Genome-scale transfer of mitochondrial DNA from legume hosts to the holoparasite *Lophophytum mirabile* (Balanophoraceae)

M. Virginia Sanchez-Puerta[a,b *]
Alejandro Edera[a]
Carolina L. Gandini[a]
Anna V. Williams[c]
Kate Howell[e]
Paul G. Nevill[c,d]
Ian Small[e]

[a]IBAM, Universidad Nacional de Cuyo, CONICET, Facultad de Ciencias Agrarias, Almirante Brown 500, M5528AHB, Chacras de Coria, Argentina.
[b]Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo, 5500, Mendoza, Argentina.
[c]School of Plant Biology, The University of Western Australia (M082), 35 Stirling Highway, 6009 Crawley, Australia.
[d]School of Molecular and Life Sciences, Curtin University, Kent Street, Bentley, Perth, Western Australia 6102, Australia.
[e]ARC Centre of Excellence in Plant Energy Biology, School of Molecular Sciences, The University of Western Australia (M316), 35 Stirling Highway, 6009 Crawley, Australia.

*Author for correspondence:
M. Virginia Sanchez-Puerta
Email: mvsanchezpuerta@fca.uncu.edu.ar
IBAM, Universidad Nacional de Cuyo, CONICET, Facultad de Ciencias Agrarias, Almirante Brown 500, M5528AHB, Chacras de Coria, Argentina.
Phone +54(261) 4135000 ext. 1307.

ABSTRACT

Angiosperm mitochondrial horizontal gene transfer (HGT) has been widely reported during the past decades. With a few exceptions, foreign sequences are mitochondrial genes or intronic regions from other plants, indicating that HGT has played a major role in shaping mitochondrial genome evolution. Host-parasite relationships are a valuable system to study this phenomenon due to the high frequency of HGT. In particular, the interaction between mimosoid legumes and holoparasites of the genus *Lophophytum* represents an outstanding opportunity to discern HGT events. The mitochondrial genome of the holoparasite *L. mirabile* has remarkable properties, the most extraordinary of which is the presence of 34 out of 43 mitochondrial protein genes acquired from its legume host, with the stunning replacement of up to 26 native homologs. However, the origin of the intergenic sequences that represent the majority (>90%) of the *L. mirabile* mtDNA remains largely unknown. The lack of mitochondrial sequences available from the donor angiosperm lineage (mimosoid legumes) precluded a large-scale evolutionary study. We sequenced and assembled the mitochondrial genome of the mimosoid *Acacia ligulata* and performed genome wide comparisons with *L. mirabile.* The *A. ligulata* mitochondrial genome is almost 700 kb in size, encoding 60 genes. About 60% of the *L. mirabile* mtDNA had greatest affinity to members of the family Fabaceae (~49% to mimosoids in particular) with an average sequence identity of ~96%, including genes but mostly intergenic regions. These findings strengthen the mitochondrial fusion compatibility model for angiosperm mitochondrion-to-mitochondrion HGT.

KEYWORDS: *Acacia*, *Lophophytum*, holoparasite, HGT, evolution, mtDNA.

1. INTRODUCTION

Horizontal gene transfer (HGT), the transmission of genetic material between non-mating organisms, has been increasingly reported among angiosperms during the last two decades (Bergthorsson *et al.* 2003; Davis & Wurdack 2004; Kim *et al.* 2014; Mower *et al.* 2004; Rice *et al.* 2013; Sanchez-Puerta *et al.* 2017; Xi *et al.* 2013; Yang *et al.* 2016). This phenomenon affects particularly plant mitochondrial genomes, which acquire, almost exclusively, mitochondrial sequences from other plants (Barkman *et al.* 2007; Bergthorsson *et al.* 2003; Cho *et al.* 1998; Mower *et al.* 2010; Rice *et al.* 2013; Sanchez-Puerta *et al.* 2008; Sanchez-Puerta *et al.* 2017). Despite the elevated frequency of HGT among flowering plants, the dynamics and mechanisms involved in plant mitochondrial HGT remains largely unknown. A mitochondrial-fusion compatibility model has been proposed, in which HGT occurs by capture of entire mitochondria from donor plants, followed by fusion of native and foreign mitochondria and mitochondrial intergenomic recombination (Rice *et al.* 2013). This model states that only mitochondrial sequences are transferred into the recipient plant mitochondria, that only species from the green lineage are compatible donors because they share a similar mitochondrial fusion mechanism with the recipient plant, and that plant mitochondrial genomes will recombine to form a chimeric mtDNA (Rice *et al.* 2013). A corollary to the model is that foreign plastid or nuclear sequences are first acquired by the donor mitochondria by intracellular gene transfer from the plastid and nuclear genomes and later horizontally-transferred to the recipient mitochondria by mitochondrion-to-mitochondrion (mt-to-mt) HGT (Gandini & Sanchez-Puerta 2017; Rice *et al.* 2013). Several natural mechanisms that enable the transfer of whole mitochondria have been proposed, including direct transmission involving tissue grafts, illegitimate pollination, or host-parasite interactions, and indirect transmission mediated by vectors such as viruses, bacteria, insects, and fungi (Keeling & Palmer 2008; Mower *et al.* 2004; Sanchez-Puerta *et al.* 2008; Stegemann & Bock 2009).

Parasitic plants form a physical connection with the vascular system of their host plant, known as a haustorium, to conduct water and nutrients (and sometimes sugars and amino acids) from hosts to parasites. The intimate association between parasites and their hosts facilitates the exchange of genetic material, making parasites particularly susceptible to HGT (Davis & Wurdack 2004; Kim *et al.* 2014; Sanchez-Puerta *et al.* 2017; Xi *et al.* 2013; Yang *et al.* 2016), although hosts can also acquire foreign genes from their parasites (Mower *et al.* 2004; Mower *et al.* 2010). Indeed, the content of the mitochondrial genomes of parasitic plants can be significantly altered by HGT (Sanchez-Puerta *et al.* 2017; Xi *et al.* 2013).

Legumes are often parasitized by other angiosperms and have been considered attractive hosts given their high nitrogen content resulting from N_2 fixation by the bacterial symbionts (Press & Phoenix 2005). The Fabaceae is the third-largest angiosperm family (ca. 19,500 species) and it was traditionally subdivided into three subfamilies: Papilionoideae, Mimosoideae, and Caesalpinioideae (LPWG 2017). Today, the mimosoids are considered part of the subfamily Caesalpinioideae (LPWG 2017). The Papilionoideae is the largest subfamily and has been the best studied because it includes many agriculturally important species, such as chickpea (*Cicer arietinum*), soybean (*Glycine max*), groundnut (*Arachis hypogaea*), lentil (L*ens culinaris*), alfalfa (*Medicago sativa*), the common bean (*Phaseolus vulgaris*), and mung bean (*Vigna radiata*). The

mimosoids have a pantropical distribution and include species-rich genera such as *Mimosa* and *Acacia* (LPWG 2017). Members of the Fabaceae have been described as donors of nuclear (Kado & Innan 2018; Vogel *et al.* 2018; Yang *et al.* 2016; Zhang *et al.* 2013) and mitochondrial (Barkman *et al.* 2007; Sanchez-Puerta *et al.* 2017) sequences horizontally transferred to different parasitic plant lineages.

Recently, a massive transfer of mitochondrial genes from a mimosoid legume host to the holoparasite *Lophophytum mirabile* (Balanophoraceae) was reported (Sanchez-Puerta *et al.* 2017). This study revealed the unparalleled acquisition of host mitochondrial genes, representing 80% of the protein-coding gene content of the parasite mitochondria (Sanchez-Puerta *et al.* 2017). The presence of foreign DNA in *L. mirabile* mtDNA is the most extensive of any eudicot examined so far and it follows the early-diverging angiosperm *Amborella trichopoda*, which carries up to six genome equivalents of foreign mitochondrial sequences (Rice *et al.* 2013). The parasitic relationship between *L. mirabile* and its mimosoid host represents an extraordinary opportunity to investigate HGT in plants because *Lophophytum* spp. have a narrow host range, the hosts are distantly related to the parasite, and the *L. mirabile* mtDNA has been fully sequenced (Sanchez-Puerta *et al.* 2017). However, the lack of mitochondrial genome sequences from mimosoid legumes precluded an in depth analysis. Complete mitochondrial genomes have been sequenced from nine legume species confined to the subfamily Papilionoideae (*Ammopiptanthus mongolicus*, *Glycine max, Lotus japonicus, Medicago truncatula, Millettia pinnata, Sophora japonica, Vicia faba, Vigna angularis,* and *V. radiata*) and recently, three species (*Senna occidentalis*, *S. tora,* and *Leucaena trichandra*) of the subfamily Caesalpinioideae (Alverson *et al.* 2011; Bi *et al.* 2016; Chang *et al.* 2013; Kazakoff *et al.* 2012; Kovar *et al.* 2018; Naito *et al.* 2013; Negruk 2013; Shi *et al.* 2018; Yu *et al.* 2018).

Based on previous knowledge (Sanchez-Puerta *et al.* 2017), we decided to gather sequence information from a mimosoid (*Acacia ligulata*) closely related to the putative ancestral host of holoparasites of the genus *Lophophytum* to uncover the mechanistic details of host to parasite HGT. We carefully compared the mitochondrial genome of the holoparasite *L. mirabile* to that of legumes and other angiosperms in order to address the following questions: (1) How much of the mimosoid mtDNAs is shared by other legumes and by *L. mirabile*? (2) What fraction of the *L. mirabile* mtDNA is derived from the mimosoid host? (3) How long/fragmented are the foreign sequences and how similar to those of the donor lineage? (4) Is there evidence for the mitochondrial-fusion compatibility model? (5) Are there chimeric genes created by homologous recombination between host and parasite alleles? (6) Are there foreign plastid or nuclear-derived regions in the *L. mirabile* mtDNA? (7) Were the foreign plastid or nuclear-derived regions acquired by mt-to-mt HGT or directly from the donor plastid or nuclear genomes?

2. MATERIALS AND METHODS

2.1 DNA extraction and sequencing

Fresh phyllodes from a seedling of *Acacia ligulata* Benth. (Fabaceae) were collected from the Western Australian Botanic Garden in Kings Park, Perth. Total genomic DNA was extracted using a CTAB protocol and

fragmented with a Covaris S220 focused ultrasonicator. The DNA from *A. ligulata* was used to construct a 400-bp paired-end library using a Trueseq DNA Sample Preparation Kit (Illumina, San Diego, USA), following the manufacturer's directions. The library was clustered on a Rapid Flow Cell v2 (Illumina), using the HiSeq Rapid PE Cluster Kit v2 (Illumina) and on instrument cluster generation on the HiSeq1500 platform, and sequenced using the Hiseq Rapid SBS Kitv2, generating over 4.7 million paired-end reads that passed filter. A PhiX library (Illumina) was spiked in at 1% as a control to provide real time analysis metrics.

2.2 Mitochondrial genome assembly and validation

Standard Illumina adaptors were removed from paired-end reads using Trimmomatic v. 0.35 (Bolger *et al.* 2014). The *A. ligulata* 150 bp paired-end reads were assembled on the Mason large-memory computer cluster at Indiana University-Bloomington (USA). To perform *de novo* assembly of the mitochondrial genome of *A. ligulata* we used Velvet 1.2.08 (Zerbino & Birney 2008) without scaffolding and with hash lengths of 41, 73, 85, 97, and 111 and a coverage cut-off of 3. The best run (hash length 85) assembled 50 *de novo* contigs larger than 2 kb, with N50 of 45,897 bp and a maximum contig size of 212,099 bp. Taking advantage of typical differences in read depths among cellular compartments (Straub *et al.* 2012), 10 putative mitochondrial contigs with total read depth >50 and <100 were further analyzed. Manual editing, joining, and closing of the mitochondrial contigs was done based on consistent paired-end reads visualized in Consed v.29 (Gordon & Green 2013).

Raw sequence data are available from the NCBI Bioproject ID PRJNA505150. The annotated mitochondrial genome was deposited in the GenBank data libraries under accession number MH933866.

2.3 Genome annotation

The mtDNA was annotated using Mitofy (Alverson *et al.* 2010), BLAST (Camacho *et al.* 2009), and the tRNAscan-SE algorithm (Lowe & Eddy 1997). The mitochondrial or plastid origin of the tRNAs was assessed by BLAST searches. In addition, gene alignments of *A. ligulata* and diverse angiosperms were constructed to assess gene boundaries and the location of splicing sites. The map of the *A. ligulata* mitochondrial genome was generated using OGDraw software (Lohse *et al.* 2007). Dispersed repeats with >90% sequence identity were identified in Consed v.29 (Gordon & Green 2013) using crossmatch. Plastid-derived mitochondrial sequences (MTPTs) were detected by BLAST searches against the *Acacia ligulata* cpDNA (NC_026134.2). The presence of paired-end reads with one mate mapping the flanking mitochondrial sequences and the other mapping the MTPT gave support to the assembly of the plastid-derived regions in the *A. ligulata* mtDNA.

2.4 Comparative and evolutionary analyses

Pairwise BLASTn analyses of the mitochondrial genomes of *A. ligulata*, diverse legumes, and *Lophophytum mirabile* were performed. Blast hits were visualized in dot plots for each pair of species and were drawn with Gepard v.1.40 (Krumsiek *et al.* 2007).

To unveil the origin of the *L. mirabile* mtDNA, we blasted each circular-mapping chromosome of *L. mirabile* (Genbank accession numbers KU992322-KU992380, KX792461) against a local database including all complete mitochondrial genomes of the green lineage available from the NCBI Organelle Genome Database as of October 2018 using the BLASTN v.2.4.0+ algorithm optimized for somewhat similar sequences (blastn) (Camacho *et al.* 2009). We also included the mtDNA of *A. ligulata* reported in this study. Only hits greater than 250 bp were considered. Because the origin of almost all mitochondrial genes from *L. mirabile* mtDNA had been previously analyzed (Sanchez-Puerta *et al.* 2017), we focused mainly on the intergenic regions. BLAST hits were plotted using the Sushi R package v.1.16.0. An *L. mirabile* mitochondrial region was considered derived from the legume host (as a result of HGT) when BLAST hits included only members of the family Fabaceae or when hits <350 bp showed a higher identity to legumes than to other angiosperms. In other cases, phylogenetic analyses were performed if the hits to legumes and to other angiosperms were of similar length. The alignments were generated based on each BLAST result and differed in taxon sampling. Such regions were considered foreign if the trees showed a relationship of *L. mirabile* with the family Fabaceae supported by bootstrap values >65%. Maximum Likelihood (ML) phylogenetic analyses were performed with RAxML v.8.2.11 (Stamatakis 2014) under the General Time Reversible model with parameters for invariable sites and gamma-distributed rate heterogeneity (GTR+Gamma with four rate categories). A hundred rapid bootstrap replicates were done under the same model of evolution using RAxML .

3 RESULTS AND DISCUSSION

3.1 The *Acacia ligulata* mitochondrial genome and a comparison to other angiosperms

About 14% of reads were assembled into one mitochondrial contig of 698,138 bp with an average read depth of 70x (Figures 1 and S1). The only exceptions represent plastid-derived mitochondrial sequences (MTPTs) that show spikes of the read-depth due to the mismapping of reads that were derived from the chloroplast genome (Figure S1). Based on read-depth and paired-end read information, we inferred that the *A. ligulata* mitochondrial genome could exist as alternative structures (Figure 1). It can be mapped as two subgenomic circular molecules of 686,972 bp and 683,146 bp that differ in a small region of <15 kb, or as head-to-tail concatemers (Figure 1). These alternative conformations are common in plant mitochondria due to recombinationally active repeats (Sloan 2013).

The *A. ligulata* mitochondrial genome is the second largest among the Fabaceae, where previously sequenced genomes ranged in size from 272 kb in *Medicago truncatula* (Bi *et al.* 2016) to 729 kb in *Leucaena trichandra (Kovar et al. 2018)*. In total, we found 3 large (>1 kb) and 5 intermediate (250-1,000 bp) repeats with >90% identity in the *A. ligulata* mtDNA (Table 1). It has a GC content of 45.06% and contains 60 unique genes, including 37 protein, 3 rRNA, and 20 tRNA genes (Tables 1 and S1). Of the 20 mitochondrial-encoded tRNAs, 11 and 9 are of mitochondrial and plastid origin, respectively. At least 3 tRNA genes, for the amino acids alanine, arginine, and threonine, are absent from the *A. ligulata* mitochondrial genome and are presumably imported from the nucleus. In agreement with other legumes, complete reading frames of the genes *rps2*, *rps7*,

rps11, and *rps13* are absent. In contrast, the genes *rpl2*, *rpl10*, *rps19*, and *sdh4* are full-length in the *A. ligulata* mtDNA, as well as in *Leucaena trichandra* (Kovar *et al.* 2018), but not in the species of the subfamily Papilionoideae that have been examined so far (Shi *et al.* 2018). The *A. ligulata* genome contains 19 cis-splicing (in the genes *ccmFc*, *nad1, nad2, nad4, nad5, nad7, cox2, rpl2, rps3,* and *rps10*) and 5 trans-splicing (in the genes *nad1, nad2*, and *nad5*) group II introns. The *A. ligulata* mtDNA also harbors a complete ORF with similarity to a viral DNA polymerase (*dpo*). Plasmid-derived sequences with similarity to DNA and/or RNA polymerases are frequent in angiosperm mitochondrial genomes (Warren *et al.* 2016). The *Acacia dpo* has only 55% similarity at the protein level to sequences in other angiosperm mtDNAs, in agreement with a study that reported greater sequence divergence between plasmid-derived sequences than between other mitochondrial genes (Warren *et al.* 2016). The *A. ligulata* mtDNA also contains 22 MTPTs encompassing 3.2% of the genome (Table 1), a similar value to that of other legume mitochondrial genomes *(Gandini & Sanchez-Puerta 2017; Sloan & Wu 2014)*. Of these, two were previously described as foreign because they were closely related to sequences in Piperales and Salicales, respectively (Gandini & Sanchez-Puerta 2017).

Pairwise nucleotide BLAST analyses of the mitochondrial genomes of *A. ligulata* and diverse legumes revealed limited synteny and a small proportion of homologous sequences. The mimosoids *A. ligulata* and *Leucaena trichandra* share 298 dispersed sequences greater than 250 bp with an average length and identity of 1,210 bp and 94.3%, respectively. Overall, 58% of the *Acacia* mtDNA has homology to *L. trichandra* mtDNA (Figure S2a). About 42% and 35% of the *A. ligulata* mtDNA has similarity to caesalpinioid and papilionoid legumes, respectively (Figure S2b,c). These findings agree with observations done among papilionoid legumes (Shi *et al.* 2018) and among other comparably related angiosperm lineages (Liu *et al.* 2013). The amount of shared sequences is generally larger between species that are more closely related (Liu *et al.* 2013).

Noticeably, a pairwise BLASTn search against the mimosoid root holoparasite *Lophophytum mirabile* mtDNA revealed that ~47% of the mimosoids *A. ligulata* and *L. trichandra* mtDNAs has similarity to *L. mirabile* mtDNA, including 300 regions larger than 250 bp (average hit length ~1,150 bp) with an average identity of ~94% (Figure S2d, e). A comparison between *Acacia* and another asterid indicated that 34% of *A. ligulata* mtDNA shows similarity *Nicotiana tabacum*, including 133 homologous regions larger than 250 bp with an average identity of 91%. The high similarity and elevated proportion of shared sequences between distantly related angiosperms, such as those from the families Fabaceae and Balanophoraceae, is highly unexpected. In a recent study, an extraordinary amount of shared sequences between *Lophophytum* and *L. trichandra* mtDNAs was reported, in comparison to the amount of shared sequences between *L. trichandra* and papilionoid legumes (Kovar *et al.* 2018).

3.2 Massive horizontal transfer of intergenic regions from a mimosoid donor to the holoparasite *Lophophytum mirabile*

Given the availability of mtDNAs from mimosoids, we evaluated the incidence of the horizontal transfer of mitochondrial sequences, particularly intergenic regions, in the parasitic relationship between mimosoid

hosts and holoparasites of the family Balanophoraceae. We performed BLASTn similarity searches of the holoparasite *L. mirabile* mtDNA against all available mitochondrial genomes in Genbank, including the *A. ligulata* mtDNA (Figure S3). BLAST hits were grouped in those from mimosoid legumes (*A. ligulata* and *L. trichandra*), from other legumes (*Ammopiptanthus mongolicus*, *Glycine max, Lotus japonicus, Medicago truncatula, Millettia pinnata, Senna occidentalis, S. tora, Sophora japonica, Vicia faba, Vigna angularis,* and *V. radiata*), or from other angiosperms (Figure S3). Depending on the length, sequence identity, and taxonomy of the BLAST hits, as well as the phylogenetic affiliation of the query, the mitochondrial regions of *L. mirabile* were considered foreign or putatively native (Figure 2 and S3). For those mitochondrial regions that found multiple hits of similar length, we conducted phylogenetic analyses to assess the evolutionary relationships of *L. mirabile* sequences. When an intergenic region of *L. mirabile* mtDNA was closely related to the legume clade with bootstrap support >65%, a horizontal transfer from the host to the parasite was inferred (Figure S4). The mitochondrial genes of *L. mirabile* were considered native or foreign based on previous analyses (Sanchez-Puerta *et al.* 2017).

Overall, we found that 49% and ~10.1% of the *L. mirabile* mitochondrial genome showed greatest similarity or evolutionary affinity to mimosoid and to other legume mtDNAs, respectively (Figure 2a). In total, we found 307 individual regions >250 bp distributed across all mitochondrial chromosomes with strongest affinity to mimosoid mtDNAs (Figure 3 and S3). Those BLAST hits had an average length of 1,312 bp and an average sequence identity of 96.87%. Less than 8% of those hits involve genic regions. The incredibly high identity of *L. mirabile* and mimosoid sequences agrees with the hypothesis that they were transferred by HGT from a mimosoid host to *Lophophytum* relatively recently. These relatively short foreign sequences identified in the *L. mirabile* mtDNA may belong to longer tracts of foreign DNA that cannot be recognized at the moment because neither *Acacia* nor *Leucaena* is the ancestral mimosoid donor. The largest continuous foreign tract identified in the *L. mirabile* mtDNA is a non-coding region of 6,992 bp transferred from mimosoids (Figure S3 chr03 16-23 kb). We predict that much longer foreign tracts in *L. mirabile* will be recognized with the additional sampling of mimosoid mitochondria. When we analyzed the arrangement of the donor sequences in *Acacia* or *Leucaena*, we found them dispersed along the their mitochondrial genomes. The foreign sequences transferred from the mimosoid host account for 58% and 55% of the mtDNA of *Acacia* and *Leucaena*, respectively.

In addition, we found 142 regions related to other legume mtDNAs. Those BLAST hits had an average length of 584 bp and an average sequence identity of 91% (Figure 3 and S3). In most of these cases, no similarity to mimosoid mtDNAs was found. The lower identity may reflect the fact that the mtDNA of the mimosoid donor containing these homologous sequences is not available for comparison. Alternatively, they could be cases of ancient HGT events from other legume donors.

About 13.6% of the *L. mirabile* mtDNA (22% are coding regions) had greatest similarity to angiosperm mitochondrial genomes other than Fabaceae (Figure 2). Because there is very limited availability of mitochondrial sequences from close relatives to *Lophophytum* (only two mitochondrial genomes from

Santalales (Skippington *et al.* 2015, 2017) and none from other Balanophoraceae) to assess the origin of these regions, we conservatively considered them as putatively native (Figure 2). A total of 208 regions were similar to other angiosperm mitochondrial genomes, with an average length and identity of 539 bp and 86.46% (Figure 3 and S3). The scarcity of comparative data and the high substitution rate in the mitochondrial genomes of the Balanophoraceae (Su *et al.* 2015) may explain the lower similarity detected in the putatively native regions.

Finally, 27.3% of the *L. mirabile* mtDNA lacks detectable similarity to any mitochondrial genome in GenBank (Figure 2). Angiosperm mtDNAs consist mostly of intergenic regions, these turn over rapidly (Mower *et al.* 2012) and there is a very small number of legume mitochondrial genomes available (nine out of 19,500 species of legumes, and only two of 3300 described mimosoids; (LPWG 2017)). Hence, our ability to recognize the origin of the non-coding sequences is limited. The sequencing of each additional mimosoid genome will improve our estimate of foreign DNA in *Lophophytum*. Indeed, in an earlier study (Sanchez-Puerta *et al.* 2017), no mimosoid mtDNAs were available for comparison, preventing the recognition of a large fraction of non-coding foreign sequences in *Lophophytum*. The availability of mimosoid mtDNAs increased the estimation of foreign sequences in *L. mirabile* mtDNA by an order of magnitude. These results highlight the impact of sequencing a close relative to the host plant of the parasite *L. mirabile* to better assess the extent and dynamics of the massive HGT between the holoparasite *Lophophytum mirabile* and its mimosoid host.

3.3 Impact of HGT in *L. mirabile* mitochondrial chromosomes

The *L. mirabile* mtDNA consists of 60 circular-mapping chromosomes (available in Genbank) that can be rearranged to form 54 chromosomes by homologous recombination across large repeats (Sanchez-Puerta *et al.* 2017). Mitochondrial genes cover <8% of the *L. mirabile* mtDNA and almost half of the chromosomes bear no intact genes and are possible non-coding molecules (Figure S3). Detailed analysis of each mitochondrial chromosome reveals a great disparity in the relative content of foreign sequences (Figure 2b and 3). In five cases, more than 80% of *L. mirabile* mitochondrial chromosomes have been likely acquired by HGT from a legume (asterisks in Figure 2b). This includes three chromosomes (chr30, 35, and 44) in which the foreign sequences represent >99% comprising mainly non-coding regions (Figure 2b). These findings raise the possibility that whole chromosomes could have been horizontally transferred from the host plant and have acquired regulatory regions to replicate in the recipient mitochondria. Except for *nad6* and *trnfM* in chr30, *trnQ* in chr35, and *atp1* in chr41, those putatively foreign chromosomes (asterisks in Figure 2b) bear no intact known genes (Figure S3). On the opposite end, chromosome 23 contains less than 10% of foreign sequences (Figure 2b and S3).

3.4 Foreign, chimeric, and putatively native genes in *L. mirabile* mtDNA

In light of the new sequence information from the mtDNAs of mimosoid legumes, we re-examined the origin of each gene encoded in the mtDNA of *L. mirabile*. The analysis consisted of searching for longer tracts of similarity to the *Acacia* mtDNA including the flanking regions of each gene.

We confirmed the origin of all foreign genes previously identified (Sanchez-Puerta *et al.* 2017). Putatively foreign genes for which the AU tests were not significant (e.g. *atp1, ccmFC, ccmFN, cob, nad6, rps3,* among others) or the AU test could not be performed due to lack of comparative data (*ccmC* and *nad2*) (Sanchez-Puerta *et al.* 2017) were analyzed here in detail. Genomic comparisons and phylogenetic analyses of surrounding sequences provided additional evidence for most of those genes to confirm that they were indeed acquired from a mimosoid donor (Table S2, Figure S3, S4).

Furthermore, we were able to identify the origin of short genes, such as tRNAs, and the presence of chimeric genes, which could not be thoroughly evaluated before. In a previous study, *atp6* was recognized as chimeric based on the results of a recombination test, in addition to *nad5* with native and foreign gene regions (Sanchez-Puerta *et al.* 2017). Here, we identified another chimeric gene (*rrnL*) formed by homologous recombination between host and parasite sequences (Figure S5). Phylogenetic analysis of the 5' end of the gene *rrnL* found no clear affiliations for *L. mirabile*, while the tree based on the 3' end showed a clade uniting legumes and *L. mirabile* with moderate support (BS=71%). The flanking sequence upstream of *rrnL* found no similarity among the Fabaceae mtDNA, while the sequence downstream found regions of similarity almost exclusively with legume mtDNAs (Sanchez-Puerta *et al.* 2017).

Finally, nine short genes can now be identified as foreign: *rrn5*, *nad5* (exon3) and seven tRNA-encoding genes (Table S2), because they are embedded within long tracts with affinity to *Acacia* mtDNA (Figure S6). Overall, out of the 56 full-length genes encoded in the *L. mirabile* mtDNA, 42 (75%) are foreign, three (5.36%) are chimeric, and 11 (19.64%) are putatively native (Table S2). In all cases, the putative donor was identified as a member of the mimosoid clade.

3.5 Foreign nuclear and plastid-derived regions in *L. mirabile* mtDNA were acquired from the legume donor via mitochondrial-to-mitochondrial HGT.

The *L. mirabile* mtDNA contains several foreign regions with similarity to nuclear and plastid sequences of legumes. A nuclear-derived region with similarity to the gene pyruvate decarboxylase (*pdc*) showed a close relationship to nuclear sequences from legumes (chr49 in Figure S3) and, in particular, to a short sequence located in the *A. ligulata* mtDNA with strong bootstrap support (Figure S7). The *L. mirabile pdc* gene piece is inserted in a 3.2 kb region with 89% identity to *A. ligulata* mtDNA. These findings suggest that *A. ligulata* mtDNA acquired the *pdc* sequence via intracellular gene transfer from its nuclear genome. The lack of introns indicates that it was mRNA-mediated. Later, a mitochondrial region of the legume donor including the *pdc* was transferred to *L. mirabile* mtDNA via mt-to-mt HGT.

In addition, the *L. mirabile* mtDNA contains eight plastid regions (MTPTs) of foreign origin, which were acquired from a legume (Gandini & Sanchez-Puerta 2017; Sanchez-Puerta *et al.* 2017). Analyses of the flanking regions of these MTPTs found evidence for mt-to-mt HGT from mimosoid legumes for five of them (Gandini & Sanchez-Puerta 2017). Here, we gathered evidence of mt-to-mt HGT for a short MTPT of 113 bp, which showed similarity to legume chloroplast intergenic sequences and was identical to an MTPT in the

Acacia mtDNA (Figure S8). This MTPT of *L. mirabile* was embedded within a ~4-kb foreign mitochondrial region highly similar to *A. ligulata* mtDNA (99% identity) (chr55 in Figure S3). We conclude that this region was most likely acquired from a mimosoid via mt-to-mt HGT. These findings reinforce the hypothesis that foreign nuclear or chloroplast sequences in angiosperm mitochondrial genomes most likely entered through mt-to-mt HGT, following intracellular transfers within the donor plant (Gandini & Sanchez-Puerta 2017; Rice *et al.* 2013).

3.6 The HGT from host to parasite strengthens the mitochondrial-fusion compatibility model

The pattern of angiosperm mt-to-mt HGT set the basis for the mitochondrial fusion compatibility model (Rice *et al.* 2013). According to this model, HGT in plant mitochondria occurs mainly by capture of entire mitochondria from foreign, donor plants (or green algae), followed by fusion of native and foreign mitochondria and the recombination of their genomes. This model is based on the fact that angiosperm mitochondria normally fuse (Arimura *et al.* 2004; Sheahan *et al.* 2005), that species of the green lineage share a similar mitochondrial fusion mechanism, which differs from that of other lineages, such as fungi or animals (Arimura 2018; Mishra & Chan 2016), and that plant mitochondrial genomes frequently undergo homologous recombination to form a chimeric mitochondrial genome in somatic hybrids (Sanchez-Puerta *et al.* 2015). The mitochondrial fusion compatibility model predicts that: i) mainly foreign mitochondrial sequences are transferred to the recipient mtDNA, that is, no chloroplast of nuclear sequences should be directly acquired by the recipient mitochondria; ii) transfers are DNA-based, instead of RNA mediated, and should include large tracts, introns, and intergenic regions; iii) only mitochondrial sequences from members of the green lineage are transferred to the recipient mtDNA, that is, sequences from bacteria, viruses, or fungal mitochondria, for example, are not expected to be horizontally transferred into plant mitochondria (Rice *et al.* 2013). The findings we report here represent strong evidence for the mitochondrial-fusion compatibility model because they show the horizontal acquisition of 486 kb of mainly intergenic regions, all foreign sequences were related to legumes, and were transferred exclusively from the legume mitochondria, with no exceptions. The foreign DNA in *L. mirabile* mtDNA may be the product of repetitive Horizontal Genome Transfers (HGT) from ancestral mimosoid (or other legume) hosts during a long period of time. Even serial HGT events could have taken place, in which newly acquired foreign sequences recombine with older foreign tracts, as observed in the *Amborella* mtDNA (Rice *et al.* 2013).

4. ACKNOWLEDGEMENTS

We thank H. A. Sato for sharing photographs of the legume and the holoparasite. This work was supported by Agencia Nacional de Promoción Científica y Tecnológica (grant number PICT1762), and Universidad Nacional de Cuyo (grant number 06/A658) to M.V.S.P, by the Australian Research Council (grant numbers IC150100041 to P.N., FL140100179 to I.S., and DE120101117 to K.A.H) and by NSF (grant number 1062432) to Indiana University, which supports the computer cluster.

5. REFERENCES

Alverson, A.J., Zhuo, S., Rice, D.W., Sloan, D., & Palmer, J.D. (2011). The mitochondrial genome of the legume Vigna radiata and the analysis of recombination across short mitochondrial repeats. *PLoS One,* 6: e16404.

Alverson, A.J., Wei, X., Rice, D.W., Stern, D.B., Barry, K., & Palmer, J.D. (2010). Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol. Biol. Evol.,* 27: 1436-48.

Arimura, S. (2018). Fission and fusion of plant mitochondria, and genome maintenance. *Plant Physiol.,* 176: 152-61.

Arimura, S., Yamamoto, J., Aida, G.P., Nakazono, M., & Tsutsumi, N. (2004). Frequent fusion and fission of plant mitochondria with unequal nucleoid distribution. *Proc. Nat. Acad. Sci. USA,* 101: 7805-08.

Barkman, T.J., McNeal, J.R., Lim, S.H., Coat, G., Croom, H.B., Young, N.D., & dePamphilis, C.W. (2007). Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evol. Biol.,* 7: 248.

Bergthorsson, U., Adams, K.L., Thomason, B., & Palmer, J.D. (2003). Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature,* 424: 197-201.

Bi, C., Wang, X.D., Xu, Y., Wei, S., Shi, Y., Dai, X., Yin, T., & Ye, N. (2016). The complete mitochondrial genome of *Medicago truncatula*. *Mitochondrial DNA Part B,* 1: 122-23.

Bolger, A., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* btu170:

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T.L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics,* 10: 421.

Chang, S., Wang, Y., Lu, J., Gai, J., Li, J., Chu, P., Guan, R., & Zhao, T. (2013). The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels. *PLoS One,* 8: e56502.

Cho, Y., Adams, K.L., Qiu, Y.L., Kuhlman, P., Vaughn, J.C., & Palmer, J.D. (1998). A highly invasive group I intron in the mitochondrial *cox1* gene, in K. Moller*, et al.* (eds.), *Plant mitochondria: from gene to function* (Leiden: Backhuys Publishers), 19-23.

Davis, C. & Wurdack, K. (2004). Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science,* 305: 676-78.

Gandini, C.L. & Sanchez-Puerta, M.V. (2017). Foreign plastid sequences in plant mitochondria are frequently acquired via mitochondrion-to-mitochondrion horizontal transfer. *Sci. Rep.,* 7: 43402.

Gordon, D. & Green, P. (2013). Consed: a graphical editor for next-generation sequencing. *Bioinformatics,* 29: 2936-37.

Kado, T. & Innan, H. (2018). Horizontal gene transfer in five parasite plant species in Orobanchaceae. *G.B.E.,* evy219: in press.

Kazakoff, S.H., Imelfort, M., Edwards, D., Koehorst, J., Biswas, B., Batley, J., Scott, P.T., & Gresshoff, P. (2012). Capturing the biofuel wellhead and powerhouse: the chloroplast and mitochondrial genomes of the leguminous feedstock tree *Pongamia pinnata*. *PLoS One,* 7: e54687.

Keeling, P.J. & Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.,* 9: 1-14.

Kim, G., LeBlanc, M.L., Wafula, E., dePamphilis, C.W., & Westwood, J.H. (2014). Genomic-scale exchange of mRNA between a parasitic plant and its hosts. *Science,* 345: 808-11.

Kovar, L., Nageswara-Rao, M., Ortega-Rodriguez, S., Dugas, D., Straub, S.C.K., Cronn, R., Strickler, S., Hughes, C., Hanley, K., Rodriguez, D. *et al.* (2018). PacBio-based mitochondrial genome assembly of *Leucaena trichandra* (Leguminosae) and an intrageneric assessment of mitochondrial RNA editing. *G.B.E.,* 10: 2501-17.

Krumsiek, J., Arnold, R., & Gepard, R. (2007). Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics,* 23: 1026-28.

Liu, G., Cao, D., Li, S., Su, A., Geng, J.N., Grover, C.E., Hu, S., & Hua, J. (2013). The complete mitochondrial genome of *Gossypium hirsutum* and evolutionary analysis of higher plant mitochondrial genomes. *PLoS One,* 8: e69476.

Lohse, M., Drechsel, O., & Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.,* 52: 267-74.

Lowe, T.M. & Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.,* 25: 955-64.

LPWG (2017). A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon,* 66: 44-77.

Mishra, P. & Chan, D.C. (2016). Metabolic regulation of mitochondrial dynamics. *J. Cell Biol.,* 212: 379-87.

Mower, J.P., Sloan, D.B., & Alverson, A.J. (2012). Plant mitochondrial genome diversity: The genomics revolution, in J.F. Wendel*, et al.* (eds.), *Plant Genome Diversity* (I; Wien: Springer-Verlag), 123-44.

Mower, J.P., Stefanovic, S., Young, G.J., & Palmer, J.D. (2004). Plant genetics: gene transfer from parasitic to host plants. *Nature,* 432: 165-66.

Mower, J.P., Stefanovic, S., Hao, W., Gummow, J.S., Jain, K., Ahmed, D., & Palmer, J.D. (2010). Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. *BMC Biol.,* 8: 150.

Naito, K., Kaga, A., Tomooka, N., & Kawase, M. (2013). *De novo* assembly of the complete organelle genome sequences of azuki bean (*Vigna angularis*) using next-generation sequencers. *Breeding Science,* 63: 176-82.

Negruk, V. (2013). Mitochondrial genome sequence of the legume *Vicia faba*. *Front. Plant Sci.,* 4: e128.

Press, M. & Phoenix, G. (2005). Impacts of parasitic plants on natural communities. *New Phytol.,* 166: 737-51.

Rice, D.W., Alverson, A.J., Richardson, A.O., Young, G.J., Sanchez-Puerta, M.V., Munzinger, J., Barry, K., Boore, J.L., Zhang, Y., dePamphilis, C.W. *et al.* (2013). Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science,* 342: 1468-73.

Sanchez-Puerta, M.V., Zubko, M.K., & Palmer, J.D. (2015). Homologous recombination and retention of a single form of most genes shape the highly chimeric mitochondrial genome of a cybrid plant. *New Phytol.,* 206: 381-96.

Sanchez-Puerta, M.V., García, L.E., Wohlfeiler, J., & Ceriotti, F. (2017). Unparalleled replacement of native mitochondrial genes by foreign homologs in a holoparasitic plant. *New Phytol.,* 214). doi: 10.1111/nph.14361.

Sanchez-Puerta, M.V., Cho, Y., Mower, J.P., Alverson, A.J., & Palmer, J.D. (2008). Frequent, phylogenetically local horizontal transfer of the *cox1* group I intron in flowering plant mitochondria. *Mol. Biol. Evol.,* 25: 1762-77.

Sheahan, M.B., McCurdy, D.W., & Rose, R.J. (2005). Mitochondria as a connected population: ensuring continuity of the mitochondrial genome during plant cell dedifferentiation through massive mitochondrial fusion. *Plant J.,* 44: 744-55.

Shi, Y., Liu, Y., Zhang, S., Zou, R., Tang, J., Mu, W., Peng, Y., & Dong, S. (2018). Assembly and comparative analysis of the complete mitochondrial genome sequence of *Sophora japonica* 'Jinhuaij2'. *PLoS One,* 13: e0202485.

Skippington, E., Barkman, T.J., Rice, D.W., & Palmer, J.D. (2015). Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad* genes. *Proc. Nat. Acad. Sci. USA,* 112: E3515-E24.

--- (2017). Comparative mitogenomics indicates respiratory competence in parasitic *Viscum* despite loss of complex I and extreme sequence divergence, and reveals horizontal gene transfer and remarkable variation in genome size. *BMC Plant Biol.,* 17: 49.

Sloan, D. (2013). One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. *New Phytol.,* 200: 978-85.

Sloan, D.B. & Wu, Z. (2014). History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. *G.B.E.,* 6: 3210-21.

Stamatakis, A. (2014). RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30: 1312-13.

Stegemann, S. & Bock, R. (2009). Exchange of genetic material between cells in plant tissue grafts. *Science,* 324: 649-51.

Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R., & Liston, A. (2012). Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Am. J. Bot.,* 99: 349-64.

Su, H.-J., Hu, J.-M., Anderson, F.E., Der, J.P., & Nickrent, D.L. (2015). Phylogenetic relationships of Santalales with insights into the origins of holoparasitic Balanophoraceae. *Taxon,* 64: 491-506.

Vogel, A., Schwacke, R., Denton, A., Usadel, B., Hollmann, J., Fischer, K., Bolger, A., Schmidt, M., Bolger, M., Gundlach, H. *et al.* (2018). Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. *Nature Comm.,* 9: 2515.

Warren, J., Simmons, M., Wu, Z., & Sloan, D.B. (2016). Linear plasmids and the rate of sequence evolution in plant mitochondrial genomes. *G.B.E.,* 8: 364-74.

Xi, Z., Wang, Y., Bradley, R., Sugumaran, M., Marx, C., Rest, J., & Davis, C.C. (2013). Massive mitochondrial gene transfer in a parasitic flowering plant clade. *PloS Genet.,* 9: e1003265.

Yang, Z., Zhang, Y., Wafula, E., Honaas, L., Ralph, P., jones, S., Clarke, C., Liu, S., Su, C., Zhang, H. *et al.* (2016). Horizontal gene transfer is more frequent with increased heterotrophy and contributes to parasite adaptation. *Proc. Nat. Acad. Sci.,* 113: E7010-E19.

Yu, T., Sun, L., Cui, H., Liu, S., Men, J., Chen, S., Chen, Y., & Lu, C. (2018). The complete mitochondrial genome of a tertiary relict evergreen woody plant *Ammopiptanthus mongolicus*. *Mitochondrial DNA Part B,* 3: 9-11.

Zerbino, D. & Birney, E. (2008). Velvet: algorithms for de *novo* short read assembly using de Bruijn graphs. *Genome Res.,* 18: 821-29.

Zhang, Y., Fernandez-Aparicio, M., Wafula, E., Das, M., Jiao, Y., Wickett, N., Honaas, L., Ralph, P., Wojciechowski, M.F., Timko, M. *et al.* (2013). Evolution of a horizontally acquired legume gene, albumin 1, in the parasitic plant Phelipanche aegyptiaca and related species. *BMC Evol. Biol.,* 13: 48.

6. TABLES

Table 1. Features of the mitochondrial genome of *Acacia ligulata*

Genome length in bp	698,138
Protein-coding genes[a]	37(38)
rRNA genes[a]	3(3)
tRNA genes[a]	20(22)
Group II introns	
cis-splicing	19
trans-splicing	5
Group I introns	0
Repeats in kb (% of genome)[b]	20.5 (2.94%)
Large repeats (>1kb) in kb (% of genome)[b]	17.2 (2.46%)
Plastid-derived sequences (% of genome)	3.23%
Mitochondrial genes (exons and cis-spliced introns)	11.37%

[a]First value excludes duplicates; value in parentheses includes them.

[b]Total length of repeats.

7. FIGURE LEGENDS

Figure 1. Linear representation of the mitochondrial genome of *Acacia ligulata*. Genes drawn above and below the main line are transcribed from opposite strands of the genome. Shown are full-length genes, large repeats (>1 kb) with >90% identity (labeled 'Repeat', followed by the repeat lengths in kb), and chloroplast-derived sequences longer than 200 bp. The mtDNA could be mapped as two subgenomic circular molecules that differ in a short region (blue and red segments). If the two circles recombine, long head-to-tail concatemers could be formed.

Figure 2. Evolutionary origin of the mtDNA of *Lophophytum mirabile*. Relative amount of sequences with affinity to mimosoid (dark blue), other legume (light blue), or other angiosperm (yellow) mitochondrial genomes, and those uncharacterized (light grey) are shown for the whole genome (a) or each mitochondrial chromosome (b). *L. mirabile* chromosomes with >80% foreign DNA are marked with an asterisk.

Figure 3. Evolutionary origin of *Lophophytum mirabile* mitochondrial chromosome sequences. For each chromosome regions of *L. mirabile* mtDNA with affinity to mimosoid, other legume, or other angiosperm mitochondrial genomes are shown. Colors depict sequence identity of BLAST hits according to the scale shown below. Genic regions are depicted with grey rectangles on the left.

8. SUPPLEMENTARY MATERIAL

Table S1. Features of mitochondrial-encoded genes in the *Acacia ligulata* mtDNA.

Table S2. Origin of genes encoded in the *Lophophytum mirabile* mtDNA.

Figure S1. Total read-depth of the *Acacia ligulata* mitochondrial genome (average 70x). Spikes of the read-depth are the result of mismapping of cpDNA-derived reads onto large MTPTs with high identity to chloroplast sequences.

Figure S2. Dot plot comparisons of the mitochondrial genomes of *Acacia ligulata* and *Leucaena trichandra* (a), *Senna tora* (b), *Glycine max* (c), or *Lophophytum mirabile* (d) and of *Leucaena trichandra* and *L. mirabile* (e) showing regions of shared synteny. The *L. mirabile* mtDNA is represented by a concatenation of the 60 chromosomes. The fraction of the mtDNAs covered by the blast hits (qcov) and the average identity of the hits >250 bp are indicated below each plot.

Figure S3. Evolutionary origin of *Lophophytum mirabile* mitochondrial chromosomes. BLAST hits between *L. mirabile* and mimosoid legumes (*Acacia ligulata* or *Leucaena trichandra*), other legumes (*Ammopiptanthus mongolicus*, *Glycine max, Lotus japonicus, Medicago truncatula, Millettia pinnata, Senna occidentalis, S. tora,*

Sophora japonica, Vicia faba, Vigna angularis, and *V. radiata*), or other angiosperms are shown in the top three rows, respectively, of each chromosome graph. Regions of *L. mirabile* with affinity to mimosoid (foreign), other legume (foreign), or other angiosperm (putatively native) mitochondrial genomes are shown inside the grey boxes. Colors depict sequence identity of BLAST hits according to the scale shown on the right. Genic regions are depicted with grey rectangles below each chromosome.

Figure S4. Phylogenetic analyses showing different regions of the mitochondrial genome of *Lophophytum mirabile* with strong affiliation to the family Fabaceae. Maximum likelihood analyses were performed with RaxML. ML bootstrap support values >50% are shown. The scale bar corresponds to substitutions per site.

Figure S5. Maximum likelihood trees of the gene *rrnL* found in the mtDNA of *Lophophytum mirabile*. Phylogenetic analyses were based on the 5' (a) and the 3' region (b) of the mitochondrial gene *rrnL*. The conflicting affiliations of the *Lophophytum mirabile rrnL* 5' and 3' regions support a chimeric origin. ML bootstrap support values >50% are shown. The scale bar corresponds to substitutions per site.

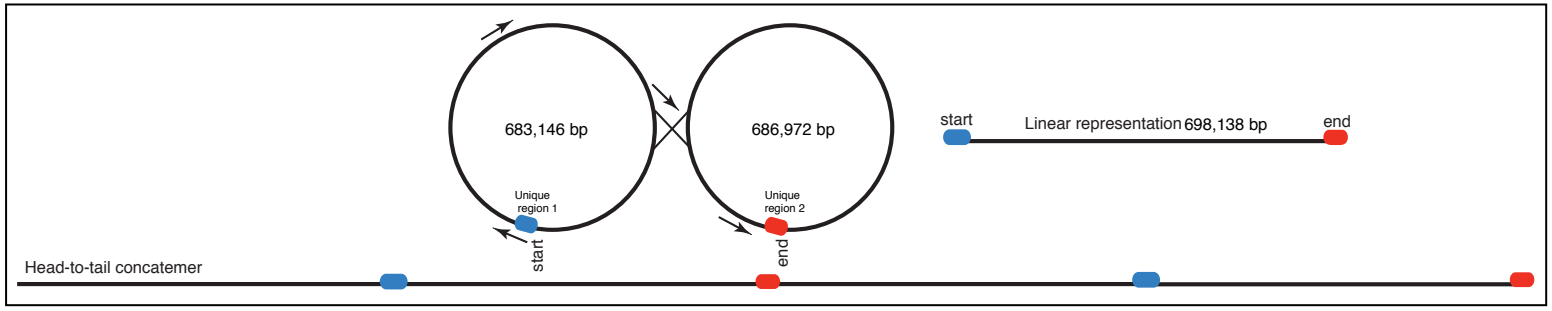
Figure S6. Analyses of short genes from *Lophophytum mirabile* mitochondrial chromosomes. Each graph represents a region of the *L. mirabile* mtDNA containing the genes under study (green squares). BLAST hits from mimosoids, other legumes and other angiosperms are shown in dark blue, light blue, or yellow, respectively, and vertically distributed according to their sequence identity.

Figure S7. Maximum Likelihood phylogenetic analysis of nuclear and mitochondrial (MT) encoded sequences with similarity to the gene pyruvate decarboxylase (*pdc*). The lack of introns in the mitochondrial-encoded sequences indicates that the intracellular transfer was RNA-mediated and then integrated into the mitochondrial genome via reverse transcription. ML bootstrap support values >50% are shown. Pseudogenes are depicted by 'Ψ'. The scale bar corresponds to substitutions per site.

Figure S8. Maximum Likelihood phylogenetic analysis of a plastid-encoded intergenic region and mitochondrial-encoded plastid-derived sequences (MTPT). ML bootstrap support values >50% are shown. The scale bar corresponds to substitutions per site.

Figure 1

Acacia ligulata
mtDNA
698,138 bp



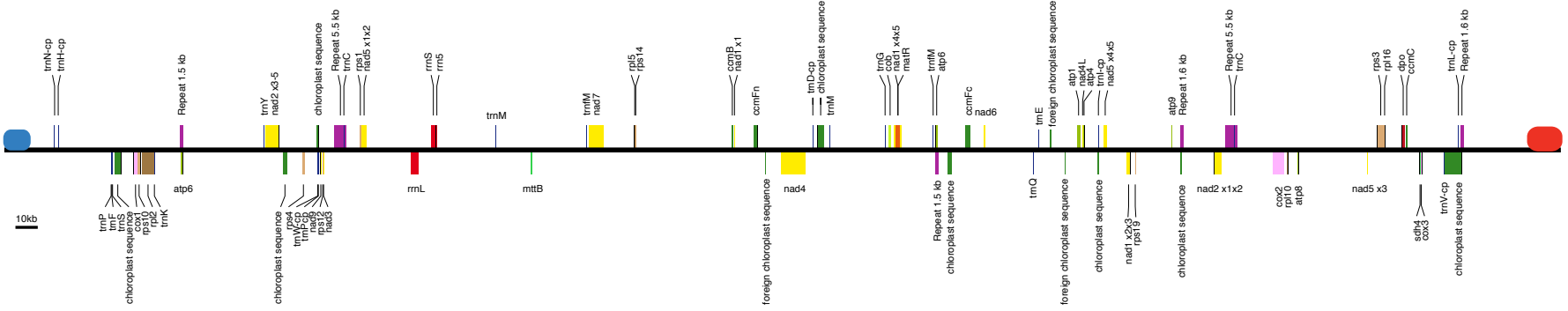


Figure 2

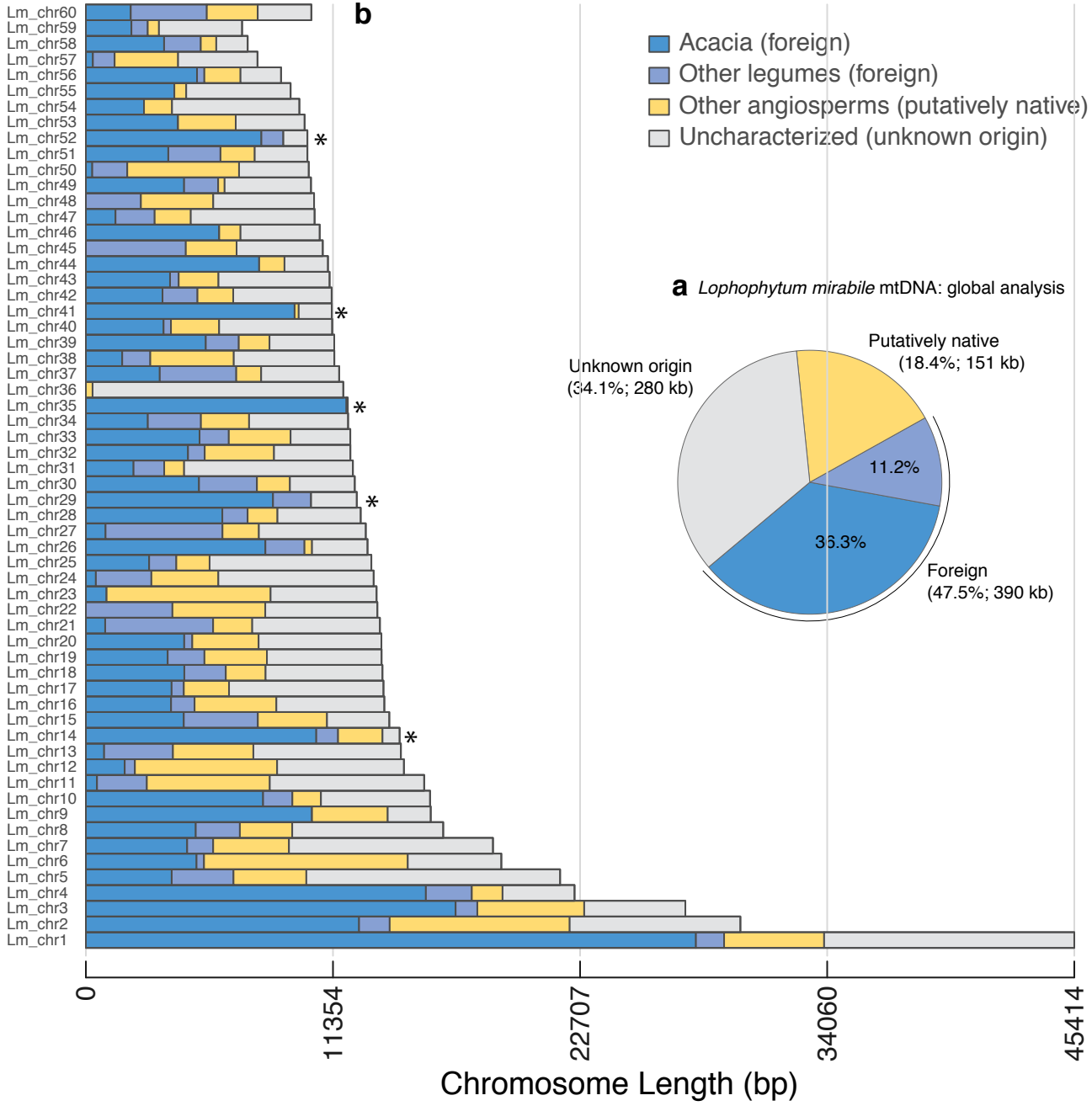

b
Acacia (foreign)
Other legumes (foreign)
Other angiosperms (putatively native)
Uncharacterized (unknown origin)
Lm_chr60
Lm_chr59
Lm_chr58
Lm_chr57
Lm_chr56
Lm_chr55
Lm_chr54
Lm_chr53
Lm_chr52
Lm_chr51
Lm_chr50
Lm_chr49
Lm_chr48
Lm_chr47
Lm_chr46
Lm_chr45
Lm_chr44
Lm_chr43
Lm_chr42
Lm_chr41
Lm_chr40
Lm_chr39
Lm_chr38
Lm_chr37
Lm_chr36
Lm_chr35
Lm_chr34
Lm_chr33
Lm_chr32
Lm_chr31
Lm_chr30
Lm_chr29
Lm_chr28
Lm_chr27
Lm_chr26
Lm_chr25
Lm_chr24
Lm_chr23
Lm_chr22
Lm_chr21
Lm_chr20
Lm_chr19
Lm_chr18
Lm_chr17
Lm_chr16
Lm_chr15
Lm_chr14
Lm_chr13
Lm_chr12
Lm_chr11
Lm_chr10
Lm_chr9
Lm_chr8
Lm_chr7
Lm_chr6
Lm_chr5
Lm_chr4
Lm_chr3
Lm_chr2
Lm_chr1
0
11354
22707
34060
45414
Chromosome Length (bp)
a Lophophytum mirabile mtDNA: global analysis
Unknown origin
(34.1%; 280 kb)
Putatively native
(18.4%; 151 kb)
11.2%
33.3%
Foreign
(47.5%; 390 kb)

Table 1. Features of the mitochondrial genome of *Acacia ligulata*

Genome length in bp	698,138
Protein-coding genes[a]	37(38)
rRNA genes[a]	3(3)
tRNA genes[a]	19(22)
Group II introns	
cis-splicing	18
trans-splicing	5
Group I introns	0
Repeats in kb (% of genome)[b]	20.5 (2.94%)
Large repeats (>1kb) in kb (% of genome)[b]	17.2 (2.46%)
Plastid-derived sequences (% of genome)	3.71%
Mitochondrial genes (exons and cis-spliced introns)	11.37%
Uncharacterized	24.40%

[a]First value excludes duplicates; value in parentheses includes them.

[b]Total length of repeats.