Genomics and Molecular Breeding 13

Scientific developments over the last two decades have led to a new phase of genetics—the plant genomics. This is the application of a wide range of novel methods and technologies for analysing the structure of the genome and its interaction with cell metabolism for protein synthesis and the use of this knowledge for better understanding of the functioning of plants or improving them into new varieties. Plant genomes are more complex than other eukaryotic genomes, and analysis reveals many evolutionary flips and turns of the DNA sequences over time. One spectacular result achieved at the beginning of this century was the entire genomic DNA sequencing of *Arabidopsis* (The Arabidopsis Genome Initiative 2000) (see Bevan and Walsh 2005) and *Oryza* (rice—International Rice Genome Sequencing Project (IRGSP)) (Eckardt 2000; Jackson 2016), the main dicotyledon and monocotyledon plant models used due to their small genomes. Some recent powerful technologies are:

(i) automatic DNA sequencing, where one machine can read two million base pairs a day (Egan et al. 2012);
(ii) microarrays and DNA chips where tens of thousands of genes can be scanned for activity levels at the same time (Pflieger et al. 2001); and
(iii) automated genotyping machines that can assay tens of thousands of DNA diagnostic points a day (Li et al. 2001).

In fact, it will soon be possible to monitor whole genomes by the use of DNA molecular genetic markers (MGMs) or analysis of gene expression on single chips. Two main fields must be distinguished: (i) MGMs, which are non-coding DNA fragments independent from the variation of the environment, and (ii) expressed genes. Genomic technologies were taken up by various research groups working with *Hevea*, in order to increase knowledge and also to identify new targets for breeding and/or complement genetic transformation and assist rubber breeders in various strategies. The group led by Huasun at the Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS) (in collaboration with Beijing Institute of Genomics, Chinese Academy of Science), Danzhou, China (see Tang et al. 2016) and Centre for Chemical Biology, Universiti Sains Malaysia (Rahman et al. 2013) and the group led by Sithichoke Tangphatsornruang at the National Center for Genetic Engineering and Biotechnology, Bangkok, are the prominent schools working on *Hevea* genomics. Tun Abdul Razak Research Centre (TARRC) (Malaysian Rubber Board) and The Genome Analysis Centre (TGAC), Norwich Research Park, UK, are also actively participating in this research. The latest advancement was during 2013, when Rahman et al. reported the draft genome sequence of *H. brasiliensis*. The assembly spans to ~1.1 Gb of the estimated 2.15 Gb haploid

P.M. Priyadarshan, *Biology of* Hevea *Rubber*, DOI 10.1007/978-3-319-54506-6_13

genome. Overall, ~78% of the genome was identified as repetitive DNA. This was refined by Lau et al. (2016).

13.1 Non-expressed Molecular Genetic Markers (MGMs)

In conventional plant breeding, many morphological traits are used as markers for analysing genetic traits and identifying cultivars, but specific genetic information on Mendelian traits has been rare in *Hevea*. In the 1980s, isozymes, the expressions of which are not modified by the environment, have been used in rubber as proteic genetic markers for (i) cultivar identification, (ii) genetic diversity analysis, (iii) control of progenies originating from hand pollination and (iv) reproductive biology (Chevallier 1988; Leconte et al. 1994; Paiva et al. 1994). Analysis of isozymes was developed at CIRAD with a set of 13 polymorphic isozymic systems to formulate a diagnostic kit associated with a clonal identification database. This kit is proved to be able to differentiate a large set of cultivated clones (Leconte et al. 1997). However, the analyses are to be carried out near the field sites due to the fragility of isozyme molecules exposed to normal temperature, or otherwise the samples need to be freeze-dried before transportation to the laboratory. Moreover, isozyme-based analyses are limited by the rather small number of marker loci available and a general lack of polymorphism for these loci.

MGMs are the ideal means for identifying genotypes and tracing the segregation and inheritance of alleles related to economically important characters. MGMs are powerful tools that could enhance the speed and effectiveness of rubber breeding as is already the case for maize and some other plants. General advantages of DNA markers include (i) their ability to reveal the sites of variation in DNA segments among many individuals, (ii) their abundance and distribution over the whole genome and (iii) their independence from the variations of the environment. There is a growing arsenal of MGMs that are currently used, notably in identifying quantitative trait loci (QTLs). The process of using such markers as selection criteria is called marker-assisted selection (MAS), the methodology of which is still at research level in rubber. There are several types of MGMs currently used: (i) restriction fragment length polymorphism (RFLP), (ii) random amplification of polymorphic DNA (RAPD), (iii) amplified fragment length polymorphism (AFLP), (iv) single sequence repeats (SSR) or sequence-tagged microsatellite sites (STMS), (v) DNA amplification fingerprinting (DAF), (vi) microsatellite-primed PCR (MP-PCR) and (vii) single nucleotide polymorphism (SNP).

RFLP involves the use of restriction enzymes to cut chromosomal DNA at specific short restriction sites; polymorphism results from variations in length of the fragments due to duplications or deletions between the sites or mutations at the restriction sites. RFLP provided the basis for most early work but requires a relatively large amount of DNA and is rather expensive in a large screening programme. RAPD utilizes low stringency PCR amplification with single primers of arbitrary sequence to generate strain-specific arrays of anonymous DNA fragments. The method requires tiny DNA samples and analyses a large number of polymorphic loci. AFLP requires digestion of cellular DNA with a restriction enzyme and then using PCR and selective nucleotides in the primers to amplify specific fragments.

The method measures up to 100 polymorphic loci and requires a relatively small DNA sample for each test. SSR analysis is based on DNA microsatellites, (short-repeat) sequences that are widely dispersed throughout the genome of eukaryotes, which are selectively amplified to detect variations in the number (and length) of simple sequence repeats. SSR analysis requires tiny DNA samples and has a low cost per analysis. SNPs are detected using PCR extension assays that efficiently pick up point mutations. The procedure requires little DNA per sample and costs little per sample once the method is established. MGMs, independent from the environment such as isozymes, have been used in the same way to determine the degree of variability within plant populations. They allow direct

access to the coding and non-coding regions of the genome, making their number potentially unlimited. Among different characteristics, they can be easily visualized and identified (such as by probes and Southern blots, PCR amplification and electrophoresis, radio-labelling, fluorescence). The more efficient MGMs generally assign one sole locus, and the diversity of their possible alleles determines the level of their polymorphism. The first applications to rubber, on nuclear and cytoplasmic genomes, were made with RFLPs. Then, when a PCR technique was developed with random primers, RAPDs were used, as well as AFLPs, which combine restriction enzymes and PCR. Microsatellites (SSRs), of PCR developed from specifically designed primers, appeared very powerful due to their high polymorphism (between 15 and 20 alleles per marker). In comparison with isozymes or mRNA, the high stability of DNA makes it possible to send leaf samples from any remote plantation site to a laboratory by normal mail for analysis.

Although RFLPs are powerful tools for studying genetic diversity and mapping, this technology is not preferred now since it is labour intensive and requires large DNA samples. Its marker index value (expressed as the number of polymorphic products per sample) is also low with only 0.10 compared with PCR-based marker systems like RAPDs (0.23), SSRs (0.60) and AFLPs (6.08) (Low et al. 1996). Ever since isozymes were utilized for clonal identification (Chevallier 1988), tools like minisatellites (Besse et al. 1993a), RFLPs (Besse et al. 1993b, 1994), mitochondrial and chloroplastic RFLPs (Luo et al. 1995), RAPDs and DAFs (Low et al. 1996; Varghese et al. 1998; Venkatachalam et al. 2001, 2002), AFLPs (Lespinasse et al. 2000a) and SSRs (Besse et al. 1993a; Atan et al. 1996; Low et al. 1996) were developed and used in detection, increasing the number of molecular markers in *H. brasiliensis*. Polymorphism in microsatellites was detected also in *H. pauciflora*, *H. guianensis*, *H. camargoana* and *H. benthamiana* (Low et al. 1996), and cross-species amplification was also done in these species (Souza et al. 2009). Microsatellite-enriched libraries were produced and led to the identification of large numbers of microsatellite markers (Atan et al. 1996; Seguin et al. 2003; Saha et al. 2005); sequences of 472 of them are currently accessible on the European Molecular Biology Laboratory (EMBL)/GenBank databases. CIRAD has also developed a database for rubber clonal identification by the use of ten microsatellite markers; microsatellite markers from rubber pathogens can also be used for distinguishing the genetic differences between the races, as has been done with 11 markers for *Microcyclus* (Le Guen et al. 2004).

Although considerable progress has been made to increase the yield in *Hevea* clones in the recent past, satisfactory resistance to biotic and abiotic stress has not been achieved because of limited genetic resources within the *Hevea* gene pool. Wind damage is one of the serious problems in rubber-growing countries; each year there is a considerable loss of rubber trees due to wind damage in rubber plantations. The incorporation of the dwarf character into high-yielding *Hevea* clones would be useful for generating a high-yielding tree with a desirable architecture (dwarf stature) (Venkatachalam et al. 2004). Information on the genetic and molecular basis of the dwarf character in this species could provide insights on the development of high-yielding dwarf clones that would eventually lead to decreasing the negative consequences of wind damage (abiotic stress) and high-density planting but might have negative effects on rubber wood productivity. The identification of molecular markers for the dwarf character would be important for isolating true-to-type high-yielding dwarf hybrid lines in the early stage of plant breeding programmes. Venkatachalam et al. (2004) identified a dwarf genome-specific RAPD marker in the rubber tree. The primer OPB-12 generated a 1.4 kb DNA marker from both natural and controlled F1 hybrid progenies (dwarf stature) derived from a cross between a dwarf parent and a normal cultivated clone as well as from the dwarf parent; it was absent in the other parent (RRII 118). To validate this DNA marker, 22 F1 hybrids (13 with a dwarf stature and 9 with a normal stature) were analysed; the dwarf genome-specific 1.4 kb RAPD marker was present in all dwarf-stature hybrids and absent in all

normal-stature hybrids. This DNA marker was cloned and characterized. DNA marker locus specificity was further confirmed by Southern blot hybridization.

13.2 Expressed Genes in *Hevea*

Non-expressed MGMs and also markers of expressed genes (single-strand conformation polymorphism (SSCP)), based on PCR, aimed at mutation detection in expressed genes, were studied by Lekawipat (2004) in 66 Amazonian and 40 Wickham accessions. It was found that microsatellites could detect higher polymorphism than gene-specific primers of SSCP in rubber accessions, although markers of expressed genes can be assumed to be more related with some putative breeding objectives. SSCP markers could not differentiate the Wickham and the Mato Grosso accessions. By the use of reverse genetics from mRNA to cDNA libraries (RT-PCR), the fields of functional genomics and molecular physiology are being developed in rubber by different teams, predominantly working on latex cells, on such themes as rubber biosynthesis, latex-cell functioning, the latex coagulation process, ethylene biosynthesis and metabolism, oxidative stress, tapping cut dryness and brown bast (the reversible or irreversible forms of TPD), allergenic proteins in the latex, heterologous genes to be expressed in the latex, drought tolerance, leaf fungus diseases, cyanogenesis metabolism or defence proteins and photosynthesis.

In reproductive biology, rubber flower and inflorescence development has been characterized: one important gene regulating flower induction and development (*leafy/floricaula*) was cloned and its expression was analysed and localized by in situ hybridization (Dornelas and Rodriguez 2005). In the field of post-germination changes in rubber seeds, proteomics (2D-PAGE and mass spectrometry methods) were implemented for examining the changes in protein expression from the mature seed to the germinated seed (Wong and Abubakar 2005). The suppression subtractive hybridization (SSH) technique is currently widely implemented between different pairs of mRNA samples for the production of molecular resources by RT-PCR in the form of subtracted cDNA libraries; microarrays or sequencing and comparison with entries from the databases will then assist in searching the functions of these expressed genes (candidate gene approach) (see Sathik et al. 2009). Expressed sequence tags (ESTs, or small and partial 5′-end sequences of expressed genes) related to various metabolic aspects are being collected to create EST banks that broadly represent the genes expressed in one tissue, such as latex cells, and this assists in the study of gene function and regulation. Entries of these banks are compared with public databases of already known genes for identifying the putative functions of the corresponding genes. These EST banks will also create the way for macroarray- or microarray-based studies of *Hevea* gene expression. The 'Latex Lambda Triplex' EST-cDNA library (Ko et al. 2003) published in the EMBL/GenBank databases (858 entries) showed that about 16% of the database matched ESTs encoding rubber biosynthesis-related proteins. Rubber biosynthesis-related genes appeared to be mostly expressed, followed by defence-related genes and other protein-related genes (Han et al. 2000). Published DNA sequences of the latex allergens were matched against these ESTs, so indirectly providing a ranking of the allergens depending on their concentration in the latex; more than 1000 ESTs matched with the sequences of *REF* (gene for rubber elongation factor, or *Hev.b.1*) and *SRPP* (gene for small rubber particle protein, or *Hev.b.3*). Cubry et al. (2014) retrieved EST sequences from public database, and these sequences were trimmed and microsatellite motifs searched using an ad hoc bioinformatic pipeline. Pairs of primers for the amplification of candidate markers were generated. A total of 10,499 unigenes from both sources of sequences and 673 microsatellites motifs were detected using the default parameters of the pipeline. Two-hundred and sixty-four primer pairs were tested and 226 (85.6%) were successfully amplified. Out of the amplified candidate markers, 164 exhibited polymorphism. A striking expansion of

the REF/SRPP (rubber elongation factor/small rubber particle protein) gene family and its divergence into several laticifer-specific isoforms seem crucial for rubber biosynthesis. Tang et al. (2016) presented a high-quality genome assembly of *Hevea* clone Reyan 7-33-97 (1.37 Gb, scaffold N50 = 1.28 Mb) that covers 93.8% of the genome (1.47 Gb) and harbours 43,792 predicted protein-coding genes (Table 13.1). The REF/SRPP family has isoforms with sizes similar to or larger than SRPP1 (204 amino acids) in 17 other plants examined, but no isoforms detected with similar sizes to REF1 (138 amino acids), the predominant molecular variant. The expansion of vast repetitive elements makes the genome size of *Hevea* apparently larger than its closely related species, such as cassava, castor oil, cottonwood and flax. REF1 is special for its smaller size and is highly expressed in latex than the others, indicating its predominant role in rubber production (see Fig. 13.1). REF1 gene expression was reported to correlate with yield levels of *Hevea* cultivars (Priya et al. 2007). There are two classes of rubber particles in *Hevea* latex: the large particles (LRPs) with REF located on their surfaces and the small particles (SRPs) that are coated with SRPP (Dennis and Light 1989; Berthelot et al. 2014). SRPs are far superior in number, accounting for 94% of all rubber particles in the latex, whereas LRPs constitute only the remaining 6% of the rubber particles. However, it is precisely this 6% of rubber particles by number that makes up 94% of the rubber by volume in the latex (Yeang et al. 1995). Tang et al. (2016) argue that the emergence of REF1, which is located on the surface of large rubber particles that account for 94% of rubber in the latex (despite constituting only 6% of total rubber particles, large and small), is pivotal to *Hevea* evolution.

Table 13.1 Details of *Hevea* genome and gene annotation

Estimate of genome size	1.46 Gb
Number of scaffolds	7453
Total length of scaffolds	1.37 Gb
N50 of scaffolds	1.28 Mb
Longest scaffolds	6.41 Mb
Number of contigs	84,285
Total length of contigs	1.29 Gb
N50 of contigs	30.6 kb
Longest contigs	312.7 kb
GC content	34.84%
Number of genes	43,792
Percentage of gene length in genome	12.47%
Mean gene length	3913 bp
Gene density	31.9 Mb^{-1}
Transcripts number	46,631
Mean transcript length	1483 bp
Mean coding sequence length	1123 bp
Mean exon length	308 bp
Exon GC content	41.54%
Mean intron length	677 bp
Intron GC content	32.61%
Masked repeat sequence length	977.5 Mb
Repeats percentage of total scaffolds	71.18%
Repeats percentage of total contigs	75.69%

After Tang et al. (2016)

Expressed sequence tags, derived as simple sequence repeats (microsatellites = EST-SSRs), are different from traditional genomic SSR (gSSR) markers. They are more likely to be embedded in the functional gene sequences, less costly and time effective, and may provide abundant information. By analysis of 10,018 expressed sequence tags (ESTs) out of 10,829 for rubber tree (*Hevea brasiliensis*) available in public domain DNA databases, 799 SSR loci were found in the 643 non-redundant SSR-ESTs (SSR-containing ESTs), corresponding to 1 SSR in every 2.25 kb of the ESTs in rubber tree transcriptome (Feng et al. 2009). Of the total 799 SSRs in these ESTs, 673 (84.2%) contained simple repeat motifs, while 126 (15.8%) represented compound motif types. Of the total EST-SSRs, 45.3% (362/799) were mononucleotide repeats (MNRs), 42.2% (337/799) were dinucleotide repeats (DNRs), 11.9% (95/799) were trinucleotide repeats (TNRs) and 0.6% (5/799) were tetranucleotide repeats (TTNRs) and hexanucleotide repeats (HNRs). The repeat motifs AAG and AG were the most abundant without regard to single nucleotide repeat. A total of 184 primer pairs were designed based on the non-redundant SSR-ESTs. Using 55 °C as annealing temperature, 110 primer pairs successfully amplified 12 *H. brasiliensis* cultivated varieties and 4 related species. Analysis on 74 alleles amplified by 30

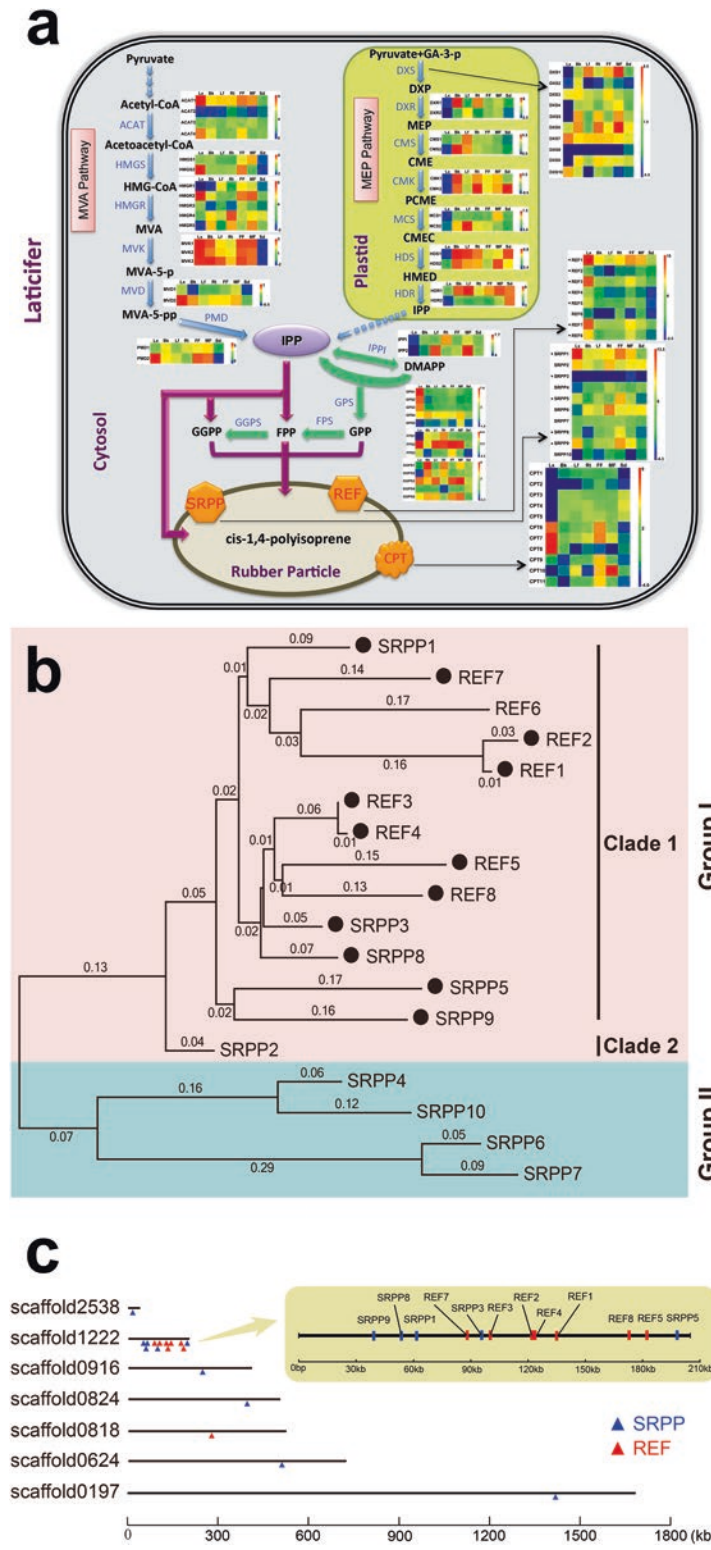

Fig. 13.1 (**a**) The rubber biosynthesis pathway and expression profiles of genes involved (*Lx* latex, *Bk* bark, *Lf* leaf, *Rt* root, *FF* female flower, *MF* male flower). (**b**) Phylogeny of REF/SRPP gene family. The tree is constructed by using MEG A5.1 with neighbour-joining model and bootstrap test with 1000 replicates. (**c**) Genomic location of the REF/SRPP genes. Scaffolds are represented as solid bars with their names on the left and length on the bottom. Note that most of the REF/SRPP genes including the four laticifer-specific ones (REF1, SRPP1, REF3 and REF7) are located in a single scaffold (scaffold 1222) (After Tang et al. 2016)

randomly selected primer pairs indicated the medium polymorphism of the EST-SSRs developed. Based on 272 alleles detected by 87 EST-SSR markers, an assessment of genetic diversity was carried out on 12 *H. brasiliensis* clones and 4 related species. In addition, investigation based on five selected EST-SSRs by cloning and sequencing across some cultivated species and related species provided evidence for cross-species/genera transferability of the EST-SSR markers (Feng et al. 2009).

Development and characterization of new genomic microsatellite markers in *H. brasiliensis* and the evaluation of their transferability to other *Hevea* species were attempted in the recent past (Mantello et al. 2012). They constructed di-and tri-nucleotide enriched libraries. From these 2 libraries, 153 primer pairs were designed and initially evaluated using 9 genotypes of *H. brasiliensis*. A total of 119 primer pairs had a good amplification product, 90 of which were polymorphic. We chose 46 of the polymorphic markers and characterized them in 36 genotypes of *H. brasiliensis*. The expected and observed heterozygosities ranged from 0.1387 to 0.8629 and 0.0909 to 0.9167, respectively. The polymorphism information content (PIC) values ranged from 0.097 to 0.8339, and the mean number of alleles was 6.4 (2–17). These 46 microsatellites were also tested in 6 other *Hevea* species. The percentage of transferability ranged from 82% to 87%. Locus duplication was found in *H. brasiliensis* and also in 5 of other species in which transferability was tested. This study reports new microsatellite markers for *H. brasiliensis* that can be used for genetic linkage mapping, quantitative trait loci identification and marker-assisted selection. The high percentage of transferability may be useful in the evaluations of genetic variability and to monitor introgression of genetic variability from different *Hevea* species into breeding programmes.

Genes responsible for the synthesis of rubber transferase, the key enzyme for polymerization of polyisoprene (natural rubber), appeared to be among the most abundantly expressed genes in the latex. *Hevein*, a chitin-binding protein, one of the defence proteins that play a crucial role in the protection of wound sites from fungal attack, is also involved in the coagulation process; it belongs to a multigene family, and the specificity of its expression in the latex is under investigation (Broekaert et al. 1990; Pujade-Renaud et al. 2005). Nearly 12.6% of the proteins available in the latex are defence related (Han et al. 2000). Among 200 distinct polypeptides (Posch et al. 1997), mainly three rubber synthesis-related genes are expressed in the latex: (i) *REF* (Dennis and Light 1989; Goyvaeerts et al. 1991), (ii) *HMGR* (Chye et al. 1992) and (iii) *SRPP* (Oh et al. 1999). The most abundantly expressed gene is *REF* (6.1%) and then *SRPP* (3.7%) (Han et al. 2000). References and partial- or full-length sequences of these cloned genes can be found in the EMBL/GenBank databases.

Unlike photosynthetic genes, transcripts involved in rubber biosynthesis are 20–100 times greater in laticifers than in leaves (Kush et al. 1990). On the other hand, transcripts for chloroplastic and cytoplasmic forms of glutamine synthetase are restricted to leaves and laticifers, respectively (Kush et al. 1990), indicating thereby that the cytoplasmic form of glutamine synthetase plays a decisive role in amino acid metabolism of laticifers. The transcript levels of hydrolytic enzymes, namely, polygalacturonase and cellulase, might be taken as indicators for a better development of the laticifers. Genes expressed in the latex of *Hevea* can be divided into three groups based on the proteins they encode: (i) defence-related proteins such as hevein, chitinase, β-1,3-glucanase and HEVER; (ii) rubber biosynthesis-related proteins such as rubber elongation factor (REF), HMGR and 3-hydroxy-3-methylglutaryl coenzyme A synthase (HMGS), *cis*-prenyltransferase (CIS), geranyl-geranyl diphosphate (GGPP) synthase, small rubber particle protein (SRPP) and isopentenyl diphosphate (IPP) isomerase; and (iii) latex allergen proteins such as Hev.b.3, Hev.b.4, Hev.b.5 and Hev.b.7. The biological functions of the allergenic proteins are largely unknown (Oh et al. 1999).

Genes expressed under stressed conditions is yet another topic that received due attention recently. Silva et al. (2014) studied leaf, panel

and latex ESTs under cold-stressed conditions. For panel and latex libraries, samples were collected from 16-year-old tree clones of PB217, PR255, GT1, PB235, RRIM701 and IAN873, and leaves of the same clones were collected from the rubber tree germplasm. PB217, PR255, GT1 and IAN873 were subjected to a 24-h cold treatment in a growth chamber and maintained at 8 °C with a 12-h photoperiod. This treatment was performed to promote the expression of genes involved in cold response and for the development of molecular markers related to this stress condition. A total of 8263 reads were assembled, generating 5025 unigenes that were analysed; 912 expressed sequence tags (ESTs) represented new transcripts, and 2 sequences were highly upregulated by cold stress. These unigenes were scanned for microsatellite (SSR) regions and single nucleotide polymorphisms (SNPs). In total, 169 novel EST-SSR markers were developed, of which 138 loci were polymorphic. Nearly 98% of this presented transferability to six other *Hevea* species. Locus duplication was observed in *H. brasiliensis* and other species. Additionally, 43 SNP markers in 13 sequences that showed similarity to proteins involved in stress response, latex biosynthesis and developmental processes were characterized. cDNA libraries are a rich source of SSR and SNP markers and enable the identification of new transcripts. Such markers will be a valuable resource for linkage mapping and QTL identification.

13.3 Transcriptome Analysis

Transcriptome analysis is one of the main approaches for identifying the complete set of active genes in a cell or tissue for a specific developmental stage or physiological condition. Transcriptome analysis is a second-generation sequencing technology offered by companies such as Illumina, Roche and Life Technologies. Salgado et al. (2014) reported on the sequencing, assembling, annotation and screening for molecular markers from a pool of *H. brasiliensis* tissues. A total of 17,166 contigs were successfully annotated. Then, 2191 single nucleotide variation (SNV) and 1397 simple sequence repeat (SSR) loci were discriminated from the sequences. This is the first study of the *Hevea* transcriptome, covering a wide range of tissues and organs, leading to the production of the first developed SNP markers. Transcriptome studies have to come a long way to yield meaningful results to tag such gene expression studies to vivid genes responsible for QTLs, resistance and other quality traits.

By assembling and analysing de novo transcriptome sequencing data, Li et al. (2012) reported the comprehensive functional characterization of rubber tree bark. This research generated a substantial fraction of rubber tree transcriptome sequences, which were very useful resources for gene annotation and discovery, molecular markers development, genome assembly and annotation, and microarrays development in rubber tree. The EST-SSR markers identified and developed in this study will facilitate marker-assisted selection breeding in rubber tree. Moreover, this study also supported that transcriptome analysis based on Illumina paired-end sequencing is a powerful tool for transcriptome characterization and molecular marker development in non-model species, especially those with large and complex genomes.

To obtain more information on the *Hevea brasiliensis* genome, Triwitayakorn et al. (2011) 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 (see Box 13.1). 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.

Box 13.1 Isotigs, Singletons and Contigs

Isotigs: Normally one gene should have one transcript, but due to splice variation one gene can have many transcripts. Suppose a gene is made of three exons, exon1, exon2 and exon3. It will generate three contigs in newbler, contig 1, contig 2 and contig 3. Due to splice variation, the final transcript can consist of exon1 + exon2 + exon3 or exon1 + exon3, etc. Thus, we get two types of variation here; these two are called isotig 1 and isotig 2. isotig 1 consists of contig 1 + contig 2 + contig 3 and isotig 2 consists of contig 1 + contig 3. These two isotigs are variation of one transcript. So these two isotigs combinedly fall into one isogroup.

Singletons: All reads are not provided as input in the final assembly. The unused reads, also called singletons, are often contaminants or insufficiently trimmed reads from the genome.

Contig: contig (originated from contiguous) is a set of overlapping DNA segments that together represent a consensus DNA.

Scaffold and contig N50s: The most widely used annotations for describing the quality of a genome assembly are its scaffold and contig N50s. A contig N50 is calculated by first ordering every contig by length from longest to shortest. Next, starting from the longest contig, the lengths of each contig are summed, until this running sum equals one-half of the total length of all contigs in the assembly. The contig N50 of the assembly is the length of the shortest contig in this list. The scaffold N50 is calculated in the same fashion but uses scaffolds rather than contigs. The longer the scaffold N50 is, the better the assembly is. However, it is important to keep in mind that a poor assembly that has forced unrelated reads and contigs into scaffolds can have an erroneously large N50. Note too that scaffolds and contigs that comprise only a single read or read pair—often termed 'singletons'—are frequently excluded from these calculations, as are contigs and scaffolds that are shorter than ~800 bp. The procedures used to calculate N50 may therefore vary between genome projects.

Reads: reads refer to somewhat digital information obtained from the sequencing machine (e.g. Illumina MiSeq) and stored in the fastq file with quality scores per base. From a single run one can get millions of reads, where each read will have a set bp size, e.g. 100 bp long.

TSS (Transcription Start Site): It is where the RNA polymerase sits down on the DNA and starts to make an RNA copy of the DNA.

CAGE (cap analysis gene expression): This is a technique used in molecular biology to produce a snapshot of the 5′ end of the mRNA population. The small fragments (usually 27 nucleotides long) from the very beginnings of mRNAs (5′ ends of capped transcripts) are extracted, reverse-transcribed to DNA, PCR amplified and sequenced.

Construction of linkage maps is crucial for genetic studies and marker-assisted breeding programmes. Recent advances in next-generation sequencing technologies allow for the generation

of high-density linkage maps, especially in non-model species lacking extensive genomic resources. Pootakham et al. (2015) constructed a high-density integrated genetic linkage map of rubber tree (*H. brasiliensis)*, the sole commercial producer of high-quality natural rubber. We applied a genotyping-by-sequencing (GBS) technique to simultaneously discover and genotype single nucleotide polymorphism (SNP) markers in two rubber tree populations. A total of 21,353 single nucleotide substitutions were identified, 55% of which represented transition events. GBS-based genetic maps of populations P and C comprised 1704 and 1719 markers and encompassed 2041 cM and 1874 cM, respectively. The average marker densities of these two maps were one SNP in 1.23–1.25 cM. A total of 1114 shared SNP markers were used to merge the 2 component maps. An integrated linkage map consisted of 2321 markers and spanned the cumulative length of 2052 cM. The composite map showed a substantial improvement in marker density, with one SNP marker in every 0.89 cM. This is the most saturated genetic map in rubber tree to date. This integrated map allowed us to anchor 28,965 contigs, covering 135 Mb or 12% of the published rubber tree genome. Pootakham et al. (2015) further demonstrated that GBS is a robust and cost-effective approach for generating a common set of genome-wide SNP data suitable for constructing integrated linkage maps from multiple populations in a highly heterozygous agricultural species.

In an attempt to develop additional microsatellite markers in rubber tree, Pootakham et al. (2012) employed a 454 pyrosequencing technology to obtain genomic shotgun sequences and subsequently identified over 24,000 putative simple sequence repeats (SSRs). A total of 418 potential SSR loci were chosen for an empirical validation. Two-hundred and twenty-four primer pairs yielded successful amplification, and 90 SSR markers exhibited polymorphism among the 18 rubber tree accessions evaluated. The number of alleles per locus ranged from 2 to 7, with an average of 3.44 alleles per SSR. The gene diversity of individual microsatellite loci displayed a broad range of values from 0.104 to 0.901 with a mean of 0.612. The polymorphism information content also varied greatly from 0.099 to 0.893 with an average of 0.577. We also identified a set of five highly informative markers that were able to unequivocally distinguish 18 rubber tree genotypes examined in this study.

In an effort to facilitate biological, biochemical and molecular research in rubber biosynthesis, Xia et al. (2011) reported the use of next-generation massively parallel sequencing technologies and de novo transcriptome assembly to gain a comprehensive overview of the *H. brasiliensis* transcriptome. The sequencing output generated more than 12 million reads with an average length of 90 nt. In total 48,768 unigenes (mean size = 436 bp, median size = 328 bp) were assembled through de novo transcriptome assembly. Out of 13,807 *H. brasiliensis* cDNA sequences deposited in GenBank of the National Center for Biotechnology Information (NCBI) (as of February 2011), 11,746 sequences (84.5%) could be matched with the assembled unigenes through nucleotide BLAST. The assembled sequences were annotated with gene descriptions, Gene Ontology (GO) and Clusters of Orthologous Group (COG) terms. In all, 37,432 unigenes were successfully annotated, of which 24,545 (65.5%) aligned to *Ricinus communis* proteins. Furthermore, the annotated unigenes were functionally classified according to the GO, COG and Kyoto Encyclopedia of Genes and Genomes databases. Xia et al. (2011) claim that this is the most comprehensive sequence resource available for the study of rubber trees as well as it demonstrates effective use of Illumina sequencing and de novo transcriptome assembly in a species lacking genomic information.

Mantello et al. (2014) performed RNA sequencing (RNA-seq) of *H. brasiliensis* bark on the Illumina GAIIx platform, which generated 179,326,804 raw reads on the Illumina GAIIx platform. A total of 50,384 contigs that were over 400 bp in size were obtained and subjected to further analyses. A similarity search against the non-redundant (nr) protein database returned 32,018 (63%) positive BLASTx hits. The transcriptome analysis was annotated using the clusters of orthologous groups (COG), gene ontology (GO),

Kyoto Encyclopedia of Genes and Genomes (KEGG) and Pfam databases. A search for putative molecular marker was performed to identify simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs). In total, 17,927 SSRs and 404,114 SNPs were detected. Finally, we selected sequences that were identified as belonging to the mevalonate (MVA) and 2-C-methyl-D-erythritol 4-phosphate (MEP) pathways, which are involved in rubber biosynthesis, to validate the SNP markers. A total of 78 SNPs were validated in 36 genotypes of *H. brasiliensis*. This new data set represents a powerful information source for rubber tree bark genes and will be an important tool for the development of microsatellites and SNP markers for use in future genetic analyses such as genetic linkage mapping, quantitative trait loci identification, investigations of linkage disequilibrium and marker-assisted selection.

Molecular mechanisms underlying high yield are not well understood. Li et al. (2015) reported the sequencing, assembly and comparative analyses of latex transcriptome from RRIM600 and RY 7-20-59. In total, 33,852 unigenes were generated with de novo assembly. The blastx results indicated that 27,886 and 15,704 unigenes showed significant similarities to known proteins from NCBI nr and Swissprot databases, respectively. Among these annotated unigenes, 21,841 and 9010 were separately assigned to Gene Ontology (GO) functional categories and Clusters of Orthologous Groups (COGs). Of 126 KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways, metabolic pathway was the biggest one, suggesting that active metabolic processes happen in rubber tree latex. In contrast to RRIM600, 2513 and 1391 genes were separately up- and down-regulated in RY 7-20-59. The expression profiles of 25 unigenes were further confirmed by real-time PCR, suggesting that the differently expressed genes (DEGs) identified by RNA-seq were accurate and reliable. The DEGs between RRIM600 and RY 7-20-59 were significantly enriched in plant-pathogen interactions, phenylalanine metabolism, ubiquinone and other terpenoid-quinone biosynthesis, biosynthesis of secondary metabolites and photosynthesis. Interestingly, the genes involved in rubber biosynthesis pathway, such as CPT, GPPS, HMGR, HMGS, FPPS and DXS, were differently expressed between RRIM600 and RY 7-20-59. Such studies really help to provide new insights into understanding latex transcriptome and molecular mechanisms underlying high yield.

13.4 Rubber Biosynthesis

Rubber biosynthesis in *Hevea* laticifer cells has become a major field of research applied to the expression and regulation of genes, with a view to possibly opening the way to genetic manipulation of the isoprenoid biosynthesis pathway enzymes. Rubber molecules (1,4 *cis*-polyisoprene) are formed from polymerization of molecules with five carbons, IPP, and aggregated as rubber particles packaged within a membrane which protects them from oxidation, in latex vessels. The general metabolic pathway of rubber biosynthesis is as follows. Sucrose from photosynthesis is actively transported into laticiferous cells through the plasmalemmic membrane and is then hydrolysed into glucose and fructose by invertase. These sugars are then converted into acetyl-coenzyme A (acetyl-CoA) through glycolysis. Three molecules of acetyl-CoA are condensed into mevalonic acid and then IPP. Polymerization of thousands of molecules of IPP leads to dimethylallyl diphosphate (DMAPP) and GGPP, with the action of the enzyme rubber transferase associated with REF, a molecule fixed on the rubber particles' membranes.

Although natural rubber is synthesized and made almost entirely of isoprene units derived from IPP, an allylic disphosphate is also required as the priming co-substrate to initiate the subsequent extensive prenyl chain elongation process for the formation of rubber macromolecules. Both the HMGS and the HMGR have been shown to be involved in the early steps of rubber biosynthesis by forming 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) using HMGS. The HMGS catalyses the condensation of acetyl-CoA and acetoacetyl-CoA to form HMG-CoA (Suwanmanee et al. 2002, 2004; Sirinupong et al. 2005).

In plants, HMG-CoA is reduced by HMGR to mevalonate (MVA) and is subsequently converted into IPP. Chye et al. (1992) reported that there are three genes encoding HMGR in *Hevea*, namely, *hmg1*, *hmg2* and *hmg3*. The *hmg1* gene is likely to be involved in rubber biosynthesis, whereas the *hmg3* is possibly involved in isoprenoid biosynthesis of another nature. Gene expression analysis by type of tissue indicated that MVA-pathway genes were highly expressed in latex, as compared with other types of tissue, and that HMGS and HMGR exist in multiple copies and have different expression patterns (Sando et al. 2008).

The first step in rubber biosynthesis is the isomerization of IPP to DMAPP by IPP isomerase. The successive head-to-tail condensation reactions of the five-carbon intermediates, catalysed by the enzyme rubber transferase, have been assumed to yield rubber. Oh et al. (2000) have isolated and characterized a cDNA clone encoding IPP isomerase from *Hevea* and showed in an in vitro rubber assay that the recombinant IPP isomerase was required for rubber biosynthesis. In order to examine possible participation of GGPP synthase in the enzymatic prenyl chain elongation in natural rubber biosynthesis, Takaya et al. (2003) studied the GGPP synthase gene from *Hevea*. Based on their investigation, GGPP synthase would catalyse the condensation of IPP with allylic diphosphates to give GGPP. Therefore, GGPP is one of the key precursors in the biosynthesis of biologically significant isoprenoid compounds. Natural rubber is thought to be made almost entirely of *cis*-isoprene units derived from IPP, and the enzyme responsible for polymerization is believed to have characteristics similar to the *cis*-prenyl diphosphate synthases. The gene responsible for the *cis*-1,4-polymerization of isoprene units has been isolated and characterized in *Hevea* recently by Asawatreratanakul et al. (2003). It was suggested that rubber biosynthesis in *Hevea* is mediated by the association of a soluble *trans*-prenyltransferase with the REF, a 14.6 kDa protein, tightly bound to the rubber particles in the laticifers (Dennis and Light 1989). Farnesyl diphosphate (FPP) is a key intermediate in the biosynthesis of at least 20,000 isoprenoids. FPP is also the allylic diphosphate initiator of rubber biosynthesis in plants. FPP is synthesized by the enzyme farnesyl diphosphate synthase (FPS), which has been cloned and characterized from *Hevea* by Adiwilaga and Kush (1996). FPP formed by FPS is assumed to be the initiating substrate for rubber formation. The REF protein was isolated and studied extensively, and the results indicate that the FPS and REF complex was responsible for the *cis*-1,4-polyprenol condensations observed in isolated rubber particles (Light and Dennis 1989). The *REF* gene was isolated and the involvement of REF on rubber formation was analysed by Attanayaka et al. (1991) and Goyvaeerts et al. (1991). Oh et al. (1999) reported a novel *Hevea* cDNA sequence associated with SRPP, and the sequence analysis revealed that this protein is highly homologous to the REF. It is possible that the rubber biosynthesis pathway is coordinately regulated by these enzymes. However, the precise mechanism for the biosynthesis of rubber molecules has not yet been established. Moreover, the exact site of the formation of new rubber molecules still remains unknown.

The genes uniquely or preferentially expressed in the latex may be important for rubber biosynthesis (Han et al. 2000). They constructed cDNA libraries from the latex of *Hevea* to investigate the genes expressed in the latex by single-run partial sequencing of the cDNA clones. Sequence analyses identified 245 expressed sequence tags (ESTs), of which 57% showed homology to previously described sequences in public databases. About 16% of the database-matched ESTs encode rubber biosynthesis-related proteins such as rubber elongation factor (REF) and small rubber particle protein (SRPP). The second most frequent transcripts next to rubber biosynthesis-related genes were defence genes and protein metabolism-related genes (12.6% each). About 27% of the database-matched ESTs had sequence homology with genes of unknown function. Among the redundantly expressed genes, REF was the most frequently expressed (6.1%), followed by SRPP (3.7%) and HbLAR (2.9%). Northern blot analyses showed that ten (71%) of the 14 ESTs studied were expressed at a higher level in latex than in leaves.

13.5 Gene Mapping

To gain a better understanding of miRNAs and their relationship with rubber tree gene regulation networks, Kanjanawattanawong et al. (2014) generated large genomic DNA insert-containing libraries to complement the incomplete draft genome sequence and applied as a new powerful tool to predict the function of interested genes. Bacterial artificial chromosome and fosmid libraries (a fosmid library is prepared by extracting the genomic DNA from the target organism and cloning it into the fosmid vector), containing a total of 120,576 clones with an average insert size of 43.35 kb, provided approximately 2.42 haploid genome equivalents of coverage based on the estimated 2.15 Gb rubber tree genome (Rahman et al. 2013). Based on these library sequences, the precursors of 1 member of rubber tree-specific miRNAs and 12 members of conserved miRNAs were successfully identified. A panel of miRNAs was characterized for phytohormone response by precisely identifying phytohormone-responsive motifs in their promoter sequences. Furthermore, the quantitative real-time PCR on ethylene stimulation of rubber trees was performed to demonstrate that the miR2118, miR159, miR164 and miR166 are responsive to ethylene, thus confirming the prediction by genomic DNA analysis. The *cis*-regulatory elements identified in the promoter regions of these miRNA genes help augment the understanding of miRNA gene regulation and provide a foundation for further investigation of the regulation of rubber tree miRNAs.

Past several years have witnessed major advances in our understanding of plant genomes and genomic information through whole genome sequencing. The increasing availability of data from several plant genome sequencing projects provides a promising direction for investigating genes and their functional and sequence homologs involved in plant development (Avraham et al. 2008). Although genome sequencing projects lead to the identification of the complete catalogue of genes of an organism, they don't consider the gene expression patterns. Large-scale end sequencing of cDNA library generates ESTs, representing genes expressed in particular tissues or under particular developmental or environmental conditions. They have also been the target of sequencing in many of the projects and found invaluable for genome assembly and annotation. Whole genome sequence information helps in many aspects of plant trait improvement through gene discovery to transgenesis and use of molecular markers in breeding. *Hevea* genome sequencing project has already been launched jointly by Tun Abdul Razak Research Centre (TARRC) of the Malaysian Rubber Board and the newly established Genome Analysis Centre at Norwich, UK, and probably RRIM600 was sequenced and draft sequence has been made available (Rahman et al. 2013). Quantum of *Hevea* genome sequencing work is a monumental task as the haploid genome size is enormous (~ 4 x 10^3 Mbp as per calculation based on the DNA content measured by Leitch et al. (1998)) and also rubber possesses a high-complexity genome with >60% repetitive sequences. Taking RRIM 600 as an example, Lau et al. (2016) assembled the genome based on ~155-fold combined coverage with Illumina and PacBio sequence data and has a total length of 1.55 Gb with 72.5% comprising repetitive DNA sequences. A total of 84,440 high-confidence protein-coding genes were predicted. Comparative genomic analysis revealed strong synteny between *H. brasiliensis* and other Euphorbiaceae genomes. Production of high levels of latex can be attributed to the expansion of rubber biosynthesis-related genes, their high expression in latex (Fig. 13.2).

The genome assembly consists of 189,316 scaffolds with an N50 size of 67.2 Kb. Using high-coverage sequence data and the application of PacBio long reads, the N50 size of *Hevea* genome assembly was increased 23-fold, and the number of scaffolds was decreased threefold compared to the previously published assembly of Rahman et al. (2013). Anchoring the genome assembly to the linkage groups of *H. brasiliensis* with 196 RFLP markers from the published genetic map (Lespinasse et al. 2000a) was done by Lau et al. (2016). In total, 189 scaffolds that account for 43.6 Mb in length (3% of the assembly size) were anchored. The assembly was anchored to 18 linkage

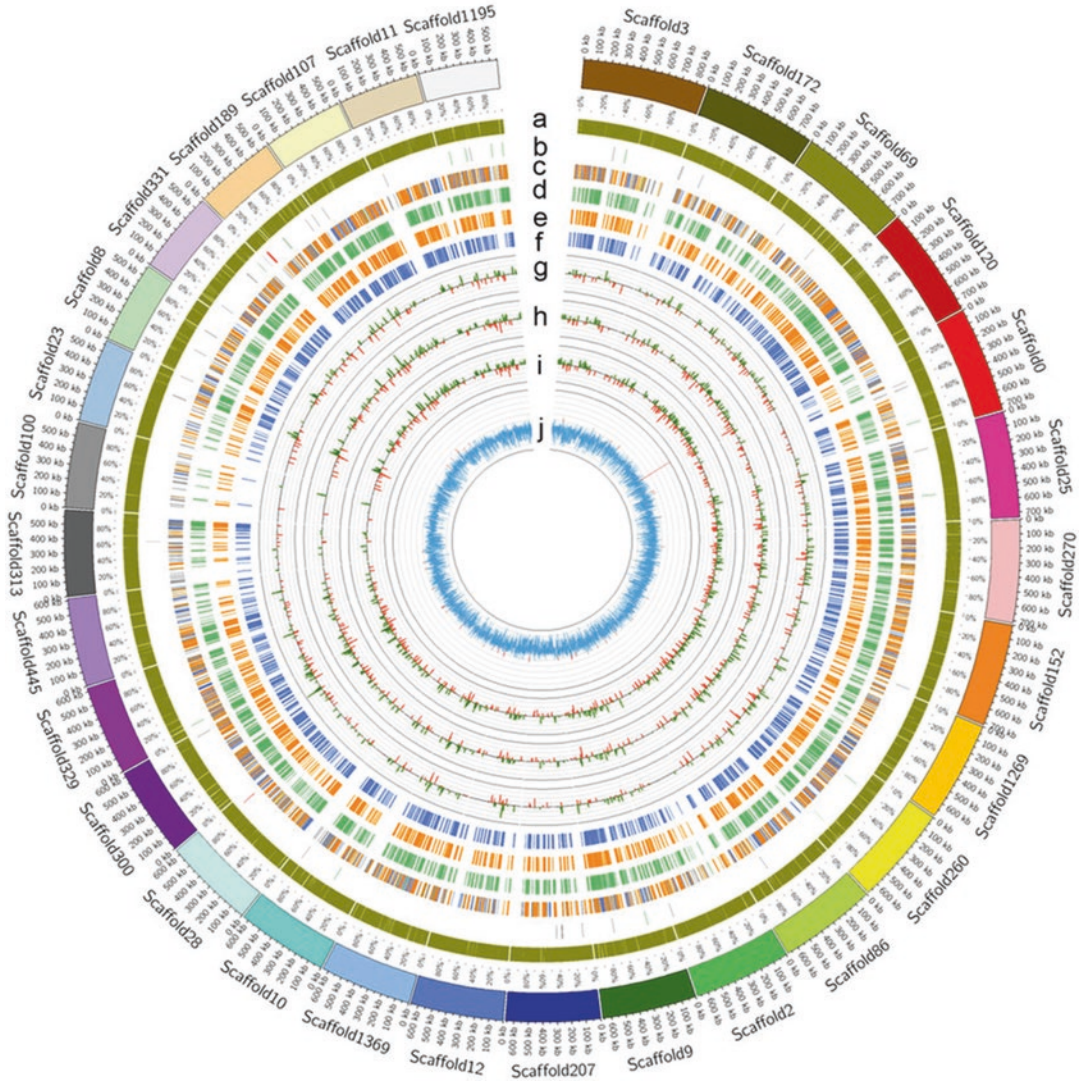

Fig. 13.2 Characterization of the *Hevea brasiliensis* genome. Circos plot of the 30 longest scaffolds. (**a**) Repeats. (**b**) Non-coding RNAs (rRNAs, tRNAs and other ncRNAs are represented by red, green and grey lines, respectively). (**c**) Gene model annotation against the NCBI non-redundant protein database (BLAST matches to *R. communis*, *M. esculenta*, *J. curcas* and other organisms are represented by blue, green, orange and grey lines, respectively). (**d**) Orthologous genes in *M. esculenta*. (**e**) Orthologous genes in *J. curcas*. (**f**) Orthologous genes in *R. communis*. (**g**) CAGE (cap analysis gene expression) tags per million (TPM) in latex (TSSs mapped to the sense and antisense strand are represented in green and red, respectively). *TSS* transcription start site. (**h**) CAGE TPM in leaf (TSSs mapped to the sense and antisense strands are represented in *green* and *red*, respectively). (**i**) CAGE TPM in bark (TSSs mapped to the sense and antisense strands are represented in *green* and *red*, respectively). (**j**) GC content (values of >50% and < = 50% are represented in *red* and *blue*, respectively) (After Lau et al. 2016) (Figure courtesy: Nyok-Sean Lau, Alexander Chong Shu-Chien, Centre for Chemical Biology, Universiti Sains Malaysia)

groups (Lau et al. 2016) (Fig. 13.3 a, b, c). The GC content of the assembled *Hevea* genome was 34.17%, similar to those of the sequenced genomes of *R. communis* (32.5%), *J. curcas* (33.7%) and *M. esculenta* (34.86%) from the Euphorbiaceae family. In addition, Lau et al. (2016) also identified 483 disease resistance genes that constitute about 0.57% of all *Hevea* genes.

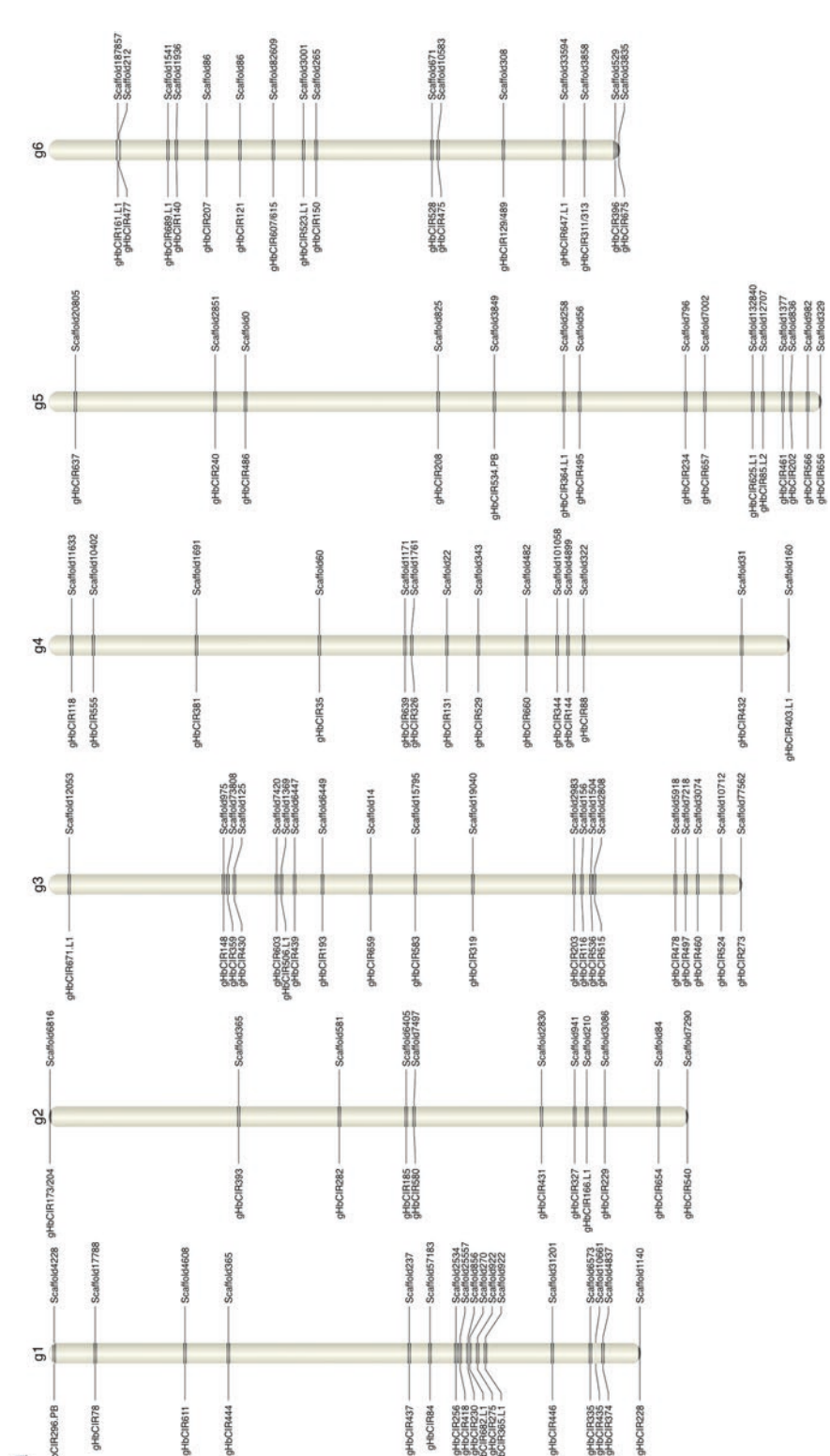

Fig. 13.3 (**a**–**c**) Anchoring the genome assembly to the linkage groups of *H. brasiliensis*. One-hundred and ninety-six RFLP markers from the published genetic map (Lespinasse et al. 2000a) were used for anchoring the scaffolds to the linkage groups. In total, 189 scaffolds that account for 43.6 Mb in length (3% of the assembly size) were anchored (After Lau et al. 2016) (Figure courtesy: Nyok-Sean Lau, Alexander Chong Shu-Chien, Centre for Chemical Biology, Universiti Sains Malaysia)

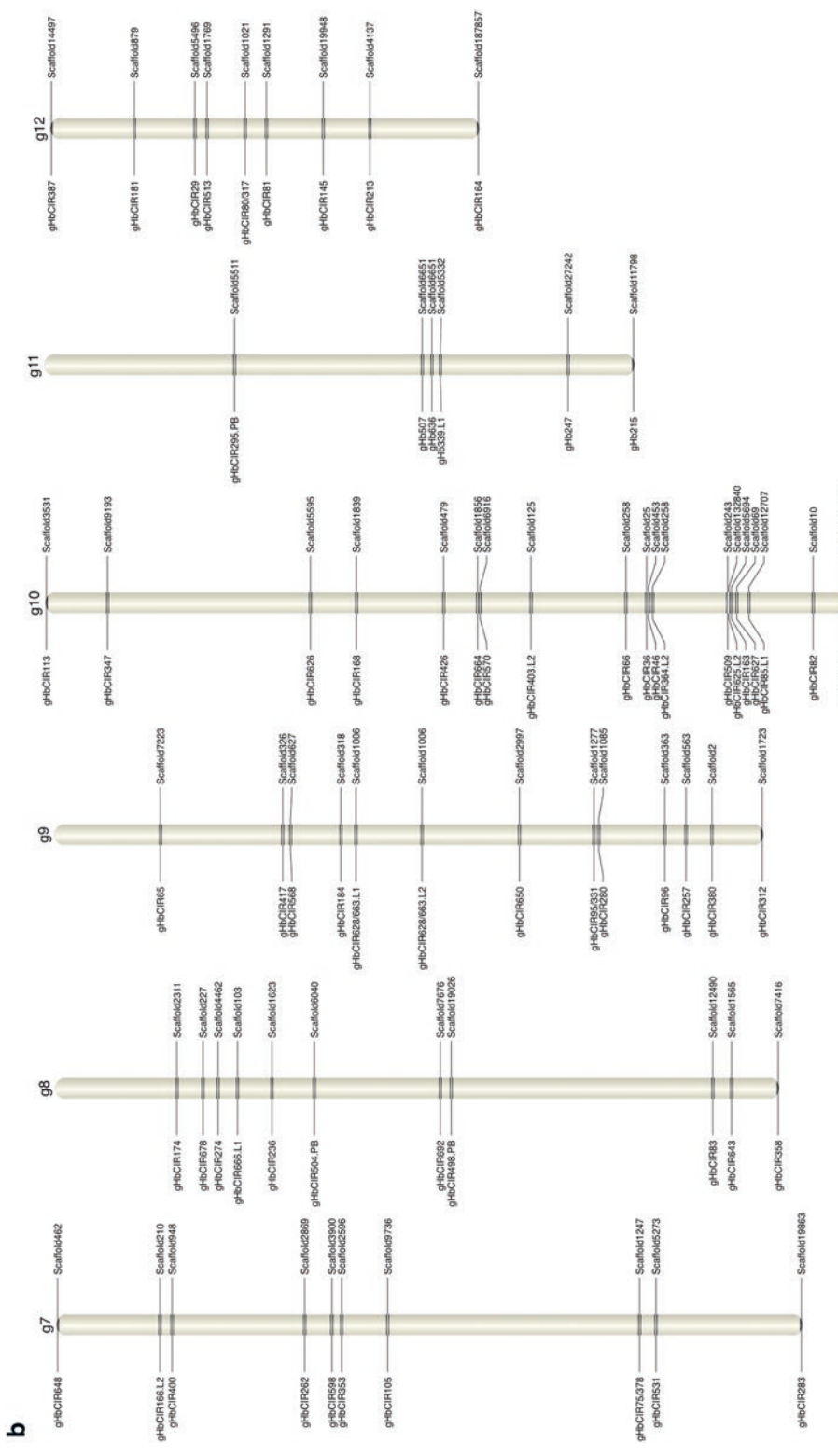

Fig. 13.3 (continued)

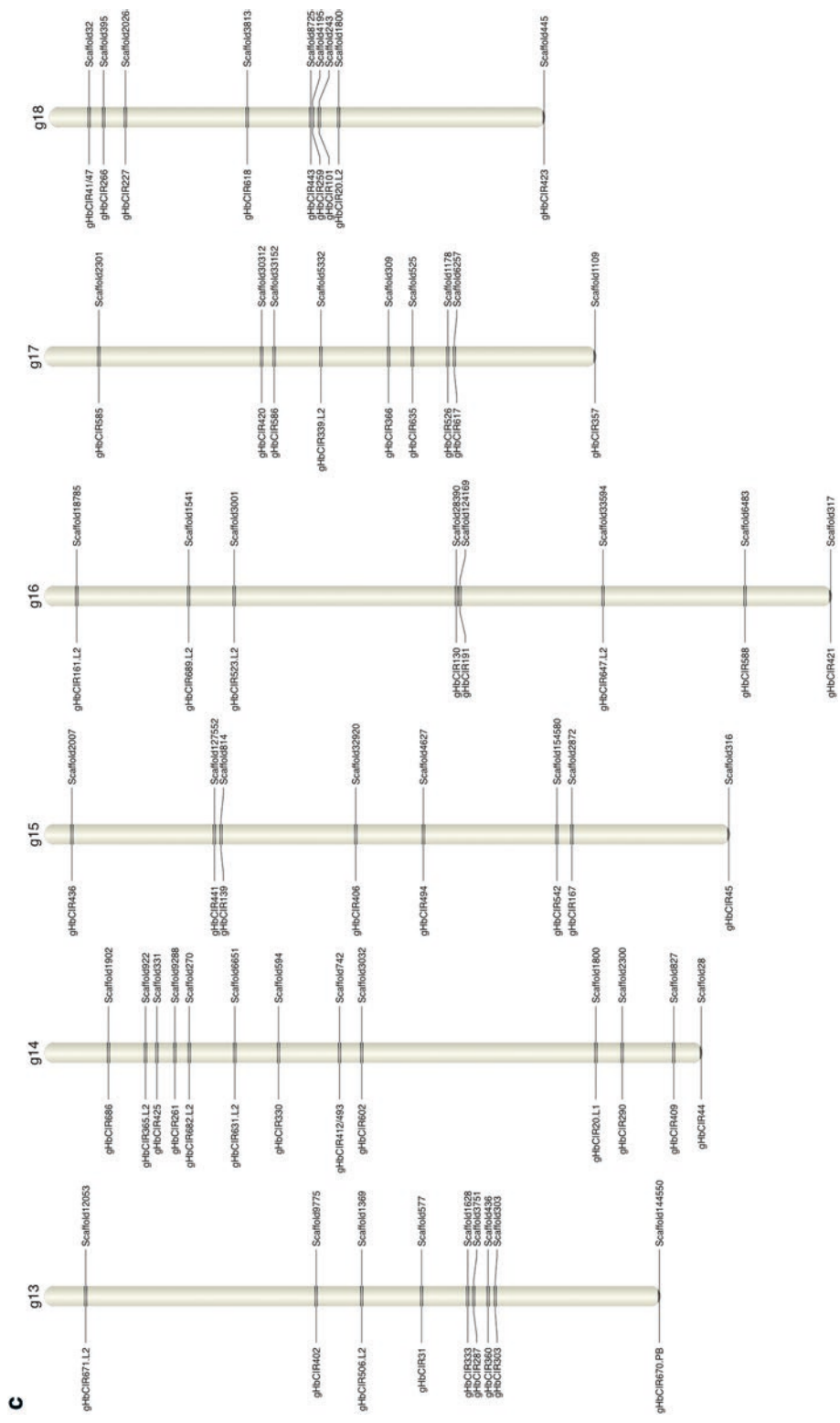

Fig. 13.3 (continued)

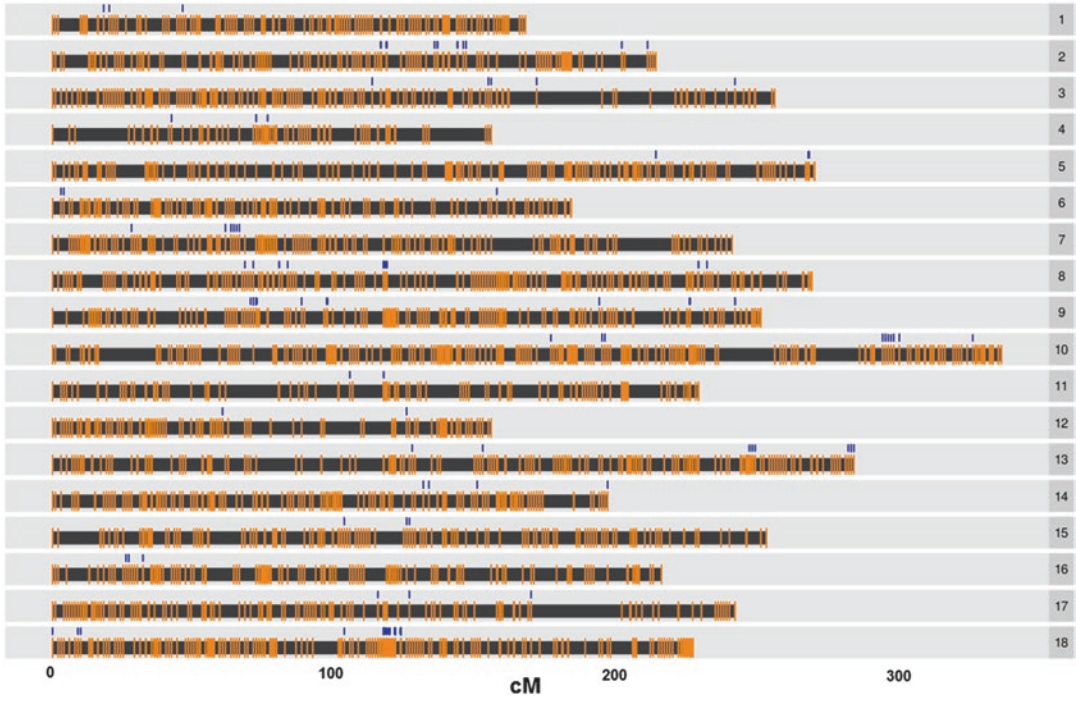

Fig. 13.4 Rubber linkage group ideograms with positions of non-redundant markers indicated as orange vertical bars and markers from within scaffolds indicated as vertical *blue* bars (After Shearman et al. 2015) (Figure courtesy: Sithichoke Tangphatsornruang, National Center for Genetic Engineering and Biotechnology, Bangkok)

According to Shearman et al. (2015), the genome assembly consists of 1,150,326 scaffolds ranging from 200 to 531,465 bp and totalling 1.1 Gb. Only 143 scaffolds, totalling 7.6 Mb, were placed into linkage groups (Rahman et al. 2013). Shearman et al. (2015) performed RNA-seq on six clones of *Hevea* (BPM24, RRII105, RRIC110, PB235, RRIT251 and RRIM600) to identify SNPs and InDels (insertion and deletion of bases in DNA) and performed target sequence enrichment to genotype a set of SNPs in 149 off-springs from a cross of RRIM600 x RRII105. They used this information to generate a linkage map by anchoring 24,424 contigs from 3009 scaffolds, totalling 115 Mb or 10.4% of the published sequence, into 18 linkage groups (Fig. 13.4). Each linkage group contains between 319 and 1367 SNPs, or 60 to 194 non-redundant marker positions, and ranges from 156 to 336 cM in length. This linkage map includes 20,143 of the 69,300 predicted genes from rubber tree and will be useful for mapping studies and improving the reference genome assembly. de Souza et al. (2016) could delineate 143 putative polymorphic positions assembled into 10,071 contigs (N50 = 3078) by a de novo assembly strategy.

Li et al. (2016a) sequenced and analysed the transcriptomes of two parent clones (RRIM600 and PR107) and their hybrids (RY 7-33-97 and RY 7-20-59) to understand their SNPs and small insertions/deletions (InDels). They could gather >31,000 genetic variations in 112,702 assembled unigenes. Their results showed that the higher yield in F_1 hybrids was positively associated with their higher genome heterozygosity in terms of SNPs and InDels. SNPs and InDels have a bearing on ethylene- and jasmonic acid–responsive genes at the transcription level.

13.6 Molecular Biology of Tapping Panel Dryness (TPD)

The two forms of TPD (reversible simple 'tapping cut dryness' related to over-exploitation of the laticifer tissue and the quite irreversible 'brown

bast' related to the development of necrosis within the bark of tapped trees) are currently addressed at the level of gene expression and protein synthesis. Dian et al. (1995) showed that (i) the latex from trees displaying tapping cut dryness exhibited five proteins specific to the cytosolic compartment of the latex cells and these were related to the disease, (ii) major changes consisted of a dramatic increase of a 14.5 kDa protein and a 26 kDa protein in diseased plants and (iii) the 26 kDa protein was linked to the coagulation process. Then Sookmark et al. (2002) observed that the two main polypeptides (here called P15 and P22) were found to accumulate in the cytosol of the TPD-affected trees; P15 and P22 were identified as *REF* (*Hev.b.1*) and *SRPP* (*Hev.b.3*), respectively. Specific molecular probes were designed for a detailed characterization of *REF* and *SRPP* gene expression and RFLP mapping. This allowed the demonstration that *REF* and *SRPP* display very similar expression profiles. They are highly overexpressed by the tapping-induced metabolic activation, although not by wounding per se, or ethylene or abscisic acid. In addition to this similarity in gene expression, they were found to share one common locus in the genome. Eventually, no significant difference in *REF* and *SRPP* gene expression was observed between healthy and TPD trees, indicating that their TPD-related accumulation in the cytosol was not transcriptionally regulated. Western blot analysis demonstrated that osmotic lysis of the sedimentable organelles (lutoids) in vitro caused the release of *REF* and *SRPP* from the rubber particle membrane into the cytosol. A mechanism of cellular delocalization as a consequence of the lutoid instability was proposed to explain *REF* and *SRPP* accumulation in the cytosol of TPD trees (Sookmark et al. 2002). In recent years, studies aimed to identify genes associated with TPD have also been carried out. Chen et al. (2003) reported that the expression of *HbMyb1* was likely to be associated with TPD syndrome. At RRII, the TPD research was focused on identification of TPD-responsive genes by SSH technology applied to mRNA isolation from latex (Venkatachalam et al. 2005). The goal of this study was to identify genes whose mRNA levels are differentially expressed in the rubber tree during TPD development. To identify the genes involved in this process, two SSH cDNA libraries were constructed. For the forward-subtracted cDNA library, healthy RNA was used as the tester and TPD RNA served as the driver, whereas TPD RNA was the tester and healthy RNA was the driver for the reverse subtracted cDNA library. A total of 1079 putatively positive clones were screened from these two libraries; 352 of these clones were positive by differential screening with forward and reverse subtracted probes and were selected for sequencing analysis. The putative functions of clones were predicted by BLASTX/BLASTN analysis. Among these, 64 were genes whose function had been previously identified while the remaining clones were genes with either unknown protein function or insignificant similarity to other protein/DNA/EST sequences in existing databases. Differentially expressed genes selected by subtractions were classified into 12 broad categories according to their putative functions generated by BLAST analysis. The possible links between the identified regulated genes and TPD syndrome were considered by dot-blot analysis and compared where two unique genes were strongly down-regulated under the TPD condition. Two genes, *Myb transcription factor* and *translationally controlled tumour-induced protein* (*TCTP*), that were unigenes to the forward-subtracted cDNA library (up-regulated) were selected for expression analysis. The expression of two selected gene transcripts was examined by Northern blot analysis using plant tissues of both healthy and TPD trees. Results from Northern analysis confirmed that the expression of these two genes was down-regulated in TPD trees. This was the first study reporting a set of suppressed genes in tapping cut dryness-affected trees by the SSH technique. Some other known genes identified in this study might provide new insights into TPD development in the rubber tree (Venkatachalam et al. 2005). Similar research based on SSH is currently being developed on brown bast (Kongsawadworakul et al. 2005). Apart from genetic engineering, studies on laticifer-specific gene expression could have important implica-

tions for selection and breeding. The use of mRNA transcript levels as molecular markers for early selection could be considered (Kush et al. 1990). It is also felt that extensive studies on the expression of genes and the regulation systems in different fields may open new paths for rubber breeding. Functional genomics in rubber will develop faster and faster, taking advantage of research developed on other species (through comparison with the information of public databases) and by focusing on specific areas of interest in order to gain a good understanding of the functioning of the network of interacting genes and regulating factors.

Eliathe et al. (2012) conducted a study to identify protein markers with yield potential and susceptibility to tapping panel dryness (TPD). Yield and susceptibility to TPD were compared in 11 clones (stimulated and non-stimulated). Their lutoid fraction polypeptides were analysed using one- and two-dimensional electrophoresis. Susceptibility to TPD appeared as a clonal trait which is not related to yield potential. TPD can occur either in stimulated or non-stimulated clones, but overstimulation increases TPD symptoms. While PB235, PB260 and IRCA130 were seen highly susceptible to TPD, IRCA41, PB217, AF261, AVROS2037 and GT1 were less susceptible. Eliathe et al. (2012) analysed 32 KDa and 35 KDa lutoidic proteins. High yielding clones with less TPD were characterized by abundant quantity of 35 KDa lutoidic polypeptide. On the contrary, clones susceptible to TPD were characterized by abundant quantity of 32 KDa polypeptide. In low-yielding clones (RO38, Tjir 1), 32 KDa protein was more abundant than 35 KDa. Overstimulation induces a decrease of 35 KDa protein intensity. This study demonstrates the utility of 32 and 35 KDa polypeptides detects yielding potential and susceptibility to TPD. Li et al. (2010) could construct forward and reverse cDNA libraries from the latex of healthy and TPD trees using suppression subtractive hybridization (SSH) method to identify the genes related to TPD. Of the 1106 clones obtained from the two cDNA libraries, 822 clones showed differential expression in two libraries by reverse Northern blot analyses. Sequence analyses indicated that the 822 clones represented 237 unique genes; and most of them have not been reported to be associated with TPD in rubber tree. The expression patterns of 20 differentially expressed genes were further investigated to validate the SSH data by reverse transcription PCR (RT-PCR) and real-time PCR analysis. According to the Gene Ontology convention, 237 unique genes were classified into 10 functional groups, such as stress/defence response, protein metabolism, transcription and post-transcription, rubber biosynthesis, etc. Among the genes with known function, the genes preferentially expressed were associated with stress/defence response in the reverse library, whereas metabolism and energy in the forward one. The genes associated with TPD were identified by SSH method in this research. Systematic analyses of the genes related to TPD suggest that the production and scavenging of reactive oxygen species (ROS), ubiquitin proteasome pathway, programmed cell death and rubber biosynthesis might play important roles in TPD. Therefore, such results not only enrich information about the genes related to TPD but also provide new insights into understanding the TPD process in rubber tree.

Trunk phloem necrosis (TPN) is known as a physiological disorder since 1980s. Distinguished from rubber tree tapping panel dryness (TPD), by its macroscopic symptoms and presumed origin, little attention has been paid to its microscopic features. de Faÿ (2011) has come out with some evidence that both syndromes could be linked to an impaired cyanide metabolism. In order to characterize TPN and compare it with TPD microscopically, the inner phloem of tapping panels was investigated by light and transmission electron microscopy in healthy trees and TPN-affected trees. TPN-affected phloem presented numerous and varied structural and ultrastructural features. Signs of cellular deterioration could be seen in a great number of specialized cells, i.e. laticifers and sieve tubes, but not in very specialized cells, i.e. parenchyma cells and companion cells. There were also signs of cellular dedifferentiation in other parenchymatous cells, e.g. in tylosoids and hyperplasic cells. These cells were derived from parenchyma cells

that ensheath laticifers in which the latex coagulated. Numerous structural features of TPN are common to TPD, notably tylosoids associated with in situ coagulated latex, which are also known to be early structural markers of TPD and cyanide-induced. de Faÿ (2011) therefore concluded that TPN is identical to or a variant of TPD and is a degenerative disorder of rubber tree trunk phloem resembling plant stress response, programmed cell death and plant tumourigenesis in some aspects.

13.7 Genomics for Changed Climates

Plant species are the best experimental material to study changes in climate since they express their reactions to such changes through significant manifestations. There will be perceptible indications in terms of morphological, physiological and genomic expressions towards such changed climates. As rubber spreads to new areas where drought and cold are conspicuous stress factors, employing genomic tools can largely help in advancing knowledge about the ability of rubber to adjust such environments. The work of Silva et al. (2014) stems promise in this direction. They studied leaf, panel and latex ESTs under cold-stressed conditions. For panel and latex libraries, samples were collected from 16-year-old tree clones of PB217, PR255, GT1, PB235, RRIM701 and IAN873, and leaves of the same clones were collected from the rubber tree germplasm. PB217, PR255, GT1 and IAN873 were subjected to a 24-h cold treatment in a growth chamber and maintained at 8 °C with a 12-h photoperiod. This treatment was performed to promote the expression of genes involved in cold response and for the development of molecular markers related to this stress condition. A total of 8263 reads were assembled, generating 5025 unigenes that were analysed; 912 expressed sequence tags (ESTs) represented new transcripts, and 2 sequences were highly up-regulated by cold stress. These unigenes were scanned for microsatellite (SSR) regions and single nucleotide polymorphisms (SNPs). In total, 169 novel EST-SSR markers were developed, of which 138 loci were polymorphic. Nearly 98% of this presented transferability to six other *Hevea* species. Locus duplication was observed in *H. brasiliensis* and other species. Additionally, 43 SNP markers in 13 sequences that showed similarity to proteins involved in stress response, latex biosynthesis and developmental processes were characterized. cDNA libraries are a rich source of SSR and SNP markers and enable the identification of new transcripts. Transcriptome analysis is one of the main approaches for identifying the complete set of active genes in a cell or tissue for a specific developmental stage or physiological condition.

13.8 Perspectives on Genomics

Application of molecular tools in rubber tree improvement was lagging behind because of limited knowledge of the genome. Initially, hybridization-based RFLP markers, providing co-dominant information, were used to characterize *Hevea* germplasm. RFLP technique was proved to be useful for genetic diversity study in wild and cultivated *Hevea* accessions using low-copy-number nuclear probes (Besse et al. 1994). RFLP analysis of organelle genomes of *Hevea* was also performed for establishing evolutionary relationships as these two genomes could reflect true evolution because of their uniparental inheritance (Luo et al. 1995). Mathew et al. (2005) studied the phylogenetic relationship among three species of rubber, *Hevea brasiliensis, H. benthamiana* and *H. spruceana*, employing different molecular marker techniques, namely, RAPD, chloroplast DNA, PCR-RFLP and heterologous chloroplast microsatellites. RAPD analysis clearly indicated a high degree of polymorphisms among the three species. For the first time, Low et al. (1996) detected microsatellites in the *Hevea* genome through the database search of some *Hevea* gene sequences. The construction of a microsatellite-enriched library in *Hevea brasiliensis* was reported by Atan et al. (1996). The detection of high degree of polymorphism indicates a way to introduce desirable variation into *H. brasiliensis* either through introgression or transformation.

Genetic linkage map presents the linear order of markers (genes and other identifiable DNA sequences) in their respective linkage groups depicting the relative chromosomal locations of DNA markers by their patterns of inheritance. The linkage map allows revelation of more and more restricted segments of the genome and undoubtedly enhances our understanding in many areas of plant systematics. A genetic map for *Hevea* spp. was constructed using a population derived from an interspecific cross between PB260 (*H. brasiliensis*) and RO38, an interspecific hybrid clone (*H. brasiliensis* × *H. benthamiana*), following the pseudo-testcross strategy (Lespinasse et al. 2000a). The markers were assembled into 18 linkage groups, thus reflecting the basic chromosome number, and covered a total distance of 2144 cM. A total of 717 loci constituted the synthetic map, including 301 restriction fragment length polymorphisms, 388 amplified fragment length polymorphisms, 18 microsatellites and 10 isoenzymes. Homologous linkage groups between the two parental maps were merged using bridge loci. Average marker density was 1 per 3 cM. Lespinasse et al. (2000b) mapped quantitative trait loci (QTLs) for resistance to South American leaf blight (SALB), a disease of the rubber tree caused by the fungus *Microcyclus ulei* using the same cross combination (PB260, a susceptible clone, and RO38, a SALB-resistant clone). Eight QTLs for resistance were identified on the RO38 map, whereas only one QTL was detected on the PB260 map. New linkage maps were added by Lau et al. (2016) and Shearman et al. (2015).

Transcriptome sequencing and development of microarrays have been undertaken recently in *Hevea* rubber (Triwitayakorn et al. 2011; Salgado et al. 2014). Sequencing of transcriptomes of bark that leads to EST-SSR markers is also of prime importance (Li et al. 2012; Cubry et al. 2014) that calls for rigorous research. Such developments are certainly welcome that elevates *Hevea* rubber research on par with other tropical tree crops. However, such innovations must help to find answers to intriguing issues like tapping panel dryness (TPD), stock-scion interactions and yield differences exhibited among trees raised through bud grafting, molecular markers for selecting high yielders at juvenile stage, delineation of parents of open-pollinated seedlings, production of natural somatic seeds and so on.

Transcriptomes of CATAS8-79 and PR107 have been sequenced to dissect the molecular mechanism for the regulation of latex regeneration and duration of latex flow (Chao et al. 2015). More than 26 million clean reads were generated and 51,829 all-unigenes were totally assembled. A total of 6726 unigenes with differential expression patterns were detected between CATAS8-79 and PR107. Expression pattern of genes upon successive tapping was analysed by quantitative PCR. Several genes related to rubber biosynthesis, cellulose and lignin biosynthesis and rubber particle aggregation were differentially expressed between CATAS8-79 and PR107. The level of endogenous jasmonates, carbohydrate metabolism, hydroxymethylglutaryl-CoA reductase (HMGR) and *Hevea* rubber transferase (HRT) are pertinent in mevalonate (MVA) pathway for latex regeneration. On the other hand, level of endogenous ethylene (ETH), lignin content of laticifer cell wall, antioxidants and glucanases are pivotal for duration of latex flow (Chao et al. 2015) (Fig.13.5). In RRIM600 alone, approximately 10,000 DNA sequences representing genes expressed in the latex have been delineated through DNA sequencing technology (Mat-isa et al. 2009). They developed *NRESTdb* (Natural Rubber EST database) to provide easy access and rapid analysis of such data as the first publicly available EST database for *H. brasiliensis*. Such studies are to be augmented further to draw feasible answers for questions on latex production, laticifer-specific gene expression and tapping panel dryness.

Lately, there was a comparative evaluation between self-rooting juvenile clones (JCs) and bud-grafted (donor) clones (DCs) at transcriptome level (Li et al. 2016b). Genes, especially encoding epigenetic modifications, are differentially expressed in JCs and DCs. Genes involved in carbohydrate metabolism, hormone metabolism and reactive oxygen species scavenging were up-regulated in JCs of CATAS7-33-97 and Haiken 2, indicating that the JCs provide

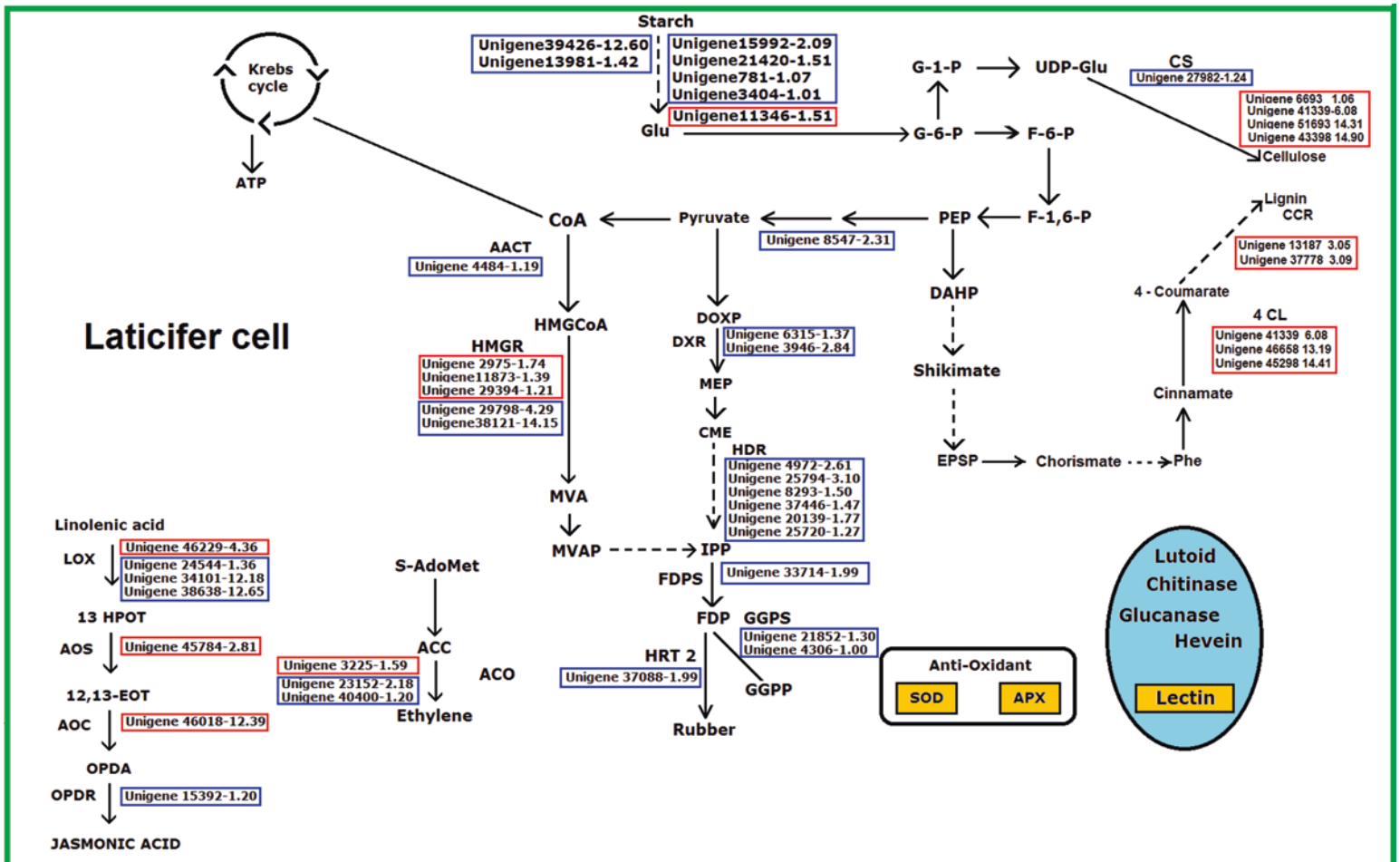

Fig. 13.5 Schematic representation of differentially expressed unigenes related to latex regeneration and duration of latex flow in laticifer cell. Unigenes up-regulated in PR107 are shown in red boxes but down-regulated in CATAS8-79. Those in *blue* boxes are unigenes up-regulated in CATAS8-79 but down-regulated in PR107. Numbers represented Log2 values in DGE data. The *dashed arrows* indicate multiple steps of enzymatic reactions (After Chao et al. 2015)

sufficient molecular basis for the increased rubber yield. Comparative trial between self-rooting JCs and DCs, self-rooting JCs exhibited better performance in rubber yield (Li et al. 2016b). Such investigation clearly indicates the intricateness of stock-scion interactions. Decoding the genes responsible for apomixis or parthenocarpy (Hand and Koltunow 2014) and their introduction into Hevea genome to achieve somatic seeds is the futuristic goal in *Hevea* genomic research.

The *Hevea* genome has now been published three times over. Yet not everyone comes up with the same findings. Rahman et al (2013) do not offer the newest *Hevea* genome, and the honour goes to the RIKEN team of Japan, working in collaboration with Universiti Sains Malaysia (USM) for a more comprehensive genome analysis (Lau et al. 2016). However, the most comprehensive genome to date comes from Chinese Academy of Tropical Agricultural Sciences (CATAS) group in China (Tang et al. 2016). CATAS declares the genome size as 1.47 Gb (Tang et al. 2016), whereas USM and RIKEN/USM both give the figure of 2.15 Gb. Despite this discrepancy, it is noteworthy that the CATAS assembly captures practically all the USM sequences contained in a purportedly larger draft genome. It appears that CATAS has done a better job of fitting contiguous sequences into a smaller number of scaffolds (Tang et al. 2016). Thus, while USM has a scaffold N50 of only 67.2 kb, CATAS weighs in with a massive scaffold N50 of 1.28 Mb. Despite three published *Hevea* genomes now in the public domain, the discrepancies between the reports have the logic that the last word is still not yet in.

Genomics, no doubt, is a science that can make inroads into the intricacies of gene actions in *Hevea*. The only care to be observed is to continue the studies with a set goal so that the plant breeder will get ample assistance to derive new clones at definite intervals, as also in tandem with circumstantial demands.