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

the successive condensation of IDP in the *trans* or *cis* configuration to produce t*rans*-polyisoprene and rubber [10,11]. Due to its important role in the isoprenoid biosynthesis process, proper understanding of the regulation and expression of the *FDPS* gene is imperative for studying the qualitative and quantitative characters of its downstream products including rubber in *Hevea*.

Until recently, the regulatory aspects of non-coding DNA regions like introns and transposons were ignored because of the notion that promoters are the sole elements responsible for gene regulation. As gene regulation is mainly facilitated by the binding of transcription factors to the *cis*-regulatory motifs within the promoter region, studies pertaining to promoter sequence variations, mutations and their impact on gene expression were given preference over sequence variations identified in other non-coding regions. Nevertheless, later studies proved that intronic regions also play an important role in gene regulation [12,13]. Their positive effects on gene expression in organisms like nematodes, insects, and mammals are well documented [14–17]. The role of regulatory motifs residing in intronic regions on transcriptional gene regulation were also reported in plants [18,19]. For example, gene expression studies in *Arabidopsis* using gene constructs have shown that inclusion of introns in a construct lead to increased accumulation of mRNA and protein relative to non-intron constructs [20,21]. Furthermore, previous studies also indicate that introns act post-transcriptionally to increase mRNA accumulation, presumably by facilitating mRNA maturation or by enhancing the stability of nascent transcripts [22]. Usually the large-sized first introns were more often found to be responsible for such effects than the other introns due to their tendency to harbour regulatory elements [23,24]. Experimental evidences proving the essentiality of first introns for strong and constitutive gene expression further ascertain this argument [25]. Therefore, any change in the sequence structure of regulatory motifs, either in the first intron or in the promoter region of key genes may result in modified gene regulation. In the case of *H. brasiliensis*, the probability of finding similar regulatory mechanism in the isoprenoid biosynthesis pathway genes is high due to the presence of large introns with abundant sequence polymorphisms despite the conserved nature of their coding sequence [26], Uthup et al., unpublished].

Intronic sequence variations are mainly due to intragenic recombination and transposon activity [27]. But intragenic recombinations rarely result in the confinement of significantly large number of SNPs and indels in to a small region as it requires several generations of meiotic cycles for this to happen. Therefore such highly variable regions are assumed to be the outcome of transposable element (TE) activity by their repeated insertion, deletion and copy functions rather than genetic recombination. Additionally, these changes are often reported to generate turbulences in the regulatory and splice elements harboured within major introns resulting in exonisation, generation of splice variants as well as development of new alleles [28–30]. Another important parameter which decides the gene regulatory role of TEs is their proximity to coding sequences and promoter regions [31]. Kapazoglou et al. (2012) [32] reports that, transposon induced mutations and successive gene regulation occur mostly towards the 5′ end of genes, preferably in the first intron which is proximal to the promoter region. The presence of intragenic TEs in the first intron, affecting the expression of the flowering locus C (*FLC*) gene have been reported by Liu et al. (2004) [33] in *Arabidopsis* and for the *knotted1* gene in maize by Greeny et al. (1994) [34].

Interestingly the recent genome-wide transcriptome sequencing (RNA-Seq) studies and whole genome sequencing using next-generation sequencing in *Hevea* reports that, repetitive sequences represent close to 75% of its genome of which, retrotransposon constitute 50% [35]. Alternatively, their existence in the genic region was reported by Saha et al. (2006) through their targeted studies on disease resistance genes in rubber [36]. Thus it seems quite likely that *Hevea* genes are structurally and functionally influenced by retrotransposon, which has to be established by extensive sequence analysis.

In the above context, the structural variants of isoprenoid biosynthesis genes prevailing in the *Hevea* population have to be thoroughly examined for understanding their role in gene regulation and evolution. Here we describe the genomic organisation of the *FDPS*, a major gene in the MVA pathway by discovering its structural variants existing in the *Hevea* gene pool. The impact of a highly polymorphic retroelement on the *FDPS* first intronic region was analysed in detail. The characterisation of this element was performed by analysing its sequence and distribution in wild germplasm accessions, popular clones and other *Hevea* species. The results reported here suggest that, the major indels and single nucleotide mutations within the first intron might have formed *via* an "imprecise" site-specific system involving a novel previously uncharacterised retroelement residing in the first intron. The induced mutations also resulted in the modification of functional elements within the first intron, which may have an impact on the regulation of the *FDPS* gene in *Hevea*. Moreover, a full sib progeny analysis of these retrotransposon harbouring alleles revealed their Mendelian mode of inheritance. The current study aims to expand our knowledge about the structural integrity of *Hevea FDPS* gene in general, with emphasis given to the impact of an intragenic retroelement on its structure and evolution.

2. Results

2.1. FDPS phylogenetic tree

The phylogenetic tree constructed based on the amino acid sequences clearly depicted the hierarchical linkage of this gene across major kingdoms and divisions. Clear differentiation based on sequence structure was observed for organisms from different strata's of life. The *H. brasiliensis FDPS* gene sequence was clustered along with its orthologs in 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 the 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 humans formed another major cluster whereas fungi and arthropods formed two separate lines (Fig. 1).

2.2. PCR amplification and primary sequence analysis of the entire FDPS gene from five popular clones

PCR amplification and sequence analysis of the entire genomic portion of the *FDPS* gene from five diverse popular *H. brasiliensis* clones was carried out with the intention of discovering SNPs. Samples used for the analysis are listed in Table 1. The analysis revealed the presence of eleven introns and twelve exons from the entire *FDPS* gene (Fig. 2). 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 FINT1-C harbouring strand of RRII-118. Interestingly, the "T" allele resulted in the introduction of a gibberellin responsive motif (GARE) [AAACAGA] in the minus strand (Supplementary material.SM.3).

Amplification of the first intronic region using the primer combination HbFDP2-F & HbFDP2-R from RRII-118, gave two bands of different size based on the electrophoretic mobility whereas, a

Fig. 1. FDPS phylogenetic tree: Phyogenetic tree constructed using the amino acid sequences of FDPS from selected plant species and representative organisms belonging to major domains of life. The different clusters depict the sequence structure diversity of the *FDPS* gene from representative organisms belonging to major domains of the living world.

single band of intermediate size was observed in the other four clones (Fig. 3). Primary sequence analysis from the cloned fragment data showed the presence of major indels within the first intronic region, which is 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 and not mere PCR artefacts or error. Interestingly, the rest of the gene from position 1380 bp onwards appeared to be conserved apart from the random 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 subjected to detailed analysis, separately from the rest of the gene.

2.2.1. Analysis of intron1 region from five popular clones

PCR amplification of the intron1 region using primers flanking the variable region (HbFDP2-F & HbFDP2-R) from the five selected clones yielded two unique alleles of approximately 760 bp and 675 bp from RRII-118 and a single allele of around 690 bp from the other four clones. The upper and lower alleles observed in RRII-118 were designated as FINT1-A and FINT1-C whereas, the common allele of intermediate size observed in the other four clones was named as FINT1-B. This disparity in size 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 (Fig. 3). Contrary to the variations exhibited by RRII-118, the region up to 1379 bp was considerably

Table 1
Samples and the type of analysis performed.

Whole gene sequencing (Both alleles)	Intron1 sequencing	Retro mapping
RRII-105 RRII-118 RRIM-600 RRIC-52 GT-1 (popular clones)	RRII-105 RRII-118 (both alleles) RRIM-600 RRIC-52 GT-1 RRII-5 (popular clones)	Forty popular clones [Includes RRII-5 and the five clones shown in column one]. Sixty wild accessions and five *Hevea* species [Includes those given in column two]. Progeny population [Sixty one plants]
	Acre-9 Acre-19 Rondonia-6 Rondonia-10 (wild accessions)	
	H.benthamiana *H.spruciana* *H.pauciflora* *H.nitida* (*Hevea species*)	

conserved in the clones RRII-105, RRIM-600, RRIC-52 and GT1 except a single SNP at position 463 where RRII-118, RRIC-52 and GT1 had "C/T" (heterozygous) while RRII-105 and RRIM-600 had "T/T" (homozygous) (Supplementary material.SM.4).

The first intron of the *FDPS* gene was a large phase-0 intron varying in size from 1358 bp to 1539 bp. 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 structure from the common allele present in the other four clones. The presence of a major insertion of 64 bp in the FINT1-A allele and a deletion of 22 bp in the FINT1-C allele was found to be responsible for the size difference and both of them formed part of the hyper variable region observed in the first 792 bp of the first intron (Fig. 4). Interestingly, the 64 bp insert in the FINT1-A allele appeared to be the fraction of a 96 bp repeat region constituted of three 32 bp repeats. The three repeats have almost 95% sequence similarity. 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 2. Apart from the above major indels, thirty five SNVs including three single nucleotide indels were identified from the first intronic region [Supplementary material SM.4]. 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. Considering the 35 single nucleotide variations alone, it was 17% (6/35) and 42% (15/35) respectively. 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 position 819 of FINT1-A, FINT1-C and FINT1-B respectively. Another multiallelic locus was identified at position 1065 (within the 22 bp indel) where FINT1-B had "T", FINT1-A had "G" and FINT1-C had "-" (the loci was within the 22 bp deleted region in FINT1-C) (Fig. 4). In addition to the above SNVs and indels, a 9 bp fragment at position 581 was found repeated two times in FINT1-B allele whereas no repetition was observed in FINT1-A and FINT1-C. Similarly a 3 bp fragment (CTA) at position 701 was found to be absent only in FINT1-B (Supplementary material.SM.3).

Blastx analysis of the first intron hyper variable region revealed its homology to a non-LTR retrotransposon of the RTE superfamily. The FINT1-A allele showed more homology to the retroelement

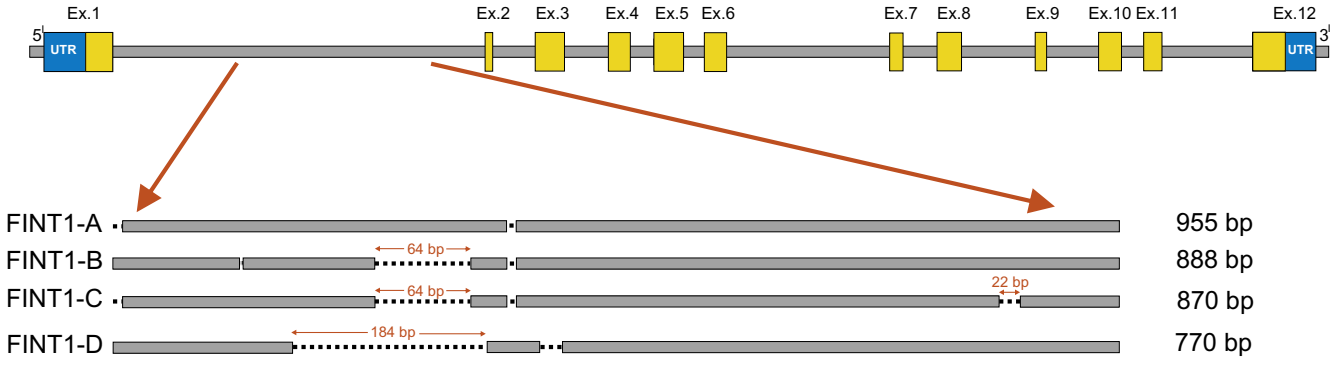

Fig. 2. Schematic representation of *FDPS* gene showing 5′ & 3′ UTRs, introns and exons: The four major intron1 alleles (FINT1-A, FINT1-B, FINT1-C, FINT1-D) with their size and location in the large-sized first intron are highlighted. The 64 bp indel in FINT1-B & FINT1-C, the 22 bp deletion in FINT1-C and the 184 bp deletion in FINT1-D are denoted by dots. The size of the alleles in base pairs is indicated in the right side. Ex = exon.

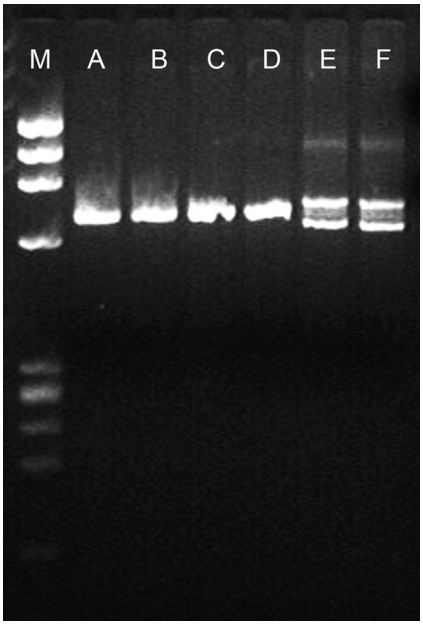

Fig. 3. Gel picture showing the *FDPS* intron1 alleles: A- RRII-105, B- RRIM-600, C-RRIC-52, D-GT1, E-RRII118, F- Mil 3/2 (one of the parents of RRII-118), M –Marker. RRII-105, RRIM-600, RRIC-52 and GT1 shared the same allele (middle band of intermediate size - FINT1-B). RRII-118 had two unique alleles (upper & lower bands – FINT1-A & FINT1-C). Mil 3/2 had the same allelic combination as that of RRII-118.

Table 2
Sequence variation within the intron1 major indels.

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 32 bp repeat	A	G	G	G	G
18th base position of the 32 bp repeat	T	T	T	A	A
19th base position of the 32 bp repeat	A	A	A	T	T
26th base position of the 32 bp repeat	G	A	G	A	G
29th base position of the 32 bp repeat	G	G	G	T	G
Status of 22 bp indel in each allele	present			present	absent
1065th base position (within the 22 bp 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 the three 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 retroelements. Deletion of 22 bp starting from position 1051 was observed only in FINT1-C allele.

than the other two alleles. Blastx analysis of this region with the maize transposable element database also showed its homology to a partial RIT_ jare like LINE retroelement which is a non-LTR retrotransposon (Table 3). The retroelement harbouring first intron also had a terminal polymorphic CT repeat having $(CT)^{13}$ in RRII-118 and $(CT)^{9}$ in the other four genotypes.

2.3. Estimation of FDPS intron1 allele status in additional popular clones, wild germplasm accessions, and Hevea species

In order to assess the distribution and status of the retrotransposon harbouring *FDPS* locus, 40 popular clones (including the initial five clones), 60 wild germplasm accessions and five *Hevea*

Fig. 4. Multiple-aligned *FDPS* intron1 allele sequences: The hyper-variable region of the four major first intronic alleles (FINT1-A, FINT1-B, FINT1-C and FINT1-D) showing the major indels. The major indel of 64 bp comprising of 32 bp repeats in FINT1-A starting from position 785 and the 22 bp deletion observed only in FINT1-C starting from position 1051 is highlighted. The core reverse transcriptase 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 deletion in FINT1-D from 977 bp resulting in two novel acceptor sites.

Table 3
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. spruceana*** (Unnamed)	888	Polyprotein, putative [*Solanum demissum*]	AAT40504.2	55.8	29%	2e-06	RIT_jare_AC204843-0	57	1e-10
6	***H. benthamiana*** (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

Blast results of the intron1 alleles from eleven representative plants (popular clones, wild germplasm accessions and *Hevea* species). NCBI blastx as well as Maize TE database tblastx results were shown in the descending order of their homology to retroelements. FINT1-D allele sequence showed no similarity to the retro sequence due to the large intron1 deletion. The smallest allele which was observed only among Rondonia accessions (Ron-6) had the maximum homology to the retro sequence.

species were analysed (Fig. 5A, B). Out of the 40 popular clones, 30, were found to be homozygous with 29 having FINT1-B allele and one (RRII-5) having FINT1-D, a new allele even smaller than FINT1-C. The smaller size was due to the deletion of a 184 bp region of the retroelement (Fig. 4). 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, additional alleles were identified exclusively from Rondonia accessions (Fig. 6A–C).

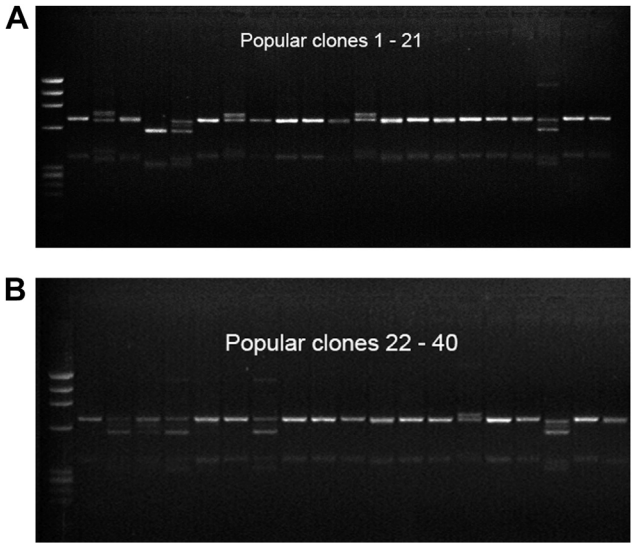

Fig. 5. A&B Gel pictures (5A and 5B) showing the *FDPS* intron1 allele status in forty popular clones including the initial five clones: All the four major alleles (FINT1-A, FINT1-B, FINT1-C, and FINT1-D) in various combinations can be observed. FINT1-D in homozygous state was seen in RRII-5 (5A-5th well).

2.4. Sequence analysis of intron1 alleles from selected plants

Intron1 alleles of selected plants from each group (popular clones, wild germplasm 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 already available allele sequences revealed that the most represented allele FINT1-B, among all the plants studied is highly conserved. Popular clones except RRII-118 and RRII-5 were clustered together (Fig. 8). Interestingly, the morphologically much different *Hevea 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 only to Acre plants. 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). Though minor single nucleotide variations were observed, *Hevea benthamiana*, *Hevea spruceana* and *Hevea nitida* were clustered together. Altogether from the above data, the 5′ end of first intron appeared to be highly destabilised.

2.5. Prediction of regulatory elements and splice sites from the first intronic region

Search for *cis*-regulatory elements in the first intronic region of the above plants exposed the presence of 19 major regulatory motifs. The *cis*-regulatory motifs present in the *FDPS* introne1 alleles of eleven representative plants are listed in Supplementary material [SM.5]. *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,

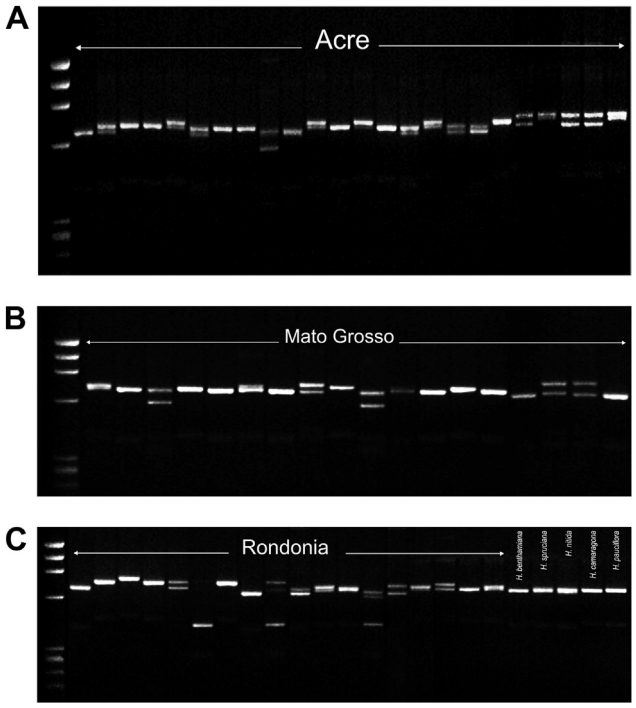

Fig. 6. A, B & C Gel pictures showing the *FDPS* intron1 allele status in sixty wild accessions and five *Hevea* species: 6A – 24 Acre accessions, 6B – 18 Mato Grosso accessions, 6C – 18 Rondonia accessions and five *Hevea* species. All the four alleles (FINT1-A, FINT1-B, FINT1-C, FINT1-D) in various combinations 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 allele 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 (Fig. 6C – 7th, 10th and 14th wells).

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. Alternatively, the *cis*-acting TCA element involved in salicylic acid responsiveness was found only in FINT1-C. A TC element involved in defence and stress responsiveness was found unique to *H. benthamiana* allele and an F-box binding domain with putative signal transduction role to FINT1-A allele.

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 (Fig. 4*) Similarly two additional acceptor sites not present in FINT1-A, FINT1-B and FINT1-C alleles were detected in FINT1-D due to an indel at position 977 (Supplementary material SM.3). Alternatively in the FINT1-C allele sequence, one acceptor site was modified to another site by a point mutation at position 1109 [Supplementary material.SM.3]. The intron1 allele sequences reported in this paper have been submitted to GenBank under the accession no KC886384 to KC886399.

2.6. Segregation analysis of intron1 alleles in a progeny population

In order to confirm that the alleles observed were of the same gene and not gene 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 of the cross, 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 of one gene with two alleles in the progeny of parental genotypes as described above, 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 respectively. Data for two plants were missing (Fig. 7). The expected and observed frequencies for the alleles and the genotypes along with the chi-square values are shown in Table 4.

2.7. Sequence analysis excluding 5′ end and intron1 hyper variable region

Altogether 25 SNPs were identified from the 3598 bp region starting from 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 GT-1 were found to be highly heterozygous with only five homozygous SNP loci. The entire region had only two exonic SNPs (4804 and 4813) and both of them were synonymous.

PHASE analysis as well as haplotype identification was carried out on the unphased genotypic data provided by using the Arlequin module of DnaSPv5 software [Supplementary material SM.6]. The software detected 25 polymorphic sites of which the SNPs at position 2127 (G/C), 2755 (A/G), 3475 (A/T), 4489 (C/T) and 4707 (C/T) were singleton variable sites with two variants. There were 20 parsimony informative sites with two variants. Seven haplotypes were identified with a haplotype diversity Hd: 0.867 (Table 5). R, between adjacent sites was estimated to be 0.0007 and the average nucleotide distance between the most distant sites was 3598.00. Three hundred pair wise comparisons were analysed out of which thirteen pairs of sites with four gametic types were detected. The minimum number of recombination event was calculated based on four-gamete test and one event was observed between site 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

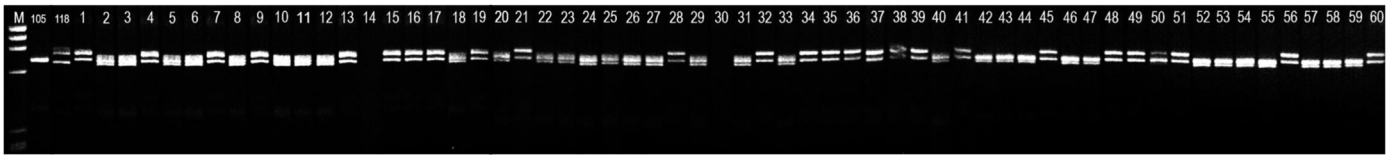

Fig. 7. *FDPS* intron1 allele segregation in a progeny population: The *FDPS* intron1 alleles, FINT1-A, FINT1-B and FINT1-C are segregating as per the Mendelian mode of inheritance in a full-sib progeny of the parents RRII-105 $^{\text{FINT1-B/FINT1-B}}$ (female) and RRII-118 $^{\text{FINT1-A/FINT1-C}}$ (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 was missing.

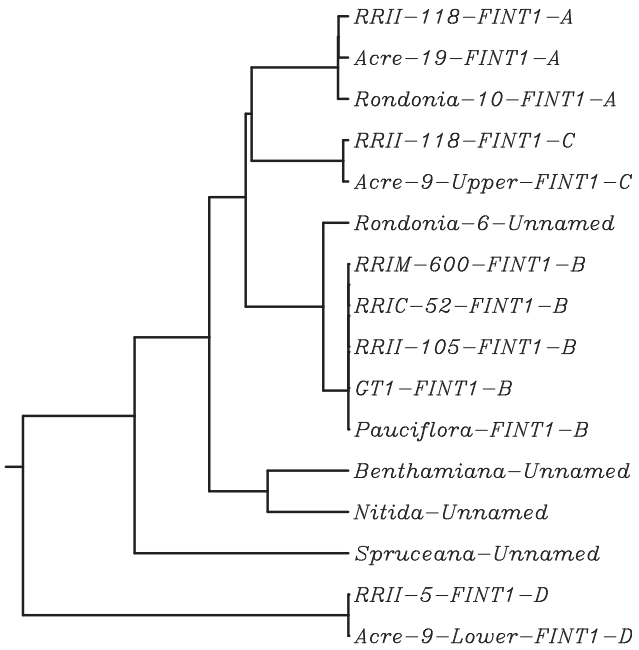

Fig. 8. Phylogenetic tree constructed using the intron1 allele sequences from representative plants of popular clones, wild accessions and *Hevea* species. (Plant name and allele name shown).

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. The linkage disequilibrium test shows that D′ between any two of the SNPs amid the 2nd to 25th equals one inferring that, SNPs found in this region have not been separated by recombination (or recurrent mutation) and presented a complete linkage disequilibrium during the evolution.

3. Discussion

Since *FDPS* codes for a major pathway enzyme involved in the synthesis of isoprenoid compounds, proper understanding of the existing structural variants of the *FDPS* gene in the *Hevea* gene pool and the functional properties associated with its variants may help

Table 4
Segregation ratios of three major *FDPS* intron1 alleles in a progeny population.

Parents			
RRII-105 (FINT1-B/FINT1-B) x RRII-118 (FINT1-A/FINT1-C)			
Allele	Expected frequency	Observed frequency	Contribution to chi-square
Allele frequency			
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			
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 the *FDPS* gene in a full sib family progeny. The alleles FINT1-A, FINT1-B and FINT1-C show no statistically significant deviation from the Mendelian inheritance mode.

Table .5
Haplotypes constructed using the 25 SNPs identified from the *FDPS* gene (excluding the first 1379 bp).

Hap_1	CCACCAAGGAAAAGTATTACGTGAG
Hap_2	CTGTGATGCGTGTGAATCTCACGGG
Hap_3	TTGTGCTGCATGAGATACTCATAGA
Hap_4	TCATCAAGGAAAATTATTACGTGAG
Hap_5	CCATCAAGGAAAATTATTACGTGAG
Hap_6	TTGTGCTCCATGAGATACTCATAGA
Hap_7	TTGTGCTGCATGAGATACTTATAGA

Haplotypes computed by DnaSP using the unphased twenty five SNP data starting from 1380 onwards. One recombination event was predicted by the software between the first and 2nd SNP (position 1380 and 1585).

in the quantitative and qualitative characterisation of rubber, a commercially important poly-isoprenoid obtained from *Hevea*. Moreover, identification of molecular mechanisms responsible for the generation of novel alleles of a gene like *FDPS* provide an evolutionary perspective for the studies on Mevalonate pathway genes in general.

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 [37]. From the phylogenetic tree constructed using FDPS protein sequences, it can be inferred that changes at genome level occurred in the organisms over millions of years. Even though the phylogram generated is a simple one based on a single protein sequence, the clustering, distribution and branching of each domain and the distance between them appeared to be very interesting due to its concordance with the evolutionary ladder of life from lower simple organisms to higher complex forms. The relationship of the organisms studied suggests that, there exist universally conserved as well as variable FDPS amino acid sequence for each group. Such variations at amino acid level occur mainly due to localised, intra-chromosomal scrambling fragmentation which was proved to be a general mechanism for generating diversity in taxa other than genetic recombination [38]. Moreover, the variations responsible for such groupings as mentioned above are often known to contribute to the evolutionary mechanism resulting in the generation of novel alleles. The same may apply for the *FDPS* gene in *H. brasiliensis* where, point mutations like SNPs, major additions and deletions may fuel microevolution within the species, thus enabling the selection of better adaptable clones. Altogether, our phylogenetic study using the FDPS sequences accentuates its suitability as a model gene for evolutionary studies of isoprenoid biosynthesis pathway genes.

Standing at a branch point in the isoprenoid pathway, FDPS synthesis is likely to be more tightly regulated than the other enzymes in the mevalonate pathway [39]. This may be the cumulative effect of various modes of gene regulatory mechanisms altogether or due to a single mechanism acting independently. In eukaryotes, gene level regulation of protein biosynthesis is achieved mainly by a complex mechanism involving *cis*-regulatory elements, transcription factors and epigenetic factors through their interaction with the promoter region of the concerned gene [40]. In plants, these factors are often triggered under stressed condition by signalling molecules like gibberellins (GAs) which are known players in mediating the effects of environmental stimuli on plant development [41]. The presence of a *cis*-acting Gibberellin responsive element (GARE) in the promoter region of an isoprenoid biosynthesis gene is a good indication of its regulation by gibberellins. The single nucleotide mutation at −11 bp upstream of the start codon resulting in the generation of a GARE site may suggest the existence of such a mechanism that might have evolved in few clones like RRII-118. However, the impact of the aforementioned SNV in

homozygous state in the clone RRII-118 has to be further investigated to prove the hypothesis.

Another element that decides the fate of a gene is the not-so-well studied non-coding region called intron. Even though introns are removed during the mRNA maturation process, they are known to enhance or regulate gene expression in numerous ways [8,42]. Recently several groups reported the presence of regulatory elements having a significant role in posttranslational modification and gene expression from the first intronic region of genes from higher plants [43–45]. 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 [46,47]. Potential mechanisms for such intron-mediated enhancement include increased transcription, splicing facilitated transcript maturation, stabilisation or export, and targeting of spliced transcripts for protein synthesis [45]. 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 having the notorious nature of high self establishment in the genic as well as non-genic regions. The repeated copy paste events during their propagation may often result in the reshuffling of targeted sites [48]. In plants, retrotransposon's often concentrate numerous sequence variations in-to short stretches of non-coding DNA by the induction of several SNPs and SSRs similar to the high SNP density observed in *FDPS* intron1 hyper variable region of *H. brasiliensis*. The similarity of this intronic region to a partial reverse transcriptase 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 position 582 (9 bp repeat), 785 (32 bp repeat) and the CT repeat at position 288 are apparent from 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 reverse transcripts initiating at the 3′end of the RNA [49]. 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, is another unambiguous indication of the presence of a non-LTR retrotransposable element [50]. 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. The above statement justifies the absence of a structurally complete TE and RT element within the *FDPS* intron1 of *H. brasiliensis*. Mutations like the 35 single nucleotide variations resulted from the copy-paste event 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 [51–53]. Evidence for this possibility is obvious from the target site duplications present in the *FDPS* intronic 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, FINT1-C in the analysed populations may not be just coincidence, but rather an indication of its stability due to the removal of a retroelement sequence. 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. But their evolutionarily advantage over other alleles if any has to be proved by further experiments like the gene expression profiling of different clones harbouring different allelic forms, and impact of splice variants leading to isoforms with differential intracellular localisation, protein structural studies etc.

The discovery of new alleles with minor to major variation from the three major alleles discussed earlier from additional 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 selected popular clones and retro-mapping of all the forty clones revealed that the intron1 region up to 1379 bp is partially conserved, since 72% of clones had only the FINT1-B allele. This preference may owe partially to the selective breeding strategy using parents of narrow genetic variability. Alternatively, this can be attributed to natural selection also, as evident from the surprisingly similar trend observed in wild accessions and *Hevea* species. Same explanation may apply to the occurrence of the smaller FINT1-D allele with moderately high frequency (20%) in popular clones. The popular clone genotyping data 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 higher similarity to the retroelement were not spotted in homozygous state and rarely found together with exceptions like in 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 in them. Furthermore, this trend indicates the existence of more genetic diversity in Rondonia accessions which could be explored further for broadening the current genetic base of *Hevea* clones.

Though several isoforms of *FDPS* from *Hevea* has been reported by earlier studies, all the alleles mentioned in the present study are supposed to be of the same gene because, analysis of individuals from several DNA collections (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 for the diploid *Hevea* [[54], Saleena et al., Un-published]. 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 Mendels law of inheritance for alleles and genotypes ($P > 0.10, 0.05$) indicating an allele segregation pattern in accordance with the Mendelian mode of inheritance. All the above results are well confirmatory for the assumption that the alleles observed were of the same gene.

3.1. Regulatory motif analysis of intron1 sequence

The presence of indels and SNVs in the regulatory region of a gene may have a significant role in gene regulation due to the disruption of its 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 [19]. In plants, Jasmonates are known signalling compounds for the production of defence related compounds like terpenoids, glycosteroids and alkaloids. They do so mainly by co-regulating the concerned gene through their interactions with its CGTCA element [55–57]. Therefore, the presence and absence of this motif in the first intronic region of the *FDPS* gene indicate the possibility of a jasmonate-mediated regulation of isoprenoid biosynthesis in *Hevea*. The role of a retroelement in the modification of this motif further led us to assume that it is an epigenetic mode of gene regulatory mechanism, because similar instances were reported previously in other systems [58,59]. Therefore, the disruption of CGTCA-motif due to a retrotransposon induced single nucleotide mutation in both the intron1 alleles of the *FDPS* gene in RRII-118 may be a retrotransposon-induced negative MeJa responsive regulatory mechanism that exists 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. On the other hand, it may also be considered as a self imposed modification by the retroelement itself for its multiplication. Nevertheless, the presence of an intact 5UTR Py-rich stretch (conferring high transcription levels) and F-box (signal transduction response domain) 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. Similarly, the additional TATA box and CAAT box observed in the intron1 is also assumed to be the regulatory elements of the retroelement although, experimental evidence lacks presently. On the other hand the presence of defence and stress responsiveness element like "TC" solely in *H. benthamiana* is justified as this species is considered to be comparatively more disease-tolerant than *H. brasiliensis* [60].

Besides the disruption of regulatory motifs, insertion of a transposable element in intronic region can also lead to exonisation as well as to new patterns of pre-mRNA processing, resulting in novel responses to developmental and environmental stimuli [61,62]. 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 [63]. 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 utilised efficiently by the spliceosome, transposons may get inserted between sequences without altering the coding sequence of the gene. This may be the reason for the highly conserved coding region of the *FDPS* gene despite its 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 [64]. Transposable elements are also known to eliminate or create canonical splice sites in regulatory region *via* single nucleotide variation [65,66]. A single nucleotide change at position 688 [G/C] (Fig. 4) resulting in the absence of a donor site "ggaggGTgagtc" in FINT1-A is a clear evidence for such a change 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.

3.2. Sequence analysis from 1380 bp onwards

Based on earlier studies from our own laboratory the frequency of one SNP in average every 143 base pairs observed in the last 3598 bp region of the *FDPS* gene in *Hevea* suggest a normal rate of mutation (Saha et al., Unpublished). Therefore it can be assumed that the TE impact is mostly limited to the first 1379 bases of the *FDPS* gene which includes mainly the large first intron. This is supported by earlier reports that TE exonizations as well as alternatively spliced exons tend to occur within or near the first introns of genes that are usually longer than the other introns. This stands true in the case of the *Hevea FDPS* intron1 as it is more than twice the length of the second longest intron. The conserved nature of the rest of the gene is evident from the phased SNP data which emphasise mainly three haplotypes (hap-1,hap-2 and hap-3) Table 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 has 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 loci. 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 high homozygosity and the presence of hap-1 in both RRII-105 and RRIM-600 can also be attributed to their close genetic relationship [Same female parent of Indonesian origin-Tjir1; Different male parents of common Malaysian origin- Gl-1 & PB-86]. The presence of hap-1 haplotype in the Sri Lankan clone RRIC-52 and the Malaysian clone GT1 further supports the assumption of a presumed common parent. The entirely different lineage of RRII-118 (Cross 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%). As their alleles (FINT1-A & FINT1-C) occurred more frequently among Acre and Rondonia accessions than in Mato Grosso, they are less likely to originate from Mato Grosso, unlike the other popular clones. Alternatively, the uniqueness of the hap-2 haplotype (coming downstream of the FINT1-C allele) may owe to historical-reasons, also based on the assumption that Sri Lanka might have retained some seedlings of the initial Wickham collection to themselves while acting as a distribution centre to south Asian countries.

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 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 by the TE present in the first intron.

Altogether, this is the first report on the presence of intragenic retroelements in the *Hevea* genome and their influence on gene

structure. Our results suggest that the indels and SNPs in the first intron were generated as a result of the retroelement activity within the same intron. The induced sequence variations eventually resulted in regulatory motif and splice site modifications, which may have some functional significance in gene regulation. Analysis of *Hevea* populations revealed that the retrotransposon activity has generated several unique inheritable *FDPS* alleles, which lends perspective to the view that transposable elements play a major role in the generation of novel alleles in plants. The functional impact of these important modifications on *FDPS* gene expression in *Hevea* with respect to selected clones, identification of the most favourable structural variant of *FDPS* and development of suitable SNP markers for screening *Hevea* populations, once marker trait association is established are the aims of future work in this direction.

4. Materials and methods

4.1. FDPS phylogenetic tree analysis

In-order to understand the sequence diversity and evolutionary relationship of *FDPS*, a phylogenetic tree was constructed using FDPS protein sequences downloaded from NCBI database. Sequence from one representative organisms each belonging to the major domains of life other than plants was taken for the analysis. In the case of plants, FDPS sequences from related as well as distant genera/family/species were considered for a comparative relationship study with *Hevea* sequence. The sequences were aligned and tree constructed using ClustalW. Details of the sequences used for the analysis are given in the Supplementary material SM.2.

4.2. Initial sampling, PCR amplification and sequencing of entire FDPS gene

The initial sample material consisted of five genetically as well as phenotypically diverse popular *Hevea* clones *viz*., RRII-105, RRII-118, RRIM-600, RRIC-52 and GT1 being cultivated extensively in the Asia pacific region. They were selected with the objective of identifying SNP markers from the genes involved in rubber biosynthesis. RRII-105 and RRII-118 are high-yielding Indian clones whereas RRIM-600 is a high-yielding Malaysian clone. Both RRIC-52 and GT1 are primary clones, the former developed in Sri Lanka and the latter in Indonesia. Leaf genomic DNA was isolated following CTAB protocol [67]. The entire *FDPS* genomic sequence of around 4.9 kb was downloaded from NCBI (EF593108.1) and eight set of overlapping primers were designed spanning the entire gene [List of primers in Supplementary material SM.1]. PCR amplification was performed in a total volume of 50 μl containing 100 ng of template DNA with 0.2 μM of each primer, 0.2 mM of each dNTP, 1X Taq DNA polymerase (Advantage Taq, Clontech) and 5 μl of DNA polymerase buffer. PCR conditions were as follows. An initial denaturation of 95 °C for 2 min was followed by 95 °C for 30 s, 54 °C for 30 s, 68 °C for 40 s for a total of 35 cycles, 4 min at 68 °C, and hold at 4 °C. The amplified products were checked on 1.5% agarose gel and documented. The bands of interest were eluted using the illustra GFX gel band purification kit (GE Healthcare). Simultaneously, the purified products were TA cloned in pGEMT easy vector (Promega USA). Direct sequencing of the PCR product as well as sequencing of the cloned product (duplicate colonies) was performed at Macrogen Inc., Seoul, South Korea. In addition, the three different intron1 alleles obtained from the initial analysis of five clones (FINT1-A & FINT1-C from RRII-118 and FINT1-B from the other four clones) were eluted separately from the gel, cloned and sequenced.

4.3. Screening of additional popular clones, wild accessions and Hevea species for intron1 alleles

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*. Popular *Hevea* clones consist of all the prominent varieties of rubber clones cultivated in the rubber growing regions of the world which were originally derived from a single collection made by Sir Henry Alexander Wickham in 1876 from a small area near the confluence of the river Tapajos with the Amazon [68]. Simultaneously, wild accessions comprise 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. The wild accessions include 24 plants from Acre, 18 from Mato Grosso and 18 from Rondonia. *FDPS* intron1 loci of latex producing *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′ & HbFDP2-R-5′- TGCCAAGAAGTTAAAGGATAACAAA-3′). The PCR products were checked on 2.5% agarose gel for allele identification and scoring. The sample details and the type of analysis performed is summarised in Table 1.

4.4. Sequence analysis of the 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 the multiple sequence alignment module of DNASIS MAX (Hitachi Solutions America, Ltd). 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 screening done in thirty five additional popular clones, 60 wild accessions and five *Hevea* species, cloning and sequencing of the *FDPS* intron1 region alone was carried out from selected plants. The selection was made based on fragment size variation observed during electrophoretic run. The additional plants include one popular clone (RRII-5), four wild accessions and four *Hevea* species. The list of plants analysed is given in Table 1.

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 from the above plants were also analysed using plantCARE software online (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to identify the putative *cis* regulatory elements [69].

4.5. Segregation analysis of intron1 alleles in a progeny population

Genotyping of a full-sib progeny (F1 progeny) of sixty one individuals derived from a controlled cross between the cultivars RRII-105 and RRII-118 was performed to assess the segregation

pattern of intron1 alleles. Genotyping was performed using the same primer combination mentioned earlier. The alleles were scored after running the PCR products in 2.5% agarose gel. The allele and genotype frequency was estimated and the contribution to chi-square was calculated from the values obtained.

4.6. Sequence analysis excluding the 5′ end and intron1 hyper variable region

The entire *FDPS* gene sequence data was obtained only from the five clones mentioned above. Since the *FDPS* gene sequence from 1380 bp onwards appeared to be stable and free from major structural variations, this portion was analysed separately for the sake of simplicity in data interpretation. The sequences were edited and aligned using DNASIS MAX software. SNPs and indels were identified by comparing direct PCR product sequence and cloned fragment sequence. Heterozygous loci were identified by locating double peaks in the chromatogram followed by confirmatory crosschecking with the cloned fragment sequence. After identification and confirmation of the SNPs, they were analysed using the DNA sequence analysis module of DnaSP for haplotype identification [70]. 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. It was also used to estimate the recombination rate along the sequences. A pure parsimony approach was used to estimate the haplotypes and the optimal solution which requires less haplotypes to resolve the genotypes was selected. The sample details and the type of analysis performed on the respective samples are summarised in Table 1.

Acknowledgements

We thank Dr James Jacob, Director of Research, Rubber Research Institute of India for his constant encouragement. We are extremely grateful to Mr. Madhusoodanan, Scientist, Rubber Technology division, RRII for the art work. Plant materials for DNA isolation was contributed by Dr. Kavitha K. Mydin, Joint Director, Crop Improvement group and Dr. Jayashree Madhavan, Scientist, Germplasm Division of RRII. No conflict of interest is declared.

Appendix A. Supplementary material

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.plaphy.2013.09.004.

References

[1] B.F. Greek, Rubber demand is expected to grow after 1991, C & E News 69 (1991) 37–54.
[2] B. Nga, S. Subramaniam, Variations in *Hevea brasiliensis*: yield and girth data of the 1937 hand pollinated seedlings, J. Rubber Res. Inst. Malayasia 24 (1974) 69–74.
[3] J. Licy, A.O.N. Panikkar, D. Premakumari, A.Y. Varghese, M.A. Nazeer, Genetic parameters and heterosis in *Hevea brasiliensis*, Indian J. Nat. Rubber Res. 5 (1992) 51–56.
[4] C. Narayanan, K.K. Mydin, Breeding for Disease Resistance in *Hevea* Spp. STatus, Potential, Threats, and Possible Strategies, 2012. Gen. Tech. Rep. PSW-GTR-240, USDA station, Albany, California, USA.
[5] C.K. Saraswathyamma, J. Licy, G. Marattukalam, Planting materials, in: P.J. George, C.K. Jacob (Eds.), Natural Rubber: Agromanagement and Crop Processing, Rubber Research Institute of India, Kottayam, 2000, pp. 59–74.
[6] P. Besse, M. Seguin, P. Lebrun, M.H. Chevallier, D. Nicolas, C. Lanaud, Genetic diversity among wild and cultivated populations of *Hevea brasiliensis* assessed by nuclear RFLP analysis, Theor. Appl. Genet. 88 (1994) 199–207.
[7] N. Lekawipat, K. Teerawatanasuk, M. Rodier-Goud, M. Seguin, A. Vanavichit, T. Toojinda, S. Tragoonrung, Genetic diversity analysis of wild germplasm and cultivated clones of *Hevea brasiliensis* Muell. Arg. by using microsatellite markers, J. Rubber Res. 6 (2003) 36–44.
[8] R.A. Hernandez, K.L. Afanador, I.R. Arango, A.M. Lobo, Analysis of genetic variation in clones of rubber (*Hevea brasiliensis*) from Asian, south and central American origin using RAPDs markers, Revista Colombia de Biotechnología 8 (2006) 29–34.
[9] K.S. Chow, K.L. Wan, M.N. Isa, A. Bahari, S.H. Tan, K. Harikrishna, H.Y. Yeang, Insights into rubber biosynthesis from transcriptome analysis of *Hevea brasiliensis* latex, J. Exp. Bot. 58 (2007) 2429–2440.
[10] A. Takaya, Y.W. Zhang, K. Asawatreratanakul, D. Wititsuwannakul, R. Wititsuwannakul, S. Takahashi, T. Koyama, Cloning, expression and characterization of a functional cDNA clone encoding geranylgeranyl diphosphate synthase of *Hevea brasiliensis*, Biochem. Biophys. Acta J. 1625 (2003) 214–220.
[11] D. Mekkriengkrai, T. Sando, K. Hirooka, J. Sakdapipanich, Y. Tanaka, E. Fukusaki, A. Kobayashi, Cloning and characterization of *farnesyl diphosphate synthase* from the rubber-producing mushroom *Lactarius chrysorrheus*, Biosci. Biotechnol. Biochem. 68 (2004) 2360–2368.
[12] A.B. Rose, The effect of intron location on intron-mediated enhancement of gene expression in *Arabidopsis*, Plant J. 40 (2004) 744–751.
[13] A.B. Rose, Intron-mediated regulation of gene expression, Curr. Top. Microbiol. Immunol. 326 (2008) 277–290.
[14] A.R. Buchman, P. Berg, Comparison of intron-dependent and intron-independent gene expression, Mol. Cell Biol. 8 (1988) 4395–4405.
[15] S. Chung, R. Perry, Importance of introns for expression of mouse ribosomal protein gene *rpL32*, Mol. Cell Biol. 9 (1989) 2075–2082.
[16] J. Meredith, R.V. Storti, Developmental regulation of the *Drosophila tropomyosin II* gene in different muscles is controlled by muscle-type-specific intron enhancer elements and distal and proximal promoter control elements, Dev. Biol. 159 (1993) 500–512.
[17] P.G. Okkema, S.W. Harrison, V. Plunger, A. Aryana, A. Fire, Sequence requirements for *myosin* gene expression and regulation in *Caenorhabditis elegans*, Genetics 135 (1993) 385–404.
[18] R. Hong, L. Hamaguchi, M.A. Busch, D. Weigel, Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing, The Plant Cell 15 (2003) 1296–1309.
[19] S.E. Schauer, P.M. Schlüter, R. Baskar, J. Gheyselinck, A. Bolaños, M.D. Curtis, U. Grossniklaus, Intronic regulatory elements determine the divergent expression patterns of *AGAMOUS-LIKE6* subfamily members in *Arabidopsis*, Plant J. 59 (2009) 987–1000.
[20] A.B. Rose, J.A. Beliakoff, Intron-mediated enhancement of gene expression independent of unique intron sequences and splicing, Plant Physiol. 122 (2000) 535–542.
[21] R. Karve, W. Liu, S.G. Willet, K.U. Torii, E.D. Shpak, The presence of multiple introns is essential for *ERECTA* expression in *Arabidopsis*, RNA 17 (2011) 1907–1921.
[22] M. Donath, R. Mendel, R. Cerff, W. Martin, Intron-dependent transient expression of the maize *GapA1* gene, Plant Mol. Biol. 28 (1995) 667–676.
[23] J. Majewski, J. Ott, Distribution and characterization of regulatory elements in the human genome, Genome Res. 12 (2002) 1827–1836.
[24] K.R. Bradnam, I. Korf, Longer first introns are a general property of eukaryotic gene structure, PLoSONE 3 (8) (2008) e3093, http://dx.doi.org/10.1371/journal.pone.0003093.
[25] Y. Jeong, J. Mun, I. Lee, J.C. Woo, C.B. Hong, S. Kim, Distinct roles of the first introns on the expression of *Arabidopsis profilin* gene family members, Plant Physiol. 140 (2006) 196–209.
[26] T. Sando, C. Takaoka, Y. Mukai, A. Yamashita, M. Hattori, N. Ogasawara, E. Fukusaki, A. Kobayashi, Cloning and characterization of mevalonate pathway genes in a natural rubber producing plant, *H. brasiliensis*, Biosci. Biotechnol. Biochem. 72 (2008) 2049–2060.
[27] B. Wang, J. Mason Depasse, W.B. Watt, Evolutionary genomics of colias *phosphoglucose isomerase* (*PGI*) introns, J. Mol. Evol. 74 (2012) 96–111, http://dx.doi.org/10.1007/s00239-012-9492-5.
[28] M.J. Varagona, M. Purugganan, S.R. Wessler, Alternative splicing induced by insertion of retrotransposons into the *maize waxy* gene, The Plant Cell 4 (1992) 811–820.
[29] A. Damert, J. Raiz, A.V. Horn, J. Löwer, H. Wang, J. Xing, M.A. Batzer, R. Löwer, G.G. Schumann, 5′-Transducing SVA retrotransposon groups spread efficiently throughout the human genome, Genome Res. 19 (2009) 1992–2008.
[30] K.C. Huang, H.C. Yang, K.T. Li, L.Y. Liu, Y.C. Charng, Ds transposon is biased towards providing splice donor sites for exonization in transgenic tobacco, Plant Mol. Biol. 79 (2012) 509–519.
[31] G. Witzany, The agents of natural genome editing, J. Mol. Cell Biol. 3 (2011) 181–189.
[32] Kapazoglou, et al., The study of two barley Type I-like *MADS-box* genes as potential targets of epigenetic regulation during seed development, BMC Plant Biol. 12 (2012) 166, http://dx.doi.org/10.1186/1471-2229-12-166.
[33] J. Liu, Y. He, R. Amasino, Chen, siRNAs targeting an intronic transposon in the regulation of natural flowering behaviour in *Arabidopsis*, Genes Dev. 18 (2004) 2873–2878.
[34] B. Greene, R. Walko, S. Hake, Mutator insertions in an intron of the maize *knotted1* gene results in dominant suppressible mutations, Genetics 1384 (1994) 1275–1285.

[35] A.Y.A. Rahman, et al., Draft genome sequence of the rubber tree *Hevea brasiliensis*, BMC Genomics 14 (2013) 75, http://dx.doi.org/10.1186/1471-2164-14-75.
[36] T. Saha, B. Roy, M. Ravindran, K. Bini, M.A. Nazeer, Existence of retroelements in rubber (*Hevea brasiliensis*) genome, J. Plant. Crops 34 (2006) 546–551.
[37] G. Emiliani, D. Paffetti, R. Giannini, Identification and molecular characterization of LTR and LINE retrotransposable elements in *Fagus sylvatica L*, iForest - Biogeosciences For. 2 (2009) 119–126.
[38] B. Kloeckner-Gruissem, M. Freeling, Transposon-induced promoter scrambling: a mechanism for the evolution of new alleles, Proc. Natl. Acad. Sci. 92 (1995) 1836–1840.
[39] V. Keim, D. Manzano, F.J. Fernández, M. Closa, P. Andrade, D. Caudepón, C. Bortolotti, M.C. Vega, M. Arró, A. Ferrer, Characterization of *Arabidopsis* FPS isozymes and *FPS* gene expression analysis provide insight into the biosynthesis of isoprenoid precursors in seeds, PLoS One 7 (11) (2012) e49109, http://dx.doi.org/10.1371/journal.pone.0049109.
[40] T.K. Uthup, M. Ravindran, K. Bini, T. Saha, Divergent DNA methylation patterns associated with abiotic stress in *Hevea brasiliensis*, Mol. Plant 4 (2011) 996–1013.
[41] Valerie Sponsel, Gibberellins: regulators of plant height, in: L. Taiz, E. Zeiger (Eds.), Plant Physiology, fifth ed., Sinauer Associates, Inc, USA, 2010.
[42] S. Morita, S. Tsukamoto, A. Sakamoto, H. Makino, E. Nakauji, H. Kaminaka, T. Masumura, Y. Ogihara, S. Satoh, K. Tanaka, 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 (2012) 115–119.
[43] U. Gowik, J. Burscheidt, M. Akyildiz, U. Schlue, M. Koczor, M. Streubel, P. Westhoff, *cis*-Regulatory elements for mesophyll-specific gene expression in the C4 plant *Flaveria trinervia*, the promoter of the C4 *phosphoenolpyruvate carboxylase* gene, Plant Cell 16 (2004) 1077–1090.
[44] L. Morello, D. Breviario, Plant spliceosomal introns: not only cut and paste, Curr. Genomics 4 (2008) 227–238.
[45] G. Parra, K. Bradnam, A.B. Rose, I. Korf, Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants, Nucl. Acids Res. 13 (2011) 5328–5337.
[46] M. Clancy, L.C. Hannah, 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 (2002) 918–929.
[47] A.B. Rose, Requirements for intron-mediated enhancement of gene expression in *Arabidopsis*, RNA 8 (2002) 1444–1453.
[48] D. Gao, J. Chen, M. Chen, B.C. Meyers, S. Jackson, A highly conserved, small ltr retrotransposon that preferentially targets genes in grass genomes, PLoS ONE 7 (2) (2012) e32010, http://dx.doi.org/10.1371/journal.pone.0032010.
[49] D.D. Luan, M.H. KORMAN, J.L. Jakubczak, T.H. Eickbush, Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition, Cell 72 (1993) 595–605.
[50] J.S. Han, A subtelomeric non-LTR retrotransposon Hebe in the bdelloid rotifer *Adineta vaga* is subject to inactivation by deletions but not 5' truncations, Mobile DNA 1 (2010) 12, http://dx.doi.org/10.1186/1759-8753-1-12.
[51] L.N. Lagemaat, L. Gagnier, P. Medstrand, D.L. Mager, Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates, Genome Res. 9 (2005) 1243–1249.
[52] M. Muñoz-López, J.L. García-Pérez, DNA transposons: nature and applications in genomics, Curr. Genomics 11 (2010) 115–128.
[53] E. Kejnovsky, J.S. Hawkins, C. Feschotte, Plant transposable elements: biology and evolution, in: J.F. Wendel, et al. (Eds.), Plant Genome Diversity, vol. 1, Springer-Verlag, Wien, 2012, pp. 17–34.
[54] K.S. Chow, M.N. Isa, A. Bahari, A. Ghazali, H. Alias, Z. Mohd.-Zainuddin, C. Hoh, K.L. Wan, Metabolic routes affecting rubber biosynthesis in *Hevea brasiliensis* latex, J. Exp. Bot. (2011), http://dx.doi.org/10.1093/jxb/err363.
[55] J. Rouster, L. Robert, J. Mundy, V. Cameron-Mills, Identification of a methyl jasmonate-responsive region in the promoter of a *lipoxygenase 1* gene expressed in barley grain, Plant J. 11 (1997) 513–523.
[56] Y. He, S. Gan, Identical promoter elements are involved in regulation of the *OPR1* gene by senescence and jasmonic acid in *Arabidopsis*, Plant Mol. Biol. 47 (2001) 595–605.
[57] V. Chinnusamy, J.K. Zhu, Epigenetic regulation of stress responses in plants, Curr. Opin. Plant Biol. 12 (2009) 133–139.
[58] E. Whitelaw, D.I.K. Martin, Retrotransposon as epigenetic mediators of phenotypic variation in mammals, Nat. Genet. 27 (2001) 361–365.
[59] K. Kashkush, M. Feldman, A.A. Levy, Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat, Nat. Genet. 33 (2002) 102–106.
[60] S.M. Jain, P.M. Priyadarshan (Eds.), Breeding Plantation Tree Crops: Tropical Species, Springer, New York, USA, 2009.
[61] C.F. Weil, S.R. Wessler, The effects of plant transposable element insertion on transcription initiation and RNA processing, Annu. Rev. Plant Physiol. Plant Mol. Biol. 41 (1990) 527–552.
[62] N. Sela, E. Kim, G. Ast, The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates, Genome Biol. 11 (2010) R59, http://dx.doi.org/10.1186/gb-2010-11-6-r59.
[63] N.J. Dibb, A.J. Newman, Evidence that introns arose at proto-splice sites, EMBO J. 8 (1989) 2015–2021.
[64] P. Yenerall, L. Zhou, Identifying the mechanisms of intron gain: progress and trends, Biol. Direct 7 (2012) 29.
[65] X. Huang, G. Lu, Q. Zhao, X. Liu, B. Han, Genome-wide analysis of transposon insertion polymorphisms reveals intraspecific variation in cultivated rice, Plant Physiol. 148 (2008) 25–40.
[66] A.M. Barbaglia, K.M. Klusman, J. Higgins, J.R. Shaw, L.C. Hannah, S.K. Lal, Gene capture by helitron transposons reshuffles the transcriptome of maize, Genetics 3 (2012) 965–975.
[67] J.J. Doyle, J.L. Doyle, Isolation of plant DNA from fresh tissue, Focus 12 (1990) 13–15.
[68] R.E. Schultes, Wild *Hevea* – an untapped source of germplasm, J. Rubber Res. Inst. Sri Lanka 54 (1977) 1–31.
[69] M. Lescot, P. Déhais, G. Thijs, K. Marchal, Y. Moreau, Y. Van de Peer, P. Rouzé, S. Rombauts, PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in-silico* analysis of promoter sequences, Nucleic Acids Res. 30 (2002) 325–327.
[70] P. Librado, J. Rozas, DnaSP v5: a software for comprehensive analysis of DNA polymorphism data, Bioinformatics 11 (2009) 1451–1452.