J Mol Evol (1992) 35:304–317

Journal of
Molecular Evolution

Rapid Evolution of the Plastid Translational Apparatus in a Nonphotosynthetic Plant: Loss or Accelerated Sequence Evolution of tRNA and Ribosomal Protein Genes

Kenneth H. Wolfe*, Clifford W. Morden†, Stephanie C. Ems, and Jeffrey D. Palmer

Department of Biology, Indiana University, Bloomington, IN 47405, USA

Summary. The vestigial plastid genome of *Epifagus virginiana* (beechdrops), a nonphotosynthetic parasitic flowering plant, is functional but lacks six ribosomal protein and 13 tRNA genes found in the chloroplast DNAs of photosynthetic flowering plants. Import of nuclear gene products is hypothesized to compensate for many of these losses. Codon usage and amino acid usage patterns in *Epifagus* plastic genes have not been affected by the tRNA gene losses, though a small shift in the base composition of the whole genome (toward A + T-richness) is apparent. The ribosomal protein and tRNA genes that remain have had a high rate of molecular evolution, perhaps due to relaxation of constraints on the translational apparatus. Despite the compactness and extensive gene loss, one translational gene (*inf*A, encoding initiation factor 1) that is a pseudogene in tobacco has been maintained intact in *Epifagus*.

Key words: tRNAs — ribosomal proteins — plastid translation — accelerated evolution — gene loss — codon usage — *Epifagus virginiana*

Introduction

Epifagus virginiana (beechdrops; family Orobanchaceae) is a nongreen parasitic flowering plant that grows on the roots of beech trees. *Epifagus* is completely nonphotosynthetic, deriving all its carbon from the host plant, but its tissues contain plastids visible by electron microscopy (Walsh et al. 1980). We have recently characterized the plastid genome of *Epifagus,* initially by filter hybridization using probes from tobacco chloroplast DNA (dePamphilis and Palmer 1990), and later by cloning and complete nucleotide sequencing (Wolfe, Morden, and Palmer, submitted). The genome is greatly reduced in size (70 kb as compared to 156 kb in tobacco; Shinozaki et al. 1986): all of the bioenergetic (photosynthetic and chlororespiratory) genes normally present in chloroplast genomes have been deleted, though fragments of a few remain as pseudogenes (Fig. 1). Of the 42 intact genes in *Epifagus* plastid DNA (ptDNA), at least 38 encode components of the genetic apparatus (chiefly ribosomal proteins, rRNAs, and tRNAs), and the remaining four encode proteins of largely unknown function. We have proposed that at least one of these four proteins must function in an unidentified nonphotosynthetic process that is the reason for the maintenance of the plastid genome (Wolfe et al., submitted).

Direct evidence that the genome is functional has been obtained by northern blot detection of ribosomal RNAs and protein gene transcripts and by sequencing PCR products derived from spliced plastid transcripts (dePamphilis and Palmer 1990; Ems and Palmer, unpublished). The evolutionary arguments in favor of functionality are also strong: (1) despite the numerous and extensive deletions that have occurred, homologs of some large protein genes (two of which are ~2,000 codons in size) present in chloroplast genomes are intact in *Epifagus;* and (2) the

* *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, USA
Offprint requests to: J.D. Palmer

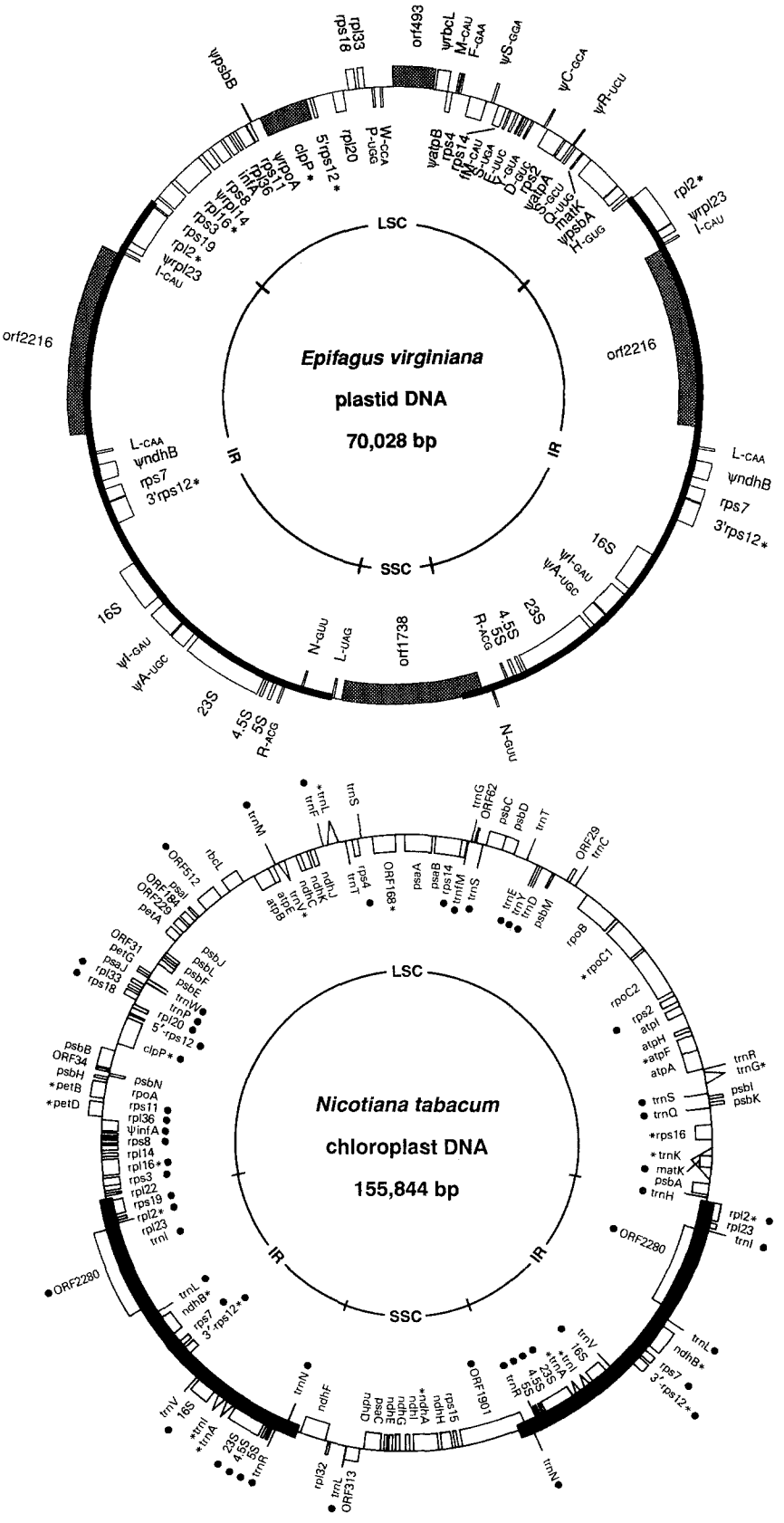

Fig. 1. Gene organization of the plastid genomes of *Epifagus virginiana* (Wolfe et al., submitted) and *Nicotiana tabacum* (modified from Shinozaki et al. 1986). The two circular maps are drawn to the same size rather than the same scale. The *Epifagus* genome contains an inverted repeat (IR) of 22,735 bp that divides the rest of the genome into large single-copy (LSC; 19,799 bp) and small single-copy (SSC; 4,759 bp) regions. In *Nicotiana*, the IR is 25,339 bp, the LSC is 86,684 bp, and the SSC is 18,482 bp. Genes drawn inside the circles are transcribed clockwise. tRNA genes are identified in *Epifagus* by the one-letter amino acid code and their anticodon sequence, and in *Nicotiana* by "trn" followed by the one-letter code. Pseudogenes are marked "ψ". Asterisks indicate genes containing introns. The four genes in *Epifagus* that are not known to be involved in gene expression are shaded. Filled circles in the tobacco map indicate the 42 genes (38 of which encode translational components) that are retained in an intact form in *Epifagus*.

distribution of deletions is heavily skewed so that most translational (i.e., ribosomal protein, rRNA, and tRNA) genes are still present whereas none of the 40 photosynthetic or chlororespiratory (*ndh*) genes remain intact (Morden et al. 1991; Wolfe et al. 1992a,b and submitted).

Despite the compelling evidence that the genome is functional, genes for several components of the

gene-expression apparatus were found to be lacking from *Epifagus* ptDNA (Morden et al. 1991; Wolfe et al. 1992a,b and submitted). All four RNA polymerase subunit genes identified in tobacco ptDNA have been "lost" (i.e., are deleted or are pseudogenes), as have six of the 21 ribosomal protein genes and 13 of the 30 tRNA genes. In this paper we describe in detail the changes in the ribosomal protein and tRNA gene sets and discuss their implications for gene expression in *Epifagus* plastids. We show that the loss of tRNA genes has had no discernable impact on codon usage in the genome, and that minor changes in the amino acid composition of proteins are more probably due to drift in the genome's base composition than to natural selection. Phylogenetic analysis of retained ribosomal protein and tRNA sequences reveals consistently faster evolution in *Epifagus* than in tobacco, as was previously found for two ribosomal proteins (Morden et al. 1991) and the 16S and 23S rRNAs (Wolfe et al. 1992b).

Materials and Methods

A set of 24 overlapping plasmid clones spanning the *Epifagus virginiana* plastid genome was isolated from libraries of total genomic DNA by standard techniques, except that polymerase chain reaction (PCR) amplification of total DNA was used to close three small gaps between clones. Details of the clone bank will be given elsewhere. Clones covering the 47.3 kb of unique sequence were sequenced on both strands by the dideoxy chain-termination method. An overview of the sequence will be presented elsewhere (Wolfe et al. submitted). The GenBank/EMBL/DDBJ accession number is M81884.

The *inf*A gene in tobacco ptDNA was sequenced from two sources. First, the 640-bp *Sal*I-11 fragment (positions 82117–82757; Shinozaki et al. 1986) that contains the whole gene was subcloned from pTBa1 (Sugiura et al. 1986) and completely sequenced. Second, a 2.4-kb region spanning the *inf*A locus was amplified by PCR from purified tobacco ptDNA (a gift from Dr. Claude dePamphilis) using primers based on sequences of the nearby genes *rpl*16 and *rps*11. The PCR product was purified by agarose gel electrophoresis and digested with *Sal*I. Two independent clones of the 640-bp fragment were partially sequenced in opposite directions on one strand each. The region of overlap includes the frameshift correction site and most of the *inf*A coding region, including the start codon. One PCR clone yielded 482 bp of sequence identical to that from the pTBa1 subclone, but the other differed by three transition mutations in the 583-bp read. We attribute these differences to PCR artifacts.

Phylogenetic trees used to examine rates of molecular evolution were constructed by parsimony analysis of amino acid or nucleotide sequences (PAUP; D. Swofford, Illinois Natural History Survey). Branch lengths were estimated assuming accelerated character transformation (ACCTRAN).

Results and Discussion

Ribosomal Protein Genes

The plastid genomes of tobacco and rice encode 21 of the approximately 60 plastid ribosomal proteins, the remainder being nuclear-encoded (Fig. 1; Shinozaki et al. 1986; Hiratsuka et al. 1989; Yokoi et al. 1990; Subramanian et al. 1991). Six of these 21 genes are nonfunctional in *Epifagus*, as a result of either complete deletion (*rps*15, *rps*16, *rpl*22, *rpl*32) or smaller mutations resulting in pseudogenes (*rpl*14, *rpl*23). The complement of ribosomal proteins varies somewhat among eubacteria (e.g., Isono and Isono 1976), so it is possible that the gene losses in *Epifagus* have been inconsequential. However, based on inference from the tertiary structure of the *Escherichia coli* ribosome, protein S15 is expected to be essential for ribosome assembly and is known to bind to 16s rRNA. [See Hill et al. (1990) and Harris et al. (1992).] Chloroplast ribosomal protein L22 binds 5S rRNA and so may also be a central component of the ribosome (Toukifimpa et al. 1989). It therefore seems unlikely that these two proteins are simply absent from *Epifagus* ribosomes and more probable that the losses are compensated for by nuclear gene products. Because the roles in *E. coli* of the homologs of the other four ribosomal proteins are poorly known (see Harris et al. 1992), little can be said about the consequences of their loss.

The nucleus can compensate for the loss of a plastid ribosomal protein gene either by the physical transfer of the gene to the nuclear genome or by functional substitution of a nuclear gene product for the missing plastid counterpart. Of the six ribosomal protein genes lost in *Epifagus*, three have also been lost in some photosynthetic lineages. *rpl*22 is missing from legume ptDNA, and phylogenetic analysis of the pea nuclear gene for plastid L22 indicates that this gene originated by duplication of the plastid gene in an ancestor of flowering plants, long before its loss from legume ptDNA (Gantt et al. 1991). The same nuclear gene may therefore have underwritten subsequent independent losses of plastid genes in legumes and *Epifagus*. *rpl23* is a pseudogene in ptDNA of the spinach family, and *rps*16 is absent from ptDNA in *Marchantia* (liverwort) and probably some angiosperm species (Zurawski and Clegg 1987; Ohyama et al. 1986; Downie and Palmer 1992). Their independent loss from *Epifagus* suggests that these may also be cases of ancient gene duplications to the nucleus. To date, loss of *rps*15, *rpl*14, or *rpl*32 has not been reported in any other angiosperm ptDNA. (See Downie and Palmer 1992.) No clear examples of functional substitution of a nuclear gene for a recently lost plastid gene have been reported, but this may have occurred in the case of *rpl*21 in angiosperms (Martin et al. 1990). Such substitution could involve nuclear genes for either cytoplasmic or mitochondrial proteins, and could occur via gene duplication (Tingey et al. 1988) or by means of a protein that functions in two cellular compartments (Surguchov 1987).

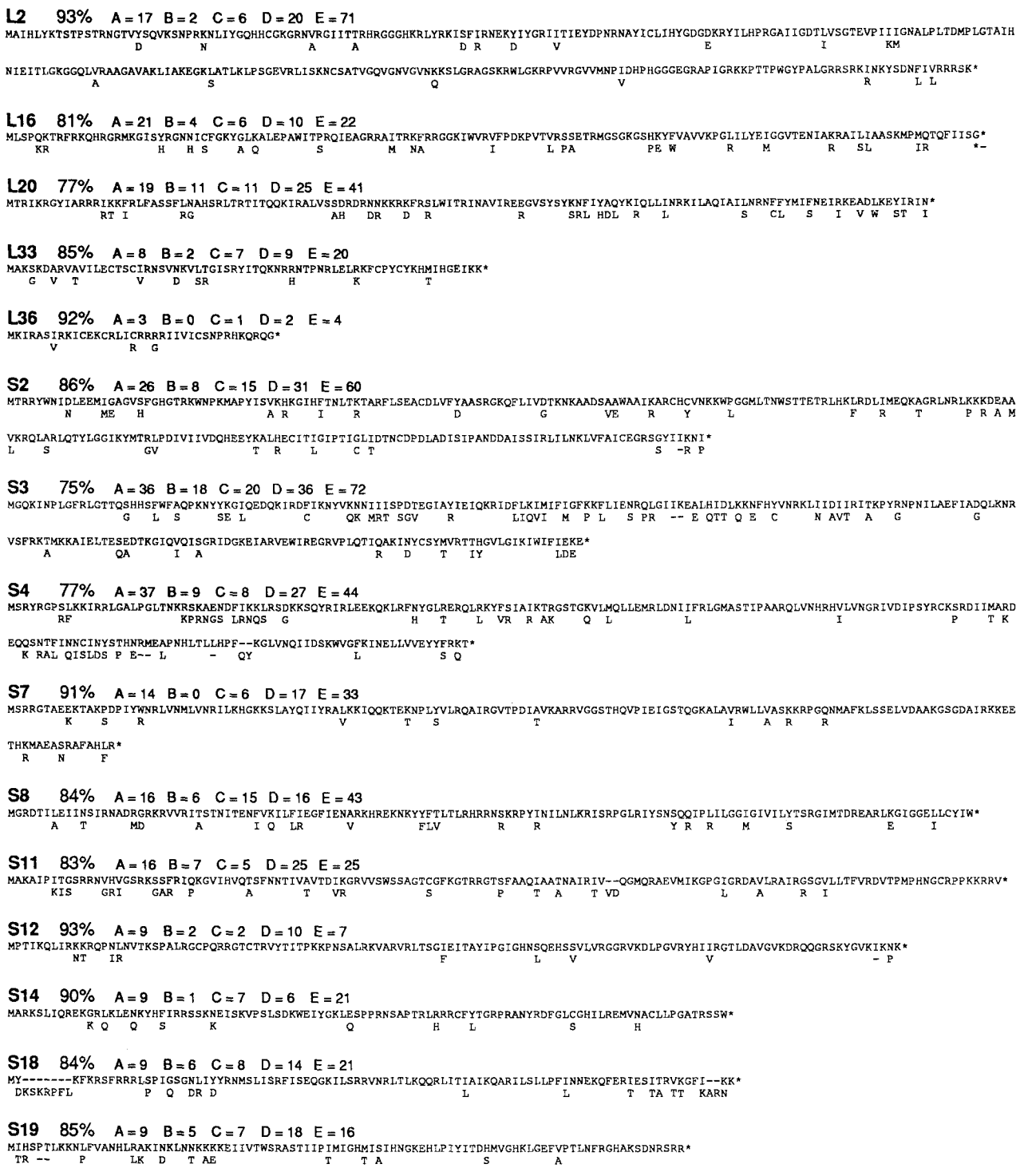

Fig. 2. Deduced amino acid sequences of the 15 ribosomal proteins encoded by *Epifagus* ptDNA. Differences in the homologous tobacco sequences (Shinozaki et al. 1986) are indicated below the *Epifagus* sequences, and the percentage of identical residues (gaps excluded) is shown. Asterisks indicate stop codons; dashes are gaps introduced to improve alignment. The numbers at A–E are branch lengths (numbers of amino acid replacements) in phylogenetic trees drawn for each protein, using sequences from *Epifagus,* tobacco, rice, and *Marchantia,* as shown in Fig. 7a.

The 15 ribosomal protein genes that remain in *Epifagus* ptDNA encode proteins with 75–93% amino acid sequence identity to tobacco, and very few gaps are needed to align them (Fig. 2). A few aspects of the alignments are noteworthy. The S3 and S4 proteins are particularly divergent, being the only sequences that contain strings of five or more consecutively mismatched residues. S3 is also poorly conserved in rice, where an internal 15-amino-acid region has been duplicated (Hiratsuka et al. 1989). Divergence at the C-terminus of S12 is due to a single-nucleotide deletion in *Epifagus*. The

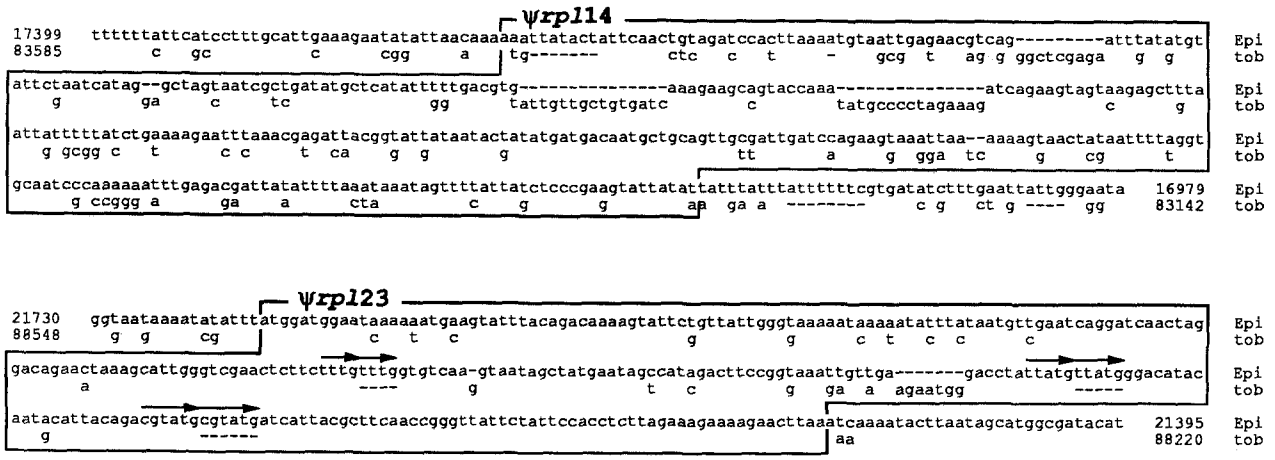

Fig. 3. Nucleotide sequence alignments between *Epifagus* pseudogenes and tobacco genes at the *rpl*14 and *rpl*23 loci. The complete *Epifagus* sequences are shown, below which are the differences in tobacco. Numbering refers to genomic coordinates (Wolfe et al., submitted; Shinozaki et al. 1986). Arrows indicate tandem repeat insertions in the *Epifagus* genome.

N-terminus of *Epifagus* S18 contains only one copy of the heptapeptide motif that is repeated two, six, and seven times in tobacco, rice, and maize, respectively (Wegloehner and Subramanian 1991). *Epifagus rps*19 conserves the GTG start codon found in most other species and, as described elsewhere (Morden, Wolfe, and Palmer, unpublished), includes six nucleotides that are shared with the 3′ end of a tRNA gene at the junction between the IR and the large single-copy (LSC) region.

Pseudogenes of rpl14 and rpl23

The *rpl*14 and *rpl*23 pseudogenes are both full-length, in contrast to the truncation that has occurred at most of the other *Epifagus* pseudogene loci. The nucleotide sequences have identities of 79% (*rpl*14) and 93% (*rpl*23) to tobacco. We have categorized them as pseudogenes on the basis of numerous frameshift mutations (Fig. 3). In ψ*rpl*14 there are two insertions and two deletions that result in frameshifts, as well as three deletions that do not change the reading frame. Mutated sites in ψ*rpl*14 include the homologs of both the start and the stop codons used in tobacco.

Frameshifts in ψ*rpl*23 result from two insertions and two deletions; there is also a 6-bp insertion. All three of the insertions (4, 5, and 6 bp) result from direct tandem duplications (arrows in Fig. 3). Zurawski and Clegg (1987) have reported ψ*rpl*23 sequences from seven species in the Caryophyllidae, and these show a similar pattern of short insertions and deletions, many associated with tandem direct repeats. Since tobacco *rpl*23 is functional (Yokoi et al. 1991), the pseudogenes in *Epifagus* and the Caryophyllidae must have independent origins. Remarkably, however, the 5-bp tandem duplication (TTATG) in *Epifagus* ψ*rpl*23 is identical (both in site and in sequence) to a mutation in *Kochia* ψ*rpl*23. In addition, a 14-bp deletion in spinach and four related species overlaps with the 4-bp insert and 1-bp deletion in *Epifagus*. This might indicate hotspots for length mutation or just be a coincidence. (Note that other length mutations in the ψ*rpl*23 sequences do not occur at the same locations.) The *rpl*23 locus in pea (Nagano et al. 1991) is interrupted by a 190-bp insertion flanked by 37-bp duplications of an internal region of the gene. Since the insert is close to the 3′ end of *rpl*23, and this region is poorly conserved among other species (Yokoi et al. 1991), it is unclear whether the pea gene is functional.

The pattern of mutation in the *Epifagus* ψ*rpl*14 and ψ*rpl*23 sequences is quite biased. Out of a total of 87 nucleotide differences within the pseudogenes (Fig. 3), there are 58 positions where a G or C in tobacco corresponds to an A or T in *Epifagus*, whereas the converse occurs at only 12 sites (the remainder being A ⟷ T or G ⟷ C changes). This is an expected consequence of loss of function in a plastid gene. Mutation pressure toward A + T-richness in functional plastid genes is normally counteracted by natural selection on amino acid sequences, so that first and second positions of codons are more G + C-rich than are third positions and intergenic regions (Osawa et al. 1988). Once selection on the gene product is removed the A + T content of a gene should increase. This effect is exacerbated by the higher A + T content of the whole *Epifagus* plastid genome (Table 1).

infA in Epifagus and Tobacco

The plastid gene *inf*A, encoding a homolog of translational initiation factor 1 (IF1) from bacteria, was first described from spinach ptDNA (Sijben-

Table 1. Base composition in angiosperm plastid genomes

Region	G + C (%)	Length (bp)	Silent-site G + C (%)[a]	No. of codons[a]
Epifagus LSC	29.2	19,799	25.0	2805
Epifagus SSC	22.7	4,759	17.5	1468
Epifagus IR	40.3	22,735	33.2	3020
Epifagus total[b]	33.9	47,293	27.7	7293
Tobacco LSC	36.0	86,684	28.3	2945
Tobacco SSC	32.1	18,482	27.5	1569
Tobacco IR	43.2	25,339	35.5	3126
Tobacco total[b]	36.8	130,505	31.3	7640
Rice LSC[c]	37.1	80,592	30.0	2482
Rice SSC[d]	33.4	12,335	—	0
Rice IR[d,e]	44.3	20,799	33.9	608
Rice total[b]	38.0	113,726	30.9	3090

[a] Only the 21 protein genes in *Epifagus* ptDNA (Wolfe et al., submitted) and their homologs in tobacco and rice are considered. Silent-site G + C content is that at fourfold degenerate sites in codons

[b] Totals include the IR only once

[c] The rice LSC lacks a homolog of *Epifagus orf*493/tobacco *orf*512 ("*zfp*A")

[d] The rice SSC and IR lack a homolog of *Epifagus orf*1738/tobacco *orf*1901

[e] The rice IR lacks a homolog of *Epifagus orf*2216/tobacco *orf*2280

Mueller et al. 1986). This gene is also present in the plastid genomes of rice and *Marchantia* (Hiratsuka et al. 1989; Ohyama et al. 1986), but Shinozaki et al. (1986) reported that tobacco *inf*A lacks a start codon and might be a pseudogene. *Epifagus inf*A is intact, has 84% amino acid sequence identity to spinach *inf*A, and is of identical length (77 codons). This is the only example of a gene that is apparently functional in *Epifagus* but not in tobacco, and runs counter to the general pattern of gene losses.

Since there are no other known pseudogenes in tobacco ptDNA, and the possibility of a sequencing error cannot be dismissed, we confirmed the pseudogene status of tobacco *inf*A by resequencing. We identified one error in the *inf*A sequence of Shinozaki et al. (1986): GG reported at positions 82453–82454 is in fact a single G, both in clone pTBa1 (Sugiura et al. 1986) and in independently cloned tobacco ptDNA. (See Materials and Methods.) This correction removes a frameshift near the 5′ end of tobacco *inf*A, keeping the tobacco and *Epifagus* sequences in register farther upstream than originally thought. However, tobacco still lacks a start codon as the result of an 11-nucleotide deletion that leaves a TAA stop codon in its place (Fig. 4). Another mutation at the position corresponding to the *Epifagus* stop codon extends the tobacco reading frame at the 3′ end. The revised tobacco *inf*A locus consists of 105 sense codons (positions 82493–82178) bounded by stop codons.

A start codon might be created for tobacco *inf*A by means of plastid RNA editing (Hoch et al. 1991; Kudla et al. 1992), but this cannot be achieved by the C → U editing mechanism reported because it would not overcome the TAA codon at the 5′ end of the gene (Fig. 4). It seems a remarkable coincidence that the gene for IF1 should lack an initiation codon, especially since an unusual start codon (ATT) occurs in *E. coli inf*C (encoding IF3) and is involved in its autoregulation. (See Butler et al. 1987.) Tobacco *inf*A is probably cotranscribed with the flanking ribosomal protein genes (Ohto et al. 1988), so it is perhaps possible that it could be translated, albeit by mechanisms not needed in other plants.

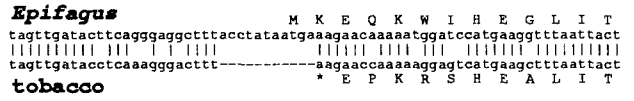

Fig. 4. Comparison of *Epifagus* and tobacco sequences at the 5′ end of *inf*A. Potential translations are shown. Dashes indicate an 11-nucleotide deletion in tobacco that results in a TAA stop codon (asterisk).

The *inf*A locus is clearly a pseudogene in *Pelargonium hortorum* (geranium; P.J. Calie and J.D. Palmer, unpublished data), which is an outgroup to *Epifagus* and tobacco. If tobacco *inf*A is also nonfunctional and if IF1 is nuclear-encoded in tobacco and *Pelargonium*, two independent transfers or activations of genes may have occurred; otherwise there is no apparent reason why *inf*A should have been maintained in *Epifagus* ptDNA. The possibility that there is an active nuclear *inf*A gene in all of these dicots and that the *Epifagus* and spinach plastid sequences are silent (i.e., "intact pseudogenes" similar to *cox*II in soybean mitochondrial DNA; Nugent and Palmer 1991) is unlikely because the majority of nucleotide substitutions between the

Epifagus and spinach *inf*A sequences are synonymous.

tRNA Genes and Pseudogenes

Epifagus ptDNA completely lacks eight of the 30 tRNA genes present in tobacco and rice (Shinozaki et al. 1986; Hiratsuka et al. 1989), and a further five tRNA loci are identified as pseudogenes (Figs. 1, 5, and 6). Distinguishing between intact tRNA genes and pseudogenes is not straightforward given their small size and the extent of variation seen in functional tRNAs (e.g., Okimoto and Wolstenholme 1990). Four of the *Epifagus* tRNA pseudogenes contain multiple mutations, including insertions and/or deletions (Fig. 5). The fifth ($\psi trnR_{UCU}$; Morden et al. 1991) is more intact but has an abnormal D-loop sequence that lacks a pair of guanine residues that are highly conserved among tRNA genes from all sources other than mitochondria (Sprinzl et al. 1989).

The 17 tRNA genes regarded as intact (Fig. 5) have no deletions, though single-base insertions in three genes ($trnS_{UGA}$, $trnW_{CCA}$, and $trnY_{GUA}$) occur at positions where some length variation is tolerable (Sprinzl et al. 1989). The stems of the 17 predicted *Epifagus* tRNAs contain a total of 15 mismatches as compared to only five in their tobacco homologs (Sugiura 1987). Four substitutions have occurred at quasi-invariant positions identified by Wakasugi et al. (1986) in their analysis of the tobacco tRNA genes, but whether these are sufficient to make these tRNAs nonfunctional is not known. These substitutions are located in the D-stem of $tRNA^{Arg}$ (ACG) and the acceptor stems of $tRNA^{Phe}$ (GAA) and $tRNA^{Ser}$ (GCU) (Fig. 5). The two substitutions in $tRNA^{Ser}$ (GCU) occur at opposing positions in the acceptor stem, replacing a G:C pair with an A:U pair. This is of interest because the normal G_2:C_{71} pair at this position forms part of the recognition site for specific aminoacylation with serine in *E. coli* and is well conserved across eubacterial serine tRNAs (Schimmel 1991; Sprinzl et al. 1989).

The 13 tRNA species encoded by ptDNA in tobacco and rice, but whose genes are no longer intact in *Epifagus,* are: both isoacceptors for Gly, Thr, and Val; the single tRNAs for Ala, Cys, and Lys; and one isoacceptor (out of two or three) for Arg, Ile, Leu, and Ser (Figs. 1 and 6). We have proposed that tRNAs are imported from the cytoplasm into *Epifagus* plastids (Morden et al. 1991), by analogy to the situation in angiosperm mitochondria (Marechal-Drouard et al. 1988; Joyce and Gray 1989). Import of nuclear-encoded tRNAs into mitochondria has also been reported in the unicellular eukaryotes *Leishmania* and *Trypanosoma,* whose mitochondrial genomes contain no tRNA genes (Simpson et al. 1989; Hancock and Hajduk 1990), and *Tetrahymena,* where only 8–10 tRNAs are mitochondrial-encoded. (See Ziaie and Suyama 1987.) One cytoplasmic tRNA is imported into yeast mitochondria but is apparently inactive (Martin et al. 1979). Import of tRNA into mitochondria is suspected in other organisms such as *Paramecium* and *Chlamydomonas* (Pritchard et al. 1990; Gray and Boer 1988) whose mitochondrial genomes lack a full set of tRNA genes.

Two of the other three hypotheses previously proposed as alternatives to tRNA import into *Epifagus* plastids (Morden et al. 1991) can be ruled out now that the complete sequence of the genome is known and the fates of all the tRNA genes have been determined. "Anticodon switching" at the DNA level (conversion of a redundant tRNA gene into another by specific point mutations; e.g., Schulman and Pelka 1988) has not occurred: all the intact tRNA genes have unmutated anticodons. There are in fact only 15 different unmodified anticodons among the 17 intact tRNA genes because the anticodon CAU occurs three times (Met, fMet, and Ile; Fig. 5).

Enhanced wobble rules for codon–anticodon interactions were also proposed (Morden et al. 1991). However, even if these rules are liberalized to the extent of ignoring the third codon position where convenient, codons for the six amino acids for which there is no tRNA species, as well as the AGR codons for Arg, could not be translated (Fig. 6). Liberal codon–anticodon wobble rules of the type seen in mitochondria and *Mycoplasma* (Andachi et al. 1989) could potentially compensate for only one of the 13 tRNA gene losses (trn^{S}_{GGA}).

Another possibility, rampant mistranslation through the use of only 17 tRNAs, is unlikely because the normal excess of synonymous-over-nonsynonymous nucleotide substitutions is still seen in codons for Ala, Cys, Gly, Lys, Thr, and Val (pooled together) in the *Epifagus* lineage. To examine this, *Epifagus,* tobacco, and rice nucleotide sequences of the 15 ribosomal protein genes held in common were joined and aligned using amino acid sequence alignments as a guide. All aligned codons were then omitted from the analysis except those where two or all three of the species specified the same amino acid and that amino acid was Ala, Cys, Gly, Lys, Thr, or Val. Extents of synonymous and nonsynonymous substitution (Li et al. 1985) were then calculated between each pair of modified sequences (totaling 682 codons) and branch lengths were computed for each lineage. The ratio of synonymous-to-nonsynonymous substitutions (K_S/K_A)

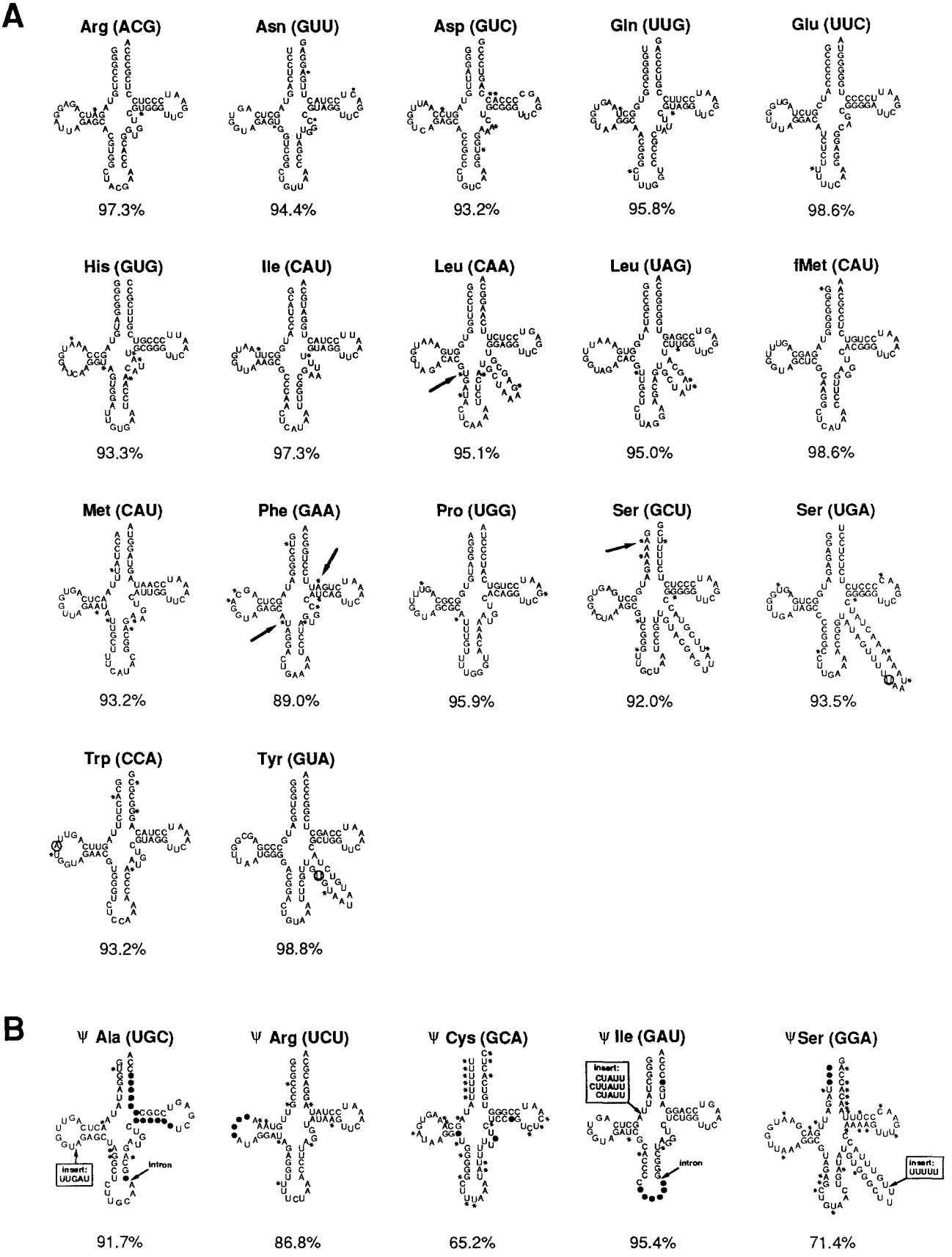

Fig. 5. Cloverleaf representations of the 17 tRNA genes (A) and five tRNA pseudogenes (B) of *Epifagus* ptDNA. Asterisks indicate positions that differ from the tobacco sequences; dots indicate deleted nucleotides and circles show insertions. Large arrows indicate four positions in the intact genes at which compensatory pairs of substitutions have occurred. Percent nucleotide identity with the tobacco homolog is indicated below each tRNA structure.

TTT F	337/247	TCT S	150/180	TAT Y	269/215	TGT C	70/ 67
TTC F ●	92/152	TCC S ⊗	77/101	TAC Y ●	57/ 52	TGC C ⊗	24/ 25
TTA L ○	253/202	TCA S ●	140/135	TAA *	12/ 12	TGA *	3/ 4
TTG L ●	124/164	TCG S	51/ 71	TAG *	6/ 5	TGG W ●	104/123
CTT L	150/173	CCT P	88/ 99	CAT H	121/144	CGT R ●	96/109
CTC L	45/ 61	CCC P	59/ 72	CAC H ●	41/ 37	CGC R	34/ 29
CTA L ●	98/ 91	CCA P ●	79/ 86	CAA Q ●	190/237	CGA R	83/138
CTG L	40/ 69	CCG P	22/ 41	CAG Q	44/ 63	CGG R	35/ 43
ATT I	315/303	ACT T	127/120	AAT N	411/358	AGT S	132/138
ATC I ⊗	90/125	ACC T ○	62/ 67	AAC N ●	101/120	AGC S ●	25/ 26
ATA I ●	328/209	ACA T ○	109/119	AAA K ○	536/438	AGA R ⊗	180/198
ATG M ●	117/142	ACG T	33/ 46	AAG K	125/159	AGG R	65/ 77
fM ●	21/ 20						
GTT V	116/109	GCT A	88/103	GAT D	219/277	GGT G	119/115
GTC V ○	37/ 48	GCC A	40/ 49	GAC D ●	54/ 68	GGC G ○	26/ 44
GTA V ○	117/107	GCA A ⊗	87/ 98	GAA E ●	280/314	GGA G ○	145/159
GTG V	35/ 50	GCG A	25/ 20	GAG E	89/116	GGG G	66/ 72

Fig. 6. Codon usage and anticodon sets in *Epifagus* and tobacco ptDNAs. To the left of the slash marks are the number of occurrences of each codon, summed over the 21 protein genes in *Epifagus* ptDNA; the numbers on the right are equivalent numbers in the 21 homologous tobacco genes (Shinozaki et al. 1986). Each box represents a group of codons thought to be recognized by a single tRNA in tobacco, and the dots indicate the codons recognized without wobble (i.e., the complements of the anticodons). Black dots represent the tRNA genes that are intact in *Epifagus;* stippled dots represent pseudogenes; open dots represent tRNA genes completely deleted. The total numbers of sense codons are 7293 in *Epifagus* and 7460 in tobacco. This slight difference in gene length was taken into account in statistical calculations. Tobacco *infA* was assumed to be functional and to begin with the first of the 105 sense codons. (See text.) $tRNA^{fMet}$ (CAU) was assumed to recognize the GTG start codon of *rps*19.

in the *Epifagus* lineage was 3.8, as compared to 7.4 for tobacco and 4.3 for rice. These results suggest selection against replacement of these residues, which in turn suggests that the codons are being translated correctly in *Epifagus*.

The only remaining alternative to tRNA import is the possibility (Morden et al. 1991) that the transcripts of the tRNA genes undergo variable extents of base modification ("RNA editing") so that a sufficient number of different tRNA species to decode all 61 sense codons can be produced from just 17 genes. However, this possibility seems unlikely, as RNA editing appears to be infrequent in plastids (Hoch et al. 1991; Kudla et al. 1992), and as no examples are known where the products of a single tRNA gene correctly translate codons for two different amino acids, even in systems such as plant mitochondria where RNA editing is quite extensive (Schuster et al. 1991). Sequencing of tRNAs and proteins from *Epifagus* plastids will ultimately be necessary to decide among hypotheses.

Extent of tRNA Import Into Plastids

If tRNAs are indeed imported into *Epifagus* plastids, questions arise as to whether this process also occurs in photosynthetic species, and whether all cytoplasmic tRNA species can be imported or just a few. We suspect that tRNA import into plastids occurs in all flowering plants because it seems unlikely that an import mechanism would have been developed especially in the parasitic lineage leading to *Epifagus* (Morden et al. 1991). However, there is no evidence that imported tRNAs play any role in the biochemistry of chloroplasts. In a survey of the distribution of the plastid $tRNA^{Cys}$ gene in species closely related to *Epifagus* we found that this tRNA gene was still intact in ptDNA of photosynthetic parasites and was lost only in two completely nonphotosynthetic species (*Epifagus* and *Conopholis*) (Taylor et al. 1991). This suggests that tRNA gene loss from ptDNA is permitted only in the absence of photosynthesis, when the total amount of translation occurring in plastids is relatively low.

While we cannot tell whether all cytoplasmic tRNA species can be imported across the plastid envelope, the apparently selective retention of some plastid tRNA genes in *Epifagus* suggests that cytoplasmic tRNAs cannot functionally substitute for all of the ptDNA-encoded tRNAs. This may be because some cytoplasmic tRNAs cannot be imported, or because they cannot interact with the

plastid's translational machinery. This interaction primarily involves the charging of the tRNA by a plastid-localized aminoacyl-tRNA synthetase and the recognition of the charged tRNA by elongation factors for incorporation into protein synthesis. Steinmetz and Weil (1986) have shown that spinach chloroplast aminoacyl-tRNA synthetases could charge some, but not all, of the cytoplasmic tRNAs tested. Such incompatibility between cytoplasmic ("eukaryotic") tRNAs and the plastid ("prokaryotic") translational apparatus is not unexpected, and may restrict the number of cytoplasmic tRNA species that can function *in organello*.

Three observations suggest that at least some of the 17 intact tRNA genes remaining in *Epifagus* ptDNA are still functional and are being maintained by natural selection. First, some tRNA genes in the single-copy regions have been retained, even though deletions of other genes have occurred in adjacent regions both upstream and downstream (Fig. 1). The retention of *trn*L_{UAG} despite loss of most of the rest of the small single-copy region is a striking example of this. (See also Morden et al. 1991.) In the large single-copy region there are seven more such sites where an intact tRNA gene or gene cluster is the only sequence remaining between two deleted regions. (These seven are, from left-to-right in Fig. 1, *trn*WP, M, F, S, EYD, S, Q.) Given the small size of the tRNA genes (as against the tens of kilobases of photosynthetic genes deleted), this is a far greater extent of retention than would be expected if these sequences were nonfunctional. Second, the intact tRNA genes have sustained numerous point mutations [parsimony analysis assigns a total of 59 nucleotide substitutions (see Fig. 7*b*) and to insertions to the *Epifagus* lineage], but none of these occurs at positions crucial to tRNA function (e.g., Schimmel 1991), and only four occur at positions highly conserved among chloroplast tRNAs (Wakasugi et al. 1986). In contrast, each of the putative pseudogenes except for ψ*trn*R_{UCU} contains damaging mutations at two or more distinct sites. Furthermore, there are four positions in the *Epifagus* tRNAs where a compensatory pair of substitutions on either side of a stem has maintained base pairing (arrows in Fig. 5). Comparison to other species indicates that in each case the substitutions have occurred in the *Epifagus* lineage (though possibly in a photosynthetic ancestor). These results suggest that there is not a continuum of levels of divergence in the tRNA genes but instead two sets of genes, one well conserved and one not. Third, a related nonphotosynthetic parasite, *Conopholis americana* (squawroot; Orobanchaceae), appears to contain a similar subset of tRNA genes (C.W. dePamphilis, S.R. Downie, and J.D. Palmer, unpublished data). While some of this is likely to be due to shared ancestry, the extent of sequence divergence between the two species is quite large (*Conopholis* and *Epifagus* plastid 16S rRNAs are as different from each other as either is from tobacco; Wolfe et al. 1992b). Sequence data from *Conopholis* are limited to the loss of *trn*C_{GCA}, *trn*I_{GAU}, and *trn*A_{UGC}, and the retention of at least part of *trn*D_{GUC} (Taylor et al. 1991; Wimpee et al. 1992).

The reasons why only a subset of tRNA genes has been lost from *Epifagus* ptDNA are probably complex. We have been unable to identify a simple rule that distinguishes all the retained tRNA genes from those lost. Criteria that fail include: size or sequence of the tRNA; properties of the amino acid; class of aminoacyl-tRNA synthetase in *E. coli;* correlation with codon or amino acid usage patterns; and comparison with tRNA species known to be imported into plant mitochondria, or with the subsets of tRNA genes in mitochondrial genomes. One striking, but likely coincidental, observation is that all six intron-containing tRNA genes are among the 13 lost, even though group II introns are still present in some protein genes in *Epifagus* ptDNA (Wolfe et al., submitted). If our hypothesis of tRNA import is correct, the discriminating factor is more likely to be a combination of properties of the cytoplasmic tRNAs, making some of them both "importable" and compatible with plastid aminoacyl-tRNA synthetases and elongation factors, and so permitting loss of the plastid counterpart.

While the above discussion has concentrated on the possibility that nuclear tRNA genes of eukaryotic origin could substitute for lost plastid genes (i.e., a gene-substitution scenario), gene transfer is also a possible explanation (Morden et al. 1991). According to this hypothesis, some of the regions of ptDNA-derived sequences known to reside in the nuclear genomes of flowering plants (Timmis and Scott 1983) would include plastid tRNA genes that could become functional and underwrite loss of the homolog in ptDNA. The imported tRNAs would then be fully compatible with the plastid translational machinery, except for possible compartment-specific base modifications.

Codon Usage, Amino Acid Usage, and Base Composition

Although codon usage and tRNA populations are correlated in chloroplasts (Pfitzinger et al. 1987), codon usage in land-plant plastid genomes appears to reflect mutational biases rather than natural selection. For example, there is little difference in codon usage patterns between photosynthetic and

ribosomal protein genes in tobacco (Ohto et al. 1988), and base composition at silent codon positions in *Marchantia* ptDNA parallels the extreme A + T-richness of noncoding regions of that genome (Ozeki et al. 1987). The loss of tRNA genes in *Epifagus* ptDNA has had little, if any, impact on its codon usage patterns as compared to tobacco (Fig. 6). This suggests that tRNA availability is not a limiting factor on the expression of *Epifagus* plastid genes, again implying import of cytoplasmic tRNAs.

Codon choice in *Epifagus* and tobacco can be compared directly (i.e., independently of amino acid choice) for the four amino acids for which *Epifagus* has retained one or more tRNA isoacceptor genes, but has lost another (Fig. 6). Of these, *Epifagus* "correctly" avoids use of the Ile codon pair ATT/ATC in favor of ATA ($p < 0.01$ by a chi-square test), but significantly ($p < 0.01$) *over*uses the Leu codon TTA for which it encodes no tRNA. Codon usage for Ser and Arg is not significantly different between *Epifagus* and tobacco.

Analysis of amino acid usage (data from Fig. 6) shows some significant ($p < 0.01$) differences between *Epifagus* and tobacco but these do not appear related to the tRNA gene content differences. *Epifagus* overuses Lys (0/1, AAR), Ile (1/2, ATH), Tyr (1/1, TAY), and Asn (1/1, AAY), and underuses Arg (1/2, CGN and AGR), Gln (1/1, CAR), and Asp (1/1, GAY). [The data in parentheses for each amino acid are the number of tRNA genes in *Epifagus* as a fraction of those in tobacco, and the codon sequences (N = any base; Y = T or C; R = A or G; H = T, C or A).]

Rather than reflecting tRNA gene contents, the changes in amino acid and codon composition appear due to a slight decrease in G + C content over the whole *Epifagus* plastid genome as compared to tobacco (Table 1). The overall G + C content of *Epifagus* ptDNA (counting the inverted repeat only once) is 34%, as against 37% in tobacco. The reduction is more pronounced in the single-copy regions than in the IR, as expected given the latter's lower rate of nucleotide substitution (Wolfe et al. 1987). These differences are also seen at silent codon positions (Table 1). Since the base composition data for rice and tobacco are more similar to each other than to *Epifagus* (Table 1), drift in base composition seems to have occurred in the *Epifagus* plastid genome in concert with the accelerated rates of nucleotide substitution and extensive DNA deletion. It is difficult to judge whether this drift is a neutral phenomenon or is another facet of the release of selective constraints on the plastid genome; we note that extensive divergence of base composition has occurred between tobacco and *Marchantia* ptDNAs [37% and 28% G + C, respectively; Shinozaki et al. (1986) and Ohyama et al. (1986)] even under the constraint of photosynthetic function.

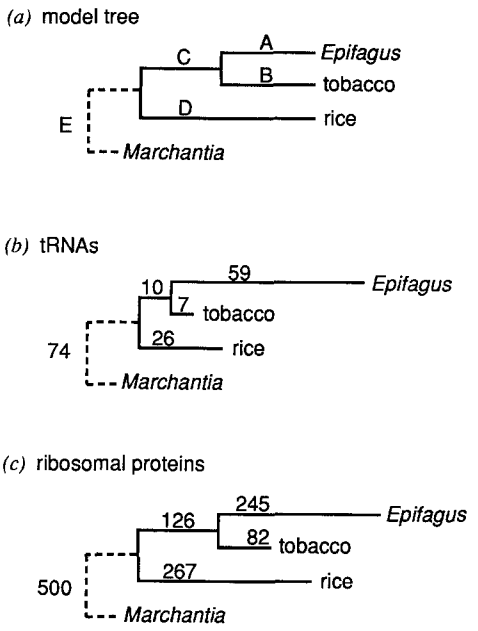

Fig. 7. (a) Model phylogenetic tree for *Epifagus*, tobacco, rice, and *Marchantia*. The phylogenetic relationships among these species (two dicots, one monocot, and one bryophyte) are completely noncontroversial. A–E refer to branch lengths. E is the length of the branch leading to the outgroup (*Marchantia*) and is not drawn to scale. The root of the tree lies on this branch but its position cannot be determined. (b) Tree obtained from the combined sequences of the 17 tRNA genes intact in *Epifagus* (Fig. 5). (c) Tree obtained from the combined sequences of the 15 ribosomal protein genes (Fig. 2). Sequence data for tobacco, rice, and *Marchantia* are from Shinozaki et al. (1986), Hiratsuka et al. (1989), and Ohyama et al. (1986), respectively.

Accelerated Evolution of the Translational Apparatus

The components of the translational apparatus encoded by the *Epifagus* plastid genome have high rates of sequence evolution. This is apparent from phylogenetic analyses in which the numbers of amino acid or nucleotide substitutions assigned to the *Epifagus* lineage can be compared to the numbers for tobacco, the closest relative to *Epifagus* for which complete sequence data are available. We have reported accelerated evolution of the *Epifagus* rRNA genes (Wolfe et al. 1992b), for which the ratio of *Epifagus* to tobacco branch lengths in a phylogenetic tree is 6.2 when the four rRNA genes are pooled together. Likewise, for the 17 intact tRNA genes pooled together, this ratio is 8.4 (Fig. 7b).

Phylogenetic trees drawn from amino acid sequences of the 15 ribosomal proteins also consistently point to rapid evolution in the *Epifagus* lineage. These results are given in Fig. 2, where the numbers A, B, C, D, and E for each protein refer to

the numbers of amino acid replacements assigned by parsimony analysis to each branch in the model tree shown in Fig. 7a. For individual ribosomal proteins, the ratio of the *Epifagus*-to-tobacco-branch length varies between 1.5 and 9 (excluding proteins for which the tobacco branch length, B, is zero). Although many of the ribosomal proteins are small and the substitution rate differences would not be judged significant by a statistical test, the fact that the *Epifagus* branch length is longer (i.e., A > B) for every one of the 15 comparisons strongly suggests that the results are not due to chance. For all 15 sequences pooled together (Fig. 7c) the branch length difference between *Epifagus* and tobacco is 3.0-fold. Inclusion of the available sequences from spinach (ten ribosomal protein genes) or legumes (nine genes) leads to similar branch length differences between *Epifagus* and tobacco (5.4- and 4.3-fold, respectively; data not shown).

Evidence that the branch length differences seen in these analyses are due to increased rates of evolution in *Epifagus* rather than to a slowdown in the tobacco lineage comes from relative rate tests using the sequences of the gene *rbc*L (R.G. Olmstead and J.D. Palmer, unpublished data) from ptDNA of photosynthetic species. *Antirrhinum majus* and *Digitalis purpurea* (Scrophulariaceae) are more closely related to *Epifagus virginiana* (Orobanchaceae) than to tobacco (Solanaceae). Using rice as an outgroup, the numbers of synonymous and nonsynonymous substitutions in *rbc*L in the tobacco lineage are very similar to those in the Scrophulariaceae (data not shown).

The ratio for *Epifagus*-to-tobacco branch lengths for each set of genes represents the outcome of the balance between mutation and selective constraints on function in each gene product in both the *Epifagus* and tobacco lineages. Analysis of synonymous and nonsynonymous nucleotide substitutions in the ribosomal protein genes indicates that the accelerated evolution in *Epifagus* is due to both an increase in mutation rate and a decrease in selective constraint. For the 15 ribosomal protein genes considered together, with rice as an outgroup, the differences between *Epifagus* and tobacco are 2.1-fold at synonymous sites and 4.5-fold at nonsynonymous sites. The difference in synonymous substitution rates probably reflects mutation rate differences. The additional rate increases seen in amino acid sequence comparisons and in tRNAs and rRNAs likely indicate a relaxation of selective constraints on translational gene products. The genetic apparatus of *Epifagus* plastids is required to produce, at most, only four nongenetic proteins (Wolfe et al., submitted) as compared to up to 52 nongenetic proteins in tobacco. Furthermore, these four proteins are probably much less abundant than photosynthetic proteins, and plastid ribosomal protein genes and unassigned ORFs are also generally less well conserved in sequence than bioenergetic genes. Relaxation of translation in terms of both efficiency and fidelity may therefore be tolerable in *Epifagus* plastids. Runaway evolution of the translational apparatus has also occurred in animal mitochondria (Brown 1983; Cann et al. 1984). This may result both from factors shared with *Epifagus* (i.e., relaxed constraints on the translational apparatus in a simple genetic system) as well ones unique to animal mitochondria (i.e., the genome's extraordinarily high mutation rate).

Conclusion

The absence of some tRNA and ribosomal protein genes from the ptDNA of *Epifagus* suggests that nuclear gene products play a greater role in the plastid translational apparatus of the parasite than of photosynthetic plants. In the case of both the tRNAs and the ribosomal proteins, relocation of function could have been achieved either by gene transfer (direct movement of a gene from ptDNA to nuclear DNA) or by gene substitution (recruitment of a nuclear gene to perform the same function as a lost plastid gene).

Gene transfer has been demonstrated for several plastid genes (Shih et al. 1986; Baldauf and Palmer 1990; Gantt et al. 1991) and is an attractive hypothesis because the necessary changes in the gene product may be quite minor. The product may then have a better chance of interacting properly with the plastid translational machinery, provided that it can be imported across the plastid envelope. The question of compatibility with the plastid translational apparatus may be more important for plastid ribosomal proteins (especially those internal to the ribosome) than for tRNAs (which interact with only a few proteins). A major objection to gene transfer as an explanation for the loss of six ribosomal protein genes from *Epifagus* ptDNA is that either multiple transfers occurred very recently or else these represent older transfers where, for unknown reasons, the plastid genes have not been lost in photosynthetic lineages (cf. Gantt et al. 1991).

Gene substitution, particularly the recruitment of nuclear genes encoding mitochondrial ribosomal proteins to serve the plastid as well (Martin et al. 1990), is perhaps a simpler explanation of the multiple ribosomal protein gene losses from *Epifagus* ptDNA. Angiosperm nuclear genomes encode perhaps 160 ribosomal proteins, directed to three subcellular compartments. The targeting of proteins to cytoplasm, mitochondrion, or plastid, like all biochemical processes, cannot be absolute. There

must be a finite frequency at which, say, a ribosomal protein precursor intended for the mitochondrion ends up in the plastid by mistake (Pfanner et al. 1988). This frequency of misdirection, and the efficiency of ribosomes composed of heterologous components, might even be sufficient to support translation at the low levels presumably required in *Epifagus* plastids, so that no special import mechanism is needed.

Direct evidence on the hypotheses of import of tRNAs and novel ribosomal proteins into *Epifagus* plastids could best be obtained by studies with isolated plastids. This is not as simple as the isolation of chloroplasts from photosynthetic plants because the number of plastids per cell in *Epifagus* tissues is lower than in green leaves, because the plastids are smaller and likely more fragile, and because the aboveground growing season of this obligate parasite is only a few weeks.

The genetic apparatus of plastids is descended from an apparatus that was once responsible for the production of thousands of cyanobacterial proteins. In *Epifagus,* however, that workload has been reduced to the expression of (at most) four protein genes in addition to self-perpetuation. Borst and Grivell (1981) likened the human mitochondrial genome to "a 1981 university department, with everything reduced to minimal size without anyone actually being fired." The *Epifagus* plastid genome might be comparable to a 1992 department where the support staff now outnumber the faculty.

Acknowledgments. We thank John Donello for help with *Epifagus* ptDNA sequencing, Claude dePamphilis for tobacco ptDNA, Dick Olmstead for *rbc*L sequences, Masahiro Sugiura for tobacco clone pTBa1, and Lib Harris for a preprint. This study was supported by grants from the NIH (GM 35087) to J.D.P. and the Alfred P. Sloan Foundation (90-3-5) to K.H.W.

References

Andachi Y, Yamao F, Muto A, Osawa S (1989) Codon recognition patterns as deduced from sequences of the complete set of transfer RNA species in *Mycoplasma capricolum*. Resemblance to mitochondria. J Mol Biol 209:37–54

Baldauf SL, Palmer JD (1990) Evolutionary transfer of the chloroplast *tuf*A gene to the nucleus. Nature 344:262–265

Borst P, Grivell LA (1981) Small is beautiful—portrait of a mitochondrial genome. Nature 290:443–444

Brown WM (1983) Evolution of animal mitochondrial DNA. In: Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer, Sunderland, MA, pp 62–88

Butler JS, Springer M, Grunberg-Manago M (1987) AUU-to-AUG mutation in the initiator codon of the translation initiation factor IF3 abolishes translational autocontrol of its own gene (*infC*) *in vivo*. Proc Natl Acad Sci USA 84:4022–4025

Cann RL, Brown WM, Wilson AC (1984) Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics 106:479–499

dePamphilis CW, Palmer JD (1990) Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature 348:337–339

Downie SR, Palmer JD (1992) Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 14–35

Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD (1991) Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO J 10:3073–3078

Gray MW, Boer PH (1988) Organization and expression of algal (*Chlamydomonas reinhardtii*) mitochondrial DNA. Phil Trans Roy Soc Lond Ser B 319:135–147

Hancock K, Hajduk SL (1990) The mitochondrial tRNAs of *Trypanosoma brucei* are nuclear encoded. J Biol Chem 265: 19208–19215

Harris EH, Gillham NW, Boynton JE (1993) Chloroplast ribosomes: genetics, biogenesis and evolutionary relationships. Microbiol Rev, submitted

Hill WE, Dahlberg A, Garrett RA, Moore PB, Schlessinger D, Warner JR (eds) (1990) The ribosome. Am Soc Microbiol, Washington, DC, pp 1–678

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun C-R, Meng B-Y, Li Y-Q, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete nucleotide sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet 217:185–194

Hoch B, Maier RM, Appel K, Igloi GL, Koessel H (1991) Editing of a chloroplast mRNA by creation of an initiation codon. Nature 353:178–180

Isono K, Isono S (1976) Lack of ribosomal protein S1 in *Bacillus stearothermophilus*. Proc Natl Acad Sci USA 73:767–770

Joyce PBM, Gray MW (1989) Chloroplast-like transfer RNA genes expressed in wheat mitochondria. Nucleic Acids Res 17:5461–5476

Kudla J, Igloi GL, Metzlaff M, Hagemann R, Koessel H (1992) RNA editing in tobacco chloroplasts leads to the formation of a translatable *psb*L mRNA by a C to U substitution within the initiation codon. EMBO J 11:1099–1103

Li W-H, Wu C-I, Luo C-C (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Marechal-Drouard L, Weil J-H, Guillemaut P (1988) Import of several tRNAs from the cytoplasm into the mitochondria in bean *Phaseolus vulgaris*. Nucleic Acids Res 16:4777–4788

Martin RP, Schneller J-M, Stahl AJC, Dirheimer G (1979) Import of nuclear deoxyribonucleic acid coded lysin-accepting transfer ribonucleic acid (anticodon C-U-U) into yeast mitochondria. Biochemistry 18:4600–4605

Martin W, Lagrange T, Li YF, Bisanz-Seyer C, Mache R (1990) Hypothesis for the evolutionary origin of the chloroplast ribosomal protein L21 of spinach. Curr Genet 18:553–556

Michel F, Umesono K, Ozeki H (1989) Comparative and functional anatomy of group II catalytic introns—a review. Gene 82:5–30

Morden CW, Wolfe KH, dePamphilis CW, Palmer JD (1991) Plastid translation and transcription genes in a nonphotosynthetic plant: intact, missing and pseudo genes. EMBO J 10:3281–3288

Nagano Y, Ishikawa H, Matsuno R, Sasaki Y (1991) Nucleotide sequence and expression of the ribosomal protein L2 gene in pea chloroplasts. Plant Mol Biol 17:541–545

Nugent JM, Palmer JD (1991) RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. Cell 66:473–481

Ohto C, Torazawa K, Tanaka M, Shinozaki K, Sugiura M (1988) Transcription of ten ribosomal protein genes from tobacco chloroplasts: a compilation of ribosomal protein genes found in the tobacco chloroplast genome. Plant Mol Biol 11:589–600

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete nucleotide sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572–574

Okimoto R, Wolstenholme DR (1990) A set of tRNAs that lack either the TψC arm or the dihydrouridine arm: towards a minimal tRNA adaptor. EMBO J 9:3405–3411

Osawa S, Ohama T, Yamao F, Muto A, Jukes TH, Ozeki H, Umesono K (1988) Directional mutation pressure and transfer RNA in choice of the third nucleotide of synonymous two-codon sets. Proc Natl Acad Sci USA 85:1124–1128

Ozeki H, Ohyama K, Inokuchi H, Fukuzawa H, Kohchi T, Sano T, Nakahigashi K, Umesono K (1987) Genetic system of chloroplasts. Cold Spring Harbor Symp Quant Biol 52:791–804

Pfanner N, Pfaller R, Nupert W (1988) How finicky is mitochondrial protein import? Trends Biochem Sci 13:165–167

Pfitzinger H, Guillemaut P, Weil J-H, Pillay DTN (1987) Adjustment of the tRNA population to the codon usage in chloroplasts. Nucleic Acids Res 15:1377–1386

Pritchard AE, Seilhamer JE, Mahalingham R, Sable CL, Venuti SE, Cummings DJ (1990) Nucleotide sequence of the mitochondrial genome of *Paramecium*. Nucleic Acids Res 18: 173–180

Schimmel P (1991) RNA minihelices and the decoding of genetic information. FASEB J 5:2180–2187

Schulman LH, Pelka H (1988) Anticodon switching changes the identity of methionine and valine transfer RNAs. Science 242:765–768

Schuster W, Wissinger B, Hiesel R, Unseld M, Gerold E, Knoop V, Marchfelder A, Binder S, Schobel W, Scheike R, Gronger P, Ternes R, Brennicke A (1991) Between DNA and protein—RNA editing in plant mitochondria. Physiol Plant 81: 437–445

Shih M-C, Lazar G, Goodman HM (1986) Evidence in favor of the symbiotic origin of chloroplasts: primary structure and evolution of tobacco glyceraldehyde-3-phosphate dehydrogenases. Cell 47:73–80

Shinozaki K, Ohme M, Tanaka T, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5: 2043–2049

Sijben-Mueller G, Hallick RB, Alt J, Westhoff P, Herrmann RG (1986) Spinach plastid genes coding for initiation factor IF-1, ribosomal protein S11 and RNA polymerase alpha-subunit. Nucleic Acids Res 14:1029–1044

Simpson AM, Suyama Y, Dewes H, Campbell DA, Simpson L (1989) Kinetoplastid mitochondria contain functional tRNAs which are encoded in nuclear DNA and also contain small minicircle and maxicircle transcripts of unknown function. Nucleic Acids Res 17:5427–5445

Sprinzl M, Hartmann T, Weber J, Blank J, Zeidler R (1989) Compilation of tRNA sequences and sequences of tRNA genes. Nucl Acids Res 17 (suppl):r1–r172

Steinmetz A, Weil J-H (1986) Isolation and characterization of chloroplast and cytoplasmic transfer RNAs. Methods Enzymol 118:212–231

Subramanian AR, Stahl D, Prombona A (1991) Ribosomal proteins, ribosomes, and translation in plastids. In: Bogorad L, Vasil IK (eds) Molecular biology of plastids, vol 7A of Vasil IK (ed-in-chief), Cell culture and somatic cell genetics of plants. Academic Press, San Diego, pp 191–215

Sugiura M (1987) Structure and function of the tobacco chloroplast genome. Bot Mag Tokyo 100:407–436

Sugiura M, Shinozaki K, Zaita N, Kusuda M, Kumano M (1986) Clone bank of the tobacco (*Nicotiana tabacum*) chloroplast genome as a set of overlapping restriction endonuclease fragments: mapping of eleven ribosomal protein genes. Plant Sci 44:211–216

Surguchov AP (1987) Common genes for mitochondrial and cytoplasmic proteins. Trends Biochem Sci 12:335–338

Taylor GW, Wolfe KH, Morden CW, dePamphilis CW, Palmer JD (1991) Lack of a functional plastid $tRNA^{Cys}$ gene is associated with loss of photosynthesis in a lineage of parasitic plants. Curr Genet 20:515–518

Timmis JN, Scott NS (1983) Sequence homology between spinach nuclear and chloroplast genomes. Nature 305:65–67

Tingey SV, Tsai FY, Edwards JW, Walker EL, Coruzzi GM (1988) Chloroplast and cytoplasmic glutamine synthetase are encoded by homologous nuclear genes which are differentially expressed *in vivo*. J Biol Chem 263:9651–9657

Toukifimpa R, Romby P, Rozier C, Ehresmann C, Ehresmann B, Mache R (1989) Characterization and footprint analysis of two 5S rRNA binding proteins from spinach chloroplast ribosomes. Biochemistry 28:5840–5846

Wakasugi T, Ohme M, Shinozaki K, Sugiura M (1986) Structures of tobacco chloroplast genes for $tRNA^{Ile}$ (CAU), $tRNA^{Leu}$ (CAA), $tRNA^{Cys}$ (GCA), $tRNA^{Ser}$ (UGA) and $tRNA^{Thr}$ (GGU): a compilation of tRNA genes from tobacco chloroplasts. Plant Mol Biol 7:385–392

Walsh MA, Rechel EA, Popovich TM (1980) Observations on plastid fine-structure in the holoparasitic angiosperm *Epifagus virginiana*. Am J Bot 67:833–837

Wegloehner W, Subramanian AR (1991) A heptapeptide repeat contributes to the unusual length variation of chloroplast ribosomal protein S18. FEBS Lett 279:193–197

Wimpee CF, Morgan R, Wrobel R (1992) An aberrant plastid ribosomal RNA gene cluster in the root parasite *Conopholis americana*. Plant Mol Biol 18:275–285

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058

Wolfe KH, Morden CW, Palmer JD (1992a) Small single-copy region of plastid DNA in the non-photosynthetic angiosperm *Epifagus virginiana* contains only two genes: differences among dicots, monocots and bryophytes in gene organization at a non-bioenergetic locus. J Mol Biol 223:95–104

Wolfe KH, Katz-Downie DS, Morden CW, Palmer JD (1992b) Evolution of the plastid ribosomal RNA operon in a nongreen parasitic plant: accelerated sequence evolution, altered promoter structure, and tRNA pseudogenes. Plant Mol Biol 18:1037–1048

Yokoi F, Vassileva A, Hayashida N, Torazawa N, Wakasugi T, Sugiura M (1990) Chloroplast ribosomal protein L32 is encoded in the chloroplast genome. FEBS Lett 276:88–90

Yokoi F, Tanaka M, Wakasugi T, Sugiura M (1991) The chloroplast gene for ribosomal protein CL23 is functional in tobacco. FEBS Lett 281:64–66

Ziaie Z, Suyama Y (1987) The cytochrome oxidase subunit I gene of *Tetrahymena:* a 57 amino acid NH_2-terminal extension and a 108 amino acid insert. Curr Genet 12:357–368

Zurawski G, Clegg MT (1987) Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure-function and phylogenetic studies. Ann Rev Plant Physiol 38:391–418

Received January 6, 1992/Revised April 21, 1992