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

Sunan Huang $^{1,2}$, Xue-jun Ge $^1$, Asuncion Cano $^3$, Betty M Salazar $^3$, Yunfei Deng$^{Corresp.1}$

$^1$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

Corresponding Author: Yunfei Deng
Email address: yfdeng@scbg.ac.cn

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 150689-150811 bp and exhibit quadripartite organizations made up of a large single copy region (LSC, 82796-82919 bp), a small single copy region (SSC, 17084-17092 bp), and a pair of inverted repeat regions (IRs, 25401-25408 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 4 ribosomal RNA (rRNA) genes. Comparative analyses of nucleotide variability (Pi) reveal the 5 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.
Comparative analysis of chloroplast genomes for five *Dicliptera* species (Acanthaceae): molecular structure, phylogenetic relationships, and adaptive evolution

Sunan Huang 1,2, Xue-jun Ge 1, Asuncion Cano 3, Betty Millan Salazar 3, and Yunfei Deng 1,*

1 Key Laboratory of Plant Resources Conservation & Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, 510650, China
2 University of Chinese Academy of Sciences, Beijing, 100049, China
3 Facultad de Ciencias Biológicas y Museo de Historia Natural, Universidad Nacional Mayor de San Marcos, Avenida Germán Amezaga 375, Lima 1, Perú

Corresponding Author:

Yunfei Deng 1

No.723, Xingke Road, Tianhe District, Guangzhou, 510650, China

Email address: yfdeng@scbg.ac.cn
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 150689-150811 bp and exhibit quadripartite organizations made up of a large single copy region (LSC, 82796-82919 bp), a small single copy region (SSC, 17084-17092 bp), and a pair of inverted repeat regions (IRs, 25401-25408 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 4 ribosomal RNA (rRNA) genes. Comparative analyses of nucleotide variability (Pi) reveal the 5 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.

Introduction

The genus *Dicliptera* Juss. belongs to tribe Justicieae 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). 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 (Buschmann & Glenn, 2010; Horacio et al., 2007; Telefo et al., 2002; Zhang et al., 2010). Species delimitation within *Dicliptera* is difficult (Balkwill et al., 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 Justicieae,
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 et al., 2017); D. peruviana is a
traditional herbal medicine used by the Andeans of Canta, Lima, Peru to alleviate stomach aches
(Horacio et al., 2007); D. mucronata is distributed mainly in Central America and is easily
confused with D. scabra (Victor et al., 2017); D. ruiziana is found throughout Peru to elevations
of about 3000 m (Antonio 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 et al., 2018; George et al., 2015; Chen et al., 2018;
Inkyu et al., 2018; Kim et al., 2004; Lin 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 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 et al., 2018; George et al., 2015; Chen et al., 2018; Inkyu et al., 2018; Kim
et al., 2004; Lin et al., 2016; Mader et al., 2018; Meng et al., 2018; Erika et al., 2015). The
complete cp genomes of more than 2400 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; Weng et al., 2013; Wicke et al., 2011). The family
Acanthaceae is a large family with approximately 230 genera and 4300 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 (Lin et al., 2016; Ma et al., 2017; Fan et al., 2018; Gao et al.,
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; Scott Phillips et al., 2014; Hall et al., 2008). The non-synonymous 
(K_\text{A})/synonymous rate (K_\text{S}) ratio (\omega = K_\text{A}/K_\text{S}) provides a measure of selective pressure at the 
amino acid level. As suggested by Makalowski and Boguski (1998), the \omega values less than one 
(K_\text{A}/K_\text{S} < 1) indicate that the gene is under negative selection and vice versa (Wojciech et al., 
1998; Meng et al., 2018). Recent studies have detected many positively selected chloroplast 
genes (K_\text{A}/K_\text{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 et al., 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.

Materials 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 et al., 2014a).

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 x 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 Brujin
125 graph files were transferred to Bandage v. 0.8.1 (Wick et al., 2015) for visualization and to
126 obtain the paths of the final “fasta” files of the cp genomes; finally, the genome structures of all
127 five species were compared to the reference genome using Mauve v. 1.1.1 software (Darling et
128 al., 2010) to determine the accuracy of the final genome. The assembled cp genome was
129 annotated using PGA v. 2019 (Qu et al., 2019) using the annotated *A. paniculata* as a reference.
130 The boundaries of the annotated genes were manually modified and coupled with CDSs in
131 Geneious v. 2019.0.3 (Kearse et al., 2012). All transfer RNA (tRNA) genes were determined
132 using tRNAscan-SE v. 2.0 (Schattner et al., 2005). The annotated cp genome files were
133 submitted to OGDRAW v. 1.3.1 ([https://chlorobox.mpimp-golm.mpg.de/OGDraw.html](https://chlorobox.mpimp-golm.mpg.de/OGDraw.html))
134 (Greiner et al., 2019) to create a circular cp genome map for each species. The five cp genomes
135 (*D. acuminata*, MK830556; *D. pruviana*, MK833945; *D. montana*, MK833946; *D. ruiziana*,
136 MK833947; *D. mucronata*, MK848596) were submitted to Genbank.

137 **Genome comparison and structural analysis**

138 Five cp genomes were compared using mVISTA v. 2.0 (Frazer et al., 2004) with Shuffle-
139 LAGAN mode and the annotation of *D. acuminata* as a reference (Brudno et al., 2003). The
140 conserved regions were visualized on an mVISTA plot. DnaSP v. 5.1 (Librado & Rozas, 2009)
141 was used to calculate the nucleotide variance (Pi) within the five *Dicliptera* species. The SC and
142 IR boundaries were compared with *A. paniculata* as a reference. The Relative Synonymous
143 Codon Usage (RSCU) of all protein-coding genes was analyzed for each species using CondoW
144 v. 1.4.2 (Sharp & Li, 1987). MISA v. 1.0 (Thiel et al., 2003) was used to identify the simple
145 sequence repeats (SSRs). The locations and lengths of long repeats (including forward,
146 palindrome, complement, and reverse repeats) were analyzed using REPuter v. 2.74 (Kurtz et al.,
147 2001) with the minimum repeat size set to 20 bp. Tandem repeats were identified using Tandem
148 Repeats Finder v. 4.09 (Benson, 1999).

149 **Phylogenetic analyses**

150 Phylogenetic analysis was conducted for all of the sequenced cp genomes (each cp genome
151 included only one IR), including the five species reported in this study and ten previously
152 reported species of Acanthaceae (Table S1). *Sesamum indicum* (NC_016433) (Pedaliaceae) and
153 *Mentha spicata* (NC_037247) (Lamiaceae) were used as outgroups. The complete cp genomes
154 were aligned using MAFFT v. 1.3.7 (Katoh & Standley, 2013) and were adjusted manually as
155 needed. The substitution models with the best fit were chosen by MrModeltest v. 2.3 (Nylander,
156 2004) based on the Akaike Information Criterion (AIC). RAxML v. 8.0.0 (Stamatakis, 2014) was
157 used to reconstruct the phylogenetic relationship with the maximum likelihood (ML) method.
158 Maximum parsimony (MP) analysis was run in Paup v. 4.0a (Swofford, 2003). Bootstrap values
159 exceeding 50% were shown next to the corresponding branches. Bayesian inference (BI) analysis
160 was conducted using MrBayes 3.2.7 (Ronquist & Huelsenbeck, 2003) with posterior
161 probabilities (PP) obtained for each branch.

162 **Selective pressure analysis**

163 OrthoMCL v. 2.0 (Li et al., 2003) was used to find the orthologous genes for the family
164 Acanthaceae. The sequences for each orthologous gene were aligned separately using MAFFT v.
165 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 (ω=K_A/K_S) was estimated using the method reported by Yang and 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 150738 bp (D. acuminata), 150811 bp (D. peruviana), 150689 bp (D. montana), 150750 bp (D. ruiziana), and 150720 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 S2; Figure 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 82796-82919 bp), a small single copy (SSC) region (range from 17084-17092 bp), and a pair of inverted repeat (IR) regions (25401-25408 bp; Table 2). The Guanine-Cytosine (GC) content of the cp genomes of the five Dicliptera 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 (Figure 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 junction 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 4543 bp and 4560 bp sequences situated at SSC in Dicliptera and...
A. paniculata, respectively. The ycf1 gene duplications located in IRa are 811-812 bp in

Dictiptera and 982 bp in A. paniculata, indicating a slight expansion of the IR regions. Likewise,

the space in Dictiptera 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

Dictiptera 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 Dictiptera species (Figure 3). The size and gene order of the
chloroplast genomes of the Dictiptera 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,
Figure 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).

Codon usage

A total of 78714-78747 bp protein-coding genes were identified in the five Dictiptera 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 2825-2828
(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%) (Figure 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., 2014b; 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
tetrinucleotides 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 (Figure 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 (Figure 6B, Table 3). Five
categories of long repeats (tandem, complement, forward, palindromic and reverse repeats) were
detected and analyzed in the five Dictiptera cp genomes (Table S6 and S7, Figure 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 (Figure 6C). The number of
repeats was highest in *D. peruviana* (56) and lowest in *D. montana* (41) (Figure 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 (Figure 7). Our phylogenetic analyses strongly support
the monophyly of the *Dicliptera* species \[BP_{(MP)}=100\%, \ BP_{(ML)}=100\%, \ PP=1.0\], 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* \[BP_{(MP)}=100\%, \ BP_{(ML)}=100\%, \ PP=1.0\].

**Selective pressure events**

There were 68 orthologous protein-coding genes found in this study. The $\omega$ 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 $\omega$ value of the *matK* gene was 0.9418, indicating a relaxed selection
(Table S8, Figure 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, Figure 1). The cp genomes ranged in size from
150689 bp to 150811 bp in *D. montana* and *D. peruviana*, respectively. These structures are
longer than the cp genome of *A. paniculata* (15249 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 et al., 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 et al., 2018; Meng et al., 2018; Li et al., 2018; Yan et al., 2019).
Khakhlova et al. (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 et al., 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 slight expansion. The size differences among the cp genome of the five Dicliptera species (150689-150811 bp) and Andrographis paniculata (150249 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; Olga et al., 2006; Meng et al., 2018; Choi et al., 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 et al., 1984; Jerzy et al., 1995; George et al., 2015; Gao et al., 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; David et al., 2009; Dong et al., 2012; Drew et al., 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 (5354-5355 bp) and A. paniculata (5542 bp). The nucleotide variability of the ycf1 gene (Pi=0.0109) was slightly higher than that of the regions matK (Pi=0.00107) and rpl16 (Pi=0.00103). The two regions are currently used in the DNA barcodes for the tribe Justicieae 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 (Pi=0.02230), petN-psbM (Pi=0.00783), psbZ-trnG-GCC (Pi=0.00697), trnG-GCC (Pi=0.00571), and trnY-GUA-trnE-UUC (Pi=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 (Figure 7). The genus *Aphelandra* was found to be the earliest diverging lineage; tribes Justicieae 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 (Mcadade et al., 2008; Huang et al., 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 1.0-1.5 × 0.8-1.0 cm versus the has subulate calyx and oblong-lanceolate leaves of 3.5-7.0 × 1.5-2.5 cm of *D. acuminata* (Table 1).

Adaptive evolution analysis

Positively selected genes are known to play a key role in adapting to different environments (Lin et al., 2016; Ma et al., 2017; Fan et al., 2018; Gao et al., 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., 2019) 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 Diciptera 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 Diciptera. 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 Diciptera (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 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.

References

Antonio Galán de Mera, Eliana Linares Perea, José Campos de la Cruz and José Alfredo Vicente Orellana. 2009. Nuevas observaciones sobre la vegetación del sur del Perú. Del Desierto Pacífico al Altiplano. Acta Botanica Malacitana 34:107-144.

Balkwill K, Norris FG, and Balkwill M-J. 1996. Systematic studies in the Acanthaceae; Diciptera in southern Africa. Kew Bulletin 51:1-61.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, and Prjibelski AD. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology 19:455-477.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic acids research 27:573-580.

Birky CW. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proceedings of the National Academy of Sciences 92:11331-11338.

Borsch T. and Quandt D. 2009. Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. Plant Systematics and Evolution 282:169-199.

Brako L. and 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 C.B., Couronne O., Dubchak I., and Batzoglou S. 2003. Glocal Alignment: Finding Rearrangements During Alignment. Bioinformatics 19S1: i54-i62.

Bryan T. Drew and Kenneth J. Sytsma. 2013. The South American radiation of Lepechinia (Lamiaceae): phyllogenetics, divergence times and evolution of dioecy. Botanical Journal of the Linnean Society 171:171-190.
Bussmann RW, and Glenn A. 2010. Medicinal plants used in Northern Peru for reproductive problems and female health. *Journal of Ethnobiology and Ethnomedicine* 6:30.

Yan C, Du J, Gao L, Li Y and Hou X. 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.

Clegg MT, Gaut BS, Learn GH, and Morton BR. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences* 91:6795-6801.

Darbyshire I. 2008. Notes on the genus Dicliptera (Acanthaceae) in eastern Africa. *Kew Bulletin* 63:361-383.

Darling AE, Mau B and Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one* 5: e11147.

David S. Gernandt, Sergio Hernández-León, Esmeralda Salgado-Hernández, and Jorge A. Pérez de la Rosa. 2009. Phylogenetic Relationships of Pinus Subsection Ponderosae Inferred from Rapidly Evolving cpDNA Regions. *Systematic botany* 34:481-491.

Xie DF, YuHX, Megan Price, Xie C, Deng YQ, Chen PJ, Yu Y, Zhou SD and He XJ. 2019. Phylogeny of Chinese *Allium* species in section Daghestanica and adaptive evolution of *Allium* (Amaryllidaceae, Allioideae) species revealed by the chloroplast complete genome. *Frontiers in Plant Science* 10:460.

Diethard Tautz and Manfred Renz. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids research* 12: 4127-4138.

Ding P, Shao Y H, Li Q, Gao J L, Zhang R J and Lai X P. 2016. The chloroplast genome sequence of the medicinal plant *Andrographis paniculata*. *Mitochondria DNA* 27:2347-2348.

Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J and Zhou S. 2015. ycf1, the most promising plastid DNA barcode of land plants. *Scientific Report* 5:8348.

Dong W, Liu J, Yu J, Wang L and Zhou S. 2012. Highly Variable Chloroplast Markers for Evaluating Plant Phylogeny at Low Taxonomic Levels and for DNA Barcoding. *Plos one* 7:e35071.

Dyole JJ, and Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11-15.

Fan W, Wu Y, Yang J, Shahzad K, and Li Z. 2018. Comparative chloroplast genomics of Dipsacales species: Insights into sequence variation, adaptive evolution, and phylogenetic relationships. *Frontiers in plant science* 9:689.

Frazer KA, Pachter L, Poliakov A, Rubin EM, and Dubchak I. 2004. VISTA: computational tools for comparative genomics. *Nucleic acids research* 32:W273-W279.
Gao C, Wang J, and Deng Y. 2018. The Complete Chloroplast Genomes of Echinacanthus Species (Acanthaceae): Phylogenetic Relationships, Adaptive Evolution, and Screening of Molecular Markers. *Frontiers in plant science* 9:1989.

Gargaud Muriel, Irvine William, Amils Ricardo, Cleaves Henderson James, Pinti Daniele L., Quintanilla Jose Cernicharo, Rouan Daniel, Spohn Tilman, Tirard Stephane and Viso Michel, eds. 2015. *Encyclopedia of Astrobiology*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1803-1803.

George B, Bhatt BS, Awasthi M, George B, and Singh AK. 2015. Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Current Genetics* 61:665-677.

Goulding SE, Wolfe K, Olmstead R, and Morden C. 1996. Ebb and flow of the chloroplast inverted repeat. *Molecular and General Genetics MGG* 252:195-206.

Greiner S, Lehwark P and Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*. 1: gkz238.

Chen H, Junjie Shao J, Zhang H, Jiang M, Huang L, Zhang Z, Yang D, He M, Mostafa R, Luo X, Botao S, Wu W and 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* 9:324.

Horacio De-la-Cruz, Graciela Vilcapoma and Percy A. Zevallos. 2007. Ethnobotanical study of medicinal plants used by the Andean people of Canta, Lima, Peru. *Journal of Ethnopharmacology* 111:284-294.

Hu J, Deng Y, John RIW, and Thomas FD. 2011. *Diciplerta in Flora of China* 19: Science Press, Beijing & Missouri Botanical Garden Press, St. Louis.

Huang S, Deng Y, and Ge X. 2019. The complete chloroplast genome of Aphelandra knappiae (Acanthaceae). *Mitochondrial DNA Part B-Resources* 4:273-274.

Inkyu P, Sungyu Y, Wook JM, Pureum N, Hyun OL and 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* 9:965.

IS Choi, R Jansen and T Ruhlman. 2019. Lost and Found: Return of the Inverted Repeat in the Legume Clade Defined by Its Absence. *Genome Biology Evolution* 11: 1321-1333.

Jerzy Jurka and Charit Pethiyagoda. 1995. Simple Repetitive DNA sequences from Primates: Compilation and Analysis. *Journal of Molecular Evolution* 40: 120-126.
Jin JJ, Yu WB, Yang JB, Song Y, Yi TS, and 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, and Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution 30:772-780.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, and Duran C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649.

Khakhlova O, and Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. The Plant Journal 46:85-94.

Kiel CA, Daniel TF, Darbyshire I, and McDade LA. 2017. Unraveling relationships in the morphologically diverse and taxonomically challenging. Taxon 66:645-674. Kim K-J, and Lee H-L. 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.

Kurt M. Neubig, W. Mark Whiten, Barbara S. Carlsward, Mario A. Blanco, Lorena Endara, Norris H. Williams, and Michael Moore. 2008. Phylogenetic utility of ycf1 in orchids: a plastid gene more variable than matK. Faculty Research & Creative Activity 257.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, and Giegerich R. 2001. REPutter: the manifold applications of repeat analysis on a genomic scale. Nucleic acids research 29:4633-4642.

Langmead B, and Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9:357.

León B. 2006. Acanthaceae endémicas del Perú (en línea). Revista peruana de Biología 13(2):23–29.

Li D, Zhao Ch and Liu X. 2018. Complete chloroplast genome sequences of Kaempferia galanga and Kaempferia elegans: molecular structures and comparative analysis. Molecules 24:474.

Li L, Stoeckert CJ, and Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome research 13:2178-2189.

Librado P, and Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.
Wang L, Wuyun T, Du H, Wang D and Cao D. 2016. Complete chloroplast genome sequences of *Eucommia ulmoides*: genome structure and evolution. *Tree Genetics & Genomes* 12:12.

Lucinda Mcdade, Thomas Daniel and Carrie Kiel. 2008. Toward a comprehensive understanding of phylogenetic relationships among lineages of Acanthaceae S.L. (Lamiales). *American Journal of Botany* 95:1136-1152.

Mabberley D. 2017. *Mabberley’s plant-book*. A portable dictionary of plants, their classifications and uses, forth edition. Cambridge University Press.

Mader M, Pakull B, Blanc-Jolivet C, Paulini-Drewes M, Bouda Z, Degen B, Small I, and Kersten B. 2018. Complete chloroplast genome sequences of four Meliaceae species and comparative analyses. *International journal of molecular sciences* 19:701.

Makalowski W and Boguski M. S. 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.

Marcondes Oliveira, Alexandre Grillo and Marcelo Tabarelli. 2004. Caracterização da flora dos remanescentes da Usina Serra Grande, Alagoas. Relatório Técnico. *Centro de Pesquisas Ambientais do Nordeste, Recife, PE*.

Meng X-X, Xian Y-F, Xiang L, Zhang D, Shi Y-H, Wu M-L, Dong G-Q, Ip S-P, Lin Z-X, and Wu L. 2018. Complete Chloroplast Genomes from *Sanguisorba*: Identity and Variation Among Four Species. *Molecules* 23:2137.

Jiang M, Chen H, He S, Wang L, Chen A and Liu C. 2018. Sequencing, Characterization, and Comparative Analyses of the Plastome of *Caragana rosea* var. *rosea*. *International Journal of Molecular Sciences* 19:1419.

Nylander J. A. A. 2004. MrModeltest v2. Program distributed by the author. *Evolutionary Biology Centre, Uppsala University*, 2.

Olmstead RG, and Palmer JD. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *American journal of botany* 81:1205-1224.

Gao QB, Li Y, Gengji ZM, RJ Gornall, Wang JL, Hai-Rui Liu HR, Jia LK and 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.

Ma Q, Li S, Changwei Bi C, Hao Z, Sun C and Ye N. 2017. Complete chloroplast genome sequence of a major economic species, *Ziziphus jujuba* (Rhamnaceae). *Current Genetics* 63:117-129.
525 Lu Q, Ye W, Lu R, Xu W and Qiu Y. 2018. Phylogenomic and Comparative Analyses of
526 Complete Plastomes of *Croomia* and *Stemona* (Stemonaceae). *International journal of
527 molecular sciences* 19: 2383.
528 Qu XJ, Moore MJ, Li DZ, and Yi TS. 2019. PGA: a software package for rapid, accurate, and
529 flexible batch annotation of plastomes. *Plant Methods* 15:1-12.
530 Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, and Jansen RK.
531 2007. Comparative chloroplast genomics: analyses including new sequences from the
532 angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC genomics* 8:174.
533 Ronquist F, and Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under
534 mixed models. *Bioinformatics* 19:1572-1574.
535 Sara El-Gebali, Jaina Mistry, Alex Bateman, Sean R. Eddy, Aurelien Luciani, Simon C. Potter,
536 Matloob Qureshi, Lorna J. Richardson, Gustavo A. Salazar, Alfredo Smart, Erik L.L.
537 Sonnhammer, Layla Hirsh, Lisanna Paladin, Damiano Piovesan, Silvio C.E. Tosatto and
538 Robert D. Finn. 2018. The Pfam protein families database in 2019. *Nucleic Acid
539 Research* gky995.
540 Särkinen T, and George M. 2013. Predicting plastid marker variation: can complete plastid
541 genomes from closely related species help? *PLoS One* 8:e82266.
542 Schattner P, Brooks AN, and Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web
543 servers for the detection of tRNAs and snoRNAs. *Nucleic acids research* 33:W686-
544 W689.
545 Scotland R, and Vollesen K. 2000. Classification of Acanthaceae. *Kew Bulletin* :513-589.
546 Sharp PM, and Li W-H. 1987. The codon adaptation index—a measure of directional synonymous
547 codon usage bias, and its potential applications. *Nucleic acids research* 15:1281-1295.
548 Shingo Kikuchi, Jocelyn Bédard, Minako Hirano, Yoshino Hirabayashi, Maya Oishi, Midori
549 Imai, Mai Takase, Toru Ide and Masato Nakai. Uncoverign the protein translocon at the
550 chloroplast inner envelope membrane. 2013. *Science* 339:571.
551 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
552 large phylogenies. *Bioinformatics* 30:1312-1313.
553 Steane DA. 2005. Complete nucleotide sequence of the chloroplast genome from the Tasmanian
554 blue gum, *Eucalyptus globulus* (Myrtaceae). *DNA research* 12:215-220.
555 Swofford DL. 2003. PAUP. Phylogenetic Analysis Using Parsimony (* and Other Methods).
556 Version 4. *Sunderland: Sinauer Associates Press*.
557 Telefo P, Moundipa P, and Tchouanguep F. 2002. Oestrogenicity and effect on hepatic
558 metabolism of the aqueous extract of the leaf mixture of *Aloe buettneri*, *Dicliptera
559 verticillata*, *Hibiscus macranthus* and *Justicia insularis*. *Fitoterapia* 73:472-478.
Thiel T, Michalek W, Varshney R, and Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and applied genetics* 106:411-422.

Victor Aredo, Jhan Carranza-Cabrera and Raúl Siche. 2017. Inventory of plant species of La Libertad (Peru) and analysis of its agro-industrial potential. *Agroindustrial Science* 7:87-104.

Wang RJ, Cheng C-L, Chang CC, Wu CL, Su TM, and Chaw S-M. 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC evolutionary biology* 8:36.

Weng ML, Blazier JC, Govindu M, and 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, and Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31:3350-3352.

Wicke S, Schneeweiss GM, Müller KF, and Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant molecular biology* 76:273-297.

Wojciech Makalowski and Mark S. Boguski. 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.

Wu Y, Liu F, Yang DG, Li W, Zhou XJ, Pei XY, Liu YG, He KL, Zhang WS, and Ren ZY. 2018. Comparative chloroplast genomics of *Gossypium* species: Insights into repeat sequence variations and phylogeny. *Frontiers in plant science* 9:376.

Feng X, Yuan X, Sun Y, Hu Y, Saman Zulfiqar, Ouyang X, Dang M, Zhou H, Keith Woeste and Zhao P. 2018. Resources for studies of iron walnut (*Juglans sigillata*) gene expression, genetic diversity, and evolution. *Tree Genetics & Genomes* 14:51.

Yang JB, Li DZ, and Li HT. 2014a. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Molecular Ecology Resources* 14:1024-1031.

Yang Y, Yuanye D, Qing L, Jinjian L, Xiwen L, and Yitao W. 2014b. Complete chloroplast genome sequence of poisonous and medicinal plant datura stramonium: organizations and implications for genetic engineering. *PLoS One* 9:e110656.
Yang Z and Willie J. Swanson. 2002. Codon-Substitution Models to Detect Adaptive Evolution that Account for Heterogeneous Selective Pressures Among Site Classes. *Molecular Biology and Evolution* 19:49-57.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* 24:1586-1591.

Yang Z, and Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular biology and evolution* 17:32-43.

Yi DK, and Kim KJ. 2012. Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PLoS One* 7:e35872.

Zhao Z, Wang X, Yu Y, Yuan S, Jiang D, Zhang Y, Zhang T, Zhong W, Yuan Q and Huang L. 2018. Complete chloroplast genome sequences of Dioscorea: Characterization, genomic resources, and phylogenetic analyses. *PeerJ* 6:e6032.

Zhang K, Zhu H, and Gao Y. 2010. Research on active extracts of *Dicliptera chinensis* on liver protection. *China journal of Chinese materia medica* 35:497-498.
Gene maps of chloroplast genomes.

(A) *Diciplerta 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.
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 (Ψycf1) 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.
Figure 3

Comparison of the five *Dicliptera* chloroplast genomes using mVISTA.

CNS indicates conserved noncoding sequences. The Y-scale represents the percent identity between 50% and 100%.
Figure 4

Comparative analysis of the nucleotide diversity (Pi) value among five *Dicliptera* chloroplast genomes.

(A) Coding regions. (B) Non-coding regions.
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%).
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.
Manuscript to be reviewed
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.
Figure 8

Synonymous (KS) substitution rates and ω values (ω=KA/KS) among all Acanthaceae species.

as shown in the column diagram, the order of genes is alphabetical, ω value of ycf15 gene (1.4453) is clearly higher than other genes.
Table 1 (on next page)

Morphological differences among *Dicliptera acuminata*, *D. peruviana*, *D. montana*, *D. mucronata* and *D. ruiziana*. 
| Species   | *D. acuminata* | *D. peruviana* | *D. montana* | *D. mucronata* | *D. ruiziana* |
|-----------|----------------|----------------|--------------|----------------|--------------|
| **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 |
| **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 thyrse | Verticillaster | Pedunculate cyme |
| **Bracts** | Lanceolate-linear, ciliate | Ovate, ciliate | Spatulate, gland-tipped pilose | Obovate-rhombic, pilose | Obovate, gland-tipped pilose |
| **Bracteoles** | Subulate, ciliate | Subulate, ciliate | 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, scarcely pilose | Purple, outside pubescent, outside pilose | Pale purple, outside pilose | Purplish red, outside pilose, scarcely pilose | Pink, outside pubescent, scarcely pilose |
| **Style** | | | | | |
**Table 2** (on next page)

Species information and Genome features of the chloroplast genomes of five *Dicliptera* species.
| Species | D. acuminata | D. peruviana | D. montana | D. ruiziana | D. mucronata |
|---------|--------------|--------------|-------------|-------------|--------------|
| 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 Carabaylo, Lima province, Peru | Lomas de Cháparra, Caravelí province, Peru | Santuario del Amancaay, 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 | MK830556 | MK833945 | MK833946 | MK833947 | MK848596 |
**Table 3** (on next page)

The polymorphic SSRs among five *Dicliptera* species.

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