Research Article

Phylogenomic analysis of Picramnia, Alvaradoa, and Leitneria supports the independent Picramniales

Maria D. Logacheva¹ and Alexey B. Shipunov^{2*}

¹A.N. Belozersky Institute for Physico-Chemical Biology, Lomonosov Moscow State University, Moscow 119234, Russian Federation ²Minot State University, Minot, ND 58707, USA

*Author for correspondence. E-mail: dactylorhiza@gmail.com

Received 6 October 2016; Accepted 22 February 2017; Article first published online xx Month 2017

Abstract Picramniales is a recently established monotypic order that has not yet acquired a stable position on the phylogenetic tree of flowering plants. While *Picramnia* and *Alvaradoa* were transferred from Simaroubaceae (Sapindales) into Picramniaceae (Picramniales), *Leitneria* was transferred from the "hamamelids" into Simaroubaceae. Using next-generation sequencing, we obtained plastome sequences of *Picramnia* and *Alvaradoa* (partial) and *Leitneria* (complete). Protein coding and rRNA genes common for flowering plant plastomes were used for phylogenetic analysis. The resulting phylogenetic trees demonstrate a robust placement of Picramniales among the rosids as a sister group to Sapindales + Brassicales + Malvales. *Leitneria* is placed as a sister to *Citrus*. We also found that the plastome of *Leitneria* carries *ycf68* gene that is absent from most other rosids and *infA* pseudogene. **Key words:** *Alvaradoa*, *Leitneria*, *Picramnia*, picramniales, plastome, sapindales.

1 Introduction

The phylogenetic tree of flowering plants is now almost resolved at the level of families. Amazingly, most families first described by the end of the 18th century have proven to be robust, stable groups. However, molecular tools have also been used to find weak, unstable families. One example of this is Simaroubaceae. This group was long suspected of being "non-natural" (Fernando et al., 1995), and finally in 1995 two former Simaroubaceae genera, *Picramnia* Sw. and *Alvaradoa* Liebm., were placed in the separate family of Picramniaceae (Fernando & Quinn, 1995).

Picramniaceae is the rosid family of ca. 50 known species (Thomas, 2004). Not much is known about many aspects of their ecology, geography and morphology. The family is restricted to the American tropics; trees of Picramniaceae play an important role in Amazonian rain forests (Thomas, 2004). Picramniaceae representatives are rich in secondary metabolites; some species are known to be important for pharmaceutical purposes (as antimalaria drugs), but their chemistry is not well studied (Jacobs, 2003). Due to their small actinomorphic flowers, bitter bark and pinnately compound leaves, they were initially placed in Simaroubaceae. However, structure of flowers within Picramniaceae is seriously different (Fernando et al., 1995); the most striking is their syncarpous ovary with 2-3 ovules per locule while gynoecium of Simauroubaceae is typically almost apocarpous, with 1-2 ovules per carpel. It was thought that Picramniaceae has two genera, Picramnia and Alvaradoa but recently the third genus, Nothotalisia W.W. Thomas (Thomas, 2011) was discovered among herbarium samples of Talisia Aubl. (Sapindaceae). Only a few species of the family have been the subject of DNArelated research, and fewer than half of the species have been

included in morphological phylogenetic analyses (Pirani, 1990). As a result, one may consider Picramniaceae as a largely underresearched plant group. The phylogenetic position of Picramniaceae among the flowering plants is still not fully resolved, and they were frequently regarded as one of the "unplaced" (Stevens, 2001 onwards; Johansson, 2013 onwards). Recent large-scale analyses place Picramniaceae as sister to Sapindales (Moore et al., 2011; Sun et al., 2016), frequently, however, without or with low support. We believe that with the application of contemporary high-throughput methods to the study of chloroplast DNA, the position of Picramniaceae will be clarified. In addition, the more data from small Rosidae groups might help to resolve other problems like the deep incongruence between chloroplast and nuclear genomes in rosids (Sun et al., 2015).

The North American monotypic genus *Leitneria* Chapm. is now agreed to be a member of the Simaroubaceae (Fernando et al., 1995; Tobe, 2011), but in the past its distinctive morphology (expressed in the significant reduction of most reproductive organs) caused it to be accepted as a separate family or even order, and placed among the "higher hamamelids" (Jarvis, 1989). We decided that study of *Leitneria* cpDNA will add to the knowledge of this unusual plant, and also provide better ground for comparison between Simaroubaceae (and Sapindales in general) and Picramniaceae.

Sequences from plastid genomes were used for inferring plant phylogeny since its very beginning (Chase et al., 1993). Chloroplast genes and intergenic regions can usually be easily amplified in wide range of species, due to the multi-copy nature of plastid DNA, the presence of highly conserved regions used for primer design, and the absence of paralogs. However, the accumulation of sequence data revealed limitations to using plastid sequences for phylogenetic analysis. Some regions, including the most well studied gene *rbcL*, are not variable enough and do not provide wellresolved phylogenetic trees. Thus it is necessary to use additional markers (e.g., Soltis et al., 1998). The advent of next-generation sequencing (NGS) made it possible to analyze large genomic segments, including complete plastid genomes, at reasonable cost and in a short time frame (Straub et al., 2012; Stull et al., 2013). This (almost) closes the long-term discussion about which is better: more genes or more taxa (Rokas & Carroll, 2005). Moreover, when compared to Sanger techniques, many NGS protocols are less dependent on DNA quality and quantity; this makes them suitable for working with older herbarium samples (e.g., Staats et al., 2012) and in other cases where the extraction of high-quality DNA is complicated.

2 Material and Methods

Leaf tissue samples (approximately 50 g of fresh leaf tissue per sample) of *Leitneria floridana* Chapm., and *Picramnia pentandra* Sw. plus *Alvaradoa amorphoides* Liebm. were received from the curators of living collections at the Missouri Botanical Garden (St. Loius, MO) and Fairchild Botanical Garden (Miami, FL). DNA extraction (from 100 mg of dried leaf material) was performed using the CTAB-based method (Doyle & Doyle, 1987). 1 microgram of total DNA was taken for library preparation. Libraries were constructed according to the instructions given in the TruSeq DNA Sample Preparation Guide (Illumina Inc., San Diego, California, USA). Unfortunately, the final stage of size selection on agarose gel had to be omitted due to low amount of DNA. Thus, all libraries have a broad distribution of insert size, from 150 to

Table 1 Primers used for gap closing in the Leitneria plastome

450 bp. Libraries were sequenced using a Hiseg2000 instrument with a read length of 100 bp from each end of the fragment. Picramnia and Alvaradoa were sequenced using Miseq with a longer read length (300 bp from each end). For trimming and assembly we used CLC Genomics Workbench software. Resulting contigs were screened for contigs of plastid origin with the blastn program (Altschul et al., 1990), using the plastid genome sequence of Citrus sinensis (L.) Osbeck (Bausher et al., 2006) as a guery. Additional searches using plant mitochondrial genes as queries were performed in order to filter out mitochondrial contigs that contained regions similar to the plastid genome. Annotation of plastid contigs has been performed using the DOGMA tool (Wyman et al., 2004) along with manual checking and correction. Sequences of plastid genes of L. floridana, P. pentandra and A. amorphoides were added to the 78-gene data set based on a study by Ruhfel et al. (2014). This study encompasses green algae and plants; and since our study is focused on Picramniaceae that were previously shown to belong to rosids (Moore et al., 2011), we included in the analysis only species from "super-rosids" (rosids, Saxifragales and Vitales). This resulted in 64-taxon set; the sequences were aligned using MAFFT (Katoh et al., 2002) (alignment is available upon request). Tree reconstruction was performed using the ML algorithm by RAxML (Stamatakis, 2014) with two datasets: unmodified alignment, and the alignment where highly divergent regions were removed using GBlocks (Castresana, 2000). Vitis vinifera is used as an outgroup. For the completion of the Leitneria plastid genome, we designed primers located at the ends of the contigs (Table 1), amplified the corresponding regions, and sequenced them using Sanger (for amplicons \leq 1000 bp) or Miseq sequencing (for longer amplicons). For Miseq, libraries were prepared using Nextera XT kit (Illumina). The resulting sequence was deposited in the NCBI GenBank under accession number KT692940.

Name	Sequence	Length
Leit-C9-F	ATG GAT ACC GGG ATT CCT ATT G	22
Leit-C2-R	CGA TTT GGC AAT AAC GAA AGA AAG	24
Leit-C2-F	TTG GTT TGG TGG TGG AAT TTG	21
Leit-C8-R	TGG TGT TTA TCA GTG GTG GTA T	22
Leit-C8-F	GGT CTC TCT ACC AGG TGT TCT	21
Leit-C4-R	CGA TGG GAA TCC CGT TTA GTT	21
Leit-C4-F	GAT GGA ATC GTC CAT TTC GAT ATA A	25
Leit-C3-R	GAT TGG TCA TAC AAT CGT GCT T	22
Leit-C3-F	AGT AAG AAG TAA CCC GTG AAT CT	23
Leit-C4-IRa-R	GAT GGA ATC GTC CAT TTC GAT ATA A	25
Leit-C4-IRa-F	TTC TTC GCC GCC GTA GTA AAT AG	23
Leit-C9-R	TAG ACC TAG CTG CTG TTG AAG	21
Leit-C9-gf-F1	AGT AAG TAC GTA ACT CAA CGA GAA A	25
Leit-C9-gf-R1	AAG TTC TCA GTT AGG TGA AGG AAG	24
Leit-C9-gf-F2	CGG AGC CAC AAA GGA CTA TC	20
Leit-C9-gf-R2	TTT CGT GAG ACT CTG CTT GG	20
Leit-C4-gf-F1	ATC TTA GCA GGG TTA TTC CAT CTT	24
Leit-C4-gf-R1	CAC TAT ACT TCG ATT CTG CCC TT	23
Leit-C4-gf-F2	GAA GAG CGT CTC GAG ATT CAG	21
Leit-C4-gf-R2	ACC AAA GGA TGC GGT CAA TA	20

3 Results and Discussion

3.1 Plastid sequence assembly and annotation

Low coverage genome sequencing and assembly allowed us to identify most of the plastid genes used for phylogenetic analysis for all three species used in this study. The assembly was most successful for Leitneria, where we were able to completely reconstruct the plastome sequence (see below). As we pointed out in a previous study (Logacheva et al., 2014), a quality of the plastome assembly from the low coverage genome sequence data depends on the relative coverage of plastid and mitochondrial genomes. Mitochondrial genomes carry many sequences that not only originated from plastids, but also retain a high similarity to the existing plastid genome. If these sequences have similar coverage, one might expect the formation of chimerical contigs (under "soft" parameters for the assembly) or breakup of the contig (under "strict" parameters). Frequently, the coverage of the plastome is much higher than the coverage of the mitochondrial; this is likely the case of Leitneria floridana, where plastid contigs had 350–392x coverage in single copy regions whereas mitochondrial contigs had only 57–75x coverage. In contrast, *Alvaradoa* had 92–101x vs. 46–65x, and *Picramnia* 58–67x vs. 18–23x coverages of plastid vs. mitochondrial contigs, respectively. Consequently, the *Picramnia* and *Alvaradoa* assemblies contain higher numbers of shorter contigs that represent plastid genomes.

3.2 Phylogenetic analysis

Phylogenetic analysis of the resulting datasets yields trees that are mostly consistent with current views on rosid phylogeny (Ruhfel et al., 2014; Su et al., 2014; Chen et al., 2016; Sun et al., 2016). As expected, Sapindales representatives *Leitneria* and *Citrus* L. group together whereas *Picramnia* and *Alvaradoa* form a separate clade with 100% support (Fig. 1). This last clade (Picramniales) is sister first to Sapindales (*Leitneria* + *Citrus*) and then to Malvales + Brassicales. The trees inferred from unmodified alignment (Fig. 1) and from alignment trimmed with GBlocks (not shown) are congruent;



Fig. 1. ML phylogenetic tree of rosids (made with RAxML, alignment without using Blocks) based on the plastome data available to date and newly obtained sequences. *Leitneria, Alvaradoa* and *Picramnia* are highlighted in bold face.



Fig. 2. The map of the chloroplast genome of *Leitneria floridana*. Genes are color-coded according to their function. Location of gene inside and outside of the circle indicates the direction of its transcription. Grey shading inside the circle reflects GC-content. LSC, IRA, IRB and SSC denote genome regions (large single copy, inverted repeat A, inverted repeat B and small single copy).

the only topological difference between these two types of trees concerns the position of Moraceae relative to Fabaceae.

There is only a little morphological ground for the separation of *Picramnia* and *Alvaradoa* from the Sapindales (Stevens, 2001 onward). Presence of anthraquinones (Jacobs, 2003) and stamens opposite to petals (Thomas, 2011) distinguish this group, but other characters like dioecy, alternate compound estipulate leaves, and small flowers bearing inconspicuous styles conform closely with the family of Sapindales. However, our plastome data suggest that Picramniales represents a separate clade of angiosperms

(Fig. 1) that does not fall within Sapindales, represented in this study by Rutaceae and Simaroubaceae (*Citrus* and *Leitneria*). This is a clear indication of the independence of the group.

3.3 Plastid genome of Leitneria floridana

The plastome of *Leitneria floridana* (Fig. 2) is a circular molecule with the length of 158 763 base pairs. The length of inverted repeat (IR) region is 27 442 bp; small and large single copy regions are 18 189 and 85 690 bp, correspondingly. GC-content is 37.6%, the most GC-rich region is IR, where GC-content is 42.6%. The length and other characteristics are

within a typical range of variation for an angiosperm plastome. It is approximately 1kb less than that of its closest relative with a sequenced plastome, Citrus sinensis (Bausher et al., 2016). It is collinear to the Citrus plastome and contains almost the same set of genes. The Leitneria plastome carries an intact ycf68 ORF (ycf68 is a putatively functional gene located in the trnl-GAU intron). Notably, this gene is represented by a pseudogene in Citrus sinensis but present in C. aurantiifolia (Christm.) Swingle (Su et al., 2014), and the level of sequence similarity between the ycf68 of Leitneria floridana and C. aurantiifolia is high (99%). For another hypothetical gene, orf56, which is present in both C. aurantiifolia and C. sinensis, there is a region with high similarity within the Leitneria plastome but it does not seem to encode protein due to multiple frame shifts and the absence of a start codon. The genes common for Leitneria and Citrus show high level of sequence similarity, from 100% (e.g., most transfer RNA genes) to 84% (e.g., ycf1 gene).

Several rosid groups possess an unusual ordering of plastid genes along with an increased substitution rate, greater fraction of repeated sequences and highly rearranged gene order. The most notable example is Pelargonium L'Hér. ex Aiton and other Geraniaceae (Chumley et al., 2006; Guisinger et al., 2011). This is not the case for the studied species. Leitneria plastome has typical structure, length and nucleotide content; the fraction of repeats and substitution rate are similar to those of Citrus species. While we did not reconstruct a complete chloroplast genome for Picramniales, the longer contigs that have typical gene order strongly suggest that plastid genomes in this group do not have rearrangements. The newly characterized plastome of Leitneria carries ycf68 that is absent in most other rosids (including malvids). It also has a sequence with high similarity to intact infA (with P. pentandra — 76%, with Francoa sonchifolia Cav. — 79%) but has three internal stop codons and thus putatively is not functional. Both Citrus sinensis and C. aurantiifolia also carry regions with similarity to infA (though they are not annotated). The loss of infA, a gene that encodes a plastid translation factor, is frequently observed in different lineages of flowering plants (Millen et al., 2001) and is associated with the presence of a functional copy in the nuclear genome. It has been suggested, on the basis of currently available plastid sequences, that a particular infA loss/pseudogenization occured in a common ancestor of Sapindales, Brassicales and Malvales (Su et al., 2014). The characterization of Leitneria plastome supports this. Picramnia and Alvaradoa which fall outside this above clade, carry intact infA genes. Ycf68 is a pseudogene in most malvids, and the only current example of an apparently functional ycf68 was Citrus aurantiifolia (Su et al., 2014). We found that Leitneria floridana is a second example. Though there are some doubts on whether vcf68 is a real gene (Raubeson et al., 2007), the high level of conservation between Leitneria and Citrus aurantiifolia suggests its functionality.

Acknowledgements

We are very grateful to Rebecca A. Sucher, Living Collections Manager of Missouri Botanical Garden; Jason F. Lopez, Living Collection Manager of the Fairchild Tropical Botanic Garden, who helped us with samples of *Leitneria*, *Picramnia* and *Alvaradoa*; and Raymond W. Krueger, who helped to improve the manuscript. This research was supported by the Department of Biology (Minot State University) (AS), and the Russian Science Foundation, project #14-50-00029 (ML).

Authors' Contributions

ML carried out genome assembly, annotation, comparison and participated in writing, constructed and sequenced DNA libraries, AS collected and identified plants and participated in the manuscript preparation. Both authors approved the final manuscript. The authors declare they have no competing interests.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.
- Bausher MG, Singh ND, Lee SB, Jansen RK, Daniell H. 2006. The complete chloroplast genome sequence of *Citrus sinensis* (L.)
 Osbeck var. 'Ridge Pineapple': Organization and phylogenetic relationships to other angiosperms. BMC Plant Biology 6: 21.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17: 540–552.
- Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL, Kron KA. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528–580.
- Chen ZD, Yang T, Lin L, Lu LM, Li HL, Sun M, Liu B, Chen M, Niu YT, Ye JF, Cao ZY. 2016. Tree of life for the genera of Chinese vascular plants. *Journal of Systematics and Evolution* 54: 277–306.
- Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK. 2006. The complete chloroplast genome sequence of *Pelargonium* ×hortorum: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular Biology and Evolution* 23: 2175–2190.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11–15.
- Fernando ES, Gadek PA, Quinn CJ. 1995. Simaroubaceae, an artificial construct: evidence from *rbcL* sequence variation. *American Journal of Botany* 82: 92–103.
- Fernando ES, Quinn CJ. 1995. Picramniaceae, a new family, and a recircumscription of Simaroubaceae. *Taxon* 44: 177–181.
- Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Molecular Biology and Evolution* 28: 583–600.
- Jacobs H. 2003. Comparative phytochemistry of Picramnia and Alvaradoa, genera of the newly established family Picramniaceae. Biochemical Systematics and Ecology 31: 773–783.
- Jarvis CE. 1989. A review of the order Leitneriales. In: Crane PR, Blackmore S eds. Evolution, Systematics, and Fossil History of Hamamelidae 2, 'Higher' Hamamelidae. Systematics Association Special Volume 40B. Oxford: Clarendon Press. 189–192.
- Johansson JT. 2013 onward. The Phylogeny of Angiosperms [online]. Available from http://angio.bergianska.se [accessed 1 June 2015].

- Katoh K, Misawa K, Kuma KI, Miyata T. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059–3066.
- Logacheva MD, Schelkunov MI, Nuraliev MS, Samigullin TH, Penin AA. 2014. The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. *Genome Biology and Evolution* 6: 238–246.
- Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW, Calie PJ. 2001. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *The Plant Cell Online* 13: 645–658.
- Moore MJ, Hassan N, Gitzendanner MA, Bruenn RA, Croley M, Vandeventer A, Horn JW, Dhingra A, Brockington SF, Latvis M, Ramdial J, Alexandre R, Piedrahita A, Xi Z, Davis CC, Soltis PS, Soltis DE. 2011. Phylogenetic analysis of the plastid inverted repeat for 244 species: Insights into deeper-level angiosperm relationships from a long, slowly evolving sequence region. International Journal of Plant Sciences 172: 541–558.
- Pirani JR. 1990. As especies de Picramnia Sw. (Simaroubaceae) do Brasil: uma sinopse. Boletim de botanica, universidade de Sao Paulo 12: 115–180.
- Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, Jansen RK. 2007. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms Nuphar advena and Ranunculus macranthus. BMC Genomics 8: 174.
- Rokas A, Carroll SB. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution* 22: 1337–1344.
- Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014. From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology* 14: 23.
- Soltis DE, Soltis PS, Mort ME, Chase MW, Savolainen V, Hoot SB, Morton CM. 1998. Inferring complex phylogenies using parsimony: An empirical approach using three large DNA data sets for angiosperms. *Systematic Biology* 47: 32–42.

- Staats M, Richardson JE, Cowan RS, Bakker FT. 2012. How to open the treasure chest? Optimising DNA extraction from herbarium specimens. *PLoS ONE* 7: e43808.
- Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.
- Stevens PF. 2001 onward. Angiosperm Phylogeny Website. Version 12, July 2012 (and more or less continuously updated since) [online]. Available from http://www.mobot.org/MOBOT/research/APweb/ [accessed 1 June 2015].
- Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.
- Stull GW, Moore MJ, Mandala VS, Douglas NA, Kates HR, Qi X, Brockington SF, Soltis PS, Soltis DE, Gitzendanner MA. 2013. A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. Applications in Plant Sciences 31: 1.
- Su HJ, Hogenhout SA, Al-Sadi AM, Kuo CH. 2014. Complete chloroplast genome sequence of Omani Lime (*Citrus aurantiifolia*) and comparative analysis within the rosids. *PLoS ONE* 9: e113049.
- Sun M, Soltis DE, Soltis PS, Zhu X, Burleigh JG, Chen Z. 2015. Deep phylogenetic incongruence in the angiosperm clade Rosidae. Molecular Phylogenetics and Evolution 83: 156–166.
- Sun M, Naeem R, Su JX, Cao ZY, Burleigh JG, Soltis PS, Soltis DE, Chen ZD. 2016. Phylogeny of the Rosidae: A dense taxon sampling analysis. Journal of Systematics and Evolution 54: 363–391.
- Thomas WW. 2004. Picramniaceae. In: Smith N, Mori SA, Henderson A, Stevenson DW, Heald SV eds. *Flowering Plants of the Neotropics*. Princeton: Princeton University Press. 294–297.
- Thomas WW. 2011. Nothotalisia, a new genus of Picramniaceae from tropical America. Brittonia 63: 51–61.
- Tobe H. 2011. Embryological evidence supports the transfer of *Leitneria floridana* to the family Simaroubaceae. *Annals of the Missouri Botanical Garden* 98: 277–293.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.