Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 10648–10652, November 1992
Evolution

Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant

(chloroplast DNA/*Epifagus virginiana*/translation/transcription/photosynthesis)

KENNETH H. WOLFE*, CLIFFORD W. MORDEN†, AND JEFFREY D. PALMER‡

Department of Biology, Indiana University, Bloomington, IN 47405

Communicated by Michael T. Clegg, August 3, 1992 (received for review April 15, 1992)

ABSTRACT **Complete nucleotide sequencing shows that the plastid genome of *Epifagus virginiana*, a nonphotosynthetic parasitic flowering plant, lacks all genes for photosynthesis and chlororespiration found in chloroplast genomes of green plants. The 70,028-base-pair genome contains only 42 genes, at least 38 of which specify components of the gene-expression apparatus of the plastid. Moreover, all chloroplast-encoded RNA polymerase genes and many tRNA and ribosomal protein genes have been lost. Since the genome is functional, nuclear gene products must compensate for some gene losses by means of previously unsuspected import mechanisms that may operate in all plastids. At least one of the four unassigned protein genes in *Epifagus* plastid DNA must have a nongenetic and nonbioenergetic function and, thereby, serve as the reason for the maintenance of an active genome. Many small insertions in the *Epifagus* plastid genome create tandem duplications and presumably arose by slippage mispairing during DNA replication. The extensive reduction in genome size in *Epifagus* reflects an intensification of the same processes of length mutation that govern the amount of noncoding DNA in chloroplast genomes. Remarkably, this massive pruning occurred with a virtual absence of gene order change.**

The known gene products of the ≈150-kilobase (kb) plastid genomes of photosynthetic flowering plants fall into two categories: those that function in gene expression and those with bioenergetic functions (1). The first category includes all RNAs thought to be necessary for gene expression (30 tRNAs and 4 rRNAs), 21 ribosomal proteins, and 4 subunits of RNA polymerase. The bioenergetic genes include 29 photosynthetic and 11 chlororespiratory genes. The functions of a further 12 conserved genes found through complete sequencing of chloroplast genomes (1) remain unknown.

Epifagus virginiana (beechdrops; Orobanchaceae) is a flowering plant that is parasitic on the roots of beech trees and is completely nonphotosynthetic. The plastid DNAs (ptDNAs) of plants such as *Epifagus* present a unique opportunity to investigate both the evolution of an organelle genome whose primary function (photosynthesis) has been rendered obsolete and the possible role of the plastid genome in metabolic processes other than photosynthesis. *Epifagus* ptDNA has been shown by Southern blot hybridization to have lost many genes, including all examined bioenergetic genes, but to have retained putatively intact genes encoding components of the genetic apparatus (2). However, sequencing of parts of the genome revealed that some RNA polymerase, ribosomal protein, and tRNA genes have also been deleted (3–5).

We report the complete nucleotide sequence of the 70,028-base-pair (bp) *Epifagus* plastid genome§ and show that the sets of tRNA, ribosomal protein, and RNA polymerase genes are all grossly incomplete compared to photosynthetic species such as tobacco (1). Nevertheless, the genome is functional: transcripts of plastid rRNA and protein genes have been identified (ref. 2; S. Ems and J.D.P., unpublished data), and the maintenance of many intact and conserved genes implies that mRNAs are translated. We hypothesize that the *Epifagus* plastid genome has remained active after the loss of photosynthesis because one or a few of its protein genes is involved in a nonbioenergetic process.

MATERIALS AND METHODS

Libraries covering the entire mapped (2) plastid genome of *E. virginiana* were made by cloning restriction fragments and PCR products in pBluescript and λ DASH II vectors. Complete sequence from both strands was obtained using exonuclease III deletion subclones, restriction subclones, or internal primers. Sequences of all PCR products were verified from at least two independent clones.

RESULTS AND DISCUSSION

DNA Loss from the *Epifagus* Plastid Genome. Massive gene loss has resulted in a plastid genome with a sequence complexity of 47.3 kb in *Epifagus* (Fig. 1), 36% of that (130.5 kb) in tobacco (1). The number of genes present is similarly reduced from 113 to 42 (Table 1). There are intact and likely functional genes for 21 proteins, 17 tRNAs, and 4 rRNAs, in addition to 14 pseudogenes. All bioenergetic genes have been lost; only 6 grossly truncated pseudogenes remain from the 29 photosynthetic and 11 chlororespiratory genes known in green plant ptDNAs (Fig. 1 and Table 1). Sequence identity between *Epifagus* and tobacco ptDNA-encoded proteins varies from 60% in the *matK* and *orf1738* products to 93% in ribosomal proteins L2 and S12.

Deletions have occurred throughout the genome. Large deletions (0.3–11.5 kb) map to at least 27 loci, resulting in truncated pseudogenes or in the loss of entire genes or groups of genes. However, the presence of multiple short deletions internal to some pseudogenes (3) and the predominance of small (<20 bp) deletions in the inverted repeat (IR) region (Fig. 2; see below) suggest that the large deleted regions are in fact the result of accumulated smaller deletions. We estimate the actual number of events responsible for the size reduction to have been in the hundreds, if not thousands. Remarkably, despite these many deletions, the genes and pseudogenes retained are found in the same relative order and orientation as in tobacco ptDNA, the only exception being the inversion of $trnL_{UAG}$ in *Epifagus*.

Abbreviations: ptDNA, plastid DNA; IR, inverted repeat; SC, single copy.

*Present address: Department of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland.
†Present address: Department of Botany/H.E.B.P., University of Hawaii, Honolulu, HI 96822.
‡To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M81884).

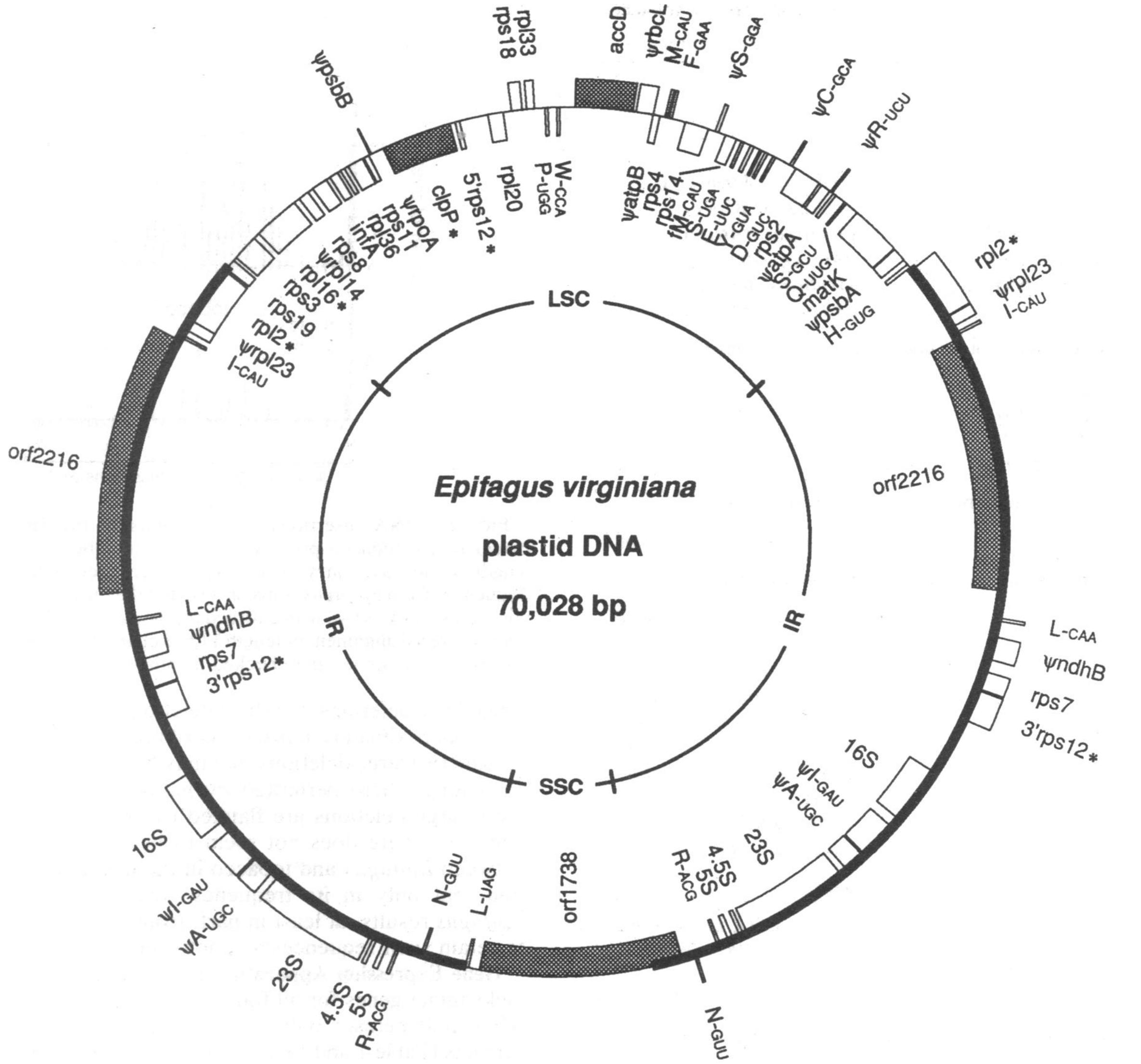

FIG. 1. Gene organization of the *E. virginiana* plastid chromosome. The IR of 22,735 bp divides the rest of the circular genome into large (LSC; 19,799 bp) and small (SSC; 4759 bp) SC regions. Genes drawn inside the circle are transcribed clockwise. tRNA genes are identified by the one-letter amino acid code and their anticodon sequence. Pseudogenes are marked "ψ." Asterisks indicate genes containing introns. The four genes that are not known to be involved in gene expression are shaded.

The greatest extent of deletion is in the single-copy (SC) regions: the large (19.8 kb) and small (4.8 kb) SC regions are 23% and 26%, respectively, of the size of their tobacco counterparts, whereas the IR (22.7 kb) is 90% of the tobacco size. The conservation of IR size, relative to the drastic changes in SC regions, could be due to two factors. (*i*) The fraction of deletions in the IR that are evolutionarily acceptable may be lower, because the IRs of photosynthetic plants contain a relatively high proportion of nonbioenergetic genes (1). (*ii*) The underlying rate at which deletion mutations are generated may be lower in the IR in all species. A precedent for a reduced mutation rate is that the rate of silent nucleotide substitutions in IR genes of photosynthetic plants is 3- to 4-fold lower than the rate in SC genes (6). This is also seen in comparisons between *Epifagus* and tobacco: the estimated numbers of synonymous substitutions per site are 0.50 ± 0.02 for SC genes (ranging from 0.24 in *rps18* to 0.67 in the SC portion of *orf1738*) and 0.12 ± 0.01 for IR genes (ranging from 0.08 in 3'-*rps12* to 0.16 in the IR part of *orf1738*). However, the proportion of DNA that is noncoding (including introns and pseudogenes) is fairly constant between IR and SC regions and between *Epifagus* and tobacco (between 36 and 44% in all cases). The *Epifagus* plastid genome is, therefore, no more streamlined than that of tobacco, and the IR of *Epifagus* is no less streamlined than its SC regions, which suggests that an equilibrium between levels of DNA deletion and insertion has been reached throughout the genome.

The slow rate of evolution of the IR permits a fine-scale analysis of the pattern of length mutation in this region. An outgroup sequence was used so that insertions could be distinguished from deletions in the *Epifagus* and tobacco lineages. Unfortunately, the only other angiosperm whose entire IR has been sequenced is rice (1); a dicot such as spinach would have been preferable. The use of a distant outgroup means that some information on length mutations is lost because of superimposed events. The analysis (Fig. 2) shows a greatly increased incidence of both insertions and deletions in the *Epifagus* lineage. The data in Fig. 2 are a composite of coding and noncoding regions and include genes such as *ndhB* for which the *Epifagus* locus is a pseudogene. Of the 84 deletions in the IR of *Epifagus* (Fig. 2), only 14 (17%) involve the loss of coding DNA (resulting in pseudogenes) and, therefore, would certainly not have been permissible in tobacco. Similarly, only 8 (15%) of the insertions in *Epifagus* disrupt genic sequences. However, it is difficult to assess the selective consequences of the many small-length

Table 1. Gene content of *Epifagus* ptDNA compared to tobacco

Present	Deleted or pseudogene
Ribosomal RNA genes	
16S, 23S, 4.5S, 5S	
Transfer RNA genes	
$trnD_{\mathrm{GUC}}$, $trnE_{\mathrm{UUC}}$, $trnF_{\mathrm{GAA}}$, $trnH_{\mathrm{GUG}}$, $trnI_{\mathrm{CAU}}$, $trnL_{\mathrm{CAA}}$, $trnL_{\mathrm{UAG}}$, $trnM_{\mathrm{CAU}}$, $trnN_{\mathrm{GUU}}$, $trnP_{\mathrm{UGG}}$, $trnQ_{\mathrm{UUG}}$, $trnR_{\mathrm{ACG}}$, $trnS_{\mathrm{GCU}}$, $trnS_{\mathrm{UGA}}$, $trnW_{\mathrm{CCA}}$, $trnY_{\mathrm{GUA}}$, $trnfM_{\mathrm{CAU}}$	$\psi trnA_{\mathrm{UGC}}$, $\psi trnC_{\mathrm{GCA}}$, $trnG_{\mathrm{GCC}}$, $trnG_{\mathrm{UCC}}$, $\psi trnI_{\mathrm{GAU}}$, $trnK_{\mathrm{UUU}}$, $trnL_{\mathrm{UAA}}$, $\psi trnR_{\mathrm{UCU}}$, $\psi trnS_{\mathrm{GGA}}$, $trnT_{\mathrm{GGU}}$, $trnT_{\mathrm{UGU}}$, $trnV_{\mathrm{GAC}}$, $trnV_{\mathrm{UAC}}$
Ribosomal protein and initiation factor genes	
rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps18, rps19, rpl2, rpl16, rpl20, rpl33, rpl36, infA	*rps15, rps16, ψrpl14, rpl22, ψrpl23, rpl32*
RNA polymerase and maturase genes	
matK	*ψrpoA, rpoB, rpoC1, rpoC2*
Photosynthetic and chlororespiratory genes	
	ψatpA, ψatpB, atpE, atpF, atpH, atpI, ndhA, ψndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK, psaA, psaB, psaC, psaI, psaJ, ψpsbA, ψpsbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, petA, petB, petD, petG, ψrbcL
Other protein genes	
clpP, accD, orf1738, orf2216	*orf29, orf31, orf34, orf62, orf168, orf184, orf229, orf313*

Pseudogenes are indicated by ψ. *Epifagus orf1738* and *orf2216* are homologs of tobacco *orf1901* (4) and *orf2280*, respectively.

mutations in RNA genes and introns, so we cannot say for certain whether the increased number of events in *Epifagus* is solely due to the loss of photosynthesis or to an increased rate of production (rather than acceptance) of length mutations as well. An increase in the rate of length mutation in *Epifagus* ptDNA would be in keeping with its increased rate of nucleotide substitution (refs. 3, 5, and 7; see below).

Approximately 30% (8 out of 27 events ≥4 bp) of the insertions in the *Epifagus* IR involve the exact tandem duplication of short sequences, typically 5 bp long. A further 9 insertions involve imperfect tandem repeats with a single mismatch. The 16-bp insert indicated in Fig. 2 involves a quadruplication of a 5- to 6-bp sequence in exon 1 of $\psi trnI_{\mathrm{GAU}}$ (5). None of the deletions examined involve the simple loss of such a repeated sequence, although two deletions in *Epifagus* (56 and 167 bp; ref. 5) are of sequences that are flanked by 6- and 12-bp direct repeats, respectively, in tobacco. An association between short tandem duplications and length mutations in ptDNA has been reported in studies on chloroplast DNAs (8, 9). However, the use of an outgroup sequence demonstrates that this association is due primarily to the formation of tandem duplications rather than the deletion of preexisting repeats. These duplications presumably arise by slippage mispairing of strands during DNA replication (8, 9). Insertions of tandem repeats are also evident in pseudogenes in SC regions of *Epifagus* ptDNA (3).

Fig. 2 shows that the pattern of DNA deletion from the IRs of *Epifagus* and tobacco is roughly trimodal, consisting of (*i*)

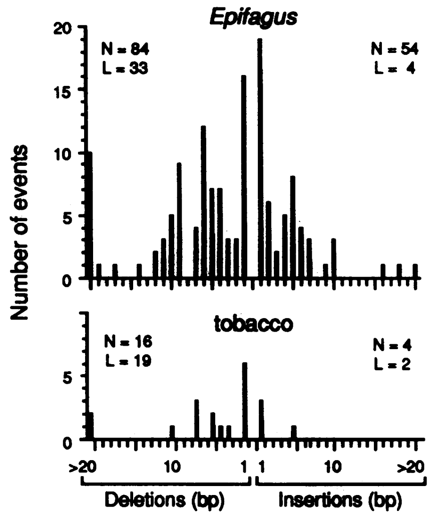

FIG. 2. DNA insertions and deletions in the IR regions of *Epifagus* and tobacco ptDNAs. The total numbers (*N*) and mean length (*L*, in base pairs) of events in each species are indicated. Sequences from *Epifagus*, tobacco (1), and the outgroup rice (1) were aligned using CLUSTAL with a forced topology and adjusted by eye to give an overall alignment of length 17,618 characters. Gaps were then counted by computer and checked manually.

single-base deletions, (*ii*) short deletions of <20 bp occurring in an approximately normal distribution centered around 6 bp, and (*iii*) large deletions that may be several hundred base pairs long where permitted by natural selection. Some of these large deletions are flanked by short direct-repeat sequences. There does not seem to be any major difference between *Epifagus* and tobacco in the mechanism(s) of deletion, but only in its frequency; the higher frequency in *Epifagus* results, at least in part, from the lifting of the need to retain such sequences as photosynthetic genes.

Gene Expression Apparatus. The *Epifagus* plastid genome lacks intact genes for all four plastid-encoded subunits of an RNA polymerase; only a truncated pseudogene for *rpoA* remains (Table 1 and Fig. 1). Since the genome is transcribed (ref. 2; S. Ems and J.D.P., unpublished data), an alternative RNA polymerase (presumably nuclear-encoded) must be active in *Epifagus* plastids, and it seems probable that two polymerases are normally responsible for transcription in chloroplasts (for discussion, see ref. 3). Loss of the ptDNA-encoded RNA polymerase might necessitate a change in plastid promoter organization, and examination of the sequences of probable promoter regions in *Epifagus* indicates that changes have occurred. There are only four intact genes (*clpP*, *trnE*, *rps2*, and *rps4*) in the large SC region of *Epifagus* ptDNA for which transcription start sites have been mapped in any photosynthetic dicot (usually spinach). The proposed eubacterial-type (10) −35 and −10 promoter regions of all four genes have either diverged or been deleted in *Epifagus*, suggesting that the parasite uses different promoters. This could be addressed experimentally by mapping transcription start sites *in vivo*. For *clpP* and *trnE*, similarity between *Epifagus* and tobacco begins a few nucleotides upstream of the inferred transcription start site, whereas for *rps2* and *rps4* the former start site has apparently been deleted. Only in the IR region, where most intergenic sequences have been retained, can homologs of chloroplast promoters be identified (and even there the rRNA promoter has been disrupted; ref. 5).

Of the 21 ribosomal protein genes in tobacco ptDNA, 15 remain intact in *Epifagus*, 4 have been completely deleted, and 2 are pseudogenes (Table 1 and Fig. 1). The sequences of these genes and the overall evolution of the plastid translational apparatus in *Epifagus* are discussed in detail in ref. 7.

Since the products of at least 2 of the missing ribosomal protein genes (*rps15* and *rpl22*) are likely to bind to rRNA (11, 12) and since S15 is required for early stages of ribosome assembly in *Escherichia coli* (11), it is unlikely that *Epifagus* plastid ribosomes can function without these proteins or some sort of replacements. These 6 genes may have migrated to the *Epifagus* nuclear genome, in the same way that *rps16, rpl21, rpl22,* and *rpl23* are thought to have been transferred in other plant lineages, but such widespread gene transfer in a single lineage would be without precedent (7).

Two other genes in *Epifagus* ptDNA are thought to encode genetic proteins: *infA* (translation initiation factor 1; see ref. 7) and *matK* (an intron maturase). In photosynthetic plants, *matK* is located within the group II intron of the $tRNA^{Lys}$ gene, but in *Epifagus, matK* is free-standing and the flanking intron and exon sequences have been lost (Fig. 1). The retention of *matK* suggests that its putative maturase product assists in splicing of other group II introns besides the one in which *matK* normally resides (C.W.M., K.H.W., and J.D.P., unpublished data). *Epifagus* ptDNA contains 6 group II introns that might be substrates for this activity, including the trans-spliced intron 1 of *rps12* (Fig. 1). The 15 other introns (1 group I and 14 group II) found in tobacco ptDNA are located in genes that have been lost from *Epifagus* ptDNA.

Each of the three sequenced chloroplast genomes encodes 30 or 31 tRNAs that are sufficient to translate all sense codons by means of fourfold wobble in some anticodons (1, 13). *Epifagus* ptDNA, however, contains only 17 intact tRNA genes, which appears too few to support translation using the 20 amino acids (for details, see ref. 7). There is no plastid-encoded tRNA species for 6 amino acids (alanine, cysteine, glycine, lysine, threonine, and valine), and 1 isoaccepting species has been lost for 4 other amino acids (Table 1). Nevertheless, all 61 sense codons are used in *Epifagus* protein genes and their frequency of use gives no indication of avoidance of the codons for which there are no tRNA species. Of the 7293 codons in the 21 *Epifagus* protein genes, 3117 (43%) have no plastid-encoded cognate tRNA (7). This is almost identical to the frequency (42%) at which these codons occur in the homologs of these genes in tobacco. Likewise, the proportions of alanine, cysteine, glycine, lysine, threonine, and valine in *Epifagus* plastid-encoded proteins are not significantly lower than those in their tobacco counterparts. The simplest explanation of these observations is that *Epifagus* plastid ribosomes have access to a full set of tRNAs, presumably as a result of tRNA import from the cytoplasm (3, 7). If this is true of *Epifagus*, it is likely also true of photosynthetic species.

Rates of Molecular Evolution. Overall, the rate of molecular evolution in the *Epifagus* plastid genome is accelerated. As anticipated, the loss of photosynthesis has led to the widespread deletion of genes in what is ordinarily a conservative molecule (14). More unexpectedly, phylogenetic trees drawn from sequences of three sets of translational components (ribosomal proteins, rRNAs, and pooled tRNAs) consistently point to faster evolution in *Epifagus* than in tobacco, by a factor of between 3 and 8. These results are presented in detail elsewhere (5, 7) and are interpreted to reflect an increase in the rate of both production and fixation of point mutations (7), the latter due to reduced selective constraints on a translational apparatus that now produces only a few nongenetic proteins (at most four).

Parsimony analysis of amino acid sequences shows accelerated evolution in *Epifagus* for three nontranslational genes: *clpP* (for which the *Epifagus* branch in a phylogenetic tree is 2.8-fold longer than the tobacco branch), *matK* (1.9-fold), and *orf2216* (4.1-fold). These accelerations are not so easily rationalized, in part because the functions of these proteins are poorly understood. For two other genes (*accD* and *orf1738*), the *Epifagus* and tobacco branch lengths are essentially equal.

Is the *Epifagus* Plastid Genome Functional? Our analysis of the structure and evolution of *Epifagus* ptDNA has been carried out under the hypothesis that the genome is active and that the intact genes are expressed, but the loss of all RNA polymerase and many translational genes from the genome must raise some doubts as to whether it is in fact functional. Transcription of all eight examined rRNA and protein genes in *Epifagus* has, however, been detected by Northern blot analysis and by sequencing PCR products derived from spliced plastid transcripts (ref. 2; S. Ems and J.D.P., unpublished data). In addition, the evident operation of selective constraints on the genome enables us to make the following three strong evolutionary arguments in favor of its being functional.

(i) Skewed deletions. If the *Epifagus* plastid genome were nonfunctional, we would expect deletions to be dispersed randomly around the genome, with the possible exception of a reduced incidence in the IR. This clearly has not been the case. For example, *Epifagus* ptDNA contains 9.1 kb of ribosomal protein gene sequences (including pseudogenes and introns), which is 80% of the amount in tobacco, whereas only 5% of photosynthetic sequences have been retained (22.2 kb in tobacco and 1.2 kb in *Epifagus*).

(ii) Maintenance of large open reading frames. All of the protein pseudogenes in *Epifagus* ptDNA (Table 1 and Fig. 1) contain numerous mutations (nonsense mutations, frameshifts, large truncations, and internal deletions), any of which alone would be sufficient to inactivate a gene. In contrast, the two largest genes in *Epifagus* ptDNA are 1738 and 2216 codons. It is impossible that open reading frames of these sizes could have been maintained intact in the face of extensive sequence divergence from tobacco (40% amino acid divergence for *orf1738* and 16% for *orf2216*) without selective filtering-out of deleterious mutations. For example, *Epifagus orf1738* differs from its tobacco homolog by 50 insertions/deletions (4) and 1253 nucleotide substitutions, none of which results in a frameshift or a stop codon.

(iii) Conservation of genes in parasites. E. virginiana and *Conopholis americana* are members of the entirely nonphotosynthetic family Orobanchaceae and share the loss of photosynthetic and chlororespiratory genes (ref. 15; S. R. Downie, C. W. dePamphilis, and J.D.P., unpublished data). All DNA sequence divergence between *Epifagus* and *Conopholis* ptDNAs represents events that have occurred in nonphotosynthetic lineages. Therefore, if these genomes are functional, gene sequences should be better conserved than those of intergenic spacers and pseudogenes, whereas if the ptDNAs of these plants are "pseudogenomes," all regions of the genomes should have diverged to a similar extent. Comparison of the rRNA operons of *Conopholis* (16) and *Epifagus* (5) shows that the rRNA genes are conserved to a much greater degree than the intergenic spacer regions (including two tRNA pseudogene loci). This is true for both point mutations (3.2% nucleotide divergence in 4.5 kb of rRNA genes and 13.4% in 0.6 kb of intergenic spacers) and length mutations (3.8 length mutations per kb in rRNA genes and 34.4 length mutations per kb in spacer DNA). From this it is clear that the sequences of the rRNA genes are being constrained by natural selection, which can only be because they are functional.

***Raison d'Être* of the Genome and its Translation.** The genetic apparatus of *Epifagus* ptDNA must be maintained to express at least one protein with a nonbioenergetic function that is essential even in a parasite. Without such a function, there would be no reason to express any gene. In addition to photosynthesis, plastids are the site of numerous metabolic processes, including the biosynthesis of amino acids, fatty acids, tetrapyrroles, isoprenoids, pyrimidine nucleotides, and vitamin B_1; the reduction of nitrite and sulfate; starch metabolism; and glycolysis (17). No genes involved in these or other nonbioenergetic processes have been mapped to the plastid genome in angiosperms, but one or more of the four unassigned protein genes retained in *Epifagus* ptDNA (Table 1) must have such a function. Homologs for these four genes have not been described in *Chlamydomonas* ptDNA, so it is not yet possible to investigate their functions genetically. Some of the eight conserved open reading frames present in tobacco ptDNA but absent from *Epifagus* (Table 1) may also have nonbioenergetic roles, but ones that have become defunct in a parasite. Several nonbioenergetic proteins, such as components of amino acid and sulfur metabolism, are encoded by ptDNAs of bryophytes or algae (14).

By assuming that none of the proteins of the gene-expression apparatus has an unrecognized nongenetic function, there are only four genes in *Epifagus* ptDNA that are even candidates for being the *raison d'être* of the genome and its translational apparatus¶ (Fig. 1). The functions of the two largest genes (*orf2216* and *orf1738*) are completely unknown, although a chromoplast-specific role has been proposed for the former in tomato (20). The third gene *accD* encodes the plastid homolog (21) of the β subunit of the carboxyltransferase component of *E. coli* acetyl-CoA carboxylase (22), which catalyzes the first committed step in fatty acid synthesis. The fourth gene *clpP* encodes the plastid homolog of the proteolytic subunit of the ATP-dependent Clp protease of *E. coli* (23). The function of plastid Clp protease—in processing, turnover, or even import of plastid proteins—is unknown, but this gene is a second strong candidate for the focus of the selective pressure that maintains the whole genome.

The most severe but least well understood case of a degenerate plastid genome is a 35-kb genome of unknown function in the malarial parasite *Plasmodium* and other members of the protist phylum Apicomplexa, which is probably derived from the plastid genome of dinoflagellates (24). A more similar situation to *Epifagus* is that of the nonphotosynthetic alga *Astasia longa*, whose plastid genome resembles that of the photosynthetic *Euglena gracilis* but has lost most photosynthetic genes and is only half the size (73 kb instead of 145 kb; ref. 25). Complete sequences are not available for the *Astasia* and *Euglena* plastid genomes, so it is not yet known whether *Astasia* lacks any gene-expression components. However, many intact translational genes have been identified in the 50% of *Astasia* ptDNA that has been sequenced, without any evidence for pseudogenes or missing translational genes (the rRNA gene spacer in *Astasia* ptDNA lacks the *trnI* and *trnA* genes present in all other plastid rRNA operons, including that of *Euglena*, but these genes have recently been found elsewhere in the *Astasia* genome; refs. 26 and 27 and W. Hachtel, personal communication). One clearly important difference in *Astasia* as compared to *Epifagus* is that the alga retains and expresses the CO_2-fixation gene *rbcL* (28). Also, *Astasia* and *Euglena* contain several, probably nongenetic, protein genes that are absent from angiosperm ptDNAs (26), and conversely, homologs of *clpP, accD, orf1738*, and *orf2216* have not been found in *Euglena* (whose ptDNA is mostly sequenced) or *Astasia*. Given all these gene content differences, the primary function(s) of the *Astasia* plastid genome is probably different from that of *Epifagus*.

The results from *Epifagus* show clearly that the products of angiosperm ptDNA have function(s) in addition to photosynthesis, chlororespiration, and gene expression. The plastid genome is still essential in this photosynthesis-deficient plant, unlike the mitochondrial genome in respiration-deficient mutants of yeast (29). Curiously, of the four genes remaining in *Epifagus* ptDNA that are not known to have genetic functions (Fig. 1), three (*accD, orf1738*, and *orf2216*) have no homologs in rice ptDNA (1), although they are present in other dicots and in the bryophyte *Marchantia* (1, 4). Their products must either be unnecessary in rice (though necessary in *Epifagus* and most other plants) or else the genes have been relocated to the nucleus. *clpP* is thus the only gene held in common by *Epifagus* and rice that is not part of the gene-expression apparatus. This raises the possibility that if *clpP* (and *trnE*) could be functionally relocated to the nuclear genome, a parasitic plant could exist without any ptDNA. In this regard, we note that *Epifagus* and the Orobanchaceae represent one of at least six lineages of flowering plants in which photosynthesis has been lost (30) and that the fate of ptDNA in these other lineages is unknown. Complete loss of a plastid genome would represent the final step in the reduction of gene content from cyanobacteria (>1000 genes not involved in gene expression) to chloroplasts (>50 genes) to the *Epifagus* plastid (1–4 genes).

¶The absence of RNA polymerase genes means that plastid *trnE*, which is intact in *Epifagus* and needed for mitochondrial heme synthesis (18), can serve as the *raison d'être* of the genome but not of the translational apparatus too. The situation of a nontranslated plastid genome retained solely because of *trnE* may exist for the grossly deleted genomes of certain pollen-derived albino plants from rice and other cereals (19).

We thank S. Ems, D. Katz-Downie, J. Donello, M. Drummond, J. Olsen, and K. Scott for technical assistance; C. dePamphilis, S. Downie, and S. Baldauf for helpful discussion; W. Hachtel for unpublished data; and the National Institutes of Health and the Alfred P. Sloan Foundation for support.

1. Shimada, H. & Sugiura, M. (1991) *Nucleic Acids Res.* **19,** 983–995.
2. dePamphilis, C. W. & Palmer, J. D. (1990) *Nature (London)* **348,** 337–339.
3. Morden, C. W., Wolfe, K. H., dePamphilis, C. W. & Palmer, J. D. (1991) *EMBO J.* **10,** 3281–3288.
4. Wolfe, K. H., Morden, C. W. & Palmer, J. D. (1992) *J. Mol. Biol.* **223,** 95–104.
5. Wolfe, K. H., Katz-Downie, D. S., Morden, C. W. & Palmer, J. D. (1992) *Plant Mol. Biol.* **18,** 1037–1048.
6. Wolfe, K. H., Li, W.-H. & Sharp, P. M. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 9054–9058.
7. Wolfe, K. H., Morden, C. W., Ems, S. & Palmer, J. D. (1992) *J. Mol. Evol.* **35,** 304–317.
8. Zurawski, G., Clegg, M. T. & Brown, A. H. D. (1984) *Genetics* **106,** 735–749.
9. Wolfson, R., Higgins, K. G. & Sears, B. B. (1991) *Mol. Biol. Evol.* **8,** 709–720.
10. Bogorad, L. (1991) in *Cell Culture and Somatic Cell Genetics of Plants: Molecular Biology of Plastids*, eds. Bogorad, L. & Vasil, I. K. (Academic, San Diego), Vol. 7A, pp. 93–124.
11. Hill, W. E., Dahlberg, A., Garrett, R. A., Moore, P. B., Schlessinger, D. & Warner, J. R., eds. (1990) *The Ribosome: Structure, Function & Evolution* (Am. Soc. Microbiol., Washington).
12. Toukifimpa, R., Romby, P., Rozier, C., Ehresmann, C., Ehresmann, B. & Mache, R. (1989) *Biochemistry* **28,** 5840–5846.
13. Pfitzinger, H., Weil, J. H., Pillay, D. T. N. & Guillemaut, P. (1990) *Plant Mol. Biol.* **14,** 805–814.
14. Wolfe, K. H., Morden, C. W. & Palmer, J. D. (1991) *Curr. Opin. Genet. Devel.* **1,** 523–529.
15. Wimpee, C. F., Wrobel, R. L. & Garvin, D. K. (1991) *Plant Mol. Biol.* **17,** 161–166.
16. Wimpee, C. F., Morgan, R. & Wrobel, R. (1992) *Plant Mol. Biol.* **18,** 275–285.
17. Weeden, N. F. (1981) *J. Mol. Evol.* **17,** 133–139.
18. Howe, C. J. & Smith, A. G. (1991) *Nature (London)* **349,** 109.
19. Harada, T., Ishikawa, R., Niizeki, M. & Saito, K. (1992) *Mol. Gen. Genet.* **233,** 145–150.
20. Richards, C. M., Hinman, S. B., Boyer, C. D. & Hardison, R. C. (1991) *Plant Mol. Biol.* **17,** 1179–1188.
21. Smith, A. G., Wilson, R. M., Kaethner, T. M., Willey, D. L. & Gray, J. C. (1991) *Curr. Genet.* **19,** 403–410.
22. Li, S.-J. & Cronan, J. E., Jr. (1992) *J. Biol. Chem.* **267,** 16841–16847.
23. Goldberg, A. L. (1992) *Eur. J. Biochem.* **203,** 9–23.
24. Palmer, J. D. (1992) *Curr. Biol.* **2,** 318–320.
25. Siemeister, G. & Hachtel, W. (1989) *Curr. Genet.* **15,** 435–441.
26. Siemeister, G., Bucholz, C. & Hachtel, W. (1990) *Mol. Gen. Genet.* **220,** 425–432.
27. Siemeister, G. & Hachtel, W. (1990) *Curr. Genet.* **17,** 433–438.
28. Siemeister, G. & Hachtel, W. (1990) *Plant Mol. Biol.* **14,** 825–833.
29. Goldring, E. S., Grossman, L. I., Krupnick, D., Cryer, D. R. & Marmur, J. (1970) *J. Mol. Biol.* **52,** 323–335.
30. Kuijt, J. (1969) *The Biology of Parasitic Flowering Plants* (Univ. Calif. Press, Berkeley).