



The Complete Chloroplast Genome of *Plukenetia volubilis* Provides Insights Into the Organelle Inheritance

Simón Villanueva-Corrales¹, Camilo García-Botero², Froilán Garcés-Cardona², Viviana Ramírez-Ríos², Diego F. Villanueva-Mejía² and Javier C. Álvarez^{1,2*}

¹ BEC Research Group, Biological Sciences Department, EAFIT University, Medellín, Colombia, ² CIBIOP Research Group, Biological Sciences Department, EAFIT University, Medellín, Colombia

OPEN ACCESS

Edited by:

Peter Poczai,
University of Helsinki, Finland

Reviewed by:

Haidong Yan,
University of Georgia, United States
Steven B. Cannon,
Agricultural Research Service,
United States Department
of Agriculture, United States
Furrukh Mehmood,
Quaid-i-Azam University, Pakistan

*Correspondence:

Javier C. Álvarez
jcorre38@eafit.edu.co

Specialty section:

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

Received: 11 February 2021

Accepted: 24 March 2021

Published: 23 April 2021

Citation:

Villanueva-Corrales S,
García-Botero C, Garcés-Cardona F,
Ramírez-Ríos V, Villanueva-Mejía DF
and Álvarez JC (2021) The Complete
Chloroplast Genome of *Plukenetia
volubilis* Provides Insights Into
the Organelle Inheritance.
Front. Plant Sci. 12:667060.
doi: 10.3389/fpls.2021.667060

Plukenetia volubilis L. (Malpighiales: Euphorbiaceae), also known as Sacha inchi, is considered a promising crop due to its high seed content of unsaturated fatty acids (UFAs), all of them highly valuable for food and cosmetic industries, but the genetic basis of oil biosynthesis of this non-model plant is still insufficient. Here, we sequenced the total DNA of Sacha inchi by using Illumina and Nanopore technologies and approached a *de novo* reconstruction of the whole nucleotide sequence and the organization of its 164,111 bp length of the chloroplast genome, displaying two copies of an inverted repeat sequence [inverted repeat A (IRA) and inverted repeat B (IRB)] of 28,209 bp, each one separating a small single copy (SSC) region of 17,860 bp and a large single copy (LSC) region of 89,833 bp. We detected two large inversions on the chloroplast genome that were not presented in the previously reported sequence and studied a promising cpDNA marker, useful in phylogenetic approaches. This chloroplast DNA (cpDNA) marker was used on a set of five distinct Colombian cultivars of *P. volubilis* from different geographical locations to reveal their phylogenetic relationships. Thus, we evaluated if it has enough resolution to genotype cultivars, intending to crossbreed parents and following marker's trace down to the F1 generation. We finally elucidated, by using molecular and cytological methods on cut flower buds, that the inheritance mode of *P. volubilis* cpDNA is maternally transmitted and proposed that it occurs as long as it is physically excluded during pollen development. This *de novo* chloroplast genome will provide a valuable resource for studying this promising crop, allowing the determination of the organellar inheritance mechanism of some critical phenotypic traits and enabling the use of genetic engineering in breeding programs to develop new varieties.

Keywords: oilseed crop, *de novo* assembly, non-model plant, Sacha inchi, long reads ONT

Abbreviations: AMSL, above mean sea level; cpDNA, chloroplast DNA; DAPI, 4',6-diamidino-2-phenylindole; Δ AICc, Akaike information criterion; Δ BIC, delta of Bayesian information criterion; GC, generative cell; GM, germination media; IR, inverted repeat; IRA, inverted repeat A; IRB, inverted repeat B; IS, intergenic sequence; LSC, large single copy; MCMC, Markov chain Monte Carlo; MGU, male germination unit; mtDNA, mitochondrial DNA; oDNA, organellar DNA; ONT, Oxford Nanopore Technologies; PBPI, potential biparental plastid inheritance; PIC, potentially informative character; SNP, single nucleotide polymorphism; SNV, single nucleotide variant; SSC, small single copy; UFA, unsaturated fatty acid; UV, ultraviolet; VC, vegetative cell.

INTRODUCTION

Plukenetia volubilis L., also known as Sacha inchi, is a climbing, perennial, semi-woody, twining oilseed plant of the Euphorbiaceae family, endemic to tropical Peruvian Amazonia that grows mainly in tropical forests at altitudes between 200 and 1500 meters above sea level (MASL) (Gillespie, 1993; Krivankova et al., 2013). It is known that Sacha inchi has a rapid growth ratio, an easy adaptation to growth in nutrient-poor soils, short production cycles, high nutritional content, and displays a potential capacity to become an essential dietary source (Hamaker et al., 1992; Chirinos et al., 2013). For instance, its amino-acid profile presents higher fractions of tryptophan, cysteine, and tyrosine than other oil-seeds sources, reaching the Food and Agriculture Organization's (FAO) highest standards for 2–5 years' children dietary supplements, except for leucine and lysine (Hamaker et al., 1992). Also, it has been shown that the crop relevance of *P. volubilis* indeed relies on its high seed unsaturated fatty acid (UFA) content, particularly of α -linolenic acid or ω 3 (12.8–16.0 g/100 g seed) and linoleic acid or ω 6 (12.4–14.1 g/100 g seed) (Chirinos et al., 2016).

Likewise, it has been reported that the establishment of commercial plantations of this species generates many positive environmental impacts because it can be installed on degraded soils (Bordignon et al., 2012). Thanks to all of the above, the Sacha inchi industry has experienced a great demand in tropical countries, among others, Colombia, Peru, Ecuador, and Brazil, where the cultivation of this species has been increasing (Valente et al., 2017), without developed varieties and ecotypes with genetic stability. In those countries, the ecotypes have been rudimentary and are too difficult to track because of the lack of studies that have been made to understand its genetic diversity (Ocelák et al., 2015). This lack of studies makes it necessary to develop molecular mechanisms that characterize the species, recognize ecotypes, and generate successful seed marketing; responses can be faced by knowing the Sacha inchi chloroplast genome. Insights obtained from the complete chloroplast genome sequence could enhance knowledge of plant biology and diversity for this species.

Chloroplast genomes have assembled notable contributions in diverse plant families, settling evolutionary relationships within phylogenetic clades. Moreover, chloroplast genome sequences have exposed considerable variation between plant species in terms of both sequence and structural variation (Daniell et al., 2016). This information advocates understanding the climatic adaptation of economically important crops, facilitating the breeding of closely related species, developing propagation technologies, genetic engineering applications, and identifying and conserving valuable traits (Wambugu et al., 2015). In addition to improving our understanding of plant biology and evolution, chloroplast genomics research has critical translational applications, such as conferring protection against biotic and abiotic stress and the development of vaccines and biopharmaceuticals in edible crops plants (Brozynska et al., 2016). All of them are significant aspects to be generated

in Sacha inchi. Contributing to baseline, Hu et al. (2018) reported the first *P. volubilis* chloroplast genome sequence (161,733 bp, still unverified GenBank accession number: MF062253) from Xishuangbanna Tropical Botanical Garden in China. However, *P. volubilis* chloroplast's genetics have not been deeply studied.

The chloroplast genome usually occurs in multiple copies within the organelle. It consists of long circular or linear DNA molecules, generally ranging from 120 to 180 kb in angiosperms and 160 to 164 kb in the Euphorbiaceae family (Rivarola et al., 2011; Tangphatsornruang et al., 2011; Zhang et al., 2019). It has a quadripartite structure characterized by two copies of a large inverted repeat A (IRA) and an inverted repeat B (IRB), separating the small single copy (SSC) and the large single copy (LSC) regions. Changes in chloroplast genomes' composition and structure, such as gene losses and rearrangements, have been documented for *Passiflora edulis*, *Cistanche deserticola*, *Hevea brasiliensis* (Tangphatsornruang et al., 2011; Li et al., 2013; Caus-Santos et al., 2017).

Based on CBOL (The Consortium for the Barcode of Life) evidence, four coding sequences (*matK*, *rbcl*, *rpoB*, and *rpoCl*) and three non-coding nucleotide inter-genic spacers (ISs) (*atpF-atpH*, *psbA-trnH^{GUG}*, and *psbK-psbI*) from the chloroplast were suggested to be the adequate plant barcodes for phylogenetic relationships (CBOL Plant Working Group, 2009). Although DNA barcode has been adopted for decades as an investigation system for interspecific taxonomic discrimination, recent evidence suggests that this method may also be applied to plant intraspecific identification and population studies (Ünsal et al., 2019). Among these, *psbA-trnH^{GUG}* was used to establish the evolutive relationships in tribe *Plukenetieae*, which used 153 accessions covering 93 species (Cardinal-McTeague and Gillespie, 2016).

In the Euphorbiaceae family and *P. volubilis*, chloroplast genes such as *rbcl*, *matK*, *ndhF*, and *trnL-F* have been used to study evolutionary relationships at higher taxonomic levels and distribution (Wurdack et al., 2005; Cardinal-McTeague et al., 2019). Nowadays, it is possible to generate entire chloroplast genomes and analyze entire chloroplast gene sequences to determine high-resolved phylogenies (Menezes et al., 2018). Third-generation sequencing technologies producing longer DNA reads have begun to produce high-quality assemblies for complex plant genomes (Li et al., 2018). Oxford Nanopore Technology (ONT) sequencing allows generating an entire chloroplast genome assembled into a single large contig, with a high degree of accuracy and much greater coverage due to longer read lengths (Belser et al., 2018).

Chloroplast DNA (cpDNA) barcodes are also helpful to prove the organelle's inheritance combined with cytological approaches. In all plant taxa, current evidence shows three possible ways by which organellar DNA (oDNA) could be inherited: maternally, paternally, or biparentally (Reboud and Zeyl, 1994). Angiosperms seem to display mainly a maternal inheritance mode of its chloroplasts (Corriveau and Coleman, 1988; Greiner et al., 2015). However, recent studies have shown that some of them are indeed showing a

potentially biparental plastid inheritance (PBPI) mode (i.e., the tendency of inheriting organelle genomes from both parents) (Zhang and Sodmergen, 2010).

In the present study, we approached a *de novo* reconstruction of the whole nucleotide sequence and the chloroplast genome organization using short and long reads technologies. We were able to localize genes, introns, and intergenic spacers and compare the structure of the cpDNA of *P. volubilis* from a Colombian cultivar with the cpDNA that has been reported in China. We found a useful cpDNA marker derived from a phylogenetic approach and used it on a set of five distinct cultivars of *P. volubilis* from different Colombian geographical locations. Thus, we evaluated if it has enough resolution to genotyping cultivars, intending to crossbreed parents and follow this marker in the F1 generation. We finally elucidated the inheritance mode of *P. volubilis* oDNA using both molecular and cytological methods on cut flower buds.

MATERIALS AND METHODS

Plant Material

This research was performed at the Plant Biotechnology and Molecular Biology Laboratories of the Department of Biological Sciences at EAFIT University, located in Medellín, Colombia. Both leaves and seeds were collected from five *P. volubilis* cultivars from farms located in Antioquia (Colombia), showing different environmental conditions and ranging in an altitudinal gradient from 685 to 1501 MASL, having permission for this gathering issued by the National Authority for Environmental Licenses (ANLA), covered in resolution 1516 of 2014 (modified through resolution 1312 of 2015). Those samples were selected by their agronomic trials (**Supplementary Figure 1**) from individuals at each cultivar. Then, samples were entirely washed with distilled water, wiped and packed in bags to avoid light degradation, and stored at -20°C . Additionally, flower buds from Santa Rosa's Cultivar 1 (named C1) were sampled in their first four developmental stages according to Cachique (2006) for cytological observations because it was regarded as promising cultivars after being assessed as suitable to be grown in dry agroclimatic areas. Sacha inchi cultivars C1–C5 are all deposited in the plant collection of the EAFIT's University Plant Biotechnology Lab and are listed in **Supplementary Table 1**.

DNA Extraction

To evaluate the cpDNA *psbA-trnH*^{GUG} IS resolution and untangle the intraspecific variation in *P. volubilis* among Colombian cultivars, total DNA was extracted from leaf tissues of five plant cultivars across Antioquia, Colombia. This was made by displaying ecological and altitudinal variation (C1–C5, **Supplementary Table 1**) using the CTAB method (Doyle and Doyle, 1990), with the addition of a washing step with pure chloroform and the total elimination of isoamyl alcohol; additionally, the centrifugation times in each step were further reduced. The integrity of DNA was determined

through 1% agarose gel electrophoresis, and the concentration was determined using Nanodrop 2000 (Thermo Scientific, Inc., CA, United States).

For ONT and Illumina sequencing, high molecular weight genomic DNA from the C1 leaves was isolated following Ramírez-Ríos et al. (2016) protocol, with some adaptations for plant DNA, as described in **Supplementary Methods 1**. Gel electrophoresis was used to evaluate the extract quality by size estimation, spectrophotometry (A260/A280 and A260/A230 ratios) was used for purity estimation, and Qubit 3 fluorometer (Invitrogen, Carlsbad, CA, United States) for total DNA extracted. DNA samples with an A260/A280 ratio close to 2 and an A260/A230 ratio above 1.5 were kept.

Next-Generation Sequencing and Genome Assembly

Total DNA was sequenced using one lane of Illumina Hi-Seq 4000 paired-end per 150 nt and two flow cells of GridION (Oxford Nanopore Technologies) at the High-Throughput Sequencing Facility of the University of North Carolina at Chapel Hill, United States. Raw whole-genome sequencing paired-end reads were *de novo* pre-assembled using Norgal (Al-Nakeeb et al., 2017) with default settings. The resulting scaffolds were filtered based on the Norgal blast report: the best hit reference was a chloroplast; minimum scaffold length of 1000 bp; minimum identity 95%; and minimum alignment length of 200 bp. The filtered scaffold was used as seed for NOVOPlasty (v2.7.2) (Dierckxsens et al., 2016), which was used to assemble the same raw paired-end reads, set with a 100–200 kb genome range, a kmer size of 39, an insert size of 370, and the remaining parameters by default. The quality of the resulting sequence was further assessed using Pilon (v1.23) (Walker et al., 2014) as follows: raw paired-end reads were trimmed using Trimomatic-PE (v0.39) (Bolger et al., 2014) with a sliding window of width 4 and quality 25 and a min length of 50. All surviving (paired and unpaired) reads were mapped using BWA-MEM (v0.7.17) (Li et al., 2013). SAMtools (v1.9) (Li et al., 2009) were used to filter the unmapped reads and sort the resulting alignments. These were provided to Pilon, along with the generated sequence, with options `-fix all` and `min depth of 60` and the remaining parameters as default.

Similarly, the structural quality of the assembly was assessed with ONT reads. Nanopore raw reads were basecalled with Guppy (v2.3.5+53a111f, Oxford Nanopore Technologies), using the flip-flop model for DNA, yielding a total of 4,699,073 reads with an N50 of 2 kbp. These reads were mapped with minimap2 (v2.17-r941) (Li, 2018) and visualized with Tablet (v1.17.08.17) (Milne et al., 2013).

No changes were suggested by Pilon, which indicates that every base in the assembled sequence is following the bases on the reads. On the other hand, misassemblies can be detected by inspecting the coverage profile of independent data on the assembly, as there would be drops in coverage in misassembled regions. A total of 956,508 mapped ONT reads generated a continuous smooth coverage of an average 3917x depth, where the LSC, SSC, and IRs could be visualized

(**Supplementary Figure 2**), indicating the structural correctness of the assembly. The assembled chloroplast genome was hereafter called SI_cpDNA_C1 (C1).

Chloroplast Genome Annotation and Comparison

The chloroplast sequence (SI_cpDNA_C1) was uploaded as a FASTA file to the Chlorobox portal and annotated using GeSeq (Tillich et al., 2017) tool using the following parameters: Circular and plastid sequence; annotate plastid IR, 85% protein search identity, and 85% rRNA, tRNA, and DNA search identity as BLAT search options; ARAGORN (v1.2.38) (Laslett and Canback, 2004) was selected as third-party tRNA annotator, with bacterial/plant chloroplast as genetic code, max intron length set to 2500 bp, and fixing introns. Since the automated annotation is error-prone, CDS were reviewed and manually curated, checking for correct start codons, lengths, and stop codons, and then the complete annotation was deposited under the GenBank accession number MW591569. SI_cpDNA_C1 was compared with MF062253.1 accession looking for nucleotide variants. Genome alignment was generated in MUMmer 3 (Kurtz et al., 2004) using the suite NUCmer, then the alignment output was filtered with delta-filter, and finally, single nucleotide polymorphisms (SNPs) were calculated with Dnadiff using default parameters. Scripts in Python and Biopython to filter and resume the data were documented in Jupyter notebooks and deposited in the Github repository for this work. All these processes and parameters are clarified in the repository associated with this article (Villanueva et al., 2020). To detect structural variants, SI_cpDNA_C1 was aligned to NC_016736 using NUCmer. The resulting alignments were used as input to MUMmerplot. Furthermore, SimpleSinteny (Veltri et al., 2016) was made using the 60 longest genes present in both chloroplasts.

CpDNA *psbA-trnH*^{GUG} Marker Amplification and Sequencing

To amplify the cpDNA *psbA-trnH*^{GUG} IS fragment from each cultivar of *P. volubilis*, *trnH*^{GUG} (5'-CGCGCATGGTGGATTCAACAATCC-3') (Tate and Simpson, 2003) and *psbA* (3'-GTTATGCATGAACGTAATGCTC-5') (Sang et al., 1997) primers were used (**Supplementary Methods**). Amplified PCR products from *psbA-trnH*^{GUG} IS (~500 bp) were sequenced using the Sanger method (Macrogen Inc., South Korea). Eventually, sequences were trimmed for base scores below 18 Phred score in a 5 base window average using Biopython libraries (Cock et al., 2009). Then, each *trnH*^{GUG} was locally aligned with the reverse complement of its *psbA* cultivar sequence. Each cultivar consensus region was reconstructed using the Smith and Waterman (1981) algorithm, with an identity more significant than 95%. Therefore, the species fragments' identity was verified running the BLAST algorithm in the NCBI database, and the sequences were deposited in the GenBank database (**Supplementary Table 1**). Subsequently, these sequences were aligned using ClustalW (Thompson et al.,

2003) on Geneious Pro ver. 11.1.5 and manually scanned the polymorphic regions.

Phylogenetic Analyses Methods

Two main phylogenetic hypotheses were inferred using maximum likelihood (ML) and Bayesian inference (BI). Analysis of the aligned matrix with JmodelTest software ver. 2.1.10 (Darriba et al., 2012) showed that the TVM+I+G4 (Motoo, 1981) substitution model was the best model based on the delta Akaike information criterion (Δ AICc) and delta of Bayesian information criterion (Δ BIC) information criteria. Bayesian information criterion (BIC) for the same matrix showed the F81+G4 substitution model as the best choice. However, topologies for both models did not change (see **Supplementary Table 4**). RAXML-NG ver. 0.9 (Kozlov et al., 2019) was used to reconstruct the best ML topology using five independent replicates; 1,000 bootstrap replicates were targeted to the best topology using DendroPy ver. 4.0 (Sukumaran and Holder, 2010) in a non-parametric bootstrapping fashion and consensus topology was visualized using FigTree ver. 1.4.3.

For BI, the aligned matrix file was edited using Beauti ver. 2.0 and updated the substitution model to TPM1uf + I and the strict molecular clock. Then, to infer the relationships, 1×10^7 generations were run using Markov chain Monte Carlo (MCMC) algorithm, sampling one tree every 1,000 generations in BEAST ver. 2.5. The adequate sample size (ESS) values for all parameters were > 200, and they reached convergence and stationarity as determined by Tracer ver. 1.7.1 (Rambaut et al., 2014). Finally, the maximum credibility tree was generated using TreeAnnotator ver. 2.5 implementing a 10% burn-in and visualized using FigTree ver. 1.4.3. Both topologies (ML and BI) were summarized and edited using Inkscape ver. 0.9. Branch lengths from BI phylogeny were conserved, but both ML bootstrap (MLB) and Bayesian posterior probability (BPP) support values were depicted in the final tree.

Inheritance Determination Based on Cytogenetic Analysis

Seeds of three cultivars (C1, C3, and C5) of *P. volubilis* were brought into cultivation in a greenhouse at EAFIT University, located in Medellin-Colombia. After germination, seedlings were placed in 11 cm pots within a mixture of 50% sand and 50% potting soil. Afterward, those plants were placed under controlled conditions (25°C, 2 months at 12 h light/12 h dark) before pollinations. Due to *P. volubilis* is an allogamous species with a high percentage of self-pollination (Cachique, 2006), controlled hand pollinations were undertaken (between November 13 and December 13, 2018) with C1, C3, and C5 (C1, C3, and C5, respectively). Before the cross-pollinations, all open flowers were removed from the inflorescence. The unopened buds were emasculated, and when the stigma became receptive (usually within two days), pollen was applied directly from the anthers of the pollen parent, following the Soda Straw Method (TNAU, 2015). The stigma of C5 was pollinated with pollen from C1 and C3. Reciprocal pollinations were also undertaken

TABLE 1 | Alignments of *psbA-trnH^{GUG}* Inter-Genic Spacer sequences (nucleotides 321–345) from Colombian cultivars of *Sacha inchi*.

| Cultivar | cpDNA Sequence (<i>psbA-trnH^{GUG}</i> IS) | |
|-------------------|---|-------|
| Parentals | 321 | 345 |
| C1 | | |
| C5 | ...G.....T A T T T A.....T A | |
| C3 |T - - - - - | |
| Generation 1 (F1) | 321 | 345 |
| ♂C1 × C5♀ | ...G.....T A T T T A.....T A | |
| ♂C3 × C5♀ | ...G.....T A T T T A.....T A | |
| ♂C5 × C1♀ | | |

Parentals; Santa Rosa cultivar (C1), Yarumal cultivar (C3), and San Luis cultivar (C5). Cross-pollinations were carried out between parentals (see section “Materials and Methods”), and F1 generation was genotyped. Dots mean similar nucleotides in the position, lines mean deletions, and letters are nucleotides with polymorphisms.

with C5 as male and C1 as female (Table 1). Once the first generation was generated, five seeds were collected and placed in 11-cm pots in a mixture of 50% sand and 50% potting soil. Afterward, the plants were placed under controlled conditions (25°C, 2 months at 12 h light/12 h dark). The leaves from the hybrids were collected and disposed of for DNA extraction and subsequent sequencing of the chloroplast’s intergenic *psbA-trnH^{GUG}* region. Likewise, to visualize cpDNA, most developed flower buds (i.e., stage IV, just before anthesis, Cachique, 2006) from different individuals from C1 were immersed in a beaker with 5% sucrose overnight and later dried with towel paper. Then, water excess was absorbed by laying out the flower buds into silica gel inside a desiccation chamber for 30 min. Later, a Zeiss Stemi DV4 stereomicroscope was used to observe, and to tape dehiscent flower buds anthers over well depression slides to release the pollen. Afterward, two drops of germination media (GM) [sucrose 10%, boric acid (100 mg/L), magnesium sulfate (200 mg/L), nitrate (100 mg/L)] (modified from Brewbaker and Kwack, 1963) were spilled into the wells. Well depression slides were stored in a dark chamber at 25 ± 3°C for 24 h.

Once the plants developed and opened their flowers, the pollen was collected to slide preparation for DAPI (NucBlue™ Fixed Cell ReadyProbes™ Reagent, ThermoFisher Scientific, United States) staining following three approaches: (i) pollen grains samples were left dried at room temperature before staining in the GM, (ii) fixation process in which pollen grains samples were treated sequentially with 10 and 30% ethanol and then incubated at 36°C to speed up evaporation, and (iii) slide preparation flower buds in four developmental stages were dehydrated by rinsing with glutaraldehyde and successive 20, 40, 60, 70, and 96% ethanol solutions (20 min each) and then dried in a BD23 Binder incubator at 37°C. Subsequently, histological sections of flower buds were carried out, embedding them in paraffin (Paraplast, 39601006, Leica) and cut at 10 μm with a microtome (RM2125 RTS, Leica). Flower buds histological slides were dehydrated at 65°C in a BD 23 Binder incubator for 1 h and then immersed in a succession of ethanol concentrations (60, 40, and 20%, 3 min each) for rehydration. Finally, slides were dehydrated with a

succession of alcohols (20, 40, and 60%, 3 min each), rinsed with one drop of DAPI solution, and dried for 1 h. Images of pollen grains and histological sections were first processed using ZEN software from AxioCam Carl Zeiss and finally with FIJI® (2012).

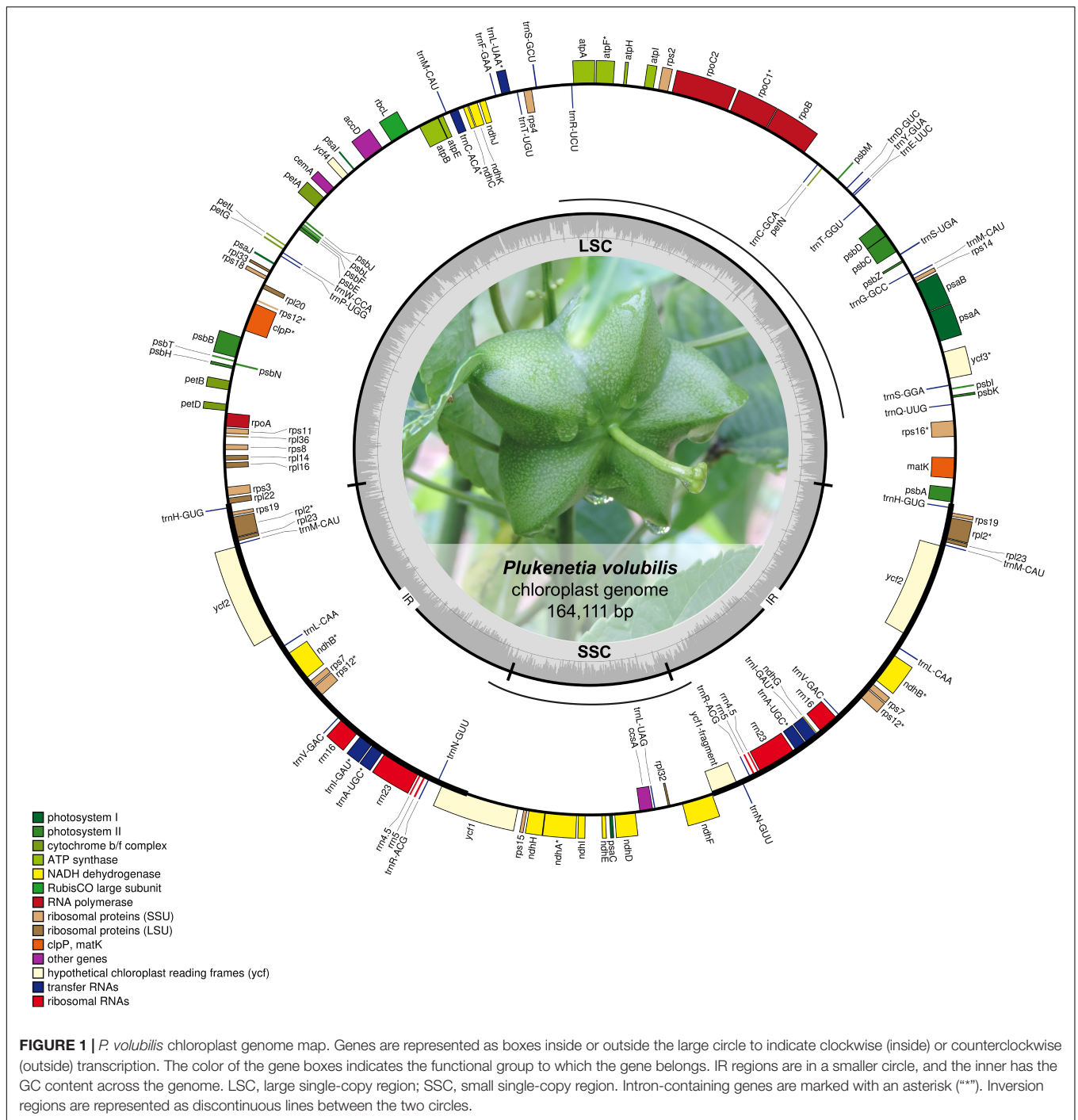
RESULTS AND DISCUSSION

Reconstruction of the Whole Chloroplast Genome of *P. volubilis* Using Next-Generation Sequencing Technologies

The hybrid strategy of ONT combined with Illumina was attempted for the first time in bacteria (Laver et al., 2015). Currently, it is a conventional hybrid method for relatively short genomes, such as cpDNA (Kang et al., 2019). This strategy of using a combination linking long and short reads may be the best approach to assembling chloroplast genomes due to its capacity to combine the benefits of the length of long reads and the accuracy of short reads (Wang et al., 2018). Previously, the chloroplast genome sequence and assembly in *Oryza coarctata* (Wang et al., 2018) and *Eucalyptus pauciflora* (Mondal et al., 2018) were reported by using ONT and Illumina. In this research, these NGS technologies were combined to reconstruct a whole chloroplast genome; first, with the short reads technology, the Illumina HiSeq 4000 system gave a total of 56,711,152 pairs of reads (according to post chloroplast assembly mapping and counting with BWA-MEM). Second, 4,699,073 reads resulted from the ONT after subsequent base-calling, with an N50 of 2 kbp. After the assembly, the average depth of coverage of the Illumina reads on the consensus sequence was 40,560x, and the average GC content was 35.8%.

Organization and Comparative Analysis of Chloroplast Genome of *P. volubilis*

The complete chloroplast genome of *P. volubilis* showed a single-circular molecule (Figure 1) similar to most other higher plants (Sato et al., 1999). The *P. volubilis* cpDNA resulted in a 164,111 bp length, 2378 bp more than Hu et al. (2018; Supplementary Table 2), displaying two copies of an inverted repeat (IRA and IRB) of 28,209 bp, each separating an SSC region of 17,860 bp and an LSC region of 89,833 bp (Figure 1). During gene annotation, the use of the accession NC_016736.1 of *Ricinus communis* as a reference, whose genome is the closest to *P. volubilis*, revealed unexpected asymmetries in the orientation of some rRNA genes on the IRs, in contrast to *Arabidopsis thaliana* (Sato et al., 1999), *Nicotiana tabacum* (Shinozaki et al., 1986), *Glycine max* (Saski et al., 2005), and others. Sequence level comparisons of these genes showed misannotations of the strand present in the *R. communis* chloroplast genome accession (see rRNA_orientation.ipynb in Villanueva et al., 2020). After reannotation without reference, the rRNA genes were found in the expected orientation. Annotation using Aragorn identified 131 genes (87 single-copy); 36 tRNAs identified eight rRNAs (four



in each IR), 20 introns (in 18 genes), and 86 protein-coding genes, and an additional one fragmented (**Supplementary Table 3**).

From the tRNAs found, the normally *trnK-UUU*, which surrounds the *matK* gene, stands out as it presents an insertion that disrupts its anticodon converting it to *trnStop-UUA*. It is unclear if this tRNA is functional. However, as there is no other *trnK*, its malfunction would imply either a shortage or an alternative mechanism for lysine supply to the translation process. Regarding other genes, *atpF* has a conserved intron of

718 bp, *ycf1* has a fragmented copy of 1916 bp in IRB and *ycf3*. Also, *clpP* has two introns, and *rps12* presents trans-splicing as it was also seen in Castor bean and found in bacterial and organellar genomes and thought to be ancestors of spliceosomal introns and retrotransposons in eukaryotes (Bonen, 2008; Lambowitz and Zimmerly, 2011).

We also reannotated the MF062253.1 genome due to mistakes assumed in the assembly with the reference *R. communis*, and the annotation pipeline followed Mummer (Kurtz et al., 2004) and

Autograph (Derrien et al., 2007) comparisons between Chinese genome and Colombian C1 genome showed two inversions in the LSC and SSC regions (Figure 1). Read coverage in this region was highly supported, and ONT reads were mapped confirming these inversions (Supplementary Figure 2). No duplications were found in the genomes, but seven GAPS in the Colombian genome summing 1782 bp were inserted, and six GAPS were deleted summing 596 bp. The variant analysis found 52 SNPs flanked by 20 exact base pair nucleotides match on both sides. Transition A: T→G: C and transversion A: T→C: G were the most abundant into the genomes with 32.7% of occurrence each. The genes with more SNPs relative to the Chinese accession were *rpoC1*, followed by the *ycf3* gene. We aligned the sequences of nucleotides and amino acids for both *rpoC1* and found several SNPs changing the protein's coding sequence. We also compared our *rpoC1* with other *Euphorbiaceae*. We found that *R. communis* presented an indel of 26 aa (FSFARPIAKKPTFLRLRGSFYEYIQS) in the amino acid 145 N-terminal of the protein. Curiously, amino acid alignment of *rpoC1* with *Jatropha curcas* had a higher identity than *R. communis*, and there was no indel detected (see Snps_analysis.ipynb in Villanueva et al., 2020). *rpoC1* and *rpoC2* are two genes that encode for two subunits of the RNA polymerase (Bergslund and Haselkorn, 1991). Palenik and Haselkorn (1992) proposed the *rpoC1* as a valuable tool to clarify phylogenetic relationships among plant groups (Liston, 1992). Since that, it has been used together with other molecular markers to resolve phylogenetic relationships in many plant families (Downie et al., 1998; Messinger et al., 1999; Watson et al., 2000).

Two large inversions were identified in the *P. volubilis* cpDNA when it was structurally compared with *R. communis* chloroplast (Figure 2). The first inversion is located in the middle of the LSC, spanning 39,426 bp, and flanking the genes *rps4* and *psbI*. A second more minor inversion of 17,493 bp changes the SSC orientation (relative to *R. communis*). This inversion affects the open reading frame of one of the *ycf1* copies and thus fragments it. This fragmented copy was annotated as a pseudogene. As no misassemblies were detected in the assembly (see section "Next-Generation Sequencing and Genome Assembly"), the observed rearrangements are from biological origin. Further assessment with Simple Synteny (Veltri et al., 2016) using the 60 longer genes showed how these inversions affected gene order mainly in the LSC (Figure 2). Both inversions are absent in the previously reported chloroplast (acc MF062253.1) because the sequence was generated with a guided assembly using *R. communis* chloroplast as a reference (Hu et al., 2018).

We performed a phylogenomic study with the chloroplast genomes available in the GenBank to the date for the family *Euphorbiaceae* members using ML reconstruction (Supplementary Figure 4). All of the *Crotonoideae*, *Euphorbioideae*, and *Acalyphoideae* subfamilies members were consistently grouped in the cladogram, and C1 was inner the tribe *Plukenetieae* next to *R. communis*. This was consistent with Hu et al. (2018), by suggesting that the general features of *P. volubilis* chloroplast genome compared to other phylogenetic relative chloroplasts have not been significantly different in terms of size or number of genes, but indeed have been significantly different in the structure of the genome

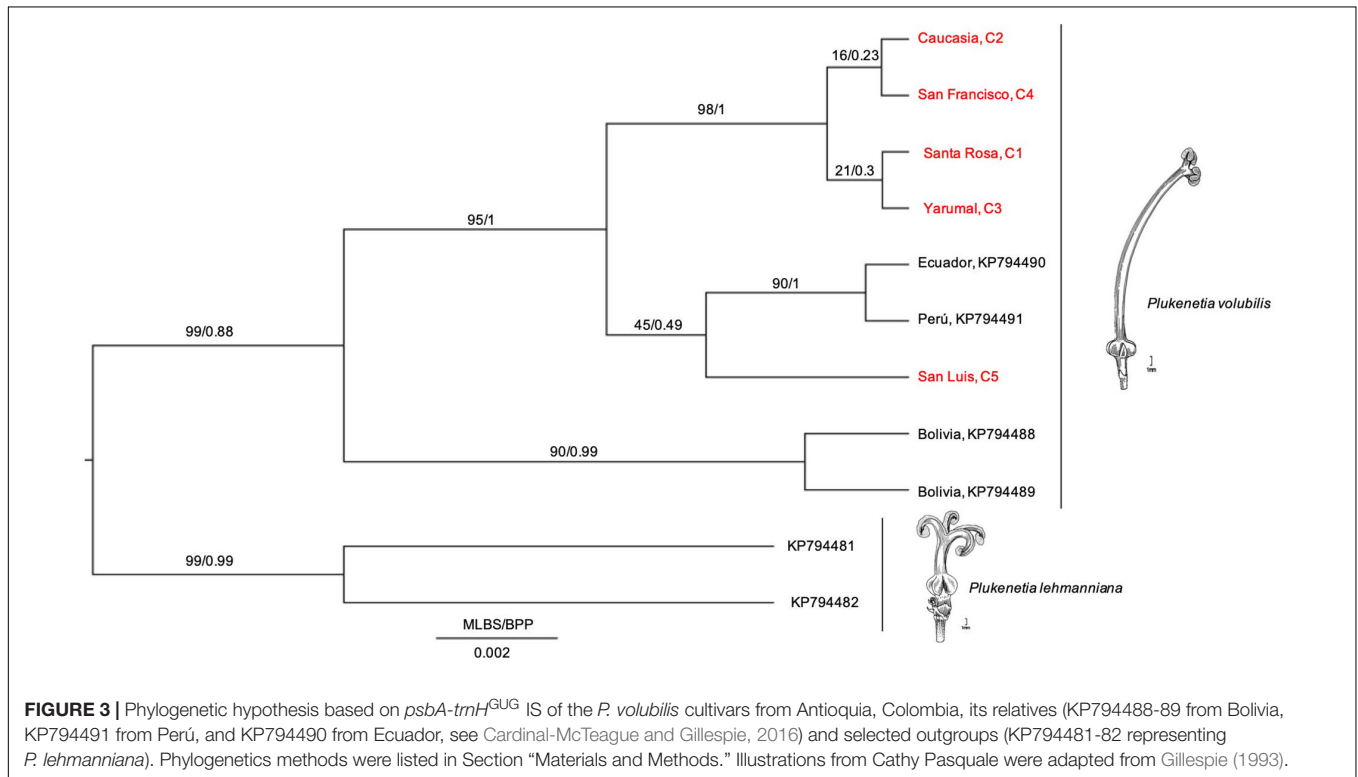
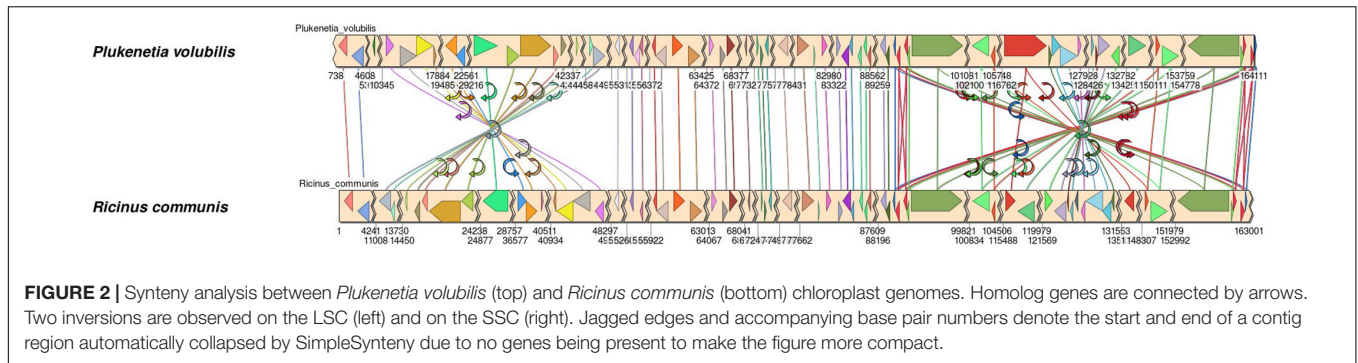
presenting rearrangements in some populations of SI as it has been documented in other chloroplasts of the tribe *Plukenetieae* (Cardinal-McTeague and Gillespie, 2016).

Phylogenetic Studies of Sacha Inchi Cultivars From Colombia

Alignments of *psbA-trnH^{GUG}* IS sequences revealed nucleotide variation between SI Colombian cultivars (Table 1). The San Luis cultivar (C5) sequence showed multiple insertions between the nucleotides 321 and 345, while C1–C4 had a deletion in the same region (Table 1). Other studies have shown that *psbA-trnH^{GUG}* IS could have potentially informative character (PIC) value relative to other cpDNA regions because it presumably shows considerable variability (>50%) between angiosperm lineages (Shaw et al., 2007). However, the marker resolution to unveil phylogenetics relationships has only been assessed at the tribe level (*Plukenetieae* in Cardinal-McTeague and Gillespie, 2016). At the genus level (*Plukenetia*), Cardinal-McTeague et al. (2019) supported two major groups (the pinnately- and palmately-veined clades) and five subclades within *Plukenetia* in perhaps the most profound phylogenetic time-dependent evolutionary framework study of the pantropical genus.

Authors mentioned above combined cpDNA (*matK* and *ndhF*) and nDNA (*KEA1* and *TEB*) markers to assess the phylogenetic relationships of *Plukenetia* spp. as well as its divergence in a geological time-scale. Our results in the *P. volubilis* cpDNA complete sequence showed that *matK* is a superimposed complex gene for the *trnK-UUU* and two antisense introns that flank the *trnK-UUU* as seen in other species such as *P. edulis* cpDNA (Cauz-Santos et al., 2017), and perhaps this nature of the gene is contributing to its resolution at this level. However, the single nucleotide variant (SNV) present in the Colombian cultivars and its associate PIC values across angiosperms lineages suggest that evolutionary changes might occur within the Colombian population *P. volubilis* due to domestication selection. Therefore, we hypothesize that the San Luis cultivar (C5) diverges from all other cultivars and that its evolutionary relationship with the other cultivars analyzed in this study could be revealed using the *psbA-trnH^{GUG}* IS as a molecular marker.

To reveal intraspecific evolutionary relationships, we performed a phylogenetic analysis with an aligned matrix of 11 taxa and 362 bp (the five Colombian cultivars + four South American accessions + two *Pterichis lehmannaiana* sequences as an outgroup). We inferred the ML and BI trees (Figure 3). Results showed that cultivars C1–C4 grouped in a unique monophyletic group, while San Luis Cultivar (C5) was found to diverge from the same ancestor as the Ecuador + Perú cultivars (Figure 3); however, this group (C5 + Perú + Ecuador) was not highly supported neither by MLB nor BPP values, indicating that their relationships could not be revealed in this study. Notwithstanding, a new supported clade (Colombia + Perú + Ecuador) suggests that cultivars diverged from an ancestor shared with Bolivian cultivars, suggesting that Colombian, Peruvian, and Ecuadorian cultivars might be sharing a common evolutionary history based on the *psbA-trnH^{GUG}* IS marker.



According to this phylogenetic hypothesis, no differences among Colombian cultivars are observed (Figure 3). Notwithstanding, other relationships between Peruvian and Bolivian and Equatorial accessions of *P. volubilis* seem to display significant divergence. Furthermore, *P. lehmanniana*, as a sister group, is consistent with preceding studies (Cardinal-McTeague and Gillespie, 2016). Remarkably, the three main groups diverged with posterior probability values greater than 0.98 (Figure 3), which illustrates *psbA-trnH^{GUG}* proper resolution for intraspecific relationships.

Determination of the Organelar Inheritance Mode by Crossbreeding

Since cpDNA paternal contribution in angiosperms has not been reported extensively, and mitochondria are crucial during pollen tube germination (Twell et al., 2006), we assume that

P. volubilis cpDNA is maternally inherited and its mtDNA is biparentally inherited. However, it is necessary to experimentally determine the cpDNA inheritance mode in *P. volubilis*, which is essential information for enabling successful genetic improvement programs, avoiding unwanted crosses with wild germplasm (Glick and Patten, 2017). It would appear suitable to develop genetically modified cultivars, harboring cpDNA modifications, avoiding gene scape. In cpDNA enhanced plants, higher expression of proteins is manifested, allowing them to overcome biotic and abiotic stress (Glick and Patten, 2017). In order to determine the cpDNA mode of transmission, *psbA-trnH^{GUG}* IS was used as a chloroplast sequence marker in crossbreeding experiments of the present study's cultivars. Therefore, solving cpDNA inheritance mode in *P. volubilis* is a stepping stone to decide how to develop modified cultivars eventually that could avoid unwanted crosses with wild germplasm (Glick and Patten, 2017).

As previously stated, *psbA-trnH*^{GUG} IS sequences from Colombian cultivars indicated nucleotide variation (**Table 1**). Since then, the San Luis cultivar's (C5) sequence has shown multiple insertions between 321 and 345 nucleotides, while C1–C4 have a deletion in the same region. This suggests the suitability of the cpDNA marker to discriminate among the local *P. volubilis* cultivars by the use of molecular techniques and particularly the cultivar C5 as the main parent in crossbreeding experiments. Many factors could be explaining the exhibited variability of *P. volubilis*. Additionally, the Peruvian Amazonia's diverse ecological structure has allowed the domestication of native plants (such as *P. volubilis*) by keeping a high genetic variability during the last centuries (Rodríguez et al., 2010). In this context, the movement of non-genotyped seed and

uncontrolled use of *P. volubilis* as a promising crop through different varied regions of South America might be possible, causing new varieties adapted to diverse habitat conditions (Correa et al., 1990; Gillespie, 1993). Besides, spotted SNVs (**Table 1**) can be tested for simultaneous occurrence in a *P. volubilis* interbreeding population and serve as additional statistical data supporting SNP discovery.

Sequences from the *psbA-trnH*^{GUG} IS region in the F1 generation, derived from genetic crosses between the three parents (C1, C3, and C5), were obtained. Sequence analysis showed a maternal inheritance of the cpDNA since whenever the C5 was the maternal parental, the “. . . G T A T T T A T A” sequence was inherited to the F1 and the reciprocal crossing inherited the

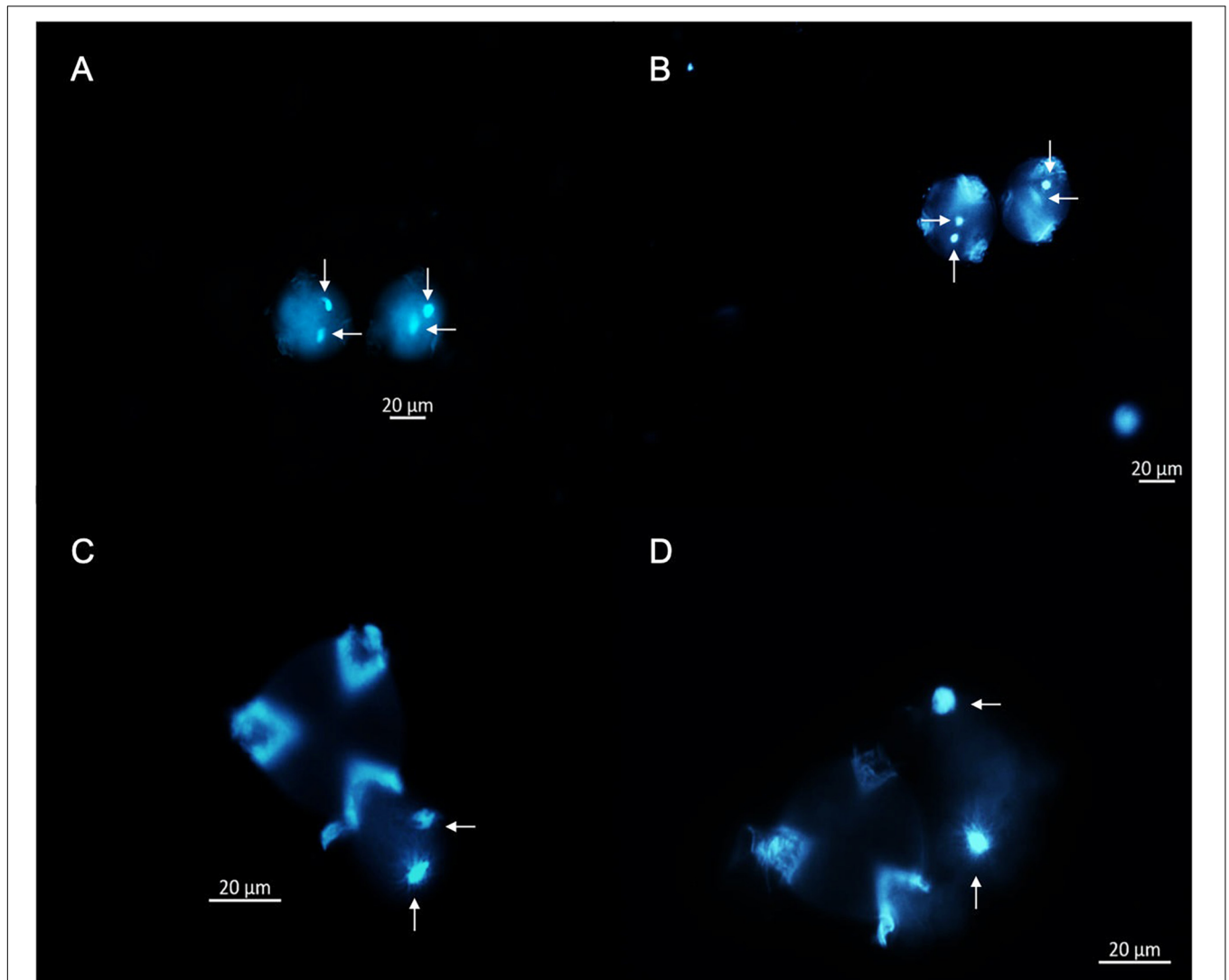


FIGURE 4 | Germination stages of *P. volubilis* pollen grains using GM, stained using DAPI fluorochrome, and visualized under fluorescent microscopy. **(A)** (Stage 1), pollen grains not completely germinated, lacking pollen tube and stained directly. **(B)** (Stage 2), pollen grains showing very conspicuous germination apertures. **(C)** (Stage 3) pollen partially germinated with an observable migration of vegetative and germinative cells. **(D)** (Stage 4), pollen grains germinated and conspicuous pollen tubes with male germination unit (MGU) migrating across the tube. Germinative cells are illustrated in vertical arrows, and vegetative nuclei are pointed by the horizontal arrows. A and B were observed at 20X and C and D at 40X.

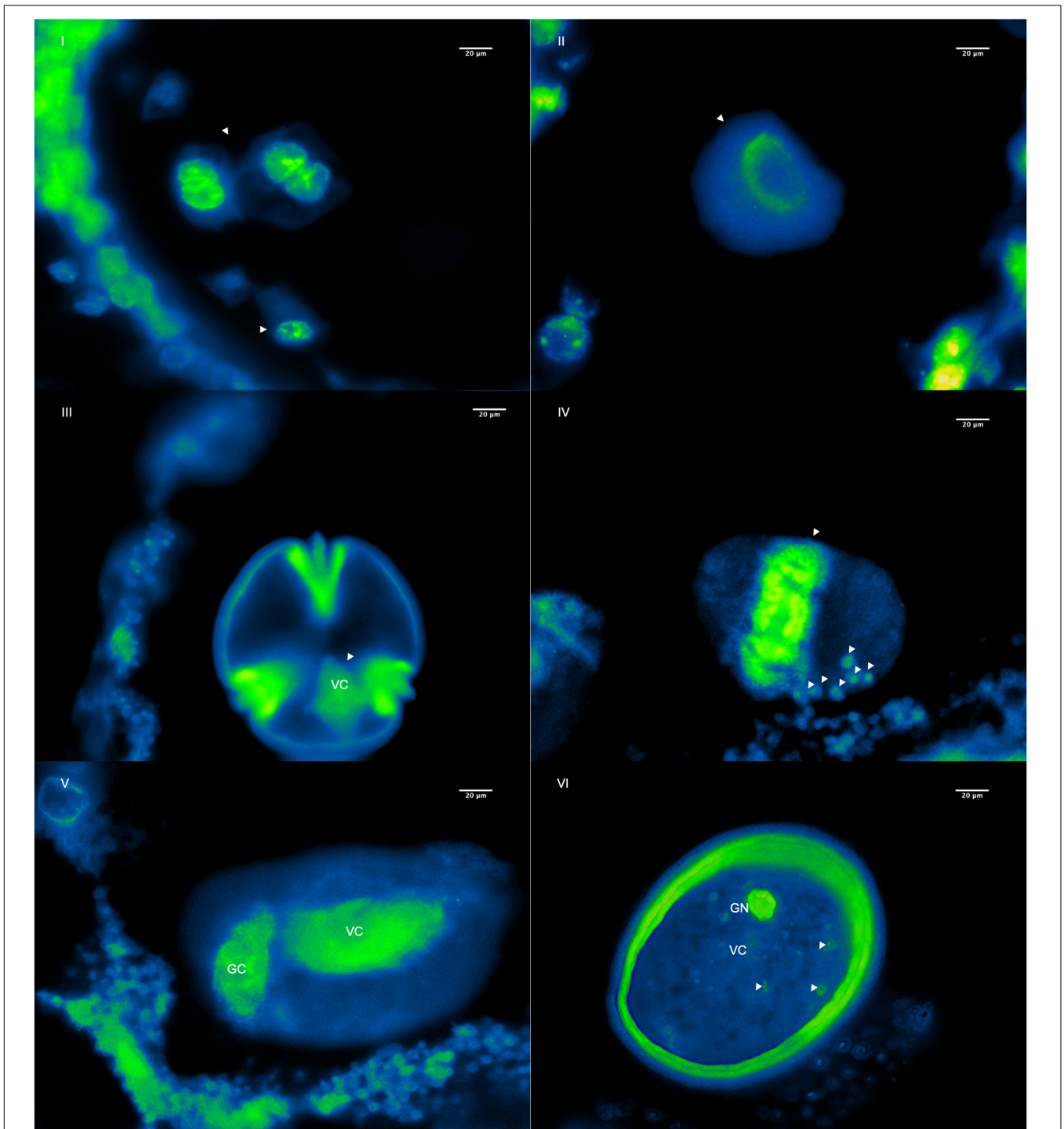


FIGURE 5 | Six morphological stages of microsporogenesis and microgametogenesis are shown during the first four flower buds developmental stages of pollen cells of *P. volubilis* using fluorescent micrograph and DAPI stain. **(I)** First meiosis and formation of tetrads, **(II)** microspore release, **(III)** microspore during interphase, **(IV)** microspore undergoing the first pollen mitosis (PMI) and trafficked fluorescent oDNA granules (pointed out with the arrows), **(V)** formation of a generative cell (GC) and vegetative cell (VC), and **(VI)** mature pollen and cytoplasmic DNA decay and formation of the final MGU before anthesis. All observations are at 100X.

genotype C1 (Table 1), indicating a maternal inheritance mode of the cpDNA. This result is consistent with most angiosperm plants where cpDNA is uniparentally transmitted

by the female parental, and thus it can be used to identify the maternal genome donor (Feitosa, 2017). Therefore, molecular understanding derived from cpDNA genome

analysis suggests that the inheritance model *P. volubilis* is uniparental maternal, which supports cytogenetic analysis carried out in parallel.

Determination of the Organellar Inheritance Mode by a Cytogenetic Approach

Around 80% of angiosperms display maternal inheritance mode of cpDNA, compared with the other 20% of angiosperms studied, which show a strong bent for plastid transmission from the male lineage, a phenomenon known as potential biparental plastid inheritance (PBPI); results mainly based on scanning plastid DNA in the male gametic cell with DAPI (Corriveau and Coleman, 1988; Zhang and Sodmergen, 2010). On the other hand, paternal inheritance in angiosperms has been observed in bare cases (Harris and Ingram, 1991; Testolin and Cipriani, 1997).

A cytological approach using DAPI staining was implemented in three steps to elucidate the oDNA inheritance mode of *P. volubilis*: (i) direct DAPI stain into germinated pollen grains; (ii) DAPI stain on alcohol fixed germinated pollen grains, and (iii) DAPI staining on histological slices of four flower bud's developmental stages. During pollen grain germination, three aperture furrows' appearance is conspicuous; these furrows expose feeble intine to the GM and are the proper places for tube germination (Figure 4A). Besides, observations of germinated pollen grains showed no evidence of plastid DNA or any oDNA granules; mtDNA is also included in any of the germination stages observed (Figure 4). Plant cytologists agree with the DAPI staining method as a fundamental approach, claiming that MGU migration through a pollinic tube could be considered the gametophyte's final and active state (Twell et al., 2006). If plastids are absent throughout MGU migration, the zygote will be missing paternal plastids (Zhong et al., 2011). Therefore, adding molecular results, this study reports for the first time that *P. volubilis* oDNA is not paternally transmitted, making this scientific report one of the first profound advances in organellar inheritance for this species.

This experiment additionally shows that *P. volubilis* has a binucleate male gametic cell based on the diagnosis of pollen grains fixed with alcohol (Figures 4C,D). It exhibits a diffuse nucleus (vegetative) and a more defined one (germinative cell) at the end of gametogenesis. Notwithstanding, almost every plant species shows binucleated stereotypical pollen grains (70% of all plant spp.). According to Twell et al. (2006), trinucleate or binucleate pollen is a random feature among *Euphorbiaceae* (Brewbaker, 1967), therefore rendering a useful characteristic regarding pollen-based taxonomy.

Considering plastid DNA or oDNA granules, mtDNA is also included. Since DAPI staining is a non-selective DNA stained method, and no stained plastids were observed (Figure 4), it is suggested that *P. volubilis* mtDNA inheritance mode seems to be non-paternally as well. The mtDNA is independently trafficked from that of cpDNA. For instance, cpDNA and mtDNA of *Medicago sativa* are inherited biparentally and maternally,

respectively (Forsthoefel et al., 1992), whereas those of *Musa acuminata* (banana) are inherited maternally and paternally, respectively (Fauré et al., 1994). These oDNA behaviors correlate well with the degradation or amplification of DNA in each organelle of the generative cell (Nagata et al., 1999); in the generative cells of *M. sativa*, cpDNA is amplified while mtDNA is degraded, whereas in *M. acuminata*, cpDNA is degraded while mtDNA is amplified (Twell et al., 2006).

Histological slices of flower buds during four developmental stages were stained with DAPI to analyze the course of oDNA along pollen gametogenesis. Results showed oDNA decay during the switch of flower buds developmental stages III and IV (Figure 5). This decrease might be happening during pollen mitosis (PMI). Loss mechanisms are several, from restriction enzyme success to total cytoplasmic content suppression in a male gametophytic cell (Sager and Lane, 1972). Indeed, many cellular mechanisms controlling oDNA inheritance transmission have been proposed clearly, for instance: (1) physical exclusion of the organelle itself during PMI; (2) elimination of the organelle by the formation of enucleated cytoplasmic bodies (ECB); (3) autophagic degradation of organelles during male gametophyte development; (4) digestion of the organelle after fertilization; and (5) digestion of oDNA in generative cells just after PMI (Nagata, 2010). However, hypotheses (1) and (5) may encompass others and also explain better the maternally or non-paternally transmitted mechanism of oDNA (Nagata, 2010).

Since results showed an organelle decay in the very early stages of mitosis (Figures 5IV–VI), it is suggested that the oDNA loss mechanism in *P. volubilis* is primarily due to an organelle physical exclusion system. Indeed, many granular bodies are bordering outside pollen grains (Figure 5IV), suggesting an oDNA removal from pollen grains. Nonetheless, these observations do not, by themselves, definitively support this hypothesis. Additional evidence is needed, such as molecular studies showing how actin filaments or microtubules traffic organelle during microgametogenesis (Twell et al., 2006). Furthermore, loss mechanisms may not be mutually exclusive, so that multiple systems might be acting during pollen formation. It is reasonable to think that later, between the last stage of pollen formation (Figure 5VI) and pollen germination (Figures 4C,D), a second mechanism such as organelle digestion may be happening. Finally, histological slide observations should be taken carefully. They could eclipse natural phenomena because tridimensional bisected pollen grains may not accurately reflect what is inside.

CONCLUSION

Supported by cpDNA genome and cytogenetic analysis, this study found that *P. volubilis* chloroplast's inheritance model is uniparental maternal. With 164,111 bp length displaying two copies of an inverted repeat sequence (IRA and IRB), the *P. volubilis* cpDNA was reported completely *de novo* for the first time, filling a significant gap and need of studies for a species with great demand due to its high UFAs seeds content. When it was compared with *R. communis* cpDNA, two large inversions were identified in the *P. volubilis* cpDNA: the first inversion located

in the middle of the LSC and the second one between the genes *trnN-GUU*. Furthermore, genomic analyses between *P. volubilis* cpDNA and other *Euphorbiaceae* cpDNA showed significant differences in the genome structure, including rearrangements. Here we reported a cpDNA *psbA-trnH^{GUG}* IS as a molecular marker, which demonstrated the suitability to discriminate intraspecific relationships among regional *P. volubilis* cultivars, adding knowledge to understanding its genetic diversity. This outcome is crucial to track all the rudimentary SI ecotypes developed in tropical countries. Moreover, we consider that this work will contribute to generating stands for plant genetic improvement, primarily to further strategies based on chloroplast genetic transformation and understanding of evolutionary dynamics between organelle and nuclear genomes.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/genbank/> – MN912383, MN912384, MN912385, MN912386, and MN912387 (for markers), and MW591569 (for the chloroplast genome).

AUTHOR CONTRIBUTIONS

JÁ and DV-M were the principal investigators. SV-C contributed by extracting high molecular weight genomic DNA for ONT sequencing and analysis. CG-B and FG-C contributed equally to this work by collecting, processing, and analyzing pollen samples, and then observing, capturing, and processing images. CG-B, FG-C, and VR-R contributed by performing the DNA extractions from cultivars and the PCR experiments. SV-C performed all the bioinformatic analyses. Já, SV-C, DV-M, CG-B, and FG-C

contributed to the manuscript's writing. Já and VR-R guided every step of the experiments and helped with invaluable and profound suggestions. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Department of Biological Sciences of EAFIT University; the Colombian general royalty system (SGR, Spanish acronym); the Secretariat of Agriculture and Rural Development of the Department of Antioquia (Spanish: Secretaría de Agricultura y Desarrollo Rural del departamento de Antioquia); the Pontifical Bolivarian University (UPB, Spanish acronym); and The National University of Colombia (UNAL, Spanish acronym), and funded by EAFIT University by convening “project code 4600 000480.”

ACKNOWLEDGMENTS

We are very grateful to Luis Alfonso Arteaga and Luisa Posada for their useful advice, text corrections, and supervision for fluorescence microscopy and DAPI staining. We also want to give special thanks to Jorge Mario Restrepo for his art illustrations, the Apolo Scientific Computing Center who provided the computational resources for this work, and the academic staff from Universidad EAFIT, who facilitated the infrastructure and resources to achieve this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.667060/full#supplementary-material>

REFERENCES

- Al-Nakeeb, K., Petersen, T. N., and Sicheritz-Pontén, T. (2017). Norgal: extraction and de novo assembly of mitochondrial DNA from whole-genome sequencing data. *BMC Bioinformatics* 18:510. doi: 10.1186/s12859-017-1927-y
- Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F.-C., Falentin, C., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* 4, 879–887. doi: 10.1038/s41477-018-0289-4
- Bergsland, K. J., and Haselkorn, R. (1991). Evolutionary relationships among Eubacteria, Cyanobacteria, and Chloroplasts: evidence from the *rpoCl* Gene of *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* 173, 3446–3455. doi: 10.1128/jb.173.11.3446-3455.1991
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bonen, L. (2008). Cis- and trans-splicing of group II introns in plant mitochondria. *Mitochondrion* 8, 26–34. doi: 10.1016/j.mito.2007.09.005
- Bordignon, S., Ambrosano, G., and Rodrigues, P. (2012). Propagação in vitro de *Sacha inchi*. *Ciênc. Rural* 42, 1168–1172. doi: 10.1590/S0103-84782012005000049
- Brewbaker, J. L. (1967). The distribution and phylogenetic significance of Binucleate and Trinucleate pollen grains in the angiosperms. *Am. J. Bot.* 54, 1069–1083. doi: 10.1002/j.1537-2197.1967.tb10735.x
- Brewbaker, J. L., and Kwack, B. H. (1963). The essential role of calcium ion in pollen germination and pollen tube growth. *Am. J. Bot.* 50, 859–865. doi: 10.2307/2439772
- Brozynska, M., Furtado, A., and Henry, R. J. (2016). Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol. J.* 14, 1070–1085. doi: 10.1111/pbi.12454
- Cachique, D. (2006). *Estudio de la Biología Floral y Reproductiva en el Cultivo de Sacha inchi (Plukenetia volubilis L.) (Ingeniero agrónomo)*. Ph.D. tesis, Universidad Nacional de San Martín, Perú.
- Cardinal-McTeague, W. M., and Gillespie, L. J. (2016). Molecular phylogeny and pollen evolution of euphorbiaceae tribe Plukenetieae. *Syst. Bot.* 41, 329–347. doi: 10.1600/036364416x691759
- Cardinal-McTeague, W. M., Wurdack, K. J., Sigel, E. M., and Gillespie, L. J. (2019). Seed size evolution and biogeography of *Plukenetia* (Euphorbiaceae), a pantropical genus with traditionally cultivated oilseed species. *BMC Evol. Biol.* 19:29. doi: 10.1186/s12862-018-1308-9
- Cauz-Santos, L. A., Munhoz, C. F., Rodde, N., Cauet, S., Santos, A. A., Penha, H. A., et al. (2017). The chloroplast genome of *Passiflora edulis* (Passifloraceae) assembled from long sequence reads: structural organization and phylogenomic

- studies in Malpighiales. *Front. Plant Sci.* 8:334. doi: 10.3389/fpls.2017.00334
- CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12794–12797.
- Chirinos, R., Zorrilla, D., Aguilar-Galvez, A., Pedreschi, R., and Campos, D. (2016). Impact of roasting on fatty acids, tocopherols, phytosterols, and phenolic compounds present in *Plukenetia huayllabambana* seed. *J. Chem.* 2016:6570935. doi: 10.1155/2016/6570935
- Chirinos, R., Zuloeta, G., Pedreschi, R., Mignolet, E., Larondelle, Y., and Campos, D. (2013). *Sacha inchi* (*Plukenetia volubilis*): a seed source of polyunsaturated fatty acids, tocopherols, phytosterols, phenolic compounds and antioxidant capacity. *Food Chem.* 141, 1732–1739. doi: 10.1016/j.foodchem.2013.04.078
- Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163
- Correa, Q., Enrique, J., and Bernal, H. Y. (1990). *Especies Vegetales Promisorias de los Países del Convenio Andrés Bello (No. LC-0346)*. Bogota: Convenio Andrés Bello, CAB Junta del Acuerdo de Cartagena, JUNAC Ministerio de Educación y Ciencia de España Secretaría Ejecutiva del Convenio Andrés Bello, SECAB.
- Corriveau, J. L., and Coleman, A. W. (1988). Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Am. J. Bot.* 75, 1443–1458. doi: 10.2307/2444695
- Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17:134. doi: 10.1186/s13059-016-1004-2
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). JModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109
- Derrien, T., André, C., Galibert, F., and Hitte, C. (2007). AutoGRAPH: an interactive web server for automating and visualizing comparative genome maps. *Bioinformatics* 23, 498–499. doi: 10.1093/bioinformatics/btl618
- Dierckx, N., Mardulyn, P., and Smits, G. (2016). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45:gkw955. doi: 10.1093/nar/gkw955
- Downie, S. R., Ramanath, S., Katzdownie, D. S., and Llanas, E. (1998). Molecular systematics of Apiaceae subfamily Apioideae: phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer and plastid rpoC1 intron sequences. *Am. J. Bot.* 85, 563–591. doi: 10.2307/2446441
- Doyle, J. J., and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* 12, 39–40.
- Fauré, S., Noyer, J. L., Carreel, F., Horry, J. P., Bakry, F., and Lanaud, C. (1994). Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas (*Musa acuminata*). *Curr. Genet.* 25, 265–269. doi: 10.1007/BF00357172
- Feitosa, M. S. S. M. (2017). *Genômica de Organelas de Cana-de-Açúcar (Saccharum spp. Cultivar RB867515)*. Available online at: <http://repositorio.bc.ufg.br/tede/handle/tede/9005> (accessed January, 2019).
- Forsthoefel, N. R., Bohnert, H. J., and Smith, S. E. (1992). Discordant inheritance of mitochondrial and plastid DNA in diverse alfalfa genotypes. *J. Hered.* 83, 342–345. doi: 10.1093/oxfordjournals.jhered.a111228
- Gillespie, L. J. (1993). A synopsis of Neotropical *Plukenetia* (Euphorbiaceae) including two new species. *Syst. Bot.* 18, 575–592. doi: 10.2307/2419535
- Glick, B. R., and Patten, C. L. (2017). *Molecular Biotechnology: Principles and Applications of Recombinant DNA*, 5th Edn. Washington, DC: ASM Press. doi: 10.1128/9781555819378
- Greiner, S., Sobanski, J., and Bock, R. (2015). Why are most organelle genomes transmitted maternally? *Bioessays* 37, 80–94. doi: 10.1002/bies.201400110
- Hamaker, B. R., Valles, C., Gilman, R., Hardmeier, R. M., Clark, D., García, H. H., et al. (1992). Amino acid and fatty acid profiles of *Inca peanut*. *Cereal Chem.* 69, 461–463.
- Harris, S. A., and Ingram, R. (1991). Chloroplast DNA and biosystematics: the effects of intraspecific diversity and plastid transmission. *Taxon* 40, 393–412. doi: 10.2307/1223218
- Hu, X. D., Pan, B. Z., Fu, Q., Chen, M. S., and Xu, Z. F. (2018). The complete chloroplast genome sequence of the biofuel plant *Sacha Inchi*, *Plukenetia volubilis*. *Mitochondrial DNA B Resour.* 3, 328–329. doi: 10.1080/23802359.2018.1437816
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., et al. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. doi: 10.7717/peerj.7359
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455. doi: 10.1093/bioinformatics/btz305
- Krivankova, B., Cepkova, P. H., Ocelak, M., Juton, G., Bechyně, M., and Lojka, B. (2013). Preliminary study of diversity of *Plukenetia volubilis* based on the morphological and genetic characteristics. *Agric. Trop. Subtrop.* 45, 140–146. doi: 10.2478/v10295-012-0023-6
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12
- Lambowitz, A. M., and Zimmerly, S. (2011). Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* 3:a003616. doi: 10.1101/cshperspect.a003616
- Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32, 11–16. doi: 10.1093/nar/gkh152
- Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., et al. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol. Detect. Quantif.* 3, 1–8. doi: 10.1016/j.bdq.2015.02.001
- Li, C., Lin, F., An, D., Wang, W., and Huang, R. (2018). Genome sequencing and assembly by long reads in plants. *Genes (Basel)* 9:6. doi: 10.3390/genes9010006
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, X., Zhang, T. C., Qiao, Q., Ren, Z., Zhao, J., Yonezawa, T., et al. (2013). Complete chloroplast genome sequence of holoparasite *Cistanche deserticola* (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host *Haloxylon ammodendron* (Chenopodiaceae). *PLoS One* 8:e58747. doi: 10.1371/journal.pone.0058747
- Liston, A. (1992). Variation in the chloroplast genes RpoC1 and RpoC2 of the genus *Astragalus* (Fabaceae): evidence from restriction site mapping of a Pcr-amplified fragment. *Am. J. Bot.* 79, 953–961. doi: 10.1002/j.1537-2197.1992.tb13679.x
- Menezes, A. P. A., Resende-Moreira, L. C., Buzatti, R. S. O., Nazareno, A. G., Carlsen, M., Lobo, F. P., et al. (2018). Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences. *Sci. Rep.* 8:2210. doi: 10.1038/s41598-018-20189-4
- Messinger, W., Hummer, K., and Liston, A. (1999). Ribes (Grossulariaceae) phylogeny as indicated by restriction-site polymorphisms of PCR-amplified chloroplast DNA. *Plant Syst. Evol.* 217, 185–195. doi: 10.1007/bf00984364
- Milne, I., Stephen, G., Bayer, M., Cock, P. J. A., Pritchard, L., Cardle, L., et al. (2013). Using tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* 14, 193–202. doi: 10.1093/bib/bbs012
- Mondal, T., Bag, I., Sncvl, P., Garikapati, K. R., Bhadra, U., and Bhadra, M. P. (2018). Two way controls of apoptotic regulators consign DmArgonaute-1 a better clasp on it. *PLoS One* 13:e0190548. doi: 10.1371/journal.pone.0190548
- Motoo, K. (1981). Estimation of evolutionary distances between nucleotide sequences. *J. Mol. Evol.* 39, 315–329. doi: 10.1007/BF00160155
- Nagata, N. (2010). Mechanisms for independent cytoplasmic inheritance of mitochondria and plastids in angiosperms. *J. Plant Res.* 123, 193–199. doi: 10.1007/s10265-009-0293-x
- Nagata, T., Kurihara, Y., Matsuda, G., Saeki, J., Kohno, T., Yanagida, Y., et al. (1999). Structure and interactions with RNA of the N-terminal UAG-specific RNA-binding domain of hnRNP DO. *J. Mol. Biol.* 287, 221–237. doi: 10.1006/jmbi.1999.2616

- Ocelák, M., Čepková, P. H., Viehmannová, I., Dvořáková, Z., Huansi, D. C., and Lojka, B. (2015). Genetic diversity of *Plukenetia volubilis* L. assessed by ISSR markers. *Sci. Agric. Bohem.* 46, 145–153. doi: 10.1515/sab-2015-0029
- Palenik, B., and Haselkorn, R. (1992). Multiple evolutionary origins of prochlorophytes, the chlorophyll b-containing prokaryotes. *Nature* 355, 265–267. doi: 10.1038/355265a0
- Rambaut, A., Suchard, M., Xie, D., and Drummond, A. (2014). *Tracer v1.5*. Available online at: <http://beast.bio.ed.ac.uk/Tracer> (accessed December, 2019).
- Ramírez-Ríos, V., Franco-Sierra, N. D., Alvarez, J. C., Saldamando-Benjumea, C. I., and Villanueva-Mejía, D. F. (2016). Mitochondrial genome characterization of *Tecia solanivora* (Lepidoptera: Gelechiidae) and its phylogenetic relationship with other lepidopteran insects. *Gene* 581, 107–116. doi: 10.1016/j.gene.2016.01.031
- Reboud, X., and Zeyl, C. (1994). Organelle inheritance in plants. *Heredity* 72, 132–140. doi: 10.1038/hdy.1994.19
- Rivarola, M., Foster, J. T., Chan, A. P., Williams, A. L., Rice, D. W., Liu, X., et al. (2011). Castor bean organelle genome sequencing and worldwide genetic diversity analysis. *PLoS One* 6:e21743. doi: 10.1371/journal.pone.0021743
- Rodríguez, Á., Corazon-Guivin, M., Cachique, D., Mejía, K., Del Castillo, D., Renno, J.-F., et al. (2010). Diferenciación morfológica y por ISSR (Inter simple sequence repeats) de especies del género *Plukenetia* (Euphorbiaceae) de la *Amazonia peruana*: propuesta de una nueva especie. *Rev. Peru. Biol.* 17, 325–330. doi: 10.15381/rpb.v17i3.7
- Sager, R., and Lane, D. (1972). Molecular basis of maternal inheritance. *Proc. Natl. Acad. Sci. U.S.A.* 69, 2410–2413. doi: 10.1073/pnas.69.9.2410
- Sang, N., Claudio, P. P., Fu, Y., Horikoshi, N., Graeven, U., Weinmann, R., et al. (1997). Transforming region of 243R E1A contains two overlapping but distinct transactivation domains. *DNA Cell Biol.* 16, 1321–1333. doi: 10.1089/dna.1997.16.1321
- Saski, C., Lee, S. B., Daniell, H., Wood, T. C., Tomkins, J., Kim, H. G., et al. (2005). Complete chloroplast genome sequence of glycine max and comparative analyses with other legume genomes. *Plant Mol. Biol.* 59, 309–322. doi: 10.1007/s11103-005-8882-0
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., and Tabata, S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res.* 6, 283–290. doi: 10.1093/dnares/6.5.283
- Shaw, J., Lickey, E. B., Schilling, E. E., and Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* 94, 275–288. doi: 10.3732/ajb.94.3.275
- Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., et al. (1986). The complete nucleotide sequence of the tobacco chloroplast genome. *Plant Mol. Biol. Rep.* 4, 111–148. doi: 10.1007/BF02669253
- Smith, T., and Waterman, M. (1981). Identification of common molecular subsequences. *J. Mol. Biol.* 147, 195–197. doi: 10.1016/0022-2836(81)90087-5
- Sukumaran, J., and Holder, M. T. (2010). DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26, 1569–1571. doi: 10.1093/bioinformatics/btq228
- Tangphatsornruang, S., Uthapaisanwong, 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
- Tate, J. A., and Simpson, B. B. (2003). Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Syst. Bot.* 28, 723–737. doi: 10.1043/02-64.1
- Testolin, R., and Cipriani, G. (1997). Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in the genus *Actinidia*. *Theor. Appl. Genet.* 94, 897–903. doi: 10.1007/s001220050493
- Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2003). Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics* Chapter 2, 2–3.
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6—W11.
- TNAU (2015). *Emasculation and Pollination Techniques Oil Seeds*. TNAU Agritech portal, Crop improvement. Available online: https://agritech.tnau.ac.in/crop_improvement/crop_imprv_emasculation_oilseeds.html (accessed January, 2019).
- Twell, D., Oh, S. A., and Honys, D. (2006). “Pollen development, a genetic and transcriptomic view,” *Plant Cell Monographs*, Vol. 3, ed. R. Malhó (Berlin: Springer), 15–45. doi: 10.1007/7089_042
- Ünsal, S. G., Çiftçi, Y. Ö., Eken, B. U., Velioglu, E., Di Marco, G., Gismondi, A., et al. (2019). Intraspecific discrimination study of wild cherry populations from North-Western Turkey by DNA barcoding approach. *Tree Genet. Genomes* 15:16. doi: 10.1007/s11295-019-1323-z
- Valente, M. S. F., Chaves, F. C. M., Lopes, M. T. G., Oka, J. M., and Rodrigues, R. A. F. (2017). Crop yield, genetic parameter estimation and selection of *Sacha inchi* in central Amazon. *Pesqui. Agropecuária Trop.* 47, 226–236. doi: 10.1590/1983-40632016v4745758
- Veltri, D., Wight, M. M., and Crouch, J. A. (2016). SimpleSynteny: a web-based tool for visualization of microsynteny across multiple species. *Nucleic Acids Res.* 44, W41–W45.
- Villanueva, S., Alvarez, J. C., and García, C. (2020). Scripts for analysis of cpDNA. *zenodo.org*. doi: 10.5281/zenodo.3932047
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963
- Wambugu, P., Brozowska, M., Furtado, A., Waters, D., and Henry, R. (2015). Relationships of wild and domesticated rices (*Oryza* AA genome species) based upon whole chloroplast genome sequences. *Sci. Rep.* 5:13957. doi: 10.1038/srep13957
- Wang, S., Song, Q., Li, S., Hu, Z., Dong, G., Song, C., et al. (2018). Assembly of a complete mitogenome of *Chrysanthemum nankingense* using Oxford nanopore long reads and the diversity and evolution of Asteraceae mitogenomes. *Genes (Basel)* 9:547. doi: 10.3390/genes9110547
- Watson, L. E., Sayed-Ahmed, H., and Badr, A. (2000). Molecular phylogeny of Old World *Trifolium* (Fabaceae), based on plastid and nuclear markers. *Plant Syst. Evol.* 224, 153–171. doi: 10.1007/bf00986340
- Wurdack, K. J., Hoffmann, P., and Chase, M. W. (2005). Molecular phylogenetic analysis of uniovulate Euphorbiaceae (Euphorbiaceae sensu stricto) using plastid rbcL and trnL-F DNA sequences. *Am. J. Bot.* 92, 1397–1420. doi: 10.3732/ajb.92.8.1397
- Zhang, Q., and Sodmergen. (2010). Why does biparental plastid inheritance revive in angiosperms? *J. Plant Res.* 123, 201–206. doi: 10.1007/s10265-009-0291-z
- Zhang, Y., Shi, Y., Duan, N., Liu, B.-B., and Mi, J. (2019). Complete chloroplast genome of *Euphorbia tirucalli* (Euphorbiaceae), a potential biofuel plant. *Mitochondrial DNA B* 4, 1973–1974. doi: 10.1080/23802359.2019.1617069
- Zhong, Z. R., Li, N., Qian, D., Jin, J. H., and Chen, T. (2011). Maternal inheritance of plastids and mitochondria in *Cycas* L. (Cycadaceae). *Mol. Genet. Genomics* 286, 411–416. doi: 10.1007/s00438-011-0653-9

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 © 2021 Villanueva-Corrales, García-Botero, Garcés-Cardona, Ramírez-Ríos, Villanueva-Mejía and Álvarez. 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.