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