A transcription map of the pea chloroplast genome

Neal W. Woodbury[1]*, Linda L. Roberts[1], Jeffrey D. Palmer[2], and William F. Thompson[1]**

[1] Carnegie Institution of Washington, 290 Panama Street, Stanford, CA 94305, USA
[2] Department of Biology, University of Michigan, Ann Arbor, MI 48109, USA

Summary. A set of 53 cloned pea chloroplast DNA fragments representing approximately 90% of the chloroplast genome was used to probe Northern blots of total pea RNA, resulting in a nearly complete chloroplast transcription map. Similar analyses were performed for RNAs extracted from pea seedlings grown under several different light regimes. We have found that at least 85 kb of the 120 bp pea chloroplast genome is represented as detectable transcripts. For many regions of the genome, we have detected multiple overlapping transcripts including both small, gene-sized RNAs and large transcripts covering entire gene clusters. All transcripts detected were more abundant (as a fraction of total cellular RNA) in light grown plants than in plants entirely in the dark. However, larger transcripts were generally more abundant in plants that had been exposed to only 24 h of white light (after germination in the dark) than in plants grown in continuous light. This study indicates that chloroplast genes are often grouped into multigene transcriptional units which can be cotranscribed, and that light-stimulated plastid development involves changes in the relative abundance of the overlapping RNAs of different length that result from transcription of these genes or gene clusters.

Key words: Chloroplast – Transcription Map – Pea – Pisum sativum

*Present address: Department of Chemistry, Arizona State University Tempe, AZ 85287, USA
**Present address: Department of Botany and Genetics, North Carolina State University, Raleigh, NC 27695, USA
Offprint requests to: N. W. Woodbury

Introduction

The 120 to 150 kb chloroplast genome of higher plants codes for about 50 polypeptides in addition to a partial set of ribosomal proteins, a full complement of rRNAs and probably a complete set of tRNAs (Whitfeld and Bottomley 1983; Posno et al. 1984; Herrmann et al. 1985; Shapiro and Tewari 1986; Shinozaki et al. 1986). Therefore, one would expect a large fraction of chloroplast DNA to be transcribed. Transcription has been detected over approximately 90% of the 146 kb chloroplast genome of Euglena by heteroduplex mapping (Koller and Delius 1984). Additionally, studies in which radio-labelled mustard chloroplast RNA was hybridized to approximately 50 chloroplast DNA fragments (Link 1984a) have confirmed that a large percentage of higher plant chloroplast DNA is transcribed as well, in agreement with earlier analyses (Whitfeld et al. 1976; Tewari 1980).

More detailed studies of transcription from selected regions of the chloroplast genome have shown that several different sized overlapping transcripts can encode the same gene sequences and that, in certain cases, large transcripts can be detected which encode two or more genes (see for example: Zurawski et al. 1982; Shinozaki et al. 1983; Alt et al. 1984; Barkan et al. 1986; Berends et al. 1986; Cozens et al. 1986). The occurrence of poly-cistronic transcripts in the chloroplast is in keeping with the prokaryotic-like organization of many chloroplast genes (Whitfeld and Bottomley 1983; Link 1984b; Gruissem and Zurawski 1985; Herrmann et al. 1985). What gives rise to the complex patterns of overlapping transcripts is as yet unclear.

The steady state levels of a number of chloroplast transcripts (relative to total cellular RNA) have been found to be greater in seedlings exposed to light than in etiolated seedlings (Smith and Ellis 1981; Thompson et

al. 1983; Link 1984a; Rodermel and Bogorad 1985; Zhu et al. 1985). More recently, the rate of chloroplast gene transcription during light-stimulated development has been compared directly to RNA abundance; these studies indicate that much of the specific regulation of chloroplast RNAs may be post-transcriptional (Deng and Gruissem 1987; Mullet and Klein 1987). If regulation of chloroplast gene expression is occurring at the level of RNA processing, one might expect to see gene-specific differences in the relative abundances of overlapping transcripts between plants grown under different light regimes. Such experiments should help to evaluate what role post-transcriptional RNA processing plays in the modulation of steady state transcript levels during light stimulated plastid development.

In this study, we use a set of small plasmid clones covering approximately 90% of the pea chloroplast genome to probe Northern blots of total RNA. Similar experiments have been performed using RNA extracted from seedlings grown under five different light regimes. The result is a nearly complete transcription map of the pea chloroplast genome as well as a genome-wide documentation of the effect that light fluence and spectral quality have on the relative levels of overlapping chloroplast RNAs.

Methods

Plant material and growth conditions. Seeds of *Pisum sativum* L. cv. Alaska (W. Atlee Burpee Co., Warmister, PA, USA) were imbibed for 6 h and germinated on two layers of water soaked Kimpac (Kimberley-Clarke, Roswell, GA), with approximately 100 seeds per 6 x 6 inch tray (Kaufman et al. 1984). One group of plants was grown under continuous white light (10^2 micromole m^{-2} s^{-1}, cool white fluorescent lamps) and a second group under continuous red light (10 micromole m^{-2} s^{-1}; light source described by Gorton and Briggs 1980) at 27 °C for 7 days. A third group was imbibed and grown for 6 days in absolute darkness at 27 °C. From this dark grown group, one set of plants was moved to continuous white light (conditions as noted above for white light grown plants) for 24 h; a second set was given a 15 min pulse of red light (10 micromole m^{-2} s^{-1}) and then returned to darkness for an additional 24 h; and the remaining dark grown seedlings were kept in darkness for the entire 7 day growth period. On day 7, apical buds (or tissues derived from apical buds in the case of light grown plants) were harvested into an ice cold vessel under white light (plants grown under continuous white light or exposed to 24 h of white light), red light (plants grown under continuous red light), or dim green safe lights (plants grown entirely in the dark or given a 15 min red pulse). The tissue was frozen in liquid nitrogen within 30 min of the beginning of harvesting and stored at −70 °C until RNA extraction.

Preparation of total pea RNA. RNA was prepared as described by Thompson et al. (1983). In this procedure, phenol extracted RNA is twice precipitated with LiCl which removes essentially all contaminating DNA. LiCl precipitation also results in the loss of transcripts less than about 300 bp.

Chloroplast DNA clones. For the Northern analysis, 53 pea chloroplast DNA fragments cloned into pUC vectors were used as hybridization probes. These clones, whose positions in the pea genome are shown in Fig. 1, were subcloned from a previously described clone bank (Palmer and Thompson 1981a), which consists of 11 of the 12 pea chloroplast PstI fragments cloned into pBR322. Complete restriction maps exist for all restriction endonucleases which were used for subcloning with the exception of BamHI, BglII, EcoRI and HindIII (Palmer and Thompson 1981b). Fragments produced by single or double digestion of the PstI clones were purified from agarose gels and cloned into appropriate pUC plasmids. Two clones were constructed that covered a portion of the only uncloned PstI fragment, P12.2 (Fig. 1B), using a 1.0 kb PstI-XhoI fragment and a 8.2 kb XhoI fragment isolated from digests of total pea chloroplast DNA. Two small regions of the pea genome are not contained in our clone banks: a 3.1 kb XhoI-PstI fragment that we were unable to clone, despite repeated attempts either by shotgun cloning of the parent 12.2 kb PstI fragment or by forced cloning of the gel-isolated 3.1 fragment and a 0.9 kb BamHI-BglII fragment in P5.7. We have no evidence to explain the unclonability of either fragment, although it is curious that the region included in the 3.1 kb fragment sustains deletions and insertions at a much higher frequency than any other region of the pea genome (Palmer et al. 1985).

Eight small, nearly gene-specific clones of spinach chloroplast DNA (Table 1) were also used as Northern probes in certain cases where only larger, less specific pea clones were available. Construction of these gene clones is described in Jansen and Palmer (1987).

Gel electrophoresis and filter hybridization. 5 μg aliquots of RNA were denatured in 6% formaldehyde, 25% formamide, 5 mM EDTA and 20 mM Na-Phosphate, pH 6.8 at 65 °C for 20 min (Thompson et al. 1983). Electrophoresis was carried out in 1.0–1.2% agarose, 6% formaldehyde, 20 mM MOPS, 5 mM Na-Acetate and 1 mM EDTA, pH 7 at room temperature. After electrophoresis, three to four 15 min washes of the gel in deionized water were followed by addition of approximately 0.5 μg/ml ethidium bromide. The gel was stained for about 10 min, destained for 2 h, photographed, and then soaked in 1 x SSC (0.15 M NaCl, 0.015 M Na-citrate) for 20 min. The RNA was transferred to Genescreen filters (New England Nuclear, Boston, MA, USA) by a modification of the method of Maniatis et al. (1982). The gel was placed on a flat surface. The filter was soaked in 1 x SSC for 10 min and placed on the top surface of the gel. Blotting paper was then stacked on top of the filter and held down by a weight of about 1 kg. Transfer was allowed to take place overnight; no buffer reservoir was used. The filters were then air-dried and baked for 2 h at 80 °C in a vacuum oven.

Filter hybridizations were performed essentially as described by Thompson et al. (1983). For each hybridization, approximately 100 ng of plasmid DNA was nick translated with 35 μCi of alpha ^{32}P dCTP (specific activity greater than 3,000 Ci/mmole), purified by gel filtration using a Sephadex G50 column, alkali denatured, and used as a probe. After 12−24 h of hybridization at 65 °C in 0.1% ficol, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin, 6 x SSC (1 x SSC = 0.15 M NaCl, 0.015 M sodium citrate), and 100 μg/ml denatured calf thymus DNA, the filters were washed extensively in 0.3 x SSC and 0.1% sodium dodecyl sulfate at 65 °C.

Transcript sizes below 4.4 kb were estimated by comparison to pUC and pBR322 DNA restriction fragments which had been denatured, electrophoretically separated and blotted in parallel with the RNA as described above. To extrapolate above 4.4 kb, the ethidium bromide staining pattern of denatured Lambda phage DNA restriction digests was used.

Table 1. Heterologous hybridization probes from spinach chloroplast DNA[a]

	Gene	Fragment[b]		Fragment location[c]	Reference
1	5′ psbE	0.65	kb (E-E)	−465	Herrmann et al. 1984
2	3′ psbE/psbF	0.5	kb (E-E)	+185	Herrmann et al. 1984
3	3′ psbC	367	bp (B-P)	+983	Alt et al. 1984
4	3′ atpE	420	bp (E-Xb)	+ 41	Zurawski et al. 1982
5	rpl2	772	bp (X-S)	+ 30	Zurawski et al. 1984
6	rps19	0.7	kb (S-P)	− 82	Zurawski et al. 1984
7	5′ petD	296	bp (B-B)	−162	Heinemeyer et al. 1984
8	5′ psbB	338	bp (B-B)	− 80	Morris and Herrmann 1984b

[a] Construction of these clones is described in Jansen and Palmer (1987)

[b] The length of the fragment is given in bp if it has been sequenced; otherwise it is given in kb. The restriction sites which define the fragment are given in parentheses. Enzyme abbreviations are as in Fig. 1 except that Xb is XbaI

[c] This number represents the position of the beginning of the DNA fragment relative to the first nucleotide of the gene's initiation codon. The fragment proceeds from this point towards the 3′ end of the gene

To control for possible *Escherichia coli* chromosomal DNA contamination of the plasmid DNA preparations of the chloroplast clones, Northern blots of pea RNA were probed with nick-translated *E. coli* DNA. Weak hybridization to chloroplast ribosomal RNA was observed. However, background hybridization with this pattern was never observed when using the purified plasmid clones as probes, even when using probes that did not themselves hybridize to any chloroplast transcripts.

Results

General properties of chloroplast transcription

In agreement with earlier studies (Koller and Delius 1984; Link 1984a), we have found that transcription occurs over a large fraction of the chloroplast DNA. Using the data summarized in Fig. 1, one can estimate that about 85 kb of the approximately 120 kb genome is represented by detectable transcripts (including transcription of the chloroplast ribosomal RNA operon, which is not shown). Other transcripts presumably exist which are below the threshold for detection. For example, there are at least three genes, rpoB, rpoC2 (Fig. 1A) and rpoA (Fig. 1C) that have been localized in the pea chloroplast genome by sequencing or by heterologous Southern hybridization for which no transcript is observed. Transcripts for these genes may be present at levels below our detection threshold, at least for the plant growth conditions and the particular tissues used in these experiments. In addition, transcripts smaller than about 300 bp are lost during our extraction procedure (see Methods). Thus, mature tRNAs and very small mRNAs would not be detected.

A generalization that can be made from Fig. 1 is that in most cases where groups of chloroplast genes are found clustered together, large transcripts are observed which

apparently encode two or more of these genes. For example, a 4.4 kb transcript hybridized throughout the region coding for atpF and atpA (Fig. 1A). Two large RNAs, 6.0 and 5.2 kb, cover the 60 kd, 20 kd, petA gene cluster (Fig. 1A). The region encoding psbC and psbD is covered by transcripts of 5.0, 4.2, 3.8 and 2.6 kb (Fig. 1B). In addition, there are three large RNAs (7.1, 6.4 and 5.6 kb) which hybridize to the DNA fragments including psaA and psaB sequences (Fig. 1B). Finally, a 5.1 kb RNA hybridized to all four of the DNA fragments which encode the gene cluster petD, petB, psbH, and psbB (Fig. 1C). Though the existence of these transcripts does not necessarily imply that these gene clusters are cotranscribed all the time, it does show that many chloroplast genes can be transcribed as part of a polycistronic RNA under at least some circumstances. It is noteworthy that genes which make up a transcriptional unit often appear to be grouped according to function. psaA and psaB, for example, are both genes whose products are thought to be components of the photosystem II reaction center; atpF, and atpA both code for ATPase components.

The relative positions of genes and gene clusters in the pea chloroplast genome is substantially different from that found in several other legumes such as mung bean and soybean. This has been interpreted to indicate that pea chloroplast DNA has undergone substantial rearrangement with respect to many other legume chloroplast DNAs (Palmer et al. 1988). In Fig. 1, the approximate positions of the rearrangement break-points between pea and mung bean are shown as long vertical arrows near the top of the restriction map (as determined by Palmer et al. 1988). In all cases, these break-points occur between cotranscribed gene clusters, indicating that many of the basic transcriptional units of the pea and mung bean chloroplast genomes have been preserved through evolution. The short vertical arrows above the restriction map
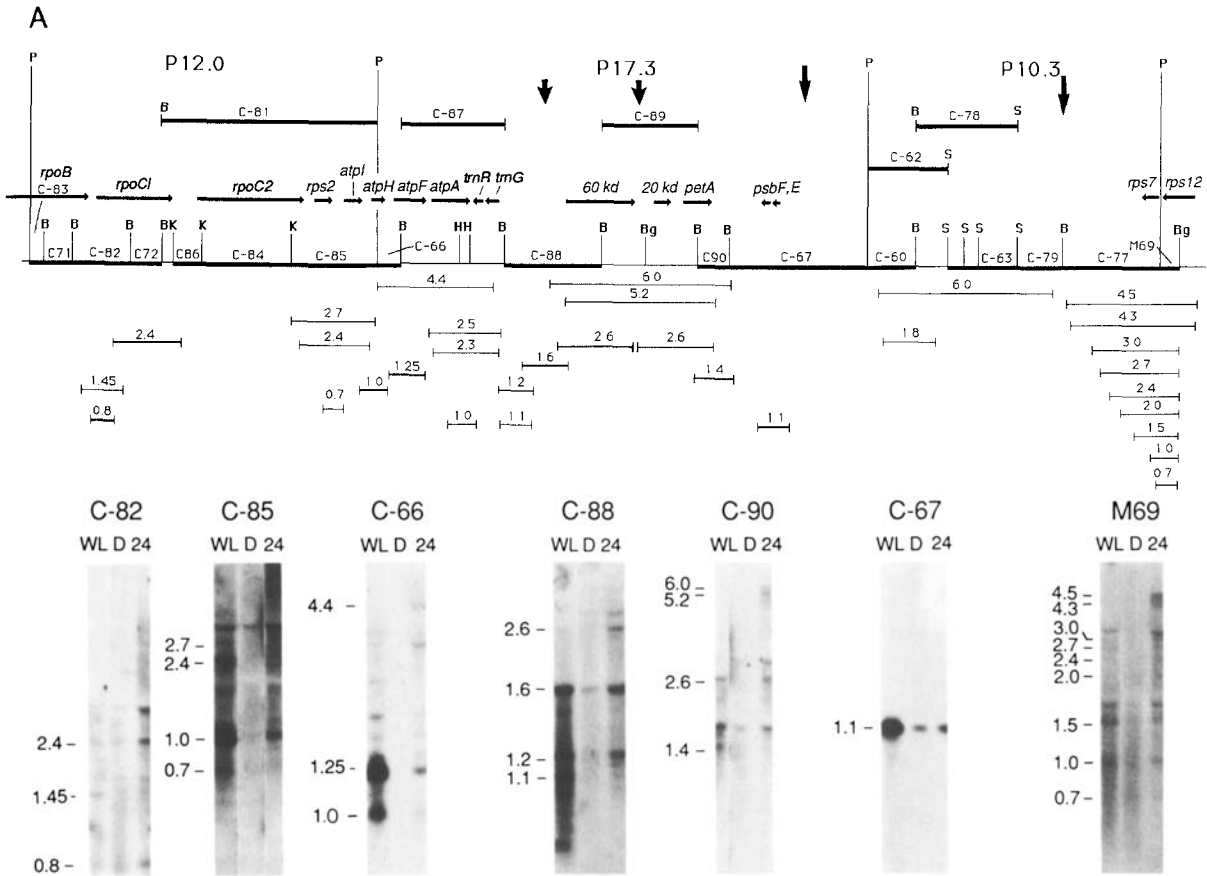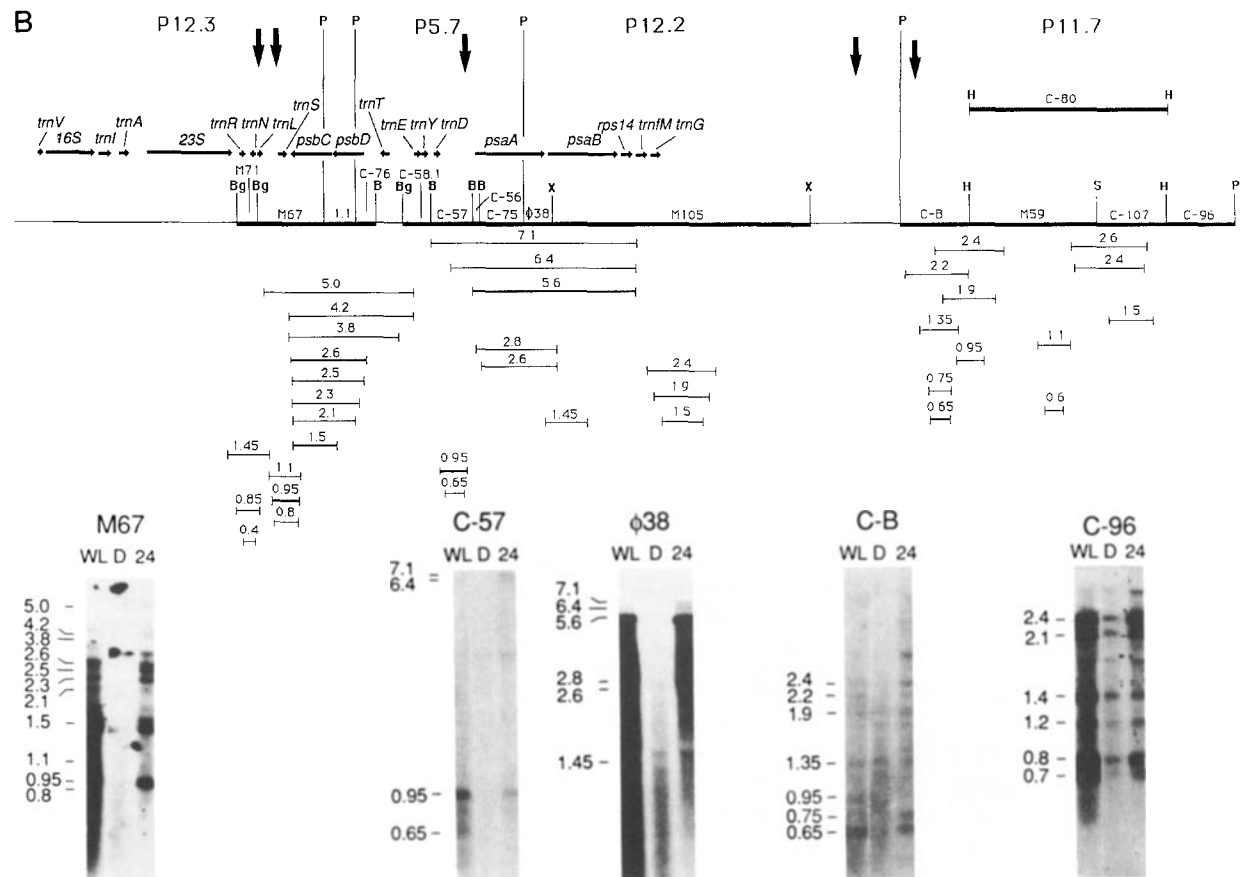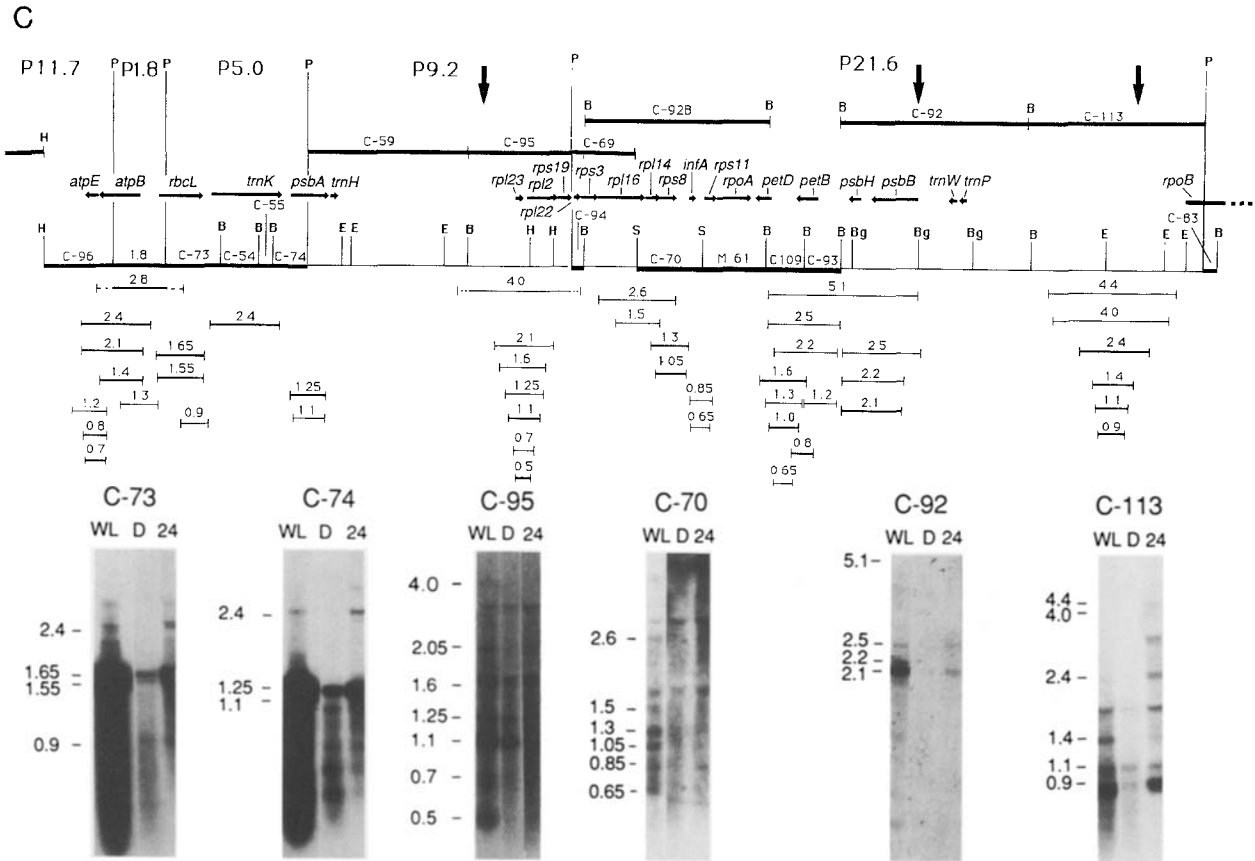
Fig. 1A

Fig. 1B

C

Fig. 1A–C. Transcription Map. Labeled chloroplast DNA clones were hybridized to Northern blots of total pea RNA. Sample blots using RNA preparations from three different sources are shown: 7 d old seedlings grown in continuous white light (WL), 7 d old dark grown (etiolated) seedlings (D), and 6 d old etiolated seedlings exposed to 24 h of white light before harvesting (24). The probe used is indicated above each autoradiogram. The source of RNA is indicated *at the top* of each lane. The RNA bands visualized on these blots are indicated by *tick marks* on the *left side* of each autoradiogram and labeled by their size in kb. The unlabeled bands which are often apparent at approximately 1.7 and 2.9 kb are exclusion artefacts due to overloading of cytoplasmic ribosomal RNA. The blots are placed approximately beneath the position of their probe sequence on the chloroplast DNA restriction map. The upper portion of each part of this figure includes a partial restriction map of the pea chloroplast genome (P: PstI, B: BamHI, K: KpnI, Bg: BglII, S: SalI, X: XhoI, H: HindIII, E: EcoRI). Each PstI fragment is labeled by its size in kb (e.g., P17.3 in part A). The clones used as probes in the Northern analysis are depicted as *bold lines* delimited by restriction sites. Chloroplast genes are shown as *horizontal arrows* above the restriction map pointing in the direction of transcription. The *long vertical arrows* in the upper portion of the figure represent rearrangement break-points determined by considering the relative arrangement of DNA sequences in pea and mung bean chloroplast genomes. The *short vertical arrows* near the top of *Fig. 1A* are the end-points of an inversion in *Pisum sativum* relative to *Pisum humile* (see text and Palmer et al. 1988). The transcripts detected are shown as thin or medium-thick lines below the restriction map and are labeled by their size in kb. Transcripts represented as partially dashed lines were larger than the total length of the clones that they hybridized to. In these cases, the transcript is depicted as a *solid line* beneath the clones which did show positive hybridization and then extended to the appropriate length by *dashed lines* on either end. The restriction map, genes, clones and transcripts are all drawn to the same scale except for extremely small genes (such as tRNA genes) which were elongated somewhat for clarity. The transcripts drawn with medium thick lines were those judged to be most abundant in the region of the map they appear. The positions of the transcript endpoints are only approximate since the Northern analyses do not permit precise mapping. In some cases, placement of the RNA with respect to the genome was guided by the strength of hybridization; greater signal intensity was interpreted as greater overlap between probe and transcript. Transcripts in the region of previously mapped genes were often drawn such that their endpoints were consistent with gene boundaries. All RNAs are drawn as though they mapped to a continuous, linear portion of the chloroplast genome. The map positions of tRNA genes shown in this figure but not mentioned in the text were taken from Shapiro and Tewari (1986), Rasmussen et al. (1984b), or Palmer et al. (1988)

in Fig. 1A indicate the endpoints of an inversion in *Pisum sativum* chloroplast DNA relative to the same region of the chloroplast genome of *Pisum humile*, the wild ancestor of pea (Palmer et al. 1985; Palmer et al. 1988). These inversion endpoints flank the 60 kd gene. Thus the 60 kd

gene is presumably not transcribed with the 20 kd and *pet*A genes in *P. humile*.

One characteristic that chloroplast transcription shares with transcription in the mitochondria of yeast (e.g., Morimoto et al. 1979) is that RNAs of many different

sizes often hybridize to the same DNA sequence (Fig. 1). Each of the cotranscribed gene clusters described above has at least five overlapping transcripts associated with it. The presence of several overlapping transcripts encoding the same gene sequences poses a serious problem if one is interested in studying the regulation of chloroplast gene expression by measuring RNA abundance. It is not at all clear which transcripts are functional (translatable in the case of protein coding genes) and which are simply the metastable intermediates or by-products of RNA processing and degradation.

Gene assignments

In order to identify the RNAs resulting from the transcription of specific chloroplast genes, we used not only the pea chloroplast clones shown in Fig. 1, but a series of nearly gene-specific chloroplast clones prepared from spinach (Table 1). In general, hybridization to the heterologous spinach clones was considerably weaker than hybridization to the homologous pea DNA clones. As a result we were only able to use the spinach probes to aid in the assignment of relatively abundant transcripts. Whether heterologous or homologous probes were used, assignments were generally made either on the basis of hybridization to probes containing only sequences internal to the coding region of the gene in question or by hybridization with two probes, one containing the 3' end and the other the 5' end of the gene. We did not attempt to assign transcripts to most tRNA genes since RNAs smaller than about 300 bp were lost during RNA isolation (Methods).

Included in Fig. 1 are samples of Northern blot autoradiograms used in the construction of the chloroplast transcription map. However, this only represents a fraction of the total Northern data that was utilized in the analysis. Northern blots are not shown for many of the clones used in this study. The blots that are shown are limited to only three of the five RNA preparations tested for each clone, and only one level of autoradiogram exposure. In addition, due to space constraints, only a portion of each blot autoradiogram could be displayed. As a result, there are several instances where RNAs that are shown on the map as hybridizing to a particular clone are not obvious on the sample autoradiogram. For example, the 5.2 and 6.0 kb RNAs which are shown extending into C-88 (Fig. 1A) do not appear in the corresponding autoradiogram because it was not possible to show both these large transcripts and the smear of small RNAs at the bottom of the blot simultaneously. Additionally, these large transcripts hybridized only very weakly to C-88 compared to the lower molecular weight RNAs and would not have shown up well on this exposure. Hybridization of C-90

to the 5.2 and 6.0 kb RNAs is shown in an adjacent autoradiogram. Another example is the blot shown which uses $\Phi 38$ as a probe (Fig. 1B). The three smallest transcripts depicted in the upper portion of the figure do not show up well in the sample autoradiogram due to overexposure of this region of the blot. This was done to clearly show the two largest RNAs which extend into this region. Thus, the sample blots shown are only a subset of a much larger bank of data used in the construction of the transcription map.

rpoB, rpoC1, rpoC2. Three genes which code for putative chloroplast RNA polymerase subunits, *rpo*B (homologous to the β subunit of *E. coli* RNA polymerase), *rpo*C1 (homologous to the amino terminal portion of the β' subunit of *E. coli* RNA polymerase) and *rpo*C2 (homologous to the carboxy terminal portion of the same subunit), cover clones C-83, C-71, C-82, C-72, C-86, and C-84 (Fig. 1A and autoradiogram of C-72 in Fig. 2B) (Cozens and Walker 1986; Palmer et al. 1988; Ohyama et al. 1986). Only three transcripts have been shown to hybridize to this region (2.4, 1.45, and 0.8 kb), and only one of those, the 2.4 kb RNA, is large enough to code for an entire polymerase subunit, *rpo*C1. (*rpo*C1 is actually slightly larger than 2.4 kb, but it contains a 0.7 kb intron). If *rpo*B and *rpo*C2 are expressed in pea chloroplasts, their transcripts must be present at very low abundance, at least under the growth conditions and in the tissue we have investigated.

rps2, atpI, atpH, atpF, atpA. rps2 encodes a protein in the small subunit of the chloroplast ribosome. This gene and a recently identified gene for a plastid encoded subunit of the chloroplast ATPase, *atp*I, are encoded on C-85 (Cozens and Walker 1986) (Fig. 1A). The *atp*I gene has been shown to be homologous to subunit a of the *E. coli* ATPase (Cozens et al. 1986). At least three transcripts hybridize to this DNA fragment, but do not hybridize appreciably to either adjacent clone. Of these, the largest transcript is 2.7 kb in length and presumably covers almost the entire 2.9 kb C-85 sequence. Stronger hybridization is seen to a 2.4 kb transcript, which must also cover most of C-85. Since both *atp*I and *rps2* are located at least several hundred bp from either boundary of C-85 (Cozens and Walker 1986; Cozens et al. 1986), their sequences are almost certainly encoded on both of these two transcripts, suggesting that these two genes are cotranscribed.

The *atp*H gene codes for the proton translocating subunit of the chloroplast ATPase, CF_0 III (Howe et al. 1982). This gene spans the junction between C-85 and C-66 (Fig. 1A) (Huttly and Gray 1984). Strong hybridization is seen between both of these clones and a 1.0 kb transcript. This RNA would be long enough to encode the entire 250 bp *atp*H gene (Howe et al. 1982), without extending into *atp*F.
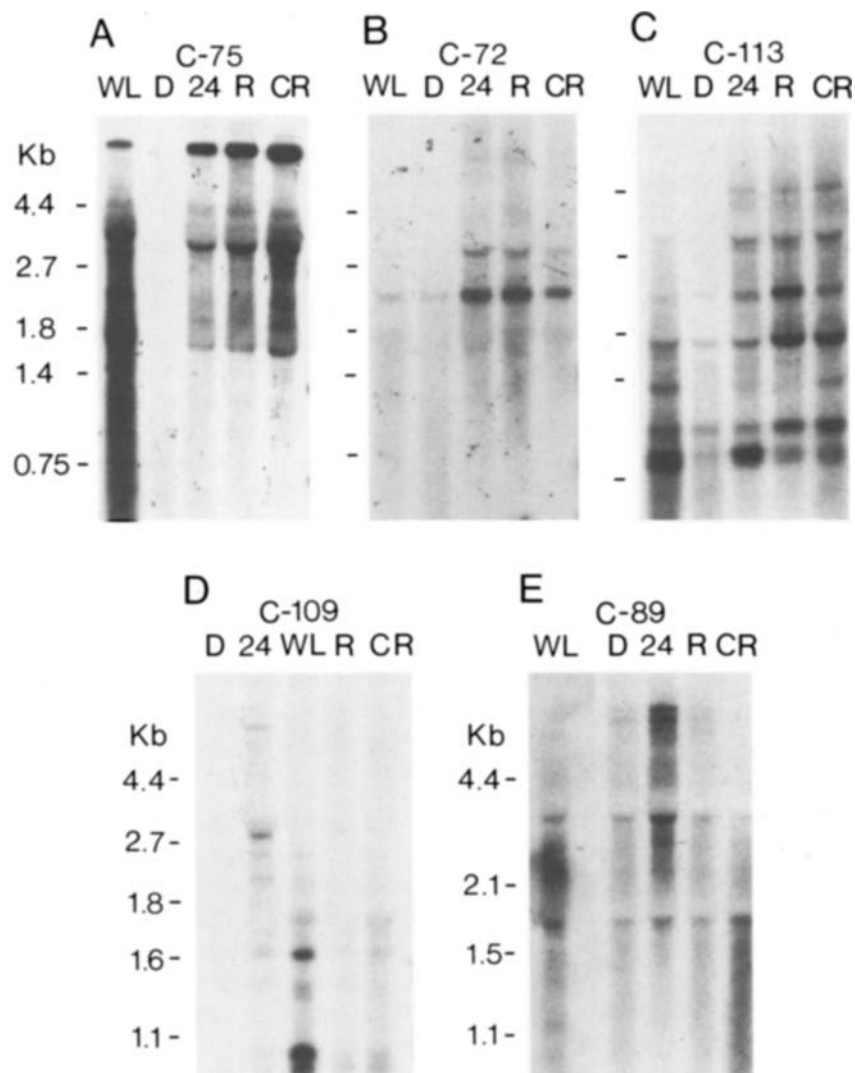
Fig. 2A–E. The effect of different light treatments on the relative abundance of overlapping chloroplast transcripts. Northern blot analysis performed and presented as in Fig. 1 except that the *tick marks to the left* of the autoradiograms show the positions of molecular weight standards. The continuous white light (WL), dark (D), and 24 h white light (24) *lanes* are as described in Fig. 1. In addition, Northern blot analyses of RNA samples from 7 d old seedlings grown in continuous red light (CR) and 6 d old etiolated seedlings given a saturating pulse of red light 24 h before harvesting (R) are also shown

The *atp*F gene, which codes for the CF_0-I subunit of the chloroplast ATPase, spans the junction between C-66 and C-87 (Fig. 1A). This gene contains a large intron (Cozens et al. 1986; Bird et al. 1985). The gene for the CF_1-α subunit of the ATPase, *atp*A, is encoded entirely within C-87 (Huttly and Gray 1984). A 4.4 kb transcript hybridizes weakly with both C-66 and C-87 and thus probably encodes both *atp*F and *atp*A. In maize (Barkan et al. 1986), spinach (Westhoff et al. 1985b), and wheat (Bird et al. 1985), Northern blot analyses of RNAs encoding *atp*A also reveal transcripts in this size range. Additionally we observe a smaller transcript of 1.25 kb which hybridizes to C-66 and C-87; the signal due to this fragment is particularly strong. This relatively abundant RNA is large enough to include the entire *atp*F gene sequence. Several other RNAs appear to hybridize exclusively to C-87 and are thus candidates for mature *atp*A transcripts. Of these,

the 2.5 and 2.3 kb transcripts are large enough to encode the entire *atp*A gene. These two transcripts are present primarily in plants exposed to white light for only 24 h, as discussed later.

A previous study of the transcription of *atp*I, *atp*H, *atp*F and *atp*A concluded that all 4 genes were cotranscribed in pea on a 6.0 kb RNA (Cozens et al. 1986). We did observe very weak hybridization to a transcript of this size (data not shown), but it was difficult to detect reproducibly. In our hands the abundance of large polycistronic transcripts is strongly dependent on growth conditions (see below). If these ATPase genes are cotranscribed, it is possible that *rps*2 is also on the same transcript; two of the transcripts hybridizing to C-85 apparently encoded both *rps*2 and *atp*I. A 6.0 kb primary transcript would be long enough to include all the sequence between the beginning of *rps*2 and the end of *atp*A.

60 kd, 20 kd, petA. An open reading frame (ORF) which potentially encodes a 60 kd protein, spans the junction between C-88 and C-89 (Fig. 1A) (Smith and Gray 1984). C-88, C-89 and C-90 all hybridize to transcripts of 2.6 kb. Since C-89 itself is larger than 2.6 kb, it seems likely that at least two different 2.6 kb transcripts are encoded by these three fragments, unless the 2.6 kb transcript is the product of a splicing event. One of the 2.6 kb RNAs probably encodes at least part of the 60 kd ORF. In addition there are two large RNAs, 6.0 and 5.2 kb, which hybridize to C-89, C-90, and probably to C-88. (Very weak signals in this size range were seen when C-88 was used as a probe.) It is probable that 60 kd ORF sequences are also present on these large RNAs.

Within C-89 there is an open reading frame which potentially codes for a 20 kd protein (Fig. 1A) (Willey et al. 1984). This sequence is probably present on the long (6.0 and 5.2 kb) transcripts which hybridize to C-89 (Fig. 2E). It could also be encoded on one of the 2.6 kb transcripts that hybridizes to this region.

The *petA* gene spans the junction between C-89 and C-90 (Fig. 1A) (Willey et al. 1984). *petA* codes for the apoprotein of cytochrome f. There are four RNAs which hybridize to both of these probes: the 6.0 and 5.2 kb transcripts discussed above, as well as a 2.6 kb and a 1.4 kb RNA (Fig. 1A and Fig. 2E). It seems likely that, at least under certain conditions, the 60 kd ORF, the 20 kd ORF, and *petA* are cotranscribed. Transcription of *petA* has also been investigated in spinach by Alt and Herrmann (1984). As we have seen for pea, large molecular weight RNAs (4.2 and 3.85 kb) were found to hybridize to *petA* gene sequences and to regions 5′ of *petA* in spinach. Additionally a 2.5 kb transcript was also shown to hybridize to *petA* sequences and to 5′ sequences (compare this to the 2.6 kb pea RNA shown in Fig. 1A covering both *petA* and the upstream 20 kd ORF) and a 1.5 kb spinach transcript was detected hybridizing only to *petA* sequences (compare this with the 1.4 kb pea RNA that hybridizes to C-90).

psbE, psbF. *psb*E and *psb*F, genes for the two components of the heme cross-linked heteromeric apocytochrome b-559 protein (Herrmann et al. 1984), are located within the C-67 chloroplast clone (Palmer et al. 1988) (Fig. 1A). This clone hybridizes very strongly to a single RNA species of 1.1 kb. In spinach, both *psb*E and *psb*F sequences hybridize to a 1.4 kb RNA implying that these two genes are cotranscribed (Westhoff et al. 1985a). In pea, we do not know where these genes are located within C-67 nor do we know which portion of C-67 is transcribed to produce the 1.1 kb RNA. However, two spinach chloroplast DNA fragments, one containing the 5′ end of *psb*E (Table 1, clone 1) and the

other containing the 3′ end of *psb*E and all of *psb*F (Table 1, clone 2) were found to hybridize to a 1.1 kb RNA, making it likely that the 1.1 kb transcript which hybridizes to C-67 does in fact encode these two genes.

rps7, 3′-rps12. In the vicinity of the PstI site which forms the junction between C-77 and M-69, there are two 30S ribosomal protein genes, *rps*7 and the 3′ end of *rps*12 (Palmer et al. 1988) (Fig. 1A). A number of RNAs hybridize to both C-77 and M69 including 4.5, 4.3, 3.0, 2.7, 2.4, 2.0, 1.5, 1.0 and 0.7 kb transcripts. In tobacco (Shinozaki et al. 1986) and soybean (von Allmen and Stutz 1987), *rps*12 is a divided gene; the 5′ exon is located more than 20 kb from the two 3′ exons. The position of the 5′ exon in pea is unknown, but some of the heterogeneity in RNA sizes detected in this region could be due to intermediates in *rps*12 transcript splicing.

rRNA. The pea chloroplast ribosomal RNA genes are encoded in the 12.3 kb PstI fragment (Fig. 1B) (pea chloroplasts contain only one copy of each ribosomal RNA gene). Small clones in this region were unavailable, thus transcription from rRNA genes is not considered here.

psbC, psbD. *psb*C, the gene for the photosystem II 44 kd polypeptide, and *psb*D, the gene for the D2 polypeptide of photosystem II are thought to be cotranscribed in pea (Berends et al. 1986; Bookjans et al. 1986). The coding sequences overlap by about 50 bp (Rasmussen 1984a). The combined sequences cover portions of M67, Pst1.1, and C-76 (Fig. 1B). There are four RNAs which hybridize to all three of these clones (5.0, 4.2, 3.8 and 2.6 kb). Strong hybridization is seen with the 2.6 kb RNA; hybridization signals from the other three large transcripts were quite weak. (Note that the 5.0 kb band is nearly invisible in the sample autorad of M67, while the 1.1, 0.95 and 0.8 kb bands are overexposed. All of these bands were visible in the original autorad.) All of these transcripts are large enough to encode both *psb*C and *psb*D, though only the 4.2 and 2.6 kb transcripts detectably hybridize to spinach chloroplast DNA fragments containing the 3′ portion of *psb*C (Table 1, clone 3).

There are several smaller transcripts which hybridize to both M67 and Pst1.1. Notably there is a 1.5 kb transcript which gives a strong signal and also hybridizes to the 3′ *psb*C probe from spinach. A putative transcription initiation site has been identified directly in front of the *psb*C coding sequence (Bookjans et al. 1986) which could give rise to the 1.5 kb transcript. Alternatively this RNA may be a processing product of one of the larger transcripts.

It is worth noting that the 4.2 kb transcript discussed above hybridizes to C-58.1, substantially 5′ of the be-

ginning of the *psb*D coding sequence. (Presumably, the 5.0 kb fragment also extends into C-58.1, but the abundance of this transcript is very low and hybridization to it was not observed when C-58.1 was used as a probe). Apparently *trn*T can be cotranscribed with *psb*C and *psb*D. However, sequences resembling prokaryotic transcription initiation consensus sequences ("−35" and "−10" like sequences) have been found just 5′ to the *psb*D coding sequence, implying that some initiation may occur in this region as well (Rasmussen et al. 1984a).

Transcription of *psb*C and *psb*D has also been characterized in spinach (Alt et al. 1984). As in pea, the coding sequences for these genes overlap by 50 bp and several large transcripts can be detected by Northern blot analysis which hybridize to sequences in both genes. Specifically, 4.6, 3.8, 3.6 and 2.7 kb RNAs were reported, very closely matching the transcription pattern we observe in pea.

*psaA, psaB. psa*A and *psa*B code for the 84 kd and 82 kd photosystem I proteins, respectively. *psa*A is contained on C-56, C-75 (autoradiogram shown in Fig. 2A) and Φ38 (the blot probed by this clone which is shown in Fig. 1B is overexposed so that the 6.4 and 7.1 kb bands are clearly visible); *psa*B has its 5′ end in Φ38 and extends into M105 (Lehmbeck et al. 1986) (Fig. 1B). The predominant RNA which hybridizes to all four of these clones is a 5.6 kb transcript. Thus, *psa*A and *psa*B are apparently cotranscribed in pea. This inference is in agreement with conclusions drawn from the DNA sequence of this region (Lehmbeck et al. 1986) and with previous studies of *psa*A and *psa*B transcription in pea (Berends et al. 1986) and in maize (Barkan et al. 1986). There are two larger RNAs which also hybridize throughout this region. One of these is approximately 6.4 kb and the other is approximately 7.1 kb. Hybridization signals from these transcripts are much weaker than those from the 5.6 kb RNA. C-57 (5′ of the *psa*A gene) also hybridizes to the 6.4 and 7.1 kb RNAs but does not hybridize to the 5.6 kb transcript. Instead a strong signal at about 0.95 kb is observed. It is tempting to speculate that the 0.95 kb and 5.6 kb RNAs result from a processing event cleaving the 6.4 kb RNA. However, DNA sequence data and S1 analyses have identified a putative promoter site about 150 bp 5′ to the beginning of the *psa*A coding sequence in maize (Fish et al. 1985). It is possible that there is a much longer 5′ transcribed region associated with the pea *psa*A/B genes than with the maize genes or that the maize S1 analysis has simply located a processing site. A processing site in approximately this position would be consistent with the formation of the 5.6 kb and 0.95 kb RNAs from one of the larger RNAs which hybridizes in this region.

atpE, atpB. In pea, the sequences coding for the CF_1-ϵ and CF_1-β subunits of the ATPase (genes *atp*E and *atp*B, respectively) span the junction between C-96 and Pst1.8 (Fig. 1B and C) (Zurawski et al. 1986a). The *atp*E and *atp*B genes are thought to be cotranscribed in spinach (Zurawski et al. 1982), tobacco (Shinozaki et al. 1983) and maize (Krebbers et al. 1982). In pea we have found three transcripts which hybridize strongly to both C-96 and Pst1.8 (2.4, 2.1 and 1.4 kb). All three also hybridize to a 420 bp spinach chloroplast DNA probe that includes the 3′ end of the *atp*E coding region (Table 1, clone 4). Either the 2.4 or 2.1 kb transcript would be large enough to contain the entire sequence for both of these genes. In tobacco, *atp*B and *atp*E are encoded on a 2.7 kb transcript (Shinozaki et al. 1983). S1 mapping of the 5′ ends of *atp*B in pea has shown only one site about 350 bp upstream from the *atp*B initiation codon (Mullet et al. 1985). The placement of the 2.4 kb RNA in Fig. 1 reflects this finding. C-96 strongly hybridizes to two small RNAs of 0.8 and 0.7 kb which are not seen when P1.8 is used as a probe. It is possible that these RNAs could come from transcription of *atp*E initiating to the left of the PstI site separating C-96 from P1.8 or from transcription of sequences 3′ of the *atp*E gene (Fig. 1C). Alternatively, the 0.8 and 0.7 kb transcripts could arise by processing of the 2.4 or 2.1 kb RNAs. In the later case, the 1.4 kb RNA which hybridizes to both C-96 and P1.8 could be the complementary processing product.

rbcL. The *rbc*L gene, which encodes the large subunit of ribulose bisphosphate carboxylase, spans the junction between Pst1.8 and C-73 (Fig. 1C) (Zurawski et al. 1986b). There are at least two transcripts which hybridize to both probes. The more abundant of the two is about 1.65 kb. A weaker transcript at about 1.55 kb can also be seen. (The sample autoradiogram of C-73 in Fig. 1C is overexposed to show the 2.4 kb band above the main *rbc*L transcript. The 2.4 kb RNA is discussed below.) These two transcripts may correspond to the processed and unprocessed *rbc*L RNAs described previously (Mullet et al. 1985). The *rbc*L transcripts in Fig. 1 were positioned using the transcription initiation site determined by Mullet et al. Two *rbc*L transcripts of about this size have also been observed in maize chloroplasts (Erion 1985).

*trnK. trn*K, a tRNA^{Lys} gene which in tobacco and mustard has been shown to include a large intron (Shinozaki et al. 1986; Neuhaus and Link 1987), occupies most of the coding sequence between *rbc*L and *psb*A (Palmer et al. 1988) (Fig. 1C). In mustard, the tRNA^{Lys} intron contains an open reading frame potentially coding for a maturase-related polypeptide (Neuhaus and Link 1987). C-73, C-54, C-55, and C-74 all hybridize to

a transcript of 2.4 kb. This RNA is large enough to encode the *trn*K intron.

psbA. *psb*A codes for a protein which is the binding site of a quinone involved in electron transport on the acceptor side of photosystem II (32 kd, Q_b protein or D1). The *psb*A sequence spans the PstI site that separates C-74 and C-59 (Fig. 1C) (Oishi et al. 1984). The strongest band that hybridizes to both probes is a 1.25 kb RNA. (The sample autoradiogram of C-74 in Fig. 1C is over-exposed to show the 2.4 kb RNA discussed above.) The transcription initiation site of this gene has been mapped as has the position of one processing site about 24 bp 3' of the initiation site (Boyer and Mullet 1986). This information was utilized in the placement of the 1.25 kb transcript in Fig. 1. There are also smaller RNAs which hybridize much more weakly to this region than the 1.25 kb transcript. These are presumably degradation products of the larger *psb*A transcript.

rpl23, rpl2, rps19, rpl22, rps3, rpl16, rpl14 and rps8. There is a cluster of eight ribosomal protein genes, *rpl*23, *rpl*2, *rps*19, *rpl*22, *rps*3, *rpl*16, *rpl*14 and *rps*8, (Shinozaki et al. 1986) which extends from the right side of P9.2 into P21.6 (Palmer et al. 1988) (Fig. 1C). C-95 hybridizes very weakly to RNAs of 2.1 and 4.0 kb and more strongly and distinctly to several smaller transcripts. The strongest hybridization is to a 1.1 kb transcript. This transcript and a 1.25 kb transcript hybridize to a spinach chloroplast DNA probe for *rpl*2 (Table 1, clone 5), making it likely that at least those two transcripts encode part or all of that gene. We have not been able to detect any transcripts which hybridize to a spinach chloroplast DNA fragment containing *rps*19 sequences (Table 1, clone 6). Southern hybridization between chloroplast DNA and the spinach probe for *rps*19 also showed little or no signal strength implying that the homology between spinach *rps*19 and pea *rps*19 is low or that *rps*19 is absent from the pea chloroplast genome (Palmer, unpublished data).

When C-69 and C-70 were used as probes, weak hybridization was observed to at least two transcripts of 2.6 and 1.5 kb. These RNAs may encode one or more of the ribosomal protein genes 3' of the PstI site. C-70 also hybridizes to four smaller RNA transcripts, two of which (1.3 and 1.05 kb) are apparently present at a significantly greater abundance than the larger RNAs. These transcripts could include *rpl*14 and/or *rps*8.

infA, rps11, rpoA. Adjacent to the ribosomal protein gene cluster is a group of three genes: *inf*A, which is thought to code for a chloroplast transcription initiation factor homologous to the *E. coli* initiation factor IF-1; *rps*11, a 30S ribosomal protein gene; and *rpo*A, a putative gene for the α subunit of a chloroplast DNA

dependent RNA polymerase (Palmer et al. 1988). The coding sequences of these genes span an approximately 2.5 kb region including the junction between C-70 and M61. Besides the 1.3 and 1.05 kb RNAs discussed above that hybridize only to C-70, two small RNAs (0.85 and 0.65) appear to hybridize to both C70 and M61. These RNAs could contain *inf*A and *rps*11 sequences. No transcripts were observed that hybridized only to M61 and not to either neighboring clone. Thus it was not possible to assign a transcript to *rpo*A. Of the four putative chloroplast RNA polymerase genes, *rpo*A, *rpo*B, *rpo*C1 and *rpo*C2, we have only been able to detect transcripts for *rpo*C1 in pea (see above). In spinach, very low level run-on transcription of *rpo*A has also been reported (Deng and Gruissem 1987).

The number and size of the RNAs which hybridize to this region of pea chloroplast DNA are not as large in pea as was seen previously by Northern blot analysis in spinach. There, more than 6 RNAs were observed with lengths as long as 6 kb (Sijben-Muller et al. 1986). (Probes specific for *rpo*A were apparently not used in this study.) It is possible that the transcriptional organization of this region of the pea chloroplast genome is different from that found in spinach. However, it is also possible that processing or degradation of these large RNAs is faster in pea than it is in spinach, making only the smaller transcripts detectable.

petD, petB, psbH, psbB. About 200 bp from the 3' end of *rpo*A is the 3' end of *pet*D, which encodes subunit 4 of the cytochrome b/f complex (Heinemeyer et al. 1984; Phillips and Gray 1984). The *pet*D coding sequence is oriented in the opposite direction from *rpo*A and spans the junction between M61 and C-109 (Fig. 1C, autoradiogram of C-109 shown in Fig. 2D). The most abundant transcript which hybridizes to both M61 and C-109 is a 1.6 kb RNA. This transcript also hybridizes to a spinach chloroplast DNA fragment containing the 5' 134 bp of *pet*D (Table 1, clone 7), and is large enough to code for the entire *pet*D gene product. Hybridization of this transcript is much more distinct using C-109 as a probe; M61 gives a weak signal which may indicate that most of the 1.6 kb RNA is encoded by C-109. The 5.1 kb transcript which hybridizes throughout this region will be discussed below.

The gene for apocytochrome b6, *pet*B, spans the BamHI site that joins C-109 and C-93 (Palmer et al. 1988) (Fig. 1C). There are at least five transcripts which hybridize to both C-109 and C-93. The most prominent of these is 0.8 kb in length, which is large enough to encode the entire *pet*B gene product.

*psb*H codes for a 10 kd phosphoprotein associated with photosystem II and is entirely contained on C-92 (Palmer et al. 1988; Westhoff et al. 1986) as is *psb*B, the gene for the photosystem II 51 kd chlorophyll a apo-

protein (Palmer et al. 1988; Westhoff et al. 1983; Morris and Herrmann 1984a) (Fig. 1C). C-92 hybridizes to four RNA transcripts (5.1, 2.5, 2.2, and 2.1 kb), with the strongest signal resulting from hybridization to the 2.1 kb transcript. (The 5.1 kb RNA hybridizes rather weakly to C-92 in Fig. 1C and can be seen only as a faint band in the lane marked "24".) A spinach DNA fragment which encodes the 5' 258 bps of *psb*B (Table 1, clone 8) was found to hybridize to both the 2.2 and 2.1 kb transcripts, suggesting that these two RNAs may include part or all of the *psb*B gene sequence. Hybridization to the 5.1 and 2.5 kb RNAs was presumably too weak to detect with this heterologous probe. Any of these RNAs would be large enough to encode both *psb*H and *psb*B.

In spinach and maize, *pet*D, *pet*B, and *psb*B are apparently cotranscribed to produce RNAs of 6.0 kb (spinach) and 5.4 kb (maize) (Heinemeyer et al. 1984; Morris and Herrmann 1984a; Westhoff 1985; Barkan et al. 1986). Further, in spinach, a 2.5 kb sequence hybridizes to both *pet*B and D but not *psb*B and a 2.1 kb transcript hybridizes to *psb*B and not to *pet*B or *pet*D (Morris and Herrmann 1984a). In addition to these transcripts, a variety of other RNAs that hybridize to this region have also been detected in both species. The transcript pattern for this region of the pea genome is quite similar to that described above for spinach and maize. An approximately 5.1 kb transcript covers the entire region between *psb*B and *pet*D. C-92 containing *psb*H and *psb*B hybridizes strongly to a 2.1 kb RNA, and *pet*B and *pet*D on C-109 and C-93 both hybridize to RNAs between 2.0 and 2.5 kb, much as was seen for spinach. Our results are in general agreement with previous studies of transcription in this region of the pea chloroplast genome (Berends et al. 1986) except that only one large molecular weight RNA was detected instead of three. The similarity of the transcription pattern in pea to that in maize and spinach suggests that in pea chloroplasts the sequence between *pet*B and *pet*D may be spliced out after transcription as it is in spinach (Herrmann et al. 1985; Sijben-Muller et al. 1986; Heinemeyer et al. 1984).

Transcripts hybridizing to regions without mapped genes

Not all of the chloroplast fragments which hybridize to RNA transcripts on Northern blots are from regions of the chloroplast genome where genes have been mapped. C-113 (between *psb*B and *rpo*B; see Fig. 1C) hybridizes to several RNAs including two large transcripts 4.4 and 4.0 kb in length, a 2.4 kb transcript, and a few smaller RNAs (Fig. 1C, Fig. 2C). The most interesting of these is the 2.4 kb transcript. This RNA is most abundant in otherwise dark grown pea seedlings which have

been given a short pulse of red light or which have been placed for 24 h in white light (Fig. 2C). It is significantly less abundant in plants which have been grown continuously in white light (see below). In the region adjacent to C-113 (including *rpo*B, *rpo*C1 and *rpo*C2) C-82, C-72 and C-86 (but not C-71) also hybridize to a 2.4 kb RNA which is most abundant in plants which have been grown in the dark and then exposed briefly to red or white light (Fig. 1A; also compare the Northern blot of C-72 in Fig. 2B to that of C-133 in Fig. 2C). The relationship between these two 2.4 kb transcripts, if any, is unclear, but the similarity of their sizes and light responses makes it tempting to speculate that they may result from a splicing event or possibly be two different but similar sized RNAs derived from the same primary transcript and thus governed by the same regulatory mechanisms. This would be consistent with the idea that *rpo*B, *rpo*C1 and *rpo*C2 are cotranscribed from a promoter in C-113 onto one large primary transcript which is then subject to processing. No RNA large enough to serve this purpose was observed in this region. However, the steady-state abundance of this transcript may be below our detection threshold due to rapid processing. This scheme does not explain why the only mature transcript observed is that for *rpo*C1.

The region of the chloroplast genome which includes C-60, C-78 and C-79 hybridizes weakly to 6.0 and 1.8 kb transcripts (Fig. 1A, P10.3). No genes have yet been localized in this region.

The only gene coding sequences known to be contained on the P11.7 (Fig. 1B and 1C) are in the clone C-96 which includes *atp*E and the 3' end of *atp*B. However, a large number of transcripts hybridize to C-B, a PstI-HindIII fragment at the opposite end of P11.7 from C-96 (Fig. 1B). In addition several transcripts have been detected hybridizing to M59 and C-107. Some of these RNAs are quite abundant, particularly a 0.65 kb transcript from C-B.

In addition to the transcripts described above, there are a number of RNAs which hybridize to DNA fragments known to contain gene coding sequences, but which cannot be unambiguously identified as transcripts from those genes. For example, there are three transcripts which hybridize to M105 (2.4, 1.9 and 1.5 kb; see Fig. 1B, P12.2) which may or may not be associated with *psa*B or *rps*14. Similarly, the 1.2, 0.8 and 0.7 kb transcripts which hybridize to C-96 (Fig. 1C, P11.7) and which we have suggested are *atp*E transcripts may code for sequences 3' to the *atp*E gene. In addition there are a number of small (less than 1.0 kb) RNAs which could arise from sequences between genes or adjacent to known genes (see, for example, the small transcripts hybridizing to C-57 or M67, Fig. 1B). Some of these RNAs may be remnants of processing events or possibly due to transcription of DNA sequences

which do not encode functional genes, but some may also result from transcription of functional genes not yet identified.

Comparison of different light treatments

As discussed above, many of the probes used in the mapping experiments hybridized to RNAs of several different sizes. The relative abundance of different transcripts hybridizing to the same probe varied when the RNA was prepared from plants grown under different light regimes. Almost all pea chloroplast transcripts represent a much smaller fraction of the total cellular RNA extracted from plants grown in the dark for 7 days than from plants grown for the same period in the light. This situation contrasts with that in mustard where most chloroplast transcripts are present at relatively high levels in dark grown plants (Link 1984a). The difference between the steady state abundance of chloroplast RNAs in pea and mustard may reflect different stages of plastid development achieved in dark grown seedlings.

There are some pea chloroplast transcripts which are present at significant levels in etiolated tissue. Three transcripts which hybridize to C-113 (2.4, 1.1 and 0.9 kb; Fig. 1C and Fig. 2C) are moderately abundant in dark grown pea buds. The same is true for the 1.6 kb transcript which is seen when C-88 is used as a probe (Fig. 1A) and the 1.35 kb transcript that hybridizes to C-B (Fig. 1B). In addition, some of the RNAs associated with the genes rpoC1 (2.4 kb; Figs. 1A and 2B), petA (6.0 and 5.2 kb; Figs. 1A and 2E), psbE and psbF (1.1 kb; Fig. 1A), atpB and atpE (2.4, 2.1, 1.4, 1.2, 0.8, and 0.7 kb; Fig. 1B and C), rbcL (1.65 and 1.55 kb; Fig. 1C), psbA (1.25 kb; Fig. 1C), and the region including rpl23 and rpl2 (1.6, 1.25 and 1.1 kb; Fig. 1C) are also present at significant levels in 7 day old dark grown pea buds. However, in every case relative RNA abundance was greater in plants which had been exposed to light than in totally dark grown material. This does not necessarily imply that the rate of transcription of chloroplast genes is greater in plants which have seen light than in dark grown plants. First, the Northern analysis revealed only the steady state transcript abundance which depends both on rates of RNA synthesis and turnover. Second, RNA abundance in these experiments was always evaluated relative to total cellular RNA; no attempt was made to correct for differences in the number of copies of the plastid genome per cell under different growth conditions. It has been reported that pea chloroplast DNA copy number is somewhat greater in light grown peas than in dark grown seedlings (Thompson et al. 1983).

In addition to 7 day dark grown plants and 7 day white light grown plants, we compared the chloroplast RNA levels produced under three other light regimes: plants grown in continuous red light, plants grown for 6 days in the dark and then given 24 h of white light or plants grown for 6 days in the dark and then given a saturating red pulse followed by 24 h of darkness. Most transcripts are more abundant in white light grown seedlings than in seedlings grown under any other light regime. However, there are a number of cases in which a higher abundance of a particular transcript was observed in dark grown plants exposed to 24 h of white light. One example, discussed above, is the 2.4 kb transcript(s) which hybridize to C-113 and rpoC1 (Fig. 1A and C and Fig. 2B and C). However, many of the transcripts which fit this pattern are the larger members of families of RNAs hybridizing to the same probe. Examples include the 5.1 kb transcript encoding petD, petB, psbH and psbB (M61, C-109, C-93 and C-92, Fig. 1C and Fig. 2D), the 4.4 and 3.4 kb transcripts for atpF and atpA (C-66 and C-87, Fig. 1A), the 6.0 and 5.2 kb transcripts for the 60 kd gene, the 20 kd gene, and petA (C-88, C-89 and C-90, Fig. 1A and Fig. 2E), the 6.0 kb RNA hybridizing to the region including C-78 (Fig. 1A), the 5.0 and 4.2 kb RNAs which encode psbC and psbD (M67, P1.1 and C-76, Fig. 1B), the 7.1 and 6.4 kb transcripts for psaA and psaB (C-57, C-56, C-75, Φ38 and M105, Fig. 1B and Fig. 2A), and the 4.0 kb transcript for covering the region around rpl2 (C-95, Fig. 1C).

In the case of the two largest transcripts encoding psbC and psbD (M67, P1.1 and C-76, Fig. 1B) and the high molecular weight transcripts which hybridize to C-113 (Figs. 1C and 2C), plants grown in continuous red light also had higher transcript levels than continuous white light grown plants. In these cases as well as several others discussed above (C-113, rpoC1, rps7 and rps12, psbC and psbD, psbA and psaB, rpl2 and rps19) a saturating red light pulse after 6 days of dark growth has an effect similar to that of 24 h of white light (see for example Fig. 2A, B and C).

In many of the cases noted above, the lower relative abundance of large RNAs in seedlings grown in continuous white light versus seedlings exposed to only 24 h of light is correlated with a greater relative abundance of low molecular weight transcripts in the continuous light plants. An example of this is shown in Fig. 2D where C-109 containing parts of petD and petB is used as a probe. Here the 5.1 kb RNA present in plants exposed to only 24 h of white light is undetectable in RNA from plants grown entirely under continuous white light. However, growth in continuous white light results in much higher levels of 1.6, 1.0 and 0.8 kb RNAs.

Discussion

Transcription of chloroplast gene clusters

Most of the protein coding genes in Fig. 1 can be grouped into gene clusters which are, at least under certain circumstances, cotranscribed. This organization is in keeping with other prokaryotic-like aspects of chloroplast gene structure and expression. However, unlike most prokaryotic operons, the transcription of chloroplast gene clusters results in a tremendous complexity of RNAs. Almost every group of two or more cotranscribed chloroplast genes in Fig. 1 gives rise to at least five RNAs that are present at significant steady state levels under one or more of the growth conditions used in this study. Perhaps the simplest explanation for this complex pattern of transcripts is that most of the RNAs observed are intermediates in the degradation of full length transcripts. However, it is difficult to understand why degradation would result in the accumulation of discrete intermediates with concentrations that are often much greater than those of the larger putative mRNAs.

One difference between gene expression in chloroplast and prokaryotes is the frequent utilization of post-transcriptional RNA processing to modify many chloroplast transcripts. For example, the primary transcript for the ribulose bisphosphate carboxylase large subunit gene (rbcL) can be cleaved at a point 113 bp from the initiation site post transcriptionally (Mullet et al. 1985). This cleavage presumably gives rise to the two rbcL transcripts observed at 1.65 and 1.55 kb (Fig. 1C). Processing of the primary transcript for psbA is thought to occur about 24 bp 3' of the transcription initiation site (Boyer and Mullet 1986), though this modification is too small to detect by Northern blot analysis. In addition, RNA splicing has been indicated in several higher plant species (Shinozaki et al. 1986; Neuhaus and Link 1987; von Allmen and Stutz 1987; Koch et al. 1981; Deno and Sugiura 1984; Torazawa et al. 1986; Fromm et al. 1986). For example, in tobacco and soybean rps12 has been shown to be encoded on three exons one of which is located more than 20 kb from the other two (Shinozaki et al. 1986; von Allmen and Stutz 1987); in mustard and tobacco, trnK contains a large intron which itself may code for a polypeptide (Shinozaki et al. 1986; Neuhaus and Link 1987). Most recently, specific processing of the 3' ends of chloroplast RNAs has been demonstrated in chloroplast extracts (Stern and Gruissem 1988). Post-transcriptional modification of chloroplast transcripts by any of these mechanisms could give rise to a variety of intermediate RNAs between the primary and mature transcript. Many of the overlapping RNAs depicted in Fig. 1 could represent intermediates in these processing pathways.

Finally, a third possible explanation for the number of overlapping chloroplast transcripts seen in Fig. 1 is that the transcription of a given chloroplast gene or gene cluster can initiate and terminate at any of a number of positions. We are presently attempting to determine which chloroplast transcripts can be capped by guanylyl transferase in vitro, and thus to identify transcripts which retain their original 5' ends. This information should aid in distinguishing between these possibilities.

This problem has been studied more extensively in yeast mitochondria. The genes of these organelles, like those of the chloroplast, are transcribed to produce many different sized overlapping RNAs (Morimoto et al. 1979). Guanylyl transferase capping experiments like those proposed above have shown that the large number of overlapping RNAs in this system is due both to genes (or groups of cotranscribed genes) with multiple transcription initiation sites and to the accumulation of RNA processing intermediates (Levens et al. 1980; Christianson and Kabinowitz 1983).

The effect of light stimulated development on the relative levels of overlapping chloroplast transcripts

Unlike mustard (Link 1984a), spinach (Deng and Gruissem 1987) and barley (Mullet and Klein 1987), pea plastids do not produce significant levels of most protein coding transcripts in the dark. This may indicate that pea chloroplast development arrests at an earlier stage in the absence of light than does the development of plastids in many other plants. Short term exposure of pea seedlings to light (after 6 days of darkness) favors the production of high molecular weight chloroplast transcripts, some of which encode several genes. Continuous growth in intense white light results in the absence of many of these large molecular weight RNAs and the accumulation of smaller transcripts (see Fig. 2A, C, D, and E for examples).

It is possible that transcription initiation or termination sites could change during light stimulated development, or that general degradation of RNA could increase with time after exposure to light. Either of these hypotheses could explain the shift in the size distribution of RNAs from large to small, but both have the disadvantage of added complexity. Switching transcription initiation sites requires invoking as yet unknown sequence specific factors which are sensitive to light or developmental stage. As noted above, general RNA degradation does not easily explain the accumulation of discrete transcripts.

Another possibility is that the decrease in the abundance of large chloroplast RNAs in plants grown in continuous light relative to plants that have been exposed to light only briefly might be explained by a regulatory

mechanism in chloroplasts which acts at the level of RNA processing. However, the shift in the size distribution of RNAs does not seem to be gene or gene cluster specific. There is no indication that this is a mechanism for controlling the relative stoichiometries of the products of one gene cluster versus another.

A simpler view is that light initiates a developmental process in peas in which the transcription rates of many genes change roughly in parallel, as seen in spinach (Deng and Gruissem 1987) and barley (Mullet and Klein 1987). In the barley system, the transcription rate of all chloroplast genes tested generally decreased with seedling age (Mullet and Klein 1987). If for a short period of time after the onset of illumination, pea plastids rapidly synthesize the RNA required for the transition from etioplast to chloroplast, then one would expect an initial increase in the abundance of primary, polycistronic transcripts. If the transcription rate then decreases as the chloroplast matures while the limiting rate constant for RNA processing remains roughly the same, the concentration of the primary transcripts should decrease and mature mRNAs should accumulate, consistent with our observations. Such a scheme requires no gene-specific regulation of either transcription rates or RNA processing.

References

Allmen von J-M, Stutz E (1987) Nucleic Acid Res 15:2387

Alt J, Herrmann RG (1984) Curr Genet 8:551–557

Alt J, Morris J, Westhoff P, Herrmann RG (1984) Curr Genet 8:597–606

Barkan A, Miles D, Taylor WC (1986) EMBO J 5:1421–1427

Berends T, Kubiced Q, Mullet J (1986) Plant Mol Biol 6:125–134

Bird CR, Koller B, Auffret AD, Huttly AK, Howe CJ, Dyer TA, Gray JC (1985) EMBO J 4:1381–1388

Bookjans G, Stummann BM, Rasmussen OF, Henningsen KW (1986) Plant Mol Biol 6:359–366

Boyer SK, Mullet JE (1986) Plant Mol Biol 6:229–243

Christianson T, Rabinowitz M (1983) J Biol Chem 258:14025–14033

Cozens AL, Walker JE (1986) Biochem J 236:453–460

Cozens AL, Walker JE, Phillips AL, Huttly AK, Gray JC (1986) EMBO J 5:217–222

Deng X-W, Gruissem W (1987) Cell 49:379–387

Deno H, Sugiura M (1984) Proc Natl Acad Sci USA 81:405–408

Erion JL (1985) Plant Mol Biol 4:169–179

Fish LE, Kuck U, Bogorad L (1985) J Biol Chem 260:1413–1421

Fromm H, Edelman M, Koller B, Goloubinoff P, Galun E (1986) Nucleic Acids Res 14:883–898

Gorton HL, Briggs WR (1980) Plant Physiol 66:1024–1026

Gruissem W, Zurawski G (1985) EMBO J 4:3375–3383

Heinemeyer W, Alt J, Herrmann RG (1984) Curr Genet 8:543–549

Herrmann RG, Alt J, Schiller B, Widger WR, Cramer WA (1984) FEBS Lett 176:239–244

Herrmann RG, Westhoff P, Alt J, Tittgen J, Nelson N (1985) In: Vloten-Doting L, Groot GSP, Hall TC (eds) Molecular form and function of the plant genome. Plenum Press, New York, London, pp 233–256

Howe CJ, Auffret AD, Doherty A, Bowman CM, Dyer TA, Gray JC (1982) Proc Natl Acad Sci USA 79:6903–6907

Huttly AK, Gray JC (1984) Mol Gen Genet 194:402–409

Jansen RK, Palmer JD (1987) Curr Genet 11:553–564

Kaufman LS, Thompson WF, Briggs WR (1984) Science 226:1447–1449

Koch W, Edwards K, Kossel H (1981) Cell 25:203–213

Koller B, Delius H (1984) Cell 36:613–622

Krebbers ET, Larrinua IM, McIntosh L, Bogorad L (1982) Nucleic Acids Res 10:4985–5002

Lehmbeck J, Rasmussen OF, Bookjans GB, Jepsen BR, Stumann BM, Henningsen KW (1986) Plant Mol Biol 7:3–10

Levens D, Ticho B, Ackerman E, Rabinowitz M (1980) J Biol Chem 256:5226–5232

Link G (1984a) Plant Mol Biol 3:243–248

Link G (1984b) EMBO J 3:1697–1704

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 382–388

Morimoto R, Locker J, Synenki RM, Rabinowitz M (1979) J Biol Chem 254:12461–12370

Morris J, Herrmann RG (1984a) Nucleic Acids Res 12:2837–2851

Morris J, Herrmann RG (1984b) Nucleic Acids Res 12:6547–6558

Mullet JE, Klein RR (1987) EMBO J 6:1517–1579

Mullet JE, Orozco EM Jr, Chua NH (1985) Plant Mol Biol 4:39–54

Neuhaus H, Link G (1987) Curr Genet 11:251–257

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) Nature 322:572–574

Oishi KK, Shapiro DR, Tewari KK (1984) Mol Cell Biol 4:2556–2563

Palmer JD, Thompson WF (1981a) Gene 15:21–26

Palmer JD, Thompson WF (1981b) Proc Natl Acad Sci USA 78:5533–5537

Palmer JD, Jorgensen RA, Thompson WF (1985) Genetics 109:195–213

Palmer JD, Osorio B, Thompson WF (1988) Curr Genet 14:65–74

Phillips AL, Gray JC (1984) Mol Gen Genet 194:477–484

Posno M, van Noort M, Debise R, Groot GSP (1984) Curr Genet 8:147–154

Rasmussen OF, Bookjans G, Stummann BM, Henningsen KW (1984a) Plant Mol Biol 3:191–199

Rasmussen OF, Stummann BM, Henningsen KW (1984b) Nucleic Acids Res 12:9143–9153

Rodermel SR, Bogorad L (1985) J Cell Biol 100:463–476

Shapiro DR, Tewari KK (1986) Plant Mol Biol 6:1–12

Sijben-Muller G, Hallick RB, Alt J, Westhoff P, Herrmann RG (1986) Nucleic Acids Res 14:1029–1044

Shinozaki K, Deno H, Kato A, Sugiura M (1983) Gene 24:147–155

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yana-guchi-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) EMBO J 5:2043–2049

Smith SM, Ellis JR (1981) J Mol Appl Genet 1:127–137

Smith AG, Gray JC (1984) Mol Gen Genet 194:471–476

Stern DB, Gruissem W (1988) Cell (in press)

Tewari KK (1980) In: Davies JW, Hall T (eds) Nucleic acids in plants. CRC Press, Cleveland, pp 41–108

Thompson WF, Everett M, Polans NO, Jorgensen RA, Palmer JD (1983) Planta 158:487–500

Torazawa K, Hayashida N, Obokata J, Shinozaki K, Sugiura M (1986) Nucleic Acids Res 14:3143

Westhoff P (1985) Mol Genet 201:115–123

Westhoff P, Alt J, Herrmann RG (1983) EMBO J 2:2229–2237

Westhoff P, Alt J, Widger WR, Cramer WA, Herrmann RG (1985a) Plant Mol Biol 4:103–110

Westhoff P, Alt J, Nelson N, Herrmann RG (1985b) Mol Gen Genet 199:290–299

Westhoff P, Farchaus JW, Herrmann RG (1986) Curr Genet 11:165–169

Whitfeld PR, Bottomley W (1983) Annu Rev Plant Physiol 34:279–310

Whitfeld P, Atchison BA, Bottomley W, Leaver CJ (1976) In: Bucher Th (ed) Genetics and biogenesis of chloroplasts and mitochondria. Elsevier/North Holland, Amsterdam, pp 361–368

Willey DL, Auffret AD, Gray JC (1984) Cell 36:555–562

Zhu YS, Kung SD, Bogorad L (1985) Plant Physiol 79:371–376

Zurawski G, Bottomley W, Whitfeld PR (1982) Proc Natl Acad Sci USA 79:6260–6264

Zurawski G, Bottomley W, Whitfeld PR (1984) Nucleic Acids Res 12:6547–6558

Zurawski G, Bottomley W, Whitfeld PR (1986a) Nucleic Acids Res 14:3974

Zurawski G, Whitfeld PR, Bottomley W (1986b) Nucleic Acids 14:3975