Higher Genomic Variation in Wild Than Cultivated Rubber Trees, *Hevea brasiliensis*, Revealed by Comparative Analyses of Chloroplast Genomes

Li-Ying Feng†, Jin Liu‡, Cheng-Wen Gao‡, Hai-Bo Wu‡, Guo-Hua Li§ and Li-Zhi Gao†,*

1 Institution of Genomics and Bioinformatics, South China Agricultural University, Guangzhou, China, 2 Plant Germplasm and Genomics Research Center, Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, 3 Yunnan Institute of Tropical Crops, Xishuangbanna, China

Rubber tree is the only commercialized natural resource to produce high-quality natural rubber with unique physical and chemical properties. They currently foster in Southeast Asia with marked morphological and productive differences with wild germplasms native to the Amazonian basin of South America. Here, we report complete chloroplast (cp) genomes of six cultivated and six wild accessions of *H. brasiliensis* using Illumina paired-end sequencing platform. The 12 *H. brasiliensis* cp genomes ranged from 161,168 to 161,254 bp. All 12 cp genomes displayed a typical quadripartite structure, which consisted of a pair of IR regions (26,787–26,804 bp) separated by a LSC region (89,216–89,284 bp) and a SSC region (18,370–18,377 bp). Phylogenomic analysis revealed that cultivated and wild rubber trees failed formed separate clades. However, we observed that wild rubber trees possessed more variable sites and ~2.8-fold higher level of nucleotide variation than cultivated rubber trees despite a short domestication history. We drew a comprehensive map of genomic variation across rubber tree plastomes, exhibiting that the density of genomic variants in wild rubber trees was slightly higher than that detected in cultivated ones. The obtainability of genomic variation across cp genomes will provide useful information for better conserving and utilizing rubber tree germplasms.

**Keywords:** Hevea brasiliensis, chloroplast genome, rubber tree, comparative genomics, phylogenetic analysis

**INTRODUCTION**

Chloroplasts are key organelles that play a key role in plant cell for photosynthesis and biochemical pathways, including the biosynthesis of starch, fatty acid, pigments, and amino acids (Raman and Park, 2015; Dong et al., 2016; Xu et al., 2017). In angiosperms, chloroplast (cp) genomes contain a circular DNA ranging from 72 to 217 kb in size (Sugiura, 1992, 1995; Shi et al., 2012), consisting of two large inverted repeat (IR) regions, a large-single-copy (LSC) region and a small-single-copy (SSC) region (Bendich, 2004; Raubeson and Jansen, 2005; Yang et al., 2010, 2017). Moreover, cp genomes usually exhibit maternal inheritance, enabling the conservation of gene
content and genome structure (Sasaki et al., 2003; Parks et al., 2009). A recent study reported that the entire plastome is transcribed in photosynthetic green plants, and that this pattern originated from prokaryotic cyanobacteria – ancestor of the chloroplast genomes that diverged about 1 billion years ago (Shi et al., 2016). Concatenating sequences from a large number of chloroplast genes may overcome the problem of multiple substitutions that results in the loss of phylogenetic information between chloroplast lineages, and thus, can reduce gene-sampling errors due to substitutions. Owing to their small sizes, relatively conserved genome structure, gene content, gene order and simple complexity compared to mitochondrial and particularly nuclear genomes, the cp genomes have recently provided valuable information for taxonomic classification, phylogenetic reconstruction and adaptive evolution as a result of sequence divergence among plant species (e.g., Huang et al., 2014; Zhang et al., 2016; Gao et al., 2019). Given decreased costs to generate increasingly large amount of genome sequences using the Next Generation Sequencing (NGS) platform, up to now (Mardis, 2008), approximately 3,869 plant cp genomes have been sequenced and deposited in the National Center for Biotechnology Information (NCBI), including the cp genome of rubber tree, Hevea brasiliensis (Wild.) Muell. Arg (Tangphatsornruang et al., 2011).

H. brasiliensis is a diploid plant (2n = 2x = 36), a member of the family Euphorbiaceae, which was most commonly known as the rubber tree (Leitch et al., 1998; Lau et al., 2016). It is a cross-pollinated tropical economical tree that grows to 30–40 m tall and can live up to 100 years in the wild (Priyadarshan and Clément-Demange, 2004). The domestication of the rubber tree, which is native to the Amazon basin in South America, began in 1896 and then spread to Southeast Asia with the relocation of H. brasiliensis seedlings (Chan, 2000). Over 100 year of traditional breeding has greatly increased rubber productivity from its wild populations (Priyadarshan and Clément-Demange, 2004). Nevertheless, although artificial selection has led to a slow growth in rubber productivity, it is even more immediately essential to raise new H. brasiliensis varieties with desired agronomic traits. Rubber trees produce natural rubber that is still required for heavy tires. Thus, there will be huge potential to be exploited through genetic breeding programs to help enlarge genomic diversity, important for the generation of more environmentally resilient and high-yielding varieties (Sneller et al., 1997). With this regard, long-standing efforts to obtain high-quality reference nuclear genome assemblies (Rahman et al., 2013; Lau et al., 2016; Tang et al., 2016; Pootakham et al., 2017; Liu et al., 2020) and the mitochondrial genome (Shearman et al., 2014) have long put to apply for genomics-assisted selection to make use of the unexploited reservoir of novel alleles from the wild.

In this study, we sequenced, assembled and characterized the cp genomes of six cultivated and six wild rubber trees using the next-generation sequencing platform. We comprehensively characterized the organization, gene content, intraspecific diversity and structural variants across the 12 cp genomes of rubber tree. The obtained results may largely help to improve our understanding of cp genome evolution of rubber tree and will provide abundant genetic resources to assist the exploration of the precious H. brasiliensis wild germplasms for future rubber tree genetic improvement program.

MATERIALS AND METHODS

Plant Material Sampling
We selected a panel of world-wide representative rubber trees, including six cultivated and six wild accessions, respectively (Table 1). The six wild H. brasiliensis accessions were originally collected from the Amazon River Basin of Brazil; three were collected from Acre, while the three other were sampled from Rondónia. The six H. brasiliensis cultivars were bred in Indonesia, Malaysia, Brazil, India, Vietnam and China, respectively (Table 1). Fresh leaves of H. brasiliensis germplasms were sampled and immediately dried with silica gel for subsequent experiments.

DNA Extraction, Genome Sequencing, and Assembly
Total genomic DNA for each sample was extracted from ~100 mg dried leaves using the Mag-MK Plant Genomic DNA extraction kit (Sanon Biotech, CA). The DNA concentration was quantified using Life Invitrogen Qubit® 3.0 (Life Invitrogen, United States), and DNA concentration >30 ng µL⁻¹ was finally determined using Bioanalyzer 2100 (Agilent Technologies). Short-insert paired-end sequencing libraries were generated according to the Illumina standard protocol. Genome sequencing was performed on Hiseq 2500 sequencing platform (Illumina, San Diego, California, United States). Approximately 5.0 GB of raw data were generated with read length of pair-end 100 bp. First, raw reads were trimmed to obtain the high-quality clean data by removing adaptor sequences and low-quality bases with Q-value ≤20 using CLC-quality trim tool1. The cp genome reads were isolated from the mixed DNA of nucleus, mitochondria and chloroplast based on the previously published rubber tree cp genome sequence (Tangphatsornruang et al., 2011). Then, the filtered reads were assembled into contigs using CLC genome assembler v4.06, which were aligned (≥90% similarity and query coverage) and ordered according to the earlier reported cp genome (Tangphatsornruang et al., 2011). Finally, the draft cp genome for each rubber tree was assembled by Geneious 9.0.5 software2, and genomic regions with ambiguous alignments were trimmed manually and considered as gaps, which were filled by sequencing fragments yielded by Polymerase Chain Reaction (PCR) using Geneious assembly software.

Genome Annotation and Whole Genome Comparisons
Initial gene annotation of the 12 rubber tree cp genomes was performed by Organellar Genome Annotator (Wyman et al., 2004). The annotation was manually corrected for start/stop

1http://www.clcbio.com/products/clc-assemblycell/
2http://www.geneious.com
TABLE 1 | Summary of the cultivated and wild rubber tree germplasms in this study.

| No. | Accession number   | Type | Origins          | No. | Accession number   | Type | Origins          |
|-----|--------------------|------|------------------|-----|--------------------|------|------------------|
| 67  | RO/PB/1 2/194      | Wild | Rondônia, Brazil | 80  | PR228              | Cultivated | Indonesia |
| 68  | AC/F/6B 40/166     | Wild | Acre, Brazil     | 84  | RRIM600            | Cultivated | Malaysia |
| 73  | RO/CM/10 44/232    | Wild | Rondônia, Brazil | 87  | IAN873             | Cultivated | Brazil   |
| 75  | RO/CM/10 44/162    | Wild | Rondônia, Brazil | 88  | RRI105             | Cultivated | India   |
| 78  | AC/B/19 56/42      | Wild | Acre, Brazil     | 89  | IRC122             | Cultivated | Vietnam |
| 79  | AC/F/5 21/258      | Wild | Acre, Brazil     | 90  | Reyan8-79          | Cultivated | Hainan, China |

FIGURE 1 | Gene map of the *H. brasiliensis* chloroplast genomes. Genes lying outside of the outer circle are transcribed in the clockwise direction whereas genes inside are transcribed in the counterclockwise direction. Genes belonging to different functional groups are color-coded. Area dashed darker gray in the inner circle indicates GC content while the lighter gray corresponds to AT content of the genome.
TABLE 2 | Statistic of chloroplast genome features of H. brasiliensis germplasms.

| Species | Genome size bp | LSC length bp | SSC length bp | IR length bp | GC content (%) | Number of genes | Protein-coding genes | Structure RNAs | GenBank accessions |
|---------|----------------|---------------|---------------|--------------|----------------|-----------------|--------------------|-------------------|-------------------|
| 67      | 161,215        | 89,267        | 18,372        | 26,788       | 35.72          | 112             | 78                 | 34                | KY363216          |
| 68      | 161,254        | 89,274        | 18,372        | 26,804       | 35.72          | 112             | 78                 | 34                | KY363217          |
| 73      | 161,232        | 89,276        | 18,373        | 26,792       | 35.71          | 112             | 78                 | 34                | KY363218          |
| 75      | 161,236        | 89,284        | 18,374        | 26,789       | 35.71          | 112             | 78                 | 34                | KY363219          |
| 78      | 161,216        | 89,266        | 18,370        | 26,790       | 35.72          | 112             | 78                 | 34                | KY363220          |
| 79      | 161,224        | 89,272        | 18,376        | 26,788       | 35.72          | 112             | 78                 | 34                | KY363221          |
| 80      | 161,222        | 89,274        | 18,374        | 26,787       | 35.72          | 112             | 78                 | 34                | KY363222          |
| 84      | 161,222        | 89,275        | 18,373        | 26,787       | 35.72          | 112             | 78                 | 34                | KY363223          |
| 88      | 161,225        | 89,277        | 18,372        | 26,788       | 35.72          | 112             | 78                 | 34                | KY419133          |
| 89      | 161,168        | 89,216        | 18,376        | 26,788       | 35.73          | 112             | 78                 | 34                | KY419135          |
| 90      | 161,170        | 89,217        | 18,377        | 26,789       | 35.72          | 112             | 78                 | 34                | KY419136          |

TABLE 3 | Genes contained in the cultivated and wild rubber tree chloroplast genomes.

| Category          | Group of genes                  | Name of genes                                                                 |
|-------------------|--------------------------------|-------------------------------------------------------------------------------|
| Self-replication  | Ribosomal RNA genes             | mr4.5, m5, m6, m23                                                           |
|                   | Small subunit of ribosome       | rps2, rps3, rps4, rps7, rps8, rps11, rps12**, rps14, rps15, rps16*, rps18, rps19 |
|                   | Large subunit of ribosome       | rpl2**, rpl14, rpl16*, rpl20, rpl22 c, rpl23, rpl32, rpl33, rpl56            |
|                   | Transfer RNA genes              | trnR-UCU, trnS-GCU, trnA-UGC*, trnC-GCA, trnF-GAA, trnG-GCC*, trnG-UGC, trnH-UCC, trnI-UCC, trnM-GUC, trnN-UUC, trnP-UAG, trnQ-UUG, trnR-ACG, trnS-GAU*, trnY-GUA, trnK-UUU*, trnL-UAA*, trnL-CAU, trnM-GAC, trnV-UAC*, trnW-CCA, trnW-UAG, trnM-CAU, trnM-CAU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-GGU |
| Photosynthesis    | DNA dependent RNA polymerase    | rpoA, rpoB, rpoC1*, rpoC2                                                  |
|                   | Subunits of photosystem?        | psaA, psaB, psaC, psaI, psaJ, ycf3**, ycf4                                   |
|                   | Subunits of NADH-dehydrogenase  | ndhA*, ndhB*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK            |
|                   | Subunits of ATP synthase        | atpA, atpB, atpE, atpF*, atpH, atpL                                        |
|                   | Subunits of photosystem?        | psaA, psaB, psaC, psaD, psaE, psbF, psbH, psbI, psbJ, psbK, psbL, psbN, psbT, psbZ |
|                   | Subunits of cytochrome complex  | petA, petB*, petD*, petG, petL, petN                                         |
|                   | Protease                       | ctpP**                                                                       |
| Other genes       | Maturase                       | matK                                                                         |
|                   | Acetyl-CoA-carboxylase c-type cytochrome synthesis gene | ccsA                                                                         |
|                   | Large subunit of rubisco        | rbcl                                                                         |
|                   | Envelop membrane protein        | cemA                                                                         |
|                   | Subunit of Acetyl-CoA-carboxylase | accD                                                                         |
|                   | Hypothetical chloroplast        | ycf1, ycf2, ycf15                                                           |

*Genes containing a single intron; **genes containing two introns.

codons and intron/exon orders. BLASTX and BLASTN searches were employed to accurately annotate protein-coding genes and identify the location of the ribosomal RNA (rRNA) and transfer RNA (tRNA) genes. The annotation results were manually checked and codon positions were adjusted by comparing to homologous genes especially from those closely related cp genomes in the public database; boundaries between introns and exons, positions of start and stop codons for protein-coding genes were accurately identified. The cp genome map was drawn using Genome Vx software (Conant and Wolfe, 2008), and cp genome sequences of the 12 rubber trees were deposited in GenBank. The shrinkage and expansion of the IR region is detected by online software IRscope (Amiryousefi et al., 2018). The comparative analysis between genomes was carried out using mVISTA (Frazer et al., 2004). The average pair wise sequence divergence within large single copy (LSC), small single copy (SSC) and inverted repeat (IR) regions was calculated between the cultivated and wild H. brasiliensis cp genomes using MEGA v6.0 (Tamura et al., 2013). The complete cp genome sequences of the four other spurge plastomes, including Euphorbia esula, Jatropha curcas, Manihot esculenta, and Ricinus communis, were downloaded from GenBank, which were further
FIGURE 2 | Number and distribution of simple sequence repeats (SSRs) in the 12 *H. brasiliensis* cp genomes. (A) Number of SSR types in the 12 *H. brasiliensis* chloroplast genomes; (B) distribution of SSRs in the 12 *H. brasiliensis* chloroplast genomes.

TABLE 4 | Levels of nucleotide variation in the cultivated and wild *H. brasiliensis* cp genomes.

| Species                  | Number of sites | Number of variable sites | Number of informative sites | Nucleotide diversity |
|--------------------------|-----------------|--------------------------|----------------------------|----------------------|
| *H. brasiliensis* Wild   | 161336          | 210 (0.13%*)             | 51 (0.03%**)               | 0.0005               |
| *H. brasiliensis* Cultivated | 161274        | 54 (0.03%*)              | 52 (0.03%**)               | 0.00018              |

*The percentage of variable sites in the number of sites. **The percentage of informative sites in the number of variable sites.

compared with the 12 wild and cultivated *H. brasiliensis* cp genomes generated in this study.

**Repeat Sequence Annotation**

Simple sequence repeats including SSRs were predicted using MISA (Thiel et al., 2003) with the parameters set to 10 repeat units ≥10 for mononucleotide SSRs, six repeat units ≥6 for dinucleotide, five repeat units ≥5 for trinucleotide, four repeat units ≥4 for tetranucleotide, and three repeat units ≥3 for pentanucleotide and hexanucleotide SSRs, respectively.

**Detection of Genomic Structural Variants**

To assess levels of genomic diversity variable and parsimony-informative base sites and nucleotide diversity ($P_i$) were calculated for the cultivated and wild *H. brasiliensis* cp genomes using DnaSP version 6.1 (Rozas et al., 2017). To examine microstructural variants among the *H. brasiliensis* cp genomes, numbers and positions of insertions/deletions (indels) and SNP across the 12 cp genome sequences were detected using Mummer\(^3\). The number of SNPs and Indels were identified using BCFTools (Li, 2011). Circos was used to display the detected chloroplast structural variants (Krzywinski et al., 2009).

**Phylogenomic Analyses**

The phylogenomic analyses were performed for complete cp genome sequences of the 12 cultivated and wild *H. brasiliensis* accessions using *J. curcasas* as outgroup. Neighbor-joining (NJ) analysis was executed using MAFFT version 7.221 (Katoh et al., 2005), while Maximum likelihood (ML) analyses of the six Euphorbiaceae species using *Gossypium barbadense* as outgroup were conducted using MEGA 7.0 (Kumar et al., 2016). RAxML (Stamatakis et al., 2008) searches relied on the general time reversible (GTR) model of nucleotide substitution with the gamma model of rate heterogeneity. Non-parametric

\(^3\)http://www.tigr.org/software/mummer/
bootstrapping test was completed using 500 replicates as implemented in the “fast bootstrap” algorithm of RAxML (Stamatakis et al., 2008). The aligned data matrices are available upon request.

RESULTS AND DISCUSSION

Chloroplast Genome Sequencing and Assembly

We employed the Illumina short-read technology with paired-end libraries on the HiSeq2000 sequencing platform to resequence them. For each *H. brasiliensis* accession, ∼5 GB raw reads were generated with an average length of ∼100 bp (Liu et al., 2020). Paired-end reads were mapped to the published cp genome of *H. brasiliensis* (Tangphatsornruang et al., 2011) (GenBank Accession Number: HQ285842). Gaps were validated using PCR-based experiments and then sequenced on ABI sequencing platform. All 12 cp genome sequences were deposited into GenBank with the following accession numbers: KY363216, KY363217, KY363218, KY363219, KY363220, KY363221, KY363222, KY363223, KY419133, KY419134, KY419135, and KY419136.

Characterization of cp Genome Features

Nucleotide sequences of the 12 *H. brasiliensis* cp genomes ranged from 161,168 bp (IRCI22) to 161,254 bp (AC/F/6B 40/166) (Figure 1 and Table 2). All 12 cp genomes displayed a typical
The chloroplast genomes of *Hevea brasiliensis* have a quadripartite structure, which consisted of a pair of IR regions (26,787–26,804 bp) separated by a LSC region (89,216–89,284 bp) and a SSC region (18,370–18,377 bp). The GC content was almost identical each other among the 12 *H. brasiliensis* cp genomes. GC contents for eight (67, 68, 78, 79, 80, 84, 87, and 88), two (73 and 75), and two (89 and 90) cp genomes were 35.72, 35.71, and 35.73%, respectively, with an average GC content of 35.72%. If duplicated genes in IR regions were counted only once, the 12 *H. brasiliensis* cp genomes harbored 112 genes in the same order, including 78 protein-coding genes and 34 RNA genes. Overall, the genomic structure, including gene number and order, were well-conserved across the 12 cp genomes (Tables 2, 3). Among these 112 unique genes, 15 (*trnA-UGC, trnI-GAU, trnG-GCC, trnK-UUU, trnV-UAC, trnL-UAA, rpl2, rpl16, petD, petB, rpoC1, atpF, ndhA, ndhB, and rps16*) had one intron, while three (*ycf3, clpP, and rps12*) contained two introns (Table 3). The rps12 gene with 3’ end exon and intron located in the IR regions, and the 5’ end exon in the LSC region. Furthermore, *matK* was located within the intron of *trnK-UUU*. Overlaps of adjacent genes were found in the chloroplast genomes, for example, *atpE-atpB* had a 4-bp intersecting region, and *psbD-psbC* had a 53-bp overlapping region. Unusual initiator codons were observed in *rps19* with TGA and *ndhD* with TGG. Of the 61 tRNAs encoded 20 amino acids, 30 tRNAs were encoded in the rubber chloroplast genome, and the remaining 31 tRNAs may be encoded in the nuclear genome.

### Repeat Sequences in the 12 *H. brasiliensis* cp Genomes
Chloroplast simple sequence repeats (cpSSRs) are usually applied to characterizing genetic diversity and performing phylogenetic analyses (Flannery et al., 2006; Yang et al., 2011; Jiao et al., 2012). In total, we detected 105–113 SSRs in the examined *H. brasiliensis* cp genomes (Figure 2A and Supplementary Table S1A). Our results showed slightly more SSRs in wild rubber tree cp genomes than cultivated rubber three cp genomes. It is clear that mononucleotides were the most abundant while tetranucleotides were the lowest across these cp genomes (Figure 2A and Supplementary Table S1A). Mononucleotide repeat (A/T) was the most abundant, ranging from 90 to 97 across the examined *H. brasiliensis* cp genomes (Figure 2B and Supplementary Table S1B).

### Genomic Structural Variation Between Cultivated and Wild Rubber Tree Plastomes
We assessed levels of nucleotide variation between cultivated and wild rubber tree plastomes. Our results showed that wild rubber trees possessed more variable sites and ~2.8-fold higher level of nucleotide variation than cultivated rubber trees (Table 4). This finding highly supports our former population genomic analysis based on nuclear genome resequencing of these rubber tree germplasms (Liu et al., 2020). With these 12 cultivated and wild rubber tree cp genomes at hand, we drew a comprehensive map of their precise genomic variation on the basis of microstructural variants. Using the wild rubber tree cp genome (67) (Table 1) as a reference, we totally identified 725 SNPs and 493 InDels (Figure 3 and Supplementary Table S2). Our results showed that the wild rubber tree (78) from Acre, Brazil harbored the largest number (193) of genomic variants, while rubber tree cultivar (Reyan8-79) had the fewest number of genomic variants (91). We interestingly observed that the density of genomic variants was as high as ~3.86/Kb in wild rubber trees, which is slightly larger than ~3.71/Kb detected in cultivated rubber trees. It is clear that...
FIGURE 5 | Alignment of the six chloroplast genome sequences. VISTA-based identity plot shows sequence identity among the six Euphorbiaceae species using *Hevea brasiliensis* cp genome (67) as a reference. The vertical scale designates the percentage of identity, ranging from 50 to 100%. The horizontal axis indicates the coordinates within the chloroplast genome. Genome regions are color-coded as exons, introns, and conserved non-coding sequences (CNS).

FIGURE 6 | Phylogenetic analyses of the 12 *H. brasiliensis* accessions and the five Euphorbiaceae species. (A) Phylogenetic tree of the 12 *H. brasiliensis* accessions inferred from complete nucleotide sequences of the chloroplast genomes; (B) ML phylogenetic tree of the five Euphorbiaceae species.
genomic variants differently occurred across cultivated and wild rubber tree cp genomes, further supporting the observed patterns of genomic variation detected in the genus *Oryza* (Gao et al., 2019). In gene regions, such as *rpoC1*, there were a number of genomic variants detected in wild rubber trees, while cultivated rubber trees showed no genomic variation.

**Expansion and Contraction Among *H. brasiliensis* and the Other Spurge Plastomes**

Although the cp genome size, structure, gene order and gene number are well-conserved, the junctions between IR regions and single-copy (SC) boundary regions were considered as a primarily mechanism causing the length variation of angiosperm cp genomes (Huang et al., 2014; Liu et al., 2018). To elucidate the potential expansion and/or contraction of IR regions, we assessed the variation in the IR/LSC and IR/SSC boundary regions across the six spurge plastomes, including cultivated and wild *H. brasiliensis* plastomes. The junctions between IR and SSC regions were highly conserved between the cultivated and wild *H. brasiliensis* plastomes. The genes (ycf1 and rps19-rpl2-rpl22-trnH) were located within the junctions of SSC/IRa and LSC/IRb regions (Figure 4). Two copies of the ycf1 gene crossed SSC/IRa and SSC/IRb. Compared to the relatively fixed location of the ycf1 gene across all cp genomes, the LSC/IR boundary regions were seemingly variable. The rps19 gene in *M. esculenta* and *H. brasiliensis* overlapped the LSC/IRb region with 186-bp and 96-bp located at the IRb region, the rpl22 gene in *R. communis* spanned the LSC/IRb region with 26-bp located at the IRb region, and the intergenic spacer of rps19-rps12 extended 18-bp or 71-bp to the LSC region in *E. esula* and *J. curcas*, respectively. The junction of LSC/IRa regions was located in the intergenic spacer of rps19-trnH (*R. communis, M. esculenta* and *H. brasiliensis*), rpl2-trnH (*J. curcas*), and rpl2-rps19 (*E. esula*). We observed that the distances between trnH and IRA region varied from 1 to 198-bp.

**Comparative Analyses Among *H. brasiliensis* and the Other Spurge Plastomes**

Pairwise chloroplast genomic alignments among six spurge plastomes showed a relatively high degree of synteny conservation. The *R. communis* cp genome was used as a reference for plotting the overall sequence identity of the cp genomes of the six spurge plastomes (Figure 5). Our results revealed relatively high sequence identities. The LSC and SSC regions exhibited less similarity than the two IR regions across all Euphorbiaceae species. All the rRNA genes were highly conserved and similar to other plant plastid genomes (Huang et al., 2014). In addition, non-coding regions displayed greater sequence divergence than protein-coding regions. These highly divergent regions contained ccsA, ndhA, rbcL, rpoC2, rpl33, ycf3, atp1-atpH, psbM-petN, and petA-psbF (Figure 5). Such hotspot regions could be developed as molecular markers and species barcoding for future phylogenetic analyses and species identification in the family Euphorbiaceae.

**Phylogenomic Analysis of Cultivated and Wild *H. brasiliensis* Germplasms**

To determine evolutionary relationships of cultivated and wild rubber trees, we compared all 12 cp genomes sequences using *J. curcas* as outgroup. Using neighbor-joining (NJ) method our results showed that cultivated and wild rubber trees failed to form separate clades (Figure 6A), further supporting our former phylogenomic analysis based on nuclear genome resequencing (Liu et al., 2020). We further investigated phylogenetic relationships of the six Euphorbiaceae species established by utilizing complete cp genomes using *Gossypium barbadense* as outgroup. Phylogenetic analyses of the aligned data matrix using Maximum Likelihood (ML) suggested that that cultivated and wild rubber trees formed together with 100% bootstrap supports, which were further grouped with *M. esculenta* with a full bootstrap support (Figure 6B). Thus, the ML tree has well-resolved phylogenetic relationships of these spurge species with ≥95% bootstrap values, which is fairly congruent with former phylogenetic analyses based on complete chloroplast and nuclear genomes (Tangphatsornruang et al., 2011; Rahman et al., 2013; Tang et al., 2016; Pootakham et al., 2017; Liu et al., 2020).

**CONCLUSION**

In the present study, we present the 12 complete chloroplast genomes of cultivated and wild *H. brasiliensis*. We characterized the microstructural variation in the five spurge plastomes, including *H. brasiliensis, Euphorbia esula, Jatropha curcas, Manihot esculenta*, and *Ricinus communis*. We show that the *H. brasiliensis* cp genome differed from the other spurge plastomes with large inversions. We find that cp genomes of wild rubber trees are more variable than cultivated rubber tree cp genomes. The detection of precise genomic variation across rubber tree cp genomes further reveals more genomic variants occurred in wild rubber trees than cultivated rubber trees. Further efforts are still needed to obtain a comprehensive map of chloroplast genomic variation in cultivated and wild rubber trees in a broad geographic range, which will help to better conserve and utilize rubber tree germplasms.

**DATA AVAILABILITY STATEMENT**

The datasets generated for this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

**AUTHOR CONTRIBUTIONS**

L-ZG: conceptualization, visualization, supervision, and funding acquisition. L-YF, C-WG, and H-BW: formal analysis. JL and G-HL: investigation. JL: resources. L-YF: data curation. L-YF and JL: writing – original draft preparation. L-ZG, L-YF, and...
**SUPPORTING MATERIAL**

The Supporting Material for this article can be found online at: [https://www.frontiersin.org/articles/10.3389/fevo.2020.00237/full#supplementary-material](https://www.frontiersin.org/articles/10.3389/fevo.2020.00237/full#supplementary-material)

**TABLE S1A** | Type of SSRs in the 12 *H. brasiliensis* chloroplast genomes.

**TABLE S1B** | Distribution of SSRs in the 12 *H. brasiliensis* chloroplast genomes.

**TABLE S2** | Nucleotide variation across the 12 *H. brasiliensis* chloroplast genomes.
Sneller, C. H., Miles, J. W., and Hoyt, J. M. (1997). Agronomic performance of soybean plant introductions and their genetic similarity to elite lines. *Crop Sci.* 37, 1595–1600. doi: 10.2135/cropsci1997.0011183X003700050032x

Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* 57, 758–771. doi: 10.1080/10635150802429642

Sugiura, M. (1992). The chloroplast genome. *Plant. Mol. Biol.* 19, 149–168. doi: 10.1007/978-94-011-2656-4_10

Sugiura, M. (1995). The chloroplast genome. *Essays Biochem.* 30, 49–57. doi: 10.1007/BF00015612

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Tang, C., Yang, M., Luo, Y., Gao, S., Xiao, X., et al. (2016). The rubber tree genome reveals new insights into rubber production and species adaptation. *Nat. Plants.* 2:16073. doi: 10.1038/plants.2016.73

Tanphatsornruang, S., Uthaipaisanwong, P., Sangsrakru, D., Chanprasert, J., Yoocha, T., Jomchai, N., et al. (2011). Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. *Gene* 475, 104–112. doi: 10.1016/j.gene.2011.01.002

Thiel, T., Michalek, W., Varshney, R., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106, 411–422. doi: 10.1007/s00122-002-1031-0

Wyman, S. K., Jansen, R. K., and Boore, J. L. (2004). Automatic annotation of organelar genomes with DOGMA. *Bioinformatics* 20, 3252–3255. doi: 10.1093/bioinformatics/bth352

Xu, C., Dong, W., Li, W., Lu, Y., Xie, X., Jin, X., et al. (2017). Comparative analysis of six *Lagerstroemia* complete chloroplast genomes. *Front. Plant Sci.* 8:15. doi: 10.3389/fpls.2017.00015

Yang, A. H., Zhang, J. J., Yao, X. H., and Huang, H. W. (2011). Chloroplast microsatellite markers in *Liriodendron tulipifera* (Magnoliaceae) and cross-species amplification in *L. chinense*. *Am. J. Bot.* 98, 123–126. doi: 10.3732/ajb.1000532

Yang, J., Yue, M., Niu, C., Ma, X. F., and Li, Z. H. (2017). Comparative analysis of the complete chloroplast genome of four endangered herbs of *Notopterygium*. *Genes* 8:124. doi: 10.3390/genes8040124

Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., et al. (2010). The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS One* 5:e12762. doi: 10.1371/journal.pone.0012762

Zhang, Y., Du, L., Liu, A., Chen, J., Wu, L., Hu, W., et al. (2016). The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. *Front. Plant Sci.* 7:306. doi: 10.3389/fpls.2016.00306

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Copyright © 2020 Feng, Liu, Gao, Wu, Li and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.**