Curr Genet (1994) 26: 256–262

Current Genetics

Genes for components of the chloroplast translational apparatus are conserved in the reduced 73-kb plastid DNA of the nonphotosynthetic euglenoid flagellate *Astasia longa*

Gabriele Gockel, Wolfgang Hachtel, Susanne Baier, Christian Fliss, Mark Henke

Botanisches Institut, Universität Bonn, Kirschallee 1, D-53115 Bonn, Germany

Received: 12 January / Accepted: 9 March 1994

Abstract. The colourless, nonphotosynthetic protist *Astasia longa* is phylogenetically related to *Euglena gracilis*. The 73-kb plastid DNA (ptDNA) of *A. longa* is about half the size of most chloroplast DNAs (cpDNAs). More than 38 kb of the *Astasia* ptDNA sequence has been determined. No genes for photosynthetic function have been found except for *rbcL*. Identified genes include *rpoB*, *tufA*, and genes coding for three rRNAs, 17 tRNAs, and 13 ribosomal proteins. Not only is the nucleotide sequence of these genes highly conserved between *A. longa* and *E. gracilis*, but a number of these genes are clustered in a similar fashion and have introns in the same positions in both species. The results further support the idea that photosynthetic genes normally encoded in cpDNA have been preferentially lost in *Astasia*, but that the chloroplast genes coding for components of the plastid translational apparatus have been maintained. This apparatus might be needed for the expression of *rbcL* and also for that of still unidentified nonphotosynthetic genes of *Astasia* ptDNA.

Key words: *Astasia longa* – Plastid DNA – Ribosomal protein genes – tRNA genes

Introduction

Astasia longa is a colourless, nonphotosynthetic flagellate protist that is phylogenetically releated to the photoautotrophic *Euglena gracilis* (Pringsheim 1942). The plastid genome of *A. longa* resembles the chloroplast genome of *E. gracilis* but has lost most photosynthetic genes and is only half the size (73 kb instead of 143 kb; Siemeister and Hachtel 1989). A similar loss of photosynthetic genes has occurred from the ptDNA of a nonphotosynthetic parasitic flowering plant, *Epifagus virginiana* (Wolfe et al. 1992).

Many intact translational genes have been identified in the 26 kb of *Astasia* plastid DNA (ptDNA) that has been sequenced so far, without any evidence for pseudogenes. The *Astasia* ptDNA is an active genome since a number of transcripts of protein-encoding genes have been detected (Siemeister and Hachtel 1990 a, b; Siemeister et al. 1990 a, b). Moreover, the protein encoded by the CO_2-fixation gene *rbcL* occurs in *A. longa* (Siemeister and Hachtel 1990 a).

Unique features shared by *Euglena* chloroplast (cp) DNA (Hallick et al. 1993) and *Astasia* ptDNA include a tandem array of three complete, and one partial, ribosomal RNA operons (Siemeister and Hachtel 1989, 1990 b), a gene for the elongation factor EF-Tu (Siemeister et al. 1990 a), a class of very small introns designated group III (Siemeister et al. 1990 b), and the absence of introns in tRNA-encoding genes (Siemeister et al. 1990 a). *Astasia* has a gene cluster with the gene order *rpl5-rps8-rpl36-trnI-rps14-trnF-trnC-rps2* (Siemeister et al. 1990 b). Not only does this same gene cluster occur in *Euglena* (Hallick et al. 1993), but three group-II and five group-III introns occur in the same positions in the same genes in both *Euglena* and *Astasia*. Other gene combinations found in both organisms are *tufA-rps7*, and *rbcL-rpl32*. *Astasia rbcL* (Siemeister and Hachtel 1990 a) has seven of the nine group-II introns in the same positions as *Euglena rbcL* (Gingrich and Hallick 1985). *Astasia rpoB* also has at least seven group-III introns (EMBL Acc. No. X75651) but their positions differ from *Euglena rpoB*. *Euglena* has a locus designated *ycf13* for a protein of 458 amino acids (Montandon et al. 1986), absent in land plants but also found in the ptDNA of *Astasia* (Siemeister et al. 1990 a). Probably absent from *Astasia* are genes for subunits of a NADH dehydrogenase complex, present in land plants but not detected in *Euglena* (Hallick et al. 1993).

In this paper we report on further genes on the ptDNA of *A. longa* that code for components of a plastid translational apparatus. The results corroborate our previous conclusion that the *Astasia* plastid genome has evolved from a *Euglena* chloroplast genome by highly specific deletions

These sequence data will appear in the EMBL/Gen Bank/DDBJ nucleotide sequence data base under accession numbers X75651, X75652 and X75653

Correspondence to: W. Hachtel

and sequence rearrangements. We hypothesize that the *Astasia* plastid genome has remained active after the loss of photosynthesis because one (or a few) of its protein genes is (are) involved in a nonphotosynthetic process which is either indispensable to, or at least of advantage for, *A. longa*.

Materials and methods

Isolation of DNA and RNA from cells of *A. longa* harvested in the late logarithmic phase of growth has been described previously (Siemeister and Hachtel 1989; Siemeister et al. 1990 a). Cloning of DNA restriction fragments, gel electrophoresis, and blotting of glyoxylated RNA followed standard procedures (Sambrook et al. 1989). Northern-blot analysis was performed as described (Siemeister et al. 1990 a). The nucleotide sequence was determined by the dideoxy chain-termination method (Sanger et al. 1977; Chen and Seeburg 1985) using T7-DNA-Polymerase (Tabor and Richardson 1987). Sequences were determined in both directions. Analysis of sequence data was performed using the Amersham Staden plus software package and the FASTP program (Lipman and Pearson 1985). Gene identification was based on screening of the EMBL database, Heidelberg, Germany. Most genes were identified by comparison with the coding sequences of *Euglena* cpDNA (EMBL Acc. No. X70810) due to the high degree of nucleotide and amino-acid sequence identity that is observed between homologous genes of *Astasia* and *Euglena* (see Table 1).

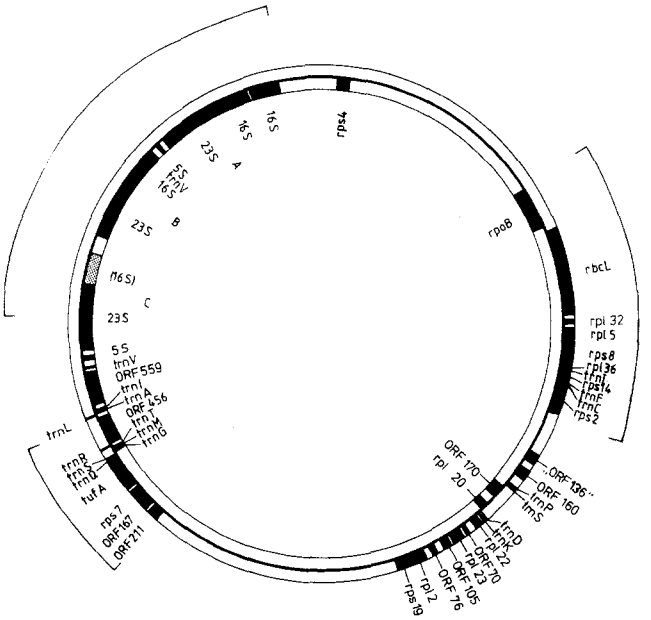

Fig. 1. Gene map of the *A. longa* 73-kb plastid DNA. Genes are represented by *filled boxes* which are proportional to gene length, including exons and introns. A truncated copy of the 16*s* rRNA gene is *shadowed.* Genes on the outer circle are transcribed clockwise. Genes on the inner circle are transcribed counterclockwise. *Brackets* indicate those parts of the sequence that have been published before (Siemeister and Hachtel 1989, 1990 a, b; Siemeister et al. 1990 a, b)

Results and discussion

The following segments of the circular ptDNA of *A. longa* were cloned and sequenced: a 4.0-kb *Xba*I fragment (X6), a 3.9-kb *Xba*I fragment (X7), a 2.9-kb *Xba*I fragment (X11), a 1.5-kb *Bgl*II fragment (B9), and a 1.9-kb *Hin*dIII fragment (H14). (For the location of these fragments see the restriction-site map presented by Siemeister and Hachtel 1989). Analysis of these sequence data identified a number of densely-packed genes. Among these are seven tRNA genes (*trnI*, *trnA*, *trnL*, *trnP*, *trnS*, *trnD*, and *trnK*), genes for six ribosomal proteins (*rps4*, *rps19*, *rpl2*, *rpl20*, *rpl22*, and *rpl23*), and several open reading frames (ORFs) encoding proteins of unknown function. An updated gene map of *Astasia* ptDNA is shown in Fig. 1, and all genes detected so far are listed in Table 1. The degree of nucleotide and amino-acid sequence identity of these genes with homologous genes of *Euglena*, the transcripts detected in *Astasia*, and the number and classification of introns, are also indicated in Table 1. Data and annotations are reported in EMBL Accession Numbers X75651, X75652 and X75653.

Genes for ribosomal RNAs

Three repeats (A, B, C) of 16*s* and 23*s* rDNA arranged in tandem, and one supplementary 16*s* rDNA adjacent to the 16*s* rDNA of repeat A, are present within an 18-kb segment of the 73-kb ptDNA of *A. longa* (Fig. 1). Repeat C contains a truncated copy of 16*s* rDNA (Siemeister and Hachtel 1989). The repeats A and B are separated by a short region containing a gene for 5*s* rRNA and a tRNA-Val (UAC) gene (Siemeister and Hachtel 1990 b). We have also now identified copies of 5*s* rRNA and the tRNA-Val (UAC) gene adjacent to the 23*s* rRNA gene of repeat C. Thus, the gene order (*5′-16s-23s-5s-trnV-3′*) of repeats A and C is identical. The 5*s* rDNAs of repeats A and C differ at three nucleotide positions, and the *trnV* of repeats A and C differ at two positions. These differences do not affect the secondary structures of the deduced RNAs.

To determine the pattern of transcripts, Northern analysis was performed using organellar RNA and a [^{32}P]-labelled *Dra*I fragment of the 3′-region of the 23*s* rDNA of repeat C. Transcripts of 7.5 kb, 5.5 kb, 3.0 kb, 2.3 kb, and 1.6 kb were detected (Fig. 2). From the sequence data (Siemeister and Hachtel 1990 b), processed 23*s* rRNA is expected to be about 3.15 kb, 3.1 kb, and 3.0 kb in size (size differences occur due to the observed length polymorphisms of 23*s* rRNA encoded in repeats A, B, and C, respectively). Therefore, the major radioactive band detected at about 3.1 kb probably represents a mixture of processed 23*s* rRNAs that were not resolved on the gel. The larger transcripts (7.5 kb and 5.2 kb) also hybridized to gene probes obtained from 16*s* rDNA (data not shown). Therefore, the 7.5-kb RNA might have originated from co-transcription of the supplementary 16*s* rDNA and repeat A (a 7.25-kb DNA segment) whereas the 5.5-kb RNA probably represents transcripts of single rDNA repeats of size 5.5 kb. The *Euglena* rRNA gene operon was shown to be transcribed as a 6.0-kb RNA which contains both the 16*s* and 23*s* rRNA sequences (Dix and Rawson 1983). The smaller RNA molecules seen in Fig. 2 might be fragments of 23*s* rRNA due to hidden breaks (Kössel et al. 1985).

Table 1. Identified genes and transcripts of the 73-kb plastid DNA of *A. longa*

Genes	Homology with *E. gracilis* cpDNA [a]			Transcript(s) detected	Intron(s) [b]
	Nucleotide sequence:	Amino-acid sequence:			
	Identical nucleotides (%)	Identical amino acids (%)	Identical amino acids plus conservative replacements (%)		
Ribosomal RNA genes					
16*s* (3 copies)	81			+	
23*s* (3 copies)	78			+	
5*s* (2 copies)	68				
Transfer RNA genes					
trnA(UGC)	82				
trnC(GCA)	90				
trnD(GUC)	81				
trnF(GAA)	89				
trnG(GCC)	81				
trnI(CAU)	94				
trnI(GAU)	89				
trnK(UUU)	90				
trnL(CAA)	78				
trnM(CAU)	74				
trnP(UGG)	89				
trnQ(UUG)	83				
trnR(UCU)	88				
trnS(GCU)	81				
trnS(UGA)	84				
trnT(UGU)	90				
trnV(UAC) (2 copies)	68				
Ribosomal protein and translation factor genes					
rps2		41	80	+	1(II) 3(III)
rps4		58	83	+	
rps7		50	88		
rps8		50	85	+	2(II) 1(III)
rps14		59	83	+	1(III)
rps19		52	85		2(III)
rpl2		68	93		
rpl5		61	90	+	
rpl20		44	82		
rpl22		42	77		
rpl23		49	83		2(III)
rpl32		60	90	+	
rpl36		62	89	+	
tufA		86	99	+	2(III)
RNA polymerase genes					
rpoB		47	83		≥7(III)
Photosynthetic genes					
rbcL		82	97	+	7(II)
Other putative protein genes					
ycf13 (=orf456)		56	84	+	
ycf14 (=orf170)		42	84		
orf559		No significant		+	
orf57, orf70, orf76, orf105		similarity with			
orf160, orf162, orf167, orf211		any plastid gene			

[a] Information provided in EMBL Accession No. X70810 (Hallick et al. 1993)
[b] For the classification of introns (group II, group III) see Christopher et al. (1988) and Christopher and Hallick (1989)

Genes for transfer RNAs

Genes encoding tRNA-Ile (GAU) and tRNA-Ala (UGC) are located in the 16*s* rDNA-23*s* rDNA spacer of the cpDNA of *Euglena* and every other cpDNA investigated so far (Kössel et al. 1985), as well as in *E. coli*, but not are present at this location in the ptDNA of *Astasia* (Siemeister and Hachtel 1990b). Both *trnI* (GAU) and *trnA* (UGC), together with *trnL* encoding tRNA-Leu (CAA), were found by sequencing the *Xba*I-fragment X7. In the very neighbourhood of the rDNA tandem repeats, these genes are clustered in a 550-bp segment between *orf559* and *ycf13* (=*orf456*); *trnL* is located on the complementary strand (Figs. 1, 3). Sequence similarity between these tRNA genes and *trnI* (Graf et al. 1980), *trnA* (Orozco et al. 1980), and *trnL* (Monfort et al. 1986), respectively, of

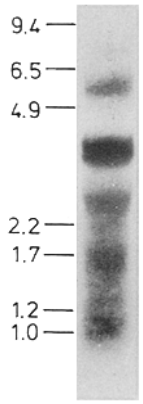

Fig. 2. Analysis of transcripts of the 23*s* rRNA genes of *A. longa* ptDNA by Northern hybridization. A 650-bp *Dra*I fragment of 23*s* rDNA was [^{32}P]-labelled and hybridized against organellar RNA blotted onto nitrocellulose filters. The position and size (kb) of single-strand marker DNA (*Bgl*I and *Hin*fI fragments of pBR328 and *Hin*dIII fragments of lambda DNA) are indicated

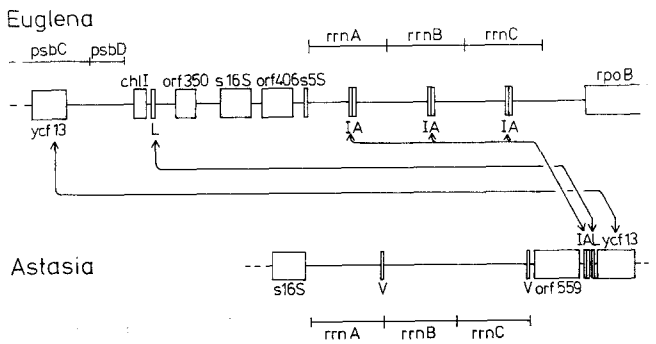

Fig. 3. Comparison of the location of the tRNA genes *trnI* (GAU), *trnA* (UGC) and *trnL* (CAA), and of *ycf13*, on the ptDNA of *A. longa* and the cpDNA of *E. gracilis*. Homologous genes are connected by *arrows*. *ycf13* is encoded in an intron of *psbC* in *Euglena*. The direction of transcription is from left to right for *Astasia* and from *chlI* for *Euglena*; *psbD*, *psbC* and *ycf13* (=*orf458*) are transcribed in the opposite direction. The maps are not true to scale (Data for *Euglena* are from Hallick et al. 1993)

Euglena, is given in Table 1. Invariant, or semi-invariant, nucleotides (see Sprinzl et al. 1989) were found in all conserved positions of the deduced tRNA-Ile (GAU) and tRNA-Leu (CAA) and in 19 out of 20 conserved positions of tRNA-Ala (UGC). In tRNA-Ala, cytidine is found instead of uridine (mostly modified to pseudouridine) at nucleotide position 55. Whether this is real or a cloning artefact needs to be clarified by sequencing a primary restriction fragment encoding this tRNA which is different from X7. The A-A mismatch (nucleotide positions 13 and 23) observed in the stem structure of the D-arm of tRNA-Leu (CAA) also occurs in the tRNA-Leu (CAA) of *E. gracilis* chloroplasts (Monfort et al. 1986).

The gene order observed in *Astasia* as compared to that in *Euglena* (Fig. 3) is putatively the result of a complicated series of deletions and sequence rearrangements. Some of these events can be tentatively specified. (1) The tRNA-Leu (CAA) gene is located about 11 kb upstream the *rrnA* operon in *Euglena* (Hallick et al. 1993). ORF *ycf13* (=*orf458*), absent in land plants but also found in *Astasia* (*orf456*), is encoded within a group-III twintron internal to the *psbC* (photosystem II CP43 chlorophyll apoprotein) gene. Assuming deletion of *chlI* (chlorophyll biosynthesis), *psbD* (photosystem II core 34-kDa protein) and *psbC* would explain the neighbourhood of *trnL* and *ycf13* in the ptDNA of *Astasia*. (2) The presence of supplementary 16*s* rRNA and 5*s* rRNA genes in *Euglena* (see Hallick and Buetow 1989) and a supplementary 16*s* rRNA gene in *Astasia* (Siemeister and Hachtel 1990b), in addition to the three rDNA repeats A–C, suggests that this is an evolutionary relic of a fourth rRNA operon (Roux and Stutz 1985). It is also reasonable to assume the presence of *trnI* and *trnA* in the spacer between the 16*s* rDNA and 23*s* rDNA of an ancestral cpDNA from which *Astasia* ptDNA has evolved. In *Astasia*, the extra 16*s* rDNA is separated from the 16*s* rDNA of the rDNA repeat A by a spacer of only 220 bp. Of these, a 74-bp segment flanking the 3′-end of the extra 16*s* rDNA is almost identical with a 77-bp sequence downstream from the 3′-end of the 16*s* rDNA of rRNA operon B, and a 160-bp segment upstream of the 16*s* rDNA of repeat A shows considerable sequence similarity to the 167-bp region upstream of the 16*s* rDNA of repeats B and C (Siemeister and Hachtel 1990b). These data might indicate that a segment of about 3.8 kb between the fourth 16*s* rRNA gene and the rRNA operon A of an ancestral DNA has been rearranged, and the flanking regions have been fused. By further events, most of this 3.8-kb segment might have been deleted except for the gene pair *trnI* and *trnA* that was inserted at its present-day position, whereas the *trnI* and *trnA* genes in the 16*s* rDNA–23*s* rDNA spacer of the ancestral rRNA operons A–C were deleted.

Further tRNA genes were detected on the *Xba*I-fragment X6: *trnP*, *trnS*, *trnD*, and *trnK* encoding tRNA-Pro (UGG), tRNA-Ser (UGA), tRNA-Asp (GUC), and tRNA-Lys (UUU), respectively. The degree of sequence similarity between these tRNA genes and *trnP* and *trnS* (Manzara and Hallick 1988) and *trnD* and *trnK* (Manzara et al. 1987) or *Euglena* is given in Table 1. In the deduced sequence of tRNA-Pro (UGG) and tRNA-Ser (UGA), all invariant and semi-invariant nucleotides (see Sprinzl et al. 1989) are conserved. The number of base pairs of the D-stem of tRNA-Ser is reduced (only two instead of four) in both *Astasia* and *Euglena* (Manzara and Hallick 1988) as compared with tobacco (Shinozaki et al. 1986), *Marchantia* (Ohyama et al. 1986), and *E. coli* (Kröger et al. 1992) tRNA-Ser. Deduced sequences of tRNA-Asp (GUC) and tRNA-Lys (UUU) show invariant or semi-invariant nucleotides at all highly-conserved positions (see Sprinzl et al. 1989). In tRNA-Lys (UUU), base pairing of the anticodon stem is incomplete as was found for tobacco (Shinozaki et al. 1986) and *Marchantia* (Ohyama et al. 1986) but not *Euglena* (Manzara et al. 1987) and *E. coli* (Yoshimura et al. 1984).

trnP and *trnS* are separated by a very short spacer (7 bp) in *Astasia*, as they are in *Euglena* (10-bp spacer), and are transcribed in the same direction as in *Euglena* (Manzara and Hallick 1988). *trnD* and *trnK* are separated by a short non-coding sequence (18 bp) in *Astasia* in contrast to the situation in *Euglena* where *petG* (encoding subunit V of the cytochrome b6/f complex) is located between *trnD* and *trnK* (Manzara et al. 1987). Further differences between *Astasia* and *Euglena* concern genes upstream of *trnD* and downstream from *trnK*. *psbI*, the gene encoding photosystem II–polypeptide I, and (on the opposite

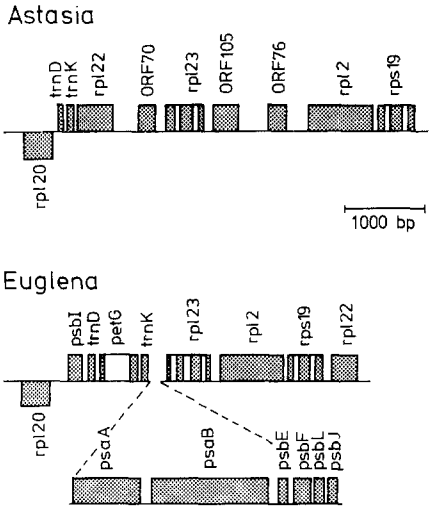

Fig. 4. Comparison of the location of *trnD* (GUC), *trnK* (UUU), *rpl22*, *rpl23*, *rpl2*, and *rps19* on the ptDNA of *A. longa* and the cpDNA of *E. gracilis*. Genes on the upper strand are transcribed from left to right, genes on the lower strand from right to left. The *Euglena* genes *psaA* to *psbJ* are not to scale. (Data for *Euglena* are from Hallick et al. 1993)

strand) *rpl20* are located upstream of the *Euglena trnD*. Downstream from *Euglena trnK*, a series of photosynthetic genes have been identified: *psaA* and *psaB*, encoding the P700 apoproteins A1 and A2, respectively, of photosystem I; *psbE* and *psbF*, encoding the cytochrome b559 α and β subunits, respectively; *psbL* and *psbJ*, encoding the photosystem II proteins L and J, respectively; and, finally, a ribosomal protein gene cluster consisting of *rpl23*, *rpl2*, *rps19*, and *rpl22*. Thus, the gene order in *Euglena* is *rpl20-psbI-trnD-petG-trnK-psaA-psaB-psbE-psbF-psbL-psbI-rpl23-rpl2-rps19-rpl22* (Manzara et al. 1987; Hallick et al. 1993), whereas in *Astasia* the gene order *rpl20-trnD-trnK-rpl22-rpl23-rpl2-rps19* was found (Figs. 1, 4). Thus, all photosynthetic genes present in this stretch of the cpDNA of *Euglena* appear to be specifically deleted in the ptDNA of *Astasia*.

Genes for ribosomal proteins

The localization of seven genes encoding proteins of the small and the large subunit of plastid ribosomes (*rps2*, *rps7*, *rps8*, *rps14*, *rpl5*, *rpl32*, *rpl36*) has been reported (Siemeister et al. 1990 a, b). Additional ribosomal protein genes were identified by sequencing *Xba*I-fragments X6 and X11 and the *Bgl*II-fragment B9. *Astasia* has a gene cluster with the gene order *rpl22-(orf70-)rpl23-(orf105-orf76-)rpl2-rps19* (Figs. 1, 4). In land plants (Fukuzawa et al. 1988; Sugiura 1992), *Euglena* (Christopher et al. 1988), and *E. coli* (Zurawski and Zurawski 1985), a similar cluster with the gene order *rpl23-rpl2-rps19-rpl22* was found (Fig. 4). However, not only is *rpl22* rearranged in *Astasia* but the transcription direction and the position of this gene cluster relative to the *rbcL* gene has also changed in *Astasia* as compared to chloroplast genomes.

Ribosomal protein genes *rps19* and *rpl23* are split in both *Euglena* (Christopher et al. 1988) and *Astasia* (Table 1). Two group-III introns occur in the same positions in the *rps19* gene of both. Amino-acid identity between the *Astasia* and *Euglena rps19* gene product is 52% in a 93 amino-acid overlap. The *Astasia rps19* homologue, however, encodes a 117 amino-acid polypeptide whereas the ribosomal protein S19 of *Euglena* is composed of 93, and that of *Marchantia*, tobacco and *E. coli* of 91, amino acids. This difference is due to a point mutation at nucleotide position 43 in the third exon of the *Astasia rps19* leading to a reading-frame shift. The gene products of *Astasia* and *Euglena rpl23* share 49% identical amino acids. Of the three introns (group III) of the *Euglena rpl23*, introns 2 and 3 were found in *Astasia* at identical positions whereas intron 1 is absent in *Astasia*. Thus, exon 1 and exon 2 of *Euglena rpl23* appear to be fused in *Astasia*. Fusion of exons in *Astasia* ptDNA as compared with *Euglena* cpDNA has also been observed for the *rbcL* gene (Siemeister and Hachtel 1990 a).

An ORF encoding a polypeptide of 117 amino acids rich in lysine residues (24%) was tentatively identified as *rpl20*. The GC content of the *Astasia rpl20* is very low (12.9%) and differs only slightly from the GC content of the flanking spacer regions (12.4%). The *Astasia rpl20* adjoins *trnD* (GUC) whereas in *Euglena rpl20* is separated from *trnD* by *psbI* which is not found in *Astasia* (Fig. 4). The N-terminal region of the *rpl20* polypeptide is much better conserved between *Astasia, Euglena,* land plants, and *E. coli* than is the C-terminus (50% identical amino acids between *Astasia* and *Euglena* in an N-terminal 66 amino-acid overlap).

A sequence encoding plastid ribosomal protein S4 was detected on *Hin*dIII-fragment H14. A striking feature of the deduced polypeptide is its high content of basic amino acids (24.5%). This compares well with the highly-basic character of the bacterial ribosome-assembly protein S4 and the S4 protein of chloroplasts (Subramanian et al. 1983). An internal fragment of the *rps4* gene hybridized to a 1.9-kb RNA (data not shown). The *Astasia rps4* gene is upstream of the rDNA repeats on the same strand, whereas it is downstream from the rRNA operons and on the opposite strand in *Euglena* (Hallick et al. 1993).

Open reading frames (ORFs)

Chloroplast genes that code for proteins of unknown function (ORFs), and are conserved in more than one organism are now designated with the gene prefix "*ycf*". Of the genes *ycf1–ycf12* occurring in land plants, none have yet been detected in *Astasia*. In *Euglena*, *ycf4*, *ycf8*, *ycf9*, and *ycf12* were found (Hallick et al. 1993). The gene locus *ycf13* is present exclusively in *Euglena* (*orf458*; Montandon et al. 1986) and *Astasia* (*orf456*; Siemeister et al. 1990 a). In addition, we detected an ORF coding for 170 amino acids that is a homologue to *Euglena orf161* (from nucleotide position 71 685 to 71 200). This locus is now designated *ycf14*.

Several ORFs are found only on *Astasia* ptDNA (Table 1). In the three large ORFs, designated *orf167*, *orf211*, and *orf559*, codons which end in T or A are used with much higher frequency than those ending in C or G. A similar

bias occurs in *tufA*, *rbcL*, the ribosomal protein genes, and *ycf13* in *Astasia* (unpublished data). This striking preference for T and A in the third position of codons reflects the extremely AT-rich composition of the *Astasia* ptDNA and has been documented also for a number of chloroplast genes in *Euglena* (see Hallick and Buetow 1989). The amino-acid sequence deduced from the hypothetical protein gene *orf559* shows a relatively high proportion of acidic residues (21%) and contains two direct repeats (66% identical residues) each of 39 amino acids that have a hydrophobic character. Transcripts (2.2 kb and 1.7 kb in size) of the hypothetical protein gene *orf559* were detected (data not shown). The size of the larger transcript suggests that it might be a cotranscript of *orf559* and at least one of the flanking tRNA genes (*trnV*, *trnI*) since the maximum size of a monocistronic *orf559* transcript should not exceed 2004 bp. Cotranscription in vivo of a gene encoding a protein and a tRNA gene has been reported for chloroplasts (Christopher and Hallick 1990).

Raison d'être of the Astasia ptDNA

It is not yet known whether *Astasia* lacks any components for plastid gene expression since the complete sequence is not available for the *Astasia* ptDNA. Complete sequencing of *Epifagus* ptDNA has demonstrated the loss of all chloroplast-encoded RNA polymerase genes and of many tRNA and ribosomal protein genes in this nonphotosynthetic parasitic plant (Wolfe et al. 1992). Since the *Epifagus* plastid genome is active (De Pamphilis and Palmer 1990; S. Ems and J. D. Palmer, unpublished data) nuclear gene products must compensate for some gene losses by means of previously unsuspected import mechanisms that may operate in all plastids (Wolfe et al. 1992).

Since the genes for photosynthetic functions – except *rbcL* – were found to be deleted in a highly specific manner from *Astasia* ptDNA, the genetic apparatus of the *Astasia* plastid genome must be maintained to express at least one protein with a nonphotosynthetic function. A merely selfish conservation of this genome does not appear to be sufficiently plausible, at least not at this high degree of conservation, if it were not needed for the synthesis of some gene product(s) that is (are) essential for a heterotrophic protist phylogenetically derived from photoautotrophic *Euglena*. By assuming that none of the proteins of the gene-expression apparatus have an unrecognized nongenetic function, there are a small number of genes in *Astasia* ptDNA that are candidates for being the *raison d'être* of the genome and its translational apparatus. (1) *Astasia* has retained and expresses the Rubisco subunit gene *rbcL* (Siemeister and Hachtel 1990a) but it is not yet known whether the small subunit of Rubisco is synthesized in *Astasia* and whether a functional Rubisco holoenzyme is assembled. If it is, one may speculate as to whether the oxygenase activity is vital for the synthesis of glycine and serine via the photorespiratory pathway. (2) Since *ycf13* is encoded within a group-III twintron in *Euglena* (it is not intron-encoded in *Astasia*), and the plastid genes of *Astasia* can contain group-III introns, the *ycf13* gene product may be required for group-III intron excision in both *Euglena* and *Astasia* (Hallick et al. 1993). (3) In addition, *Astasia* has some large ORFs that are absent in *Euglena* and do not show significant similarity with any plastid gene (Table 1). Transcripts of one of these ORFs (*orf559*) have been detected.

A similar situation to *Astasia* is that of the nonphotosynthetic parasitic flowering plant, *E. virginiana*, whose 70-kb plastid genome is completely sequenced and lacks all genes for photosynthesis present in the chloroplast genomes of green plants (Wolfe et al. 1992). One clearly important difference in *Epifagus* as compared to *Astasia* is that the parasite has not retained the *rbcL* gene. Conversely, homologues of *clpP*, *accD*, *orf1738*, and *orf2216* encoded by *Epifagus* ptDNA have not been found in *Astasia* or *Euglena*. *clpP* encodes the plastid homologue of the proteolytic subunit of the ATP-dependent Clp protease of *E. coli*, and *accD* encodes the plastid homologue of the β subunit of the carboxyltransferase component of *E. coli* acetyl-CoA carboxylase, whereas the functions of the two largest genes (*orf1738* and *orf2216*) are unknown. Given all these gene content differences, the primary function(s) of the *Epifagus* plastid genome is probably different from that of *Astasia*.

Acknowledgements. We thank G. Siemeister for helpful discussions, E. Raschke for assistance with database searches, H. Geithmann for photography and artwork, D. Lemke for typing the manuscript, and the Deutsche Forschungsgemeinschaft for financial support (grant Ha817/10-1 to W. H.).

References

Chen EY, Seeburg PH (1985) DNA 4: 165–170
Christopher DA, Hallick RB (1989) Nucleic Acids Res 17: 7591–7608
Christopher DA, Hallick RB (1990) Plant Cell 2: 659–671
Christopher DA, Cushman JC, Price CA, Hallick RB (1988) Curr Genet 14: 275–286
De Pamphilis CW, Palmer JD (1990) Nature 348: 337–339
Dix KP, Rawson JRY (1983) Curr Genet 7: 265–272
Fukuzawa H, Kohchi T, Sano T, Shirai H, Umesono K, Inokuchi H, Ozeki H, Ohyama K (1988) J Mol Biol 203: 333–351
Gingrich JC, Hallick RB (1985) J Biol Chem 260: 16156–16161
Graf L, Kössel H, Stutz E (1980) Nature 286: 908–910
Hallick RB, Buetow DE (1989) In: Buetow DE (ed) The biology of *Euglena*, vol 4. Academic Press, San Diego, pp 351–414
Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E (1993) Nucleic Acids Res 21: 3537–3544
Kössel H, Natt E, Strittmatter G, Fritzsche E, Gozdzicka-Jozefiak A, Przybyl D (1985) In: Vloten-Doting L van, Groot GSP, Hall TC (eds) Molecular form and function of the plant genome. Plenum Press, New York, pp 183–198
Kröger M, Wahl R, Schachtel G, Rice P (1992) Nucleic Acids Res 20: 2119–2144
Lipman DJ, Pearson WR (1985) Science 227: 1435–1441
Manzara T, Hallick RB (1988) Nucleic Acids Res 16: 9866
Manzara TB, Hu J, Price CA, Hallick RB (1987) Plant Mol Biol 8: 327–336
Monfort A, Rutti B, Stutz E (1986) Nucleic Acids Res 14: 3971
Montandon PE, Vasserot A, Stutz E (1986) Curr Genet 11: 35–39
Ohyama H, Fukuzawa H, Kohchi T, Shirai H, Sano S, Sano T, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Nature 322: 572–574
Orozco ME, Rushlow KE, Dodd JR, Hallick RB (1980) J Biol Chem 255: 10997–11003

Pringsheim EG (1942) New Phytol 41: 171–205
Roux E, Stutz E (1985) Curr Genet 9: 221–227
Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
Sanger F, Nicklen S, Coulson AR (1977) Proc Natl Acad Sci USA 74: 5463–5467
Shinozaki K, Ohme M, Tanaka M, 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, Kamagashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Shimada H, Sugiura M (1986) EMBO J 5: 2043–2049
Siemeister G, Hachtel W (1989) Curr Genet 15: 435–441
Siemeister G, Hachtel W (1990a) Plant Mol Biol 14: 825–833
Siemeister G, Hachtel W (1990b) Curr Genet 17: 433–438
Siemeister G, Buchholz C, Hachtel W (1990a) Mol Gen Genet 220: 425–432
Siemeister G, Buchholz C, Hachtel W (1990b) Curr Genet 18: 457–464
Sprinzl M, Hartmann T, Weber J, Blank J, Zeidler R (1989) Nucleic Acids Res 17: r1–r172
Subramanian AR, Steinmetz A, Bogorad L (1983) Nucleic Acids Res 11: 5277–5286
Sugiura M (1992) Plant Mol Biol 19: 149–168
Tabor S, Richardson CC (1987) Proc Natl Acad Sci USA 84: 4767–4771
Wolfe KH, Morden CW, Palmer JD (1992) Proc Natl Acad Sci USA 89: 10648–10652
Yoshimura M, Kimura M, Ohno M, Inokuchi H, Ozeki H (1984) J Mol Biol 177: 609–625
Zurawski G, Zurawski SM (1985) Nucleic Acids Res 13: 4521–4526

Communicated by K. Esser