The loss of photosynthesis pathway in a holoparasitic plant *Aeginetia indica* revealed by plastid genome and transcriptome sequencing

Jingfang Chen

School of Life Sciences, Sun Yat-sen University

Runxian Yu

School of Life Sciences, Sun Yat-sen University

Jinhong Dai

School of Life Sciences, Sun Yat-sen University

Ying Liu

School of Life Sciences, Sun Yat-sen University

Renchao Zhou (✉ zhrench@mail.sysu.edu.cn)

Research article

Keywords: *Aeginetia indica*, plastid genome, transcriptome

Posted Date: November 27th, 2019

DOI: <https://doi.org/10.21203/rs.2.17795/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Plant Biology on May 8th, 2020. See the published version at <https://doi.org/10.1186/s12870-020-02415-2>.

Abstract

Background With three origins of holoparasitism, Orobanchaceae provides an ideal system to study the evolution of holoparasitic lifestyle in plants. The evolution of holoparasitism can be revealed by plastid genome degradation and the coordinated changes in the nuclear genome, since holoparasitic plants lost the capability of photosynthesis. Among the three clades with holoparasitic plants in Orobanchaceae, only Clade VI has no available plastid genome sequences for holoparasitic plants.

Results In this study, we sequenced the plastome and transcriptome of *Aeginetia indica*, a holoparasitic plant in Clade VI of Orobanchaceae, to study its plastome evolution and the corresponding changes in the nuclear genome as a response of the loss of photosynthetic function. Its plastome is reduced to 86,212 bp in size, and almost all photosynthesis-related genes were lost. Most protein coding genes in the plastome showed the signal of relaxation of purifying selection. Plastome and transcriptome analyses indicated that the photosynthesis pathway is completely lost, and that the porphyrin and chlorophyll metabolism pathways are partially retained, although chlorophyll synthesis is not possible.

Conclusions Our study suggests the loss of photosynthesis-related functions in *A. indica* in both the nuclear and plastid genomes. The *Aeginetia indica* plastome also provides a resource for comparative studies on the repeated evolution of holoparasitism in Orobanchaceae.

Background

Chloroplast (plastid) is an organelle of plants that conducts photosynthesis, and the structure and gene content of chloroplast genomes are highly conserved in most flowering plants [1]. Typical chloroplast DNA (cpDNA) is circular, ranging mainly from 110 to 160 kb in length [2], and it contains two inverted repeat (IR) sequences separated by a large single-copy region (LSC) and a small single-copy region (SSC) [3]. Holoparasitic plants offer a good system to study plastid genome evolution due to their loss of photosynthetic capacity. They usually display a reduction of plastid genome including genome size and gene content. The family Orobanchaceae is especially suitable for studying chloroplast (plastid) genome evolution because it contains a full trophic spectrum from autotrophic plants, to hemiparasites and holoparasites. In Orobanchaceae, holoparasites occur in three of six well supported clades, namely, Clade III (Orobanchaeae, ~ 180 holoparasitic species), Clade V (Rhinanthaeae, 5–7 holoparasitic species) and Clade VI (Buchnereae, ~ 70 species) [4–6]. So far, plastid genome sequences of holoparasites in Orobanchaceae were mainly from Clade III [7–12]. In Clade V, the plastid genome of only one holoparasite, *Lathraea squamaria*, has been sequenced recently [12]. Although Clade VI includes ~ 70 species from four holoparasitic genera (*Hyobanche*, *Harveya*, *Aeginetia* and *Christisonia*), no plastid genome sequences of holoparasitic plants from this clade have been characterized.

Plastid genomes of holoparasites in Clade III and Clade V of Orobanchaceae differ markedly in genome size and gene content. Plastid genome sizes of holoparasites in Clade III range from 45,673 (*Conopholis americana*) to 120,840 bp (*Orobanche californica*) [9]. However, the plastid genome size of *Lathraea*

squamaria from Clade V is 150,504 bp [12], much larger than those in Clade III. The number of intact genes in the plastid genomes of *Conopholis americana* and *Orobanche* species ranges from 21 to 34 [9], and almost all genes related to photosynthesis (*pet*, *psa*, *psb*, and *rbcL*) were lost or became pseudogenes. Whereas in the plastid genome of *Lathraea squamaria*, there are 46 intact genes including many genes related to photosynthesis (such as *psa*, *pab* and *pet*). This might be due to holoparasitic lineages in Clade V is younger than those in Clade III [12].

In addition to plastome degradation, the nuclear genomes of holoparasitic plants are also expected to evolve as a response of the loss of photosynthesis capability, since the genes related to photosynthesis in the plastid genome interact with many genes in the nuclear genome [7, 13]. The expressional changes of nuclear genes could be revealed by transcriptome sequencing. For example, the expression of genes in the photosynthesis and chlorophyll synthesis pathways has been examined in some parasitic plants [7, 14, 15].

Aeginetia is a small holoparasitic genus of Orobanchaceae and it consists of about four species distributed in southern and southeastern Asia [16]. According to the phylogenetic analyses of Orobanchaceae, *Aeginetia*, along with *Hyobanche*, *Harveya* and *Christisonia*, forms a monophyletic holoparasitic lineage in Clade VI [4, 5]. *Aeginetia indica* is the most widespread species in this genus [17]. It usually parasitizes on the roots of Poaceae plants like *Miscanthus* and *Saccharum* [18]. In a recent study, transcriptome data of *A. indica* have been used to detect horizontally transferred genes from Fabaceae and Poaceae species [19]. So far, plastid genome sequence and the degradation of photosynthesis related pathways have not been studied in this holoparasitic plant.

In this study, we assembled the plastid genome of *A. indica* using Illumina short reads produced by genome skimming. We also sequenced the transcriptomes from multiple tissues to examine the expressional changes of genes involved in photosynthesis. The results of this study will contribute to our understanding of the coordinated evolution of plastid and nuclear genomes and also facilitate comparative analysis of convergent evolution of holoparasitism in Orobanchaceae.

Results

Complete plastid genome of *A. indica*

The complete plastid genome of *A. indica* has a typical quadripartite structure, and it is 86,212 bp in length, with 22,301 bp of the LSC region, 529 bp of the SSC region, and 31,691 bp each of the IR regions (Fig. 1). AT content of this plastid genome was 65.64%. Based on the DOGMA and GeSeq annotation, the plastid genome of *A. indica* contains 54 putative intact genes and three pseudogenes. These intact genes contain 24 tRNA genes, 4 rRNA genes, 8 rpl genes, 12 rps genes and 6 other genes, namely, *ycf1*, *ycf2*, *accD*, *matK*, *infA* and *clpP* (Table 1). The three pseudogenes are *ψatpA*, *ψatpI* and *ψndhB*. *ψatpA* and *ψatpI* genes in the LSC region of *A. indica* plastome became pseudogenes because of being truncated at the 88nd codon and a premature stop codon at the 32nd codon, respectively. *ψndhB* gene in the IR region became a pseudogene due to an internal stop codon at the 53rd codon.

Table 1
Summary of genes in the *Aeginetia indica* plastome

Function	Genes
Ribosomal proteins large subunit	rpl2, rpl14, rpl16, rpl20, rpl22, rpl23, rpl33, rpl36
Ribosomal proteins small subunit	rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, rps16, rps18, rps19
Transfer RNA genes	trnH-GUG, trnQ-UUG, trnS-GCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnS-UGA, trnG-UCC, trnM-CAU, trnS-GGA, trnL-UAA, trnA-UGC, trnF-GAA, trnW-CCA, trnL-UAG, trnN-GUU, trnL-CAA, trnM-CAU, trnI-CAU, trnV-GAC, trnI-GAU, trnT-GGU, trnP-UGG
Ribosomal RNA genes	rrn4.5, rrn5, rrn16, rrn23
Other protein-coding genes	ycf1, ycf2, accD, clpP, matK, infA
Pseudogenes	ψ ndhB, ψ atpA, ψ atpI
Supplementary information	
Additional file 1: Figure S1. Maximum likelihood tree of seven species in Orobanchaceae based on sequences of 20 plastid genes shared among them. Numbers in the nodes are bootstrap values. Scale in substitutions per site.	
Additional file 2: Figure S2. The expression of genes in the photosynthesis pathway observed in the <i>Aeginetia indica</i> transcriptome. Genes with detected expression were in the red boxes. With courtesy of © www.genome.jp/kegg/kegg1.html .	
Additional file 3: Figure S3. The expression of genes in the porphyrin and chlorophyll metabolism pathway detected in the <i>Aeginetia indica</i> transcriptome. Genes with detected expression were in the red boxes. With courtesy of © www.genome.jp/kegg/kegg1.html .	
Additional file 4: Table S1. Relaxation of purifying selection in parasitic plants of Orobanchaceae based on branch model analysis of 20 protein coding genes shared by seven species of Orobanchaceae. The likelihood ratio test was used to compare the three models (M0: one ratio model; M2: two ratio model; M3: three ratio model). P-values are in bold when they are less than 0.05.	
Additional file 5: Table S2. Expression level of unigenes of <i>Aeginetia indica</i> in the photosynthesis pathway based on transcriptome analysis.	

The SSC region in plastome of *A. indica* shows a severe reduction in size and only two genes, rpl15 and trnL-UAG, were found in this region (Fig. 1). The two IR regions undergone expansions which towards both the LSC and SSC regions. In *L. philippensis* and other autotrophic plants, an intact ycf1 gene usually spans the IR and SSC regions, and rps8, rpl14, rpl16, rps3, rpl22 and rps19 genes were in the LSC region. Whereas, in *A. indica*, there is an intact ycf1 gene in each of the IR regions, and rps8, rpl14, rpl16, rps3, rpl22 and rps19 genes were all shifted into the IR regions.

Gene loss in the *A. indica* plastid genome

Compared with *L. philippensis*, there is substantial loss of genes in the *A. indica* plastid genome. Ten *ndh* (*ndhA*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ* and *ndhK*) genes were lost, and *ndhB* gene became a pseudogene, they encode subunits of NADH-dehydrogenase complex. All five *psa* (*psaA*, *psaB*, *psaC*, *psaI* and *psaJ*) and 15 *psb* (*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT* and *psbZ*) genes involved in photosystem I and photosystem II, were lost. Also, all six *pet* (*petA*, *petB*, *petE*, *petF*, *petH* and *petI*) genes, which encode cytochrome *b6/f* complex subunits with function in photosynthetic electron transport, were missing. In addition, four *atp* (*atpB*, *atpE*, *atpF* and *atpH*) genes encoding F-type ATPase subunits, four genes encoding DNA dependent RNA polymerase (*rpoA*, *rpoB*, *rpoC1* and *rpoC2*), and genes encoding envelop membrane protein (*cemA*), large subunit of Rubisco (*rbcl*), haem attachment factor (*ccsA*), and photosystem assembly factors (*ycf3* and *ycf4*) were lost as well.

Plastid genome rearrangements in *A. indica*

With Mauve 2.4.0, sequence alignment for the plastomes of *A. indica* and *L. philippensis* was shown in Fig. 2. We identified four locally co-linear blocks (LCBs) for the two species, and *A. indica* plastid genome has undergone two major inversions relative to *L. philippensis*. One is a 1,452 bp inversion which contains an intact *accD* gene and occurred in the LSC region, the other is a large inversion of 60,255 bp in length and it contains an intact *infA* gene at the boundary of the LSC and *IR_B* regions, complete SSC and *IR_B* region, and most of the *IR_A* region.

Relaxed purifying selection of *A. indica* plastid genes

A total of 20 protein coding genes shared among the seven species in Orobanchaceae, including 10 *rps* genes, 7 *rpl* genes, and *accD*, *infA* and *matK* genes were used for phylogenetic analysis. The maximum likelihood tree was strongly supported, with bootstrap values of all branches being 100 (Figure S1). Three *Striga* species were clustered into one clade, and *Buchnera americana* was sister to them. *Aeginetia indica* was sister to the clade consisting of the former four species.

Non-synonymous (dN)/synonymous (dS) substitution rate ratio (ω) can be considered as an indicator for selection pressure. Two-ratio model (M2) was first compared with one-ratio model (M0). ω values of all genes but *rpl20* and *rps18* in the parasitic plant branch were larger than those of the nonparasitic plant branch (Table S1), and the likelihood ratio test showed that M2 is significantly better than M0 at nine genes, i.e. *accD*, *infA*, *rpl22*, *rps11*, *rps14*, *rps19*, *rps2*, *rps3* and *rps7*, suggesting that these genes were under relaxed purifying selection in parasitic plants. Using three-ratio branch model (M3), we found that hemiparasitic species had higher or much higher ω than holoparasitic species at 13 of 18 genes (ω values of the remaining two genes are not available), while holoparasitic species had slightly higher ω than hemiparasitic species at only five genes (Table S1). This suggests that protein-coding genes retained in the plastome of *A. indica* still play important functional roles rather than experiencing more relaxed selective pressure than hemiparasitic species.

Transcriptome analysis for *A. indica*

We obtained 21.05, 19.04, 18.34 and 18.02 Gb clean reads for four tissues, i.e. flower, sepal, fruit, and stem, respectively. By de novo assembly of read data from the four tissues, we obtained a total of 205,380 transcripts, among which 153,986 were extracted as unigenes. The average length and N50 of these unigenes were 623.18 and 880 bp, respectively. There were 47,480 ORFs (Open Reading Frames) predicted from all unigenes by TransDecoder, and 42,007 of them could be annotated in Swissprot database, among 42,007 Swissprot annotations, 8,466 could be assigned to 131 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

The photosynthesis pathway (ko00195) from the KEGG pathway database contains 63 genes (30 plastid genes and 33 nuclear genes). In the *A. indica* plastome, genes involved in photosystem I and II, cytochrome b6f complex, and photosynthetic electron transport are completely lost. The only two F-type ATPase related genes (*atpA* and *atpI*) in its plastome are pseudogenes. Based on the transcriptome analysis, only 14 unigenes in the photosynthesis pathway had expression (Table S2). The 14 genes included one gene encoding PSII 6.1 kDa protein, seven involving in photosynthetic electron transport and six being components of F-type ATPase (Figure S2). Expression of other genes in this pathway was not detected, indicating that these genes were either lost or non-expressional. The results from plastome and transcriptome analyses indicated that the photosynthesis pathway in *A. indica* was completely lost.

The porphyrin and chlorophyll metabolism pathway (ko00860) is complicated in plants. Porphyrins are intermediates of heme and chlorophyll, and heme is required for chlorophyll biosynthesis [20]. In the pathway from alanine to protoporphyrin IX, expression of genes encoding the intermediate products including HemA, HemB, HemC, HemD, HemE, HemF, HemL and HemY, were observed in the transcriptome of *A. indica* (Figure S3). However, because of the absence of expression of divinyl chlorophyllide a 8-vinyl-reductase [EC:1.3.1.75], the chlorophyll synthesis pathway appears to end at divinyl-proto-chlorophyllide production in *A. indica* (Figure S3). Obviously, the chlorophyll synthesis pathway is not complete at the later stage and chlorophyll can not be synthesized in *A. indica*.

Discussion

The plastome structure of *A. indica*

Among the available plastid genomes in Orobanchaceae, the plastid genome of *A. indica* is much smaller than those of *Striga* species (*S. aspera*, *S. forbesii* and *S. hermonthica*) from Clade VI of Orobanchaceae, but similar to those of *Orobanche* species, such as *O. pancicii* (88,525 bp) and *O. crenata* (87,529 bp) from Clade III [10]. However, their plastid genome structures are quite different: the sizes of the LSC (22,301 bp) and SSC (529 bp) regions in *A. indica* are much smaller than those of *O. pancicii* (42,763 and 7,608 bp) and *O. crenata* (43,054 and 7,925 bp), while the size of IR (31,691 bp) is much larger than those of *O. pancicii* (19,077 bp) and *O. crenata* (18,275 bp). The IR and SSC regions in plastome of *A. indica* undergone expansions and shrinkage, respectively. These phenomena were also observed in four hemiparasites (*Striga*) from the same clade VI in Orobanchaceae [1]. The IR region in chloroplast genome

is usually highly conserved in size and gene content, and it plays an important role in stabilizing chloroplast genome structure [21]. In the present study, IR expansion occurred in *A. indica* suggests that dramatic changes in the plastome accompanied by the loss of photosynthesis.

Gene content of the *A. indica* plastome

All 54 genes in plastome of *A. indica* had sequence length similar to the autotrophic *Lindenbergia philippensis*. The protein-coding genes in *A. indica* play fundamental roles in plastid function, including large and small ribosome protein subunits (rpl and rps genes), acetyl-CoA carboxylase beta subunit involved in lipid biosynthesis (accD), intron splicing (matK), translational initiation factor (infA), ATP-dependent protease subunit P (clpP), and protein import and turnover (ycf1) [10, 22]. The function of ycf2 gene, which has a conserved open reading frame, is still unknown. Four genes, atpA, clpP, rpl2 and rpl23, contain introns, which is consistent with the retention of matK's function as intron splicing.

The loss of photosynthesis related genes is a common phenomenon in holoparasitic plants, such as *Aphyllon* and *Orobanche* [10, 11]. Loss of housekeeping genes was also observed in other holoparasitic plants, for example, the plastid genome of *Balanophora laxiflora* is only 15,505 bp in size, with most genes being lost [23], and *Rafflesia lagascae* has even lost its whole plastid genome [24]. Some housekeeping genes have been transferred to the nuclear genome and their proteins can move back to the plastid to perform their functions [25]. Previous studies proposed models of plastome evolution in parasites and the order of gene losses [26–28]. The five stages in these models include "Photosynthetic", "Degradation I", "Stationary", "Degradation II" and "Absent" stages. The order of gene losses starts with *ndh* genes, followed by *psa/psb* genes and *rpo* genes, then *atp* genes, *rbcL* gene, nonessential housekeeping genes and other metabolic genes like *accD*, *clpP*, *ycf1* and *ycf2*, ends with the remaining housekeeping genes like *rpl* and *rps* genes. According to their models, plastome of *A. indica* is in the "Stationary" stage.

Rearrangement of the *A. indica* plastome

Rearrangement of *A. indica* plastome relative to *L. philippensis* chloroplast genome contained two inversions, one is a small fragment with an intact *accD* gene in the LSC region, while the other is a very large fragment across the most part of IR_A, intact SSR and IR_B regions. Large inversions around the IR regions were also observed in the plastomes of two other holoparasitic plants (*Phelipanche ramosa* and *P. purpurea*), and such large-scale rearrangements may be caused by relaxed selective pressure and progressive plastome nonfunctionalization [9].

The loss of photosynthesis pathway in *A. indica*

Aeginetia indica has no photosynthetic activity and obtains all carbon through connection with its host [19]. In the present study, the loss of photosynthesis pathway in *A. indica* was confirmed based on the loss of photosynthesis genes in its plastome and no detected expression of many genes in the photosynthesis pathway from its transcriptome. Chlorophyll is the reaction center of PSII and absorbs

light energy, playing an important role in photosynthesis [29]. Chlorophyll synthesis is impossible in *A. indica* because some key genes in the later stage of the porphyrin and chlorophyll metabolism pathway was not detected with expression. In contrast, an intact chlorophyll synthesis pathway was ever found in a holoparasitic plant *Phelipanche aegyptiaca*, suggesting that the expression of the chlorophyll synthesis pathway is for other functions other than photosynthesis [30].

Conclusions

The plastid genome of *Aeginetia indica*, a holoparasitic plant from Clade VI of Orobanchaceae, was assembled in the present study. Its plastid genome shows a reduction in size, accompanied with loss or pseudogenization of almost all photosynthesis related genes and some structural rearrangements. Transcriptome analysis from multiple tissues indicates that the photosynthesis pathway of *A. indica* was completely lost, while the porphyrin and chlorophyll metabolism pathway was partially retained, although chlorophyll synthesis is not possible. Our results suggest coordinated loss of photosynthesis related functions in the plastid and nuclear genomes of a holoparasitic plant. The plastome and transcriptome data of *A. indica* in the present study provides genetic resources for future studies and will facilitate comparative analysis of convergent evolution of holoparasitism in Orobanchaceae.

Methods

Plant materials, DNA isolation and Illumina sequencing

Aeginetia indica was collected from Shimentai Forest Park, Yingde, Guangdong, China. There were no specific permits required for collecting tissue samples of this species for research purpose. The plants were sampled and frozen immediately in liquid nitrogen, then kept at -80°C for further analysis. Total DNA was isolated from the flower stalk with a HiPure Plant DNA Mini Kit (Magen Company, Guangzhou, China) following the manufacturer's instructions. The quality and quantity of DNA were detected with 1% agarose gel electrophoresis and Qubit 3.0 Fluorometer (Invitrogen Corporation, USA), respectively. Shotgun genome sequencing with paired-end reads of 150 bp was performed on an Illumina HiSeq X Ten platform (IGE Biotechnology Ltd, Guangzhou, China). The sequencing data were deposited in NCBI Sequence Read Archive under accession number SRR9878563.

Assembly, annotation and alignment of plastid genome

The plastid genome of *A. indica* was assembled from Illumina sequencing data using NOVOPlasty [31], with parameters of insert size (300 bp), K-mer (37) and coverage cut off (1500). Annotation of plastid genome was performed by combining the DOGMA program [32] and GeSeq in OGDRAW [33]. Genes which contain one or more premature stop codons or frameshift mutations were considered as potential pseudogenes. The annotated plastid genome sequence of *A. indica* was deposited in GenBank under accession number MN529629. The circular map of plastid genome was drawn with OGDRAW. To seek any potential genomic rearrangement, the plastid genome sequences of *A. indica* and its autotrophic

relative *Lindenbergia philippensis* from the same family were aligned using Mauve 2.4.0[34] with default parameters except that Min LBC weight was set to 150.

Testing signatures of relaxed purifying selection for plastid genes in *A. indica*

We downloaded plastid genomes of six species of Orobanchaceae from GenBank, including four hemiparasitic plants *Striga aspera* (MF780872.1), *Striga forbesii* (MF780873.1), *Striga hermonthica* (MF780874.1), and *Buchnera americana* (MF780871.1) from Clade VI, and two autotrophic plants *Lindenbergia philippensis* (NC_022859.1) and *Rehmannia glutinosa* (NC_034308.1). Phylogeny of the six species and *A. indica* was performed using 20 protein-coding genes (7 *rpl* genes, 10 *rps* genes, *accD*, *infA* and *matK*) shared in their plastid genomes. After sequence alignment using MAFFT [35], phylogeny was reconstructed using the maximum likelihood algorithm in RaxML [36] with 1,000 bootstrap replicates, with *Rehmannia glutinosa* served as an outgroup. The maximum likelihood tree was shown with FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

We then used the phylogenetic tree as the input tree to test relaxed purifying selection for plastid genes in *A. indica*. The coding regions of each gene shared among the seven species were aligned by ClustalW (Codons) with default settings in MEGA-X version 10.0.5 [37]. The ratios (ω) of non-synonymous (dN) to synonymous (dS) substitution rate for 20 shared genes were estimated using codon-based analysis (codeml) in the PAML v.4.8a package [38]. Different branch models were used to analyze selective pressures among these species. The null one-ratio model (M0, it hypothesizes that all branches have one ω) was firstly performed, and then the likelihood of a two-ratio model (M2), with a foreground ω_1 for parasitic species and a background ω_2 for autotrophic species, was compared with that of M0. Moreover, a branch model with three ratios (M3) which assumes three different ω values for holoparasitic, hemiparasitic and autotrophic species, respectively, was compared with M2. The likelihood ratio test for M0 vs M2, and M2 vs M3, was conducted with the Chi-square distribution, with the degree of freedom equal to the difference in the number of parameters for the models, to evaluate the fit of the data to alternative branch models.

Transcriptome sequencing

Total RNA was isolated from flower, sepal, fruit, and stem tissues of *A. indica*, respectively. The quality and concentration of RNA were determined using 1% agarose gel electrophoresis and a Qubit spectrophotometer, respectively. mRNAs of these four tissues were purified with Oligo dT, and then used to construct cDNA libraries. Four cDNA libraries were sequenced on an Illumina HiSeq2000 platform (IGE Biotechnology Ltd, Guangzhou, China) with PE150 (paired-end 150 bp) strategy. The raw RNA sequencing data of flower, sepal, fruit, and stem tissues of *A. indica* were deposited in NCBI Sequence Read Archive under accession number SRR9959046, SRR9959049, SRR9959048 and SRR9959047, respectively.

Transcriptome analysis

Raw sequencing data were filtered by removing the adaptors and low quality reads. Transcripts were assembled using clean reads from four tissues through Trinity v2.8.4 [39], and then unigenes were extracted from these transcripts using a perl script `get_longest_isoform_seq_per_trinity_gene.pl` in Trinity. Expression level of all unigenes was determined using RSEM [40]. For annotation, all unigenes were aligned to NCBI non-redundant database (Nr, <http://www.ncbi.nlm.nih.gov/genbank/>) using DIAMOND (<http://ab.inf.uni-tuebingen.de/software/diamond/>). Open reading frames (ORFs) were predicted and putative protein sequences were obtained using TransDecoder tools (<http://transdecoder.sourceforge.net/>). Putative protein sequences were annotated in Swiss-Prot (<http://www.uniprot.org/>) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (<http://www.genome.jp/kegg/pathway.html>) databases with the criterion of E value < 1e-5 using Blastx and GhostKOALA (<https://www.kegg.jp/ghostkoala/>), respectively. Within the KEGG pathways, we focused only on the photosynthesis pathway (ko00195) and porphyrin and chlorophyll metabolism pathway (ko00860), and expression of genes in these two pathways were recorded.

Declaration

Ethics approval and consent to participate:

Not applicable

Consent for publication:

Not applicable

Availability of data and materials

The shotgun genome sequencing data of *Aeginetia indica* were deposited in NCBI Sequence Read Archive under accession number SRR9878563; the annotated plastid genome sequence of *A. indica* was deposited in GenBank under accession number MN529629; and the raw RNA sequencing data of flower, sepal, fruit, and stem tissues of *A. indica* were deposited in NCBI Sequence Read Archive under accession number SRR9959046, SRR9959049, SRR9959048 and SRR9959047, respectively. The data sets supporting the results of this study are included in this manuscript and its additional files.

Competing interests:

The authors declare that they have no competing interests.

Fundings:

This study was supported by the National Natural Science Foundation of China (31811530297), and Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences (AB2018014), and Chang Hungta Science Foundation of Sun Yat-sen University. The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' Contributions:

RZ conceived the study and designed the experiments; JC, RY and JD performed the experiments and data analysis; JC wrote the manuscript; RZ and YL revised the manuscript. All authors read and approved the final manuscript.

Acknowledgments:

Not applicable

Abbreviations

cpDNA
Chloroplast DNA
IR
Inverted Repeat
LSC
Large single-copy region
SSC
Small single-copy region
DOGMA
Dual Organellar GenoMe Annotator
LCB
Locally co-linear block
ORF
Open Reading Frame
KEGG
Kyoto Encyclopedia of Genes and Genomes
PSII
Photosystem II

References

1. Frailey DC, Chaluvadi SR, Vaughn JN, Coatney CG, Bennetzen JL. Gene loss and genome rearrangement in the plastids of five Hemiparasites in the family Orobanchaceae. BMC Plant Biol 2018, 18(1):30.
2. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. Plant Mol Biol 2011, 76(3-5):273-297.
3. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The Tortoise and the hare III. Am J Bot 2007, 94(3):275-288.
4. Bennett JR, Mathews S. Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. Am J Bot 2006, 93(7):1039-1051.
5. McNeal JR, Bennett JR, Wolfe AD, Mathews S. Phylogeny and origins of holoparasitism in Orobanchaceae. Am J Bot 2013, 100(5):971-983.
6. Fu W, Liu X, Zhang N, Song Z, Zhang W, Yang J, Wang Y. Testing the hypothesis of multiple origins of holoparasitism in orobanchaceae: Phylogenetic evidence from the last two unplaced holoparasitic genera, Gleadowia and phacellanthus. Front Plant Sci 2017, 8:1380.
7. Wolfe KH, Morden CW, Palmer JD. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc Natl Acad Sci USA 1992, 89(22):10648-10652.
8. Li X, Zhang TC, Qiao Q, Ren Z, Zhao J, Yonezawa T, Hasegawa M, Crabbe MJC, Li J, Zhong Y. Complete Chloroplast Genome Sequence of Holoparasite *Cistanche deserticola* (Orobanchaceae) Reveals Gene Loss and Horizontal Gene Transfer from Its Host Haloxylon ammodendron (Chenopodiaceae). PLoS ONE 2013, 8(3):3.
9. Wicke S, Müller KF, de Pamphilis CW, Quandt D, Wickett NJ, Zhang Y, Renner SS, Schneeweiss GM. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. Plant Cell 2013, 25(10):3711-3725.
10. Cusimano N, Wicke S. Massive intracellular gene transfer during plastid genome reduction in nongreen Orobanchaceae. New Phytol 2016, 210(2):680-693.
11. Schneider AC, Chun H, Tefanovic S, Baldwin BG. Punctuated plastome reduction and host-parasite horizontal gene transfer in the holoparasitic plant genus Aphyllon. P Roy Soc Lond B: Bio 2018, 285(1887):1887.
12. Samigullin TH, Logacheva MD, Penin AA, Vallejo-Roman CM. Complete plastid genome of the recent holoparasite lathraea squamaria reveals earliest stages of plastome reduction in orobanchaceae. PLoS ONE 2016, 11(3):3.
13. Barkan A, Goldschmidt-Clermont M. Participation of nuclear genes in chloroplast gene expression. Biochimie 2000, 82(6-7):559-572.
14. Ng SM, Lee XW, Mat-Isa MN, Aizat-Juhari MA, Adam JH, Mohamed R, Wan KL, Firdaus-Raih M. Comparative analysis of nucleus-encoded plastid-targeting proteins in *Rafflesia cantleyi* against

- photosynthetic and non-photosynthetic representatives reveals orthologous systems with potentially divergent functions. *Sci Rep-UK* 2018, 8(1):17258.
15. Ravin NV, Gruzdev EV, Beletsky AV, Mazur AM, Prokhortchouk EB, Filyushin MA, Kochieva EZ, Kadnikov VV, Mardanov AV, Skryabin KG. The loss of photosynthetic pathways in the plastid and nuclear genomes of the non-photosynthetic mycoheterotrophic eudicot *Monotropa hypopitys*. *BMC Plant Biol* 2016, 16:238.
 16. Zhang Z. Orobanchaceae. In: Wang W, editor. *Flora Republicae Popularis Sinicae*. Beijing, China: Science Press; 1990. p. 69-124.
 17. Morawetz JJ, Randle CP, Wolfe AD. Phylogenetic relationships within the tropical clade of Orobanchaceae. *Taxon* 2010, 59(2):416-426.
 18. Kato Y, Hisano K. In vitro culture of a root parasite, *Aeginetia indica* L. - I. General survey. *The Botanical Magazine Tokyo* 1983, 96(3):203-209.
 19. Kado T, Innan H. Horizontal Gene Transfer in Five Parasite Plant Species in Orobanchaceae. *Genome Biol Evol* 2018, 10(12):3196-3210.
 20. Oborník M, Green BR. Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Mol Biol Evol* 2005, 22(12):2343-2353.
 21. Perry AS, Wolfe KH. Nucleotide Substitution Rates in Legume Chloroplast DNA Depend on the Presence of the Inverted Repeat. *Journal of Molecular Evolution* 2002, 55(5):501-508.
 22. Cho WB, Choi BH, Kim JH, Lee DH, Lee JH. Complete Plastome Sequencing Reveals an Extremely Diminished SSC Region in Hemiparasitic *Pedicularis ishidoyana* (Orobanchaceae). *Ann Bot Fenn* 2018, 55(1-3):171-183.
 23. Su HJ, Barkman TJ, Hao W, Jones SS, Naumann J, Skippington E, Wafula EK, Hu JM, Palmer JD, DePamphilis CW. Novel genetic code and record-setting AT-richness in the highly reduced plastid genome of the holoparasitic plant *Balanophora*. *Proc Natl Acad Sci USA* 2019, 116(3):934-943.
 24. Molina J, Hazzouri KM, Nickrent D, Geisler M, Meyer RS, Pentony MM, Flowers JM, Pelser P, Barcelona J, Inovejas SA *et al*. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Mol Biol Evol* 2014, 31(4):793-803.
 25. Börner T, Zhelyazkova P, Legen J, Schmitz-Linneweber C. Chloroplast gene expression-RNA synthesis and processing. In: editor. *Plastid Biology*; 2014. p. 3-47.
 26. Wicke S, Müller KF, DePamphilis CW, Quandt D, Bellot S, Schneeweiss GM. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proc Natl Acad Sci USA* 2016, 113(32):9045-9050.
 27. Wicke S, Naumann J. Chapter Eleven - Molecular Evolution of Plastid Genomes in Parasitic Flowering Plants. In: Chaw S-M, Jansen RK, editor. *Advances in Botanical Research*: Academic Press; 2018. p. 315-347.
 28. Naumann J, Der JP, Wafula EK, Jones SS, Wagner ST, Honaas LA, Ralph PE, Bolin JF, Maass E, Neinhuis C *et al*. Detecting and characterizing the highly divergent plastid genome of the

- nonphotosynthetic parasitic plant *Hydnora visseri* (Hydnoraceae). *Genome Biol Evol* 2016, 8(2):345-363.
29. Baker NR. Chlorophyll Fluorescence: A Probe of Photosynthesis In Vivo. *Annual Review of Plant Biology* 2008, 59(1):89-113.
 30. Wickett NJ, Honaas LA, Wafula EK, Das M, Huang K, Wu B, Landherr L, Timko MP, Yoder J, Westwood JH *et al.* Transcriptomes of the parasitic plant family orobanchaceae reveal surprising conservation of chlorophyll synthesis. *Curr Biol* 2011, 21(24):2098-2104.
 31. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 2017, 45(4):4.
 32. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 2004, 20(17):3252-3255.
 33. Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* 2013, 41(Web Server issue):W575-581.
 34. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004, 14(7):1394-1403.
 35. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 2013, 30(4):772-780.
 36. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014, 30(9):1312-1313.
 37. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018, 35(6):1547-1549.
 38. Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007, 24(8):1586-1591.
 39. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011, 29(7):644-652.
 40. Li B, Dewey CN. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011, 12:323.

Figures

Figure 1

The plastid genome map of *Aeginetia indica*. Genes shown outside and inside the outer circle are transcribed counterclockwise and clockwise, respectively. GC and AT contents across the chloroplast genome are shown with the dark and light shading, respectively, inside the inner circle.