Coding For Medicine Club
Houston, Texas Headquarters
550 Dulles Ave, Sugar Land 77478
Dulles High School

Novel genetic code and record-setting AT-richness in the highly reduced plastid genome of the holoparasitic plant Balanophora - Huei-Jiun Su, Todd J. Barkman, Weilong Hao, Samuel S. Jones, Julia Naumann, Elizabeth Skippington, Eric K. Wafula, Jer-Ming Hu, Jeffrey D. Palmer, and Claude W. dePamphilis

Name: Emily Christman
Date: 11/22/2021

Significance

- Flowering plants became parasites -> cannot perform photosynthesis
    - Plastid genomes are smaller in size
- Balanophora- mushroom like parasitic plant
    - Most A+T rich protein coding genes known
    - Helps us understand the mutational and selective forces that drive radical genome evolution

Abstract

- Plastid genomes (plastomes) vary enormously in size and gene content among the many lineages of nonphotosynthetic plants, but key lineages remain unexplored. Balanophora plastomes are exceptionally compact, with numerous overlapping

genes, highly reduced spacers, loss of all cis-spliced introns, and shrunken protein genes. Most plastid protein genes in Balanophora consist of ≥90% AT, with several between 95% and 98% AT, resulting in the most biased codon usage in any genome described to date.

- Mycoheterotrophic, the term used for non photosynthetic plants that parasitize fungi, have arisen in 10 angiosperm families. The many independent lineages of holoparasitic and mycoheterotrophic angiosperms provide numerous test cases to explore the limits and potential outcomes of plastid genome (plastome) reduction in heterotrophic plants and the extent to which parallel evolution of gene content occurs.
- The >2,000 sequenced plastomes of photosynthetic angiosperms are highly conserved in size, gene content, and base composition; the great majority are 135–165 kb in size (11), contain 112 or 113 different genes, and are 35–40%. As opposed to the supporting process of gene expression, plastomes of photosynthetic and nonphotosynthetic angiosperms contain only one to four protein genes thought to be involved in core plastid processes other than photosynthesis and ATP synthesis.
- Extreme plastome reduction has also occurred in several lineages of mycoheterotrophic plants (13), most notably in Sciaphila thaidanica, whose 13-kb genome is stuffed with 20 genes as a result of severe compaction pressure (21), and Thismia tentaculata, whose 16-kb genome contains only 12 genes (22). Of the various parasitic and mycoheterotrophic lineages of angiosperms for which no plastome sequences are available, the holoparasitic Balanophoraceae are one of the most biologically interesting. To determine whether the plastome in Balanophoraceae is also bizarre, we sequenced and examined the expression of the plastid sequences of two Balanophora species

Results

- The near identity in size of the two Balanophora plastomes is the fortuitous result of the accumulation of many indels in each genome that happen to virtually balance out in aggregate length; indeed, all but 2 of their 19 genes differ in length . The Balanophora genomes are the smallest known plastomes except for the 11.3- and 15.2-kb genomes of two holoparasitic species of Pilostyles and the 12.8-kb genome of the mycoheterotroph.
- The two Balanophora plastomes have an identical and highly reduced set of 19 putatively functional genes, consisting of three rRNA genes (rrn16, rrn23, rrn4.
- 5), one tRNA gene (trnE), 11 ribosomal protein genes (rps2, rps3, rps4, rps7, rps11, rps12, rps14, rps18, rps19, rpl2, rpl14), and four protein genes of varying or unknown function
- Intergenic spacer regions in the Balanophora plastomes are exceptionally short (as detailed in the next section) and lack any ORFs of appreciable length, making it unlikely that any protein genes remain unannotated.
- However, given the small size of these RNA genes and the extreme divergence and AT-richness of the Balanophora plastomes (as detailed later), we can not rule out the

presence of one or a few so-far unrecognizable small RNA genes (only three spacers in both plastomes are large enough to contain a tRNA gene; Fig.

- reflexa trnE sequence must be nonfunctional with respect to protein synthesis, as the UUC anticodon is absent
- To the best of our knowledge, there is no precedent for an asymmetric location of a functional anticodon within an anticodon loop of a canonical length of 7 nt, even among the incredibly divergent and bizarre tRNA genes found in certain mitochondrial genomes (28⇓–30).
- In contrast, it probably still functions in heme biosynthesis, as it is highly conserved (as detailed later) and contains seven of the eight sequence determinants for correct charging of tRNAGlu(UUC) in Escherichia coli that are present in Nicotiana and Schoepfia, a hemiparasitic member of the Santalales
- reflexa trnE sequence lacks the UUnA motif at its canonical position, the motif is present only a few nucleotides away (SI Appendix, Fig.
- laxiflora) is strongly supported by the fact that trnE is more highly conserved (96% identity, gaps excluded), and also more GC-rich (29%), between the two Balanophora plastomes than any of their 18 other genes (SI Appendix, Tables S1 and S2).
- Taking both Balanophora plastomes together, 27 of the 30 protein genes were annotated as starting with ATG, with ATA or ATT annotated as start codons in the other three cases.
- All protein genes were annotated as terminating with TAA except for rpl2, which appears to terminate with TGA in both species (SI Appendix, Table S3).
- Moreover, despite being an order of magnitude smaller in size and gene content than virtually all plastomes of photosynthetic land plants
- S1A), the two Balanophora plastomes are colinear with those of Schoepfia and Nicotiana (and most other angiosperms) except for the location of a single gene, rpl14
- Plastid protein genes in Balanophora are compact too, having lost introns and sustained a net reduction in coding-sequence length.
- Balanophora has lost both clpP introns, the single rpl2 intron, and intron 2 of rps12, all of which are present and cis-spliced in Schoepfia and most other land-plant plastomes.
- The frequency of indels, especially deletions, is elevated in Balanophora protein genes and has led to a notable shortening of many genes
- Only 2 of the 15 protein genes are essentially the same size as homologs in the hemiparasitic relative Schoepfia, whereas all others are ≥5% shorter, 9 are ≥10% shorter, and 5 are extremely reduced in length (32–88%; SI Appendix, Table S1).
- Although their small-subunit and large-subunit rRNA genes are relatively GC-rich (19–24%), most protein genes are even more AT-rich than the genome as a whole, with five or six of them ≤5% GC and ycf2 an astounding 2% GC (Fig.
- 2 and SI Appendix, Table S1).
- To put these base compositional biases in perspective, we compared, in four ways, the Balanophora plastomes with 28 of the most AT-rich genomes of plastids, mitochondria, and bacteria (SI Appendix, SI Materials and Methods includes genome-selection criteria).

- First, the Balanophora plastomes are the most compositionally biased (toward AT or GC) plastid genomes sequenced to date and are surpassed only by the mitochondrial genome of the yeast Nakaseomyces bacillisporus (33), whose 10.
- 9% GC content is largely the result of its high content of extremely AT-rich noncoding spacer DNA (SI Appendix, Figs.
- Second, at 8.
- 7% and 8.
- 9% GC, the Balanophora plastid protein genes are the most AT-rich of any gene set analyzed, with the plastome of the malarial parasite Plasmodium falciparum next at 11.
- 0% GC, followed by another apicomplexan (Babesia) at 12.
- 1% (SI Appendix, Table S5).
- Third, at 1.
- 1% and 1.
- 2% GC, Balanophora also has the most extreme compositional bias at third-position synonymous sites (GC3), with Plasmodium next at 2.
- 1% (SI Appendix, Table S5).
- However, the comparable usage of A and T at third-position synonymous sites for eight of the nine codon families for which such a choice is possible (SI Appendix, Table S6) is inconsistent with translational efficiency.
- Three lines of evidence argue against the hypothesis that the extreme codon-usage bias results from selection driven by nitrogen availability and/or energetic costs (36, 37).
- Although this selection hypothesis does predict the observed bias in AT over GC, it also predicts a predominance of T over A at third-position synonymous sites, which is clearly not the case in Balanophora (SI Appendix, Table S6).
- Finally, one might expect selective pressure on nitrogen availability and/or energetic costs to operate on all Balanophora genomes, perhaps foremost on its vastly larger nuclear genome, yet the Balanophora nuclear genome possesses only a trivial codon-usage bias compared with the plastome (SI Appendix, Fig.
- S8).
- Examination of the 15 Balanophora protein genes revealed 18 internal, in-frame TAG codons in B.
- TAG is, of course, a stop codon in the canonical genetic code, the code that is employed by all land-plant plastomes examined heretofore.
- reflexa, these TAG codons occur at positions at which TGG (tryptophan in the canonical code) is present in most or all of the diverse photosynthetic land plants in the sequence alignments shown in SI Appendix, Fig.
- Conversely, there is not a single TGG codon in any Balanophora plastid protein gene.
- The four partial protein-gene sequences from Balanophora fungosa (as detailed later) contain four in-frame TAG codons, two each in clpP and rpl2.
- In both sequenced Balanophora plastomes, all occurrences of TAG are internal, and, conversely, no annotated stop codons are TAG; all are TAA or, for one gene, TGA (SI Appendix, Table S3).
- Arrows mark internal TAG codons present in one or both Balanophora plastomes and inferred to encode W; note that, at five of these six positions, most or all non-Balanophora land plants contain TGG (W in the standard genetic code).

- The positions of these TAG codons in the complete alignments are shown in SI Appendix, Fig.
- In-frame TAG codons in Balanophora plastid genes One hypothesis is that these codons do in fact serve as premature stop codons, in which case 9 of the 15 protein genes (Table 1) are pseudogenes in one or both Balanophora plastomes.
- (i) It is entirely inconsistent with sliding-window analysis of the ratios of nonsynonymous substitution rate (dN) to synonymous substitution rate (dS) for the five longest and best-conserved genes for which the two Balanophora species share internal TAG codons (Fig.
- 3C and SI Appendix, Fig.
- S9): for all five genes, the signature of strong purifying selection is evident downstream of their internal TAG codons.
- (ii) As a group, these five genes (accD, clpP, rpl2, rps2, ycf1) show higher levels of purifying selection within Balanophora (mean dN/dS = 0.
- 26, median = 0.
- 14) than the six protein genes that lack TAGs in both plastomes (mean dN/dS = 0.
- 33, median = 0.
- 36; SI Appendix, Table S2).
- reflexa divergence (SI Appendix, Table S2), one would expect these five internal-TAG presumptive pseudogenes to contain frame shifts, yet, as annotated, they have none.
- laxiflora cDNA sequences (as detailed later), which cover 10 of its 16 internal TAG codons, definitively rejects the editing hypothesis.
- As seen for certain other holoparasitic angiosperms, especially Hydnora and Pilostyles (9, 19), Balanophora plastid genes are, in aggregate, extremely divergent in sequence relative to a diverse range of photosynthetic land plants (SI Appendix, Fig.
- S10).
- Rapid sequence evolution in a 16-gene concatenate (3 genes present in Balanophora were excluded; Materials and Methods) is evident from both the extremely long branch leading to Balanophora and the relatively high divergence between B.
- Despite the extreme divergence, phylogenetic analysis placed the Balanophora sequences within the Santalales (SI Appendix, Fig.
- S10), consistent with an analysis of three nuclear genes, one mitochondrial gene, and three plastid genes from a large number of relevant taxa (24).
- When analyzed individually, the 14 protein genes in Fig.
- 4 show considerable variation within Balanophora in levels of amino acid divergence (56–90% identity, gaps excluded; SI Appendix, Table S2).
- Because of the highly biased nucleotide composition of these genomes, nucleotide identity is actually higher than amino acid identity for all protein genes, a situation that is rarely observed (SI Appendix, Table S2).
- The long dS branches for these genes, together with the high dS values shown in Fig.
- 4B for all 14 protein genes, indicate that a high mutation pressure is operating throughout both Balanophora plastomes.
- Moreover, there is clear evidence of saturation at synonymous sites on the branch leading to the Balanophora common ancestor, with dS > 1.
- 5 for 13 of the 14 protein genes examined and >3.
- 0 for 3 genes (Fig.

- 4B).
- There is, however, clear evidence for purifying selection on the Balanophora stem-branch for 13 of the 14 protein genes, for which dN/dS is at most 0.
- 40 (Fig.
- Pairwise comparison between the two Balanophora plastomes reveals that their genes have continued to evolve under purifying selection, albeit with what appears to be a modest overall relaxation of selective constraints (Fig.
- 4B and SI Appendix, Table S2).
- laxiflora plastid genes selected for RT-PCR amplification showed evidence of transcription, with a very strong correlation between predicted and experimentally determined cDNA product sizes (three examples are provided in SI Appendix, Fig.
- S11).
- laxiflora because comparison of rps12 gene and cDNA sequences revealed an expected size difference as a result of splicing across sites predicted to generate a contiguous ORF.
- A lack of editing is not surprising because, in those plants for which plastid editing has been comprehensively determined (41), there are only very few edit sites in the protein genes present in Balanophora plastomes, but it does provide important evidence that the observed internal TAG codons are not altered by editing to a sense codon in the extremely divergent Balanophora plastome.
-

Discussions

- Although many lineages of holoparasitic and fully mycoheterotrophic plants share with Balanophora a highly reduced plastid genome and gene set, as well as highly elevated substitution rates (9, 13, 19, 22), only Pilostyles (19) must also import all tRNAs for plastid protein synthesis and only S.
- However, none approach Balanophora in AT-richness and codon-usage bias or have evolved a noncanonical genetic code, much less a novel one
- The radicalism of Balanophora plastomes is, in several respects, reminiscent of that found in “apicoplasts” of Apicomplexans, which are also extremely AT-rich and codon-biased, highly compact (including tiny spacers and many overlapping genes), and highly divergent in sequence, with a genetic-code change in a large subgroup of these pernicious parasites (45, 46).
- However, apicoplast genomes differ in two important ways from the highly reduced plastomes of Balanophora and many other lineages of nonphotosynthetic plants.
- The GC content of Balanophora plastomes is extremely reduced, to just 11.
- 6% and 12.
- 2%, making them the most AT-rich plastid genomes found to date.
- TAG occasionally encodes leucine, tyrosine, or glutamic acid (58, 59), but its use for tryptophan is a novel code variant that makes the Balanophora plastome unique.
- Indeed, UGG is probably no longer deciphered as tryptophan in Balanophora, as no TGG codons were found in the 15 protein genes of either plastome.

- The mitochondrial and nuclear genomes of Balanophora still use UGG for Trp and UAG as a stop codon (SI Appendix, Fig.
- S8).
- We presume that the plastid tRNATrp and accompanying tryptophanyl-tRNA synthetase that charge it are nuclear encoded, in which case a complicated translational scenario must have arisen during the evolution of the novel use of UAG for Trp in Balanophora.
- Therefore, the above scenario predicts that Balanophora has a unique plastid-targeted tryptophanyl-tRNA synthetase that charges the putative plastid-specific tRNATrp(CUA).
- One partial Balanophora tryptophanyl-tRNA synthetase assembly was found by BLAST analyses, but it appears to be orthologous to the cytosolic protein of other plants and, as such, would not be predicted to participate in plastid tRNA charging.
- Other mechanisms could also allow for decoding of UAG as Trp in Balanophora, such as read-through by plastid ribosomes using a “near-cognate” tRNA (a tRNA that can pair with stop codons at two of the three positions of a codon–anticodon sequence); such read-through has been shown to occur in eukaryotes (66).
- In addition, if other Balanophoraceae plastomes are also extremely AT-rich, they may have traveled a different pathway of genetic-code evolution, e.
- , instead of the novel TAG-for-Trp change, they may have sustained the relatively common TGA-for-Trp change.
- At 91% AT in coding regions, and 99% AT at third-position synonymous sites, the Balanophora plastome is, in these respects, the most extreme genome known in any organism or genetic compartment.
- Because nuclear and mitochondrial genes in Balanophora possess relatively modest codon-usage bias and AT richness (SI Appendix, Fig.
- S8), the forces responsible for the extreme plastid biases are presumably compartment-specific.
- However, it may nonetheless be revealing to fully examine the Balanophora mitochondrial genome, as angiosperm plastids and mitochondria share a number of dual-targeted nuclear genes for nucleotide metabolism and DNA replication, recombination, and repair (67).
- These patterns are clearly evident in Balanophora plastomes, and thus their unprecedented codon-usage bias is probably driven by a genome-wide AT mutation pressure, and possibly also AT-biased gene conversion (69), operating largely free of natural selection (70, 71).
- On top of these forces, there is the evident pressure toward genome streamlining in Balanophora plastid DNA (as detailed in the next section), as well as a highly elevated rate of synonymous substitutions.
- The extraordinary divergence—in length, in primary sequence, and in base composition and codon usage—of ycf2 and, to a somewhat lesser extent, several other plastid protein genes in Balanophora, may make them useful models to elucidate the limits of overall change and the nature of specific alterations that these proteins can sustain and still be functional.

- We hypothesize that the intense streamlining in Balanophora is, as with its AT content and codon-usage bias, driven predominantly by neutral, runaway mutational forces, in this case by exceptionally high deletion rates.
- Moreover, looking across all plastid genomes, the Balanophora situation may be the most extreme case known in terms of percentage of overlapping genes and spacer size, rivaled only by the plastomes of apicomplexan parasites, the green algal parasite Helicosporidium spp.
- , and the photosynthetic red alga Cyanidioschyzon merolae (SI Appendix, Table S9).
- Data on protein size are unavailable for most of the highly compact plastomes analyzed in SI Appendix, Table S9, and therefore an outstanding question is whether, as in Balanophora and Sciaphila, genome compaction pressures generally lead to shrinkage of plastid-encoded proteins.
- Most highly compact lineages of plastomes are, unlike Balanophora and apicomplexans, not particularly AT-rich in base composition (SI Appendix, Table S9).
- The Balanophora plastome is clearly still functional, as its 11 examined genes are all transcribed, its only intron is correctly spliced, most if not all of its 15 protein genes are evolving under selective constraint, and none of these highly divergent genes are riddled with TAA stop codons as would be the case if they were nonfunctional.
- Furthermore, if we are correct that Balanophora trnE still functions in heme synthesis, it nicely illustrates some of the divergence that this gene tolerates when selection on protein synthesis has been lost
- Functions for three of the five raison d'être (i.
- , nontranslational) genes in Balanophora plastomes are known: clpP encodes part of the Clp protease complex involved in protein degradation and import (77); accD encodes a subunit of acetyl-CoA carboxylase (ACCase), which is necessary for fatty-acid biosynthesis in most plants (78); and, as already emphasized, trnE plays an essential role in heme biosynthesis (15, 16).
- The presence in Balanophora of an accD gene evolving under strong purifying selection and the potential role of ycf1 in ACCase assembly suggest that one of the primary functions of its plastome is biosynthesis of lipids.
- Most of the four raison d'être protein genes in Balanophora plastomes contain multiple TAG/Trp codons; for a transferred copy of such a gene to become functional, all of its TAG codons would have to revert to TGG before the copy sustains any disabling mutations.
- In combination, these two standout features of Balanophora plastid DNA may permanently lock in some of its remaining genes, and thus the genome itself.
- Slowly evolving lineages could reveal important antecedent conditions to the genetic-code switch, the increase in AT content, and other unusual features of the Balanophora plastome.
-

Material and Methods

Questions

Problem to Solve