



Comparative analysis of chloroplast genomes for five *Dicliptera* species (Acanthaceae): molecular structure, phylogenetic relationships, and adaptive evolution

Sunan Huang^{1,2}, Xuejun Ge¹, Asunción Cano³, Betty Gaby Millán Salazar³ and Yunfei Deng¹

¹ Key Laboratory of Plant Resources Conservation & Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, Guangdong, China

² University of Chinese Academy of Sciences, Beijing, China

³ Facultad de Ciencias Biológicas y Museo de Historia Natural, Universidad Nacional Mayor de San Marcos, Lima, Peru

ABSTRACT

The genus *Dicliptera* (Justicieae, Acanthaceae) consists of approximately 150 species distributed throughout the tropical and subtropical regions of the world. Newly obtained chloroplast genomes (cp genomes) are reported for five species of *Dicliptera* (*D. acuminata*, *D. peruviana*, *D. montana*, *D. ruiziana* and *D. mucronata*) in this study. These cp genomes have circular structures of 150,689–150,811 bp and exhibit quadripartite organizations made up of a large single copy region (LSC, 82,796–82,919 bp), a small single copy region (SSC, 17,084–17,092 bp), and a pair of inverted repeat regions (IRs, 25,401–25,408 bp). Guanine-Cytosine (GC) content makes up 37.9%–38.0% of the total content. The complete cp genomes contain 114 unique genes, including 80 protein-coding genes, 30 transfer RNA (tRNA) genes, and four ribosomal RNA (rRNA) genes. Comparative analyses of nucleotide variability (Pi) reveal the five most variable regions (*trnY*-GUA-*trnE*-UUC, *trnG*-GCC, *psbZ*-*trnG*-GCC, *petN*-*psbM*, and *rps4*-*trnL*-UUA), which may be used as molecular markers in future taxonomic identification and phylogenetic analyses of *Dicliptera*. A total of 55–58 simple sequence repeats (SSRs) and 229 long repeats were identified in the cp genomes of the five *Dicliptera* species. Phylogenetic analysis identified a close relationship between *D. ruiziana* and *D. montana*, followed by *D. acuminata*, *D. peruviana*, and *D. mucronata*. Evolutionary analysis of orthologous protein-coding genes within the family Acanthaceae revealed only one gene, *ycf15*, to be under positive selection, which may contribute to future studies of its adaptive evolution. The completed genomes are useful for future research on species identification, phylogenetic relationships, and the adaptive evolution of the *Dicliptera* species.

Submitted 27 June 2019

Accepted 23 December 2019

Published 6 February 2020

Corresponding author

Yunfei Deng, yfdeng@scbg.ac.cn

Academic editor

Jasmine Janes

Additional Information and
Declarations can be found on
page 16

DOI 10.7717/peerj.8450

© Copyright
2020 Huang et al.

Distributed under
Creative Commons CC-BY 4.0

Subjects Evolutionary Studies, Genomics, Molecular Biology, Plant Science

Keywords *Dicliptera*, Chloroplast genome, Species identification, Phylogeny, Adaptive evolution, Molecular markers

OPEN ACCESS

INTRODUCTION

The genus *Dicliptera* Juss. belongs to tribe Justiceae of the family Acanthaceae; it consists of approximately 150 species, which are typically found in the tropical and subtropical regions of the world (Scotland & Vollesen, 2000; Mabberley, 2017; Hu et al., 2011). It is readily recognized by umbellately arranged, rarely solitary, cymose inflorescence units (cymules) subtended by conspicuously paired bracts, anthers with two partially or completely superposed thecae and, in the Palaeotropics, resupinate corollas that lack a rugula (Darbyshire, 2008). Eleven species of *Dicliptera* are found in Peru, most of which are located in the Andes (Brako & Zarucchi, 1993; León, 2006). Species such as *Dicliptera chinensis*, *D. peruviana*, and *D. verticillata* are used in traditional herbal medicines in China and Peru (Bussmann & Glenn, 2010; Horacio, Graciela & Percy, 2007; Telefo, Moundipa & Tchouanguep, 2002; Zhang, Zhu & Gao, 2010). Species delimitation within *Dicliptera* is difficult (Balkwill, Norris & Balkwill, 1996) due, in part, to the remarkable uniformity of the floral morphology in the majority of the taxa. Its taxonomy is also confounded by the presence of several widespread species complexes (Darbyshire, 2008). The taxonomic difficulties make it important to analyze the species using molecular analysis, infrageneric classification, and the relationships within *Dicliptera*. Kiel et al. (2017) conducted the only phylogenetic analysis of the tribe Justiceae, using five chloroplast regions (*ndhF-trnL*, *trnT-trnL-UAG*, *trnS-trnG*, *ndhA*, *rpl16*) and one nuclear region (*nrITS*). However, the interspecific relationships within *Dicliptera* have not been determined because of the limited number of samples available.

Five common Peruvian species were collected, representing the genus *Dicliptera* (*D. acuminata*, *D. peruviana*, *D. montana*, *D. mucronata*, and *D. ruiziana*). *D. acuminata*, *D. montana*, and *D. peruviana* are used in the agroindustrial industry (Victor, Jhan & Raúl, 2017); *D. peruviana* is a traditional herbal medicine used by the Andeans of Canta, Lima, Peru to alleviate stomach aches (Horacio, Graciela & Percy, 2007); *D. mucronata* is distributed mainly in Central America and is easily confused with *D. scabra* (Victor, Jhan & Raúl, 2017); *D. ruiziana* is found throughout Peru to elevations of about 3,000 m (Antonio Galán et al., 2009). All five species were collected in Southeast Peru and were distinguished from each other by the character of their leaves, bracts, bracteoles, and calyxes (Table 1). However, species delimitation is difficult without the aid of the flowers and the distribution of the five *Dicliptera* species is often overlapping. Species determinations in Central America are predominantly made by morphological comparisons as opposed to molecular comparisons. The study of the complete cp genomes of five *Dicliptera* species may encourage more effective species identification within the genus *Dicliptera*, especially in Central America.

The chloroplast genome (cp genome) is an independent genome that has been used in many evolutionary studies (Fan et al., 2018; Gao, Wang & Deng, 2018; George et al., 2015; Chen et al., 2018; Inkyu et al., 2018; Kim & Lee, 2004; Wang et al., 2016; Mader et al., 2018; Meng et al., 2018; Ma et al., 2017; Raubeson et al., 2007; Wang et al., 2008; Wu et al., 2018). It has a simple structure with a low molecular weight and multiple copies. Most of the cp genomes have circular structures with quadripartite organizations composed

Table 1 Morphological differences among *Dicliptera acuminata*, *D. peruviana*, *D. montana*, *D. mucronata* and *D. ruiziana*.

Species	<i>D. acuminata</i>	<i>D. peruviana</i>	<i>D. montana</i>	<i>D. mucronata</i>	<i>D. ruiziana</i>
Plant height	Ca. 60 cm	Ca. 60 cm	Ca. 50 cm	60–130 cm	Ca. 30 cm
Stem	Erect, branched, sulcate, hirsute	Erect, branched, sulcate, hirsute	Erect, branched, sulcate, pubescent	Erect, branched, sulcate, pubescent	Erect, branched, sulcate, pubescent
Leaf blade	Oblong-lanceolate, 3.5–7.0 × 1.5–2.5 cm, villous	Ovate, 3.5–6.0 × 2.5–4.0 cm, pubescent	Ovate, 1.0–1.5 × 0.8–1.0 cm, pilose when young, then glabrescent	Ovate, 3.0–3.5 × 1.5–2.0 cm, scarcely pilose	Ovate, 1.0–1.5 × 0.8–1.0 cm, pubescent
Leaf apex	Acuminate-acute	Acute	Acuminate	Acuminate-acute	Acute
Inflorescence	Verticillaster	Verticillaster	Spikelike thyrses	Verticillaster	Pedunculate cyme
Bracts	Lanceolate-linear, ciliate	Ovate, ciliate	Spatulate, gland-tipped pilose	Obovate-rhombic, pilose	Obovate, gland-tipped pilose
Bracteoles	Subulate, ciliate	Subulate,	Hyaline, asymmetrical, minute pilose	Linear-subulate, pilose	Subulate
Calyx lobes	Subulate, ciliate	Linear, hirsute	Lanceolate, minute pilose	Linear-lanceolate, margin minutely pubescent	Lanceolate, gland-tipped pilose
Corolla	Purple, outside pubescent	Purple, outside pubescent	Pale purple, outside pilose	Purplish red, outside pilose	Pink, outside pubescent
Style	scarcely pilose	glabrous	scarcely pilose	glabrous	scarcely pilose

of one large single copy region (LSC), one small single copy region (SSC), and a pair of inverted regions (IRs). However, there are numerous exceptions to the common structure, like the IR-lacking clade (IRLC) in Fabaceae (Fan et al., 2018; Gao, Wang & Deng, 2018; George et al., 2015; Chen et al., 2018; Inkyu et al., 2018; Kim & Lee, 2004; Wang et al., 2016; Mader et al., 2018; Meng et al., 2018; Schwarz et al., 2015). The complete cp genomes of more than 2,400 plants have been published, to date, in the NCBI database (<http://www.ncbi.nlm.nih.gov/genome>). The majority of plant cp genomes are 110 to 170 kb in length (Olmstead & Palmer, 1994; Wang et al., 2013; Wicke et al., 2011). The family Acanthaceae is a large family with approximately 230 genera and 4,300 species, yet only ten species from this family have fully sequenced cp genomes (Table S1).

The complete chloroplast genome is widely used for species identification, phylogenetic studies, and studies in adaptive evolution (Wang et al., 2016; Ma et al., 2017; Fan et al., 2018; Gao, Wang & Deng, 2018). Adaptive evolution is defined as the suitability for the improvement of a species during its evolutionary processes. It is always driven by evolutionary processes such as natural selection and leads to biological pressures and biodiversity at all levels of biological organization (Yang & Swanson, 2002; Scottphillips et al., 2013; Hall & Strickberger, 2008). The non-synonymous (K_A)/synonymous rate (K_S) ratio ($\omega = K_A/K_S$) provides a measure of selective pressure at the amino acid level. As suggested by Makalowski & Boguski (1998), the ω values less than one ($K_A/K_S < 1$) indicate that the gene is under negative selection and vice versa (Wojciech & Mark, 1998; Meng et al., 2018). Recent studies have detected many positively selected chloroplast genes ($K_A/K_S > 1$), such as the *ndhC*, *ndhJ*, *psbK*, *psbN*, *rpl14*, *rpl16*, *rps4*, *rps15*, *rps18*, *rps19*, *infA*, and

rpoB genes in *Echinacanthus* and the *petA*, *psbD*, *psbE*, *ycf3*, *psaI*, *rps4*, *psbM*, *ndhE*, *ndhG* and *rpoC1* genes in *Allium* (Gao, Wang & Deng, 2018; Xie et al., 2019).

The cp genomes of five *Dicliptera* species were sequenced, compared, and reported for the first time in this study. The five most variable regions were identified through genome comparison analysis and nucleotide variability; these were chosen as candidate molecular markers for taxonomic identification and systematic analysis in the future. Codon usage analysis was conducted to find the codon bias in the genus *Dicliptera*. 285 simple sequence repeats (SSRs), 21 polymorphic SSRs, and 229 long repeats were detected and described. The phylogenetic relationships of the five species and other members of the family Acanthaceae were analyzed. Finally, the orthologous protein-coding genes were identified in the family Acanthaceae and the selective pressure for these genes was analyzed. This work may contribute to future adaptive evolution analysis of the Acanthaceae species.

MATERIAL AND METHODS

Plant materials and DNA extraction

Fresh leaf tissues were collected during several botanical surveys conducted by South China Botanical Gardens, Chinese Academy of Sciences and Facultad de Ciencias Biológicas y Museo de Historia Natural, Universidad Nacional Mayor de San Marcos in Peru. The samples were dried in silica gel immediately after collection. Voucher specimens were deposited at the Museo de Historia Natural, Universidad Nacional Mayor de San Marcos (USM) and the herbarium of South China Botanical Garden, Chinese Academy of Sciences (IBSC) (Table 2). The specimens were visually identified by Deng Yunfei and the total genomic DNA was extracted using a modified CTAB method (Doyle & Doyle, 1987) that included 4% CTAB with 2% polyvinyl polypyrrolidone (PVP) (Yang, Li & Li, 2014).

Genome sequencing, assembly, and annotation

Short-insert (300–500 bp) libraries were constructed using the Nextera XT DNA Library Prep Kit (Illumina) following the manufacturer's instructions. Illumina × Ten instruments at BGI-Wuhan were used to perform paired-end (PE) sequencing for each sample. GetOrganelle v. 1.6.2 (Jin et al., 2018) was used to assemble the sequenced PE reads. *Andrographis paniculata* (GenBank accession no. NC_022451) served as a reference and the sequenced reads were filtered using Bowtie2 v. 2.3.5.1 (Langmead & Salzberg, 2012); SPAdes v. 3.13.1 (Bankevich et al., 2012) was used to assemble the filtered plastid reads and the final “fastg” files were reduced using the “slim_fastg.py” script in GetOrganelle to retain the pure plastid contigs; the filtered De Bruijn graph files were transferred to Bandage v. 0.8.1 (Wick et al., 2015) for visualization and to obtain the paths of the final “fasta” files of the cp genomes; finally, the genome structures of all five species were compared to the reference genome using Mauve v. 1.1.1 software (Darling, Mau & Perna, 2010) to determine the accuracy of the final genome. The assembled cp genome was annotated using PGA v. 2019 (Qu et al., 2019) using the annotated *A. paniculata* as a reference. The boundaries of the annotated genes were manually modified and coupled with CDSs in Geneious v. 2019.0.3 (Kearse et al., 2012). All transfer RNA (tRNA) genes were determined using tRNAscan-SE v. 2.0 (Schattner, Brooks & Lowe, 2005). The annotated cp genome files were submitted

Table 2 Species information and genome features of the chloroplast genomes of five *Dicliptera* species.

Species	<i>D. acuminata</i>	<i>D. peruviana</i>	<i>D. montana</i>	<i>D. ruiziana</i>	<i>D. mucronata</i>
Location	9.05°S, 77.81°W	11.41°S, 77.23°W	11.79°S, 77.05°W	15.87°S, 74.15°W	12.21°S, 76.82°W
Geographic region	Caraz, Paron, Peru	Lomas de Iguanil, Huaral province, Peru	Lomas de Carabayllo, Lima province, Peru	Lomas de Cháparra, Caravelí province, Peru	Santuario del Amancay, Lima Province, Peru
Voucher specimens No.	P10-091	P170099	P170177	P170209	P170492
Assembled reads	3745800	2103300	2272793	2108293	1927500
Mean coverage	3727.5	2092.0	2262.4	2097.8	1918.3
Size (bp)	150738	150811	150689	150750	150720
LSC length (bp)	82844	82919	82796	82843	82834
SSC length (bp)	17092	17090	17091	17091	17084
IR length (bp)	25401	25401	25401	25408	25401
CDSs total length	78714	78714	78714	78747	78717
Number of total genes	114	114	114	114	114
Number of Protein-coding genes	80	80	80	80	80
Number of tRNA genes	30	30	30	30	30
Number of rRNA genes	4	4	4	4	4
Overall GC content (%)	38.0%	38.0%	38.0%	38.0%	38.0%
GC content in LSC (%)	36.0%	36.0%	36.0%	36.0%	36.0%
GC content in SSC (%)	31.9%	31.7%	31.9%	31.9%	31.6%
GC content in IR (%)	43.3%	43.4%	43.3%	43.3%	43.4%
Genbank accession number	MK830556	MK833945	MK833946	MK833947	MK848596

to OGDRAW v. 1.3.1 (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) (Greiner, Lehwark & Bock, 2019) to create a circular cp genome map for each species. The five cp genomes (*D. acuminata*, [MK830556](#); *D. pruviana*, [MK833945](#); *D. montana*, [MK833946](#); *D. ruiziana*, [MK833947](#); *D. mucronata*, [MK848596](#)) were submitted to Genbank.

Genome comparison and structural analysis

Five cp genomes were compared using mVISTA v. 2.0 (Frazer et al., 2004) with ShuffleLAGAN mode and the annotation of *D. acuminata* as a reference (Brudno et al., 2003). The conserved regions were visualized on an mVISTA plot. DnaSP v. 5.1 (Librado & Rozas, 2009) was used to calculate the nucleotide variance (Pi) within the five *Dicliptera* species. The SC and IR boundaries were compared with *A. paniculata* as a reference. The Relative Synonymous Codon Usage (RSCU) of all protein-coding genes was analyzed for each species using CondoW v. 1.4.2 (Sharp & Li, 1987). MISA v. 1.0 (Thiel et al., 2003) was used to identify the simple sequence repeats (SSRs). The locations and lengths of long repeats (including forward, palindrome, complement, and reverse repeats) were analyzed using REPuter v. 2.74 (Kurtz et al., 2001) with the minimum repeat size set to 20 bp. Tandem repeats were identified using Tandem Repeats Finder v. 4.09 (Benson, 1999).

Phylogenetic analyses

Phylogenetic analysis was conducted for all of the sequenced cp genomes (each cp genome included only one IR), including the five species reported in this study and ten previously reported species of Acanthaceae (Table S1). *Sesamum indicum* (NC_016433) (Pedaliaceae) and *Mentha spicata* (NC_037247) (Lamiaceae) were used as outgroups. The complete cp genomes were aligned using MAFFT v. 1.3.7 (Katoh & Standley, 2013) and were adjusted manually as needed. The substitution models with the best fit were chosen by MrModeltest v. 2.3 (Nylander, 2004) based on the Akaike Information Criterion (AIC). RAxML v. 8.0.0 (Stamatakis, 2014) was used to reconstruct the phylogenetic relationship with the maximum likelihood (ML) method. Maximum parsimony (MP) analysis was run in Paup v. 4.0a (Swofford, 2003). Bootstrap values exceeding 50% were shown next to the corresponding branches. Bayesian inference (BI) analysis was conducted using MrBayes 3.2.7 (Ronquist & Huelsenbeck, 2003) with posterior probabilities (PP) obtained for each branch.

Selective pressure analysis

OrthoMCL v. 2.0 (Li, Stoeckert & Roos, 2003) was used to find the orthologous genes for the family Acanthaceae. The sequences for each orthologous gene were aligned separately using MAFFT v. 1.3.7 (Katoh & Standley, 2013). The nonsynonymous (K_A) and synonymous (K_S) substitution rates were calculated using PAML v. 4.9 with the codeml program to analyze the selective pressures of every orthologous gene sequence (Yang, 2007). The ω value ($\omega = K_A/K_S$) was estimated using the method reported by Yang & Nielsen (2000). The genes under positive selection were confirmed by computing the likelihood ratio tests (LRTs).

RESULTS

Chloroplast genome features

The average assemblies of the five cp genomes varied from 1918.3 to 3727.5 bp. The cp genome sequences were 150,738 bp (*Dicliptera acuminata*), 150,811 bp (*D. peruviana*), 150,689 bp (*D. montana*), 150,750 bp (*D. ruiziana*), and 150,720 bp (*D. mucronata*) in length (Table 2). Each of the sequences encoded 80 protein-coding genes, 30 transfer RNA (tRNA) genes, and 4 ribosomal RNA (rRNA) genes (Table 2 and Table S2; Fig. 1), of which three protein-coding genes had two introns and nine had one intron. The *Rps12* gene was trans-spliced because of the locations of the first exon at the LSC and the other two exons at the IRs. Six protein-coding genes (*rpl2*, *rpl23*, *ycf2*, *ycf15*, *ndhB*, and *rps7*), seven tRNA (*trnI*-CAU, *trnI*-GAU, *trnL*-CAA, *trnN*-GUU, *trnR*-ACG, *trnV*-GAC, and *trnA*-UGC), and all four rRNA (*rrn4.5*, *rrn5*, *rrn16* and *rrn23*) had two copies because of their location at the IR regions. The *rps12* gene was identified as the pseudogene in *D. mucronata* by the existence of internal stop codons.

The five cp genomes displayed a typical quadripartite structure, including a large single copy (LSC) region (range from 82,796–82,919 bp), a small single copy (SSC) region (range from 17,084–17,092 bp), and a pair of inverted repeat (IR) regions (25,401–25,408 bp; Table 2). The Guanine-Cytosine (GC) content of the cp genomes of the five *Dicliptera*

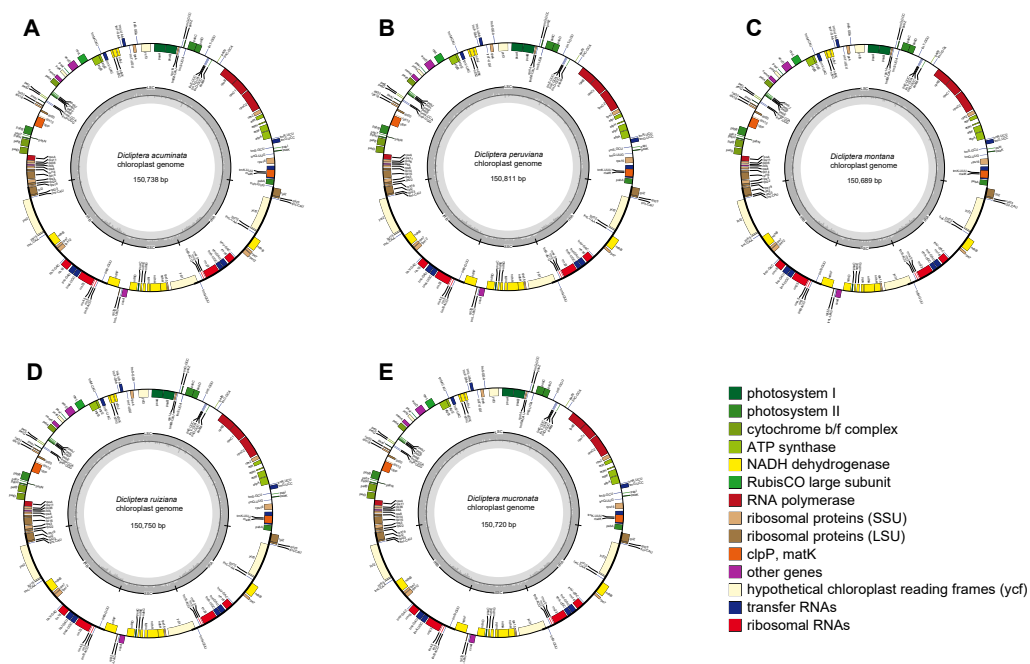


Figure 1 Gene maps of chloroplast genomes. (A) *Dicliptera acuminata*; (B) *D. peruviana*; (C) *D. montana*; (D) *D. ruiziana*; (E) *D. mucronata*. Genes shown outside of the circle are transcribed clockwise, whereas genes inside of the circle are transcribed counterclockwise. The colored bars indicate known protein-coding genes, tRNA and rRNA. The dark gray area in the inner circle indicates GC content, while the light gray area indicates AT content. LSC, large single copy; SSC, small single copy; IR, inverted repeats.

Full-size DOI: 10.7717/peerj.8450/fig-1

was approximately 38.0%. The GC content in the IR regions (43.3–43.4%) was noticeably above that of the LSC (36.0%) and SSC (31.6–31.9%) regions in each cp genome.

IR contraction and expansion

The IR/LSC and IR/SSC borders of the cp genomes of the five *Dicliptera* species and the *A. paniculata* were compared to identify the expansion or contraction of the IR (Fig. 2). The genes, *rps19*, *rpl2*, *ndhF*, *ycf1*, and *psbA* were present at the juncture of the LSC/IRA, IRA/SSC, SSC/IRb, and IRb/LSC borders. The five *Dicliptera* species have identical IR/SC borders with the exception of the *ycf1* gene at the SSC/IRb border in *D. montana*, which varies from those in *A. paniculata*. There are 167 bp from the border of *rpl2* to the juncture of LSC/IRA of *Dicliptera*, while this distance is just 57 bp in *A. paniculata*. The *rps19* gene is on the LSC/IRA border in all five *Dicliptera* species, indicating that this border has moved toward the LSC region when compared to *A. paniculata*. The *ndhF* gene in *Dicliptera* is situated at the juncture of the IRA/SSC region and has 117 bp sequences located at IRA, however, the comparable region in *A. paniculata* is 40 bp long, indicating that the IRA/SSC boundary has moved toward the SSC region. The *ycf1* gene is located at the IRb/SSC junction and the border has moved toward the SSC region because there are 4,543 bp and 4,560 bp sequences situated at SSC in *Dicliptera* and *A. paniculata*, respectively. The *ycf1* gene duplications located in IRA are 811–812 bp in *Dicliptera* and 982 bp in *A. paniculata*,

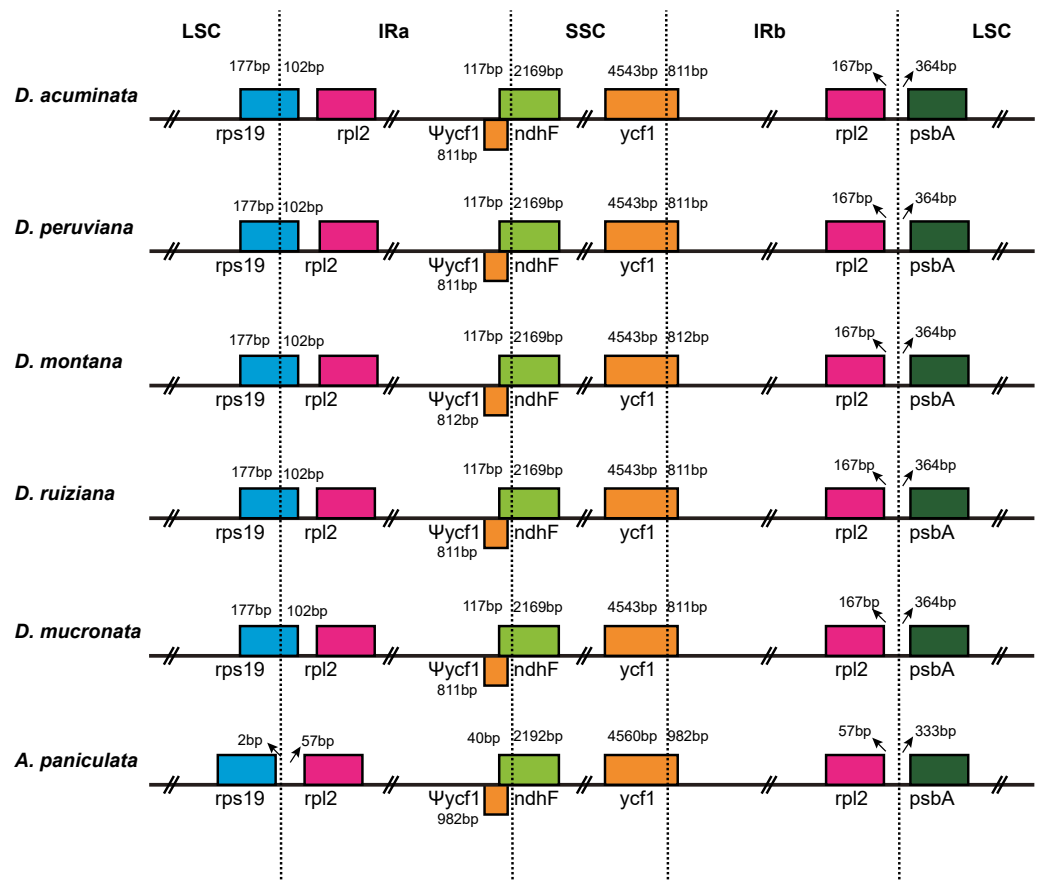


Figure 2 Comparison of the border regions of the LSC, SSC and IR among six Acanthaceae chloroplast genomes. The IRb/SSC junction extended into the *ycf1* genes creating various lengths of *ycf1* pseudogenes (Ψ *ycf1*) among the six cp genomes. The number above, below or adjacent to genes shows the distance between the ends of genes and the boundary sites. The figure features are not to scale.

Full-size [DOI: 10.7717/peerj.8450/fig-2](https://doi.org/10.7717/peerj.8450/fig-2)

indicating a slight expansion of the IR regions. Likewise, the space in *Dicliptera* from *psbA* to the IRb/LSC boundary (364 bp) is enlarged compared to that of *A. paniculata* (333 bp). These findings reveal that the IR regions in the cp genome of *Dicliptera* have expanded compared to those of *A. paniculata*.

Comparative chloroplast genome analysis

The annotated *D. acuminata* cp genome was used as a reference in mVISTA for the alignment of the cp genome among the five *Dicliptera* species (Fig. 3). The size and gene order of the chloroplast genomes of the *Dicliptera* species are conserved, but some divergent regions were identified, including the *trnH*-GUG, *rpl16*, *petN-psbM*, *trnS-trnG*, *trnT-trnF*, *ndhC-trnV*, *petA-psbJ*, and *rps12-clpP* genes. The nucleotide variability (Pi) was calculated for the coding and non-coding regions, respectively, in order to further confirm the sequence variations (Table S3, Fig. 4). However, the Pi values are rather low among the five species (0 to 0.02230) and a total of 5 hotspot regions were identified with Pi >0.005 (*trnY*-GUA-*trnE*-UUC, *trnG*-GCC, *psbZ-trnG*-GCC, *petN-psbM*, and *rps4-trnL*-UUA).

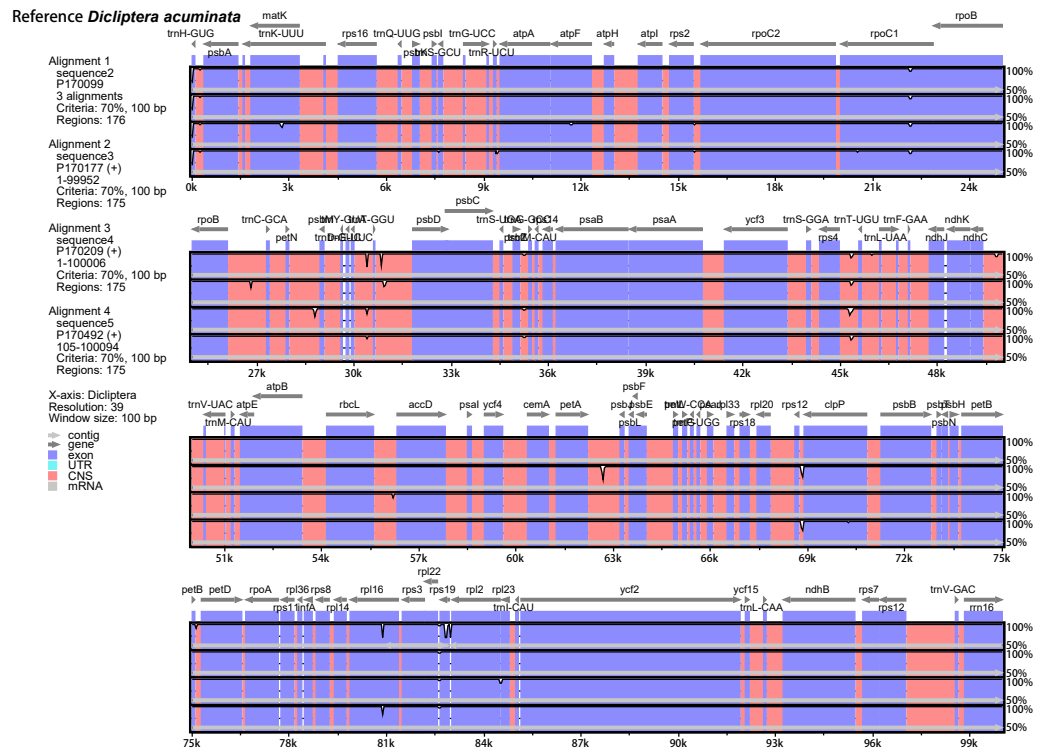


Figure 3 Comparison of the five *Dicliphora* chloroplast genomes using mVISTA. CNS indicates conserved noncoding sequences. The Y-scale represents the percent identity between 50% and 100%.

Full-size [DOI: 10.7717/peerj.8450/fig-3](https://doi.org/10.7717/peerj.8450/fig-3)

Codon usage

A total of 78714-78747 bp protein-coding genes were identified in the five *Dicliphora* cp genomes, accounting for 52.19%-52.24% of the entire genome sequence. These genes are encoded in 26238-26249 codons. Leucine (Leu, encoded by UUA, UUG, CUU, CUC, CUA and CUG) was the most frequent amino acid encoded by these codons, comprising 2,825-2,828 (10.8%) of the total number of codons; cysteine (Cys, encoded by UGU and UGC) was the least frequently encoded amino acid, with 303-305 codons (1.2%) (Fig. 5). For all codons, from the first to third position, the AU contents are 55.8%-60.2%, 60.6%-64.1%, and 64.9%-65.8%, respectively. The majority of the preferred codons (RSCU >1) ended with A or U, with the exception of UUG (RSCU = 1.25). This phenomenon is congruent with the results from other plant studies (Table S4) (Lu et al., 2018; Yang et al., 2014; Yi & Kim, 2012).

Simple Sequence Repeats (SSRs) and long repeat analysis

A total of 285 SSRs were identified in this study. The numbers of mono-, di-, tri-, and tetranucleotides were 157, 37, 55, and 36, respectively (Tables S5 and S7). Mononucleotide repeats were the most common repeats, accounting for 55.1% of the total repeats, while dinucleotides repeats accounted for 13.0%, and other SSRs occurred less frequently (Fig. 6A). The SSR varied in number and type depending on the species; *D. acuminata* and *D. montana* (58) had the most SSRs and *D. mucronata* (55) had the least (Fig. 6B, Table 3).

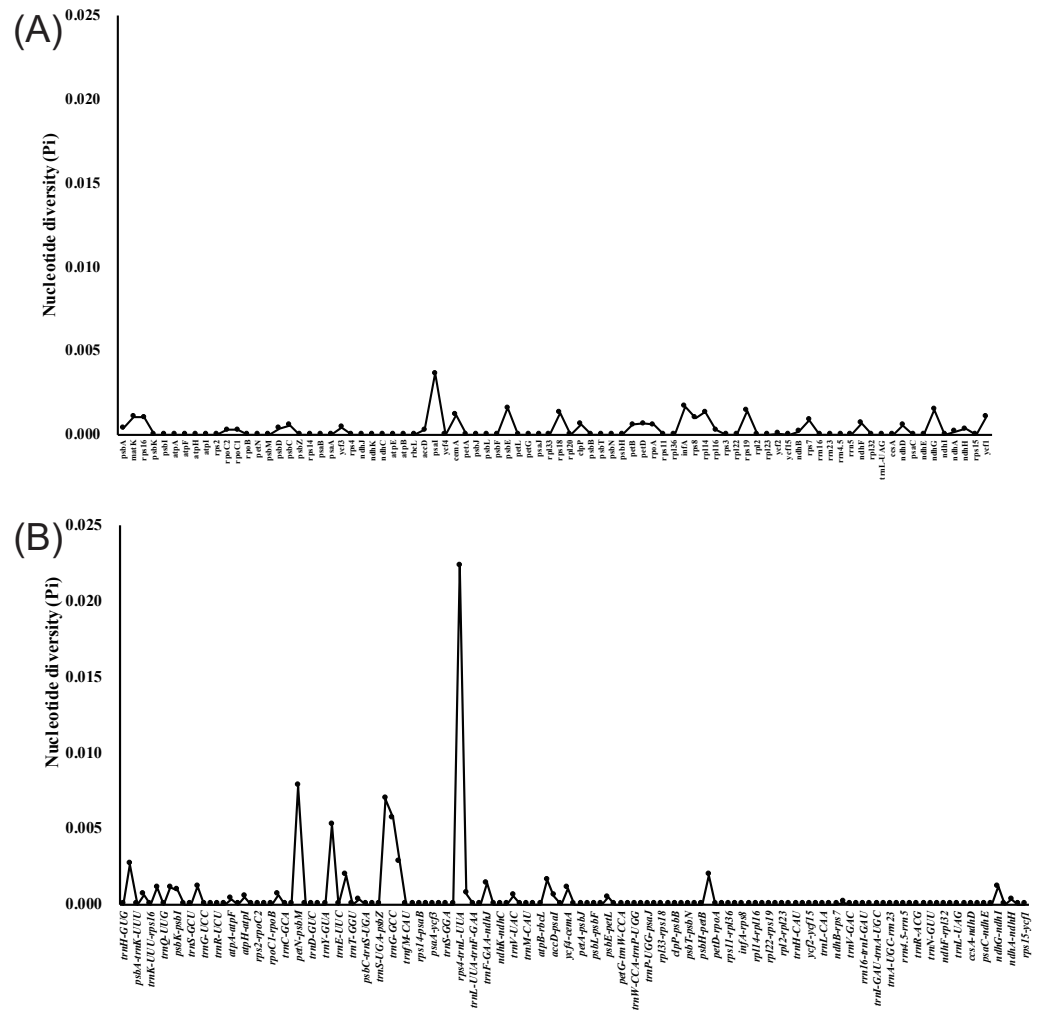


Figure 4 Comparative analysis of the nucleotide diversity (Pi) value among five *Dicliptera* chloroplast genomes. (A) Coding regions. (B) Non-coding regions.

Full-size [DOI: 10.7717/peerj.8450/fig-4](https://doi.org/10.7717/peerj.8450/fig-4)

Five categories of long repeats (tandem, complement, forward, palindromic and reverse repeats) were detected and analyzed in the five *Dicliptera* cp genomes (Tables S6 and S7, Fig. 6C). 229 long repeats were identified and were composed of 128 tandem repeats, 8 complement repeats, 53 forward repeats, 38 palindromic repeats, and 2 reverse repeats (Fig. 6C). The number of repeats was highest in *D. peruviana* (56) and lowest in *D. montana* (41) (Fig. 6D).

Phylogenetic analyses

GTR and SYM+G were the best fit models used for the ML and BI trees to display the completed cp genomes. The data matrix for all of the MP, ML, and BI analyses revealed trees with highly congruent topologies. The phylogenetic relationships within the 17 cp genomes sequences analyzed were well-resolved (Fig. 7). Our phylogenetic analyses strongly support the monophyly of the *Dicliptera* species [$BP_{(MP)} = 100\%$, $BP_{(ML)} = 100\%$, $PP = 1.0$],

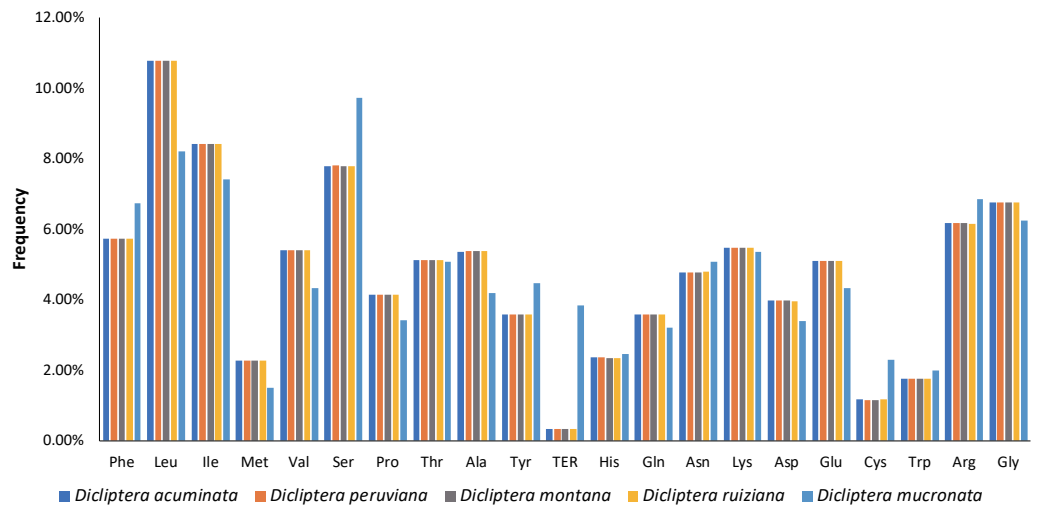


Figure 5 Amino acid frequencies in five *Dicliptera* species protein-coding sequences. As shown in the column diagram, Leucine was the most frequent amino acid (10.8%), Cysteine was the least (1.2%).

Full-size [DOI: 10.7717/peerj.8450/fig-5](https://doi.org/10.7717/peerj.8450/fig-5)

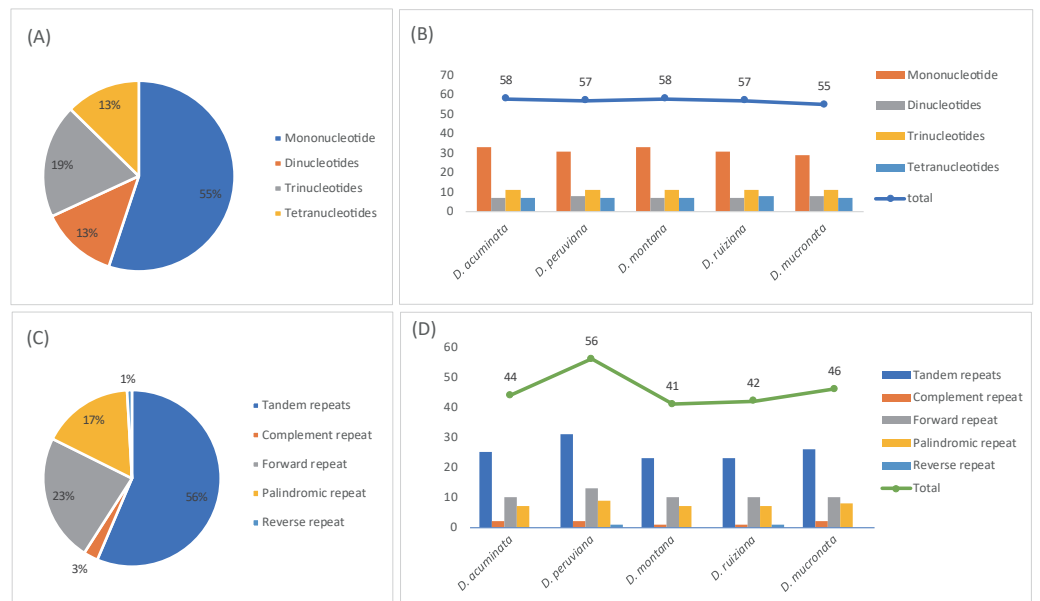


Figure 6 The type and presence of simple sequence repeats (SSRs) and long repeated sequences in the chloroplast genomes of five *Dicliptera* species. (A) Percentage of SSR types; (B) number of SSRs and their types; (C) percentage of five repeat types; (D) number of five repeats types.

Full-size [DOI: 10.7717/peerj.8450/fig-6](https://doi.org/10.7717/peerj.8450/fig-6)

in which *D. ruiziana* has the closest relationship with *D. montana* [$BP_{(ML)} = 52\%$, $PP = 0.96$], followed by *D. acuminata* [$BP_{(ML)} = 69\%$, $PP = 0.97$], *D. peruviana* [$BP_{(MP)} = 99.8\%$, $BP_{(ML)} = 100\%$, $PP = 1.0$], and *D. mucronata* ($BP_{[MP]} = 100\%$, $BP_{[ML]} = 100\%$, $PP = 1.0$).

Table 3 The polymorphic SSRs among five *Dicliptera* species. LSC, large single copy; SSC, small single copy.

Type	<i>D. acuminata/D. peruviana/ D. montana/D. ruiziana/D. mucronata</i>	Location	Regions
A	11/11/10/11/13	<i>psbI-trnS-GCU</i>	LSC
T	0/0/0/10/0	<i>trnS-GCU-trnG-UCC</i>	LSC
T	13/13/14/13/13	<i>trnR-UUCU-atpA</i>	LSC
A	12/11/11/10/11	<i>aptF</i>	LSC
A	10/10/10/0/0	<i>atpI-rps2</i>	LSC
TA	0/0/0/0/8	<i>trnE-UUC-trnT-GGU</i>	LSC
AT	0/9/0/0/0	<i>trnE-UUC-trnT-GGU</i>	LSC
T	13/10/12/16/14	<i>psbZ-trnG-GCC</i>	LSC
G	11/11/12/9/8	<i>psbZ-trnG-GCC</i>	LSC
T	10/10/10/10/11	<i>rps4-trnT-UGU</i>	LSC
ATAA	3/3/3/6/3	<i>rps4-trnT-UGU</i>	LSC
TA	6/7/7/7/7	<i>rps4-trnT-UGU</i>	LSC
T	12/11/13/13/11	<i>ndhC-trnV-UAC</i>	LSC
T	12/11/11/12/11	<i>psaI-ycf4</i>	LSC
T	11/10/11/11/10	<i>petG-trnW-CCA</i>	LSC
T	11/10/10/10/0	<i>clpP</i> intron	LSC
TA	7/6/6/6/6	<i>rpl22-rps19</i>	LSC
T	10/10/10/10/11	<i>ndhF-rpl32</i>	SSC
A	10/11/10/10/10	<i>ndhD-psaC</i>	SSC
G	11/0/11/11/10	<i>ndhG-ndhI</i>	SSC
A	11/10/10/10/0	<i>ndhA</i> intron	SSC

Selective pressure events

There were 68 orthologous protein-coding genes found in this study. The ω values of most genes were low ($\omega < 1$), approaching zero, except for the *ycf15* gene found in the SSC region, which had a ratio of 1.4453. The ω value of the *matK* gene was 0.9418, indicating a relaxed selection (Table S8, Fig. 8).

DISCUSSION

Sequence variation among five *Dicliptera* species

The results of our study showed that the cp genomes of five Peruvian *Dicliptera* species were similar in structure, content, and order (Table 2, Fig. 1). The cp genomes ranged in size from 150,689 bp to 150,811 bp in *D. montana* and *D. peruviana*, respectively. These structures are longer than the cp genome of *A. paniculata* (15,249 bp) (Ding et al., 2016). The genome size of all *Dicliptera* is relevant to LSC variation (Table 2) and this phenomenon has also been identified in other species (Zhao et al., 2018; Li, Zhao & Liu, 2018; Meng et al., 2018). mVISTA revealed a low divergence between the genomes of the five *Dicliptera* species, suggesting that the cp genomes were conserved. The IR regions were more highly conserved than the SC regions and the coding regions were less variable than the non-coding regions, which is also found in other angiosperms (Gao, Wang &

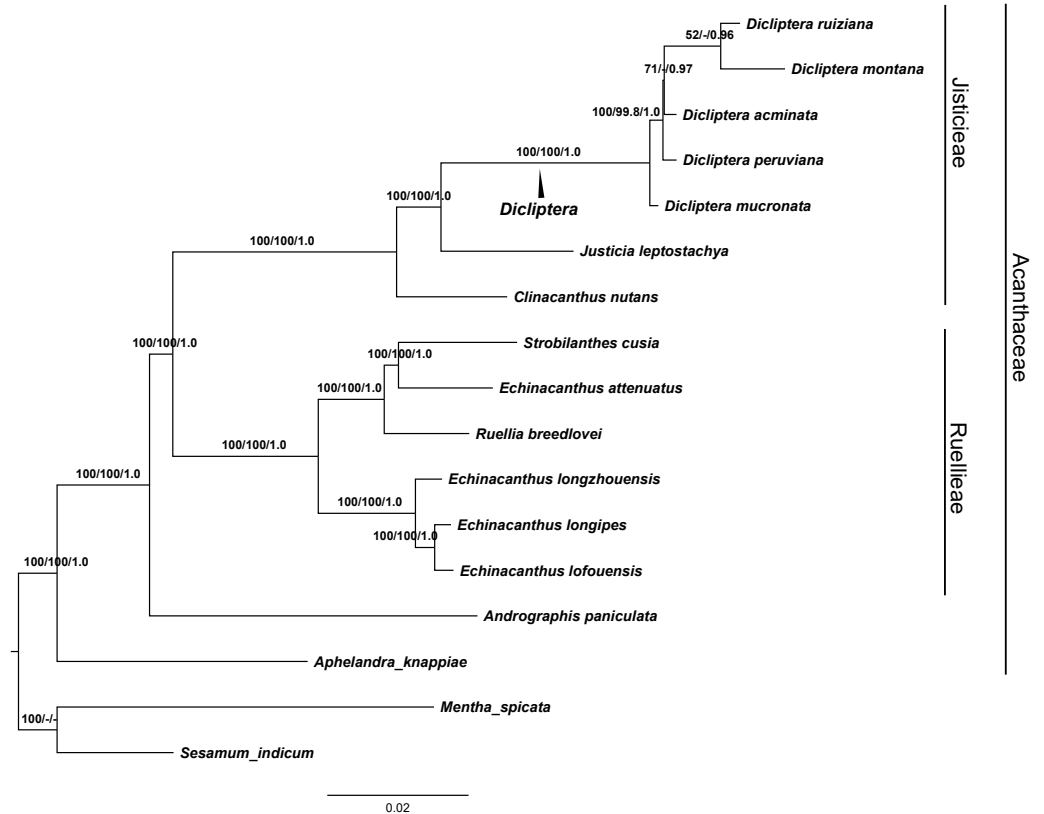


Figure 7 The maximum Likelihood (ML) tree of Acanthaceae. Numbers associated with branches are ML bootstrap values, MP bootstrap values and Bayesian posterior probabilities, respectively. Hyphens indicate the bootstrap support or posterior probability lower than 50% or 0.5. *Mentha spicata* (NC_037247) and *Sesamum indicum* (NC_016433) were used as outgroups.

Full-size DOI: 10.7717/peerj.8450/fig-7

Deng, 2018; Meng et al., 2018; Li, Zhao & Liu, 2018; Yan et al., 2019). Khakhlova & Bock (2006) suggested that gene divergence with less variability in the IR and coding regions may be a result of copy corrections during gene conversion, which can correct or delete the mutation. Codons were shown to have a strong tendency toward A or U at the third codon position, which is similar to the expression of an A/U ending in other plants (Gao et al., 2017; Clegg et al., 1994; Mader et al., 2018; Meng et al., 2018). This phenomenon may explain why the Adenine-Thymine (AT) content is slightly higher than the GC content in the cp genome of *Dicliptera*.

IR expansion analysis

IR regions are the most conserved regions in the cp genomes. Frequent expansions and contractions at the junctions of SSR and LSC with IRs illustrate the relationships among taxa and have been recognized as evolutionary signals (Khakhlova & Bock, 2006; Inkyu et al., 2018; Lu et al., 2018; Raubeson et al., 2007; Wang et al., 2008). In this study, only a few variations were found among the five *Dicliptera* species. When compared with the cp genome of *A. paniculata*, the IR regions of the cp genome of *Dicliptera* revealed a

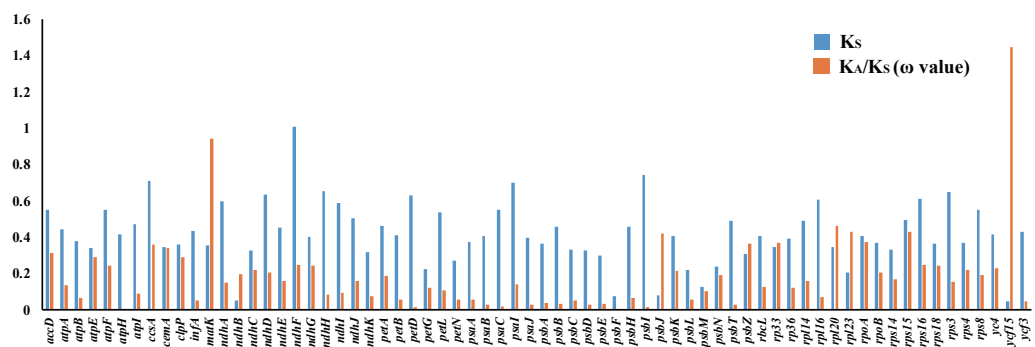


Figure 8 Synonymous (KS) substitution rates and ω values ($\omega = KA/KS$) among all Acanthaceae species. As shown in the column diagram, the order of genes is alphabetical. The ω value of the *ycf15* gene (1.4453) is clearly higher than other genes.

Full-size [DOI: 10.7717/peerj.8450/fig-8](https://doi.org/10.7717/peerj.8450/fig-8)

slight expansion. The size differences among the cp genome of the five *Dicliptera* species (150,689–150,811 bp) and *Andrographis paniculata* (150,249 bp) are congruent with the results of previous studies. The contractions and expansions at the LSC/IRs and SSC/IRs junctions contribute to the size variations of the cp genomes (*Kim & Lee, 2004; Raubeson et al., 2007*). Gene conversion during speciation is thought to be responsible for small IR expansions or contractions. (*Wang et al., 2008; Goulding et al., 1996; Khakhlova & Bock, 2006; Meng et al., 2018; Choi, Jansen & Ruhlman, 2019*).

Molecular markers

Simple sequence repeats (SSRs), known as microsatellites, are short stretches of DNA which consist of only one, or a few, tandemly repeated nucleotides. Polymorphic SSRs are the same units with different unit numbers located in the homologous regions; these are frequently used to identify variable species complexes (*Diethard & Manfred, 1984; Jerzy & Charit, 1995; George et al., 2015; Gao, Wang & Deng, 2018*). 21 SSRs were identified as polymorphic SSRs among the five *Dicliptera* species; these may be used as candidate genetic markers for further phylogenetic studies in the genus *Dicliptera* (Table 3). The presence of these repeats indicates that these regions are important hotspots for genome recombination. All polymorphic SSRs are located in LSC/SSC regions. Polymorphic SSRs are mainly distributed in non-coding regions, which are also highly variable regions in the chloroplast genomes (*Asaf et al., 2017*). The presence of long sequence repeats are indicators of mutational hotspots (*Borsch & Quandt, 2009; Jiang et al., 2018*).

The *ycf1* gene was previously reported for its use in DNA barcodes due to its abundance of variable sites (*Kurt et al., 2008; Gernandt et al., 2009; Dong et al., 2012; Drew & Sytsma, 2013; Dong et al., 2015*). *Shingo et al. (2013)* concluded that the *ycf1* gene is crucial for plant viability because it encodes the *Arabidopsis* protein, Tic214, which is essential for photosynthetic protein import. A substantial size difference was noted between the *ycf1* gene of the five *Dicliptera* species (5,354–5,355 bp) and *A. paniculata* (5,542 bp). The nucleotide variability of the *ycf1* gene ($P_i = 0.0109$) was slightly higher than that of the regions *matK* ($P_i = 0.00107$) and *rpl16* ($P_i = 0.00103$). The two regions are currently used in

the DNA barcodes for the tribe Justiceae and other angiosperms (Kiel *et al.*, 2017; Särkinen & George, 2013). Therefore, the *ycf1* gene should be a potential molecular marker for the *Dicliptera* species as well. The most divergent regions among the *Dicliptera* species, as determined by a comparison of nucleotide variability, are *rps4-trnL-UUA* ($P_i = 0.02230$), *petN-psbM* ($P_i = 0.00783$), *psbZ-trnG-GCC* ($P_i = 0.00697$), *trnG-GCC* ($P_i = 0.00571$), and *trnY-GUA-trnE-UUC* ($P_i = 0.00526$). The variability in these regions was much higher than that in the coding regions and the highly variable regions identified here could be validated and used as molecular markers in future species delimitation and phylogenetic studies.

Phylogenetic analyses

The phylogenetic trees (MP, ML and BI) demonstrated a significant relationship among Acanthaceae with high bootstrap values and posterior probabilities (Fig. 7). The genus *Aphelandra* was found to be the earliest diverging lineage; tribes Justiceae and Ruellieae are strongly supported as monophyletic groups [$BP_{(MP)} = 100\%$, $BP_{(ML)} = 100\%$, $PP = 1.0$] that form sister groups with each other. The results are consistent with previous studies (Mcdade, Daniel & Kiel, 2008; Huang, Deng & Ge, 2019). Phylogenetic analysis strongly supports *Dicliptera* as a monophyletic group. The clade formed by all five *Dicliptera* species is a sister to the species *Justicia leptostachya* [$BP_{(MP)} = 100\%$, $BP_{(ML)} = 100\%$, $PP = 1.0$], which supports the conclusion by Kiel *et al.* (2017) that the genus *Dicliptera* should be placed in the justicioid lineage. *D. mucronata* and *D. peruviana* are the first and second diverging clades among the five *Dicliptera* species; *D. acuminata*, *D. ruiziana*, and *D. montana* are species that can confidently be assigned to one clade. Trees with the same topology were retrieved from the ML and BI analyses. *D. ruiziana* was most closely related to *D. montana*, followed by *D. acuminata*. However, the relationships among *D. acuminata*, *D. ruiziana* and *D. montana* were not resolved using MP analysis. The sister relationship between *D. ruiziana* and *D. montana* is supported by their shared morphological characteristics, including a lanceolate calyx and ovate leaves of $1.0\text{--}1.5 \times 0.8\text{--}1.0$ cm versus the has subulate calyx and oblong-lanceolate leaves of $3.5\text{--}7.0 \times 1.5\text{--}2.5$ cm of *D. acuminata* (Table 1).

Adaptive evolution analysis

Positively selected genes are known to play a key role in adapting to different environments (Wang *et al.*, 2016; Ma *et al.*, 2017; Fan *et al.*, 2018; Gao, Wang & Deng, 2018; Wu *et al.*, 2018) and it is important to understand the adaptive evolutionary history of Acanthaceae. Orthologous genes are a particular class of homologous genes that diverged following the speciation of their host species; they are ideal markers for analyzing evolutionary history (Gargaud *et al.*, 2015). 68 protein-coding genes were found to be orthologous in the family Acanthaceae and the selective pressure of these genes was measured. The resulting measurements found that most genes in the family Acanthaceae were under negative selection ($\omega < 1$) except for *ycf15* ($\omega = 1.4453$). According to previous studies, the *ycf15* gene is a member of the PFAM protein family accession PF10705 (Sara *et al.*, 2018) and was not considered to be a protein-coding gene because of its unknown function (Steane, 2005; Feng *et al.*, 2018). The *ycf15* gene acts as a pseudogene in some species because of

its premature stop codons ([Chen et al., 2018](#); [Jiang et al., 2018](#)). The *ycf15* gene should be further investigated for its role in adaptive evolution and gene function.

CONCLUSIONS

Our study sequenced and analyzed the complete cp genomes of five Peruvian *Dicliptera* species (*D. acuminata*, *D. peruviana*, *D. montana*, *D. ruiziana*, and *D. mucronata*) for the first time. The identification of the chloroplast genomes and the new molecular markers of these five species contributes to the genetic resources available for future identification and phylogenetic studies. The goal of this study was to determine the appropriate DNA barcode for the identification of the *Dicliptera* species, especially those found in Peru. The genes *ycf1*, *rps4-trnL-UUA*, *petN-psbM*, *psbZ-trnG-GCC*, *trnG-GCC*, and *trnY-GUA-trnE-UUC* were found to be the most suitable DNA barcode for the species *Dicliptera*. The interspecies relationships among the five species were resolved. However, further phylogenetic analysis using additional genes from the nucleus will have to be conducted in order to understand how gene introgression and hybridization affects the phylogeny of *Dicliptera* ([Birky, 1995](#); [Meng et al., 2018](#); [Lu et al., 2018](#)). A single gene, *ycf15*, was found to be positively selected among all of the protein coding genes that were identified. This gene may play an important role in the adaptive evolution of the Acanthaceae species and its function should to be further studied. Our genome data enhances the cp genome resources for the family Acanthaceae and our understanding of its species identification, phylogeny, and evolutionary history.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was financially supported by the International Partnership Program of Chinese Academy of Sciences (Grant No. 151644KYSB20160005, GJHZ1620), the National Natural Science Foundation of China (Grant no. 31470302, 31670191), and “One-Three-Five” Strategic Planning of SCBG, CAS to Yunfei Deng and Xuejun Ge. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

International Partnership Program of Chinese Academy of Sciences: 151644KYSB20160005, GJHZ1620.

National Natural Science Foundation of China: 31470302, 31670191.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Sunan Huang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Xuejun Ge conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Asunción Cano and Betty Gaby Millán Salazar performed the experiments, prepared figures and/or tables, apply the collecting permission, and approved the final draft.
- Yunfei Deng conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The chloroplast sequences are available at GenBank: [MK830556](#), [MK833945](#), [MK833946](#), [MK833947](#), [MK848596](#).

Data Availability

The following information was supplied regarding data availability:

Data is available at NCBI SRA: [PRJNA531624](#), [PRJNA531625](#), [PRJNA531630](#), [PRJNA531639](#), [PRJNA531641](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.8450#supplemental-information>.

REFERENCES

- Antonio Galán DM, Eliana Linares P, José CDLC, José Alfredo Vicente O. 2009.** Nuevas observaciones sobre la vegetación del sur del Perú. Del Desierto Pacífico al Altiplano. *Acta Botanica Malacitana* **34**:107–144.
- Asaf S, Khan AL, Khan MA, Waqas M, Kang SM, Yun BW, Lee IJ. 2017.** Chloroplast genomes of *Arabidopsis halleri* ssp. *gemmifera* and *Arabidopsis lyrata* ssp. *petraea*: Structures and comparative analysis. *Scientific Reports* **7**(1):Article 7556 DOI [10.1038/s41598-017-07891-5](#).
- Balkwill K, Norris FG, Balkwill M-J. 1996.** Systematic studies in the Acanthaceae; *Dicliptera* in southern Africa. *Kew Bulletin* **51**:1–61 DOI [10.2307/4118744](#).
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD. 2012.** SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**:455–477 DOI [10.1089/cmb.2012.0021](#).
- Benson G. 1999.** Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**:573–580 DOI [10.1093/nar/27.2.573](#).
- Birky CW. 1995.** Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences of the United States of America* **92**:11331–11338 DOI [10.1073/pnas.92.25.11331](#).

- Borsch T, Quandt D. 2009.** Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. *Plant Systematics and Evolution* **282**:169–199 DOI [10.1007/s00606-009-0210-8](https://doi.org/10.1007/s00606-009-0210-8).
- Brako L, Zarucchi J. 1993.** Catálogo de las Angiospermas y Gimnospermas del Perú. *Monograph in Systematic Botany from Missouri Botanic Gardens* **45**:1–1286.
- Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, Batzoglou S. 2003.** Glocal alignment: finding rearrangements during alignment. *Bioinformatics* **19**S1:i54–i62.
- Bussmann RW, Glenn A. 2010.** A medicinal plants used in Northern Peru for reproductive problems and female health. *Journal of Ethnobiology and Ethnomedicine* **6**:Article 30 DOI [10.1186/1746-4269-6-30](https://doi.org/10.1186/1746-4269-6-30).
- Chen HM, Shao JJ, Zhang H, Jiang M, Huang LF, Zhang Z, Yang D, He M, Mostafa R, Luo X, Botao S, Wu WW, Liu C. 2018.** Sequencing and analysis of strobilanthes cusia (nees) kuntze chloroplast genome revealed the rare simultaneous contraction and expansion of the inverted repeat region in angiosperm. *Frontiers in Plant Science* **9**:Article 324 DOI [10.3389/fpls.2018.00324](https://doi.org/10.3389/fpls.2018.00324).
- Choi IS, Jansen R, Ruhlman T. 2019.** Lost and found: return of the inverted repeat in the legume clade defined by its absence. *Genome Biology Evolution* **11**:1321–1333 DOI [10.1093/gbe/evz076](https://doi.org/10.1093/gbe/evz076).
- Clegg MT, Gaut BS, Learn GH, Morton BR. 1994.** Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America* **91**:6795–6801 DOI [10.1073/pnas.91.15.6795](https://doi.org/10.1073/pnas.91.15.6795).
- Darbyshire I. 2008.** Notes on the genus Dicliptera (Acanthaceae) in eastern Africa. *Kew Bulletin* **63**:361–383 DOI [10.1007/s12225-008-9053-7](https://doi.org/10.1007/s12225-008-9053-7).
- Darling AE, Mau B, Perna NT. 2010.** progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLOS ONE* **5**:e11147 DOI [10.1371/journal.pone.0011147](https://doi.org/10.1371/journal.pone.0011147).
- Diethard T, Manfred R. 1984.** Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research* **12**:4127–4138 DOI [10.1093/nar/12.10.4127](https://doi.org/10.1093/nar/12.10.4127).
- Ding P, Shao YH, Li Q, Gao JL, Zhang RJ, Lai X. 2016.** The chloroplast genome sequence of the medicinal plant *Andrographis paniculata*. *Mitochondria DNA* **27**:2347–2348.
- Dong WP, Liu J, Yu J, Wang L, Zhou SL. 2012.** Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLOS ONE* **7**:e35071 DOI [10.1371/journal.pone.0035071](https://doi.org/10.1371/journal.pone.0035071).
- Dong WP, Xu C, Li CH, Sun JH, Zuo YJ, Shi S, Cheng T, Guo JJ, Zhou SL. 2015.** ycf1, the most promising plastid DNA barcode of land plants. *Scientific Report* **5**:8348 DOI [10.1038/srep08348](https://doi.org/10.1038/srep08348).
- Doyle JJ, Doyle JL. 1987.** A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**:11–15.
- Drew BT, Sytsma KJ. 2013.** The South American radiation of *Lepechinia* (Lamiaceae): phylogenetics, divergence times and evolution of dioecy. *Botanical Journal of the Linnean Society* **171**:171–190 DOI [10.1111/j.1095-8339.2012.01325.x](https://doi.org/10.1111/j.1095-8339.2012.01325.x).

- Fan WB, Wu Y, Yang J, Shahzad K, Li ZH. 2018. Comparative chloroplast genomics of Dipsacales species: Insights into sequence variation, adaptive evolution, and phylogenetic relationships. *Frontiers in Plant Science* 9:Article 689 DOI 10.3389/fpls.2018.00689.
- Feng XJ, Yuan XY, Sun YW, Hu YH, Zulfiqar S, Ouyang XH, Dang M, Zhou HJ, Keith W, Zhao P. 2018. Resources for studies of iron walnut (*Juglans sigillata*) gene expression, genetic diversity, and evolution. *Tree Genetics & Genomes* 14:Article 51 DOI 10.1007/s11295-018-1263-z.
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Research* 32:W273–W279 DOI 10.1093/nar/gkh458.
- Gao CM, Wang J, Deng YF. 2018. The complete chloroplast genomes of *Echinacanthus* species (acanthaceae): phylogenetic relationships, adaptive evolution, and screening of molecular markers. *Frontiers in Plant Science* 9:Article 1989.
- Gao QB, Li Y, Gengji ZM, Gornall RJ, Wang JL, Hai-Rui Liu HR, Jia LK, Chen SL. 2017. Population genetic differentiation and taxonomy of three closely related species of *Saxifraga* (saxifragaceae) from Southern Tibet and the Hengduan Mountains. *Frontiers in Plant Science* 8:1325 DOI 10.3389/fpls.2017.01325.
- Gargaud M, Irvine W, Amils R, Cleaves HJ, Pinti Daniele L, Quintanilla Jose C, Rouan D, Spohn T, Tirard S, Viso M (eds.) 2015. *Encyclopedia of astrobiology*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1803–1803.
- George B, Bhatt BS, Awasthi M, George B, Singh AK. 2015. Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Current Genetics* 61:665–677 DOI 10.1007/s00294-015-0495-9.
- Gernandt DS, Hernández-León S, Salgado-Hernández E, Pérez de la Rosa JA. 2009. Phylogenetic relationships of pinus subsection ponderosae inferred from rapidly evolving cpDNA regions. *Systematic Botany* 34:481–491 DOI 10.1600/036364409789271290.
- Goulding SE, Wolfe K, Olmstead R, Morden C. 1996. Ebb and flow of the chloroplast inverted repeat. *Molecular and General Genetics* 252:195–206 DOI 10.1007/BF02173220.
- Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47:W59–W64.
- Hall B, Strickberger BW. 2008. *Strickberger's evolution*. Burlington: Jones & Bartlett Learning.
- Horacio DLC, Graciela V, Percy AZ. 2007. Ethnobotanical study of medicinal plants used by the Andean people of Canta, Lima, Peru. *Journal of Ethnopharmacology* 111:284–294 DOI 10.1016/j.jep.2006.11.018.
- Hu JQ, Deng YF, John RIW, Thomas FD. 2011. *Dicliptera* in Flora of China, vol. 19. St. Louis: Science Press, Beijing & Missouri Botanical Garden Press.
- Huang SN, Deng YF, Ge XJ. 2019. The complete chloroplast genome of *Aphelandra knappiae* (Acanthaceae). *Mitochondrial DNA Part B-Resources* 4:273–274 DOI 10.1080/23802359.2018.1541718.

- Inkyu P, Sungyu Y, Wook JM, Pureum N, Hyun OL, Byeong CM. 2018.** The complete chloroplast genomes of Six *Ipomoea* species and indel marker development for the discrimination of authentic pharbitidis semen (seeds of *I. nil* or *I. purpurea*). *Frontiers in Plant Science* **9**:Article 965 DOI [10.3389/fpls.2018.00965](https://doi.org/10.3389/fpls.2018.00965).
- Jerzy J, Charit P. 1995.** Simple repetitive DNA sequences from primates: compilation and analysis. *Journal of Molecular Evolution* **40**:120–126 DOI [10.1007/BF00167107](https://doi.org/10.1007/BF00167107).
- Jiang M, Chen HM, He SB, Wang LQ, Chen A, Liu C. 2018.** Sequencing, characterization, and comparative analyses of the plastome of *Caragana rosea* var. *rosea*. *International Journal of Molecular Sciences* **19**:Article 1419 DOI [10.3390/ijms19051419](https://doi.org/10.3390/ijms19051419).
- Jin JJ, Yu WB, Yang JB, Song Y, Yi TS, Li DZ. 2018.** GetOrganelle: a simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *bioRxiv* 256479.
- Katoh K, Standley DM. 2013.** MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**:772–780 DOI [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).
- Kearse M, Moir R, Wilson A, Stoneshavas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012.** Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**:1647–1649 DOI [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199).
- Khakhlova O, Bock R. 2006.** Elimination of deleterious mutations in plastid genomes by gene conversion. *The Plant Journal* **46**:85–94 DOI [10.1111/j.1365-3113.2006.02673.x](https://doi.org/10.1111/j.1365-3113.2006.02673.x).
- Kiel CA, Daniel TF, Darbyshire I, McDade LA. 2017.** Unraveling relationships in the morphologically diverse and taxonomically challenging. *Taxon* **66**:645–674 DOI [10.12705/663.8](https://doi.org/10.12705/663.8).
- Kim KJ, Lee HL. 2004.** Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Research* **11**:247–261 DOI [10.1093/dnares/11.4.247](https://doi.org/10.1093/dnares/11.4.247).
- Kurt M, Mark Whiten NW, Carlswald BS, Blanco MA, Endara L, Williams NH, Moore M. 2008.** Phylogenetic utility of *ycf1* in orchids: a plastid gene more variable than *matK*. *Faculty Research & Creative Activity* **277**:75–84.
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001.** REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* **29**:4633–4642 DOI [10.1093/nar/29.22.4633](https://doi.org/10.1093/nar/29.22.4633).
- Langmead B, Salzberg SL. 2012.** Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**:357–359 DOI [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923).
- León B. 2006.** Acanthaceae endémicas del Perú (en línea). *Revista peruana de Biología* **13**(2):23–29.
- Li D, Zhao CH, Liu X. 2018.** Complete chloroplast genome sequences of *Kaempferia galanga* and *Kaempferia elegans*: molecular structures and comparative analysis. *Molecules* **24**:Article 474.
- Li L, Stoeckert CJ, Roos DS. 2003.** OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**:2178–2189 DOI [10.1101/gr.1224503](https://doi.org/10.1101/gr.1224503).

- Librado P, Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**:1451–1452
DOI [10.1093/bioinformatics/btp187](https://doi.org/10.1093/bioinformatics/btp187).
- Lu QX, Ye WQ, Lu RS, Xu WQ, Qiu YX. 2018.** Phylogenomic and comparative analyses of complete plastomes of *Croomia* and *Stemona* (stemonaceae). *International Journal of Molecular Sciences* **19**:2383 DOI [10.3390/ijms19082383](https://doi.org/10.3390/ijms19082383).
- Ma QY, Li SX, Bi CW, Hao ZD, Sun CR, Ye N. 2017.** Complete chloroplast genome sequence of a major economic species, *Ziziphus jujuba* (Rhamnaceae). *Current Genetics* **63**:117–129 DOI [10.1007/s00294-016-0612-4](https://doi.org/10.1007/s00294-016-0612-4).
- Mabberley D. 2017.** *Mabberley's plant-book. A portable dictionary of plants, their classifications and uses.* forth edition. Cambridge: Cambridge University Press.
- Mader M, Pakull B, Blanc-Jolivet C, Paulini-Drewes M, Bouda Z, Degen B, Small I, Kersten B. 2018.** Complete chloroplast genome sequences of four Meliaceae species and comparative analyses. *International Journal of Molecular Sciences* **19**:Article 701 DOI [10.3390/ijms19030701](https://doi.org/10.3390/ijms19030701).
- Makalowski W, Boguski MS. 1998.** Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. *Proceedings of the National Academy of Sciences of the United States of America* **95**:9407–9412 DOI [10.1073/pnas.95.16.9407](https://doi.org/10.1073/pnas.95.16.9407).
- Mcdade L, Daniel T, Kiel C. 2008.** Toward a comprehensive understanding of phylogenetic relationships among lineages of Acanthaceae S.L. (Lamiales). *American Journal of Botany* **95**:1136–1152 DOI [10.3732/ajb.0800096](https://doi.org/10.3732/ajb.0800096).
- Meng XX, Xian YF, Xiang L, Zhang D, Shi YH, Wu ML, Dong GQ, Ip SP, Lin ZX, Wu L. 2018.** Complete chloroplast genomes from *Sanguisorba*: identity and variation among four species. *Molecules* **23**:Article 2137 DOI [10.3390/molecules23092137](https://doi.org/10.3390/molecules23092137).
- Nylander JAA. 2004.** MrModeltest v2 Program distributed by the author. Evolutionary Biology Centre, Uppsala University. Available at <https://github.com/nylander/MrModeltest2>.
- Olmstead RG, Palmer JD. 1994.** Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany* **81**:1205–1224
DOI [10.1002/j.1537-2197.1994.tb15615.x](https://doi.org/10.1002/j.1537-2197.1994.tb15615.x).
- Qu XJ, Moore MJ, Li DZ, Yi TS. 2019.** PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* **15**:Article 12 DOI [10.1186/s13007-018-0385-5](https://doi.org/10.1186/s13007-018-0385-5).
- Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, Jansen RK. 2007.** Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* **8**:174 DOI [10.1186/1471-2164-8-174](https://doi.org/10.1186/1471-2164-8-174).
- Ronquist F, Huelsenbeck JP. 2003.** MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574
DOI [10.1093/bioinformatics/btg180](https://doi.org/10.1093/bioinformatics/btg180).
- Sara EG, Jaina M, Alex B, Sean RE, Aurelien L, Simon CP, Matloob Q, Lorna JR, Gustavo AS, Alfredo S, Sonnhammer ELL, Layla H, Lisanna P, Damiano P, Tosatto**

- SCE, Finn RD. 2018. The Pfam protein families database in 2019. *Nucleic Acid Research* **D1**:D427–D432.
- Särkinen T, George M. 2013. Predicting plastid marker variation: can complete plastid genomes from closely related species help? *PLOS ONE* **8**:e82266 DOI [10.1371/journal.pone.0082266](https://doi.org/10.1371/journal.pone.0082266).
- Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research* **33**:W686–W689 DOI [10.1093/nar/gki366](https://doi.org/10.1093/nar/gki366).
- Schwarz EN, Ruhlman TA, Sabir JSM, Hajrah NH, Alharbi NS, Almalki AL, Bailey CD, Jansen RK. 2015. Plastid genome sequences of legumes reveal parallel inversions and multiple losses of *rps16* in papilionoids. *Journal of Systematics and Evolution* **53**:458–468 DOI [10.1111/jse.12179](https://doi.org/10.1111/jse.12179).
- Scotland R, Vollesen K. 2000. Classification of Acanthaceae. *Kew Bulletin* **55**:513–589.
- Scottphillips TC, Laland KN, Shuker DM, Dickins TE, West ST. 2013. The niche construction perspective: a critical appraisal. *Evolution* **68**:1231–1243.
- Sharp PM, Li WH. 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research* **15**:1281–1295 DOI [10.1093/nar/15.3.1281](https://doi.org/10.1093/nar/15.3.1281).
- Shingo K, Jocelyn B, Minako H, Yoshino H, Maya O, Midori I, Mai T, Toru I, Masato N. 2013. Uncover the protein translocon at the chloroplast inner envelope membrane. *Science* **339**:571–574 DOI [10.1126/science.1229262](https://doi.org/10.1126/science.1229262).
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313 DOI [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033).
- Steane DA. 2005. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). *DNA Research* **12**:215–220 DOI [10.1093/dnares/dsi006](https://doi.org/10.1093/dnares/dsi006).
- Swofford DL. 2003. PAUP. Phylogenetic analysis using parsimony (* and other methods) version 4. Sunderland: Sinauer Associates Press.
- Telefo P, Moundipa P, Tchouanguép F. 2002. Oestrogenicity and effect on hepatic metabolism of the aqueous extract of the leaf mixture of *Aloe buettneri*, *Dicliptera verticillata*, *Hibiscus macranthus* and *Justicia insularis*. *Fitoterapia* **73**:472–478 DOI [10.1016/S0367-326X\(02\)00177-6](https://doi.org/10.1016/S0367-326X(02)00177-6).
- Thiel T, Michalek W, Varshney R, Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* **106**:411–422 DOI [10.1007/s00122-002-1031-0](https://doi.org/10.1007/s00122-002-1031-0).
- Victor A, Jhan C-C, Raúl S. 2017. Inventory of plant species of La Libertad (Peru) and analysis of its agro-industrial potential. *Agroindustrial Science* **7**:87–104 DOI [10.17268/agroind.sci.2017.02.05](https://doi.org/10.17268/agroind.sci.2017.02.05).

- Wang L, Wuyun TN, Du HY, Wang DP, Cao DM. 2016. Complete chloroplast genome sequences of *Eucommia ulmoides*: genome structure and evolution. *Tree Genetics & Genomes* 12:Article 12 DOI 10.1007/s11295-016-0970-6.
- Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evolutionary Biology* 8:36 DOI 10.1186/1471-2148-8-36.
- Weng ML, Blazier JC, Govindu M, Jansen RK. 2013. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Molecular Biology and Evolution* 31:645–659.
- Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31:3350–3352 DOI 10.1093/bioinformatics/btv383.
- Wicke S, Schneeweiss GM, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* 76:273–297 DOI 10.1007/s11103-011-9762-4.
- Wojciech M, Mark SB. 1998. Evolutionary parameters of the transcribed mammalian genome: An analysis of 2,820 orthologous rodent and human sequences. *Proceedings of the National Academy of Sciences of the United States of America* 95:9407–9412 DOI 10.1073/pnas.95.16.9407.
- Wu Y, Liu F, Yang DG, Li W, Zhou XJ, Pei XY, Liu YG, He KL, Zhang WS, Ren ZY. 2018. Comparative chloroplast genomics of *Gossypium* species: insights into repeat sequence variations and phylogeny. *Frontiers in Plant Science* 9:Article 376 DOI 10.3389/fpls.2018.00376.
- Xie DF, Yu HX, Price M, Xie C, Deng YQ, Chen PJ, Yu Y, Zhou SD, He XJ. 2019. Phylogeny of Chinese *Allium* species in section Daghestanica and adaptive evolution of *Allium* (Amaryllidaceae, Alliioideae) species revealed by the chloroplast complete genome. *Frontiers in Plant Science* 10:Article 460.
- Yan C, Du JC, Gao L, Li Y, Hou XL. 2019. The complete chloroplast genome sequence of watercress (*Nasturtium officinale* R. Br.): Genome organization, adaptive evolution and phylogenetic relationships in Cardamineae. *Gene* 699:24–36 DOI 10.1016/j.gene.2019.02.075.
- Yang JB, Li DZ, Li HT. 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Molecular Ecology Resources* 14:1024–1031.
- Yang Y, Dang YY, Li Q, Lu JJ, Li XW, Wang YT. 2014. Complete chloroplast genome sequence of poisonous and medicinal plant *Datura stramonium*: organizations and implications for genetic engineering. *PLOS ONE* 9:e110656 DOI 10.1371/journal.pone.0110656.
- Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24:1586–1591 DOI 10.1093/molbev/msm088.

- Yang ZH, Nielsen R. 2000.** Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* 17:32–43 DOI [10.1093/oxfordjournals.molbev.a026236](https://doi.org/10.1093/oxfordjournals.molbev.a026236).
- Yang ZH, Swanson WJ. 2002.** Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Molecular Biology and Evolution* 19:49–57 DOI [10.1093/oxfordjournals.molbev.a003981](https://doi.org/10.1093/oxfordjournals.molbev.a003981).
- Yi DK, Kim KJ. 2012.** Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PLOS ONE* 7:e35872 DOI [10.1371/journal.pone.0035872](https://doi.org/10.1371/journal.pone.0035872).
- Zhang KF, Zhu H, Gao Y. 2010.** Research on active extracts of *Dicliptera chinensis* on liver protection. *China Journal of Chinese Materia Medica* 35:497–498.
- Zhao ZY, Wang X, Yu Y, Yuan SB, Jiang D, Zhang YJ, Zhang T, Zhong WH, Yuan QJ, Huang LQ. 2018.** Complete chloroplast genome sequences of *Dioscorea*: characterization, genomic resources, and phylogenetic analyses. *PeerJ* 6:e6032 DOI [10.7717/peerj.6032](https://doi.org/10.7717/peerj.6032).