mevalonate pathway and results from the aforementioned studies prompted us to carry out further investigations.

Assessment of genetic diversity using molecular markers is required, not only for crop improvement programmes but also for suitable management and conservation of plant genetic resources in gene banks (Varshney *et al.* 2007). Among the different types of molecular marker used, single nucleotide polymorphisms (SNPs) have gained popularity recently because of their abundance and ease of automation. Furthermore, an SNP is of great importance if it affects gene function, thereby directly influencing the phenotype in terms of yield or stress response. Such information can be used in breeding programmes to develop better varieties using approaches such as marker-assisted recurrent selection (MARS) or gene pyramiding. In *Hevea*, the advent of next generation sequencing platforms significantly broadened the scope of SNP markers through identification of thousands of SNPs from RNA

Plant Physiology and Biochemistry 73 (2013) 176–188

Contents lists available at ScienceDirect

Plant Physiology and Biochemistry

journal homepage: www.elsevier.com/locate/plaphy

Research article

Impact of an intragenic retrotransposon on the structural integrity and evolution of a major isoprenoid biosynthesis pathway gene in *Hevea brasiliensis*

Thomas Kadampanattu Uthup*, Thakurdas Saha, Minimol Ravindran, K. Bini

Genome Analysis Laboratory, Rubber Research Institute of India, Rubber Board, P O, Kottayam, Kerala Pin-686009, India

ARTICLE INFO

Article history:
Received 2 May 2013
Accepted 10 September 2013
Available online 24 September 2013

Keywords:
Farnesyl diphosphate synthase
Hevea brasiliensis
Introns
Isoprenoid biosynthesis
Retrotransposons
SNPs

ABSTRACT

Isoprenoids belong to a large family of structurally and functionally different natural compounds found universally from prokaryotes to higher animals and plants. In *Hevea brasiliensis*, the commercially important *cis*-polyisoprene (rubber) is synthesised as part of its defence mechanism in addition to other common isoprenoids like phytosterols, growth hormones etc. Farnesyl diphosphate synthase (*FDPS*) is a key enzyme in this process which catalyses the conversion of isoprene units into polyisoprene. Although prior sequence information is available, the structural variants of the *FDPS* gene presently existing in *Hevea* population are largely unknown. Since gene structure has a major role in gene regulation, extensive sequence analysis of this gene from different genotypes was carried out to identify the prevailing structural variants. We identified several SNPs and large indels which were associated with a partial transposable element (TE). Modification of key regulatory motifs and splice sites induced by the retroelement was also identified in the first intron. Screening of popular rubber clones, wild germplasm accessions and *Hevea* species revealed that the retroelement is responsible for the generation of new alleles with varying degrees of sequence homology. Segregation analysis of a progeny population confirmed that the alleles are not paralogs and are inherited in a Mendelian mode. Our findings suggest that the first intron of the *FDPS* gene has been subjected to various chromosomal rearrangements due to the interaction of a retrotransposon, resulting in novel alleles which may substantially contribute towards the evolution of this major gene in rubber. Moreover, the results indicate the possible existence of a retrotransposon-mediated epigenetic gene regulatory mechanism in *Hevea*.

1. Introduction

Hevea brasiliensis (Willd. ex A. Juss.) Müll. Arg is a tropical rubber producing tree that yields 90% of the natural rubber needed by the worldwide rubber industry [1]. The plant is extensively cultivated in Asia pacific countries like Malaysia, Indonesia, Thailand, Vietnam and parts of India and China. Even though the presently cultivated *Hevea* clones are considered to have a narrow genetic base due to several years of selective breeding from a few original seedlings, earlier studies shows that there exist significant variation among the clones in characters like disease resistance, abiotic stress tolerance and latex yield. Variation in phenotypic characters like yield, girth and other secondary characters are well established in *Hevea* clones by previous studies [2–4]. Moreover, the popular clones are reported to be of divergent nature in terms of disease resistance also [5]. These variations may be attributed mainly to the epigenetic as well as the still existing genetic diversity within them, already established by molecular markers studies by several groups [6–8]. Natural rubber is a polyisoprene (*cis*-1,4-polyisoprene) making part of isoprenoids, the oldest known bio-molecules with diverse families of organic compounds that are widespread in the three domains of life. Although they are produced by the condensation of the same precursors universally (isopentenyl diphosphate (IDP) and dimethylallyl diphosphate (DMAPP)), the genes involved in their biosynthesis have evolved independently in various ways to satisfy the specific needs of the concerned organism. The isoprenoids including *Hevea cis*-polyisoprene are primarily synthesised by the mevalonate pathway (MVA) in plants *via* isopentenyl diphosphate (IDP) as a common intermediate [9]. Farnesyl diphosphate synthase (FDPS) plays a key role in this pathway by mediating the catalysis of the sequential 1–4 condensations of IDP with DMAPP to produce geranyl diphosphate (GDP) and with GDP to give Farnesyl diphosphate (FDP), eventually used for the synthesis of sterols, prenylated proteins etc. Furthermore, it is the allylic diphosphate initiator for

* Corresponding author. Tel.: +91 4812353311x202; fax: +91 4812353311.
E-mail address: thomasku79@gmail.com (T.K. Uthup).

0981-9428/$ – see front matter © 2013 Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.plaphy.2013.09.004

Appendix 1

REAGENTS AND SOLUTIONS

1. DNA Extraction Buffer

   CTAB - 2% (W/V)

   NaCl - 1.4M

   EDTA - 20mM

   Tris HCl - 100mM (pH 8.o)

   β-mercaptoethanol - 0.3% (added just before use)

2. Chloroform:Isoamyl alcohol (24:1)

3. RNase A-(10 mg/ml)

   RNase powder - 10mg

   Sodium acetate - 0.5 M (pH 8.0)

   Autoclaved Milli Q water

4. TE Buffer (10:1)

   Tris-HCl – 10 mM (pH 8.0)

   EDTA – 1 mM

5. TAE Buffer per litre (50X)

   Tris base – 242 g

   Glacial acetic acid - 57.1 ml

   EDTA (0.5 M) – 100 ml (pH 8.0)

6. Ethidium Bromide

10 mg/ml in water

7. DNA Molecular Weight Marker

λ-DNA cleaved with *Eco* R1 and *Hin*d III

8. Gel Loading Buffer (6X)

Bromophenol Blue - 0.25%

Xylene Cyanol FF - 0.25%

Glycerol - 50%

MilliQ water - 10ml

9.CTAB buffer

2 % hexadecyl triethyl ammonium bromide,

1.4 M NaCl

20 mM EDTA (pH 8.0)

0.1 M Tris-HCl (pH 8.0)

1 % (w/v) polyvinyl poly pyrolidone (PVPP)

0.3 % β-mercaptoethanol (added just before use).

Acetic acid – 10%

10. LB medium broth (1litre)

Yeast Extract – 5 g

Sodium Chloride – 10 g

Tryptone – 10 g

NaOH (for adjusting pH)

MilliQ water

11. LB medium (1litre)

Yeast Extract – 5 g

Sodium Chloride – 10 g

Tryptone – 10 g

NaOH (for adjusting pH)

MilliQ water

Agar – 15 g

12. Ampicillin: Stock solution

Ampicillin sodium salt - 2.5 g

Deionized water - 50 ml

Filter sterilized and stored as aliquots at (-20) °C.

13. IPTG (0.1 M) solution

1.2 g IPTG

MilliQ water to make up final volume of 50 ml

Filter sterilized and stored at -20^{0}C

14. X-Gal solution

X-Gal - 100 mg

Dimethyl formamide - 2 ml

Covered with aluminium foil and stored at -20^{0}C

15. LB-Amp-Glycerol medium

LB broth - 100 ml

ampicillin - 100 µl

Glycerol - 7.5%

REFERENCES

Ackermann M, Sikora-Wohlfeld W, Beyer A (2013) Impact of natural genetic variation on gene expression dynamics. PLos Genet, doi.org/10.1371/journal.pgen. 1003514

Adiwilaga K, Kush A (1996) Cloning and characterization of cDNA encoding farnesyl diphosphate synthase from rubber tree (*Hevea brasiliensis*). Plant Mol. Biol, 30:935–946.

Akhtar TA., Matsuba Y, Schauvinhold I, Yu G, Lees HA, Klein SE, Pichersky E (2013), The tomato *cis*–prenyltransferase gene family. Plant J, 73:640–652.

Alex D, Bach TJ, Chye ML (2000) Expression of *Brassica juncea* 3- hydroxy-3-methylglutaryl-CoA synthase is developmentally regulated and stress-responsive. Plant J, 22:415–426.

Andreassi JL, Dabovic K, Leyh TS (2004) *Streptococcus pneumoniae* isoprenoid biosynthesis is downregulated by diphosphomevalonate: an antimicrobial target. Biochem, 43:16461–16466.

Andreassi JL, Vetting MW, Bilder PW, Roderick SL, Leyh TS (2009) Structure of the ternary complex of phosphomevalonate kinase: the enzyme and its family. Biochem, 48:6461–6468.

Arigoni D, Sagner S, Latzel C, Eisenreich W, Bacher A, Zenk MH (1997) Terpenoid biosynthesis from 1-deoxy-D-xylulose in higher plants by intramolecular skeletal rearrangement. Proc Natl Acad Sci USA, 94:10600-10605.

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The swiss-model workspace: a web-based environment for protein structure homology modeling. Bioinfo, 22: 195-201.

References

Asawatreratanakul K, Zhang YW, Wititsuwannakul D *et al.*, (2003) Molecular cloning, expression and characterization of cDNA encoding *cis*-prenyltransferases from *Hevea brasiliensis*. A key factor participating in natural rubber biosynthesis. Eur J Biochem, 270:4671–4680

Atan S, Low FC, Saleh NM (1996) Construction of a microsatellite enriched library from *Hevea brasiliensis*. J. Nat. Rubb. Res, 11:247–255.

Backhaus RA (1985) Rubber formation in plants - mini review. Israel J Bot, 34:283-293.

Baes P, Custsem V (1993) Electrophoretic analysis of eleven isozyme system and their possible use as biochemical markers in breeding chicory (*Chychorium intybus L.*). Plant Breed, 110: 16-23.

Barbaglia AM, Klusman KM, Higgins J, Shaw JR, Hannah LC, Lal SK (2012) Gene capture by helitron transposons reshuffles the transcriptome of maize. Genetics, 3: 965-975.

Bardel C, Danjean V, Hugot JP, Darlu P, Genin E (2005) On the use of haplotype phylogeny to detect disease susceptibility loci. BMC Genet, doi: 10.1186/1471-2156-6-24.

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinfo, 21: 263-265.

Beckmann JS, Soller M (1986) Restriction fragment length polymorphisms and genetic improvement of agricultural species. Euphytica, 35:111-124.

Begum H, Spindel JE, Lalusin A, *et al.*, (2015) Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). PLoS ONE, doi:10.1371/journal.pone.0119873.

Bennett MD, Leitch IJ (1995) Nuclear DNA amounts in angiosperms. Annals Bot , 76: 113-176.

Bennetzen J (2000) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell, 12: 1021–1029.

Bennetzen JL, Sanmiguel P, Chen M, Tikhonov A, Francki M *et al* (1998) Grass genomes. Pro Natl Acad Sci USA, 95: 1975–1978.

References

Bergmann F, Gregorius H-R, Scholz F (1989) Isoenzymes, indicators of environmental impacts on plants or environmentally stable gene markers? In: Scholz, F., Gregorius H-R, Rudin, D (eds.) Genetic effects of air pollutants in forest tree populations. Springer-Verlag Heidelberg, New York, Tokyo, 17-28.

Berthelot K *et al.* (2012). Rubber elongation factor (*REF*), a major allergen component in *Hevea brasiliensis* latex has amyloid properties. PloS ONE, doi:10.1371/journal.pone.0048065

Besse P, Lebrun P, Seguin M, Lanaud C (1993b) DNA fingerprints in *Hevea brasiliensis* (rubber tree) using human minisatellite probes. Heredity, 70:237–244.

Besse P, Seguin M, Lebrun P, Chevallier MH, Nicolas D, Lanaud C (1994) Genetic diversity among wild and cultivated populations of *Hevea brasiliensis* assessed by nuclear RFLP analysis. Theor Applied Genet, 88:199–207.

Besse P, Seguin M, Lebrun P, Lanaud C (1993a) Ribosomal DNA variations in wild and cultivated rubber tree (*Hevea brasiliensis*). Genome, 36:1049–1057.

Bhangu-Uhlmann A (2011) The Mevalonate Pathway: A monitoring approach in plants by systems biology tools. Ph.D Thesis, Technical University of Munich

Bhattramakki D, Dolan M, Hanafey M, Wineland R, Vaske D, Register JC, Tingey SV, Rafalski A (2002) Insertion deletion polymorphisms in 3 regions of maize genes occur frequently and can be used as highly informative genetic markers. Plant Mol. Biol, 48: 539–547.

Biasini M *et al.* (2014) Swiss-model: modeling protein tertiary and quaternary structure using evolutionary information. Nucleic. Acids Res, 42: 252-258.

Bini K (2013) Development of a genetic linkage map using molecular markers in Para rubber tree (*Hevea brasiliensis*). Ph.D thesis, University of Kerala.

Blackmore T, Thomas I, McMahon R, Powell W, and Hegarty M (2015) Genetic–geographic correlation revealed across a broad European ecotypic sample of perennial ryegrass (*Lolium perenne*) using array-based SNP genotyping. Theor Applied Genet, 128: 1917–1932.

Bloch K, Chaykin S, Phillips AH, Dewaard A (1959) Mevalonic Acid Pyrophosphate and Isopentenylpyrophosphate, J Biol Chem, 234: 2595-2604.

References

Bokma E, Rozeboom HJ, Sibbald M, Dijkstra BW, Beintema JJ (2002) Expression and characterization of active site mutants of hevamine, a chitinase from the rubber tree *Hevea brasiliensis*. Eur J Biochem, 269: 893-901.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am. J. Hum. Genet, 32: 314-331.

Bourdon V, Ladbrooke ZL, Wickham A, Lonsdale D M, Harwood WA (2002) Homozygous transgenic wheat plants with increased luciferase activity do not maintain their high level of expression in the next generation. Plant Sci, 163:297-305. doi: 10.1016/S0168-9452(02)00095-X.

Bourguiba H, Audergon J, Krichen L, Trifi-Farah N, Mamouni A, Trabelsi S, D'Onofrio C, Asma B M, Santoni S, Khadari B (2012) Loss of genetic diversity as a signature of apricot domestication and diffusion into the Mediterranean Basin. BMC Plant Biol, doi:10.1186/1471-2229-12-49

Boylan MT, Quail PH (1991) Oat phytochrome is biologically active in transgenic tomatoes. Proc Natl Acad Sci. USA, 88:10806–10810.

Broun P, Somerville C (2001) Progress in plant metabolic engineering. Proc Natl Acad Sci. USA, 98: 8925-8927.

Brown SM, Kresovich S (1996) In Genome mapping in plants (ed. Paterson, AH), Clandes, New York, 85–93.

Byrne PF, McMullen MD, Snooks ME, Musket TA, Theuri JM, Widstrom NW, Wiseman BR, Coe EH (1996) Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. Proc Natl Acad Sci. USA, 93: 8820–8825.

Campos JL, Halligan DL, Haddrill PR, Charlesworth B (2014) The relation between recombination rate and patterns of molecular evolution and variation in *Drosophila melanogaster*. Mol Biol Evol, 31:1010-1028.

Carron, MP, Enjalric, LL, Deschamps A (1989). Rubber (*Hevea brasiliensis* Muell. Arg.). p. 222–245. In: Bajaj YPS (ed.), Biotechnology in agriculture and forestry. Vol, V. Springer-Verlag, Germany.

References

Carron, MP, Lardet L, Montoro P (2005) Different ways of integrating *in vitro* culture in *Hevea* planting material propagation. IRRDB Conference, Cochin, Nov. 7–11.

Carron, MP, Lardet L, Julien J, Boko C (2001) Somatic embryogenesis in *Hevea brasiliensis* (Muell.Arg.); Current advances and limits. In: J. Sainte-Beuve (ed.), Annual IRRDB meeting. CIRAD, Montpellier, France.

Chagné D, Gasic K, Crowhurst RN, Han Y, Bassett HC, Bowatte DR, Lawrence TJ, Rikkerink EHA, Gardiner SE, Korban SS (2008)Development of a set of SNP markers present in expressed genes of the apple. Genomics, 92:353–358.

Chaykin S, Law J, Phillips AH, Tchen TT, Bloch K (1958) Phosphorylated intermediates in the synthesis of squalene, Pro Natl Acad Sci USA, 44: 998-1004.

Chen A, Kroon PA, Poulter CD (1994) Isoprenyl diphosphate synthases: protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure. Protein Sci, 3:600–607.

Chen A, Poulter CD (1994) Isolation and Characterization of *idsA*: The gene for the short chain isoprenyl diphosphate synthase from *Methanobacterium thermoautotrophicum*. Arch Biochem Biophys, 314:399–404.

Chen Q, Sun F (2013) A unified approach for allele frequency estimation, SNP detection and association studies based on pooled sequencing data using EM algorithms. BMC genomics, doi: 10.1186/1471-2164-14-S1-S1.

Chevallier MH (1988) Genetic variability of *Hevea brasiliensis* germplasm using isozyme markers. J Nat Rubb Res, 3:42–53.

Chin PS, Chang WP, Lau CM, Pong KS (1974) Deproteinized natural rubber (DPNR). Proc Rub Res Inst Malaysia, Planters' conference, Kuala Lumpur, Malaysia: 252-262.

Cho MH, Ciulla D, Klanderman BJ, Raby BA, Silverman EK (2008) High-resolution melting curve analysis of genomic and whole-genome amplified DNA. Clin Chem, 54:2055–2058.

Choi IY *et al.*, (2007) A soybean transcript map:gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics, 176:685-696.

Chow KS, Isa MN, Bahari A, Ghazali A, Alias H, Mohd.-Zainuddin Z, Hoh C, Wan K L (2012) Metabolic routes affecting rubber biosynthesis in *Hevea brasiliensis* latex. J Exp Bot, 63:1863-1871

Chow KS, Wan KL, Isa MN, Bahari A, Tan SH, Harikrishna K, Yeang HY (2007) Insights into rubber biosynthesis from transcriptome analysis of *Hevea brasiliensis* latex. J Exp Bot, 58:2429-2440.

Chu Y, Su X, Huang Q, Zhang X (2009) Patterns of DNA sequence variation at candidate gene loci in black poplar (*Populus nigra L.*) as revealed by single nucleotide polymorphisms. Genetica, 137:141-150.

Clancy M. and Hannah, LC (2002) Splicing of the maize Sh1 first intron is essential for enhancement of gene expression, and a T-rich motif increases expression without affecting splicing. Plant Physiol, 130: 918–929.

Clark AG (2004) The role of haplotypes in candidate gene studies. Genet Epidemiol, 27: 321–333.

Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, *et al.,* (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana.* Science, 317:338-342.

Clément-Demange A, Legnate H, Seguin M, Carron MP, Le Guen V, Chapuset T, Nicolas D (2000) Rubber Tree. In: Tropical Plant breeding, Charrier A, Jacquot M, Hamon S, Nicolas D (ed.), 455-480, Collection Repères, ISBN 978-1-57808-144-8, Montpellier, France.

Concibido VC, Young ND, Lange DA, Denny RL, Danesh D, Orf JH (1996) Targeted comparative genome analysis and qualitative mapping of a major partial-resistance gene to the soybean cyst nematode. Theor Applied Genet, 93: 234-241.

Cong B, Barrero LS, Tanksley SD (2008) Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. Nat Genet. 40:800-804.

References

Cornish K (1993b) The separate roles of plant cis and trans prenyl transferases in cis-1, 4-polyisoprene biosynthesis. Eur J Biochem, 218: 267-271.

Cornish K (2001) Similarities and differences in rubber biochemistry among plant species. Phytochemistry, 57: 1123-1134.

Cornish K, Blakeslee JJ (2011) Rubber biosynthesis in plants. American Oil Chemist Society, The Lipid Library.

Cornish K, Siler DJ, Grosjean O, Goodman N (1993a) Fundamental similarities in rubber particle architecture and function in three evolutionarily divergent plant species. J Nat Rubb Res, 8:275–285.

Crabtree B, Newsholme EA (1987) The derivation and interpretation of control coefficients. Biochem. J, 247: 113–120.

Cunillera N, Arro M, Fores O, Manzano D, Ferrer A (2000) Characterization of dehydrodolichyl diphosphate synthase of *Arabidopsis thaliana*, a key enzyme in dolichol biosynthesis. FEBS Lett, 477: 170–174.

Da Costa BMT, Keasling JD, McMahan CM, Cornish K (2006). Magnesium ion regulation of *in vitro* rubber biosynthesis by *Parthenium argentatum* Gray. Phytochemistry, 67: 1621-1628.

Dennis MS, Light DR (1989) Rubber elongation factor from *Hevea brasiliensis.* Identification, characterization, and role in rubber biosynthesis. J Biol Chem, 264: 18608–18617.

Dereeper A, Nicolas S, Lecunff L, Bacilieri R, Doligez A, Peros JP, Ruiz M, This P (2011) SNiPlay: a web-based tool for detection management and analysis of SNPs. Application to grapevine diversity projects. BMC Bioinfo, 12:134.

Dibb NJ, Newman AJ (1989) Evidence that introns arose at proto-splice sites. EMBO J, 8:2015-2021.

Dimas AS, Stranger BE, Beazley C, Finn RD, Ingle CE, et al. (2008) Modifier effects between regulatory and protein-coding variation. Plos Genetics, doi.org/10.1371/journal.pgen.1000244

Dimas, AS *et al.*, (2009) Common regulatory variation impacts gene expression in a cell type–dependent manner. Science, 325:1246–1250

References

Dohm JC, Lange C, Reinhardt R, Himmelbauer H (2009) Haplotype divergence in *Beta vulgaris* and microsynteny with sequenced plant genomes. Plant J, 57:14-26.

Domazet-Loso T, Tautz D (2003) An evolutionary analysis of orphan genes in *Drosophila*. Genome Res., 13: 2213-2219.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus, 12:13-15.

Dufour P, Johnsson C, Antoine-Michard S, Cheng R, Murigneux A, Beckert M (2001) Segregation distortion at marker loci: variation during microspore embryogenesis in maize. Theor Applied Genet, 102: 993-1001.

Etienne C, Rotham C, Moing A, Plomion C, Bodenes C, Svanella-Dumas L, Cosson P, (2002)evaluation In: Applications of RAPD technology to plant breeding. Proc Joint Plant Breeding Symposia Series, Minneapolis, Minnesota, 9-11.

Feng SP, Li WG, Yu F, Wang JY, Wu YT (2010). Construction of genetic linkage map for rubber tree *(Hevea brasiliensis*) based on SSR markers. Hereditas, 8: 857-863

Fernando DM (1973) Trends in the improvement of rubber planting material with particular reference to Sri Lanka. Q J Rub Res Inst Sri Lanka (Ceylon), 50: 84-90.

Foolad MR, Arulsekar S, Becerra V, Bliss FA (1995) A genetic map of *Prunus* based on an interspecific cross between peach and almond. Theor Applied Genet, 91: 262-269.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, *et al.* (2002) The structure of haplotype blocks in the human genome. Science, 296: 2225–2229.

Gao D, Chen J, Chen M, Meyers BC, Jackson S (2012) A highly conserved, small LTR retrotransposon that preferentially targets genes in grass genomes. PLoS ONE, doi:10.1371/journal.pone.0032010.

Garg K, Green P, Nickerson DA (1999) Identification of candidate coding region single nucleotide polymorphisms in 165 human genes using assembled expressed sequence tags. Gemome. Res, 9: 1087-1092.

References

Garritano S, Gemignani F, Voegele C, Nguyen-Dumont T, Le Calvez-Kelm F, De Silva D, Lesueur F, Landi S , Tavtigian SV (2009) Determining the effectiveness of High Resolution Melting analysis for SNP genotyping and mutation scanning at the TP53 locus. BMC Genetics, doi: 10.1186/1471-2156-10-5

George PJ, Nair VKB, Panikkar AON (1980) Yield and secondary characters of the clone RRII 105 in trial plantings. International rubber conference, Kottayam.

Gibson G (2012) Rare and common variants: twenty arguments. Nat Rev Genet, 13:135–145.

Gore MA, Wright MH, Ersoz ES, Bouffard P, Szekeres ES, Jarvie TP *et al.*, (2009) Large-scale discovery of gene-enriched SNPs. Plant Genome, 2: 121–133.

Gowik U, Burscheidt J, Akyildiz M, Schlue U, Koczor M, Streubel M, Westhoff P (2004) *cis*-regulatory elements for mesophyll-specific gene expression in the C4 plant *Flaveria trinervia*, the promoter of the C4 *phosphoenolpyruvate carboxylase* gene. The Plant Cell, 16:1077-1090.

Graham MA, Silverstein KA, Cannon SB, Bosch Vanden KA(2004) Computational identification and characterization of novel genes from legumes. Plant Physiol, 135: 1179-1197.

Grandillo S, Tanksley SD (1996) QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium.* Theor Applied Genet, 92:935-51..

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics, 137: 1121–1137

Gray J, Kekwick R (1972) The inhibition of plant mevalonate kinase preparations by prenyl pyrophosphates. Biochem Biophys Acta Gen Subj, 279:290–296.

Greenwood TA, Kelsoe JR (2003) Promoter and intronic variants affect the transcriptional regulation of the human dopamine transporter gene. Genomics, 82:511-520.

Gronover CS, Wahler D, Prufer D (2011) Natural rubber biosynthesis and physic-chemical studies on plant derived latex, biotechnology of biopolymers, Prof. Magdy Elnashar (Ed.), ISBN: 978-953-307-179-4

Guex N, Peitsch MC (1996) Swiss-PdbViewer: a fast and easy-to-use PDB viewer for Macintosh and PC. Protein Data Bank Quaterly Newsletter, 77:7.

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis, 18:2714-2723.

Gupta PK, Varshey RK, Sharma PC, Ramesh B (1999) Molecular markers and their applications in wheat breeding. Plant breeding 118:369-390.

Guyon-Debast A, Lécureuil A, Bonhomme S, Guerche P, Gallois J (2010) A SNP associated with alternative splicing of RPT5b causes unequal redundancy between RPT5a and RPT5b among *Arabidopsis thaliana* natural variation. BMC Plant Biol, doi: 10.1186/1471-2229-10-158

Hahn MW, Kern AD (2005) Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol, 22: 803–806.

Hamon SC, Kardia SL, Boerwinkle E, Liu K, Klos KL, *et al.*, (2006) Evidence for consistent intragenic and intergenic interactions between SNP effects in the APOA1/C3/A4/A5 gene cluster. Hum Hered, 61: 87–96.

Hampton R, Dimster-Denk D, Rine J (1996) The biology of HMG-CoA reductase: the pros of contra-regulation. Trends Biochem Sci, 21: 140–145.

Han SJ (2010) Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. Mobile DNA, 1:15, doi: 10.1186/1759-8753-1-15

Han Y, Khu DM, Monteros MJ (2012) High-resolution melting analysis for SNP genotyping and mapping in tetraploid alfalfa (*Medicago sativa* L.) Mol Breed, 29: 489–50

Hare MP, Palumbi SR (2003) High intron sequence conservation across three mammalian orders suggests functional constraints. Mol Biol Evol, 20:969–978.

References

He G, Luo X, Tian F, Li K, Zhu Z, Su W, Qian X, Fu Y, Wang X, Sun C, Yang J (2006) Haplotype variation in structure and expression of a gene cluster associated with a quantitative trait locus for improved yield in rice. Genome Res, 16: 618-626.

He Y, Gan S (2001) Identical promoter elements are involved in regulation of the *OPR1* gene by senescence and jasmonic acid in *Arabidopsis*. Plant Mol Biol, 47: 595–605.

Heid C A, Stevens J, Livak K J *et al., (*1996) Real time quantitative PCR. Genome Res, 6: 986-994.

Hemmerlin A, Harwood JL, Bach TJ (2012) A raison d'e^tre for two distinct pathways in the early steps of plant isoprenoid biosynthesis? Prog Lipid Res, 51:95–148.

Herdendorf TJ, Miziorko HM (2006) Phosphomevalonate kinase: functional investigation of the recombinant human enzyme. Biochem, 45:3235–3242.

Hirakawa H, Shirasawa K, Ohyama A, Fukuoka H, Aoki K, Rothan C, Sato S, Isobe S, Tabata S (2013) Genome-wide SNP genotyping to infer the effects on gene functions in tomato. DNA Res, 20:221-233

Hospital F, Charcosset A (1997) Marker-assisted introgression of quantitative trait loci. Genetics, 147: 1469-1485.

Huang X, Lu G, Zhao Q, Liu X, Han B (2008) Genome-wide analysis of transposon insertion polymorphisms reveals intraspecific variation in cultivated rice. Plant Physiol, 148: 25-40.

Huq M A, Akter Shahina, Ill Sup Nou , Kim HT, Jung YJ, Kang K K. (2016) Identification of functional SNPs in genes and their effects on plant phenotypes. J Plant Biotechnol, 43:1-11.

Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. Genetics, 180:329-340.

References

Ishiguro S, Nishimori Y, Yamada M, Saito H , Suzuki T , Nakagawa T , Miyake H, Okada K, Nakamura K (2010) The Arabidopsis FLAKY POLLEN1 gene encodes a 3-Hydroxy-3-Methylglutaryl-Coenzyme A Synthase required for development of tapetum-specific organelles and fertility of pollen grains. Plant Cell Physiol, 51: 896–911.

Jain M, Khurana P, Tyagi AK, Khurana JP (2008) Genome-wide analysis of intronless genes in rice and *Arabidopsis*. Funct Integr Genomics, 8:69-78.

Jain SM, Priyadarshan PM PM eds (2009) Breeding plantation tree crops: tropical species. Springer, New York, USA.

James A, Avart C, Worland B, Snape JW, Vain P (2002) The relationship between homozygous and hemizygous transgene expression levels over generations in populations of transgenic rice plants. Theor Applied Genet, 104:553-561.

Jayashree R, Rekha K, Venkatachalam P, Uratsu SL, Dandekar AM, Jayasree PK, Kala RG, Priya P, Sushamakumari S, Sobha S, Asokan MP, Sethuraj MR, Thulaseedharan A (2003) Genetic transformation and regeneration of rubber tree (*Hevea brasiliensis* Muell. Arg.) transgenic plants with a constitutive version of an anti-oxidative stress superoxide dismutase gene. Plant Cell Rep 22: 201-209

Jones KP, Allen PW (1992) Historical development of the world rubber industry.in ed: Sethuraj MR and Mathew NM. Natural rubber: biology, cultivation and technology 1-26.

Joobeur T, Periam N, de Vicente MC, King GJ, Arús P (2000) Development of a second generation linkage map for almond using RAPD and SSR markers. Genome, 43:649-655.

Jung JH, Kim S, Ryu S, Kim MS, Baek YS, Kim SJ, Choi JK, Park JK, Min GS (2012) Development of single-nucleotide polymorphism- based phylum-specific pcr amplification technique: application to the community analysis using ciliates as a reference organism. Molecules and Cells, 34: 383–391.

Kasahara H, Takei K, Ueda N, Hishiyama S, Yamaya T, Kamiya Y, Yamaguchi S, Sakakibara H (2004) Distinct isoprenoid origins of cis- and trans-zeatin biosyntheses in *Arabidopsis*. *J Biol Chem,* 279: 14049-14054.

References

Kashkush K, Feldman M, Levy AA (2002) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat Genet, 33:102- 106.

Kauppi L, Jasin M, Keeney S (2007) Meiotic crossover hotspots contained in haplotype block boundaries of the mouse genome. Proc Natl Accad Sci USA, 104:13396–13401.

Kejnovsky E, Hawkins JS, Feschotte C (2012) Plant transposable elements: biology and evolution. In: Wendel JF *et al.,* (eds.) Plant Genome Diversity, 1:17-34, Springer-Verlag, Wien.

Kelly RD, Rodda AE, Dickinson A, Mahmud A, Nefzger CM, Lee W, Forsythe JS, Polo JM, Trounce IA, McKenzie M, Nisbet DR, St John JC (2013) Mitochondrial DNA haplotypes define gene expression patterns in pluripotent and differentiating embryonic stem cells. Stem Cells, 31:703-16.

Kera K, Takahashi S, Sutoh T, Koyama T, Nakayama T (2012) Identification and characterization of a cis,trans–mixed heptaprenyl diphosphate synthase from *Arabidopsis thaliana*. FEBS J, 279: 3813–3827.

Kersten B, Voss MM, Fladung M (2015) Development of mitochondrial SNP markers in different *Populus* species. Trees, 29: 575-582.

Kharabian-Masouleh A, Waters DLE, Reinke RF, Ward R, Henry RJ (2012) SNP in starch biosynthesis genes associated with nutritional and functional properties of rice. Sci Rep, 2:557.

Kim EY, Seo YS, Lee H, Kim WT (2010) Constitutive expression of *CaSRP1*, a hot pepper small rubber particle protein homolog, resulted in fast growth and improved drought tolerance in transgenic *Arabidopsis* plants. Planta, 232: 71–83.

Kim IJ, Ryu SB, Kwak YS, Kang H (2004) A novel cDNA from *Parthenium argentatum* Gray enhances the rubber biosynthetic activity *in vitro*. J Exp Bot, 55: 377–385.

Kloeckner-Gruissem B and Freeling M (1995) Transposon-induced promoter scrambling: A mechanism for the evolution of new alleles. Proc Natl Acad Sci USA, 92: 1836-1840.

Kloer DP, Welsch R, Beyer P, Schulz GE (2006) Structure and reaction geometry of geranylgeranyl diphosphate synthase from *Sinapis alba*. Biochem, 45:15197–15204.

Ko JH, Chow KS, Han KH (2003) Transcriptome analysis reveals novel features of the molecular events occurring in the laticifers of *Hevea brasiliensis* (para rubber tree). Plant Mol Biol, 53: 479-492.

Koehn RK, Eanes WF (1976) An analysis of allelic diversity in natural populations of *Drosophila*: the correlation of rare alleles with heterozygosity. In: Karlin S., Nevo E. (eds), Population genetics and ecology. Acad Press, New York, 377–390.

Koeyer D, Douglass K, Murphy A, Whitney S, Nolan L, Song Y, Jong W (2010) Application of high-resolution DNA melting for genotyping and variant scanning of diploid and autotetraploid potato. Mol Breed, 25:67–90.

Kota R, Rudds S, Facius A, Kolesov G, Thicl T, Zhang H, Stein N, Mayer K, Graner A (2003) Snipping polymorphisms from large EST collection in barley. Mol Genet Genomics, 270: 24-33.

Külheim C, Yeoh HS, Maintz J, Foley WJ, Moran GF (2009) Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. BMC Genomics, doi.org/10.1186/1471-2164-10-452

Kumari S, Priya P, Misra G, Yadav G (2013) Structural and biochemical perspectives in plant isoprenoid biosynthesis. Phytochem Rev, 12:255–291.

Kush A, Goyvaerts E, Chye ML, Chua NH (1990) Laticifer-specific gene expression in *Hevea brasiliensis* (rubber tree). Pro Nat Acad Sci , USA , 87:1787–1790

Kwok PY (2001) methods for genotyping single nucleotide polymorphisms. Annu Rev Genomics Hum Genet, 2:235–258.

Lambrides CJ, Godwin ID, Lawn RJ, Imrie BC (2004) Segregation distortion for seed testa color in Mungbean (*Vigna radiata* L. Wilcek). J Hered, 95:532-535.

References

Lange BM, Rujan T, Martin W, Croteau R (2000) Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. Pro Nat Acad Sci , USA, 97: 13172–13177.

Latorra D, Campbell K, Wolter A, Hurley JM (2003) Enhanced allele-specific PCR discrimination in SNP genotyping using 3' locked nucleic acid (LNA) primers. Hum Mutat, 22:79-85.

Le Guen V, Gay C, Xiong TC, Souza LM, Rodier-Goud M, Seguin M (2011) Development and characterization of 296 new polymorphic microsatellite markers for rubber tree (*Hevea brasiliensis*). Plant Breed, 130: 294–296.

Lefebvre V, Kuntz M, Camara B, Palloix A (1998) The capsanthin-capsorubin synthase gene: a candidate gene for the *y* locus controlling the red fruit colour in pepper. Plant Mol Biol, 36: 785–789.

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A Schulze-Lefert P (1998) Rapid reorganization of resistance gene homologues in cereal genomes. Pro Nat Acad Sci , USA , 95: 370–375.

Leitch AR, Lim KY , Leitch IJ , O'Neill M, Chye M, Low F (1998) Molecular cytogenetic studies in rubber, *Hevea brasiliensis* Muell. Arg. (*Euphorbiaceae*). Genome, 41: 464-467.

Lekawipat NK, Teerawatannasuk M, Rodier-Goud M, Seguin A, Vanavichit T, Toojinda, Tragoonrung S (2003) Genetic diversity analysis of wild germplasm and cultivated clones of *Hevea brasiliensis* Muell. Arg. by using microsatellite markers. J Rubb. Res, 6:36–47.

Leonardo RS *et al.*, *De novo* transcriptome analysis of *Hevea brasiliensis* tissues by RNA-seq and screening for molecular markers. BMC Genomics, doi: 10.1186/1471-2164-15-236

Leonforte A, Sudheesh S, Cogan NOI, Salisbury *et al.,* (2013) SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea(*Pisum sativum L.*). BMC Plant Biol, 3: 161.

References

Lespinasse D, Rodier-Goud M, Grivet L, Leconte A ,Legnate H, Seguin M (2000) A saturated genetic linkage map of rubber tree (*Hevea spp.)* based on RFLP, AFLP, microsatellite and isozyme markers, Theor Applied Genet, 100:127–138.

Lewinsohn TM (1991) The geographical distribution of plant latex. Chemo Ecol, 2: 64–68.

Li D, Zeng , Li , Zhao M, Chao J, Li Y, Wang K, Zhu L, Tian W-M, Liang C (2016) Gene expression analysis and SNP/InDel discovery to investigate yield heterosis of two rubber tree F1 hybrids. Scientific Rep, doi: 10.1038/srep24984

Li W, Liu W, Wei H, He Q, Chen J, Zhang B *et al.,* (2014) Species-specific expansion and molecular evolution of the 3-hydroxy-3-methylglutaryl coenzyme a reductase (*HMGR*) gene family in plants. PLoS ONE, doi:10.1371/journal. pone.0094172

Li Z, Zheng X, Ge S (2011) Genetic diversity and domestication history of African rice (*Oryza glaberrima*) as inferred from multiple gene sequences. Theor Applied Genet, 123:21–31

Liang MH, Liang YJ, Jin HH, Jiang JG (2015). Characterization and functional identification of a gene encoding geranylgeranyl diphosphate synthase from *Dunaliella bardawil.* J Agric Food Chem, 63:7805-7812.

Liang PH, Ko TP, Wang AHJ (2002) Structure, mechanism and function of prenyltransferases. Eur J Biochem, 269: 3339-3354.

Librado P, Rozas J (2009) DnaSP v5: software for comprehensive analysis of DNA polymorphism data. Bioinfo, 11:1451-1452.

Licy J (1997) Variability correlation and heterosis for yield and yield components in certain hybrid clones of para rubber tree *Hevea brasiliensis* Muell. Arg. Ph.D thesis, University of Kerala

Liew M, Pryor R, Palais R, Meadows C, Erali M, Lyon E, Wittwer C (2004) Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. Clin Chem, 50:1156–1164.

References

Liu JP, Zhuang YF, Guo XL, Li YJ (2016) Molecular mechanism of ethylene stimulation of latex yield in rubber tree (*Hevea brasiliensis*) revealed by de novo sequencing and transcriptome analysis. BMC Genomics, doi.org/10.1186/s12864-016-2587-4

Liu A and Burke J M (2006) Patterns of nucleotide diversity in wild and cultivated Sunflower. Genetics, 173: 321–330.

Livingstone K, Anderson S (2009) Patterns of variation in the evolution of carotenoid biosynthetic pathway enzymes of higher plants. J Hered, 100: 754–761.

Lorenz AJ, Hamblin MT, Jannink JL (2010) Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. PLoS ONE, doi:10.1371/journal.pone.0014079

Lorieux M, Perrier X, Goffinet B, Lanaud C, deLeon Go (1995) Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F_2 populations. Theor Applied Genet, 90: 81-89.

Low FC, Atan S, Jaafar H, Tan H (1996) Recent advances in the development of molecular markers for *Hevea* studies. J Nat Rubb Res, 11:32–44.

Low FC, Bonner J (1985) Characterization of the nuclear genome of *Hevea brasiliensis* Int Rubb Conf, Kuala Lumpur, Malaysia, 1–9.

Lu YQ, Rausher MD (2003) Evolutionary rate variation in anthocyanin pathway genes. Mol Biol Evol, 20: 1844–1853.

Luan DD, Korman MH, Jakubczak JL, Eickbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell, 72:595–605.

Luo H, Boutry M (1995) Phylogenetic relationships within *Hevea brasiliensis* as deduced from a polymorphic mitochondrial DNA region. Theor Applied Genet, 91:876–884.

Lynen FH, Eggerer U, Kessel HI (1958) Farnesyl-pyrophosphat und 3-Methyl-delta-3-butenyl-1-pyrophosphat, die biologischen Vorstufen des Squalens. Zur Biosynthese der Terpene, III, Angewandte Chemie, 70:738-742.

References

Majumder A, Datta S, Choudhary B K, Majumdar K (2014) Do extensive rubber plantation influences local environment? A case study from Tripura, northeast India. Curr World Environ, doi.org/10.12944/CWE.9.3.25

Mandrou E, Hein RPG, Villar E, Vigneron P, Plomion C, Gion JM (2012) A candidate gene for lignin composition in *Eucalyptus*: cinnamoyl-CoA reductase (CCR). Tree Genet & Genomes, 8:353-364.

Mantello CC, Suzuki FI, Souza LM, Gonçalves PS, Souza AP (2012) Microsatellite marker development for the rubber tree (*Hevea brasiliensis*): characterization and cross-amplification in wild *Hevea* species. BMC Res Notes, doi: 10.1186/1756-0500-5-329

Mantello CC, Cardoso-Silva CB, da Silva CC, de Souza LM, Scaloppi Junior EJ, de Souza Gonçalves P, *et al.,* (2014) *De Novo* assembly and transcriptome analysis of the rubber tree (*Hevea brasiliensis*) and SNP markers development for rubber biosynthesis pathways. PLoS ONE, doi:10.1371/journal.pone.0102665

Marattukalam JG, Panikkar AON, Saraswathyamma CK (1992) Long term performance of a few clones of *Hevea brasiliensis* in large scale trial. J Plantation Crops 20: 170-174.

Mariette *et al.,* (2015) Genome-wide association links candidate genes to resistance to Plum Pox Virus in apricot (*Prunus armeniaca*). New Phytol,: 209:773-84.

Markert CL, Moller F (1959) Multiple forms of enzymes: Tissue,ontogenetic and species specific patterns. Proc Natl Acad Sci USA, 45: 753-763.

Martin D, Tholl D, Gershenzon J, Bohlmann J (2002) Methyl Jasmonate induces traumatic resin ducts, terpenoid resin biosynthesis, and terpenoid accumulation in developing xylem of Norway Spruce stems. Plant Physiol, 129: 1003–1018.

Mccouch SR (2001) Genomics and synteny. Plant Physiol, 125: 152–155.

McDaniel SF, Willis JH, Shaw AJ (2007) A Linkage map reveals a complex basis for segregation distortion in an interpopulation cross in the moss *Ceratodon purpureus*. Genetics, 176:2489–2500.

References

McDowell JM, Dhandaydham M, Long TA, Aarts MG, Goff S, Holub EB, Dangl JL (1998) Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of *Arabidopsis*. Plant Cell, 10:1861-74.

Melchinger AE (1990) Use of molecular markers in breeding for oligogenic disease resistance. Plant Breed, 104:1-19.

Mendel G (1866) Experiments in plant hybridization in Proc Baunn Natural history society. Harward University Press, Cambridge.

Michael PH, Stumpf (2004) Haplotype diversity and SNP frequency dependence in the description of genetic variation. Europ J Hum Genet, 12: 469–477.

Morello L and Breviario D (2008) Plant spliceosomal introns: not only cut and paste. Curr Genomics, 4: 227–238.

Morgante M, Hanafey H, Powell W, (2002) Micro satellites are preferentially associated with non-repetiture DNA in plant genome. Nature Gent, 30: 194-200.

Morita S, Tsukamoto S, Sakamoto A, Makino H, Nakauji E, Kaminaka H, Masumura T, Ogihara Y, Satoh S, Tanaka K (2012) Differences in intron-mediated enhancement of gene expression by the first intron of cytosolic superoxide dismutase gene from rice in monocot and dicot plants. Plant Biotechnol, 29: 115–119.

Nackley A G, Shabalina SA, Tchivileva IE, Satterfield K, Korchynskyi O, Makarov SS, Maixner W, Diatchenko, L (2006) Human catechol-o-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. Science, 314:1930-1933

Nagata N, Suzuki M, Yoshida S, Muranaka T (2002) Mevalonic acid partially restores chloroplast and etioplast development in *Arabidopsis* lacking the non-mevalonate pathway. Planta, 216: 345-350.

Nagegowda DA, Ramalingam S, Hemmerlin A, Bach TJ, Chye M (2005) *Brassica juncea* HMG-CoA synthase: localization of mRNA and protein. Planta, 221: 44-856

References

Nair VKB, George PJ (1969) Indian clones-RRII 100 series. Rubber Board Bulletin, 10: 115-140.

Nazeer MA, Marcose VC, George PJ, Panikkar AON (1987) Performance of a few *Hevea* clones from RRII 100 series in a large sale trial. J Plantation Crops, 14:99-104.

Norma C, Manrique-Carpintero, Tokuhisa JG, Ginzberg I, Holliday JA, Richard EV (2013) Sequence diversity in coding regions of candidate genes in the glycoalkaloid biosynthetic pathway of wild potato species, G3 (Bethesda), 4:1467-1479.

Novaes E, Drost D, Farmerie W, Pappas G, Grattapaglia D, Sederoff R, Kirst M (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMC Genomics, 9:312.

Novalina N, Sagala AD (2013) Construction of *Hevea brasiliensis* genetic linkage map and identification of quantitative trait loci using RAPD markers. Internatio journa advance science engineering and information technology, 3: 71-75.

Oh SK *et al.*, (1999) Isolation, characterization, and functional analysis of a novel cDNA clone encoding a small rubber particle protein from *Hevea brasiliensis*. J Biol Chem, 274:17132–17138.

Oh SK, Han KH, Ryu SB, Kang H (2000) Molecular cloning, expression, and functional analysis of a cis–prenyltransferase from *Arabidopsis thaliana*: implications in rubber biosynthesis. J Biol Chem, 275: 18482–18488.

Ohnuma S, Suzuki M., Nishino T (1994) Archaebacterial ether-linked lipid biosynthetic gene. Expression cloning, sequencing, and characterization of geranylgeranyl-diphosphate synthase. J Biol Chem, 269:14792–14797.

Okada K, Kasahara H, Yamaguchi S, Kawaide H, Kamiya Y, Nojiri H, Yamane H (2008) Genetic evidence for the role of isopentenyl diphosphate isomerases in the mevalonate pathway and plant development in *Arabidopsis*. Plant Cell Physiol, 49:604–616.

References

Okada K, Saito T, Nakagawa T, Kawamukai M, Kamiya Y (2000) Five geranylgeranyl diphosphate synthases expressed in different organs are localized into three subcellular compartments in *Arabidopsis*. Plant Physiol, 122: 1045–1056

Oliveira, RP, Aguilar-Vildoso CI, Cristofani M, Machado MA (2004) Skewed RAPD markers in linkage maps of Citrus. Genet Mol Biol, doi.org/10.1590/S1415-4757200400030002

Olsen M, Hood L, Cantor C, Botstein D (1989) A common language for physical mapping of the human genome. Science, 245:1434-1435.

Ong SH (1976) Chromosome morphology at the pachytene stage in *Hevea brasiliensis*: A preliminary report. Proceedings of International Rubber Conference, KualaLumbur, Malaysia, 1-12.

Orita M, Suzula Y, Sekiya T, Hayarshi K, Rapid and Semitoori (1989) Detection of point mutations and DNA polymorphisms using the polymerace chain reaction. Genomics, 5: 874-879.

Paiva JR, Kageyama PY, Vencorsky R, Contel PB (1994) Genetics of rubber tree *Hevea brasiliensis* (Willd. ex Adr. de Juss.) Muell. Arg. Genetic variation in natural populations. Silvae Gen, 43:307–312.

Palle SR, Seeve CM, Eckert AJ, Wegrzyn JL, Neale DB, Loopstra C A (2013). Association of loblolly pine xylem development gene expression with single-nucleotide polymorphisms. Tree Physiol, 33: 763-774.

Pan J, Zhang S, Borchers CH (2016) Protein species-specific characterization of conformational change induced by multisite phosphorylation. J Proteomics, 134:138-143.

Panikkar AON, Nair VKB, Markose VC (1980) Breeding and tree improvement. In: Handbook of Natural Rubber Production in India. (eds) Radhakrishna Pillay PN). Rubber Research Institute of India, Kottayam, India, 35-43.

Parra G, Bradnam K, Rose AB, Korf I (2011) Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. Nucl Acids Res, 13:5328-5337.

References

Pawsoi N, Phumichai T, Teerawatanasuk K, Wongkaew A and Phumichai1 C (2013) Microsatellite paternity analysis used for evaluation of outcrossing rate among five *Hevea* rubber clones in a systematic seed orchard. Kasetsart J (Nat Sci.), 47: 407 - 415

Perrier X, Flori A, Bonnot F (2003) Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC ed. Genetic diversity of cultivated tropical plants. Enfield Science Publishers, Montpellier, 43 – 76.

Perseguini JMKC, Romão LR de Castro, Briñez B, Erivaldo JS, Gonçalves Pde Souza, Benchimol LL (2012) Genetic diversity of cultivated accessions and wild species of rubber tree using EST SSR markers. Pesq agropec bras Brasília, 47:1087-1094.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem, 25:1605-1612.

Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Mol Breed, 7: 275–291.

Phillips MA, Savage TJ, Croteau R (1999) Monoterpene synthases of loblolly pine (*Pinus taeda*) produce pinene isomers and enantiomers. Arch Biochem Biophys, 372:197-204.

Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs from EST data bases. Genome Res, 9: 167-174.

Plomion C, Frigerio J-M, Ridolfi M, Pot D, Pionneau C, Bodenes C, Kremer A, Hurme (1999) Polymorphism discovery In Association mapping in plants. ed. Oraguzie *et al.,* Springer science media, New York, 41-52.

Pojer F, Ferrer JL, Richard SB, Nagegowda DA, Chye ML, Bach TJ, Noel JP (2006) Structural basis for the design of potent and species-specific inhibitors of 3-hydroxy-3-methylglutaryl CoA synthases. Proc Natl Acad Sci USA, 103:11491–11496.

References

Ponciano G, McMahan CM, Xie W, Lazo GR, Coffelt TA, Collins-Silva J, NuralTaban A, Gollery M, Shintani DK, Whalen MC (2012) Transcriptome and gene expression analysis in cold-acclimated guayule (*Parthenium argentatum*) rubber-producing tissue. Phytochemistry, 79: 57–66.

Pootakham W, Chanprasert J, Jomchai N, Sangsrakru D, Yoocha T *et al.,* (2011) Single nucleotide polymorphism marker development in the rubber tree, *Hevea brasiliensis* (*Euphorbiaceae*). Am J Bot 98: 337–338.

Pootakham, Wirulda *et al.*, (2015) Construction of a high-density integrated genetic linkage map of rubber tree (*Hevea Brasiliensis*) using genotyping-by-sequencing (GBS) Front Plant Sci, doi: 10.3389/fpls.2015.00367.

Priya P, Venkatachalam P, Thulaseedharan A (2007) Differential expression pattern of rubber elongation factor (REF) mRNA transcripts from high and low yielding clones of rubber tree (*Hevea brasiliensis* Muell. Arg.). Plant Cell Rep, 26: 1833–1838.

Priya,P (2005) Molecular cloning, characterization and expression of rubber elongation factor gene from *Hevea brasiliensis*. Muell. Arg. Ph.D Thesis, Mahatma Gandhi University.

Qin L, Hao C, Hou J, Wang Y, Li T, Wang L, Ma Z, Zhang X (2014) Homologous haplotypes, expression, genetic effects and geographic distribution of the wheat yield gene TaGW2. BMC Plant Biol, doi: 10.1186/1471-2229-14-107.

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

Rahman AYA *et al.,* (2013) Draft genome sequence of the rubber tree *Hevea brasiliensis*. BMC Genomics, doi: 10.1186/1471-2164-14-75

Rajesh PN, O'Bleness M, Roe BA, Muehlbauer FJ (2008) Analysis of genome organization, composition and microsynteny using 500 kb BAC sequences in chickpea. Theor Applied Genet, 117: 449-458.

Ratnaparkhe MB, Tekeoglu M, Muehlbauer FJ (1998) Intersimple-sequence-repeat (ISSR) polymorphisms are useful for finding markers associated with disease resistance gene clusters. Theor Applied Genet, 97: 515–519.

References

Rausher M, Lu Y, Meyer K (2008) Variation in constraint vs. positive selection as an explanation for evolutionary rate variation among anthocyanin genes. J Mol Evol, 67: 137–144.

Rausher MD, Miller RE, Tiffin P (1999) Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol Biol Evol, 16: 266–274.

Rekha K, Nazeem PA, Venkatachalam P, Jayashree R, Sobha S, Sushama Kumari S (2014) Development of osmotin transgenics in *Hevea brasiliensis* Muell. Arg. using explants of zygotic origin. J Tropical Agri 52 : 7-20.

Remmert M, Biegert A, Hauser A, Söding J (2011) HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods, 9: 173-175.

Rensink WA, Lee Y, Liu J, Iobst S, Ouyang S, Buell CR (2005) Comparative analyses of six *solanaceous* transcriptomes reveal a high degree of sequence conservation and species-specific transcripts. BMC Genomics, doi:10.1186/1471-2164-6-124.

Romanowski MJ, Bonanno JB, Burley SK (2002) Crystal structure of the Streptococcus pneumoniae phosphomevalonate kinase, a member of the GHMP kinase superfamily. Proteins, 47:568–571.

Rose, AB. (2002) Requirements for intron-mediated enhancement of gene expression in *Arabidopsis*. RNA, 8:1444-1453.

Rouster J, Robert L, Mundy, J, Cameron-Mills V (1997) Identification of a methyl jasmonate-responsive region in the promoter of a *lipoxygenase 1* gene expressed in barley grain. Plant J, 11: 513–523.

Roy CB, Nazeer M A, Saha T (2004) Identification of simple sequence repeats in rubber (*Hevea brasiliensis*). Curr Sci, 87:807–811.

Roy CB, Ravindran M and Saha T (2012) Efficient screening of AFLP primer combinations for evaluating genetic diversity among cultivated rubber (*Hevea brasiliensis*) clones. Natural Rubber Res, 25:21-30.

Sacchettini JC, Poulter CD (1997) Creating isoprenoid diversity. Science, 277: 1788–1789.

Saha T, Roy CB, Nazeer MA (2005) Microsatellite variability and its use in the characterization of cultivated clones of *Hevea brasiliensis*. Plant Breed, 124: 86–92.

Saleena A, Supriya R, Thulaseedharan A (2006) Isolation and characterisation of Hevein gene promoter from *Hevea brasiliensis* (clone: RRII 105) (1934 bp.) Unpublished,Accession number DQ859572.

Saleena A, Supriya R, Vineetha M, Thulaseedharan A (2010) Cloning and molecular characterisation of an intronless *cis*-prenyltransferase from *Hevea brasiliensis.* Placrosym XIX, 7-10 December 2010, RRII Kottayam, India.

Sando T *et al.* (2008a) Cloning and characterization of the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway genes of a natural-rubber producing plant, *Hevea brasiliensis*. Biosci Biotech Biochem, 72: 2903–2917.

Sando T, Hayashi T, Takeda T, Akiyama Y, Nakazawa Y, Fukusaki E, Kobayashi A (2009) Histochemical study of detailed laticifer structure and rubber biosynthesis-related protein localization in *Hevea brasiliensis* using spectral confocal laser scanning microscopy. Planta, 230: 215-225.

Sando T. *et al.* (2008b) Cloning and characterization of mevalonate pathway genes in a natural rubber producing plant, *Hevea brasiliensis*. Biosci Biotech Biochem 72:2049–2060

Saraswathyamma CK, Licy J, Marattukalam G (2000) Planting materials, In: George PJ, Jacob CK (eds.), Natural rubber: Agromanagement and crop processing, Rubber Research Institute of India, Kottayam, 59-74.

Saraswathyamma CK, Markose VC, Licy J, Panikkar AON (1984) Cytomorphological studies in an induced polyploidy of *Hevea brasiliensis* Muell.Arg. Cytologia, 49: 725-729.

Sauer S, Lechner D, Berlin K, Lehrach H, Escary JL, Fox N, Gut IG (2000) A novel procedure for efficient genotyping of single nucleotide polymorphisms. Nucleic Acids Res, 28:e13.

Schaal BA, Gaskin JF, Caicedo AL (2003) Phylogeography, haplotype trees, and invasive plant species. J Heredity, 94:197–204.

Schauer SE, Schlüter PM, Baskar R, Gheyselinck J, Bolaños A, Curtis MD, Grossniklaus U (2009) Intronic regulatory elements determine the divergent expression patterns of Agamous-like6 subfamily members in *Arabidopsis*. Plant J, 59: 987-1000.

Schmidt T, Lenders M *et al.,* (2010) Characterization of rubber particles and rubber chain elongation in *Taraxacum koksaghyz.*BMC Biochem, doi:10.1186/1471-2091-11-11.

Schreiber L, Nader-Nieto AC, Schönhals EM, Walkemeier B, Gebhardt C (2014) SNPs in genes functional in starch-sugar interconversion associate with natural variation of tuber starch and sugar content of potato. G3 (Bethesda), 4: 1797-1811.

Schulte AE, Heijden RV, Verpoorte R (1999) Purification and characterization of phosphomevalonate kinase from *Catharanthus roseus*. Phytochemistry, 52: 975–983.

Schultes RE (1977) Wild Hevea: An untapped source of germplasm. J Rubber Res Inst Sri Lanka. 54: 227-257.

Schwede T, Kopp J, Guex N, Peitsch MC (2003) Swiss-model: an automated protein homology-modeling server. Nucleic. Acids Res, 31:3381-3385.

Sela N, Kim E, Ast G (2010) The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. Genome Biol, doi: 10.1186/gb-2010-11-6-r59.

Seo SG, Kim JS, Yang YS, Jun BK, Kang SW, Lee GP, Kim W, Kim JB, Lee HU, Kim SH, (2010) Cloning and characterization of the new multiple stress responsible gene I (MuSI) from sweet potato. Genes Genom, 32: 544–552.

Semikhov VF, Aref'ev LP, Zolkin SY, Timoshchenko AS, Novozhilova OA, Kostrikin DS (2004) Relationships of the gymnosperms and angiosperms estimated on the basis of data obtained by biochemical methods. Biology Bulletin of the Russian Academy of Sciences, 31: 21-35.

References

Shearman JR, Sangsrakru D, Jomchai N, Ruang-Areerate P, Sonthirod C, Naktang C, Theerawattanasuk K, Tragoonrung S, Tangphatsornruang S (2015) SNP identification from RNA sequencing and linkage map construction of rubber tree for anchoring the draft genome. PLoS One, doi: 10.1371/journal.pone.0121961

Sicard A, Kappel C, Josephs EB, Lee YW, Marona C, Stinchcombe JR, Wright SI, Lenhard M (2015) Divergent sorting of a balanced ancestral polymorphism underlies the establishment of gene-flow barriers in *Capsella.* Nat Commun, doi: 10.1038/ncomms8960.

Simmonds NW (1989) Rubber breeding. In: Rubber (eds.) Webster CC and Baulkwill WJ). Longman Scientific and Technical, Essex, 85-124.

Sirinupong N, Suwanmanee P, Doolittle RF, Suvachitanont W (2005) Molecular cloning of a new cDNA and expression of 3-hydroxy-3-methylglutaryl-CoA synthase gene from *Hevea brasiliensis*. Planta, 221: 502–512.

Skilleter DN, Kekwick RGO (1971) The enzymes forming isopentenyl pyrophosphate from 5-phosphomevalonate (mevalonate-5-phosphate) in the latex of *Hevea brasiliensis*. Biochem J, 124: 407-415.

Smith GD, Chadwick BE, Willmore-Payne C, Bentz JS (2008) Detection of epidermal growth factor receptor gene mutations in cytology specimens from patients with non-small cell lung cancer utilising high-resolution melting amplicon analysis. J Clin Pathol, 61:487–493.

Sobha S, Sushamakumari S, Thanseem I, Jayasree PK, Rekha K, Jayashree R, Kala RG, Asokan MP, Sethuraj MR, Dandekar AM, Thulaseedharan A (2003) Genetic transformation of *Hevea brasiliensis* with the gene coding for superoxide dismutase with FMV 34S promoter. Current Science 85: 1767-1773.

Souza LM, Gazaffi R, Mantello CC, Silva CC, Garcia D, Le Guen V, Cardoso SE, Garcia AA, Souza AP (2013) QTL mapping of growth-related traits in a full-sib family of rubber tree (*Hevea brasiliensis*) evaluated in a sub-tropical climate. PLoS One, doi: 10.1371/journal.pone.0061238.

References

Souza LM *et al.,* (2011) Development of a genetic linkage map of rubber tree (*Hevea braziliensis*) based on microsatellite markers. BMC Proceedings, doi: 10.1186/1753-6561-5-S7-P39.

Sunderasan ER, Wickneswari MZ, Aziz A., Yeang HY (1994) Incidence of self- and cross-pollination in two *Hevea brasiliensis* clones. J Nat Rubb Res, 9:253–257.

Surmacz L, Swiezewska E (2011) Polyisoprenoids – secondary metabolites or physiologically important superlipids? Biochem Biophys Res Commun, 407: 627–632.

Sushamakumari, SS, Shoba K, Rekha K, Jayashree R, Asokan MP (2000) Influence of growth regulators and sucrose on somatic embryogenesis and plant regeneration from immature inflorescence of *Hevea brasiliensis*. Indian J Nat Rubb Res, 13:19–29.

Suwanmanee P, Sirinupong N, Suvachittanont W (2004) Regulation of the expression of 3-hydroxy-3-methylglutaryl-CoA synthase in *Hevea brasiliensis* (HBK) Mull. Plant Sci. 166:531–537.

Suwanmanee P, Suvachittanont W, Fincher GB (2002) Molecular cloning and sequencing of a cDNA encoding 3-hydroxy-3-methylglutaryl-CoA synthase from *Hevea brasiliensis* (HBK) Mull. Arg. Sci Asia, 28:29–36.

Syed NH, Kalyna M, Marquez Y, Barta A, Brown J W (2012) Alternative splicing in plants--coming of age. Trends Plant Sci, 17: 616–623.

Tang C, *et al* (2016) The rubber tree genome reveals new insights into rubber production and species adaptation. Nat Plants, doi: 10.1038/Nat PLANTS.2016.73.

Tang C, Xiao X, Li H, Fan Y, Yang J, Qi J, Li H (2013) Comparative analysis of latex transcriptome reveals putative molecular mechanisms underlying super productivity of *Hevea brasiliensis*. PLoSONE, doi.org/10.1371/journal.pone. 0075307

Tanksley SD (1984) Linkage relationships and chromosomal locations of enzyme-coding genes in pepper, *Capsicum annuum*. Chromosoma, 89: 352–360.

References

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature, 408: 796-815.

Thulaseedharan A *et al.,* (2008) Rubber tree. In: Compendium of transgenic crop plants: Transgenic plantation crops, ornamentals and turf grasses. V8 (eds. Chittaranjan Kole and Timothy CH) Wiley- Blackwell, USA, pp.117-140.

Toka HR, Genovese G, Mount DB, Pollak MR, Curhan GC (2013) Frequency of rare allelic variation in candidate genes among individuals with low and high urinary calcium excretion. Plos ONE, doi: 10.1371/journal.pone 0071885

Tompa M (2001) Identifying functional elements by comparative DNA sequence analysis. Genome Res, 11: 1143–1144.

Triwitayakorn K *et al.,* (2011). Transcriptome Sequencing of *Hevea brasiliensis* for development of microsatellite markers and construction of a genetic linkage map. DNA Res, 18: 471–482

Uthup TK, Rajamani A, Ravindran M, Saha T (2016) Molecular evolution and functional characterisation of haplotypes of an important rubber biosynthesis gene in *Hevea brasiliensis*. Plant Biol, doi:10.1111/plb.12433.

Uthup TK, Saha T, Ravindran M, Bini K (2013) Impact of an intragenic retrotransposon on the structural integrity and evolution of a major isoprenoid biosynthesis pathway gene in *Hevea brasiliensis*. Plant Physiol Biochem, 73: 176-188.

Varghese YA, Knaak C, Sethuraj MR, Ecke W (1997) Evaluation of random amplified polymorphic DNA (RAPD) markers in *Hevea brasilinesis*. Plant Breed, 116:47–52.

Venkatachalam P, Sailasree R, Priya P, Saraswathyamma CK, Thulaseedharan A (2001) Identification of a DNA marker associated with dwarf trait in *Hevea brasiliensis* Muell. Arg. through random amplified polymorphic DNA analysis. In: J. Sainte-Beuve (ed.), Annual IRRDB Meeting. CIRAD, Montpellier, France.

Venkatachalam P, Thomas S, Priya P, Thanseem I, Gireesh T, Saraswathyamma CK, Thulaseedharan A (2002) Identification of DNA polymorphism with the cultivated clones of rubber tree (*Hevea brasiliensis* Muell. Arg.). Indian J Nat Rubber Res, 15:72–181.

References

Venkatachalama P, Geetha N, Thulaseedharan A, Shivendra VS (2016) Molecular cloning and characterization of an intronless farnesyl diphosphate synthase (FDP) gene from Indian rubber clone (*Hevea brasiliensis* Muell. Arg. RRII105): A gene involved in isoprenoid biosynthesis. Gene Rep, 4: 153–161.

Vidal RO, Nascimento LC, Mondego JMC, Pereira GAG, Carazzolle MF (2012) Identification of SNPs in RNA-seq data of two cultivars of *Glycine max* (soybean) differing in drought resistance. Genet Mol Biol, 35:331-334

Vollack KU, Dittrich B, Ferrer A, Boronat A, Bach TJ (1994) Two radish genes for 3-hydroxy-3-methylglutaryl-CoA reductase isozymes complement mevalonate auxotrophy in a yeast mutant and yield membrane-bound active enzyme. J Plant Physiol, 143:479–487.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Friters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23: 4407-4414.

Wall JD, Pritchard JK (2003) Assessing the performance of haplotype block models of linkage disequilibrium. Am J Hum Genet, 73: 502-515.

Wall JD, Pritchard JK (2003) Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet, 4: 587–597.

Wang CH, Lia W, Tiana YK, Houa D L, Bai MD (2016) Development of molecular markers for genetic and physical mapping of the PcDw locus in pear (*Pyrus communis L.*) The J Horti Sci and Biotec, 91:299-307

Wang X, Hughes A L, Tsukamoto T, Ando T, Kao T (2001) Evidence that intragenic recombination contributes to allelic diversity of the S-RNase gene at the self-incompatibility (*S*) Locus in *Petunia inflata*. Plant Physiol, 125:1012-1022.

Watt WB (1972) Intragenic recombination as a source of population genetic variability. The Am Naturalist, 106: 737-753.

Weber D, Helentjaris T (1989) Mapping RFLP loci in maize using B-A translocations. Genetics, 121: 583–590.

Weil CF, Wessler SR (1990) The effects of plant transposable element insertion on transcription initiation and RNA processing. Annu Rev Plant Physiol Plant Mol Bio, 41: 527-552.

Welsh J, Petersen C, McClelland M (1990) Polymorphism generated by arbitarily primed PCR. Nucleic Acids. Res, 19: 6823-6831.

Weng J *et al.*, (2013) A non-synonymous SNP within the isopentenyl transferase 2 locus is associated with kernel weight in Chinese maize inbreds (*Zea mays* L.) BMC Plant Biol, doi: 10.1186/1471-2229-13-98.

Whitelaw E , Martin DIK (2001) Retrotransposon as epigenetic mediators of phenotypic variation in mammals. Nat Genet, 27: 361–365.

Wilding EI, Brown JR, Bryant AP, *et al.* (2000) Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci. Journal of Bacteriol, 182:4319-4327.

Wu SB, Tavassolian I, Rabiei G, Hunt P, Wirthensohn M, Gibson J, Ford C, Sedgley M (2009) Mapping SNP-anchored genes using high-resolution melting analysis in almond. Mol Genet Genomics, 282:273–281.

Xia Y, Li R, Ning Z, Bai G, Siddique KH, Yan G, Baum M, Varshney RK, Guo P (2013) Single nucleotide polymorphisms in HSP17.8 and their association with agronomic traits in barley. PLoS One, doi.org/10.1371/journal.pone.0056816

Xie W, Mcmahan CM, Whalen MC, Distefano M, Degraw A, Cornish K, Shintani D (2008) Initiation of rubber biosynthesis: in vitro comparisons of benzophenone-modified diphosphate analogue structure in three natural rubber-producing species. Phytochemistry, 69: 2539-2545.

Xu H, Jiang B, Cao Y, *et al.* (2015) Detection of epistatic and gene-environment interactions underlying three quality traits in rice using high-throughput genome-wide data, BioMed Res Intl, doi:10.1155/2015/135782

Xu Q, Zhu C, Fan Y, Song Z, Xing S, Liu W, Yan J, Sang T (2016). Population transcriptomics uncovers the regulation of gene expression variation in adaptation to changing environment. Scientific Rep, doi: 10.1038/srep25536

Yan H, Zhang W, Lin Y, Dong Q, Peng X, Jiang H, Zhu S, Cheng B (2014) Different evolutionary patterns among intronless genes in maize genome Biochem Biophys Res Commun, 449:146-150.

Yeang H, Yip E, Hamzah S (1995) Characterisation of Zone 1 and Zone 2 rubber particles in *Hevea brasiliensis* latex. J Nat Rubb Res, 10:108–123.

Yenerall P, Zhou L (2012) Identifying the mechanisms of intron gain: progress and trends. Biol Direct, 7: 29.

Zhang J, Singh A, Mueller DS, Singh AK (2015) Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. Plant J, 84: 1124-1136.

Zhang L, Li H, Wang J (2015) Linkage analysis and map construction in genetic populations of clonal F1 and double cross. G3, 5: 427–439

Zhao H, Pfeiffer R, Gail MH (2003) Haplotype analysis in population genetics and association studies. Pharmacogenomics, 4:171–178.

Zhou L, Myers AN, Vandersteen JG, Wang L, Wittwer CT (2004) Closed-tube genotyping with unlabeled oligonucleotide probes and a saturating DNA dye. Clin Chem, 50:1328–1335.

Zhu YL, Hyatt S, Quigley C, Song QJ, *et al.* (2001). Single nucleotide polymorphisms (SNPs) in soybean genes, cDNAs, and random genomic sequence. Plant and Animal Genome IX Conference, San Diego, 13-17.

Zietkiewicz E, Rafalski A , Labuda D (1994) Genomic fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics, 20: 176-183.