Complete plastid genome sequences of two species of the Neotropical genus *Brunellia* (Brunelliaceae)

Janice Valencia-D.[1], José Murillo-A.[2], Clara Inés Orozco[2], Carlos Parra-O.[2], Kurt M. Neubig[1]

[1] School of Biological Sciences, Southern Illinois University at Carbondale, Carbondale, Illinois, United States

[2] Instituto de Ciencias Naturales, Universidad Nacional de Colombia, Bogotá D.C., Colombia

Corresponding Author:
Janice Valencia-D.[1]
1125 Lincoln Dr, Carbondale, Illinois, 62901-6509, USA
Email address: janice.valenciaduarte@siu.edu

Abstract

Here we present the first two complete plastid genomes for Brunelliaceae, a Neotropical family with a single genus, *Brunellia*. We surveyed the entire plastid genome in order to find variable cpDNA regions for further phylogenetic analyses across the family. We sampled morphologically different species, *B. antioquensis* and *B. trianae,* and found that the plastid genomes are 157,685 and 157,775 bp in length and display the typical quadripartite structure found in angiosperms. Despite the clear morphological distinction between both species, the molecular data show a very low level of divergence. The amount of nucleotide substitutions per site is one of the lowest reported to date among published congeneric studies (π= 0.00025). Plastid genomes have gene order and content coincident with other COM (Celastrales, Oxalidales, Malpighiales) relatives plastid genomes. Phylogenetic analyses of selected superrosid representatives show high bootstrap support for the ((C,M)O) topology. The N-fixing clade appears as sister group of the COM clade and Zygophyllales as the sister to the rest of the fabids group.

Introduction

Brunellia Ruiz & Pav., with about 60 species, is the only genus within the Neotropical family Brunelliaceae Engl. Its range of distribution is from southern Mexico to Bolivia, and a single species reaches the Greater Antilles (Orozco et al., 2017). They are evergreen trees, some of which can reach 40 m in height and 1 m DBH in high mountain forests, or grow less than 10 m high at over 3,600 m of elevation (Cuatrecasas, 1970). They are also important elements of the Andean ecosystems where they have diversified, and comprise part of the high endemism of these zones (Orozco, 2001). The narrow range of distribution of the vast majority of the species

52 make *Brunellia* an interesting group to study from phylogenetic and biogeographical
53 perspectives.
54
55 Brunelliaceae belong to the order Oxalidales in a clade formed with Cephalotaceae, Cunoniaceae
56 and Elaeocarpaceae (Zhang & Simmons, 2006; Soltis et al., 2011; Heibl & Renner, 2012; Sun et
57 al., 2016a; Li et al., 2019). Oxalidales belong to the commonly called 'COM clade' together with
58 Celastrales and Malpighiales (Matthews & Endress, 2006), and it has been broadly recognized as
59 monophyletic based on plastid data (but see (Zhao et al., 2016; Zeng et al., 2017; Leebens-Mack
60 et al., 2019)). However, despite support for monophyly of each of these three orders, there is still
61 some controversy regarding the relationships among them.
62
63 For Brunelliaceae, the first phylogenetic analysis was made by Orozco (2001) based on
64 morphological characters. In that study, *Brunellia* was recognized as monophyletic with five
65 morphological characters as synapomorphies. Orozco (2001) proposed five sections for the
66 genus, but the phylogenetic tree had poor resolution and low support. A phylogenetic analysis of
67 *Brunellia* using molecular characters, both from nuclear and plastid DNA is currently ongoing
68 (Murillo *et al.* in prep.). In a preliminary survey, plastid DNA regions (*atpB-matK, ndhF, psaB-*
69 *rps14, psaI-accD, psbJ-petA, psbA-trnH, rbcL, trnS-trnG*) were used, but they showed low
70 sequence variation, making them unsuitable for resolving relationships within this genus.
71
72 With the purpose of identifying variable cpDNA regions, we surveyed the entire plastid genome
73 of two morphologically different species, *Brunellia antioquensis* (Cuatrec.) Cuatrec. and *B.*
74 *trianae* Cuatrec., which belong to the biggest sections in the genus, Sect. *Brunellia* and Sect.
75 *Simplicifolia*, respectively. *Brunellia antioquensis* has an ochraceous to fulvous pubescence,
76 stipules 3-5 mm long, compound leaves and fruits with U-shaped endocarp, whereas *B. trianae*
77 has an appressed, lanate to arachnoid, ochraceous indument, stipules 0.5-2 mm long, simple
78 leaves and fruits with boat-shaped endocarp (Fig. 1).
79
80 Here we explore the variability and utility of the plastid DNA in Brunelliaceae and document for
81 first-time complete cp-genome sequences for two species. We characterize the plastid genome of
82 each species and compare them in terms of divergence hotspots in coding and non-coding
83 regions. We also analyze the plastid gene organization of *Brunellia* and ten other COM
84 representatives. Finally, using 75 protein-coding regions available for 41 selected superrosid
85 species; we establish the phylogenetic position of Brunelliaceae and produce a hypothesis of
86 relationships among the COM clade and related orders.
87

Materials & Methods
DNA sampling, extraction and sequencing
90 The samples used in this study were collected under the institutional Universidad Nacional de
91 Colombia collection permit (Number: 0255 March 2014). Collections were made in Colombia in

92 Cerro del Padre Amaya, Antioquia department, rural areas of Medellín, in October 2012 (*B.*
93 *antioquensis C.I. Orozco 4001*, *B. trianae C.I. Orozco 4015*). Sections of leaves were dried on
94 silica gel. Vouchers specimens of both collections were deposited at the Herbario Nacional
95 Colombiano (COL). These species belong to two different clades that were identified using ITS
96 and ETS regions (Murillo et al. pers. obs.).
97
98 DNA was extracted using a CTAB method (Doyle & Doyle, 1987), followed by a silica
99 purification column step and elution in Tris-EDTA (Neubig et al., 2014). DNA samples were
100 adjusted to 50 ng/μL to be sheared to fragments of approximately 500 bp. Library preparation,
101 barcoding and sequencing on an Illumina HiSeqX were conducted at Rapid Genomics LLC
102 (Gainesville, FL, USA).
103

104 **Plastid genome assembly and annotation**
105 Sequencing process produced, on average, 7.5 Gb (s.d. = 0.2 Gb) reads per sample. Paired-end
106 reads were trimmed (quality at 0.05 probability) and later assembled in Geneious 10.2.3
107 (Biomatters Ltd, Auckland, New Zealand). The assembly was performed by a combination of
108 reference and *de novo* assemblies using as a reference the plastid genome of *Averrhoa*
109 *carambola,* Oxalidaceae (GenBank accession KX364202). Annotations of gene-regions and
110 rRNAs were transferred in Geneious from *Averrhoa* reference. The tRNAs were annotated using
111 tRNAscan-SE v2.0 (Lowe & Chan, 2016) implemented in CHLOROBOX
112 (https://chlorobox.mpimp-golm.mpg.de/geseq.html). All annotations were manually reviewed
113 and, if necessary, edited. The circular map of the plastid genome was generated with OGDRAW
114 (https://chlorobox.mpimp-golm.mpg.de/cite-OGDraw.html, Lohse, Drechsel, Kahlau, & Bock,
115 2013) and modified manually.
116

117 **Comparison of plastid genomes of *Brunellia antioquensis* and *B. trianae***
118 Comparisons of the boundaries between the large single-copy (LSC) and the small single-copy
119 (SSC) regions with the inverted repeats (IRA and IRB) were performed using Unipro UGENE
120 v1.31.0 (Okonechnikov et al., 2012). Content percentages of A, T, C, G, A/T and G/C were
121 estimated for LSC, SSC and IR regions, and for the rRNA, tRNA, protein coding regions,
122 intergenic regions, and introns of both species with Bioedit v7.1.11 (Hall, 1999).
123

124 **Repeated sequences**
125 Identification and location of complement, forward, palindromic, and reverse repeat sequences
126 were conducted using REPuter program (Kurtz et al., 2001). Simple sequence repeats (SSRs)
127 were identified using the MISA program with parameters set by default (Thiel et al., 2003).
128

129 **Comparative genome analysis**
130 We compared *Brunellia* plastid genomes regarding nucleotide diversity (π), insertions/deletions
131 (InDel) and base substitutions with the program DnaSP6 (Rozas et al., 2017). Transitions (Ts)

132   and transversions (Tv) events were identified across all loci and for protein-coding regions, they
133   were also classified based on the effect in the amino acid chain as synonymous (S) or non-
134   synonymous (N) substitutions. Furthermore, the whole plastid genome sequences of *Brunellia*
135   and eight other COM clade representatives were compared using Mauve (Darling et al., 2004)
136   and mVISTA programs (Frazer et al., 2004). The eight sequences retrieved from GenBank are
137   listed in Table S1. Plastid sequences of *Ceratopetalum apetalum, Sloanea australis* and *Tetraena*
138   *mongolica* assembled by Foster et al. (2016) were annotated using the same procedure we
139   implemented for the *Brunellia* samples.
140
141   **Phylogenomic Analyses**
142   We extracted and aligned by translation the protein-coding genes of the ten COM representatives
143   in Geneious. We added 33 superrosid representatives (Table S1) from previously published
144   matrices by Ruhfel et al. (2014a), and pruned our data accordingly. The genes considered were
145   *atpE-F, atpH-I, ccsA, cemA, clpP, infA, matK, ndhA-K, petA-B, petD, petG, petL, petN, psaA-C,*
146   *psaI, psaJ, psbA-F, psbH-N, psbT, psbZ, rbcL, rpl2, rpl14, rpl16, rpl20, rpl22-23, rpl32-33,*
147   *rpl36, rpoA-B, rpoC1-C2, rps2-4, rps7-8, rps11-12, rps14-16, rps18-19* and *ycf2-4*. The *accD*
148   gene alignment was removed from the final concatenated matrix due to its high variability that
149   produced conflicts in the alignment. The *accD* has been reported as missing in species of
150   Fabaceae (Fabids) and Geraniaceae (Malvids), and in some cases it has transfer to the nucleus
151   (Rousseau-Gueutin et al., 2013; Liu et al., 2016). Phylogenetic analyses were performed using
152   Maximum Likelihood using IQ-Tree web server (Nguyen et al., 2014; Trifinopoulos et al., 2016)
153   with 1000 ultrafast bootstrap replicates (Hoang et al., 2017) and the GTR+F+R3 model selected
154   with ModelFinder under BIC (Kalyaanamoorthy et al., 2017).
155
156   *Information availability*- The plastid genomes of *B. antioquensis* and *B. trianae* were prepared
157   for submission for NCBI using GB2sequin (Lehwark & Greiner, 2018) and are available in the
158   GenBank portal with the numbers MN585217 and MN615725. The nucleotide alignments are
159   available in supplemental information.
160

161   # Results
162   **Plastid genome features**
163   A total of 73,003 for *B. antioquensis* and 97,054 for *B. trianae* paired-end reads were assembled
164   to produce a mean coverage of 104.8× and 138.3×, respectively. Plastid genome size of *B.*
165   *antioquensis* and *B. trianae* is 157,685 bp and 157,775 bp, respectively (Table 1) and consists of
166   a single circular, double-stranded DNA sequence, showing the typical quadripartite structure
167   (Fig. 2); both sequences include a pair of IR (26,376 bp in *B. antioquensis* and 26,389 bp in *B.*
168   *trianae*), one SSC (17,538 bp in *B. antioquensis* and 17,542 bp in *B. trianae*), and one LSC
169   (87,395 in *B. antioquensis* and 87,455 bp in *B. trianae*) (Table 1). The proportion of each region
170   in the plastid genome for *B. antioquensis* is LSC 55.42%, SSC 11.12% and IRs 33.46%; for *B.*

171 *trianae* these percentages are quite similar (Table 1). Content percentages of A, T, C, G, A/T and
172 G/C are similar for both species (Table 1).
173
174 Both genomes consist of 52.13-52.14% protein-coding genes, 40.67-40.71% intergenic regions,
175 10.32-10.33% introns, 5.73% rRNAs, and 1.85% tRNAs (Table S2). Plastid genomes of
176 *Brunellia* species has 134 coding regions, including 38 tRNAs, 8 rRNAs, and 88 protein-coding
177 genes. Ninety-five genes are unique and 19 are duplicated in the IRs regions (Table S3); of the
178 19 duplicate genes, seven are tRNAs, four are rRNAs, and seven are fully included and one
179 interrupted protein-coding genes. Coding regions are distributed as follows: LSC included 22
180 tRNAs and 59 protein-coding regions, IRs has 4 rRNAs, 7 tRNAs, and 7 protein-coding regions,
181 and SSC contains one tRNA and 11 protein-coding regions; additionally, *ycf1* is found between
182 IRA and SSC border region, and *rps12* appears three times (twice at the IRs and other one at
183 LSC region). Most of the protein-coding regions have one exon, 15 genes have one intron,
184 whereas *clpP*, *rps12*, and *ycf3* present two introns (Table S4). Size of introns is between 544 to
185 2549 bp, being *trnK-UUU* the longer on the whole cpDNA. *Brunellia antioquensis* and *B.*
186 *trianae* have similar IRA-SSC and SSC-IRB boundaries, but they are somewhat different by the
187 number of base pairs at the LSC-IR borders (Fig. 3).
188
189 **Repeated sequences and SSRs**
190 Twenty-six repeated sequences are found in the genome of *B. antioquensis*, whose lengths vary
191 between 20 and 30 bp (Table S5). Twelve of those sequences are located in intergenic regions.
192 Twelve repeats are palindromic, eight are forward repeats, three are complement repeats, and
193 three are reverse repeats. Fourteen repeats are found in the LSC, six in the SSC, and four in the
194 IR. Number, size and locations of repeated sequences in *B. trianae* are different (Table S6). This
195 species has 18-repeated sequences with a length that varies between 21 and 88 bp; ten of these
196 repeats are located in intergenic regions. Of the 18 repeats, six are palindromic, 11 are forward
197 repeats, and one is a complement repeat; no reverse repeats are found in the cpDNA of this
198 species. Thirteen repeats are found in the LSC, one in the SSC, and four in the IR regions.
199
200 The numbers of SSRs are quite similar between the two *Brunellia* species. *Brunellia*
201 *antioquensis* has 223 SSRs, including 124 mononucleotides (66 poly T, 56 poly A, 1 poly C, 1
202 poly G), 19 dinucleotides, 69 trinucleotides, seven tetranucleotides, one pentanucleotide, one
203 hexanucleotide, one 13-nucleotide, and one 21-nucleotide. *Brunellia trianae*, with 225 SSRs,
204 differs from *B. antioquensis* by the number of dinucleotides (18) and the number of poly T (69).
205 Dinucleotide SSRs are only represented by units containing TA/AT (82.4%) and TC/CT (17.6%)
206 in both species.
207
208 **Analysis of polymorphisms among the plastid genome sequences of *Brunellia***
209 *Brunellia antioquensis* and *B. trianae* has very similar sequences. These species share 99.85%
210 sequence similarity. In the analysis of the 157,844 bp aligned (gaps included), only 243 bp are

211   different corresponding to 204 indels and 39 nucleotide substitutions (Table S7). The nucleotide
212   variability expressed as the level of divergence among the sequences is $\pi=0.00025$. An
213   examination by sliding window of the alignment shows π values range from 0 to 0.0067 for a
214   600 bp gradation (Fig. 4). Hotspots correspond with the intergenic regions *rps2-rpoC2, petB-*
215   *ropA* and *ndhF-ccsA*. The SSC has the highest amount of substitutions per site ($\pi=0.00068$),
216   followed by the LSC ($\pi=0.00026$); the IRs have the lowest value with only two substitutions in
217   each ($\pi=7.6 \times 10^{-5}$). Twenty-two transversions and 11 transitions evenly distributed among the
218   regions of the genome are found, with a ratio of 2:1 of Tv to Ts. Fourteen substitutions occur in
219   protein coding sequences of ten genes, including five transversions and nine transitions, giving a
220   ratio of 1:1.8 of Tv to Ts. Only five substitutions are synonymous, four of them due to the
221   replacement of thymine (T) by cytosine (C).

223   Among the sequences, two thirds of the indel events (20) involve the gain or loss of single
224   nucleotides of adenine (A) or thymine (T), and they are associated with regions where they
225   appear repeated more than six times. Twenty-four indels are located in intergenic regions, six
226   indels in introns and one in the *rpoC2* exon. This last indel creates a change in nine nucleotide
227   sites and causes three amino acid changes. The longest indel was 83 bp long and is located in the
228   LSC region between *petN* and *psbM*.

230   **Comparison of the plastid genomes of COM clade representatives**
231   Three Locally Collinear Blocks are identified as the LSC, IR, and SSC (Fig. S1). The possible
232   lack of information of the *Sloanea australis* plastid genome in the regions adjacent to the LSC-
233   IR and IR-SSC junctions may be the reason why there is not one longer collinear block. This is
234   also detectable in the Mauve output as two inversions in the LSC and SSC. Those apparent
235   inversions denote a different starting point for the reported *Sloanea* sequence, which might be
236   caused by a possible problem with the assembly. In the mVISTA analysis (Fig. S2), using *B.*
237   *antioquensis* as a reference, there is high similarity among *Brunellia* and other Oxalidales plastid
238   genomes.

240   **Phylogenomic analyses**
241   Seventy-five protein-coding regions were aligned in a matrix of 54,367 bp length. ML analysis
242   places Brunelliaceae in a clade with Cunoniaceae and Elaeocarpaceae, where Oxalidaceae is
243   sister to these 3 families (Fig. 5). The monophyly of the COM clade, as well as Fabids and
244   Malvids, is recovered with 100% bootstrap support (BS). Oxalidales appear as sister group of the
245   Celastrales-Malpighiales clade with 89% BS. The sister group of the COM clade is the N-fixing
246   clade, and Zygophyllales is sister to them with 100% BS in both cases. The Malvid clade is sister
247   to Fabids and is supported by 100% BS.

Discussion
250   We report the first two plastid genomes for Brunelliaceae that are added to the previously

251   published data for Oxalidales: the fully annotated plastid genome of *Averrhoa carambola* (Jo et
252   al., 2016), the preliminary plastid sequences of *Ceratopetalum apetalum* and *Sloanea australis*
253   (Foster et al., 2016) and the plastid regions of *Oxalis latifolia* (Moore et al., 2010). Structure,
254   organization, and AT/GC content are similar to the *A. carambola*, COM group, and other
255   Eudicot plastid genomes (e.g. Ravi, Khurana, Tyagi, & Khurana, 2008).

257   Although *B. antioquiensis* and *B. trianae* have different morphological characters in their leaves,
258   indument and fruits, the plastid sequence of them is surprisingly similar. *Brunellia trianae* has a
259   plastid genome that is only 90 bp longer than *B. antioquensis* (Table 1). Differences among the
260   size of the sequences have been related to the expansion and contraction of the IR regions
261   (Downie & Jansen, 2015), but in *Brunellia* it does not seem to be the case; variations of the
262   boundaries occur only between the LSC region and the IRs (Fig. 3), and at these boundaries the
263   variations are small (12 to 13 nucleotides). Therefore, IR contractions or expansions do not affect
264   the size of the genome in the two species studied. Plastid genome variation size between species
265   is mostly explained by an indel of 88 bp located in the LSC region (between *psbM* and *petN*) in
266   *B. antioquensis*. Similar results were found in *Tilia*, another rosid with low variability among
267   congeneric genomes (Cai et al., 2015).

269   Variations between *B. antioquensis* and *B. trianae* comprise a difference of 39 nucleotides
270   substitutions and 204 indels. The sequences show relatively low nucleotide variability
271   (π=0.00025), with IRs less variable (π=7.6x10^{-5}) as expected due to their effect in keeping the
272   structure and stabilization of the chromosome (e.g. Sun et al., 2016; Wicke, Schneeweiss,
273   dePamphilis, Müller, & Quandt, 2011). The nucleotide variability found among *Brunellia*
274   species is also one of the lowest reported. In general, the π values in trees tend to be lower than
275   in herbs; for example, in trees: *Machilus* (π=0.00154, Song et al., 2015) and *Juglans* (π=0.00219,
276   Hu, Woeste, & Zhao, 2017); in herbs: *Nicotiana* (π=0.00321, Asaf et al., 2016), *Aconitum*
277   (π=0.00549, Kong, Liu, Yao, & Gong, 2017) and *Papaver* (π=0.00895, Zhou et al., 2018).
278   Discrepancies in the rate of molecular substitutions can be explained by variations of generation
279   time, size and habit, the age of the clade, and molecular mechanisms of DNA repair, among
280   others (e.g. Lanfear et al., 2013; Smith & Donoghue, 2008). In the case of *Brunellia* trees, likely
281   rapid and recent diversification in the last 5 mya (Murillo et al. *in prep.*) might explain a π value
282   ten times smaller than the one recorded for other trees. High morphological diversity and low
283   molecular divergence has also been found in other Andean clades that speciated during the Plio-
284   Pleistocene (e. g. Hughes & Eastwood, 2006; Nürk, Scheriau, & Madriñán, 2013).

286   Previously, we analyzed eight plastid regions and found low variability and no phylogenetic
287   signal (Murillo et al. *in prep.*). This fact is not surprising, given the overall low variability in the
288   complete plastid genome. In this study we found three variable intergenic regions (*rps2-rpoC2,*
289   *petB-ropA* and *ndhF-ccsA,* π= 0.005-0.0067); although with low sequence divergence, they could
290   represent the most variable plastid regions useful in phylogenetic analysis of *Brunellia*.

291

292   The use of either complete plastid genomes or many coding regions have proved to increase the

293   resolution and support of the relationships in phylogenetic analyses (e.g. Jansen et al., 2007;

294   Moore, Bell, Soltis, & Soltis, 2007). In our analyses, the inclusion of *Brunellia* sequences, the

295   data produced by Foster et al. (2016; *Sloanea australis* and *Ceratophyllum apetalum*), and

296   genome published by Jo et al. (2016; *Averrhoa carambola*), Gu, Tembrock, Zheng, & Wu (2018;

297   *Catha edulis*) and Xia et al. (2018; *Parnassia brevistyla*) gave us the opportunity to improve the

298   dataset for Oxalidales and Celastrales in terms of amount of characters and taxon sampling.

299   Despite the possible assembly issues in the *S. australis* plastid genome, the coding regions placed

300   it in the expected phylogenetic position, so we decided to include it in the complete analysis. To

301   date, plastid sequences of *Averrhoa* and *Oxalis* were the only ones used to represent the whole

302   order (Sun et al., 2015; Jo et al., 2016).

303

304   The phylogenetic reconstruction placed Brunelliaceae in the clade Oxalidales and sister to the

305   clade Cunoniaceae-Elaeocarpaceae with 76% BS, which is congruent with Soltis et al. (2011).

306   These relationships disagree with Zhang & Simmons (2006) in which Brunelliaceae is sister to

307   Cunoniaceae. Sun et al. (2016) and Heibl & Renner (2012) include in their analyses

308   Cephalotaceae (not present in our sampling) and do not recover the clade Cunoniaceae-

309   Elaeocarpaceae. In these, Brunelliaceae is sister to Cephalotaceae, this clade sister to

310   Elaeocarpaceae, and Cunoniaceae sister to all three. Further work is needed with a wider

311   sampling to disentangle the phylogenetic relationships among these four families.

312

313   Our results depict that the COM clade is monophyletic with 100% BS, which is consistent with

314   many phylogenetic analyses from plastid data (see Table S8). However, nuclear data sets have

315   not shown this monophyly, and support the relationships (C,M)(O,Malvids p.p.) (Zhao et al.,

316   2016; Zeng et al., 2017; Leebens-Mack et al., 2019). Inside the COM clade, the relationships

317   among the three orders have been conflicting. Phylogenetic analyses that support each one of the

318   three possible topologies among the orders are compiled in Table S8. We found that Celastrales

319   and Malpighiales form a clade with 89% BS making the ((C,M)O) topology the most likely

320   according to our data. These relationships had been found by Li et al. (2019) using 80 plastid

321   genes, by Cauz-Santos et al. (2017) using 43 plastid coding regions, by Moore et al. (2011) using

322   all IR data, and by Zhang & Simmons (2006), Soltis, Gitzendanner & Soltis (2007) and Burleigh

323   et al. (2009) using few plastid and nuclear genes. The (C(O,M)) topology was found by Ruhfel et

324   al. (2014b) using 1,2 codon positions and Wu et al. (2014) using all three, however both show

325   low bootstrap support. The third possible topology ((C,O)M) has been also recovered using

326   plastid coding regions (Ruhfel et al., 2014b; Sun et al., 2015), but the change in the taxon

327   sampling leads to the ((C,M)O) topology (Fig. 5). Therefore, conflicting results with other

328   plastid-based phylogenies may be caused by the small sampling of Celastrales (i.e. *Euonymus*

329   *americanus*) and Oxalidales (i.e. *Oxalis latifolia* or *Averrhoa carambola*) in the analyses (Ruhfel

330 et al., 2014b; Wu et al., 2014; Sun et al., 2015; Jo et al., 2016) or by the use of an amino acid
331 dataset (Ruhfel et al., 2014b; Gitzendanner et al., 2018).

332

333 The phylogenetic analysis shows that the N-fixing clade is sister to the COM clade with 100%
334 BS. This relationship has been found in other studies based on plastid data (Sun et al. 2015,
335 Table 1). However, when nuclear and mitochondrial data are analyzed the COM clade is sister to
336 Malvids or even not monophyletic. Sun et al. (2015) developed a hypothesis of the possible
337 origin of the COM clade genome by the action of introgression and hybridization processes from
338 both malvid and fabid ancestors. This scenario, in which the mitochondrial and most of the
339 nuclear genomes would have their origin from a Malvid ancestor, and the plastid genome from a
340 Fabid ancestor, would explain the observed discrepancies in the phylogenies (Sun et al., 2015).
341 Therefore, deep analyses based on molecular data from all three genomes will give us a better
342 understanding of the evolutionary processes within the diversity of rosids.

343

344 Based on our analysis of plastid data, the Fabid clade appears monophyletic (including
345 Zygophyllales) with 100% BS. Zygophyllales is sister to the rest of Fabids, as has been
346 previously found with large datasets (e.g. Soltis et al., 2011; M. Sun et al., 2015), but further
347 analysis including more representatives of Zygophyllales is necessary because other studies have
348 found the order sister to Malvids (Zhao et al., 2016) or in an uncertain position (Ruhfel et al.,
349 2014b).

350

351

Conclusions

353 Assemblies of complete plastid genomes have proved to be useful for the identification of
354 variable sites that can work as barcodes for specific groups; however, *Brunellia* sequences do not
355 show high levels of variation and are not particularly useful for this purpose. Our results agree
356 with previous analyses in which we tested eight plastid regions for phylogenetic purposes. Here,
357 we present for the first time, plastid genomes for Brunelliaceae. We analyzed the genomes of
358 two species to look for regions with high variability, but we found a very low level of nucleotide
359 diversity. The three main hotspots identified among them were mainly composed by indel
360 regions, with some potential in further systematic analyses. Phylogenetic analyses of *Brunellia*
361 plastid regions plus another 41 superrosid representatives corroborate the placement of the
362 family in Oxalidales, and as sister group of a clade formed by Cunoniaceae and Elaeocarpaceae.
363 Our data also support Oxalidales as sister of the Celastrales-Malpighiales clade. This topology
364 disagrees with most common scenario of Celastrales and Oxalidales as sister groups, however,
365 all three possible topologies have been found with both small and large datasets. Possible causes
366 of incongruence may be related to taxon sampling and the lack of information for many species
367 in the less diverse orders, as Celastrales and Oxalidales, which tend to be poorly represented in
368 the phylogenies. Future studies focused on the COM group should include a better representation

369 of families and genera in order to understand other evolutionary trends in this relatively recently
370 recognized clade.

371
372

Acknowledgements
374 We would like to thank Jorge Vélez and Juan Pablo Tobón at the Universidad Nacional de
375 Colombia - sede Medellín, for assisting with fieldwork.

376
377

References
379 Asaf S, Khan AL, Khan AR, Waqas M, Kang S-M, Khan MA, Lee S-M, Lee I-J. 2016. Complete
380   chloroplast genome of *Nicotiana otophora* and its comparison with related species. *Frontiers in*
381   *Plant Science* 7. DOI: 10.3389/fpls.2016.00843.
382 Burleigh JG, Hilu KW, Soltis DE. 2009. Inferring phylogenies with incomplete data sets: a 5-gene, 567-
383   taxon analysis of angiosperms. *BMC Evolutionary Biology* 9:61. DOI: 10.1186/1471-2148-9-61.
384 Cai J, Ma P-F, Li H-T, Li D-Z. 2015. Complete plastid genome sequencing of four *Tilia* species
385   (Malvaceae): a comparative analysis and phylogenetic implications. *PLoS One* 10:e0142705.
386 Cauz-Santos LA, Munhoz CF, Rodde N, Cauet S, Santos AA, Penha HA, Dornelas MC, Varani AM,
387   Oliveira GCX, Bergès H, Vieira MLC. 2017. The chloroplast genome of *Passiflora edulis*
388   (Passifloraceae) assembled from long sequence reads: Structural organization and phylogenomic
389   studies in Malpighiales. *Frontiers in Plant Science* 8. DOI: 10.3389/fpls.2017.00334.
390 Cuatrecasas J. 1970. Brunelliaceae. *Flora Neotropica* 2:1–188.
391 Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic
392   sequence with rearrangements. *Genome research* 14:1394–1403. DOI: 10.1101/gr.2289704.
393 Downie SR, Jansen RK. 2015. A comparative analysis of whole plastid genomes from the Apiales:
394   expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and
395   identification of highly divergent noncoding regions. *Systematic Botany* 40:336–351.
396 Doyle J, Doyle J. 1987. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem*
397   *Bull* 19:11–15.
398 Foster CSP, Sauquet H, van der Merwe M, McPherson H, Rossetto M, Ho SYW. 2016. Evaluating the
399   impact of genomic data and priors on bayesian estimates of the Angiosperm evolutionary
400   timescale. *Systematic Biology*:syw086. DOI: 10.1093/sysbio/syw086.
401 Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for
402   comparative genomics. *Nucleic Acids Research* 32:W273–W279.
403 Gitzendanner MA, Soltis PS, Wong GK, Ruhfel BR, Soltis DE. 2018. Plastid phylogenomic analysis of
404   green plants: a billion years of evolutionary history. *American Journal of Botany* 105:291–301.
405 Gu C, Tembrock L, Zheng S, Wu Z. 2018. The complete chloroplast genome of *Catha edulis*: A
406   comparative analysis of genome features with related species. *International Journal of Molecular*
407   *Sciences* 19:525. DOI: 10.3390/ijms19020525.
408 Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for
409   Windows 95/98/NT. In: [London]: Information Retrieval Ltd., c1979-c2000., 95–98.

410 Heibl C, Renner SS. 2012. Distribution models and a dated phylogeny for Chilean *Oxalis* species reveal
411   occupation of new habitats by different lineages, not rapid adaptive radiation. *Systematic Biology*
412   61:823–834.
413 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Le SV. 2017. UFBoot2: Improving the Ultrafast
414   Bootstrap Approximation. *Molecular biology and evolution*:msx281.
415 Hu Y, Woeste KE, Zhao P. 2017. Completion of the chloroplast genomes of five Chinese *Juglans* and
416   their contribution to chloroplast phylogeny. *Frontiers in Plant Science* 7. DOI:
417   10.3389/fpls.2016.01955.
418 Hughes C, Eastwood R. 2006. Island radiation on a continental scale: Exceptional rates of plant
419   diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences*
420   103:10334–10339. DOI: 10.1073/pnas.0601928103.
421 Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Müller KF, Guisinger-
422   Bellian M, Haberle RC, Hansen AK. 2007. Analysis of 81 genes from 64 plastid genomes
423   resolves relationships in angiosperms and identifies genome-scale evolutionary patterns.
424   *Proceedings of the National Academy of Sciences* 104:19369–19374.
425 Jo S, Kim HW, Kim YK, Cheon SH, Kim K-J. 2016. Complete plastome sequence of *Averrhoa
426   carambola* L. (Oxalidaceae). *Mitochondrial DNA Part B* 1:609–611. DOI:
427   10.1080/23802359.2016.1209095.
428 Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model
429   selection for accurate phylogenetic estimates. *Nature Methods*.
430 Kong H, Liu W, Yao G, Gong W. 2017. A comparison of chloroplast genome sequences in *Aconitum*
431   (Ranunculaceae): a traditional herbal medicinal genus. *PeerJ* 5:e4018. DOI: 10.7717/peerj.4018.
432 Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001. REPuter: the
433   manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* 29:4633–
434   4642.
435 Lanfear R, Ho SYW, Davies TJ, Moles AT, Aarssen L, Swenson NG, Warman L, Zanne AE, Allen AP.
436   2013. Taller plants have lower rates of molecular evolution. *Nature Communications* 4:1879.
437   DOI: 10.1038/ncomms2836.
438 Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I,
439   Li Z, Melkonian M, Mirarab S, Porsch M, Quint M, Rensing SA, Soltis DE, Soltis PS,
440   Stevenson DW, Ullrich KK, Wickett NJ, DeGironimo L, Edger PP, Jordon-Thaden IE, Joya S,
441   Liu T, Melkonian B, Miles NW, Pokorny L, Quigley C, Thomas P, Villarreal JC, Augustin MM,
442   Barrett MD, Baucom RS, Beerling DJ, Benstein RM, Biffin E, Brockington SF, Burge DO,
443   Burris JN, Burris KP, Burtet-Sarramegna V, Caicedo AL, Cannon SB, Çebi Z, Chang Y, Chater
444   C, Cheeseman JM, Chen T, Clarke ND, Clayton H, Covshoff S, Crandall-Stotler BJ, Cross H,
445   dePamphilis CW, Der JP, Determann R, Dickson RC, Di Stilio VS, Ellis S, Fast E, Feja N, Field
446   KJ, Filatov DA, Finnegan PM, Floyd SK, Fogliani B, García N, Gâteblé G, Godden GT, Goh F
447   (Qi Y, Greiner S, Harkess A, Heaney JM, Helliwell KE, Heyduk K, Hibberd JM, Hodel RGJ,
448   Hollingsworth PM, Johnson MTJ, Jost R, Joyce B, Kapralov MV, Kazamia E, Kellogg EA,
449   Koch MA, Von Konrat M, Könyves K, Kutchan TM, Lam V, Larsson A, Leitch AR, Lentz R, Li
450   F-W, Lowe AJ, Ludwig M, Manos PS, Mavrodiev E, McCormick MK, McKain M, McLellan T,
451   McNeal JR, Miller RE, Nelson MN, Peng Y, Ralph P, Real D, Riggins CW, Ruhsam M, Sage
452   RF, Sakai AK, Scascitella M, Schilling EE, Schlösser E-M, Sederoff H, Servick S, Sessa EB,
453   Shaw AJ, Shaw SW, Sigel EM, Skema C, Smith AG, Smithson A, Stewart CN, Stinchcombe JR,

454   Szövényi P, Tate JA, Tiebel H, Trapnell D, Villegente M, Wang C-N, Weller SG, Wenzel M,
455   Weststrand S, Westwood JH, Whigham DF, Wu S, Wulff AS, Yang Y, Zhu D, Zhuang C,
456   Zuidof J, Chase MW, Pires JC, Rothfels CJ, Yu J, Chen C, Chen L, Cheng S, Li J, Li R, Li X,
457   Lu H, Ou Y, Sun X, Tan X, Tang J, Tian Z, Wang F, Wang J, Wei X, Xu X, Yan Z, Yang F,
458   Zhong X, Zhou F, Zhu Y, Zhang Y, Ayyampalayam S, Barkman TJ, Nguyen N, Matasci N,
459   Nelson DR, Sayyari E, Wafula EK, Walls RL, Warnow T, An H, Arrigo N, Baniaga AE,
460   Galuska S, Jorgensen SA, Kidder TI, Kong H, Lu-Irving P, Marx HE, Qi X, Reardon CR,
461   Sutherland BL, Tiley GP, Welles SR, Yu R, Zhan S, Gramzow L, Theißen G, Wong GK-S, One
462   Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the
463   phylogenomics of green plants. *Nature* 574:679–685. DOI: 10.1038/s41586-019-1693-2.
464   Lehwark P, Greiner S. 2018. GB2sequin-A file converter preparing custom GenBank files for database
465   submission. *Genomics*.
466   Li H-T, Yi T-S, Gao L-M, Ma P-F, Zhang T, Yang J-B, Gitzendanner MA, Fritsch PW, Cai J, Luo Y,
467   Wang H, van der Bank M, Zhang S-D, Wang Q-F, Wang J, Zhang Z-R, Fu C-N, Yang J,
468   Hollingsworth PM, Chase MW, Soltis DE, Soltis PS, Li D-Z. 2019. Origin of angiosperms and
469   the puzzle of the Jurassic gap. *Nature Plants* 5:461–470. DOI: 10.1038/s41477-019-0421-0.
470   Liu T-J, Zhang C-Y, Yan H-F, Zhang L, Ge X-J, Hao G. 2016. Complete plastid genome sequence of
471   *Primula sinensis* (Primulaceae): structure comparison, sequence variation and evidence for *accD*
472   transfer to nucleus. *PeerJ* 4:e2101. DOI: 10.7717/peerj.2101.
473   Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDRAW—a suite of tools for
474   generating physical maps of plastid and mitochondrial genomes and visualizing expression data
475   sets. *Nucleic acids research* 41:W575–W581.
476   Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer
477   RNA genes. *Nucleic Acids Research* 44:W54–W57. DOI: 10.1093/nar/gkw413.
478   Matthews ML, Endress PK. 2006. Floral structure and systematics in four orders of rosids, including a
479   broad survey of floral mucilage cells. *Plant Systematics and Evolution*. DOI: 10.1007/s00606-
480   006-0443-8.
481   Moore MJ, Bell CD, Soltis PS, Soltis DE. 2007. Using plastid genome-scale data to resolve enigmatic
482   relationships among basal angiosperms. *Proceedings of the National Academy of Sciences*
483   104:19363–19368.
484   Moore MJ, Hassan N, Gitzendanner MA, Bruenn RA, Croley M, Vandeventer A, Horn JW, Dhingra A,
485   Brockington SF, Latvis M, Ramdial J, Alexandre R, Piedrahita A, Xi Z, Davis CC, Soltis PS,
486   Soltis DE. 2011. Phylogenetic analysis of the plastid inverted repeat for 244 species: Insights into
487   deeper-level angiosperm relationships from a long, slowly evolving sequence region.
488   *International Journal of Plant Sciences* 172:541–558. DOI: 10.1086/658923.
489   Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. 2010. Phylogenetic analysis of 83 plastid genes
490   further resolves the early diversification of eudicots. *Proceedings of the National Academy of*
491   *Sciences* 107:4623–4628. DOI: 10.1073/pnas.0907801107.
492   Neubig KM, Whitten WM, Abbot JR, Elliott S, Soltis DE, Soltis P. 2014. Variables affecting DNA
493   preservation in archival plant specimens. In: Applequist W, Campbell L eds. *DNA Banking for*
494   *the 21st Century: Proceedings of the US Workshop on DNA Banking*. St. Louis, Miss: William L.
495   Brown Center, Missouri Botanical Garden, 81–112.

496 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: a fast and effective stochastic
497   algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*
498   32:268–274.
499 Nickrent DL, Der JP, Anderson FE. 2005. Discovery of the photosynthetic relatives of the" Maltese
500   mushroom" *Cynomorium*. *BMC Evolutionary Biology* 5:38.
501 Nürk NM, Scheriau C, Madriñán S. 2013. Explosive radiation in high Andean *Hypericum*—rates of
502   diversification among New World lineages. *Frontiers in Genetics* 4. DOI:
503   10.3389/fgene.2013.00175.
504 Okonechnikov K, Golosova O, Fursov M, Ugene Team. 2012. Unipro UGENE: a unified bioinformatics
505   toolkit. *Bioinformatics* 28:1166–1167.
506 Orozco. 2001. *Evolutionary biology of* Brunellia *Ruiz & Pavon (Brunelliaceae, Oxalidales). Ph.D.*
507   *Dissertation*. Amsterdam: Universiteit van Amsterdam.
508 Orozco, Perez AJ, Romoleroux K, Murillo Aldana J. 2017. The discovery of a new species of *Brunellia*
509   (Brunelliaceae) with ephemeral petals from Llanganates National Park, Ecuador. *Phytotaxa*
510   311:263–269.
511 Ravi V, Khurana JP, Tyagi AK, Khurana P. 2008. An update on chloroplast genomes. *Plant Systematics*
512   *and Evolution* 271:101–122. DOI: 10.1007/s00606-007-0608-0.
513 Rousseau-Gueutin M, Huang X, Higginson E, Ayliffe M, Day A, Timmis JN. 2013. Potential functional
514   replacement of the plastidic acetyl-CoA carboxylase subunit (*accD*) gene by recent transfers to
515   the nucleus in some angiosperm lineages. *Plant Physiology* 161:1918–1929. DOI:
516   10.1104/pp.113.214528.
517 Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-
518   Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular*
519   *biology and evolution* 34:3299–3302.
520 Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014a. Data from: From algae to
521   angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes.
522   DOI: 10.5061/dryad.k1t1f.
523 Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014b. From algae to angiosperms–
524   inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC*
525   *Evolutionary Biology* 14:23.
526 Smith SA, Donoghue MJ. 2008. Rates of molecular evolution are linked to life history in flowering
527   plants. *Science* 322:86–89. DOI: 10.1126/science.1163197.
528 Soltis DE, Gitzendanner MA, Soltis PS. 2007. A 567-taxon data set for angiosperms: the challenges
529   posed by Bayesian analyses of large data sets. *International Journal of Plant Sciences* 168:137–
530   157.
531 Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF,
532   Walker JB, Moore MJ, Carlsward BS, Bell CD, Latvis M, Crawley S, Black C, Diouf D, Xi Z,
533   Rushworth CA, Gitzendanner MA, Sytsma KJ, Qiu Y-L, Hilu KW, Davis CC, Sanderson MJ,
534   Beaman RS, Olmstead RG, Judd WS, Donoghue MJ, Soltis PS. 2011. Angiosperm phylogeny: 17
535   genes, 640 taxa. *American Journal of Botany* 98:704–730. DOI: 10.3732/ajb.1000404.
536 Song Y, Dong W, Liu B, Xu C, Yao X, Gao J, Corlett RT. 2015. Comparative analysis of complete
537   chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae*
538   in the family Lauraceae. *Frontiers in Plant Science* 6. DOI: 10.3389/fpls.2015.00662.

539 Sun Y, Moore MJ, Zhang S, Soltis PS, Soltis DE, Zhao T, Meng A, Li X, Li J, Wang H. 2016a.
540   Phylogenomic and structural analyses of 18 complete plastomes across nearly all families of
541   early-diverging eudicots, including an angiosperm-wide analysis of IR gene content evolution.
542   *Molecular Phylogenetics and Evolution* 96:93–101. DOI: 10.1016/j.ympev.2015.12.006.
543 Sun M, Naeem R, Su J-X, Cao Z-Y, Burleigh JG, Soltis PS, Soltis DE, Chen Z-D. 2016b. Phylogeny of
544   the Rosidae: A dense taxon sampling analysis: Phylogeny of the Rosidae. *Journal of Systematics*
545   *and Evolution* 54:363–391. DOI: 10.1111/jse.12211.
546 Sun M, Soltis DE, Soltis PS, Zhu X, Burleigh JG, Chen Z. 2015. Deep phylogenetic incongruence in the
547   angiosperm clade Rosidae. *Molecular Phylogenetics and Evolution* 83:156–166. DOI:
548   10.1016/j.ympev.2014.11.003.
549 Thiel T, Michalek W, Varshney R, Graner A. 2003. Exploiting EST databases for the development and
550   characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and*
551   *Applied Genetics* 106:411–422.
552 Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic
553   tool for maximum likelihood analysis. *Nucleic acids research* 44:W232–W235.
554 Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid
555   chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology*
556   76:273–297. DOI: 10.1007/s11103-011-9762-4.
557 Wu Z, Gui S, Quan Z, Pan L, Wang S, Ke W, Liang D, Ding Y. 2014. A precise chloroplast genome of
558   *Nelumbo nucifera* (Nelumbonaceae) evaluated with Sanger, Illumina MiSeq, and PacBio RS II
559   sequencing platforms: insight into the plastid evolution of basal eudicots. *BMC Plant Biology*
560   14:289.
561 Xia M, Zhang F, Rao H, Chi X, Khan G, Zhang Y, Yu J, Chen S. 2018. Complete chloroplast genome
562   sequence of *Parnassia brevistyla* (Celastraceae) and phylogenetic analysis with related species.
563   *Mitochondrial DNA Part B* 3:1187–1188. DOI: 10.1080/23802359.2018.1524725.
564 Zeng L, Zhang N, Zhang Q, Endress PK, Huang J, Ma H. 2017. Resolution of deep eudicot phylogeny
565   and their temporal diversification using nuclear genes from transcriptomic and genomic datasets.
566   *New Phytologist* 214:1338–1354. DOI: 10.1111/nph.14503.
567 Zhang L-B, Simmons MP. 2006. Phylogeny and delimitation of the Celastrales inferred from nuclear and
568   plastid genes. *Systematic Botany* 31:122–137. DOI: 10.1600/036364406775971778.
569 Zhao L, Li X, Zhang N, Zhang S-D, Yi T-S, Ma H, Guo Z-H, Li D-Z. 2016. Phylogenomic analyses of
570   large-scale nuclear genes provide new insights into the evolutionary relationships within the
571   rosids. *Molecular Phylogenetics and Evolution* 105:166–176. DOI:
572   10.1016/j.ympev.2016.06.007.
573 Zhou J, Cui Y, Chen X, Li Y, Xu Z, Duan B, Li Y, Song J, Yao H. 2018. Complete chloroplast genomes
574   of *Papaver rhoeas* and *Papaver orientale*: Molecular structures, comparative analysis, and
575   phylogenetic analysis. *Molecules* 23:437. DOI: 10.3390/molecules23020437.
576

Figure 1

Brunellia antioquensis (Cuatrec.) Cuatrec. and *B. trianae* Cuatrec.

Brunellia antioquensis (A) Habit. (B) twig with infructescences. (C) young leaflets and stipels. *Brunellia trianae* (D) twig with adult leaves and vegetative buds. (E) twig with infructescences. [(A)–(C) from Orozco et al. 4021 (COL), (D)–(E) from Orozco et al. 4019 (COL)]. Scale for (A) = 2 m, (B) = 2 cm, (C) = 1 cm, (D)–(E) = 2 cm. Photos by Clara I. Orozco.

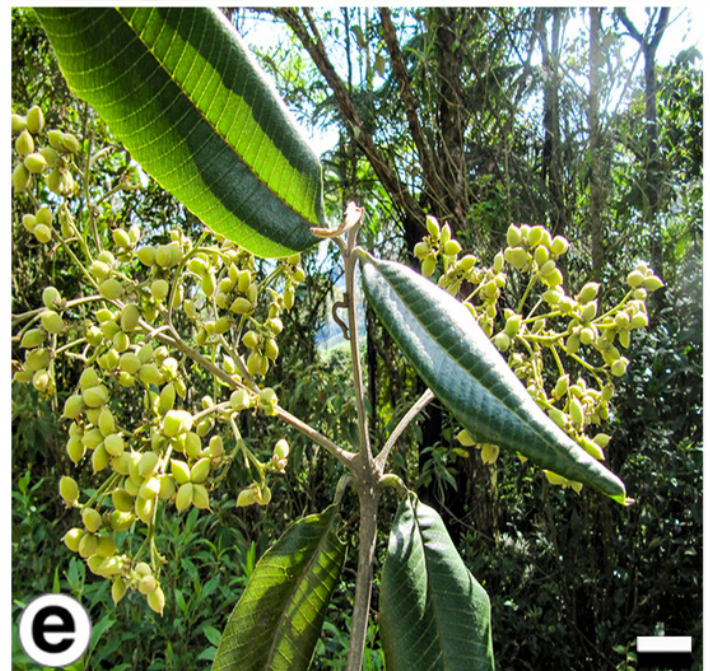



Figure 2

Gene map of the plastid genome of *Brunellia antioquensis* and *B. trianae*.

Genes placed inside of the outer circle are transcribed in counterclockwise direction whereas genes outside are transcribed in the clockwise direction. Colors refer to genes from different functional groups. The lighter grey area in the inner circle indicates AT content, while the darker grey area indicates GC content.

Brunellia antioquensis
plastid genome
157,685 bp

Brunellia trianae
plastid genome
157,775 bp

photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

Figure 3

Boundaries among different plastid regions.

(A) *Brunellia antioquensis*. (B) *B. trianae*. Apparent size of each region, gene, and distance in bp at ends (not to scale).

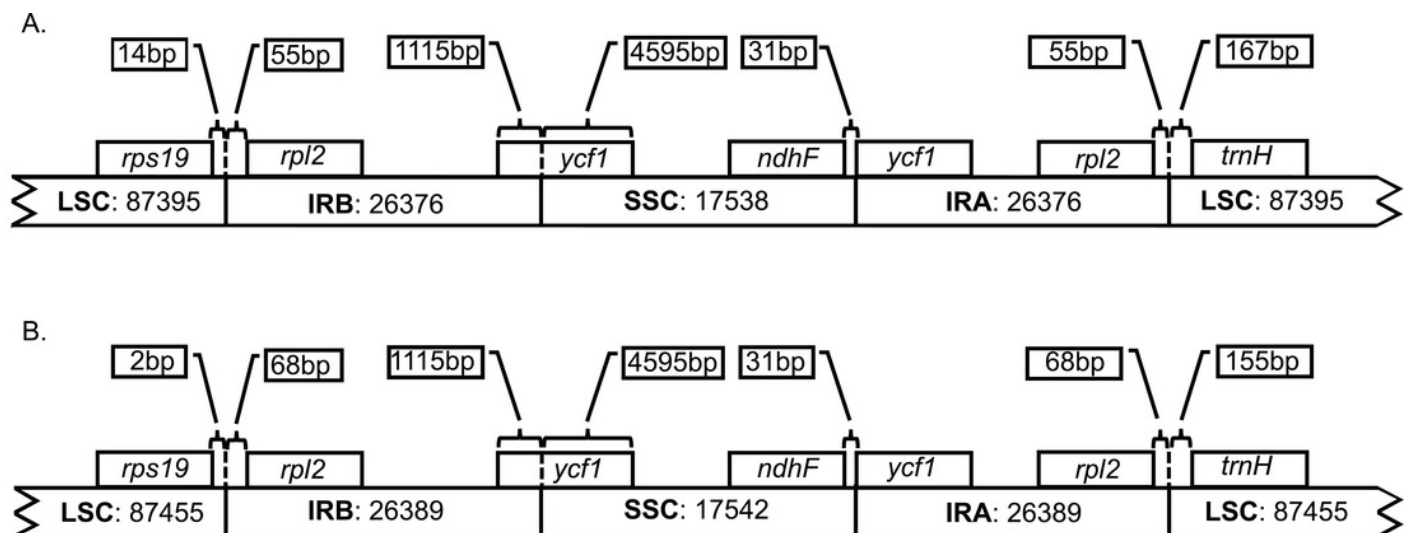



Figure 4

DNA polymorphism analysis of the plastid genome sequences of *Brunellia antioquensis* and *B. trianae.*

Nucleotide diversity (π) evaluated with DnaSP6 (settings: window length: 600 bp, step size: 200 bp).

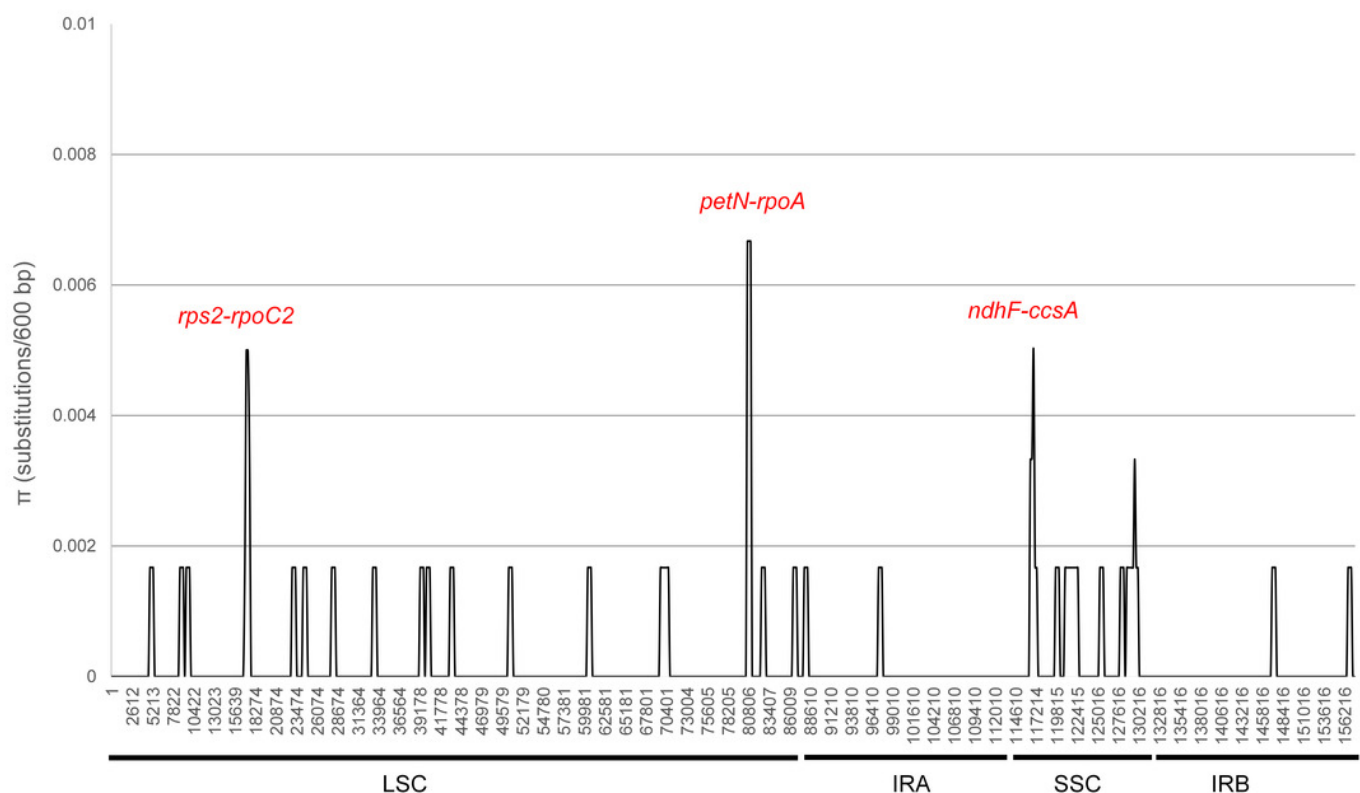
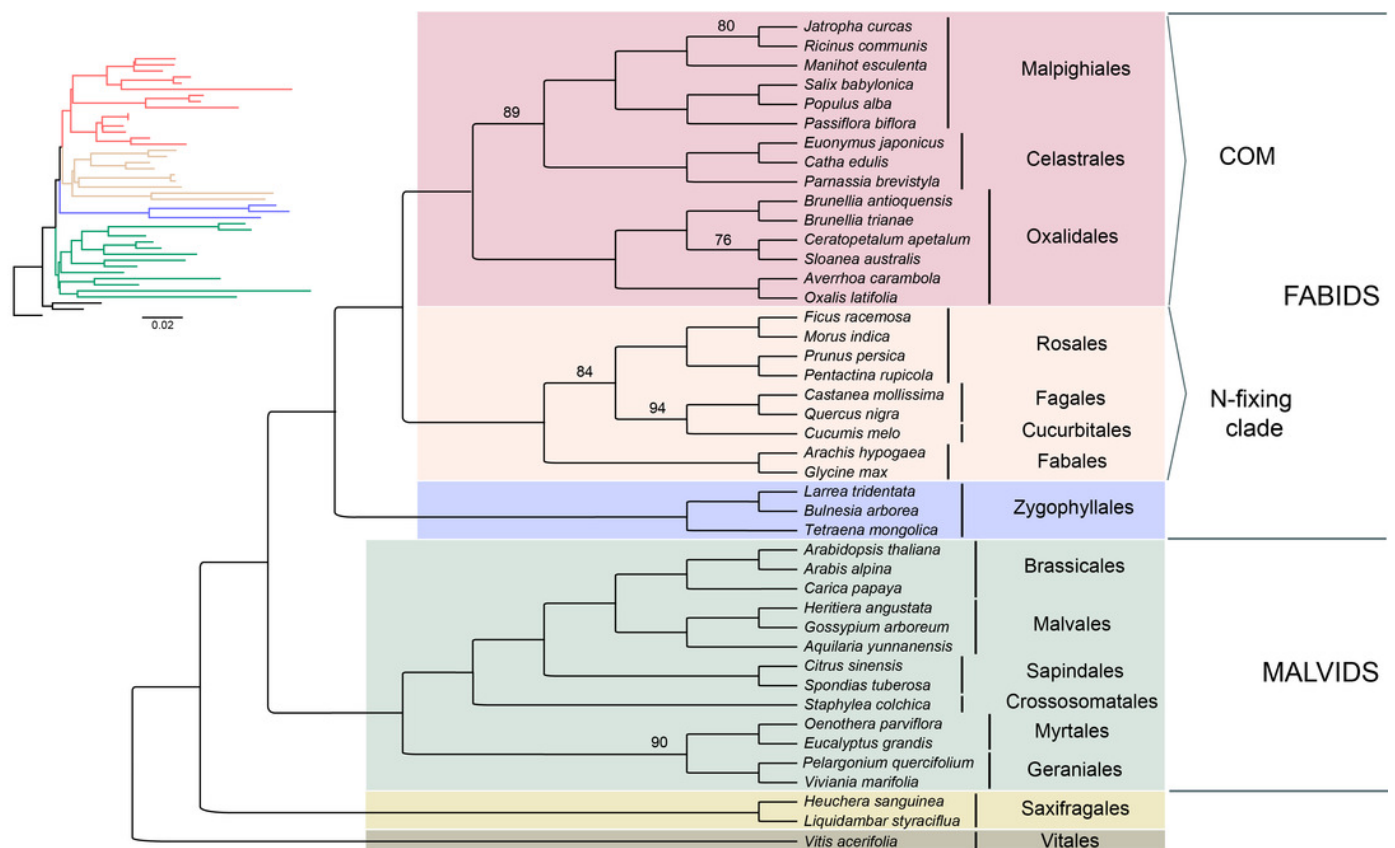



Figure 5

Phylogenetic relationships of 43 superrosid representatives.

Relationships inferred from the 75-plastid protein-coding regions using maximum likelihood (ML) analysis. Bootstrap values are 100% except where indicated.

Plastid genome characteristics of *Brunellia antioquensis* and *B. trianae.*

1 **Table 1. Plastid genome characteristics of *Brunellia antioquensis* and *B. trianae***

B. antioquensis	Size (bp)	Length (%)	A (%)	C (%)	G (%)	T (%)	C/G (%)	A/T (%)
LSC	87,395	55.42	31.90	17.9	17	33.3	34.85	65.15
IRB	26,376	16.73	28.7	20.6	22.2	28.6	42.75	57.25
SCC	17,538	11.12	34.4	15.0	16.2	34.4	31.22	68.78
IRA	26,376	16.73	28.6	22.2	20.6	28.7	42.75	57.25
Total	157,685	100	31.08	18.73	18.36	31.83	37.09	62.91
B. trianae								
LSC	87,455	55.43	31.91	17.86	16.98	33.25	34.84	65.16
IRB	26,389	16.73	28.7	20.6	22.2	28.6	42.75	57.25
SCC	17,542	11.11	34.4	15	16.2	34.4	31.23	68.77
IRA	26,389	16.73	28.6	22.2	20.64	28.7	42.75	57.25
Total	157,775	100	31.08	18.72	18.36	31.83	37.09	62.91