

## Transcriptome Sequencing of *Hevea brasiliensis* for Development of Microsatellite Markers and Construction of a Genetic Linkage Map

KANOKPORN Triwitayakorn<sup>1</sup>, PORNSUPA Chatkulkawin<sup>1</sup>, SUPANATH Kanjanawattanawong<sup>1,2</sup>, SUPAJIT Sraphet<sup>1</sup>, THIPPAWAN Yoocha<sup>3</sup>, DUANGJAI Sangsrakru<sup>3</sup>, JUNTIMA Chanprasert<sup>3</sup>, CHUMPOL Ngamphiw<sup>3</sup>, NUKOON Jomchai<sup>3</sup>, KANIKAR Therawattanasuk<sup>4</sup>, and SITHICHOKE Tangphatsornruang<sup>3,\*</sup>

*Institute of Molecular Biosciences, Mahidol University, Salaya, Nakhon Pathom 73170, Thailand<sup>1</sup>; Department of Biotechnology, Faculty of Science, Mahidol University, Phayathai, Bangkok 10400, Thailand<sup>2</sup>; National Center for Genetic Engineering and Biotechnology, 113 Phaholyothin Rd., Klong 1, Klong Luang, Pathumthani 12120, Thailand<sup>3</sup> and Department of Agriculture, Ministry of Agriculture and Cooperatives, Rubber Research Institute of Thailand, Chachoengsao, Thailand<sup>4</sup>*

\*To whom correspondence should be addressed. Tel. +66 2-5646700; Fax. +66 2-5646707.  
E-mail: sithichoke.tan@biotec.or.th

Edited by Satoshi Tabata  
(Received 30 June 2011; accepted 29 August 2011)

### Abstract

**To obtain more information on the *Hevea brasiliensis* genome, we sequenced the transcriptome from the vegetative shoot apex yielding 2 311 497 reads. Clustering and assembly of the reads produced a total of 113 313 unique sequences, comprising 28 387 isotigs and 84 926 singletons. Also, 17 819 expressed sequence tag (EST)-simple sequence repeats (SSRs) were identified from the data set. To demonstrate the use of this EST resource for marker development, primers were designed for 430 of the EST-SSRs. Three hundred and twenty-three primer pairs were amplifiable in *H. brasiliensis* clones. Polymorphic information content values of selected 47 SSRs among 20 *H. brasiliensis* clones ranged from 0.13 to 0.71, with an average of 0.51. A dendrogram of genetic similarities between the 20 *H. brasiliensis* clones using these 47 EST-SSRs suggested two distinct groups that correlated well with clone pedigree. These novel EST-SSRs together with the published SSRs were used for the construction of an integrated parental linkage map of *H. brasiliensis* based on 81 lines of an F1 mapping population. The map consisted of 97 loci, consisting of 37 novel EST-SSRs and 60 published SSRs, distributed on 23 linkage groups and covered 842.9 cM with a mean interval of 11.9 cM and ~4 loci per linkage group. Although the numbers of linkage groups exceed the haploid number (18), but with several common markers between homologous linkage groups with the previous map indicated that the F1 map in this study is appropriate for further study in marker-assisted selection.**

**Key words:** transcriptome sequencing; marker development; rubber tree (*Hevea brasiliensis*); linkage map construction

### 1. Introduction

*Hevea brasiliensis*, commonly known as rubber tree, is almost the sole source of natural rubber production. Natural rubber has a wide range of industrial applications and is under increasing global demand. *Hevea brasiliensis* is a perennial cross-pollinating and

monoecious plant that belongs to the Euphorbiaceae family. The observation of tetraivalents during meiosis has led to the conclusion that *H. brasiliensis* is a stabilized amphidiploid ( $2n = 4x = 36$ ).<sup>1</sup> However, the pattern of marker ratios segregating in a population of over 100 trees suggests that *H. brasiliensis* behaves as a diploid ( $2n = 36$ ).<sup>2</sup> Several research groups have

developed molecular markers to study the genetic diversity of *H. brasiliensis*,<sup>3–8</sup> including isozymes,<sup>9</sup> restriction fragment length polymorphisms (RFLPs),<sup>10</sup> amplified fragment length polymorphisms (AFLPs),<sup>2</sup> microsatellites (or simple sequence repeats, SSRs)<sup>2,8,11,12</sup> and expressed sequence tag (EST)-SSRs.<sup>3</sup> These markers have also been used to construct the linkage maps<sup>2</sup> and quantitative trait loci (QTL) maps.<sup>13–16</sup> Lespinasse *et al.*<sup>14</sup> produced the first rubber tree linkage map containing 301 RFLPs, 388 AFLPs, 18 SSRs and 10 isozymes, which was later used to identify the QTL variants conferring resistance to the South American leaf blight.<sup>13–15</sup> Recently, Le Guen *et al.*<sup>16</sup> constructed the linkage maps based on SSR and AFLP markers and were able to identify the QTL conferring resistance to *Microcyclus ulei*.

In many organisms, ESTs have been useful for the annotation of genes during genome sequencing efforts,<sup>17</sup> for comparative genome studies<sup>18</sup> and for the production of a genetic linkage map.<sup>19</sup> To date, there are only 12 365 ESTs from *H. brasiliensis* in GenBank, restricting the quality of research that can be performed on this important plant species. Previous transcriptome studies of *H. brasiliensis* have been limited in range, focusing mainly on latex in order to gain insight into the rubber biosynthesis pathways.<sup>20–22</sup> Of the available *H. brasiliensis* ESTs, 11 256 ESTs are from latex, 1091 ESTs are from bark and 18 ESTs are from leaves. In addition to gene discovery, EST resources enable the identification of markers such as EST-SSRs and single-nucleotide polymorphism. Since these markers are directly linked to functional genes, they are useful for assessing genetic diversity and mapping phenotypic traits. Feng *et al.*<sup>3</sup> identified 799 SSRs in 10 829 ESTs available in the GenBank database and carried out the genetic diversity assessment of *H. brasiliensis* using 87 EST-SSR markers. The result provided evidence for cross-taxa transferability and indicated moderate polymorphisms of EST-SSR markers in *Hevea* species.<sup>3</sup> However, additional markers are desirable to enable quality research into the genetic basis of commercially relevant traits that can be used in marker-assisted breeding programs.

Genomic and transcriptomic resources for *H. brasiliensis* can greatly benefit from the application of the recent high-throughput sequencing technology, such as the 454 pyrosequencer,<sup>23</sup> which has been instrumental in the development of genetic databases for several economical crops.<sup>24–27</sup> The purpose of the present study is therefore to sequence the transcriptome of the shoot apical tissue, which is a highly dynamic structure, to discover genes, expand the EST database and develop EST-SSR markers that can be used for assessing genetic diversity,

constructing linkage maps and identifying traits of commercial interest.

## 2. Materials and methods

### 2.1. Plant materials

The shoot apical meristem (SAM; 1-cm long from the vegetative shoot apex) of *H. brasiliensis* (clone RRIM600) was collected for RNA extraction from an experimental field at the Rubber Research Institute of Thailand, Ministry of Agriculture and Cooperatives, Thailand. The sample was immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until RNA extraction. For the analysis of SSR markers, leaf samples from 20 clones of *H. brasiliensis*, 2 accessions of *Manihot esculenta*, 3 accessions of *Jatropha curcas* and 1 accession of *Jatropha gossypifolia* (Supplementary data, File S1) were collected and DNA was extracted using a DNeasy Plant Mini Kit (Qiagen). For genetic linkage map construction, 81 samples of an F<sub>1</sub> mapping population were developed from a cross between RRIM600 as a female parent and RRII105 as a male parent. The plants were grown at Chachoengsao Rubber Research Center Office of Agriculture Research and Development, Department of Agriculture, Ministry of Agriculture and Cooperatives, Thailand. DNA samples were extracted using a DNeasy Plant Mini Kit (Qiagen). The concentration of each sample was calculated from the OD measurement using a nanodrop ND1000 (NanoDrop Technologies).

### 2.2. cDNA library preparation and sequencing

Total RNA was extracted using Concert™ Plant RNA Reagent (Invitrogen). Two hundred nanograms of the poly-A mRNA sample was isolated using an Absolutely mRNA Purification Kit (Stratagene) and fragmented in 10× fragmentation buffer (0.1 M ZnCl<sub>2</sub>, 0.1 M Tris-HCl, pH 7.0) at 70°C for 30 s. The reaction was stopped by adding 2 μl of 0.5 M EDTA and 28 μl of 10 mM Tris-HCl, pH 7.5. The mRNA sample was cleaned using Agencourt RNAClean reagent (Beckman Coulter), washed with 200 μl of 70% EtOH, air dried and eluted in 20 μl of 10 mM Tris-HCl, pH 7.5. Fragmented mRNA samples were converted to double-stranded cDNA with the cDNA Synthesis System Kit (Roche Applied Sciences) using random primers and AMV Reverse Transcriptase. A cDNA library for 454 pyrosequencing was prepared according to the October 2008 version of the cDNA Rapid Library Preparation protocol (Roche Applied Sciences). The cDNA library was amplified in emulsion PCR and subject to pyrosequencing on two full picotiter plates of the Genome Sequencer (GS) FLX

Titanium platform using the October 2008 version of the titanium chemistry protocol (Roche Applied Sciences).

### 2.3. Sequence analysis

Poly-A/T(18) and 454-adaptor sequences were trimmed off. Sequence reads with low quality (average quality scores <20), short reads (<100 bp), rRNAs and tRNAs were removed. Sequence assembly was performed using the cDNA option of Newbler 2.5, the *de novo* sequence assembly software. The cDNA option assembles reads into contigs much like a genomic assembly; however, this option allows/expects contigs to have multiple joining contigs representing alternately spliced genes. These contigs (which essentially represent exons) are then assembled into isotigs (representing processed mRNA) and isotigs utilizing the overlapping subsets of contigs are grouped into isogroups (representing genes, isoforms or gene families). When the maximum number of contigs in an isogroup is exceeded, the contigs are output as un-traversed contigs in the isotigs file; all further mention of isotigs will include these un-traversed contigs. Unique sequences were searched for sequence homology against the Uniprot plant protein database ([www.uniprot.org](http://www.uniprot.org)); and reference protein sequences of *Manihot*, *Ricinus*, *Arabidopsis* and *Oryza* ([www.phytozome.net](http://www.phytozome.net)) using the BLASTx program with a cutoff at E-6.<sup>28</sup> The assignment of functionality via gene ontology (GO) was performed using Blast2GO.<sup>29</sup>

The MISA-Micro Satellite identification tool (<http://pgrc.ipk-gatersleben.de/misa/misa.html>) was used to search for SSR from the EST data set. For the searches and comparison of microsatellites, SSRs were defined as being mononucleotide repeats (MNRs)  $\geq 10$  repeats and di- (DNRs), tri- (TNRs), tetra- (TTNRs), penta- and hexanucleotide repeats  $\geq 6$  repeats; criteria for composite SSRs was an interval of bases  $\leq 100$ . For the purpose of marker evaluation, we increased stringency to reduce the number of candidates. We designed primer pairs overlapping DNRs and TNRs  $\geq 8$ , TTNRs  $\geq 7$ , pentanucleotide repeats (and more)  $\geq 6$  or containing complex SSRs  $\geq 30$  nucleotides. In some cases, candidate SSRs that passed the criteria suggested by Feng *et al.*<sup>3</sup> were prioritized based on the presence of motif size polymorphisms in the sequence alignment results.

### 2.4. SSR markers from the previous reports

Primer pairs designed for the amplification of genomic SSR markers from the NCBI database (AY486558.1–AY486910.1) and previously described<sup>8</sup> were used to construct a linkage map together with novel EST-SSR markers.

### 2.5. SSR analysis

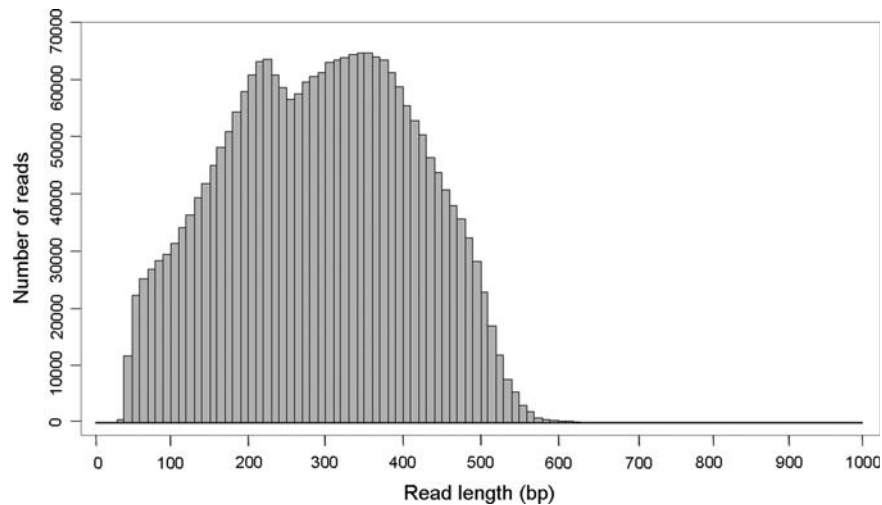
DNA samples were extracted from young leaf tissue using a DNeasy Plant Mini Kit (Qiagen). Primer pairs were designed to amplify SSR regions using PRIMER3.<sup>30</sup> PCR was carried out in a total volume of 10  $\mu$ l containing 2 ng of DNA template, 1  $\times$  *Taq* buffer, 2 mM MgCl<sub>2</sub>, 0.2 mM dNTPs, 1 U *Taq*-DNA polymerase (Fermentas) and 0.5  $\mu$ M each of forward and reverse primers. Amplification was performed in a GeneAmp PCR 9700 thermocycler (Applied Biosystems) programmed as follows: 94°C for 2 min followed by 35 cycles of 94°C for 30 s, 52°C for 30 s, 72°C for 1 min and a final extension step at 72°C for 10 min. Amplified products were separated on 5% denaturing polyacrylamide gels and visualized by silver staining.

### 2.6. Analysis of polymorphic loci

Twenty accessions of *H. brasiliensis*, as listed in Supplementary data, File S1, were used for the polymorphism analysis of SSR markers. Details of primer pairs for amplifiable EST-SSR markers are listed in Supplementary data, File S2. Scored data from polymorphic loci were used to calculate the polymorphism information content (PIC).<sup>31</sup> Observed heterozygosity and expected heterozygosity were calculated using the PowerMarker 3.25 software.<sup>32</sup> The cross-taxa transferability of *H. brasiliensis* SSR loci were evaluated using six other taxa of Euphorbiaceae plants, including two accessions of *M. esculenta*, three accessions of *J. curcas* and one *J. gossypifolia* (Supplementary File S1). The percentage of transferability was calculated for each taxon by dividing the number of successfully amplified SSR loci by the total number of loci analysed. A genetic similarity matrix was prepared for the 20 *H. brasiliensis* genotypes at 47 EST-SSR loci (Supplementary File S3) using the NTSYSpc 2.2 software.<sup>33</sup> UPGMA (un-weighted pair group method with arithmetic mean) cluster analysis was conducted using usizSpc 2.2 software.<sup>33</sup>

### 2.7. Linkage map construction

Eighty-one of *H. brasiliensis* progenies derived from a cross between RRIM600 and RRII105 were used as mapping population. Genomic DNA of individual samples was used to genotype with informative primers and genotypic data were scored as codominant markers under the cross-pollination model, e.g. <abxcd>, <efxeg>, <lmxll>, <nnxnp> and <hkhk> with up to four distinguishable alleles as described by Van Ooijen and Voorrips.<sup>34</sup> The integrated parental genetic linkage map was constructed using the double pseudo-testcross strategy by JoinMap 3.0.<sup>34</sup> The Mendelian segregation ratio of all markers was evaluated using the chi-square test



**Figure 1.** Read length distribution of 454 reads of the *H. brasiliensis* transcriptome.

**Table 1.** Isotig and singleton sequence length distribution

Sequence length (bp)	Number of singleton	Number of isotig
101–500	83 630	2670
501–1000	1296	9553
1001–1500	0	7018
1501–2000	0	4371
2001–2500	0	2271
2501–3000	0	1306
>3000	0	1198
Total	84 926	28 387

( $\chi^2$ ) and distorted markers ( $P \leq 0.1$ ) were excluded. The map was constructed with an LOD score threshold of 3.0 and the mapping parameters were set with a recombination threshold of 0.4, a jump threshold of 5.0 and a minimum LOD score threshold of 1.0. The map distance between two markers in centiMorgan (cM) was calculated using Kosambi's mapping function.<sup>35</sup> The linkage map was drawn using MapChart 2.2.<sup>36</sup>

### 3. Results and discussion

#### 3.1. Transcriptome sequencing of rubber tree

A total of 2 311 497 filtered sequence reads were generated from the vegetative shoot apical tissue with an average read length of 294 bases totalling 676.5 Mb. The majority of the reads were in the range 200–400 bp (Fig. 1). All reads were deposited in DDBJ Read Archive (ID = DRA000170). There were 191 369 (8.27%) reads with homology to plant tRNAs and rRNAs and 139 838 (6.04%) reads shorter than 100 bp which were removed before

sequence assembly. Raw sequencing reads were assembled by Newbler 2.5, the *de novo* sequence assembly software,<sup>23</sup> currently the most robust software for 454 transcriptome assembly.<sup>37</sup> A total of 28 387 isotigs from 19 152 isogroups and 84 926 singletons were obtained from the assembly. An isogroup theoretically represents a single gene; however, genes with high sequence similarity may be grouped together and therefore isogroups may also represent gene isoforms or gene families. The isotigs had an average length of 1326 bp and the majority of isotigs were between 500 and 1000 bp (Table 1). The largest isotig (isotig08423) was 9041 bp, which showed sequence similarity to the *M. esculenta* chloroplast polycistronic transcript *psaA-psaB* (YP\_001718437.1;  $E$ -value = 0). The most highly represented isotig (isotig00002) was assembled from 22 641 reads and the highest blast match was the *Hevea* hydroxynitrile lyase Chain A (AAC49184.1;  $E$ -value =  $1E-59$ ), reflecting the high level of cyanogenesis activity in the tissue sample. It is well recognized that all living tissues of *H. brasiliensis*, including seeds, are strongly cyanogenic accumulating high quantities of cyanogenic precursors such as linamarin and lotaustralin.<sup>38</sup> The average GC content of the *H. brasiliensis* transcriptome generated in this study was 42.16%, which is similar to the GC content of *H. brasiliensis* sequences in the GenBank EST database (42.18%). The GC content of *H. brasiliensis* coding sequences is slightly lower than the average GC content of *Arabidopsis* coding sequences (44.5%) and rice coding sequences (51.5%);<sup>39</sup> but much higher than *Arabidopsis* intergenic regions (32.9%; <http://gi.kuicr.kyoto-u.ac.jp>). All unique transcripts were annotated and characterized according to GO using BLAST2GO<sup>29</sup> and the result is available at <http://www4a.biotec.or.th/rubber>.

In total, 61 625 isotigs and singletons were assigned one or more GO terms. Under the biological process domain, 71 071 assignments were made, with a large proportion of assignments falling into the categories metabolic process (31.07%) and cellular process (29.94%). A total of 60 927 assignments were made to the molecular function domain, with the majority falling into the categories binding activity (46.52%) and catalytic activity (39.82%). This distribution of GO terms is similar to the previous study in the pea SAM transcriptome sequencing.<sup>40</sup> The large number of annotated sequences shows that ESTs generated by high-throughput sequencing are more likely to represent mRNA than ESTs generated by lower-throughput methods, such as previous studies performed on the latex transcriptome.<sup>20–22</sup> The reason for this is that high-throughput sequencing generates sequence data supported by multiple reads, often representing complete mRNAs. The majority of EST sequences currently in GenBank from Chow *et al.*<sup>20</sup> are mostly un-annotated singletons with only 904 sequences (26%) that have GO terms assigned. Also, the diversity of the latex transcriptome is limited as Han *et al.*<sup>22</sup> pointed out that the genes expressed in latex are mainly associated with rubber biosynthesis pathways, defence mechanisms and allergenic proteins. A larger diversity of genes was considered more likely to be found in the vegetative shoot apex tissue than in latex, and that is what has been found here.

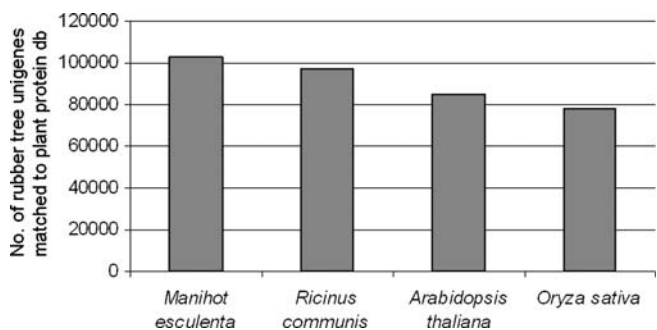
A unique characteristic of the shoot apical tissue is the maintenance of the SAM via intercellular communication involving a complex signalling network such as epigenetic control,<sup>41,42</sup> transcriptional gene regulation<sup>43–45</sup> and hormonal regulation.<sup>46</sup> Key regulatory genes controlling SAM maintenance, such as *WUSCHEL* (*WUS*) and *SHOOT MERISTEMLESS* (*STM*) genes, were identified as putative full-length cDNAs in this data set. Recent studies revealed that *WUS* plays the key role in SAM maintenance through the regulatory loop of *WUS*–*CLAVATA* (*CLV*) feedback<sup>41,47,48</sup> and interacts with *STM* via phytohormone signalling pathways.<sup>49</sup> Furthermore, the regulation of *WUS* expression is also controlled by auxin signalling, chromatin remodelling and positive and negative transcriptional regulators.<sup>50</sup> Transcripts for a positive transcriptional regulator of *WUS*, such as *APETELA2*,<sup>45</sup> *SPLAYED* and *BARD1*, which bind to the *WUS* promoter sequence<sup>43</sup> were found in this data set. Whereas, *WUS* negative transcriptional regulators which are required for development of floral organs, such as *ULTRAPETALA1*<sup>51</sup> and *HANABA TARANU*<sup>52</sup> transcripts, were not detected.

Also identified were genes from the *KNOX* (*Knotted-like homeobox*) family such as *STM* and *KNOTTED-LIKE FROM ARABIDOPSIS THALIANA* (*KNAT*) genes which are

essential in maintaining the balance between organ primordia growth and stem cell maintenance in the SAM.<sup>53</sup> *KNOX* transcription factors have roles in suppressing gibberellins (GAs) in the SAM by inhibiting *GA-20 oxidase*, which is required for GA biosynthesis, and promoting *GA-2 oxidase*, which inactivates the active GA.<sup>54</sup> Thus, *KNOX* proteins prevent the accumulation of GA in the central zone of the SAM, consequently preventing the differentiation of stem cells. *KNOX* proteins also promote cytokinin activity in the SAM central zone stimulating division and maintenance of undifferentiated stem cells.<sup>46</sup> Flanking the SAM, high levels of auxin and GA activities have roles in development and growth of lateral organ primordia.<sup>46</sup> The tissue sample in this study contained both undifferentiated meristem and differentiated organ primordia; therefore, it was likely to identify transcripts of genes involving in many phytohormone biosynthesis and signalling pathways.

### 3.2. Similarity of rubber tree ESTs to other plant proteins

To investigate the efficiency of gene discovery in the *H. brasiliensis* transcriptome, the isotigs and singletons were searched for homology using BLAST<sup>28</sup> against other plant reference sequences such as *M. esculenta* (Euphorbiaceae, Rosids), *Ricinus communis* (Euphorbiaceae, Rosids), *Arabidopsis thaliana* (Brassicaceae, Rosids) and *Oryza sativa* (Poaceae, Liliopsida). The majority of *H. brasiliensis* unigenes matched against proteins from *Manihot* (102 936 or 48.1%), followed by *Ricinus* (97 089 or 45.4%), *Arabidopsis* (84 643 or 39.5%), then *Oryza* (77 805 or 36.3%) as shown in Fig. 2. *Hevea*, *Manihot* and *Ricinus* are phylogenetically related and grouped together in the Euphorbiaceae family; therefore, it was expected that a large number of *H. brasiliensis* isotigs and singletons would match proteins from *Manihot* and *Ricinus*. These observations agree with previous studies on cross transferability of EST markers which demonstrated a high level of genome



**Figure 2.** Homology results of *H. brasiliensis* isotigs and singletons that matched to proteins in the plant reference databases (*Manihot esculenta*, *Ricinus communis*, *Arabidopsis thaliana* and *Oryza sativa*) using BLASTx.

**Table 2.** Distribution of identified SSRs using the MISA software according to SSR motif types and repeat numbers

Repeats	Number of repeat units											Total
	6	7	8	9	10	11	12	13	14	15	>15	
MNR	N/A	N/A	N/A	N/A	2999	1915	1435	1127	813	688	4393	12 682
DNR	991	535	414	307	237	200	123	132	110	94	315	3458
TNR	700	292	192	114	103	45	42	31	14	15	45	1593
TTNR	33	10	5	2	1	0	0	0	0	0	0	51
≥PTNR	24	5	4	0	2	0	0	0	0	0	0	35

MNR, mononucleotide repeat; DNR, dinucleotide repeat; TNR, trinucleotide repeat; TTNR, tetranucleotide repeat; PTNR, pentanucleotide repeat.

conservation among plants in Euphorbiaceae, especially *Hevea* and *Manihot*.<sup>3,55,56</sup> However, it should also be noted that sequence homology analysis by BLAST can be biased by the number and quality of query database and the reference databases.

### 3.3. EST-SSR: distribution and frequencies

A total of 17 819 SSRs were identified in the isotigs and singletons from the *H. brasiliensis* transcriptome (Table 2). This represents an average frequency of one EST-SSR in every 3383 bp, which is a lower frequency than the previous report (1 SSR per 2.25 kb) by Feng *et al.*<sup>3</sup> Among plant species, SSR frequencies range from 1 per 1.5 kb in coffee<sup>57</sup> to 1 per 67 kb in mungbean.<sup>25</sup> The distribution and frequencies of EST-SSRs significantly vary between different studies due to SSR search criteria, the size of the EST database and software tools,<sup>58</sup> so the MISA software was used with the same criteria as the previous study on *H. brasiliensis* EST-SSRs<sup>3</sup> to allow for direct comparability. From 17 819 SSRs identified, there were 12 682 MNRs, 3458 DNRs, 1593 TNRs, 51 TTNRs and 35 SSRs with pentanucleotide repeats or more. It should be noted that the number of MNRs may not be accurate due to the limitations of the 454 technology in reading long homopolymer sequences. MNRs were described here for the purpose of comparison with previous studies but they were not used for polymorphism analysis. The most common type of DNR was AG/CT which accounted for 64.43% of the repeats, followed by AT/TA (26.75%), AC/GT (8.48%) and GC/CG (0.3%). The most common type of TNR was AAG/CTT (32.53%), followed by AAT/ATT (19.17%) and ACC/GGT (18.33%). The least frequent DNR and TNR motif types were GC-rich motifs, which were found at only 33 loci. SSRs with GC-rich motif repeats are rare in many plants, such as *H. brasiliensis*,<sup>3</sup> rice, corn, soybean,<sup>59</sup> wheat,<sup>60</sup> *Arabidopsis*, apricot, peach<sup>61</sup> and coffee.<sup>57</sup>

### 3.4. Polymorphism test in EST markers

Four hundred and thirty EST-SSRs were selected from EST-SSR present in the data set to give a range of repeat units and motif types, and primers were designed to

amplify them. Based on the sequence homology search, 16 of 430 (3.7%) SSR primer pairs were mapped to the EST reads with GenBank accession numbers reported for the development of EST-SSR markers by Feng *et al.*<sup>3</sup> The primary PCR screening in three clones of *H. brasiliensis* (RRIM600, RRIC110 and BPM24) showed that 323 primer pairs (75.11%) were amplifiable (primer sequences were listed in Supplementary File S2). We selected 47 primer pairs that flanked long SSR motifs (criteria mentioned above), gave clear PCR bands in the primary amplification screening and represented in all classes of nucleotide repeats for polymorphism evaluation among 20 different *H. brasiliensis* clones (Table 3). The number of alleles observed at each locus ranged from two to six, with an average of 3.85. Although *H. brasiliensis* is believed to be a stabilized amphidiploid, only five SSRs (EHB61, EHB100, EHB109, EHB115 and EHB144) gave more than two alleles supporting the report by Lespinasse *et al.*<sup>14</sup> that *H. brasiliensis* behaves as a diploid. The value of expected heterozygosity varied from 0.1349 to 0.7494 with an average of 0.5594, while direct count heterozygosity ranged from 0.1429 to 0.8095 with an average of 0.5076. PIC values ranged from 0.13 to 0.71 with an average of 0.50, which is higher than the previous report (PIC = 0.38) by Feng *et al.*<sup>3</sup> The higher PIC value in this study was probably due to a larger sample size of *H. brasiliensis* clones tested ( $n = 20$ ) than the previous study ( $n = 12$ ). The average number of alleles and PIC values of EST-SSR markers are lower than those of genomic SSR markers,<sup>8</sup> as expected for functional sequences.

The 47 polymorphic EST-SSR markers were used to evaluate the genetic relatedness among 20 different *H. brasiliensis* clones which were classified into two groups at the level of genetic similarity 0.44 (Fig. 3). This generally corresponded well to the clone pedigree (Supplementary File S1). The first group (Group I) contains clones: PB260, PB310, RRIM605, PB217, PB5/51 and PB235. The majority of these clones have PB5/51 or PB49 clones as one of the parental lines. Group II contains a mixture of primary clones and cultivated clones from various rubber

**Table 3.** Characteristics of the 47 primer pairs targeting polymorphic microsatellite loci analysed in 20 different clones of *H. brasiliensis*.

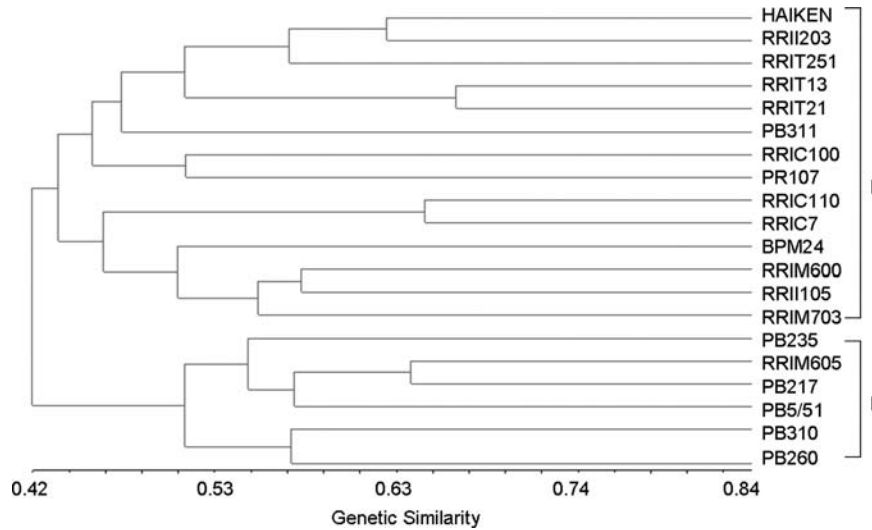
Marker	Forward sequence	Reverse sequence	Repeat motif and count	PCR product size (bp)	Number of alleles	Expected heterozygosity	Observed heterozygosity	PIC
EHB012	AAGATTGAACTAGGGTTGAACTGG	CCAAATGTTTCATTTAATTGTGGA	(CAA) 8, (TAA) 6	250–300	5	0.6746	0.3333	0.6186
EHB013	AAGCAAGGAAGAGGAAGGGA	CAAGAAGTTGCCCATTTTCA	(TTTTA) 8	225–255	3	0.4206	0.5238	0.3824
EHB025	ACCGTCCACCATAACCACAT	AAAGGCCATGCCTACATTTG	(CT) 10, (CA) 12	245–250	3	0.4989	0.2857	0.447
EHB033	ATACCCAGACCTATGTGGCG	AATGGGCTCGGAGATTCTTT	(TC) 16	225–240	3	0.2166	0.1429	0.2051
EHB034	ATAGCCGACCCCAAATCTTT	GGACAGCAAGACATGAAGAGTG	(AGTTG) 6	148–155	5	0.7154	0.7619	0.6783
EHB061	CCACAGCAACACCACCATTA	TCATCCATCCAATGAAGCAA	(CAGCAA) 6	150–200	4	0.568	0.5238	0.4876
EHB063	CCAGCTGGTTGTGTTAGAAGG	GAGCTCATCTTCCAGGGACTT	(CTT) 12	160–270	4	0.6565	0.4286	0.605
EHB065	CCAGTGAGCAcAGGCATAAT	TGGAGAGTGCagATGAATGC	(AAT) 10	300–350	3	0.5408	0.4286	0.4648
EHB069	CCCATTTCTACAACACACTTTTC	TGCTAGGGCCTTGTGCGATAC	(AAAAAT) 5	100–110	4	0.5884	0.2381	0.5459
EHB070	CCCCACATGCGATTTAAGTT	TGGGCTGTGTTGTGCTATTC	(AAG) 10	230–250	5	0.7494	0.4762	0.7051
EHB079	CCTATCCTTCTGCTCGTTCCG	TTTCCACAGAAGGGAAGGTG	(ATC) 11	150–165	6	0.6837	0.6667	0.64
EHB081	CCTCTTGCTCTGAAAGCCAC	AACCAACCAACTGGGATCAA	(CACCGG) 5	235–245	4	0.6213	0.7619	0.5455
EHB085	CGATTAGGTACGTGATCCCA	AAGTTGTTGAGGAATGATCAGGA	(TCATGC) 5	110–120	6	0.7472	0.5238	0.7061
EHB086	CGCATCCCAACAAGCTAAAT	CAGAAAGCAATCACAAACACACA	(TC) 10, (GTTT) 7	245–250	3	0.1349	0.1429	0.13
EHB087	CGGAGCTAAGTTCGAGTCTCT	CTGGAACCGTATTTCCAGGT	(ATT) 13	180–200	4	0.6202	0.5714	0.5693
EHB088	CGGAGGCTCCAATTAGACAA	AAGATGGTCTGTGATCGTGCT	(TGAGT) 7	160–250	4	0.2959	0.2381	0.2825
EHB100	CTGCCGATGTGCTCTTCATA	AAATGAGGTTGGTCGTCGTC	(GCTTCT) 6, (CTT) 8	240–270	3	0.5612	0.7143	0.465
EHB109	GAAaGCTAACGGTGGACTCG	ACGAATCGGACTTTGGTGT	(ATC) 10	252–254	6	0.6984	0.619	0.6613
EHB110	GAATCCTGCCAGTGGGACTA	GAGAAGGTGCCGAAGAAGAA	(TCT) 10	200–230	2	0.4592	0.619	0.3538
EHB112	GACATTACCATCCCCTCCC	TCAGTTACCAGCAGCCATTG	(ATA) 10	180–188	2	0.4444	0.4762	0.3457
EHB113	GAGGCACTTGAGCTCCAAAC	CGAATCCGGAATTTTCTTCA	(GCT) 10	170–175	4	0.6859	0.6667	0.6308
EHB115	GATCAAGCTGAAAAGCACCC	GAGtcaAGAATCCACGAGC	(CTTT) 8	225–250	3	0.6315	0.381	0.5584
EHB118	GCAAATAATGGCGAGCTGTT	TGGTTGATGGCAGAACAAAG	(CAA) 11	160–178	4	0.5862	0.381	0.542
EHB120	GCAACCGTTCTCTTCACAT	TCCTCCGTCACCAAAGACTC	(AGA) 10	134–140	2	0.4082	0.4762	0.3249
EHB122	GCATGATTGGGAAACCAGAT	GAGTCAACCTGGAAATTAGCG	(TCT) 13	230–250	4	0.6134	0.5238	0.5486
EHB125	GCTTCCAGTCCACAAAGCAG	TCATCAGACAGAAATAGTAATAGCCG	(TC) 9, (AC) 9	152–154	4	0.5181	0.381	0.442
EHB126	GCTTCCTCTTTCCGTGTTTG	ACCAATTGAAAGGCACTGCT	(ATTTC) 6	115–140	4	0.5941	0.619	0.5396
EHB127	GGAATTTCTGCTGGCACTGT	TCGTGACCCAACAGAATAAAGA	(ATTA) 8	190–220	4	0.6723	0.6667	0.61
EHB133	GGCCATCACTCAACATCCTT	CTCACCTTTTGAAAGCGAA	(CTT) 10	210–225	3	0.5159	0.5238	0.4233
EHB135	GGGGACGCTTCATGGTAGTA	ACTTGCAATTGGTGGCACA	(TTA) 12	114–125	2	0.3628	0.4762	0.297

Continued

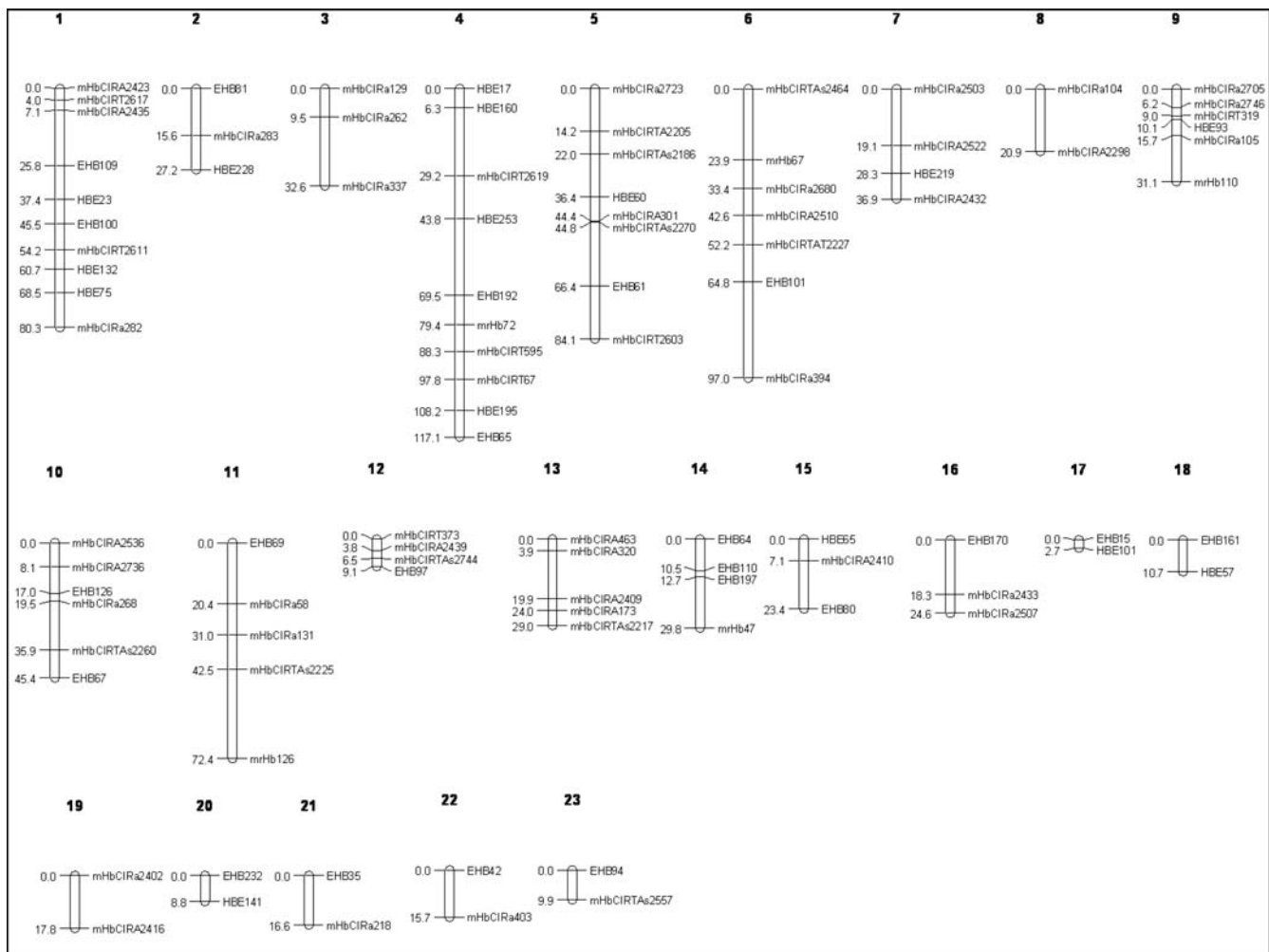
**Table 3.** Continued

Marker	Forward sequence	Reverse sequence	Repeat motif and count	PCR product size (bp)	Number of alleles	Expected heterozygosity	Observed heterozygosity	PIC
EHB136	GGGTATGGATGTGGTGAAGG	ATGGTTTGGTTTCATCCCA	(GTG) 10	253–255	4	0.6224	0.619	0.5605
EHB140	GGTAGAGGTTTGGAGGGGAG	TGATGGCAGCTATGCTGAAG	(TGAGAC) 5	140–153	4	0.602	0.5238	0.5528
EHB143	GGTGGTAAAAGTGGCAATGG	CTCCATTTTGTACCACCACT	(TGG) 8	175–220	3	0.5442	0.8095	0.4393
EHB144	GGTCTTTGCCGATCTACA	CTgGGCATGAGAGATTTGT	(CAG) 6, (AAG) 6	160–182	3	0.2098	0.2381	0.1878
EHB148	GGTTTTCAAAATCTTTTCTATACATCC	TGCAGAAGCATCAACAAACC	(ATT) 16	160–180	6	0.7392	0.5238	0.7014
EHB151	GTCCGGTGAATGAGATGCT	AGGCGGAAACAGACTCTGAA	(ATT) 15	225–245	3	0.3571	0.381	0.3254
EHB157	GTTGGCCTGGTCAATCTCAT	GATTAATTCAGTGGTGGCGG	(CCCAAT) 5	200–215	2	0.2778	0.3333	0.2392
EHB159	TACCAAGCATGTTGCCATA	TCTCAGAAACAAGGGTTGGG	(CA) 17	185–210	5	0.7063	0.6667	0.6497
EHB160	TAGAAGCTGCCACAATGC	TTGACGCCAAATGTTTATGC	(AAT) 13	210–235	6	0.6338	0.7143	0.5817
EHB161	TAGGATGAGTTTTGGCTGC	TGGCTCCTTGAACTGCTCT	(CATCGT) 5	250–270	3	0.4751	0.3333	0.3826
EHB168	TCAAGCGCATCACAGGTATC	TGGTCACCGAACAACAACAT	(TCA) 10	118–120	3	0.4524	0.3333	0.3845
EHB169	TCACTTTTACAACCCACCA	GGCAAACCAGGAAATCAACA	(TCT) 11	200–225	4	0.7336	0.619	0.6845
EHB177	TCGCTTTCTCCATATAGAGTTTCA	CAGCAAGAAATCCCTCAACC	(GAA) 7, (TTC) 8	209–212	6	0.7188	0.8095	0.6674
EHB178	TCGTGACCCAACAGAATAAAGA	GGAAATCTGCTGGCACTGT	(ATTA) 8	190–215	4	0.7302	0.619	0.68
EHB190	TGATCCCAAGAAGTACTTGC	TAGGAATGGTACCGACCCAC	(TCATGC) 7	130–140	4	0.6655	0.5714	0.6043
EHB197	TGGAAGTGAGAAATGAcGGTTT	CGAAGACTTGTGTCAGCAGC	(GAT) 7, (GAG) 7	250–258	4	0.6134	0.381	0.5486
EHB198	TGGCATTCCCACTAATTCAA	CGGTGGAAATGGTAAGCTGT	(AAACCAG) 5	194–200	5	0.7245	0.8095	0.6754
Mean					3.8511	0.5594	0.5076	0.5026





**Figure 3.** Similarity relationships of 20 different *H. brasiliensis* clones based on 47 EST-SSR loci.



**Figure 4.** The genetic linkage map of rubber tree F1 population (*H. brasiliensis*) developed from EST-SSR and SSR markers. The map is composed of 97 loci covering 842.9 cM on 23 linkage groups.

research institutes. In one branch of Group II, RR11 203, RRIT21 and RRIC100 were clustered together and shared PB86 as one of the parents. Clone RRIT21, which is a descendant of PB86 × RRIT13, was closely grouped together with RRIT13 with a genetic similarity of 0.66. Another branch of Group II contained descendants of the Tjir 1 clone (RRIM600, RR11105 and RRIM703). Although RRIM605 has Tjir 1 as a female parent, RRIM605 was classified in Group 1 because it shares the same male parent (PB49) with PB260. Since these markers are able to reproduce the relationship that was already known from pedigree information, they can be used to provide reliable genotypic information for clonal identification and the selection of parents in breeding programs. The minimal set of five highly informative microsatellites (EHB85, EHB109, EHB169, EHB177 and EHB178) was able to distinguish each of the *H. brasiliensis* clones included in this study.

Genotyping of the 93 EST-SSR makers that were amplifiable in *H. brasiliensis* was performed across genera with two accessions of *M. esculenta*, three accessions of *J. curcas* and one accession of *J. gossypifolia*, all belonging to the same family as *H. brasiliensis*, Euphorbiaceae. The results showed that 47 of 93 primer pairs (40%) gave successful amplifications in *Manihot* species. Fourteen primer pairs (15%) and nine primer pairs (10%) were successfully amplified in *J. curcas* and *J. gossypifolia*, respectively. The higher rate of cross-transferability in *Manihot* species compared with that in *Jatropha* species suggests a closer relationship of *Hevea* with *Manihot* than with *Jatropha*. Six primer pairs, EHB61, EHB63, EHB85, EHB115, EHB116 and EHB156, were able to amplify unique products from all plant taxa tested.

### 3.5. Genetic linkage map

From 323 novel EST-SSR primer pairs amplifiable in *H. brasiliensis*, 59 primer pairs were polymorphic between RRIM600 and RR11105 parental clones. All of these polymorphic primers were used to genotype with 81 individual F<sub>1</sub> samples together with 98 published SSR markers. Genotypic data were scored and subjected to linkage analysis. Of 157 polymorphic markers, 124 markers (78.9%) revealed the expected Mendelian segregation ratio and used to construct the genetic linkage map. The F<sub>1</sub> map consisted of 97 loci distributed on 23 linkage groups. Of these, 37 loci were novel EST-SSRs. The total map distance covered 842.9 cM with a mean interval of 11.9 cM and the average loci per linkage group were approximately four loci (Fig. 4). The number of linkage groups exceed the expected haploid number of linkage groups (18 linkage groups), suggesting that more markers are required to fill the gap between adjacent

**Table 4.** Comparison of common SSR marker positions between linkage maps

Common SSR loci	Position in Le Guen <i>et al.</i> <sup>16</sup>	Position in this study
mHbCIRTAs2557	LG1 (96.7 cM)	LG23 (9.9 cM)
mHbCIRTAs2510	LG2 (48.2 cM)	LG6 (42.6 cM)
mHbCIRTAs2186	LG5 (19 cM)	LG5 (22 cM)
mHbCIRTAs2603	LG5 (97.6 cM)	LG5 (84.1 cM)
mHbCIRT67	LG8 (106.6 cM)	LG4 (97.8 cM)
mHbCIRTAs2260	LG11 (15.6 cM)	LG10 (35.9 cM)
mHbCIRA268	LG11 (24.7 cM)	LG10 (19.5 cM)
mHbCIRA2736	LG11 (34.1 cM)	LG10 (8.1 cM)
mHbCIRA2536	LG11 (38.3 cM)	LG10 (0 cM)
mHbCIRA282	LG14 (19 cM)	LG1 (80.3 cM)
mHbCIRA2435	LG14 (53.9 cM)	LG1 (7.1 cM)
mHbCIRA2423	LG14 (58 cM)	LG1 (0 cM)
mHbCIRA2298	LG9 (9.5 cM)	LG8 (20.9 cM)
mHbCIRA104	LG9 (54 cM)	LG8 (0 cM)
mHbCIRA2432	LG9 (119.9 cM)	LG7 (36.9 cM)
mHbCIRTAs2225	LG16 (0 cM)	LG11 (42.5 cM)
mHbCIRA131	LG16 (6.3 cM)	LG11 (31 cM)
mHbCIRA2410	LG16 (101.1 cM)	LG15 (7.1 cM)
mHbCIRA463	LG18 (44 cM)	LG13 (0 cM)
mHbCIRA320	LG18 (46.4 cM)	LG13 (3.9 cM)
mHbCIRA2409	LG18 (54.5 cM)	LG13 (19.9 cM)
mHbCIRTAs2217	LG18 (64. cM)	LG13 (29 cM)
mHbCIRT373	LG18 (94 cM)	LG12 (0 cM)
mHbCIRA2439	LG18 (100.6 cM)	LG12 (3.8 cM)
mHbCIRTAs2744	LG18 (109.5 cM)	LG12 (6.5 cM)

markers. Moreover, the comparison between the F<sub>1</sub> map in this study and the map constructed by Le Guen *et al.*<sup>16</sup> using different parents revealed 25 common markers on nine homologous linkage groups between both maps (Table 4). A total of seven marker intervals of 20 markers showed co-linearity between homologous linkage groups. Some linkage groups in this study (LG7-LG8, LG11-LG15 and LG12-LG13) could be joined together based on the linked markers in Le Guen *et al.*<sup>16</sup> The common and collinear markers indicated the reliability between different maps.<sup>62</sup> Therefore, the F<sub>1</sub> map of this study is appropriate for further studies in marker-assisted selection.

**Supplementary data:** Supplementary data are available at [www.dnaresearch.oxfordjournals.org](http://www.dnaresearch.oxfordjournals.org).

### Funding

We acknowledge the funding support by the National Center for Genetic Engineering and

Biotechnology (Thailand). The database server is supported from the national infrastructure project by National Science and Technology Development Agency (NSTDA, Thailand).

## References

- Ong, S.H. 1975, Chromosome morphology at the pachytene stage in *Hevea brasiliensis*: a preliminary report, *Proc. Int. Rubb. Conf.*, **2**, 3–12.
- Lespinasse, D., Rodier-Goud, M., Grivet, L., Leconte, A., Legnate, H. and Seguin, M. 2000, A saturated genetic linkage map of rubber tree (*Hevea spp.*) based on RFLP, AFLP, microsatellite and isozyme markers, *Theor. Appl. Genet.*, **100**, 127–38.
- Feng, S.P., Li, W.G., Huang, H.S., Wang, J.Y. and Wu, Y.T. 2009, Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*), *Mol. Breed.*, **23**, 85–97.
- Besse, P., Seguin, M., Lebrun, P., Chevallier, M.H., Nicolas, D. and Lanaud, C. 1994, Genetic diversity among wild and cultivated populations of *Hevea brasiliensis* assessed by nuclear RFLP analysis, *Theor. Appl. Genet.*, **88**, 199–207.
- Seguin, M., Besse, P., Lespinasse, D., Lebrun, P., Rodier-Goud, M. and Nicolas, D. 1996, *Hevea* molecular genetics, *Plant. Res. Dev.*, **3**, 77–88.
- Le Guen, V., Doare, F., Weber, C. and Seguin, M. 2009, Genetic structure of Amazonian populations of *Hevea brasiliensis* is shaped by hydrographical network and isolation by distance, *Tree Genet. Genomes*, **5**, 673–83.
- Lekawipat, N., Teerawatanasuk, K., Rodier-Goud, M., et al. 2003, Genetic diversity analysis of wild germplasm and cultivated clones of *Hevea brasiliensis* Muell. Arg. by using microsatellite markers, *J. Rubber Res.*, **6**, 36–47.
- Le Guen, V., Gay, C., Xiong, T.C., Souza, L.M., Rodier-Goud, M. and Seguin, M. 2010, Development and characterization of 296 new polymorphic microsatellite markers for rubber tree (*Hevea brasiliensis*), *Plant Breed.*, doi:10.1111/j.1439-0523.2010.01774.x.
- Chevallier, M.H. 1988, Genetic variability of *Hevea brasiliensis* germplasm using isozyme markers, *J. Nat. Rubber Res.*, **3**, 42–53.
- Besse, P., Lebrun, P., Seguin, M. and Lanaud, C. 1993, DNA fingerprints in *Hevea brasiliensis* (rubber tree) using human minisatellite probes, *Heredity*, **70**, 237–44.
- Roy, C.B., Nazeer, M.A. and Saha, T. 2004, Identification of simple sequence repeats in rubber (*Hevea brasiliensis*), *Curr. Sci.*, **87**, 807–11.
- Yu, F., Wang, B., Feng, S., Wang, J., Li, W. and Wu, Y. 2010, Development, characterization, and cross-species/genera transferability of SSR markers for rubber tree (*Hevea brasiliensis*), *Plant Cell Rep.*, doi:10.1007/s00299-010-0908-7.
- Le Guen, V., Garcia, D., Mattos, C.R., Doare, F., Lespinasse, D. and Seguin, M. 2007, Bypassing of a polygenic *Microcyclus ulei* resistance in rubber tree, analyzed by QTL detection, *New Phytol.*, **173**, 335–45.
- Lespinasse, D., Grivet, L., Troispoux, V., Rodier-Goud, M., Pinard, F. and Seguin, M. 2000, Identification of QTLs involved in the resistance to South American leaf blight (*Microcyclus ulei*) in the rubber tree, *Theor. Appl. Genet.*, **100**, 975–84.
- Le Guen, V., Lespinasse, D., Oliver, G., Rodier-Goud, M., Pinard, F. and Seguin, M. 2003, Molecular mapping of genes conferring field resistance to South American Leaf Blight (*Microcyclus ulei*) in rubber tree, *Theor. Appl. Genet.*, **108**, 160–7.
- Le Guen, V., Garcia, D., Doare, F., et al. 2011, A rubber tree's durable resistance to *Microcyclus ulei* is conferred by a qualitative gene and a major quantitative resistance factor, *Tree Genet. Genomes*, doi:10.1007/s11295-11011-11038-11297.
- Emrich, S.J., Barbazuk, W.B., Li, L. and Schnable, P.S. 2007, Gene discovery and annotation using LCM-454 transcriptome sequencing, *Genome Res.*, **17**, 69–73.
- Vera, J.C., Wheat, C.W., Fescemyer, H.W., et al. 2008, Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing, *Mol. Ecol.*, **17**, 1636–47.
- Kunkeaw, S., Yoocha, T., Sraphet, S., et al. 2010, Construction of a genetic linkage map using simple sequence repeat markers from expressed sequence tags for cassava (*Manihot esculenta* Crantz), *Mol. Breed.*, doi:10.1007/s11032-11010-19414-11034.
- Chow, K.S., Wan, K.L., Isa, M.N., et al. 2007, Insights into rubber biosynthesis from transcriptome analysis of *Hevea brasiliensis* latex, *J. Exp. Bot.*, **58**, 2429–40.
- Ko, J.H., Chow, K.S. and Han, K.H. 2003, Transcriptome analysis reveals novel features of the molecular events occurring in the laticifers of *Hevea brasiliensis* (para rubber tree), *Plant Mol. Biol.*, **53**, 479–92.
- Han, K.H., Shin, D.H., Yang, J., Kim, I.J., Oh, S.K. and Chow, K.S. 2000, Genes expressed in the latex of *Hevea brasiliensis*, *Tree Physiol.*, **20**, 503–10.
- Margulies, M., Egholm, M., Altman, W.E., et al. 2005, Genome sequencing in microfabricated high-density picolitre reactors, *Nature*, **437**, 376–80.
- Novaes, E., Drost, D.R., Farmerie, W.G., et al. 2008, High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome, *BMC Genomics*, **9**, 312.
- Tangphatsornruang, S., Somta, P., Uthapaisanwong, P., et al. 2009, Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.)Wilczek), *BMC Plant Biol.*, **9**, 137.
- Parchman, T.L., Geist, K.S., Grahnen, J.A., Benkman, C.W. and Buerkle, C.A. 2010, Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery, *BMC Genomics*, **11**, 180.
- Tangphatsornruang, S., Sangsrakru, D., Chanprasert, J., et al. 2010, The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships, *DNA Res.*, **17**, 11–22.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990, Basic local alignment search tool, *J. Mol. Biol.*, **215**, 403–10.

29. Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. and Robles, M. 2005, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, *Bioinformatics*, **21**, 3674–6.
30. Rozen, S. and Skaletsky, H. 2000, Primer3 on the WWW for general users and for biologist programmers, *Methods Mol. Biol.*, **132**, 365–86.
31. Botstein, D., White, R.L., Skalnick, M.H. and Davies, R.W. 1980, Construction of a genetic linkage map in man using restriction fragment length polymorphism, *Am. J. Hum. Genet.*, **32**, 314–31.
32. Liu, K. and Muse, S.V. 2005, PowerMarker: an integrated analysis environment for genetic marker analysis, *Bioinformatics*, **21**, 2128–9.
33. Rohlf, F.J. 2005. *NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System Version 2.2*. Exeter Publishing Ltd: New York.
34. Van Ooijen, J.W. and Voorrips, R.E. 2001, *JoinMap 3.0, Software for the calculation of genetic linkage maps*, Plant Research International, Wageningen, The Netherlands.
35. Kosambi, D.D. 1944, The estimation of map distance from recombination values, *Ann. Eugen. Lond.*, **12**, 172–5.
36. Voorrips, R.E. 2002, MapChart: software for the graphical presentation of linkage maps and QTLs, *J. Hered.*, **93**, 77–8.
37. Kumar, S. and Blaxter, M.L. 2010, Comparing de novo assemblers for 454 transcriptome data, *BMC Genomics*, **11**, 571.
38. Lieberei, R. 2007, South American leaf blight of the rubber tree (*Hevea* spp.): new steps in plant domestication using physiological features and molecular markers, *Ann. Bot.*, **100**, 1125–42.
39. Wang, H.C. and Hickey, D.A. 2007, Rapid divergence of codon usage patterns within the rice genome, *BMC Evol. Biol.*, **7**(Suppl. 1), S6.
40. Wong, C.E., Bhalla, P.L., Ottenhof, H. and Singh, M.B. 2008, Transcriptional profiling of the pea shoot apical meristem reveals processes underlying its function and maintenance, *BMC Plant Biol.*, **8**, 73.
41. Busch, W., Miotk, A., Ariel, F.D., et al. 2010, Transcriptional control of a plant stem cell niche, *Dev. Cell*, **18**, 849–61.
42. Shen, W.H. and Xu, L. 2009, Chromatin remodeling in stem cell maintenance in *Arabidopsis thaliana*, *Mol. Plant*, **2**, 600–9.
43. Baurle, I. and Laux, T. 2005, Regulation of WUSCHEL transcription in the stem cell niche of the *Arabidopsis* shoot meristem, *Plant Cell*, **17**, 2271–80.
44. Kwon, C.S., Chen, C. and Wagner, D. 2005, WUSCHEL is a primary target for transcriptional regulation by SPLAYED in dynamic control of stem cell fate in *Arabidopsis*, *Genes Dev.*, **19**, 992–1003.
45. Wurschum, T., Gross-Hardt, R. and Laux, T. 2006, APETALA2 regulates the stem cell niche in the *Arabidopsis* shoot meristem, *Plant Cell*, **18**, 295–307.
46. Veit, B. 2009, Hormone mediated regulation of the shoot apical meristem, *Plant Mol. Biol.*, **69**, 397–408.
47. Schoof, H., Lenhard, M., Haecker, A., Mayer, K.F., Jurgens, G. and Laux, T. 2000, The stem cell population of *Arabidopsis* shoot meristems is maintained by a regulatory loop between the CLAVATA and WUSCHEL genes, *Cell*, **100**, 635–44.
48. Fletcher, J.C. and Meyerowitz, E.M. 2000, Cell signaling within the shoot meristem, *Curr. Opin. Plant Biol.*, **3**, 23–30.
49. Leibfried, A., To, J.P., Busch, W., et al. 2005, WUSCHEL controls meristem function by direct regulation of cytokinin-inducible response regulators, *Nature*, **438**, 1172–5.
50. Dodsworth, S. 2009, A diverse and intricate signalling network regulates stem cell fate in the shoot apical meristem, *Dev. Biol.*, **336**, 1–9.
51. Carles, C.C., Choffnes-Inada, D., Reville, K., Lertpiriyapong, K. and Fletcher, J.C. 2005, ULTRAPETALA1 encodes a SAND domain putative transcriptional regulator that controls shoot and floral meristem activity in *Arabidopsis*, *Development*, **132**, 897–911.
52. Zhao, Y., Medrano, L., Ohashi, K., et al. 2004, HANABA TARANU is a GATA transcription factor that regulates shoot apical meristem and flower development in *Arabidopsis*, *Plant Cell*, **16**, 2586–600.
53. Scofield, S. and Murray, J.A. 2006, KNOX gene function in plant stem cell niches, *Plant Mol. Biol.*, **60**, 929–46.
54. Kyojuka, J. 2007, Control of shoot and root meristem function by cytokinin, *Curr. Opin. Plant Biol.*, **10**, 442–6.
55. Tangphatsornruang, S., Sraphet, S., Singh, R., Okogbenin, E., Fregene, M. and Triwitayakorn, K. 2008, Development of polymorphic markers from expressed sequence tags of *Manihot esculenta* Crantz, *Mol. Ecol. Resources*, **8**, 682–5.
56. Raji, A.A., Anderson, J.V., Kolade, O.A., Ugwu, C.D., Dixon, A.G. and Ingelbrecht, I.L. 2009, Gene-based microsatellites for cassava (*Manihot esculenta* Crantz): prevalence, polymorphisms, and cross-taxa utility, *BMC Plant Biol.*, **9**, 118.
57. Aggarwal, R.K., Hendre, P.S., Varshney, R.K., Bhat, P.R., Krishnakumar, V. and Singh, L. 2007, Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species, *Theor. Appl. Genet.*, **114**, 359–72.
58. Varshney, R.K., Graner, A. and Sorrells, M.E. 2005, Genic microsatellite markers in plants: features and applications, *Trends Biotechnol.*, **23**, 48–55.
59. Gao, L.F., Tang, J.F., Li, H.W. and Jia, J.Z. 2003, Analysis of microsatellites in major crops assessed by computational and experimental approaches, *Mol. Breed.*, **12**, 245–61.
60. Nicot, N., Chiquet, V., Gandon, B., et al. 2004, Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs), *Theor. Appl. Genet.*, **109**, 800–5.
61. Jung, S., Abbott, A., Jesudurai, C., Tomkins, J. and Main, D. 2005, Frequency, type, distribution and annotation of simple sequence repeats in Rosaceae ESTs, *Funct. Integr. Genomics*, **5**, 136–43.
62. Okogbenin, E., Marin, J. and Fregene, M. 2006, An SSR-based molecular genetic map of cassava, *Euphytica*, **147**, 433–40.