New Phytologist

Research

Horizontal gene transfers dominate the functional mitochondrial gene space of a holoparasitic plant

Laura E. Garcia[1,2]*, Alejandro A. Edera[1]*, Jeffrey D. Palmer[3], Hector Sato[4] and M. Virginia Sanchez-Puerta[1,2]

[1]IBAM, Universidad Nacional de Cuyo, CONICET, Facultad de Ciencias Agrarias, Almirante Brown 500, Chacras de Coria M5528AHB, Argentina; [2]Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo, Padre Jorge Contreras 1300, Mendoza M5502JMA, Argentina; [3]Department of Biology, Indiana University, Bloomington, IN 47405, USA; [4]Facultad de Ciencias Agrarias (UNJu), Cátedra de Botánica General–Herbario JUA, Alberdi 47, Jujuy CP 4600, Argentina

Author for correspondence:
M. Virginia Sanchez-Puerta
Email: mvsanchezpuerta@fca.uncu.edu.ar

Received: *26 June 2020*
Accepted: *2 September 2020*

New Phytologist (2021) **229:** 1701–1714
doi: 10.1111/nph.16926

Key words: Balanophoraceae, holoparasite, *Lophophytum*, mitochondrial DNA, RNA editing, transcriptome.

Summary

- Although horizontal gene transfer (HGT) is common in angiosperm mitochondrial DNAs (mtDNAs), few cases of functional foreign genes have been identified. The one outstanding candidate for large-scale functional HGT is the holoparasite *Lophophytum mirabile*, whose mtDNA has lost most native genes but contains intact foreign homologs acquired from legume host plants.
- To investigate the extent to which this situation results from functional replacement of native by foreign genes, functional mitochondrial gene transfer to the nucleus, and/or loss of mitochondrial biochemical function in the context of extreme parasitism, we examined the *Lophophytum* mitochondrial and nuclear transcriptomes by deep paired-end RNA sequencing.
- Most foreign mitochondrial genes in *Lophophytum* are highly transcribed, accurately spliced, and efficiently RNA edited. By contrast, we found no evidence for functional gene transfer to the nucleus or loss of mitochondrial functions in *Lophophytum*. Many *functional* replacements occurred via the *physical* replacement of native genes by foreign genes. Some of these events probably occurred as the final act of HGT itself.
- *Lophophytum* mtDNA has experienced an unprecedented level of functional replacement of native genes by foreign copies. This raises important questions concerning population-genetic and molecular regimes that underlie such a high level of foreign gene takeover.

Introduction

The evolution of angiosperm mitochondrial DNA (mtDNA) is often impacted by horizontal gene transfer (HGT) events that introduce mitochondrial sequences from unrelated plant species (Bergthorsson *et al.*, 2003; Won & Renner, 2003; Davis *et al.*, 2005; Mower *et al.*, 2010; Rice *et al.*, 2013; Park *et al.*, 2015; Sanchez-Puerta *et al.*, 2017). Mitochondria of parasitic plants are especially prone to HGT (Davis & Wurdack, 2004; Mower *et al.*, 2004; Barkman *et al.*, 2007; Xi *et al.*, 2013; Sanchez-Puerta *et al.*, 2017; Schneider *et al.*, 2018), as the intimate contact between host and parasite through haustorial structures facilitates the exchange of nutrients, as well as macromolecules (Kim *et al.*, 2014; Yang *et al.*, 2016; Wickell & Li, 2020) and even whole mitochondria (Sanchez-Puerta *et al.*, 2019).

Two features of angiosperm mitochondrial gene expression are critical for evaluating the potential functionality of foreign genes acquired by HGT, and, for that matter, of native genes too. First, despite consisting predominantly of intergenic spacer DNA, angiosperm mtDNAs are nearly entirely transcribed, but with transcript levels so low over most of the genome as to be considered functionally insignificant (Holec *et al.*, 2006; Fujii *et al.*, 2011; Grimes *et al.*, 2014; Wu *et al.*, 2015b; Silva *et al.*, 2017; Lloyd Evans *et al.*, 2019). Second, efficient C-to-U RNA editing occurs at hundreds of nonsynonymous sites and is essential to produce the correct amino acid sequence for the great majority of mitochondrial protein genes in land plants (Small *et al.*, 2019). Furthermore, even pseudogenes can be edited, usually with low efficiency, in plant mitochondria (Rice *et al.*, 2013). Therefore, assigning functionality to foreign genes requires RNA sequencing (RNA-seq) that is both deep and genome-wide, in order to detect high transcript levels that significantly exceed basal levels of genomic expression and are comparable to those of native genes, as well as efficient RNA editing at nonsynonymous sites.

Most foreign genes in plant mitochondria are pseudogenes (Won & Renner, 2003; Bergthorsson *et al.*, 2004; Davis & Wurdack, 2004; Mower *et al.*, 2004; Davis *et al.*, 2005; Rice *et al.*, 2013; Park *et al.*, 2015; Forgione *et al.*, 2019) or are intact but not detectably transcribed or properly edited at the RNA level

*These authors contributed equally to this work.

(Davis *et al.*, 2005; Mower *et al.*, 2010; Hepburn *et al.*, 2012; Rice *et al.*, 2013; Sinn & Barrett, 2020). These nonfunctional foreign genes almost always coexist with functional native homologs. There are also a number of cases where intact foreign and native homologs coexist but where expression has not been examined (Xi *et al.*, 2013; Park *et al.*, 2015; Roulet *et al.*, 2020). For only 10 putatively foreign protein genes in plant mitochondria has a functional role been inferred based on demonstrated transcription and RNA editing (Bergthorsson *et al.*, 2003; Hao *et al.*, 2010; Rice *et al.*, 2013; Xi *et al.*, 2013). Critical assessment of these 10 cases, however, indicates that only four are reasonable candidates for being functional foreign genes, with two being considered strong candidates (Supporting Information Fig. S1; Notes S1).

Rampant HGT has transformed the mitogenome of the holoparasitic angiosperm *Lophophytum mirabile* (Balanophoraceae). At least 60% of this 822 kb genome was acquired by HGT in a series of transfers from various legume donors (Sanchez-Puerta *et al.*, 2017, 2019; Roulet *et al.*, 2020). Most of these donors trace back to the mimosoid clade of legumes, woody members of which currently serve as the host for the root parasite *L. mirabile* and other species of *Lophophytum* (Sanchez-Puerta, *et al.*, 2017). The *Lophophytum* mitogenome consists of 60 circular chromosomes, 29 of which are devoid of intact genes. The remaining chromosomes contain 44 intact protein genes, encoding 35 functionally distinct proteins (Sanchez-Puerta *et al.*, 2017). Only nine of these proteins are represented solely by native genes, whereas 23 are represented solely by foreign or chimeric genes and three by both foreign and native genes (Sanchez-Puerta, *et al.*, 2017). If these foreign genes have functionally replaced the missing native ones, then this would be an unprecedented proportion of replacement HGT for any mitogenome, plant or otherwise, and even for prokaryotes, where HGT mostly adds new kinds of genes. The *Lophophytum* situation stands in striking contrast to the other extraordinary case of mitochondrial HGT in angiosperms. Even though the mtDNA of *Amborella trichopoda* contains nearly 200 foreign mitochondrial genes, it retains a full set of intact and expressed native protein genes, and its foreign genes are overwhelmingly pseudogenes or poorly expressed, with but a single foreign protein gene a reasonable candidate for being functional (Rice *et al.*, 2013).

At least three processes could contribute to the exceptional loss of native genes from *Lophophytum* mtDNA. One process is HGT-driven *functional* replacement of some or all of the missing native mitochondrial genes by their foreign homologs. To assess this requires deep and broad sequencing of the mitochondrial transcriptome, as described above. The foreign genes are irrelevant – a red herring – to the other two processes: Some missing native genes may have been replaced by nuclear homologs, mostly likely through functional gene transfer from the mitogenome. Or they may have been lost entirely from the cell because they no longer have a biochemical role to play in the context of the extreme parasitism displayed by *Lophophytum*. Angiosperms have exceptionally high rates of functional transfer to the nucleus of their ancestrally 17 mitochondrial ribosomal protein and *sdh* genes (Adams *et al.*, 2002; Adams & Palmer, 2003). It is therefore imperative to search for functionally transferred copies of these (and other) genes and determine if these nuclear copies correlate with nonfunctionality of their foreign (or even native) mitochondrial homologs. A hallmark of parasitism is reductive evolution of mitochondrial function and genome (Zíková *et al.*, 2016; Santos *et al.*, 2018). A striking example of this in parasitic plant mitochondria is the loss of respiratory complex I and its associated nine mitochondrial genes in *Viscum* mistletoes (Skippington *et al.*, 2015; Skippington *et al.*, 2017), with this loss leading to a highly remodeled respiratory chain and greatly diminished oxidative phosphorylation (OXPHOS; Maclean *et al.*, 2018; Senkler *et al.*, 2018). Compared with hemiparasitic, photosynthetic *Viscum*, holoparasite *Lophophytum* is substantially reduced in morphological and biochemical complexity, lacking photosynthesis, leaves, stomata, epidermis, and trichomes (Gonzalez & Sato, 2016). It would not be surprising, therefore, if *Lophophytum* has also lost normally essential mitochondrial functions and genes. Evidence for this would be fundamentally negative; that is, failure to find meaningful expression of foreign and native mitochondrial genes in *Lophophytum* coupled with failure to find functional homologs of these genes in the nucleus. To determine the relative contributions of these three processes to the unprecedented genetic make-up of the mitogenome in *Lophophytum*, we examined its mitochondrial and nuclear transcriptomes by deep strand-specific RNA sequencing.

Materials and Methods

Plant collection and RNA sequencing

Inflorescences composed of male and female flowers were collected in February 2017 from an individual of *Lophophytum mirabile* growing in Parque Nacional Calilegua (23°55′34.9″S, 64°55′58.5″W, 733 m amsl) that was parasitizing the mimosoid legume *Anadenanthera colubrina*. We produced RNA depleted of ribosomal RNA (rRNA) because poly-A tracts represent a signal for RNA degradation in plant mitochondria (Kuhn *et al.*, 2001; Holec *et al.*, 2006; Hirayama *et al.*, 2013). Total RNA was extracted using a protocol for highly viscous samples rich in polysaccharides (Zeng & Yang, 2012), purified using the RNAqueous kit (Invitrogen), which uses columns with glass fiber filters, treated with Ribo-Zero Plant Leaf Kit (Illumina), and sequenced on an Illumina Hiseq2500 at Macrogen (Korea) using the TruSeq-Stranded-Total-RNA-LT-Sample Prep Kit. The quality of the 591 066 570 paired-end reads (101 bp read lengths, with insert sizes of 138 bp) was assessed using FastQC v.0.10.1. Read adapters and low-quality sequences were trimmed using Trimmomatic (Bolger *et al.*, 2014), leading to 562 330 328 paired-end reads of length 91 bp. The RNA-seq data are available as NCBI Bioproject PRJNA601125.

RNA read alignments

Alignment was performed using Bowtie 2 (Langmead & Salzberg, 2012), built from source code (v.2.3.4.3) to improve Bowtie time-execution performance, using a sensitive and end-to-end mode along with the no-discordant, no-mixed, and

$R = 10$ options as well as the fr and nofw options to enforce strand-specific alignment. RNA reads were aligned to two sets of DNA sequences. One contained both strands of all 60 mitochondrial chromosomes of *Lophophytum* (KU992322–KU992380 and KX792461). The other contained all intact protein genes annotated in the *Lophophytum* mitogenome, with these sequences extended by 100 bp at their 5′ and 3′ ends using SeqKit v.0.8.0 (Shen *et al.*, 2016) to obtain good read coverage at their ends. The mapping strategy allowed the alignment of edited reads to unedited reference sequences, even for those rich in editing sites. To be sure that the read depth was not an underestimation, we replaced each editing site with N in the reference sequences of highly edited genes, such as *ccmC* or *nad4L*, but their read depths did not change substantially.

Statistical test for detecting significant RNA read depth

We used the method of Wu *et al.* (2015b) to determine whether the RNA read depth of a genomic region was significant. This method was developed for large plant mtDNAs that consist predominantly of intergenic sequences (92% in the case of *Lophophytum*) that are poorly expressed. The method calculates the probability of having an RNA read depth, averaged from 500 bp genomic windows, that is greater than or equal to the 5% tail of the distribution of read depth across a mitogenome of interest. In *Lophophytum* mtDNA, this tail corresponds to an average read depth of 3960. This means that, in *Lophophytum* mtDNA, there is a probability of $\leq 5\%$ that a region of 500 bp in length has an average read depth ≥ 3960 RNA reads.

Open reading frame and long noncoding RNA identification

We used getorf (Emboss:6.6.0.0) to identify open reading frames (ORFs) starting with ATG and of length > 300 bp in *Lophophytum* mtDNA. For long noncoding RNA (lncRNA) prediction, a sliding window of 200 bp with stride of 1 bp was used to extract candidate regions located outside of annotated regions, large repeats, and untranslated regions (UTRs). Pairs of significant candidate regions were concatenated if they overlapped and the average of their transcript levels was significant. The RNA read depth of the ORFs and lncRNAs was calculated using the genome-wide RNA read alignment.

Identification of RNA editing events

An in-house script was implemented to count the number and type of RNA nucleotides at each position in the two DNA-sequence sets already described. DNA/RNA mismatches were assigned at those sites that were covered by 100 or more RNA reads, at least 10% of which gave a different nucleotide than did the DNA. For example, C-to-U RNA editing sites were identified as cytidine positions in the DNA sequence at which $\geq 10\%$ of the aligned RNA reads had uracils (Notes S2). For editing site inference in UTRs, 2 kb of sequence 5′ and 3′ of each protein gene was treated as a UTR. The mitochondrial protein genes of *Lophophytum* were aligned (Dataset S1) with those of 18 diverse angiosperms for which RNA-seq data are available (Edera *et al.*, 2018) using Clustal Omega v.1.2.4 (Sievers *et al.*, 2011). *Silene conica* editing sites were identified based on the following transcriptomes: SRR2090183, SRR2093550, SRR2093554, SRR2093555, SRR2093557, SRR2093573, SRR2093650, SRR2093676 (Havird *et al.*, 2017).

Trinity assembly and transcriptome analyses

The 562 million paired-end reads used for measuring mitochondrial transcript levels were subjected to a further rRNA-depletion step with SortMeRNA 2.1 (Kopylova *et al.*, 2012) using several rRNA databases (silva-bac-16s-id90, silva-bac-23s-id98, silva-euk-18s-id95, silva-euk-28s-id98, rfam-5.8s-database-id98, and rfam-5s-database-id98) and –blast 1 –paired_in –paired_out settings. This resulted in an rRNA-depleted data set of 193 166 478 paired-end reads. Assembly of this data set was performed using Trinity v.2.8.4 on the Carbonate large-memory computer cluster at Indiana University–Bloomington (USA) with the following parameters: SS_lib_type RF –min_contig_length 100. We conducted tBlastn searches of the assembled transcriptome for mitochondrial genes that may have been functionally transferred to the nuclear genome in *Lophophytum*, using each mitochondrial protein gene from *Liriodendron tulipifera* (MK340747) as query sequences.

Results

Lophophytum mirabile mitochondrial chromosomes are entirely transcribed

We obtained 562 million, high quality, paired-end RNA reads, of which *c.* 30 million aligned strand specifically to the mitochondrial genome of *Lophophytum*. These reads cover 99% of both strands of the mitogenome (Fig. S2). The predominantly short (< 100 bp) regions lacking any reads (see chromosomes KU992323 and KU992339 in Fig. S2 for longer exceptions) probably reflect mtDNA indels or rearrangements in the individual used for RNA extraction relative to the individual used for mitogenome sequencing (Notes S2). Hence, both strands of the *Lophophytum* mitogenome are probably entirely transcribed (Fig. S2).

To assess the basal transcript level (i.e., the average RNA level for those regions that probably have no function but are transcribed nonetheless), RNA reads derived from noncoding chromosomes of *Lophophytum* were analyzed. The average read depth, on both strands, of the 26 chromosomes lacking known genes, pseudogenes, and plastid-derived sequences > 200 bp (to exclude mismapped reads from plastid transcripts) is 137 with an SD = 93 (Table S1). This compares with an average read depth of 3960 established using the method of Wu *et al.* (2015b) as the threshold for statistically significant RNA level in *Lophophytum*.

Transcript levels vary enormously across the lengths of many mitochondrial chromosomes, as well as between them (Fig. S2). Significantly high transcript levels correspond to regions containing known mitochondrial genes, with the exception of one

plastid-derived sequence, three lncRNAs, and two unidentified ORFs – these last five sequences are treated in Notes S3 and S4 and in Tables S2 and S3.

Lophophytum mtDNA contains 56 intact genes (including duplicates) of known identity and 135 unidentified ORFs. The minimum, maximum, and average read depths of these genes and ORFs are highly variable (Table S4). Very short genes (i.e. transfer RNA (tRNA) genes (*trn*) and the 5S rRNA gene (*rrn5*)) show a very low read depth; this is almost certainly an artefact of very poor recovery of such short RNAs by the column-based RNA extraction method employed (Stone & Storchova, 2015) and the fact that the mature transcripts are shorter than the library size (i.e. the identified tRNA transcripts probably correspond to precursor transcripts).

The expression levels of known protein genes differ widely between genes, between copies of the same gene, and within genes (Figs 1, S3). Drops in read depth near the ends of *nad6* and *ccmC* (Fig. S3) have been reported for other plants (Wu *et al.*, 2015b; Tsujimura *et al.*, 2019; Varré *et al.*, 2019) and are probably caused by 3′ endonucleolytic processing events signaled by tRNA-like elements embedded in the 3′ ends of these genes (Forner *et al.*, 2007).

Expression of single-copy intact protein genes

The *Lophophytum* mitogenome contains 29 intact and effectively single-copy protein genes, of which nine are native, 17 were acquired from legumes by HGT, and three (*rpl2*, *nad5*, and *rps3*) are chimeric as a result of homologous recombination between native and foreign copies or because they contain foreign and native *trans*-spliced exons (Sanchez-Puerta *et al.*, 2017, 2019; Roulet *et al.*, 2020). The use of 'effectively' in the preceding sentence refers to the absence from the genome of the *L. mirabile* individual used for RNA-seq of one of the two foreign *cox3*/*sdh4* loci (these genes overlap) present in the individual used for genome sequencing (see Notes S2). Consequently, we could measure *cox3*/*sdh4* expression only for the locus, on chromosome KU992331, that is present in both individuals.

For all but three of the 29 intact and effectively single-copy genes, transcript levels were significant (Fig. 1; Table 1). The three exceptions, *ccmC* (foreign), *nad4L* (foreign), and *rpl2* (chimeric), were just barely insignificant ($P = 0.06$). In other transcriptome studies, these three genes have also shown a relatively low RNA level (Picardi *et al.*, 2010; Islam *et al.*, 2013; Grewe *et al.*, 2014; Grimes *et al.*, 2014; Wu *et al.*, 2015b). In addition, the relatively low expression level of *ccmC* is partly produced by its 3′ drop in read depth already described (Fig. S3). In sum, we view the transcript levels of these three genes as evidence that they are most likely functional.

Relative expression of protein genes present in multiple intact copies

Excluding the *cox3*/*sdh4* duplication (see earlier), we were able to assess relative expression levels for members of the five two-copy families of intact protein genes present in *Lophophytum* mtDNA as well as its one three-copy family. These copies differ at both synonymous and nonsynonymous sites. The copies differ in origin in three cases (*atp8*, *atp9* and *rpl10*), with native and foreign copies coexisting (*atp9* has one native and two foreign copies). In the other three cases (*rps4*, *rps12* and *rps19*), both copies are foreign (Fig. 1; Table 1).

In all but one of these six cases the copies have sufficient polymorphisms across the gene to estimate RNA read depth separately for each copy (Fig. 2; Table 1). The native copies of *atp8* and *rpl10* have significant read depths, whereas the foreign copies have low, insignificant read depths (Fig. 2; Table 1). Likewise, only one of the two different foreign copies of *rps4* and of *rps19* has a significant transcript level (Fig. 2; Table 1). Of the triplicates of *atp9*, the foreign copy on chromosome KU992341, has a very low, insignificant read depth (Table 1). Although the other two *atp9* copies both have significant transcript levels, the native copy has a 16-fold higher read depth than the KU992340 foreign copy (Fig. 2).

Evaluation of *rps12* expression was challenging because its two foreign copies are identical over most of their lengths such that cross-mapping of RNA reads precluded measurement of transcript levels over most of *rps12* for each copy individually. To overcome this problem, we examined the one polymorphic region in *rps12*, at the 3′ end of the gene, and found that only one *rps12* gene has significant transcript levels in this region (Fig. 2; Table 1).

Highly transcribed genes are accurately spliced and efficiently edited

Mapping of RNA reads against intron–exon junctions revealed accurate splicing of all 24 mitochondrial introns, nine of which are foreign or chimeric in origin (Table 1). Transcript levels for most introns, both *cis* and *trans*-spliced, are comparable to those of their flanking exons, with a few introns showing either somewhat reduced or higher expression and only one showing notably lower transcript levels (Notes S5).

A total of 583 C-to-U editing sites were identified in the 35 putatively functional protein genes in *Lophophytum* (Fig. 3). The number and density of editing sites differ among genes, and these sites are distributed unevenly among first, second, and third codon positions (32%, 49% and 19%, respectively). Editing efficiency varies by codon position, with averages of 77%, 83%, and 42% for first, second, and third codon positions, respectively (Fig. S4). These results are in agreement with observations in other angiosperms, with higher editing extents at nonsynonymous sites, present at first and second codon positions, in order to efficiently restore the expected amino acid sequence (Edera *et al.*, 2018). In four *Lophophytum* genes (*atp6*, *ccmFc*, *cox2* and *matR*), stop codons were post-transcriptionally generated by high-efficiency (> 93%) RNA editing (Table S5), resulting in shorter proteins (by 9–24 amino acids). No editing sites were identified within *cis*-spliced introns, and only 13 editing sites were detected in the 5′ and 3′ UTRs of mitochondrial genes (Table S6).

Thirty-two protein genes with significant transcript levels (i.e. all but atp9-KU992340) are, in aggregate, efficiently edited (i.e.

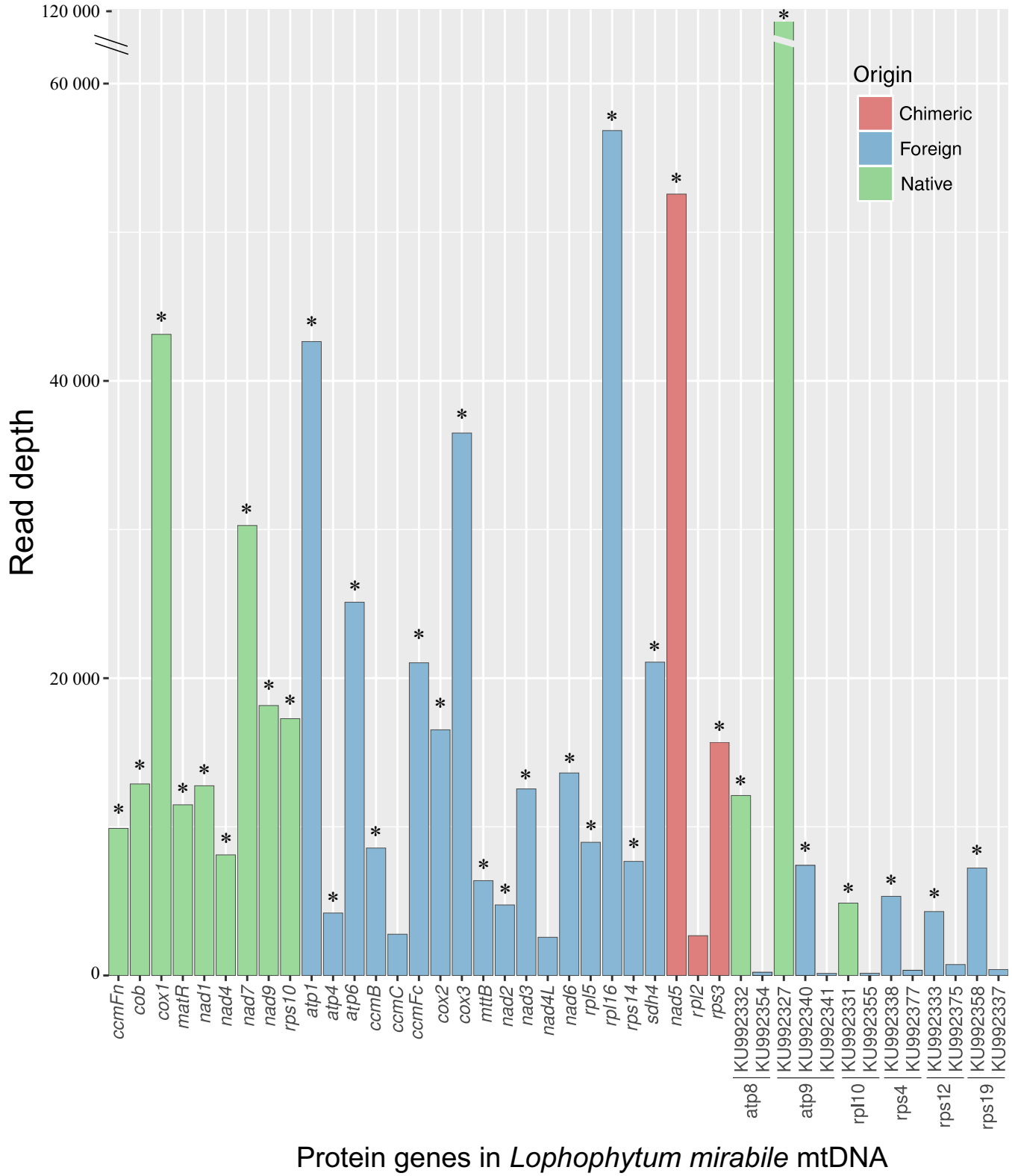

Fig. 1 RNA read depths of intact mitochondrial protein genes in *Lophophytum mirabile*. *, $P \leq 0.05$ level. For the two copies of *rps12*, average read depth is shown only for their polymorphic regions. Chromosome numbers are given for genes with multiple copies. The (interrupted) read depth for *atp9*_KU992327 is 118 967. mtDNA, mitochondrial DNA.

New Phytologist

Table 1 RNA read depth and RNA-editing data for intact mitochondrial protein genes in *Lophophytum mirabile*.

Gene	Origin[a]	Average read depth	*P*-value on read depth[b]	No. 1st and 2[b] codon position editing sites		No. introns, if any[d]	Coding region length (bp)	Chromosome location
				Total	> 70% editing[c]			
atp1	Foreign	42 644	**0.006**	3	2		1527	KU992362 (6163–7689)
atp4	Foreign	4201	**0.050**	6	6		579	KU992379 (1441–2019)
atp6	Foreign	25 110	**0.013**	18	18		1200	KU992360 (7945–9144)
atp8	Native	12 106	**0.025**	5	4		528	KU992332 (10 069–9542)
atp8	Foreign	217	0.362	3	0		447	KU992354 (1–447)
atp9	Native	118 967	**0.001**	1	1		225	KU992327 (15 163–15 387)
atp9	Foreign	7412	**0.034**	2	1		225	KU992340 (8154–7930)
atp9	Foreign	141	0.561	0	0		225	KU992341 (8428–8652)
ccmB	Foreign	8572	**0.031**	30	22		621	KU992333 (6769–6149)
ccmC	Foreign	2775	0.060	30	21		744	KU992330 (4511–5254)
ccmFc	Foreign	21 035	**0.015**	20	15	1	1305	KU992352 (10 682–8693)
ccmFn	Native	9893	**0.028**	24	21		1755	KU992372 (2129–3883)
cob	Native	12 885	**0.022**	14	11		1179	KU992375 (4576–5754)
cox1	Native	43 131	**0.006**	9	9	1	1587	KU992332 (5294–2700)
cox2	Foreign	16 526	**0.018**	13	11		771	KU992333 (8312–9082)
cox3[e]	Foreign	36 491	**0.007**	9	8		798	KU992331 (7413–8210)
matR	Native	11 481	**0.026**	11	10		1818	KU992322 (43 621–41 804)
mttB	Foreign	6379	**0.037**	31	13		807	KU992364 (2771–1965)
nad1	Native	12 763	**0.023**	22	21	4	978	KU992331 (10 277–10 663), KU992371 (4930–3671), KU992322 (44 315–44 257), KU992344 (11 098–11 356)
nad2	Foreign	4744	**0.046**	33	20	4	1470	KU992343 (8773–7268), KU992361 (4192–8592)
nad3	Foreign	12 552	**0.023**	14	12		363	KU992375 (2972–3334)
nad4	Native	8112	**0.032**	44	40	3	1485	KU992344 (9099–2689)
nad4L	Foreign	2568	0.063	9	8		303	KU992379 (810–1112)
nad5	Chimeric	52 567	**0.004**	20	20	4	2013	KU992327 (15 580–17 785), KU992342 (6250–6271), KU992334 (10 771–12 598)
nad6	Foreign	13 623	**0.021**	9	8		618	KU992351 (7605–8222)
nad7	Native	30 272	**0.010**	25	25	4	1176	KU992333 (99–5806)
nad9	Native	18 159	**0.016**	9	8		576	KU992375 (1656–2231)
rpl2	Chimeric	2676	0.061	1	1	1	1047	KU992364 (5296–3416)
rpl5	Foreign	8958	**0.030**	12	6		558	KU992331 (3708–4265)
rpl10	Native	4868	**0.044**	7	6		438	KU992331 (8965–9402)
rpl10	Foreign	151	0.528	2	0		492	KU992355 (7825–8316)
rpl16	Foreign	56 841	**0.004**	6	4		474	KU992358 (3505–3978)
rps3	Chimeric	15 683	**0.019**	3	3	1	1752	KU992358 (613–3572)
rps4	Foreign	5324	**0.041**	20	17		1056	KU992338 (12 495–13 450)
rps4	Foreign	350	0.236	14	6		1053	KU992377 (5867–6919)
rps10	Native	17 276	**0.016**	2	2	1	357	KU992332 (6539–5443)
rps12-full[f]	Foreign		**0.035**	7	4		450	KU992375 (3383–3832)
rps12-3′[f]	Foreign	771	0.125				82	KU992375 (3751–3832)
rps12-full[f]	Foreign		**0.034**	6	4		471	KU992333 (12 228–12 698)
rps12-3′[f]	Foreign	4120	**0.049**				103	KU992333 (12 596–12 698)
rps14	Foreign	7675	**0.033**	1	0		303	KU992331 (4269–4571)
rps19	Foreign	8678	**0.030**	4	2		291	KU992358 (311–601)
rps19	Foreign	567	0.161	2	1		300	KU992337 (7180–7479)
sdh4[e]	Foreign	21 078	**0.015**	6	4		402	KU992331 (8138–8539)

[a]Based on Sanchez-Puerta *et al*. (2017) and Roulet *et al*. (2020).
[b]$P \leq 0.05$ are in bold.
[c]Number of first and second codon position editing sites at which ≥ 70% of RNA reads show C-to-U editing.
[d]*nad1* has three *trans*-spliced introns, *nad2* has one, and *nad5* has two. All other introns are *cis*-spliced. Genes with foreign or chimeric introns are *ccmFc* (one such intron), *nad2* (four), *nad5* (three), and *rps3* (one).
[e]The *cox3* and *sdh4* copies located on chromosome KU992323 are not included (see Supporting Information Notes S2).
[f]Read depth over much of the length of the two intact *rps12* genes could not be distinguished because of cross-mapping of reads over their 382 bp region of 100% identity. The polymorphic 3′ regions included here were used to estimate read depth of each copy separately (Fig. 2).

≥ 70% of RNA reads show C-to-U editing at a given site) at most nonsynonymous editing sites (Table 1; Table S5), thus supporting the functional role of these genes. In addition, and of particular importance, the three putatively functional genes (*ccmC*, *nad4L*, and *rpl2)* with relatively low transcript levels (Table 1) are efficiently edited at > 90% of phylogenetically conserved editing sites (Fig. S5; Table S5). This reinforces the functional role already attributed herein to these genes.

For five of the six protein genes present in two or three intact copies (Notes S6), editing efficiency helped distinguish between functional and nonfunctional copies. Poorly transcribed copies are inefficiently edited, with little or no editing (Fig. 2; Table S7) at evolutionarily conserved positions across angiosperms (Fig. S5). These results confirm the nonfunctional nature of duplicate genes with insignificant transcript levels. The two copies of *atp9* with significant transcript levels display differential editing efficiency (Notes S6) such that, in conjunction with the 16-fold higher transcript level of native *atp9*, this gene probably produces the overwhelming majority of fully functional *atp9* messenger RNAs (mRNAs).

Clusters of co-transcribed genes

We identified nine clusters of intact genes that may be co-transcribed based on their close positioning and a continuously elevated RNA read depth throughout the cluster (Fig. S2). Five of these gene clusters (*rps10-cox1*, *nad4L-atp4*, *rps19-rps3-rpl16*, *cox3-sdh4*, *rpl5-rps14*) are widely present among angiosperms (Richardson *et al.*, 2013) and are co-transcribed where examined. These clusters are therefore highly likely to be co-transcribed in *Lophophytum*. Other clusters appear to be *Lophophytum* specific, and therefore northern blot or reverse transcription PCR analysis is desirable to confirm co-transcription. These clusters include three two-gene clusters (*atp9-nad5x1x2*, *rpl2-mttB*, *nad9-nad3*) and a large cluster (*rpl5-rps14-ORF166-cox3-sdh4-rpl10-nad1x1*) that encompasses two of the evolutionarily conserved clusters already listed.

Physical replacement of native genes by foreign genes

Four of the aforementioned gene clusters have a mixed evolutionary history. Two clusters (*rpl2-mttB* and *rps19-rps3-rpl16*) consist of foreign and chimeric genes, and two (*nad9-nad3* and *rpl5-rps14-ORF166-cox3-sdh4-rpl10-nad1x1*) consist of foreign and native genes. These findings raise the possibility that part or all of a native gene was erased by foreign DNA that served as donor (converting sequence) in one or more gene-conversion-like events. Strong evidence in support of this hypothesis comes from comparison of mtDNAs of *Lophophytum* and *Ombrophytum subterraneum*, a holoparasitic member of the sister genus to *Lophophytum* (Roulet *et al.*, 2020). These genomes are syntenic for the first three of the aforementioned four clusters, with synteny extended for the *nad9-nad3* cluster to include *rps12* and *cob*. An additional syntenic cluster consists of three genes (*ccmB-sdh3Ψ-cox2*) that are not co-transcribed. All four syntenically conserved clusters consist entirely of native genes in *Ombrophytum* (Roulet *et al.*, 2020). To our knowledge, three of these four clusters do not exist in any other angiosperm mtDNA, and the *rps19-rps3-rpl16* cluster is absent from mimosoid legumes (the likely donors of the foreign sequences in this cluster) and their caesalpinioid relatives. These observations lead us to conclude that *Lophophytum* mtDNA once possessed fully native forms of these four gene clusters, that these clusters were not present in any legume donors, and, therefore, that the foreign regions in the clusters most likely replaced the cognate native regions by one or more conversion events. To account for the chimeric nature of these four gene clusters, we postulate (Notes S7) a minimum of seven conversion events (the actual number of conversions could be much higher) whereby nine foreign genes partially or fully replaced their native homologs.

Missing native genes were not functionally transferred to the nucleus of *Lophophytum*

The *de novo* assembled transcriptome of *Lophophytum* consists of 1 068 587 contigs of ≥ 200 bp, with 172 469 ORFs ≥ 300 bp identified using TransDecoder. Blast searches against this assembly failed to find nuclear homologs of any of the intact mitochondrial genes of *Lophophytum*. However, an intact nuclear copy of the mitochondrial *sdh3* pseudogene was recovered, with this gene having been acquired by nuclear-to-nuclear HGT from mimosoid legumes (Fig. S6; Notes S8).

To validate the negative results, we assessed the completeness of the *Lophophytum* transcriptome using the Busco v.3 pipeline (Waterhouse *et al.*, 2018). We obtained full-length transcript coverage for 93.4%, 61.73% and 63.4% of the typically single-copy genes that comprise the conserved orthogroups of Eukaryotes (303), Embryophyta_odb9 (1440), and Eudicotyledons_odb10 (2121 genes), respectively. These results are to be expected for a relatively complete transcriptome of a holoparasitic plant, in which most genes involved in photosynthesis-related functions were probably lost. Furthermore, the vegetative form of *Lophophytum* is a highly derived underground tuber that lacks leaves, stomata, epidermis, and trichomes (Gonzalez & Sato, 2016); therefore, many genes related to these structures were probably lost too.

Discussion

Massive functional replacement of native mitochondrial genes by foreign and chimeric homologs

Twenty-three of the 30 full-length foreign and chimeric protein genes in *Lophophytum* mtDNA possess the RNA-level attributes expected for functional mitochondrial genes. In 20 cases, transcript levels are significantly higher than the background level of this entirely transcribed genome, whereas the three genes with marginally significant transcript levels in *Lophophytum* also show relatively low RNA levels in other plants (Picardi *et al.*, 2010; Islam *et al.*, 2013; Grewe *et al.*, 2014; Grimes *et al.*, 2014; Wu *et al.*, 2015b). These 23 genes contain a total of nine foreign or chimeric introns, all of which are accurately spliced. These genes

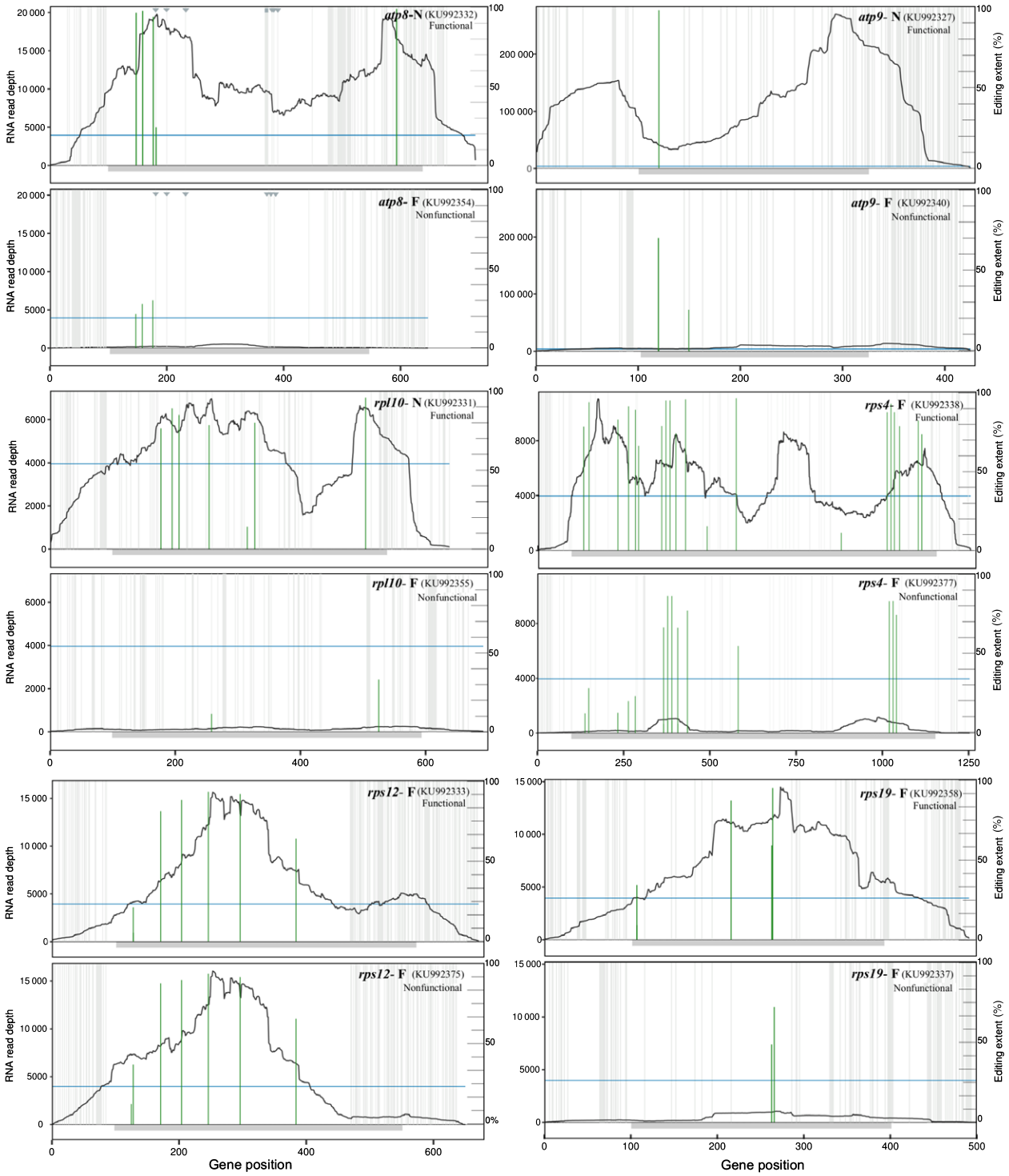

Fig. 2 Expression of six duplicated protein genes in the *Lophophytum mirabile* mitogenome. The black curves plot RNA read depth for each intact duplicated gene (grey horizontal bar) and the 100 nucleotides flanking it on both sides (not included is the third copy of *atp9*, KU992341, with very low read depth). Blue horizontal lines indicate the 3960 read depth threshold for significance at the $P \leq 0.05$ level. Editing efficiency at nonsynonymous editing sites is indicated by the green vertical lines (see the scale on the right side of each plot). Gray vertical lines mark nucleotide and insertional differences between gene copies, whose alignments are thrown out of register after each indel. For purposes of illustration, the first six single-nucleotide variants within the gene *atp8* are marked by triangles; the first three variants are in register, whereas the second three are out of register owing to the 6 bp indel (marked by the rectangle in the top *atp8* plot) located between the two sets of variants. Gene origin is indicated by N (native) or F (foreign) after each gene name.

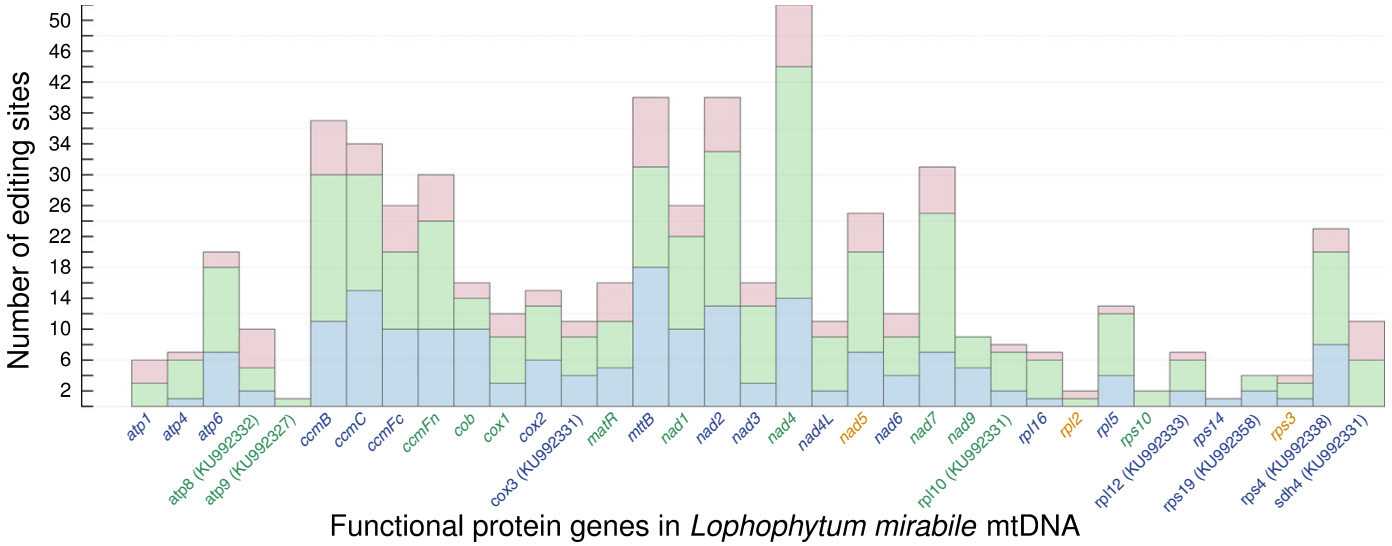

Fig. 3 Number of C-to-U editing sites in the 35 putatively functional protein genes in the *Lophophytum mirabile* mitogenome. Blue indicates the number of editing sites at first-codon positions, green at second, and pink at third. Gene names are colored according to their phylogenetic origin (blue, foreign; green, native; orange, chimeric). Chromosome numbers are given for genes with multiple copies.

are efficiently edited at 224 nonsynonymous editing sites, most of which are present as editing sites across angiosperms. Further support for functionality comes from analysis of synonymous and nonsynonymous substitution rates on the *Lophophytum* branch, which indicates that some of these foreign genes have continued to evolve under purifying selection (Sanchez-Puerta *et al.*, 2017).

The *Lophophytum* mitogenome lacks native copies for all 23 of these intact and expressed foreign and chimeric protein genes (Sanchez-Puerta *et al.*, 2017). Nor did we detect nuclear transcripts for any of these genes, or for any of the 12 intact and expressed native mitochondrial genes, under conditions of effective transcript recovery for genes expected to be present in the *Lophophytum* nucleus. Therefore, it is likely that these 35 genes are present only in the mitogenome. In sum, all evidence leads to the same fundamental conclusion, that 23 legume-derived foreign genes have functionally replaced their native homologs in the mitogenome of *Lophophytum*. A corollary conclusion follows: unlike the mistletoe hemiparasite *Viscum* (Skippington *et al.*, 2015, 2017) and many aerobic and anaerobic protistan parasites (Zíková *et al.*, 2016; Santos *et al.*, 2018), the loss of native mitochondrial genes from the holoparasite *Lophophytum* is not a symptom of loss of biochemical function.

Although we have not examined expression at the protein level in *Lophophytum*, given the evidence herein, we regard it as a foregone conclusion that these 23 foreign mRNAs are translated into functional proteins.

To our knowledge, replacement HGT of the scale found in *Lophophytum*, involving nearly two-thirds of a genome's protein genes, is unprecedented for any genome or organism. This even includes prokaryotes, where HGT is a dominating evolutionary force, but one that mostly adds new kinds of genes and gene functions rather than replacing pre-existing genes.

Lophophytum overcame all potential barriers to foreign gene expression

There are many potential barriers at the level of gene expression to a foreign gene functionally supplanting its native counterpart. *Lophophytum* faced the challenges of recognizing foreign promoters, accurately splicing foreign introns, and editing numerous foreign editing sites.

Transcription initiation in angiosperm mitochondria shows few specificity constraints; genes are transcribed from promoters with either core or deviant motifs and often from multiple promoters (Kuhn *et al.*, 2005). Despite this flexibility, only conserved nine-nucleotide promoter sequences were recognized in a heterologous *in vitro* transcription system (Binder *et al.*, 1995). In *Lophophytum*, three functional foreign genes are preceded by long foreign regions, and therefore are probably transcribed using foreign promoters (Notes S9). By contrast, transcription of at least one functional foreign gene is likely initiated at a native promoter. The promoter situation for other functional foreign genes is unclear. One nonfunctional foreign gene is also preceded by a long foreign region, raising the possibility that a subset of legume-derived promoters is ineffectively recognized by the *Lophophytum* transcriptional machinery.

The 25 group II introns present in the ancestral angiosperm mitogenome have all lost the ability to self-splice and are dependent on many, mostly nuclear-encoded, splicing factors to achieve accurate splicing (Brown *et al.*, 2014). Accurate splicing in foreign mitochondria, occurring reciprocally between monocots and eudicots, has been shown for two of these introns (Bolle & Kempken, 2006; Choury & Araya, 2006). Our results extend these findings to the nine introns that *Lophophytum* acquired, in whole or in part, from legumes. This raises the possibility that splicing determinants, both *cis* and *trans*-acting, are conserved

across angiosperms for most, if not all, of their 25 group II introns.

Transcript levels for most of the 24 introns present in *Lophophytum* mtDNA are comparable to those of their flanking exons, with only a single intron showing a markedly reduced level. This finding surprised us, as previous evidence indicates that spliced-out introns are present in low abundance, presumably due to rapid degradation (Grimes *et al.*, 2014; Wu *et al.*, 2015b; Lloyd Evans *et al.*, 2019). This unexpected finding merits follow-up study.

Most sites of C-to-U RNA editing that are highly conserved in angiosperms are either C in *Lophophytum* mtDNA and efficiently edited or have changed to T, which equally well results in the proper amino acid. Given the substantial variation in mitochondrial editing content among angiosperms (Richardson *et al.*, 2013; Edera *et al.*, 2018) and the distant relationship between *Lophophytum* and legumes, it is likely that the host-acquired functional genes contain, in aggregate, numerous editing sites that were not present in the native genes they replaced and which are nonetheless edited in *Lophophytum*. These findings are in line with studies showing that exogenous editing sites are successfully edited when electroporated into phylogenetically distant mitochondria (Bolle & Kempken, 2006; Choury & Araya, 2006) or when the nuclear background of a mitogenome is changed (Wu *et al.*, 2019). Given that individual editing factors can act on multiple editing sites in plant mitochondria (Glass *et al.*, 2015), pre-existing factors in *Lophophytum* could have operated on host-specific editing sites introduced by HGT.

In all six cases examined where *Lophophytum* contains multiple intact copies of a protein gene, only one copy is a strong candidate for being functional based on transcript levels and/or RNA editing data. This situation is not unexpected, as it should minimize gene dosage effects (Sorek *et al.*, 2007).

Complete but largely spurious transcription of the *Lophophytum* mitogenome

The *Lophophytum* mitogenome is entirely transcribed. This finding was not unexpected, as transcription initiation and termination are largely unconstrained in plant mitochondria (Holec *et al.*, 2008) and RNA-seq studies have shown substantial transcription of intergenic DNA in mitogenomes from a diversity of angiosperms (Fujii *et al.*, 2011; Grimes *et al.*, 2014; Wu *et al.*, 2015b; Silva *et al.*, 2017; Lloyd Evans *et al.*, 2019), with two of these genomes completely transcribed (Lloyd Evans *et al.*, 2019). Our study extends these findings by showing that both strands of an angiosperm mitochondrial genome are pervasively transcribed. This may be the case in other plants, where transcription of both strands has been shown for a few regions (Holec *et al.*, 2006; Wu *et al.*, 2015b; Lloyd Evans *et al.*, 2019), but previous RNA-seq studies have not addressed this issue on a genome scale.

The overwhelming majority of the intergenic sequences in *Lophophytum* mtDNA (which comprise over 90% of the genome) show nonsignificant transcript levels, with these levels averaging over 140-fold lower than transcript levels for the 35 functional protein genes. Thirty-two of the 60 mitochondrial chromosomes contain no significantly expressed regions. It is possible that some of these 32 chromosomes harbor important RNA species of unidentified function, with such RNAs, like tRNAs and 5S rRNA, too short to be recovered by our RNA-isolation procedure. More likely, these are empty neutral chromosomes evolving according to the whims of genetic drift and on their way to extinction, as shown for many noncoding chromosomes in *Silene noctiflora* (Wu *et al.*, 2015a,b).

Gene replacement driven by horizontal gene transfer

Nine native protein genes in *Lophophytum* were probably lost (entirely in seven cases and in part for two chimeric genes) by gene-conversion-like events that resulted in the physical replacement of these native sequences by foreign homologs. Fig. 4 shows two pathways by which replacement conversion can occur in the context of HGT. In pathway A, as part of an HGT event itself, a newly arrived foreign gene enters the *Lophophytum* genome by replacing its native homolog through homologous recombination. In pathway B, the foreign gene integrates apart from the native gene, most likely by homologous recombination between a sequence flanking the gene and a similar sequence in the recipient genome (Cappadocia *et al.*, 2010; Chevigny *et al.*, 2020). This foreign copy eventually replaces the native gene by serving as the donor in gene-conversion-like events between the two loci, followed by loss of the foreign locus. Although both conversion pathways are compatible with our data, pathway A is more parsimonious, requiring only one event rather than three (Fig. 4). This suggests that most of these nine replacement conversions took place as the final act of the HGT event itself. These conversions probably occurred by break-induced replication, a common mechanism of homologous recombination in plant mitochondria (Davila *et al.*, 2011; Gualberto *et al.*, 2014; Garcia *et al.*, 2019). Break-induced replication repairs either double-strand breaks (DSBs), in which only a single DSB end is homologous with a donor sequence, or one-ended DSBs created by the collapse of the replication fork, resulting in both cases in a nonreciprocal translocation that can overwrite the native information (Anand *et al.*, 2013; Mehta *et al.*, 2017).

Homologous recombination between native and horizontally acquired genes has been described in many angiosperm mtDNAs (Cho *et al.*, 1998; Bergthorsson *et al.*, 2003; Barkman *et al.*, 2007; Sanchez-Puerta *et al.*, 2008; Hao *et al.*, 2010; Hepburn *et al.*, 2012; Park *et al.*, 2015; Roulet *et al.*, 2020). However, this is the first evidence in plants that this process can lead to the replacement of an entire native gene, much less do so as part of an HGT event itself. Most of the published cases involve short conversion tracts, with some chimeric genes containing many alternating regions of foreign and native DNA (Hao *et al.*, 2010; Mower *et al.*, 2010; Park *et al.*, 2015). We suspect, therefore, that, in addition to the relatively large conversional replacements of native genes by foreign ones reported here, *Lophophytum* mtDNA contains a number of smaller, less obvious conversions and that these occurred in both directions between foreign and native copies. To test this hypothesis requires sequencing mtDNAs from many more legumes and Balanophoraceae

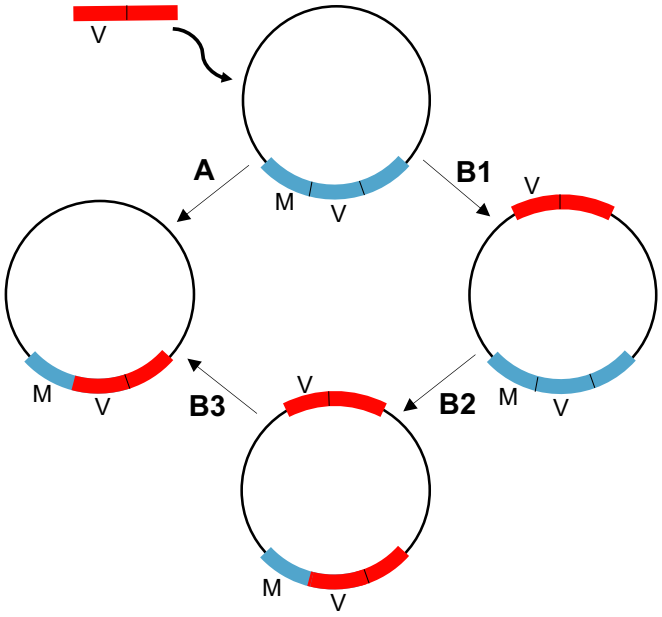

Fig. 4 Pathways for physical replacement of native genes by foreign counterparts. In pathway A, a newly arrived foreign region containing gene V plus flanking spacer DNA (all in red) replaces the cognate native region (in blue) through homologous recombination, most likely break-induced replication-mediated gene conversion. In pathway B, the foreign V region integrates apart from the native region (B1), eventually replacing it by serving as the donor in homologous recombination between the two loci (B2) and is ultimately lost (B3). M, a gene adjacent to V that stays native.

members and close visual examination of alignments to detect small-scale conversions, which are largely refractory to recombination-detection programs in the context of the low substitution rates in most angiosperm mitogenomes (Hao *et al.*, 2010).

Why is *Lophophytum* so exceptional?

Most foreign genes in angiosperm mitochondria are pseudogenes or poorly expressed; these derelict genes, as well as intact foreign genes, almost always coexist with intact native homologs, and there are very few cases of foreign genes that appear to be functional (see Introduction and Notes S1). The situation in *Lophophytum* mtDNA is utterly different, with most native protein genes having vanished and been functionally replaced by foreign genes from its legume hosts. Standing in sharpest contrast to *Lophophytum* is *O. subterraneum*, which belongs to the sister genus of *Lophophytum*, is also a root holoparasite, and has also experienced massive bouts of mitochondrial HGT yet retains a full set of native genes (Roulet *et al.*, 2020).

At least three nonselective factors could contribute to these stark differences. First, rates of HGT may have been exceptionally high for millions of years in the lineage leading to *L. mirabile*, with the total flux of foreign DNA substantially higher than even in *Ombrophytum* (Roulet *et al.*, 2020) or *Amborella* (Rice *et al.*, 2013). Hyper-elevated HGT could result from (1) unusually frequent mixing of the contents of haustorial cells of *Lophophytum* and root cells of its host tree, this mixing essential to the fusion of donor and recipient mitochondria thought to be the driving force of mitochondrial HGT in plants (Rice *et al.*, 2013; Sanchez-Puerta *et al.*, 2019); (2) unusually high rates of mitochondrial fusion in *Lophophytum*; and (3) elevated rates of integration of foreign DNA into the *Lophophytum* mitogenome. Second, barriers to the successful expression of foreign mitochondrial genes may be unusually low at multiple levels in *Lophophytum* mitochondria. Finally, rates of both DSBs and their repair by homologous recombination may be unusually high in *Lophophytum* mitochondria, leading to very high rates of conversional replacement of native genes by foreign homologs.

Selective forces and genetic drift are probably crucial to the *Lophophytum* situation. All 35 proteins encoded by *Lophophytum* mtDNA form coevolved partnerships within multisubunit complexes, largely those of OXPHOS and the ribosome. Therefore, legume mitochondrial proteins may not, as a rule, work as well in *Lophophytum* – because they will disturb lineage-specific coevolved fit and function – as in their native coevolved context, in which case we would expect replacement HGTs to be largely eliminated by purifying selection. This is what is seen in prokaryotes, where genes that encode members of macromolecular complexes, the ribosome being the type example, are relatively refractory to replacement HGT compared with, say, single-subunit enzymes (Jain *et al.*, 1999; Sorek *et al.*, 2007; Cohen *et al.*, 2011). So, why do we not see the same thing here? That is, why *do* we see so exceptionally much replacement HGT in *Lophophytum*? Part of the answer may lie at the feet of genetic drift, which is generally weak in bacteria but probably strong in *Lophophytum*. In addition, as a holoparasite, *Lophophytum* may be evolving under altered selective constraints on OXPHOS and mitochondrial protein synthesis.

Lophophytum mirabile reproduces predominantly if not solely by parthenogenesis (Sato & Gonzalez, 2017) and probably has a small effective population size. It might also have a history of recurrent bottlenecks, both at the population level and within an individual, with bottlenecks in numbers of both cells and mitochondria upon germination of the minute *Lophophytum* seed and the early formation of haustorial connections with host plants increasing the chances that a newly integrated legume gene is initially fixed. Strong genetic drift may therefore have led to the fixation of deleterious replacement HGTs in *Lophophytum*. However, though drift could certainly account for some deleterious gene swaps, it is hard to imagine an organism with *normally* functioning OXPHOS complexes and mitochondrial ribosomes swapping out most of its mitochondrial protein gene set without incurring some impairment of OXPHOS and protein synthesis. Perhaps these crucial mitochondrial processes were impaired *prior* to replacement HGT; indeed, many parasites have reduced respiratory capabilities (Zíková *et al.*, 2016; Santos *et al.*, 2018). If selection were relaxed on OXPHOS, the predominant consumer of mitochondrial translational activity, then constraints on mitochondrial translation may also have been relaxed. Relaxed situation could result in most HGT replacements being neutral or only mildly deleterious, and thus readily fixed by genetic drift. At the extreme, constraints on OXPHOS may be so altered and

its resulting impairment so great that many legume genes actually work better in OXPHOS and translation in *Lophophytum* than do its own genes. Such beneficial foreign genes would then be fixed by selection rather than by drift.

The many replacement HGTs in *Lophophytum* probably result from the combined, if not synergistic, action of various of the foregoing molecular, cellular, and population-genetic forces, with the relative intensity and interplay of these forces varying over time and across genes and protein complexes. We propose that the major drivers of massive replacement HGT in *Lophophytum* are very high levels of HGT, unusually frequent conversional replacement of native genes by foreign ones, strong genetic drift, and altered selection on OXPHOS.

Acknowledgements

We thank L. F. Ceriotti and M. E. Roulet for help with the computational analyses and A. J. Alverson, W. Hao, J. P. Mower, and D. B. Sloan for constructive feedback on parts of the manuscript. This work was supported by grants to MVS-P from Agencia Nacional de Promoción Científica y Tecnológica (PICT0691) and Universidad Nacional de Cuyo (06/A724). It was also supported in part by the Lilly Endowment Inc., through its support for the Indiana University Pervasive Technology Institute.

Author contributions

LEG, AAE, and MVS-P designed the study. HS identified and collected *L. mirabile*. LEG extracted and purified RNA. AAE wrote scripts for data analyses. LEG, AAE, JDP and MVS-P analyzed data and wrote the manuscript. LEG and AAE contributed equally to this work.

ORCID

Alejandro A. Edera https://orcid.org/0000-0001-9864-3159
Laura E. Garcia https://orcid.org/0000-0002-9324-6872
Jeffrey D. Palmer https://orcid.org/0000-0002-4626-2220
M. Virginia Sanchez-Puerta https://orcid.org/0000-0003-2511-5093
Hector Sato https://orcid.org/0000-0003-2070-3339

References

Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution* **29**: 380–395.

Adams KL, Qiu Y-L, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proceedings of the National Academy of Sciences, USA* **99**: 9905–9912.

Anand RP, Lovett ST, Haber JE. 2013. Break-induced DNA replication. *Cold Spring Harbor Perspectives in Biology* **5**: a0103978.

Barkman TJ, McNeal JR, Lim SH, Coat G, Croom HB, Young ND, dePamphilis CW. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evolutionary Biology* **7**: e248.

Bergthorsson U, Adams KL, Thomason B, Palmer JD. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **424**: 197–201.

Bergthorsson U, Richardson A, Young G, Goertzen L, Palmer JD. 2004. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proceedings of the National Academy of Sciences, USA* **101**: 17747–17752.

Binder S, Hatzack F, Brennicke A. 1995. A novel pea mitochondrial *in vitro* transcription system recognizes homologous and heterologous mRNA and tRNA promoters. *Journal of Biological Chemistry* **270**: 22182–22189.

Bolger A, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Bolle N, Kempken F. 2006. Mono and dicotyledonous plant-specific RNA editing sites are correctly edited in both in organello systems. *FEBS Letters* **580**: 4443–4448.

Brown GG, Colas des Francs-Small C, Ostersetzer-Biran O. 2014. Group II intron splicing factors in plant mitochondria. *Frontiers Plant Science* **5**: e35.

Cappadocia M, Marechal A, Parent J, Lepage E, Sygusch J, Brisson N. 2010. Crystal structures of DNA-whirly complexes and their role in *Arabidopsis* orgenelle genome repair. *Plant Cell* **22**: 1849–1867.

Chevigny N, Schatz-Daas D, Lofti F, Gualberto JM. 2020. DNA repair and the stability of the plant mitochondrial genome. *International Journal of Molecular Sciences* **21**: e328.

Cho Y, Qiu YL, Kuhlman P, Palmer JD. 1998. Explosive invasion of plant mitochondria by a group I intron. *Proceedings of the National Academy of Sciences, USA* **95**: 14244–14249.

Choury D, Araya A. 2006. RNA editing site recognition in heterologous plant mitochondria. *Current Genetics* **50**: 405–416.

Cohen O, Gophna U, Pupko T. 2011. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Molecular Biology and Evolution* **28**: 1481–1489.

Davila J, Arrieta-Montiel M, Wamboldt Y, Cao J, Hagmann J, Shedge V, Xu Y, Weigel D, Mackenzie S. 2011. Double-strand break repair processes drive evolution of the mitochondrial genome in *Arabidopsis*. *BMC Biology* **9**: e64.

Davis C, Andersen W, Wurdack K. 2005. Gene transfer from a parasitic flowering plant to a fern. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **272**: 2237–2242.

Davis C, Wurdack K. 2004. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science* **305**: 676–678.

Edera A, Gandini CL, Sanchez-Puerta MV. 2018. Towards a comprehensive picture of RNA editing in angiosperm mitochondria. *Plant Molecular Biology* **97**: 215–231.

Forgione I, Bonavita S, Regina TMR. 2019. Mitochondria of *Cedrus atlantica* and allied species: a new chapter in the horizontal gene transfer history. *Plant Science* **281**: 93–101.

Forner J, Weber B, Thuss S, Wildum S, Binder S. 2007. Mapping of mitochondrial mRNA termini in *Arabidopsis thaliana*: t-elements contribute to 5′ and 3′ end formation. *Nucleic Acids Research* **35**: 3676–3692.

Fujii S, Toda T, Kikuchi S, Suzuki R, Yokoyama K, Tsuchida H, Yano K, Toriyama K. 2011. Transcriptome map of plant mitochondria reveals islands of unexpected transcribed regions. *BMC Genomics* **12**: e279.

Garcia L, Zubko MK, Zubko EI, Sanchez-Puerta MV. 2019. Elucidating genomic patterns and recombination events in plant cybrid mitochondria. *Plant Molecular Biology* **100**: 433–450.

Glass F, Härtel B, Zehrmann A, Verbitskiy D, Takenaka M. 2015. MEF13 requires MORF3 and MORF8 for RNA editing at eight targets in mitochondrial mRNAs in *Arabidopsis thaliana*. *Molecular Plant* **8**: 1466–1477.

Gonzalez A, Sato H. 2016. Anatomía vegetativa de *Lophophytum mirabile* subsp. *bolivianum* (Balanophoraceae) y efecto de su parasitismo en la anatomía de las raíces de su hospedante *Anadenanthera colubrina* var. *cebil*. *Anales del Jardín Botánico de Madrid* **73**: e039.

Grewe F, Edger P, Keren I, Sultan L, Pires JC, Ostersetzer-Biran O, Mower JP. 2014. Comparative analysis of 11 Brassicales mitochondrial genomes and the mitochondrial transcriptome of *Brassica oleracea*. *Mitochondrion* **19B**: 135–143.

Grimes BT, Sisay AK, Carroll HD, Cahoon A. 2014. Deep sequencing of the tobacco mitochondrial transcriptome reveals expressed ORFs and numerous editing sites outside coding regions. *BMC Genomics* **15**: e31.

Gualberto JM, Mileshina D, Wallet C, Niazi A, Weber-Lotfi F, Dietrich A. 2014. The plant mitochondrial genome: dynamics and maintenance. *Biochimie* **100**: 107–120.

Hao W, Richardson AO, Zheng Y, Palmer JD. 2010. Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. *Proceedings of the National Academy of Sciences, USA* **14**: 21576–21581.

Havird J, Trapp P, Miller C, Bazos I, Sloan DB. 2017. Causes and consequences of rapidly evolving mtDNA in a plant lineage. *Genome Biology and Evolution* **9**: 323–336.

Hepburn N, Schmidt D, Mower JP. 2012. Loss of two introns from the *Magnolia tripetala* mitochondrial *cox2* gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss. *Molecular Biology and Evolution* **29**: 311–320.

Hirayama T, Matsuura T, Ushiyama S, Narusaka M, Kurihara Y, Yasuda M, Ohtani M, Seki M, Demura T, Nakashita H *et al.* 2013. A poly(A)-specific ribonuclease directly regulates the poly(A) status of mitochondrial mRNA in *Arabidopsis*. *Nature Communications* **4**: 1212–1215.

Holec S, Lange H, Canaday J, Gagliardi D. 2008. Coping with cryptic and defective transcripts in plant mitochondria. *Biochimica et Biophysica Acta* **1779**: 566–573.

Holec S, Lange H, Kuhn K, Alioua M, Borner T, Gagliardi D. 2006. Relaxed transcription in *Arabidopsis* mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase. *Molecular and Cellular Biology* **26**: 2869–2876.

Islam M, Studer B, Byrne SL, Farrell JD, Panitz F, Bendixen C, Møller I, Asp T. 2013. The genome and transcriptome of perennial ryegrass mitochondria. *BMC Genomics* **14**: e202.

Jain R, Rivera MC, Lake JA. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proceedings of the National Academy of Sciences, USA* **96**: 3801–3806.

Kim G, LeBlanc ML, Wafula E, dePamphilis CW, Westwood JH. 2014. Genomic-scale exchange of mRNA between a parasitic plant and its hosts. *Science* **345**: 808–811.

Kopylova E, Noe L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**: 3211–3217.

Kuhn J, Tengler U, Binder S. 2001. Transcript lifetime is balanced between stabilizing stem–loop structures and degradation-promoting polyadenylation in plant mitochondria. *Molecular and Cellular Biology* **21**: 731–742.

Kuhn K, Weihe A, Borner T. 2005. Multiple promoters are a common feature of mitochondrial genes in *Arabidopsis*. *Nucleic Acids Research* **33**: 337–346.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

Lloyd Evans D, Hlongwane TT, Joshi SV, Riaño Pachón DM. 2019. The sugarcane mitochondrial genome: assembly, phylogenetics and transcriptomics. *PeerJ.* **7**: e7558.

Maclean A, Hertle A, Ligas J, Bock R, Balk J, Meyer E. 2018. Absence of complex I is associated with diminished respiratory chain function in European mistletoe. *Current Biology* **28**: 1614–1619.

Mehta A, Beach A, Haber JE. 2017. Homology requirements and competition between gene conversion and break-induced replication during double-strand break repair. *Molecular Cell* **65**: 515–526.

Mower JP, Stefanovic S, Hao W, Gummow JS, Jain K, Ahmed D, Palmer JD. 2010. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. *BMC Biology* **8**: e150.

Mower JP, Stefanovic S, Young GJ, Palmer JD. 2004. Plant genetics: gene transfer from parasitic to host plants. *Nature* **432**: 165–166.

Park S, Grewe F, Zhu A, Ruhlman T, Sabir J, Mower JP, Jansen RK. 2015. Dynamic evolution of *Geranium* mitochondrial genomes through multiple horizontal and intracellular gene transfers. *New Phytologist* **208**: 570–583.

Picardi E, Horner DS, Chiara M, Schiavon R, Valle G, Pesole G. 2010. Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. *Nucleic Acids Research* **38**: 4755–4767.

Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, dePamphilis CW *et al.* 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* **342**: 1468–1473.

Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. 2013. The 'fossilized' mitochondrial genome of *Liriodendron tulipifera*: an ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biology* **11**: e29.

Roulet ME, Garcia LE, Gandini CL, Sato H, Ponce G, Sanchez-Puerta MV. 2020. Multichromosomal structure and foreign tracts in the *Ombrophytum subterraneum* (Balanophoraceae) mitochondrial genome. *Plant Molecular Biology* **103**: 623–638.

Sanchez-Puerta MV, Cho Y, Mower JP, Alverson AJ, Palmer JD. 2008. Frequent, phylogenetically local horizontal transfer of the *cox1* group I intron in flowering plant mitochondria. *Molecular Biology and Evolution* **25**: 1762–1777.

Sanchez-Puerta MV, Edera A, Gandini CL, Williams AV, Howell KA, Nevill PG, Small I. 2019. Genome-scale transfer of mitochondrial DNA from legume hosts to the holoparasite *Lophophytum mirabile* (Balanophoraceae). *Molecular Phylogenetics and Evolution* **132**: 243–250.

Sanchez-Puerta MV, García LE, Wohlfeiler J, Ceriotti F. 2017. Unparalleled replacement of native mitochondrial genes by foreign homologs in a holoparasitic plant. *New Phytologist* **214**: 376–387.

Santos H, Makiuchi T, Nozaki T. 2018. Reinventing an organelle: the reduced mitochondrion in parasitic protists. *Trends in Parasitology* **34**: 1038–1055.

Sato H, Gonzalez A. 2017. Embryogenesis, endospermogenesis and fruit development in *Lophophytum* (Balanophoraceae): focus on endosperm and embryo initiation. *Flora* **233**: 79–89.

Schneider A, Chun H, Stefanovic S, Baldwin BG. 2018. Punctuated plastome reduction and host–parasite horizontal gene transfer in the holoparasitic plant genus *Aphyllon*. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **285**: e20181535.

Senkler J, Rugen N, Eubel H, Hegermann J, Braun HP. 2018. Absence of complex I implicates rearrangements of the respiratory chain in European mistletoe. *Current Biology* **28**: 1606–1613.

Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**: e0163962.

Sievers F, Wilm A, Dineen D, Gibson T, Karplus K, Li W. 2011. Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**: e539.

Silva SR, Alvarenga DO, Aranguren Y, Penha HA, Fernandes CC, Pinheiro DG, Oliveira MT, Michael TP, Miranda VFO, Varani AM. 2017. The mitochondrial genome of the terrestrial carnivorous plant *Utricularia reniformis* (Lentibulariaceae): structure, comparative analysis and evolutionary landmarks. *PLoS ONE* **12**: e0180484.

Sinn B, Barrett C. 2020. Ancient mitochondrial gene transfer between fungi and the orchids. *Molecular Biology and Evolution* **37**: 44–57.

Skippington E, Barkman TJ, Rice DW, Palmer JD. 2015. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad* genes. *Proceedings of the National Academy of Sciences, USA* **112**: E3515–E3524.

Skippington E, Barkman TJ, Rice DW, Palmer JD. 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 Biology* **17**: e49.

Small I, Schallenberg-Rudinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. 2019. Plant organellar RNA editing: what 30 years of research has revealed. *The Plant Journal* **101**: 1040–1056.

Sorek R, Zhu Y, Creevey C, Francino M, Bork P, Rubin E. 2007. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**: 1449–1452.

Stone JD, Storchova H. 2015. The application of RNA-seq to the comprehensive analysis of plant mitochondrial transcriptomes. *Molecular Genetics and Genomics* **290**: 1–9.

Tsujimura M, Kaneko T, Sakamoto T, Kimura S, Shigyo M, Yamagishi H, Terachi T. 2019. Multichromosomal structure of the onion mitochondrial genome and a transcript analysis. *Mitochondrion* **46**: 179–186.

Varré J, D'Agostino N, Touzet H, Gallina S, Tamburino R, Cantarella C, Ubrig E, Cardi T, Drouard L, Gualberto JM *et al.* 2019. Complete sequence, multichromosomal architecture and transcriptome analysis of the *Solanum tuberosum* mitochondrial genome. *International Journal of Molecular Sciences* **20**: e4788.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* **35**: 543–548.

Wickell DA, Li FW. 2020. On the evolutionary significance of horizontal gene transfers in plants. *New Phytologist* **225**: 113–117.

Won H, Renner S. 2003. Horizontal gene transfer from flowering plants to *Gnetum*. *Proceedings of the National Academy of Sciences, USA* **100**: 10824–10829.

Wu Z, Cuthbert JM, Taylor DR, Sloan DB. 2015a. The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. *Proceedings of the National Academy of Sciences, USA* **112**: 10185–10191.

Wu Z, Hu K, Yan M, Song L, Wen J, Ma C, Shen J, Fu T, Yi B, Tu J. 2019. Mitochondrial genome and transcriptome analysis of five alloplasmic male-sterile lines in *Brassica juncea*. *BMC Genomics* **20**: e348.

Wu Z, Stone JD, Štorchová H, Sloan DB. 2015b. High transcript abundance, RNA editing, and small RNAs in intergenic regions within the massive mitochondrial genome of the angiosperm *Silene noctiflora*. *BMC Genomics* **16**: e938.

Xi Z, Wang Y, Bradley R, Sugumaran M, Marx C, Rest J, Davis CC. 2013. Massive mitochondrial gene transfer in a parasitic flowering plant clade. *PLoS Genetics* **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. *Proceedings of the National Academy of Sciences, USA* **113**: E7010–E7019.

Zeng Y, Yang T. 2012. RNA isolation from highly viscous samples rich in polyphenols and polysaccharides. *Plant Molecular Biology Reporter* **20**: 417.

Zíková A, Hampl V, Paris Z, Tyc J, Lukes J. 2016. Aerobic mitochondria of parasitic protists: Diverse genomes and complex functions. *Molecular and Biochemical Parasitology* **209**: 46–57.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.**Dataset S1** Alignments of each protein gene in the *Lophophytum mirabile* mitogenome.

Fig. S1 Reanalysis of the phylogenetic history of the *cox1* gene of *Rafflesia cantleyi*.

Fig. S2 RNA read depth of the 60 chromosomes in the *Lophophytum mirabile* mitogenome.

Fig. S3 RNA read depth of intact protein genes in *Lophophytum mirabile* mtDNA.

Fig. S4 Editing efficiency at C-to-U editing sites in the 35 putatively functional protein genes in *Lophophytum mirabile* mtDNA.

Fig. S5 Phylogenetic conservation of sites of C-to-U RNA editing at first and second codon positions in angiosperm mitochondrial genes.

Fig. S6 Foreign origin of the nuclear *sdh3* gene in *Lophophytum mirabile*.

Notes S1 Reassessment of published studies on putatively functional HGT in plant mitochondria.

Notes S2 Differences in mtDNA of the samples used for DNAseq and RNAseq.

Notes S3 Long non-coding RNAs in *Lophophytum* mtDNA.

Notes S4 Unidentified ORFs in *Lophophytum* mtDNA.

Notes S5 Nonsignificant expression of the homing endonuclease-encoding intron in *cox1*.

Notes S6 Differential editing efficiency of duplicated genes.

Notes S7 Minimum number of gene replacement events.

Notes S8 Evolution of nuclear-encoded *sdh3* in *Lophophytum*.

Notes S9 Promoters.

Table S1 RNA read depth of the 60 mitochondrial chromosomes in *Lophophytum mirabile*.

Table S2 Location and RNA read depth of long non-coding RNAs (lncRNAs) in *Lophophytum mirabile* mtDNA.

Table S3 C-to-U RNA editing sites in the unidentified ORF166 located between positions 5279 and 5776 in strand 1 of chromosome KU992331.

Table S4 RNA read depth of intact genes and unidentified ORFs in *Lophophytum mirabile* mtDNA.

Table S5 C-to-U RNA editing sites in putatively functional protein genes in *Lophophytum mirabile* mtDNA.

Table S6 C-to-U RNA editing sites in untranslated regions (UTRs) of mitochondrial genes in *Lophophytum mirabile*.

Table S7 C-to-U RNA editing sites in intact but putatively non-functional genes in *Lophophytum mirabile* mtDNA.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.