AN EVOLUTIONARY GENETIC APPROACH TO UNDERSTANDING PLASTID GENE FUNCTION: LESSONS FROM PHOTOSYNTHETIC AND NONPHOTOSYNTHETIC PLANTS

Jeffrey D. Palmer, Patrick J. Calie*, Claude W. dePamphilis, John M. Logsdon, Jr., Deborah S. Katz-Downie, and Stephen R. Downie

Department of Biology, Indiana University, Bloomington, IN 47405
*Department of Biology, University of Michigan, Ann Arbor, MI 48109

1. INTRODUCTION

The biogenesis of chloroplasts and other plastid types requires the coordinated expression of nearly a thousand different genes, most of which now reside in the nucleus. The 90 or so protein genes and 34 ribosomal and transfer RNA genes that remain in the chloroplast genome are compactly arranged in a small circular chromosome whose size is typically about 150 kb (1). The prokaryotic, specifically cyanobacterial, ancestry of chloroplast DNA (cpDNA) is reflected in many of the ways its genes are organized (often into operons), transcribed (using eubacterial-like promoters and RNA polymerase), and translated (on eubacterial-like 70S ribosomes) (1,2). Certain aspects of chloroplast gene expression, however, appear to reflect the long period of the organelle's existence within and adaptation to the environment of the eukaryotic cell. These include the common occurrence of introns, an extensive array of splicing and other kinds of RNA processing, the long half-lives of chloroplast mRNAs, and a great reliance on post-transcriptional processes, especially translational mechanisms, for regulating the abundance of plastid proteins (1-3).

The complete sequences of three chloroplast genomes have now been determined - one from a bryophyte (Marchantia polymorpha; 4) and two from flowering plants (tobacco and rice; 1). With the sequences of all chloroplast genes known, several major challenges remain. First, the identity of many chloroplast genes is only tentatively established based on sequence similiarity to identified bacterial genes and without any direct evidence from biochemistry or genetics. Second, many chloroplast genes were initially identified only as open reading frames (ORFs) of potentially significant length. The elucidation of the biochemical function of these ORFs is currently limited to a few components of the thylakoid membrane (1) and chloroplast RNA polymerase (5). Of particular relevance to this challenge is the fact that the metabolic diversity of plastids is belied by the apparent simplicity of their gene products - limited thus far to components of the photosynthetic and genetic apparatus (1). A further challenge is to identify which parts of a given chloroplast gene are essential for its product's function and which are dispensable.

M. Baltscheffsky (ed.), Current Research in Photosynthesis, Vol. III, 475–482.
© 1990 Kluwer Academic Publishers. Printed in the Netherlands.

We are taking an evolutionary genetic approach to address some of these issues. In one set of studies, we are exploiting the unusual instability of cpDNA in the photosynthetic plant geranium (Pelargonium hortorum) to provide a wealth of natural mutations affecting chloroplast gene structure and function. We describe several dramatically altered (fragmented and deleted) genes encoding components of the genetic apparatus. In a second set of studies, we are exploiting one of nature's grand experiments - the permanent loss of photosynthesis - as a means of unmasking the potential contribution of plastid genes to nonphotosynthetic processes within the plastid. Surprisingly, our results suggest that not only are most photosynthetic genes missing from the highly reduced plastid genome of the nonphotosynthetic parasite beechdrops (Epifagus virginiana), but that most of the ORFs and some of the genes encoding genetic functions are absent too.

2. CHLOROPLAST GENE MUTATIONS IN PELARGONIUM

Our initial study of the chloroplast genome of the garden geranium, Pelargonium hortorum, showed it to be highly rearranged in structure compared to most other land plant cpDNAs (6). Its structural alterations include a tripling in size (from 25 kb to 76 kb) of the large inverted repeat common to most cpDNAs, a series of large inversions, and the origin and dispersion of two families of novel repeated sequences. We now describe a fine-scale analysis of a region of the geranium genome that has undergone an unprecendented amount of gene-shattering deletion and fragmentation.

2.1. Massive fragmentation and deletion of the rpoA gene.

Figure 1 shows the gene arrangement, based on complete DNA sequencing, of a 21 kb region of the geranium chloroplast genome that is highly rearranged compared to tobacco (7) and most other land plants. The most drastic alterations are those affecting rpoA, the gene encoding the alpha subunit of chloroplast RNA polymerase (5). Only about 60% of the normally 1 kb long coding region of rpoA is present in the sequenced DNA of geranium and Southern hybridizations fail to detect any rpoA-like sequences elsewhere in the chloroplast genome. The retained 60% of rpoA is fragmented into three regions (labeled A-C in Figs. 1 and 2), whose locations relative to the intact rpoA gene of tobacco are shown in Fig. 2. Part or all of these three regions is present in multiple copies (Fig. 2) that are scattered over a 12 kb region of the chromosome (Fig. 1).

This massive fragmentation and deletion of a gene thought (5) to be a necessary component of the chloroplast RNA polymerase raises questions as to the location and nature of the gene for the alpha subunit polypeptide. Has the functional rpoA gene been transferred from the chloroplast to the nucleus? The answer appears to be no, because Southern hybridizations using an intact rpoA gene from a related genus (Geranium) to Pelargonium fail to detect homology to nuclear DNA from Pelargonium. Do the residual rpoA pieces in

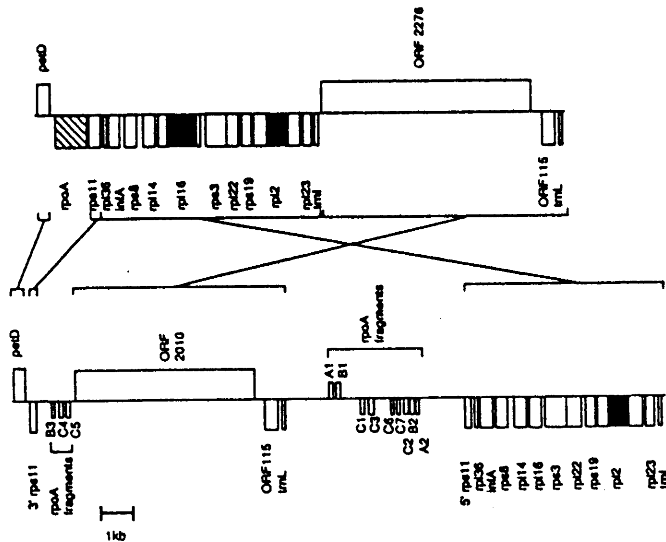

FIGURE 1. Gene rearrangement and fragmentation in a 21 kb segment of geranium cpDNA. The top map shows the standard arrangement (here represented by tobacco; 7) of a cluster of 17 chloroplast genes and the bottom map shows the highly altered arrangement of these genes in geranium cpDNA. "A1" through "C7" mark the 12 fragments of rpoA in geranium (cf. Fig. 2). Solid boxes within genes represent introns.

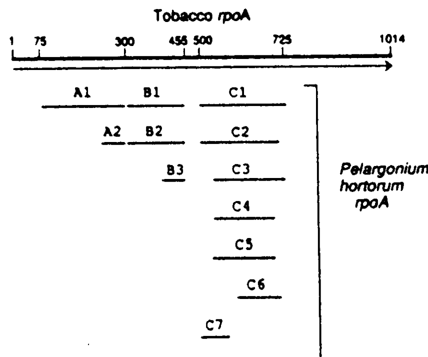

FIGURE 2. Location, relative to the intact rpoA gene of tobacco, of the fragments of rpoA that remain in the geranium chloroplast genome.

Pelargonium cpDNA produce a truncated but functional alpha polypeptide, perhaps through some newly created pattern of cis- and trans-splicing? We have little evidence to bear on this question; thus far we have failed to detect transcripts from the rpoA regions of geranium in Northern blot experiments. Another possibility is the substitution within the chloroplast RNA polymerase of an entirely different polypeptide for the alpha subunit.

The rearrangements of rpoA are not only extensive but also recent. As mentioned above, an intact rpoA gene is found in the chloroplast of another genus in the geranium family (Geraniaceae) and is flanked by intact copies of the genes - rps11 and petD - that normally flank it (Fig. 1). Furthermore, the rpoA alterations appear to be isolated events as far as the rest of the chloroplast-encoded subunits of RNA polymerase are concerned. Southern mapping experiments indicate that rpoB, rpoC1, and rpoC2 are most likely intact genes in Pelargonium.

2.2. Mutations of ribosomal protein genes.

In tobacco and most other plants, rpoA is the last gene of a 12-gene cluster thought to constitute an operon. Two of the 10 ribosomal protein genes that precede rpoA in this cluster show noteworthy changes in geranium. First, rps11 shows rpoA-like fragmentation and deletion, albeit on a smaller scale. Only 58% of the standard rps11 coding region is found in the 21 kb region of geranium cpDNA sequenced, and this is split into two segments (a 5' region of 44 codons and a 3' region of 37 codons) separated by 14 kb of other rearranged and transposed genes (Fig. 1). The function of the rps11 gene pieces is as unclear as for rpoA; we have yet to examine their possible expression in the chloroplast or to look for an rps11-homolog in the nucleus.

Second, rpl16 of geranium has sustained a clean deletion of the 1 kb intron that characterizes the gene in all other plants examined (1,4,7), including all other genera in the geranium family. Although intron losses are not uncommon in nuclear genes, this is the first reported case for chloroplasts and has important implications. The best theory to account for the precise loss of introns is via gene conversion of an intron-containing genomic gene with an intron-lacking form of the gene generated by reverse transcription of its mature, spliced mRNA (8). Thus, this intron loss provides indirect evidence for the presence of a reverse transcriptase activity in chloroplasts.

2.3. Extreme size polymorphism in the largest chloroplast gene.

The largest gene in the chloroplast genome is an enigma. Not only is the identity of its product unknown, but it is also variable in size (2131-2276 codons in spinach, tobacco, and Marchantia; 4,7,9), even among isolates of a single strain of Oenothera hookeri (10). This gene is unusual in geranium in that it has become transposed from its normal location (as in tobacco) to a novel one (Fig. 1). In addition, the geranium gene, at 2010 codons, is considerably smaller than in any other plant examined as the result of a series of small deletions throughout the gene. Northern blot studies reveal that ORF2010 is

transcribed in geranium, with some of the detected transcripts spanning the entire 6 kb gene. These observations dramatize the degree to which sequences in this gene are dispensable for the function of its (unknown) product and further delimit the potentially important sequences within the gene.

3. STRUCTURE AND FUNCTION OF PLASTID DNA IN NONPHOTOSYNTHETIC PLANTS

Nonphotosynthetic plants, most of which are parasitic on photosynthetic ones, have evolved repeatedly in flowering plants and algae (11). Although lacking chlorophyll, many or all of these nonphotosynthetic plants still contain plastids, at least some of which are known to store starch, pigments or lipids and contain plastid DNA and ribosomes (11). These plants provide important genetic systems in which to study the role of the plastid genome in processes other than photosynthesis. Below we describe our physiological and genetic characterization of the plastid of beechdrops, _Epifagus virginiana_, an obligate root parasite of its sole host, beech trees (_Fagus americana_), and then extend these findings to two other groups of parasitic plants.

3.1. Genetic and physiological loss of photosynthesis in the nonphotosynthetic parasite beechdrops.

Physiological and biochemical measurements reveal that beechdrops shoots are completely lacking in photosynthetic activity. Chlorophyll is undetectable and gas exchange measurements show an absence of any net CO_2 uptake. The plants lack any detectable Rubisco (ribulose-1,5-bisphosphate carboxylase/oxygenase) protein and activity, but do have measurable levels of PEPCase (phosphoenol pyruvate carboxylase) activity. However, carbon isotope composition studies show that beechdrops acquires its carbon via Rubisco rather than PEPCase. Moreover, the $\delta^{13}C$ value of beechdrops is essentially identical to that of its host beech tree. Taken together, these results indicate that beechdrops is incapable of net CO_2 fixation and that it acquires all of its carbon from its host.

Mapping studies of the beechdrops plastid genome show it to be a highly degenerate molecule that has lost most, if not all, of its photosynthetic genes (Table 1). The circular genome is only 71 kb in size, by far the smallest naturally occurring plastid genome known. Both single copy regions are highly reduced in size relative to related photosynthetic plants (the large and small single copy regions are 87 kb and 18.5 kb, respectively, in tobacco, but only 18 kb and 3.6 kb in beechdrops). Interestingly, beechdrops retains a full-sized inverted repeat of 25 kb. Overall, nearly two-thirds of the genetic complexity of tobacco (131 kb) is missing in beechdrops (46 kb). Figure 3 provides examples of hybridization experiments demonstrating that, under conditions in which total tobacco cpDNA hybridizes strongly to beechdrops DNA, tobacco probes specific to three photosynthetic genes (_psb_A, _atp_B, _pet_A) fail to hybridize. We emphasize that the

TABLE 1. Genes found in cpDNA of tobacco and ones (indicated with *) partly or entirely deleted from cpDNA of beechdrops.

Probable function	Gene name*	Number of Genes[a]
Photosynthesis		17/26
photosystem I	_psa_ A*,B*,C	2/3
photosystem II	_psb_ A*,B,C*,D*,E*,F*,G,H*,I,K,L*	7/11
cytochrome b/f	_pet_ A*,B*,C,E	2/4
ATP synthase	_atp_ A*,B*,E,F*,H*,I*	5/6
Calvin cycle	_rbc_ L*	1/1
chlorophyll synthesis	_trn_ G	0/1
Transcription and Translation		8/59
RNA polymerase	_rpo_ A,B*,C1*,C2*	3/4
ribosomal RNA	_rrn_ 4.5S,5S,16S,23S	0/4
transfer RNA	_trn_ 30 different (C*,K*,F*,L*)	4/30
ribosomal protein, 50S	_rpl_ 8 different	0/8
ribosomal protein, 30S	_rps_ 12 different (_rps_16*)	1/12
Miscellaneous		28/38
NADH dehydrogenase	_ndh_ A*,B,C*,D*,E*,F*,G*,H*	7/8
other ORFs	at least 30	21/30

[a]Number of genes missing from beechdrops cpDNA (left of slash); number of genes present in tobacco cpDNA (right of slash).

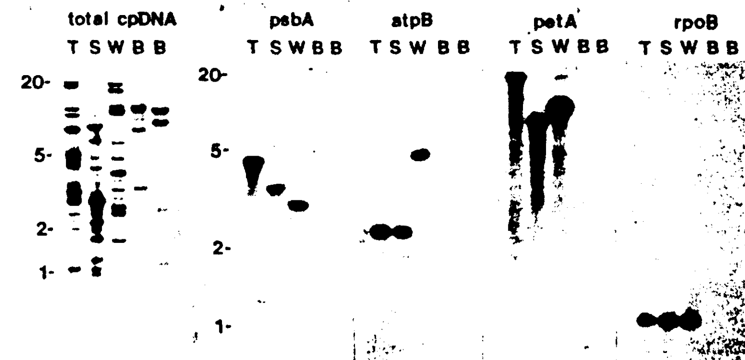

FIGURE 3. Absence of three photosynthetic genes and one RNA polymerase gene from beechdrops plastid DNA. Hybridization probes used are total cpDNA from tobacco and tobacco clones internal to the four indicated genes. Probes were hybridized to Southern blots containing BamHI digests of cpDNAs from tobacco (T), snapdragon (S), and witchweed (W), and ClaI and HindIII digests of beechdrops (B) total DNA.

negative hybridization results obtained in experiments like these have been confirmed by the mapping studies described above, which rule out the possibility of beechdrops having retained more than a small portion of these genes. The absence of most normally cpDNA-encoded genes from the plastid genome of beechdrops (Table 1) indicates that the physiological absence of photosynthetic activity discussed in the preceding paragraph reflects a permanent and irreversible loss of the genetic ability to produce a functional photosynthetic apparatus.

3.2. Deletions of nonphotosynthetic genes in beechdrops: A nonfunctional genome?

Plastids carry out a diversity of metabolism unrelated to photosynthesis, including certain steps of amino acid biosynthesis, lipid synthesis, nitrite and sulfate reduction, chlorophyll and terpenoid synthesis, carbon storage, and glycolysis, not to mention the maintenance of their own genome (DNA replication, repair, and recombination). Do any of these activities require plastid gene products? Our studies on beechdrops tentatively suggest that, at least in the physiological context of a parasitic plant that acquires carbon and other nutrients from its host, the answer may well be no. For not only are most of the photosynthetic genes missing from the beechdrop plastid genome, but so too are most of the NADH dehydrogenase genes and unidentified ORFs (Table 1). In fact, the possibility exists that the beechdrop plastid genome is completely nonexpressed and nonfunctional, in some sense a "pseudogenome". Evidence leading to this preliminary conclusion is of two types. First, beechdrops plastid DNA lacks genes for three of the four cpDNA-encoded subunits of RNA polymerase (e.g. Fig. 3) and probably, at a minimum, one ribosomal protein gene and four tRNA genes (Table 1). Second, we are thus far unable to detect any plastid transcripts, including ribosomal RNAs, in beechdrops.

3.3. Parallel trends in plastid genome reduction in different groups of nonphotosynthetic plants.

Are the highly reduced size and possibly nonfunctional nature of the beechdrops plastid genome unique to this particular plant, or have parallel reductions occurred in other, independently derived lineages of nonphotosynthetic plants? Hybridization studies similar to those shown in Fig. 3 suggest that the latter is the case. Most of the genes missing from beechdrops plastid DNA also fail to hybridize to plastid DNAs of two other nonphotosynthetic lineages (Cuscuta sp., Cuscutaceae, and Corallorhiza odontorhiza, Orchidaceae).

4. CONCLUSIONS

This paper is a first progress report of our efforts to apply an evolutionary genetic approach to understanding chloroplast gene function. The results reported here are promising for both the photosynthetic and nonphotosynthetic systems described. In the case of

the photosynthetic plant geranium, we have identified two genes, one encoding an RNA polymerase subunit and the other a ribosomal protein, that were recently shattered by means of deletion and fragmentation. Although the full implications of these findings for the function of these genes and their products remain to be elucidated, it is clear that major changes must have taken place in either the coding site of these proteins (gene transfer), the identity of the gene producing them (gene substitution), or their expression and ultimate sizes (if the gene pieces within the chloroplast retain function). In the case of the nonphotosynthetic plant beechdrops, our results reveal that the plastid is physiologically and genetically incompetent to perform photosynthesis; indeed, most of the plastid photosynthetic genes have been lost entirely. Surprisingly, most other plastid genes are also missing in beechdrops, including ones encoding important components of the gene expression machinery. If future studies confirm our working hypothesis that the plastid genome in beechdrops is nonfunctional, then we will have to conclude that, in a parasitic plant at least, the plastid genome does not contribute to any of the diverse metabolic and biochemical processes that take place in plastids.

ACKNOWLEDGEMENTS

This work was supported by NIH R01-GM35087 to JDP, NIH F32-GM1194[8] to RJC, NSF DCB-8710614 to CWD, and NSERC PDF to SRD. We are grateful to C. Cartwright for assistance with the geranium cpDNA sequencing studies, to M. Chase for helping to collect beechdrops, and to J. Teeri, J. Seeman, N. Bowlby, and C. Yocum for assistance with the physiological and biochemical studies of beechdrops.

REFERENCES

1 Sugiura, M. (1989) Ann. Rev. Cell. Biol. 5, 51-70
2 Gruissem, W. (1989) Cell 56, 161-170
3 Mullet, J.E. 1988. Ann. Rev. Plant Physiol. 39, 475-502
4 Ohyama, K., Kohchi, T., Fukuzawa, H., Sano, T., Umesono, K. and Ozeki, H. (1988). Photosyn. Res. 16, 7-22
5 Purton, S. and Gray, J.C. (1989) Mol. Gen. Genet. 217, 77-84
6 Palmer, J.D., Nugent, J.M. and Herbon, L.A. (1987) Proc. Natl. Acad. Sci. USA 84, 769-773
7 Shinozaki, K., Hayashida, N. and Suigura, M. (1988) Photosyn. Res. 18, 7-31
8 Fink, G.R. (1987) Cell 49, 5-6
9 Zhou, D.X., Massenet, O., Quigley, F., Marion, M.J., Moneger, F., Huber, P. and Mache, R. (1988) Curr. Genet. 13, 433-439
10 Blasko, K., Kaplan, S.A., Higgins, K.G., Wolfson, R. and Sears, B.B. (1988) Curr. Genet. 14, 287-292
11 dePamphilis, C.W. and Palmer, J.D. (1989) in Physiology, Biochemistry, and Genetics of Nongreen Plastids (Boyer, C.D., Shannon, J.C. and Hardison, R.C., eds.), in press, Amer. Soc. Plant Physiol., Rockville, Maryland