Theor Appl Genet (1992) 85:229–238

THEORETICAL AND APPLIED GENETICS

Tobacco nuclear DNA contains long tracts of homology to chloroplast DNA

M. A. Ayliffe and J. N. Timmis

Department of Genetics, University of Adelaide, GPO Box 498, Adelaide, South Australia 5001

Received February 15, 1992; Accepted April 7, 1992
Communicated by R. Hagemann

Summary. Long tracts of DNA with high sequence homology to chloroplast DNA were isolated from nuclear genomic libraries of *Nicotiana tabacum*. One lambda EMBL4 clone was characterised in detail and assigned to nuclear DNA. The majority of the 15.5-kb sequence is greater than 99% homologous with its chloroplast DNA counterpart, but a single base deletion causes premature termination of the reading frame of the *psa*A gene. One region of the clone contains a concentration of deleted regions, and these were used to identify and quantify the sequence in native nuclear DNA by polymerase chain reaction (PCR) methods. An estimated 15 copies of this specific region are present in a 1c tobacco nucleus.

Key words: *Nicotiana tabacum* DNA transfer – Plastid DNA – Nuclear DNA – Promiscuous DNA

Introduction

Evidence from a variety of experimental sources has confirmed the hypothesis (Margulis 1970) that organelles within eukaryotic cells are descendants of engulfed prokaryotes. This evidence has come from comparative anatomy, comparative biochemistry (Weeden 1981), the presence of organelle/cytoplasm isozyme pairs (Weeden 1981) and from genomic sequence comparisons (Pace et al. 1986). The evolutionary behaviour of the cohabiting genomes since the initiation of endosymbiosis, however, is poorly understood.

Full sequence data are available for three chloroplast genomes, from *Nicotiana tabacum* (Shinozaki et al. 1986), *Marchantia polymorpha* (Ohyama et al. 1986) and *Oryza sativa* (Hiratsuka et al. 1989). Many products of chloroplast (cp) DNA have been identified, and most of the proteins produced have some function in photosynthesis or in chloroplast transcription and replication. The size of the chloroplast genome is similar in all plants (about 150 kb) and is small compared with the genome of the simplest free-living prokaryote. One explanation for this is that the evolutionary history of the chloroplast precursor since its first initial endosymbiosis has involved many genes from the prokaryote being transferred to the nuclear genome where they have become active (Weeden 1981). Alternatively, the nuclear genes which control chloroplast development were at first duplicated in the cell as a result of the endosymbiosis and those in the ancestral chloroplast were lost.

Evidence favours the origin of the extant compartmental gene arrangements by large-scale transfer of DNA. Southern hybridisations in spinach (Timmis and Scott 1983; Scott and Timmis 1984), in several members of the Chenopodiaceae (Ayliffe et al. 1988) and in potato (du Jardin 1990) indicate widespread homology of highly methylated *Hpa*II-resistant nuclear DNA to chloroplast DNA probes. In spinach these *Hpa*II-resistant or unpredicted fragments have been ascribed to promiscuous DNAs (Ellis 1982) present in nuclear DNA (Timmis and Scott 1983; du Jardin 1990). This technique suggested that cpDNA integrants in spinach were numerous, included most of the chloroplast genome and were generally smaller than 2 kb (Timmis and Scott 1983). In other species some of the cpDNA integrants in nuclear (n) DNA appeared larger, with 13.5 kb suggested in potato (du Jardin 1990) and insertions of over 6 kb in two members of the Chenopodiaceae (Ayliffe et al. 1988). These integrant sizes are based on the apparent maintenance of a restriction pattern identical with cpDNA in *Hpa*II-resistant nuclear DNA.

Correspondence to: J. N. Timmis

Evidence of cpDNA incorporation into nDNA independent of molecular hybridisation came unexpectedly from the analysis of a nuclear gene from cultivated tomato. Sequence comparisons revealed that two short tracts in the third intron of *Cab-7*, a nuclear gene coding for a plastid protein, had high homology to the chloroplast gene *psb*G (Pichersky and Tanksley 1988). A subsequent systematic search of two lambda libraries, covering more than 95% of the tomato genome, for homology to probes covering 58 kb of the plastid genome yielded only two short integrants with flanking nuclear sequences (Pichersky et al. 1991). A detailed characterisation of one genomic clone from the spinach nucleus (Cheung and Scott 1989) revealed the presence of nuclear DNA linked to three abutting chloroplast-related tracts that, surprisingly, were not contiguous in the chloroplast genome. The sequence of a total of 3.4 kb of this spinach nuclear clone was more than 99% homologous with that of the corresponding region of *bona fide* chloroplast DNA.

Promiscuous cpDNA homology is also found in mitochondrial genomes of plants where the phenomenon was first discovered (Stern and Lonsdale 1982). Many sequences have been identified that are common to all three genetic compartments of the plant cell (Pichersky et al. 1991; du Jardin 1990; Timmis and Scott 1983). Genomic sequences of mitochondrial origin have been detected in a wide variety of plants, animals and fungi (Schuster and Brennicke 1988), but as yet no DNA from other genomes has been identified in cpDNA.

Nicotiana tabacum was chosen to characterise nuclear sequences with homology to cpDNA because of the availability of the complete nucleotide sequence of the chloroplast genome (Shinozaki et al. 1986).

Materials and methods

Isolation of plant nucleic acids

Total DNA. Total DNA was isolated from young leaf tissue obtained from glasshouse-grown *Nicotiana tabacum* var 'W38' as described by Scott and Possingham (1980).

Nuclear DNA. Tobacco nuclei were isolated from young leaf tissue by the method of van Loon et al. (1975). The nuclei were then lysed in 2 x SSC, 50 m*M* EDTA, 1% sarkosyl and the lysate phenol/chloroform/isoamyl alcohol extracted, followed by precipitation of nucleic acids with ethanol.

Mitochondrial DNA. Fresh tobacco leaf tissue was chilled on ice prior to homogenisation in 0.3 *M* sucrose, 25 m*M* tetrasodium pyrophosphate, 20 m*M* EDTA, 1% polyvinylpyrrolidone and 1% BSA, pH7.5. The crude lysate was filtered through miracloth and centrifuged at 700 *g* for 5 min at 4 °C. The supernatant was retained and centrifuged at 11,000 *g* for 20 min at 4 °C to pellet mitochondria, and the pellet resuspended in homogenisation buffer and respun. The crude mitochondrial pellet was then spun through a Percoll gradient, and the purified mitochondria treated with DNase grade 1 (Boehringer) to remove contaminating plastid and nuclear DNA. The mitochondria were finally lysed in 2 × SSC, 50 m*M* EDTA and 1% sarkosyl. The lysate was phenol/chloroform/isoamyl alcohol extracted, and the mitochondrial DNA precipitated with 2.5 volumes of ethanol.

Construction and screening of a tobacco genomic library

Tobacco total DNA was partially restricted with *Sau3*A and 15- to 20-kb DNA fragments isolated by sucrose gradient centrifugation. These fragments were further restricted with *Hpa*II, which results in extensive cleavage of the unmethylated cp and mtDNA fragments, while the heavily methylated nuclear genomic fragments remain essentially intact. The remaining 15- to 20-kb fragments were purified by a second sucrose gradient, ligated into the *Bam*HI site of lambda EMBL4 and *in vitro* packaged using a Packagene (Promega) packaging system. The resultant genomic library of 1.0×10^6 pfu was plated onto agarose plates using *E. coli* NW2 (*mcrA*$^-$, *mcrB*$^-$) as a phage host. Duplicate plaque lifts were performed as described by Sambrook et al. (1989) onto nitrocellulose filters and the library screened with three radioactively labelled cpDNA probes: the spinach *rbc*L gene (Zurawski et al. 1981), the spinach cpDNA 7.7-kb *Pst*I fragment (Palmer and Thompson 1981), which contains the *psa*A gene (Westhoff et al. 1983) and a portion of *psb*C (Holschuh et al. 1984) and the tobacco cpDNA *Bam*HI fragment 8, which contains *trn*I, *rpl*23, *rpl*2, *trn*H, *psb*A and part of *trn*K (Shinozaki et al. 1986). These probes are referred to as *prbc*L, *pPst*7.7 and *pBam*8, respectively, in the following text. Hybridisation with *pPst*7.7 was at normal stringency (Tm −25 °C), and other probes were hybridised at reduced stringency (Tm −30 °C).

Southern blotting and hybridisation

DNA samples were restricted with restriction endonucleases under conditions recommended by the manufacturer (Boehringer) and electrophoresed on 0.8% agarose (Sigma) gels. Following electrophoresis the DNA was transferred to a nitrocellulose membrane (Schleicher and Schuell) by the method of Southern et al. (1975). Nitrocellulose filters were hybridised as previously described (Scott and Timmis 1984) with ^{32}P-labelled *pPst*7.7 and p*rbc*L. Probes were labelled with alpha-labelled [^{32}P]-adenosine triphosphate using a Bresatec oligo-labelling kit. Autoradiography was carried out at −80 °C with an intensifying screen, except when autoradiographs were used for quantification by laser densitometry.

Sequencing

Double-stranded DNA templates were sequenced by the dideoxy chain termination method of Sanger et al. (1977) using a Sequenase Version 2.0 kit (United States Biochemical Corp). DNA sequencing reactions were electrophoresed in a 5% acrylamide, 7 *M* urea and 1 × TBE denaturing polyacrylamide gel using a BRL sequencing gel electrophoresis system (BRL). Sequencing utilised a combination of overlapping subclones and oligonucleotide primers. Sequence comparison to the published tobacco chloroplast genome sequence (Shinozaki et al. 1986) was done using the FASTA program of the Genetic Computer Group Sequence Analysis Software Package (Devereux et al. 1984).

Polymerase chain reaction (PCR)

PCR reactions were as described by Sambrook et al. (1989). One microgram of genomic template DNA underwent PCR in *Taq* DNA polymerase buffer, 1.5 m*M* $MgCl_2$, 1 pmol of each primer, 1.25 m*M* of each deoxynucleoside triphosphate and 2.5 units of *Taq* DNA polymerase (Bresatec). Initial template denaturation was at 94 °C for 5 min followed by primer annealing at

50 °C for 2 min and polymerisation at 72 °C for 3 min. A further 30 similar PCR cycles were carried out except that the denaturation step was reduced to 1 min.

Results

Figure 1 shows that the 1.8-kb *Eco*RI fragment (Zurawski et al. 1981) from spinach cpDNA (p*rbc*L) which codes for the large subunit of ribulose bisphosphate carboxylase/oxygenase hybridises strongly to a 4.4-kb *Eco*RI fragment in tobacco total DNA (Fig. 1 a, track 7), which is consistent with its arrangement in tobacco cpDNA (Shinozaki et al. 1986). Further digestion of the samples with *Hpa*II yields visible bands of 966, 424 and 1,700 bp (Fig. 1 a, track 2), which hybridise p*rbc*L as predicted from the cpDNA sequence (Shinozaki et al. 1986 and Fig. 2 b). In addition to these major fragments there is a proportion of the 4.4-kb band that remains undigested by *Hpa*II together with a range of additional hybridizing bands that are not predicted from the cpDNA sequence.

Hybridisation to the surviving 4.4-kb band is intense relative to similar observations in other species (Timmis and Scott 1983; Ayliffe et al. 1988) and may be explained either by incomplete *Hpa*II digestion or by the occurrence of large tracts of cpDNA homology in methylated DNA with conservation of flanking restriction sites or co-migration of a repeated nuclear or mitochondrial DNA sequence with cpDNA homology. Figure 1 a, tracks 8 and 9, show that a *Kpn*I site is conserved in the *Hpa*II-resistant, 4.4-kb *Eco*RI fragment, suggesting that it is entirely homologous to cpDNA. These homologies may originate in nuclear DNA or in a subpopulation of plastids containing DNA with extensive cytosine methylation. Comparisons of the intensity of this 4.4-kb band with a dilution series of total DNA (Fig. 1 a, tracks 3–7) suggests that, in tobacco, this methylated DNA contributes 1.3% of the hybridisation to this probe.

Digests using several other restriction enzymes in combination with *Hpa*II (Fig. 1 a, tracks 10–15) indicate that a proportion of the *bona fide* chloroplast DNA band produced by *Xho*I (Fig. 1 a, track 11), *Pvu*II (Fig. 1 a, track 13) and *Sal*I (Fig. 1 a, track 15) remains after *Hpa*II digestion (Fig. 1 a, tracks 10, 12 and 14, respectively). As the size of the resistant fragment increases, its intensity of hybridisation is apparently reduced and the proportion of the signal appearing at higher molecular sizes is increased (compare Fig. 1 a, tracks 10 and 14).

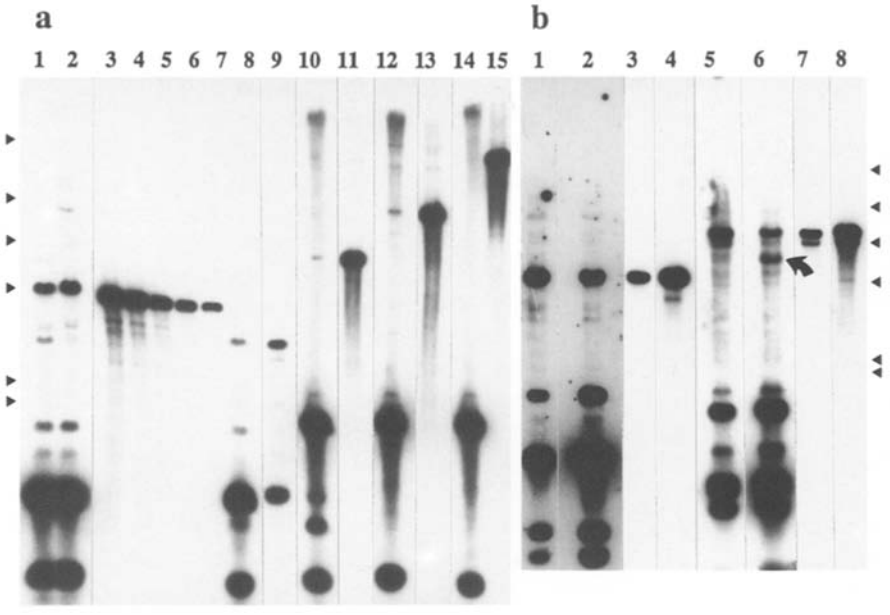

Fig. 1 a, b. Hybridisation of cpDNA probes to nuclear DNA. **a** Hybridisation of the spinach cp *rbc*L gene to tobacco total leaf DNA. Each track was loaded with 5 µg of DNA except for lanes 3–7 and 9, which were loaded with 0.5, 0.25, 0.125, 0.063, 0.031 and 0.5 µg of DNA, respectively. DNA samples were restricted with the following restriction endonucleases: *lane 1 Msp*I/*Eco*RI, *2 Hpa*II/*Eco*RI, *3–7 Eco*RI, *8 Hpa*II/*Eco*RI/*Kpn*I, *9 Eco*RI/*Kpn*I, *10 Hpa*II/*Xho*I, *11 Xho*I, *12 Hpa*II/*Pvu*II, *13 Pvu*II, *14 Hpa*II/*Sal*I, *15 Sal*I. *Markers* on the *left* represent molecular weights of 23, 9.6, 6.4, 4.4, 2.3 and 2.0 kb. **b** Hybridisation of cpDNA probes to nuclear and total leaf DNA preparations of tobacco. Hybridisation of the spinach plastid *rbc*L gene to tobacco nuclear DNA (*lanes 1* and *3*) and total DNA (*lanes 2* and *4*). *Lanes 1* and *2* were loaded with 5 µg of DNA that was restricted with *Hpa*II and *Eco*RI, while *lanes 3* and *4* contained 1 µg of DNA restricted with *Eco*RI. *Tracks 5–8* rehybridisation of the nitrocellulose filter used in tracks 1–4 with the spinach plastid 7.7-kb *Pst*I DNA fragment. The *arrow* indicates a homology, putatively assigned to mtDNA. *Markers* on the *right* represent the same molecular weights as in Fig. 1 a

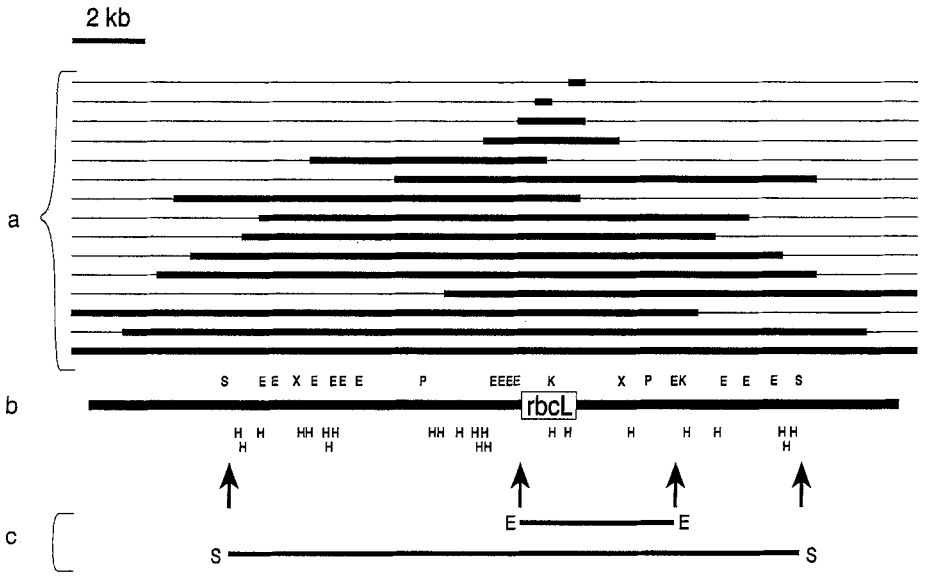

Fig. 2a–c. Schematic diagram of hypothetical tobacco plastid DNA fragments inserted into various regions of the nuclear genome. Inserts **a** contain at least a proportion of the *rbc*L gene. Promiscuous plastid sequences are represented by a *bold line* with flanking nuclear sequences represented by a *fine line*. The size and extent of homology of these sequences to the plastid genome can be determined by a direct comparison of **a** to the restriction map of the tobacco plastid *rbc*L gene and flanking regions **b**. In **b** letters *S*, *E*, *X*, *P*, *K* and *H* refer respectively to the position of *Sal*I, *Eco*RI, *Xho*I, *Pvu*II, *Kpn*I and *Hpa*II restriction sites in this region of the plastid genome. The position on this restriction map of the 4.4-kb *Eco*RI and 15-kb *Sal*I fragments discussed in relation to Fig. 1a are shown in **c**

These large *Hpa*II-resistant fragments contain several CCGG sequences that would be available for digestion in cpDNA. The number of *Hpa*II sites in cpDNA varies from 3 for the *Eco*RI fragment containing *rbc*L to 23 for the *Sal*I fragment (Fig. 2b). Partial digestion is therefore an unlikely explanation for the presence of fragments undigested by *Hpa*II in double digests, as the enzyme would be expected to cleave at least some of these numerous sites, producing a range of partial and complete digestion products. When *Msp*I, an isoschizomer of *Hpa*II that will not cut CCGG when the '5 C is methylated, replaced *Hpa*II, an apparently identical 4.4-kb *Eco*RI band survived together with a modified but similar range of additional fragments (Fig. 1a, track 1) suggesting methylation at both CXG and CG sites (Doerfler et al. 1990).

When the probe was removed from the filter that gave rise to Fig. 1, a second hybridisation with p*Pst*7.7 gave parallel results, but suggested that about 0.8% of the hybridisation to this chloroplast probe may be attributed to highly methylated DNA sequences with a very similar restriction map to the corresponding *bona fide* cpDNA (result not shown).

Partially purified tobacco nuclear DNA was prepared to determine whether the *Hpa*II-resistant fragments were increased relative to bands of known chloroplast origin when hybridised with p*Pst*7.7. Hybridisation to total tobacco leaf DNA yielded the expected homologies to *Eco*RI bands of 7.3 and 6.6 kb (Fig. 1b, track 8), most of which was further digested by *Hpa*II into the expected 5 cpDNA bands but also leaving about 0.8% of the *Eco*RI fragment undigested as previously (Fig. 1b, track 6). Track 6 contains more DNA than track 8 and a further series of methylated bands is observed. *Eco*RI restriction of nuclear DNA produces an identical hybridisation pattern (Fig. 1b, track 7), but the 7.3- and 6.6-kb bands remaining after *Hpa*II digestion are of greater intensity compared with the *Hpa*II restriction products, the majority of which must represent cpDNA contamination of this nuclear preparation (Fig. 1b, track 5). Other methylated fragments are also present in enriched quantities in the nuclear preparation. However, the 5.2-kb band in total leaf DNA (arrowed in Fig. 1b, track 6) is not as prominent in the nuclear DNA preparation and was tentatively ascribed to a promiscuous sequence in tobacco mtDNA. Parallel results were obtained with p*rbc*L (Fig. 1b, tracks 1–4) but with no evidence of the involvement of mtDNA.

One explanation for these observations is that there is a minor population of plastids which contain genomes that are highly methylated at *Hpa*II and *Msp*I sites. The specific and localised methylation of plastid genomes that has been reported (Ngernprasirtsiri et al. 1989) would not account for these observations. In addition, a subset of plastids with highly methylated genomes is unlikely to co-purify with tobacco nuclei and remain intact

after treatment with TritonX-100 (Fig. 1b). The most plausible explanation of these results is that the tobacco nuclear genome contains a number of regions that are complementary to cpDNA probes. The majority of such complementary tracts in tobacco appear to be large compared with those reported for spinach (Cheung and Scott 1989) or tomato (Picherski et al. 1991). The nucleotide sequence of these putative nuclear tracts must be very similar to that of the true chloroplast genome because of the frequent conservation of the restriction sites (Fig. 1).

Figure 2 shows a model that explains the above results. The section of the chloroplast genome in the vicinity of the *rbc*L gene is bounded by *Eco*RI sites 4.4 kb apart, and this fragment contains three *Hpa*II sites which would be digested in unmethylated DNA but remain uncleaved if they were modified. The 4.4-kb fragment therefore often remains in *Eco*RI + *Hpa*II-digested DNA because the majority of the sequences are sufficiently long and conserved to retain the *Eco*RI sites (Fig. 2c). When other restriction enzymes are used in a similar regime the proportion of the *Hpa*II-resistant DNA falls, presumably reflecting the greater rarity of longer regions of cpDNA in the nucleus (Fig. 2c). A 15.5-kb *Sal*I fragment of cpDNA hybridises p*rbc*L, and this complete fragment is much more rarely found in methylated DNA than the 4.4-kb *Eco*RI fragment (Fig. 2c and Fig. 1a, cf. tracks 2 and 14). However, differential methylation sensitivity of these six-base-cutting enzymes could also account for this observation.

A library of DNA resistant to digestion with *Hpa*II was constructed in lambda EMBL4 (Sambrook et al. 1989). Considerable difficulties were initially experienced in isolating cpDNA homologous clones from nuclear libraries, as has been experienced elsewhere (Cheung and Scott 1989). These problems were overcome when the *E. coli* host strain was free from modified cytosine restriction systems *mcr*A and *mcr*B (Raleigh et al. 1988). Approximately 200-fold more recombinant clones were obtained with the host strain NW2 (Woodcock et al. 1989), which lacks the McrA and McrB restriction systems compared with strains NM538 or NM539, which both contain the McrB system (Raleigh et al. 1988).

Approximately 300 clones with homology to p*rbc*L, p*Pst*7.7 and p*Bam*8 were selected from 1.0×10^6 lambda clones. A restriction map of one of these clones (3D), 13.8 kb in length, is shown in Fig. 3. Probe p*Pst*7.7 is homologous to DNA adjacent to the long arm of the vector and extends about 1.5 kb into the clone. The *Eco*RI map of this clone is very similar to that of tobacco cpDNA except that a 2.9-kb fragment contains an additional *Eco*RI site (boxed in Fig. 3) to give 1.6- and 1.3-kb products. The appropriate tobacco cpDNA clones (Shinozaki et al. 1986) hybridised to all of the lambda clone 3D *Eco*RI fragments in proportion to their size, (result not shown) suggesting that apart from the possibility of minor insertions of nuclear DNA this clone is entirely homologous to cpDNA. This clone therefore contains no borders with nDNA or mtDNA. Consequently, there are tracts of methylated DNA in tobacco which are at least 13.8 kb in size and which are entirely homologous to cpDNA.

To rule out promiscuous cpDNA in mtDNA as the origin of this homology, clone 3D DNA was restricted with *Eco*RI, the fragments Southern blotted and hybridised with labeled tobacco mtDNA. Only three of the ten *Eco*RI bands in the insert were homologous to mtDNA, which indicates that only part of the plastid sequence is present in mitochondria and precludes mtDNA as the origin of clone 3D.

To characterise the relationship between clone 3D and the complementary region of cpDNA four separate regions of the clone were sequenced in both directions

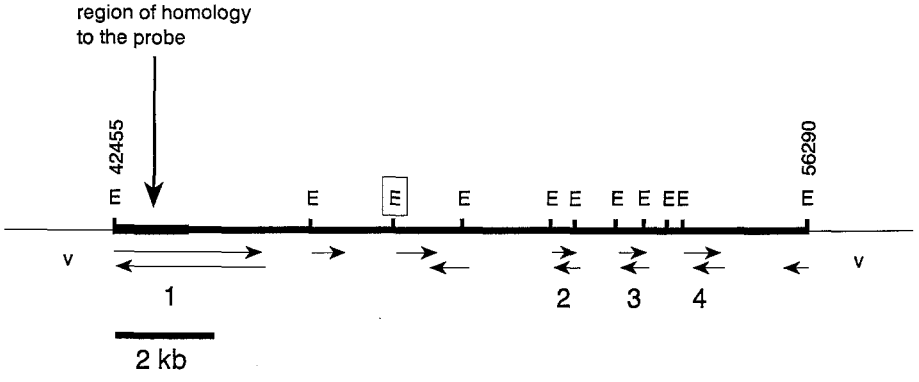

Fig. 3. Restriction map of lambda clone 3D. The tobacco DNA fragment inserted into lambda clone 3D is represented by a *bold line* with flanking vector (*v*) arms represented as a *fine line*. The region of homology of this clone to the spinach plastid 7.7-kb *Pst*I fragment is illustrated as a *black rectangle* and the position of *Eco*RI restriction sites on this clone are marked with an *E*. *Regions 1–4* represent portions of the lambda clone that have been sequenced in both directions as indicated by *arrows*. Other *arrows* indicate portions of the clone that were sequenced on one strand only. Clone 3D is homologous to the published tobacco plastid genome sequence (Shinozaki et al. 1986) from nucleotide 42,455–56,290

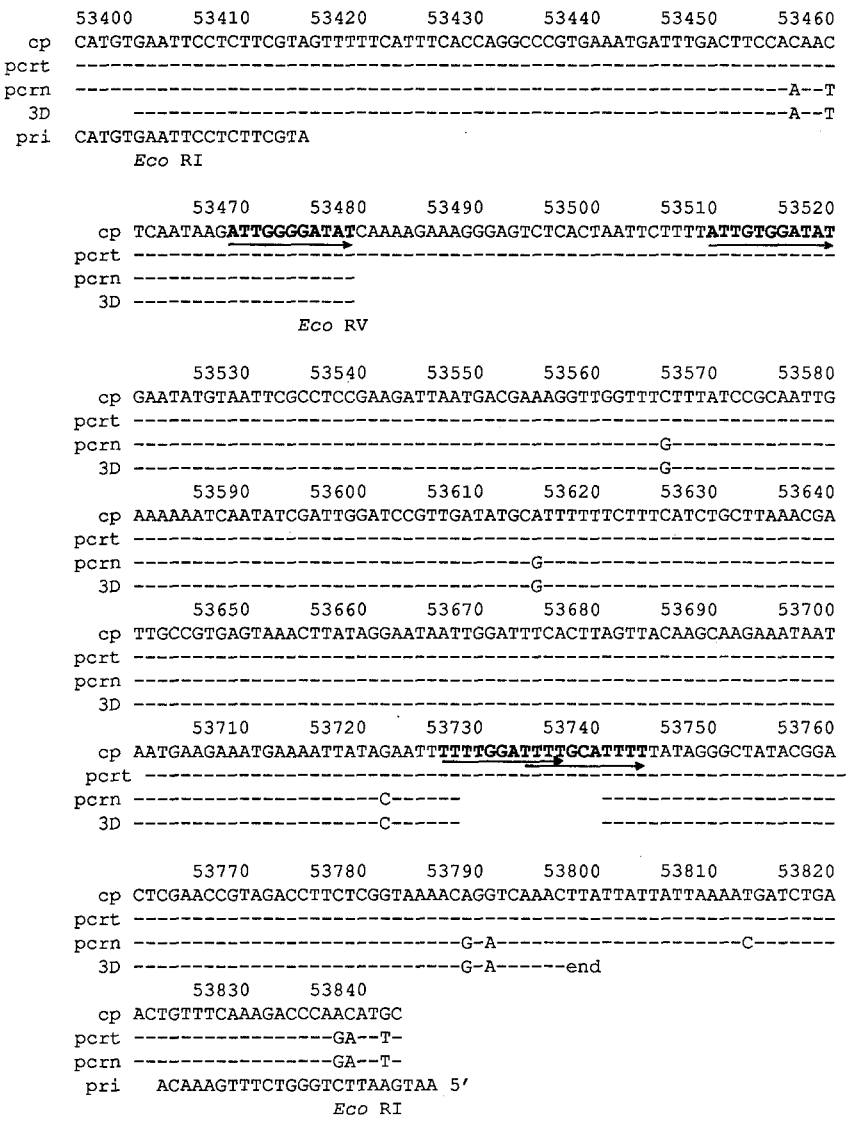

Fig. 4. Nucleotide sequence comparisons of homologous chloroplast and nuclear DNA. The partial nucleotide sequences of tobacco cpDNA (*cp*) (Shinozaki et al. 1986), the PCR product from tobacco total leaf DNA (*pcrt*), the PCR product from tobacco nuclear DNA restricted with *Eco*RV (*pcrn*) and region 4 of lambda clone 3D (*3D*) are shown. Nucleotides identical to the published tobacco plastid genome sequence are indicated with a *dashed line*. Deleted nucleotides are shown by a *blank space*. Direct repeats in cpDNA flanking the deletions in nuclear DNA are in *bold face* and marked with *arrows*. The positions and sequence of the PCR primers are shown (*pri*)

and three regions were sequenced on one strand only, as indicated in Fig. 3. All the data confirmed that clone 3D contains an insert which in its entirety has high homology to a contiguous cpDNA tract but which has several features that unequivocally distinguish it from true cpDNA. It contains the region between nucleotides 42,455 and 56,290 of cpDNA (Shinozaki et al. 1986) that encompasses nuclear equivalents of the chloroplast genes for *psa*A (part), *trn*S, *rps*4, *trn*T, *trn*L *trn*F, *psb*G, *ndh*C, *trn*V, *trn*M and part of *atp*E in addition to five unidentified open reading frames. A total of 3 kb of the largest 3.8-kb *Eco*RI fragment (region 1, Fig. 3) contains striking similarity to the chloroplast genome. However, a deletion of a single nucleotide at position 43,152 (Shinozaki et al. 1986) destroys the reading frame of the *psa*A gene, resulting in the occurrence of a stop codon 26 bases downstream and precluding the production of a functional protein. When all of the sequences obtained from regions 1, 2 and 3 (Fig. 3), which totals 3,835 bp, are compared with the cpDNA sequence of Shinozaki et al. (1986), only minor differences are observed. These include 23 base substitutions, three deletions and four additions indicating 99% similarity between the DNAs located in the two different genetic compartments.

The more conspicuous sequence differences occur in region 4 (Fig. 3) towards the 3′ end of the clone and include deletions of 12 and 41 base pairs (Fig. 4). Each of these deletions is flanked by an 11-bp direct repeat in the chloroplast DNA (Fig. 4). Homologous pairing and recombination between similar direct repeats has been postulated by Moon et al. (1988) to account for a large deletion in a promiscuous sequence in the mtDNA of rice. These rearrangements may have occurred as an artifact of cloning, and this possibility was tested by searching for the deletions directly in genomic DNA.

To determine whether a sequence identical to clone 3D is present in tobacco nuclear DNA, PCR primers containing *Eco*RI restriction sites were prepared that were designed to amplify a 390-bp region spanning the 12- and 41-bp deletions (Fig. 4). The 41-bp deletion which is absent in putative nuclear DNA is present in cpDNA, where it contains an *Eco*RV restriction site (Fig. 4). A partially purified nuclear DNA preparation was therefore restricted with *Eco*RV to cleave any contaminating cpDNA present in the sample. Exponential amplification of the 390-bp region was expected to be possible only in sequences lacking the *Eco*RV site. If *Rco*RV did not digest all of the cpDNA molecules a 443-bp product would also be exponentially amplified. Gel analysis indicated a PCR product from *Eco*RV-digested nuclear DNA of the expected length consistent with the presence of a tobacco genomic sequence identical to clone 3D (Fig. 5a). This PCR product was cloned (Fig. 5b, track 1), and its sequence was identical to that of 3D (Fig. 4). When total leaf DNA, undigested by *Eco*RV, was amplified using the same primers a 443-bp product was obtained as predicted from the tobacco cpDNA sequence (Shinozaki et al. 1986). This product was cloned (Fig. 5b, track 10), and three inserts were shown to have the published cpDNA sequence (Shinozaki et al. 1986) (Fig. 4). When the same PCR protocol was applied to purified mtDNA no products were synthesised, indicating that the region between the primers is not to be found in the mitochondrial genome.

The Southern hybridisations (Fig. 1a) suggest that the highly methylated nuclear sequences homologous to cpDNA probes are repetitious. It is difficult for many technical reasons to quantify these sequences in the nucleus, and only very approximate estimates have been made (Scott and Timmis 1984; du Jardin 1990). However, the degree of repetition of the specific genomic region containing the 41-bp deletion may be accurately estimated by comparative PCR quantification (Sambrook et al. 1989) as the nuclear PCR product can be unambiguously distinguished from the plastid counterpart. To provide a standard for quantification, a 100-bp *Hae*III fragment from pUC19 was inserted into one of the cloned 443-bp PCR products from cpDNA amplification. Relative to each tobacco genome, a range of copy numbers of this construct was added to *Eco*RV-digested nuclear DNA, and the samples amplified using the same primers as before. Gel analysis of the products (Fig. 5b) indicated between 10 and 20 copies of the genomic sequence containing the deletions were present in each tobacco nuclear genome.

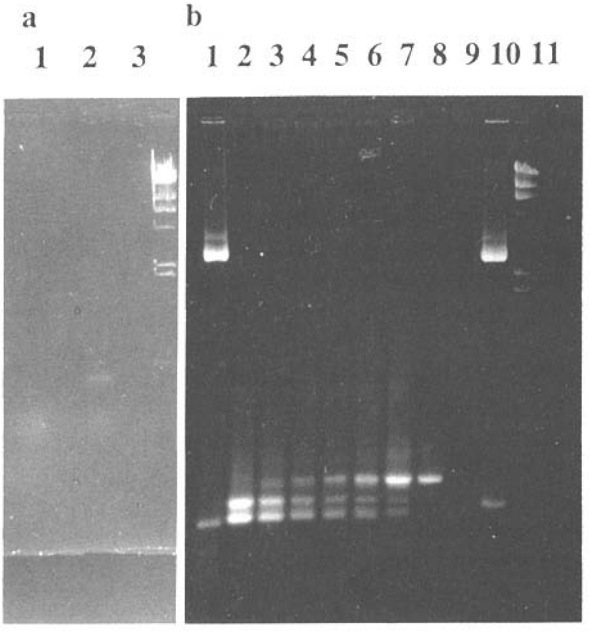

Fig. 5a, b. Detection and quantification of nuclear genomic sequences with homology to cpDNA using the polymerase chain reaction. **a** Agarose gel showing the PCR product obtained from 1 µg of tobacco nuclear DNA restricted with *Eco*RV (*lane 2*). *Lane 1* shows that no PCR product was obtained from a control reaction containing no DNA. *Lane 3* is 0.5 µg of lambda DNA restricted with *Hin*dIII. **b** A 2% agarose gel showing PCR products used to quantify the number of tobacco nuclear DNA sequences containing 12- and 41-bp deletions compared with plastid DNA. *Lanes 2–7* are PCR products obtained from 1 µg of nuclear DNA restricted with *Eco*RV and having 0, 1, 5, 10, 20, and 100 copies, respectively, of a plasmid construct added per tobacco 1 c nuclear genome. *Lane 8* is the product obtained from a PCR reaction containing only construct plasmid, and *lane 9* is a control showing no PCR products from a reaction with neither plasmid nor nuclear DNA added. *Lanes 1* and *10* are PCR products that have been cloned into pUC19 and sequenced, with the 390-bp fragment in lane 1 being identical in sequence to region 4 of lambda clone 3D and the 443-bp fragment in lane 10 being identical to the published tobacco plastid genome from nucleotide 53,401–53,841. *Lane 11* contains *Hin*dIII-restricted lambda DNA

Expansion within this range of added construct with [^{32}P]dATP supplemented during the final two rounds of PCR followed by resolution of the products on agarose and denaturing polyacrylamide gels enabled estimates to be made that were based upon both ethidium bromide staining and densitometry of autoradiographs. Taking into account the relative size and adenine composition of the PCR products we estimated this region of the 3D

clone to be present 15 ± 5 times in each tobacco genome. An initial estimate of the number of tobacco genomes added to the PCR reactions, based on microdensitometric measurements of Feulgen-stained nuclei [3.76×10^6 kb per 1 c nucleus (Bennett and Smith 1976)] was of 128,000. The value used to calculate copy numbers was based on dot blots of the DNA samples probed with both of the two different cloned chitinase genes from tobacco (Neale et al. 1990). This method gave an independent estimate of 100,000 genomes loaded in each PCR reaction, giving an estimate of 4.81×10^6 kb of DNA per 1 c tobacco nucleus. *N. tabacum* is an allotetraploid with two diploid ancestors, *N. sylvestris* and *N. tomentosiformis*, each contributing a different chitinase gene. The genome size here is calculated as half the telophase nuclear DNA amount (Bennett and Smith 1976).

Similar PCR quantification of the amount of cpDNA in total leaf DNA gave an estimate of $4{,}350 \pm 680$ copies of the chloroplast genome per 1 c nucleus. This indicates that total DNA contains about 14% cpDNA. Southern hybridisations (Fig. 1 a) indicated that 1.3% of the homology of the 4.4-kb *Eco*RI band to p*rbc*L was due to nuclear DNA, suggesting that there are 56 copies of this sequence in a 1 c nucleus. The corresponding estimate for the p*Pst*7.7 probe is 35 copies.

Discussion

These results indicate that the tobacco nuclear genome contains large sections of DNA that are closely related to its cpDNA genome. Clone 3D is one of six EMBL4 clones we have examined in some detail which vary between 14 and 18 kb (unpublished results). Only one of these clones contains a flanking sequence of nuclear DNA, the other five being highly, but not completely, homologous in their entirety to various regions of cpDNA. Experimental approaches to a reliable estimate of the maximum size of cpDNA integrants into the tobacco nuclear genome are technically difficult and have so far been unsuccessful in our laboratory. Limited calculations based on a random distribution of *Sau*3A cloning sites suggest that our finding only one nuclear border in a total of almost 100 kb of cloned DNA indicates the presence of nuclear integrants of cpDNA in excess of 155 kb. This suggests the possibility of entire chloroplast genomes in the nuclear DNA of tobacco.

The large size of cpDNA homology in nuclear clone 3D contrasts with other results where sequence analyses are available (Cheung and Scott 1989; Pichersky et al. 1988, 1991), although Southern hybridisations have suggested large integrants in some species (Ayliffe et al. 1988; du Jardin 1990). The most conspicuous contrast with our observations is in tomato (Pichersky et al. 1991) where, in a systematic search for plastid sequences integrated into the nuclear genome, only two very short integrants were isolated. As this search of two clone libraries together covered 95% of the tomato nuclear genome and 33% of the plastid genome, both the rarity and the small size of the integrants in tomato contrast with our findings in tobacco. The nuclear genome of tobacco [4.8×10^6 kb (Bennett and Smith 1976)] is much larger than that of tomato [7×10^5 [Tanksley and Pichersky 1988)], and the former species is a tetraploid whereas the latter is a diploid. Both species are members of the Solanaceae and are therefore closely related, but there are differences in the way these different genomes are organised and these are reflected in the contrasting results. Hybridisation with poly(dA) (Pichersky 1990) gives a very strong signal to Southern blots of tobacco DNA and almost none to tomato, an observation which was interpreted to indicate the presence of many more pseudogenes in tobacco. The gene duplications present in polyploid species may allow much more genomic experimentation in these species than in diploids where more intense selection for genomic conservation and economy may apply. High ploidy is common in primitive plants, possibly offering a environment conducive to rapid early evolution. Most of the possible functional chloroplast gene transfers to the nucleus were completed early in evolution.

The cloning strategy employed by Pichersky et al. (1991) would almost certainly have excluded our clone 3 D as being of chloroplast origin. It is also possible that the two tomato libraries screened were constructed using *mcr*A or B expressing *E. coli* hosts and may consequently have been deficient in clones of methylated DNA. Conversely, digestion with *Hpa*II, utilised during our procedure for constructing a cpDNA-free library, may have removed sequences containing active, possibly undermethylated, plant genes of the type reported by Pichersky et al. (1991). It must be noted, however, that we have analysed only 6 of about 300 clones hybridising cpDNA probes and have concentrated on clones with strong homology to cpDNA. It may be, as suggested by Pichersky et al. (1991), that only hypermethylated nuclear genomic regions contain relatively large plastid DNA integrants.

As clone 3D does not contain any flanking nuclear regions little information about possible mechanisms of transfer and integration can be derived from this sequence. It is noteworthy, however, that the sequence includes DNA that is thought not to be transcribed in the chloroplast, suggesting that an RNA intermediate (Nugent and Palmer 1991; Schuster and Brennicke 1987) was not involved in transposition in this case.

The absolute identification of these highly methylated sequences as being located in the tobacco nucleus, as opposed to location in an unidentified minor population of plastids with highly methylated DNA, awaits the analysis of YAC clones or *in situ* or hybridisation. This latter method is problematical given the unamenability of *N.*

tabacum chromosomes and the constant difficulty of removing contaminating cpDNA from chromosome spreads. Promiscuous plastid sequences have been identified in maize YAC libraries, but their absolute extent has not been determined (Gupta and Hoo 1991). One YAC clone examined in detail contained the entire maize chloroplast genome in addition to an extra 7.3 kb of DNA at the vector/clone junction. The origin of this extra DNA is unknown (Gupta and Hoo 1991).

It is generally assumed that stringent sequence conservation is the rule within the interrelated plastid types and that in an individual, all plastids are generally derived from a small number of maternal proplastids in the fertilised egg. The extensive alterations detected in clone 3D are most unlikely to have occurred during development given that they include a change that destroys an open reading frame in an essential chloroplast gene. In addition 2 of the other putative nuclear clones examined recently in our laboratory (unpublished results) contain sequences that are highly homologous to part of 4D, but they are different from cpDNA and different again from 3D in significant respects. This implies that any possible aberrant plastid subpopulation accounting for these results must possess both a highly heterogeneous and extensively methylated genome, and this is considered improbable.

The simplest explanation for our results is that large sections of the chloroplast genome have been transferred to the nucleus where gradual changes have occurred with time. Assuming a rate of sequence divergence similar to mammalian pseudogenes [5×10^{-9} changes/site/year (Li et al. 1985)] the 3D sequence is estimated to have transferred to nuclear DNA only 700,000 years ago. This sort of divergence which occurs with time in unexpressed portions of the genome is one possible source of evolutionary change. Such sequences may drift into genetic oblivion in all but the significant cases where, in a specific genomic environment, they achieve transcription, processing and expression (Scott et al. 1991). Genes transferred and activated in this way could then be lost from the chloroplast genome, a loss which may be paralleled by the rapid cpDNA deletions observed in parasitic plants (de Pamphilis and Palmer 1990). The gene activation events seem much less probable, but at least one clear case has been described. The large subunit protein *rpl22* of the chloroplast ribosome is present in cpDNA in all flowering plants except legumes where it is located in the nucleus (Gantt et al. 1991). Compared with homologous chloroplast genes this nuclear gene contains an additional exon presumed to code for a chloroplast-directed transit peptide separated from the unequivocally identified ancestral chloroplast gene by an intron (Gantt et al. 1991). The genomic mechanisms responsible for this sort of gene rearrangement in the nucleus are unknown but may account for the large size of eukaryotic genomes as this may allow successful gene manipulation, including exon shuffling, to occur (Gantt et al. 1991; Timmis and Scott 1984). Only a few other cases of differing genetic compartments for the same gene have been observed; one is the gene *tuf*A, which is in algal cpDNA but in the nucleus in land plants (Baldauf and Palmer 1990).

The copy number of 15 for the 3D-like sequence estimated to be present in tobacco nuclear DNA may represent the minimum degree of repetition of this region of the plastid genome. Related sequences not carrying the 12- and 41-bp deletion may also be present which are not true cpDNA but which would be cut by *Eco*RV and not amplified in our quantitative PCR assay. These 15 copies may have been individually transferred to the nuclear genome or be derived from a single sequence that has been amplified within the nucleus. In the latter case, which appears most likely given that each copy has an identical deletion, only 1 out of 15 clones will be present at or near the site of the initial transposition event. Care should therefore be exercised when speculating on the mechanisms of transfer of plastid sequences to the nucleus by analysis of cpDNA/nDNA junctions because many of these may represent secondary rather than primary transpositions.

Acknowledgements. We are grateful to Dr. Nigel Scott for discussions throughout the course of these experiments. We thank Ian Menz for technical assistance, Dr. A. D. Neale for providing the *N. tabacum* chitinase clones, Dr. David Woodcock for providing *E. coli* strain NW2, Dr. K. Shinozaki for the *N. tabacum* cpDNA clones. We thank The Cotton Seed Distributors Pty Ltd, Wee Waa, New South Wales, Australia for a Post-Graduate Research Award to MAA.

References

Ayliffe MA, Timmis JN, Scott NS (1988) Homologies to chloroplast DNA in the nuclear DNA of a number of Chenopod species. Theor Appl Genet 75:282–285

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

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Philos Trans R Soc London Ser B 274:227–274

Cheung WY, Scott NS (1989) A contiguous sequence in spinach nuclear DNA is homologous to three separated sequences in chloroplast DNA. Theor Appl Genet 77:625–633

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

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Doerfler W, Toth M, Kochanek S, Achten S, Freisem-Rabien U, Behb-Krappa A, Orend G (1990) Eukaryotic DNA methylation: facts and problems. FEBS Lett 268:329–333

du Jardin P (1990) Homologies to plastid DNA in nuclear and mitochondrial genomes of potato. Theor Appl Genet 79:807–812

Ellis J (1982) Promiscuous DNA-chloroplast genes inside plant mitochondria. Nature 299:678–679

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

Gupta M, Hoo B (1991) Entire maize genome is stably maintained in a yeast artificial chromosome. Plant Mol Biol 17:361–369

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

Holschuh K, Bottomley W, Whitfeld PR (1984) Structure of the spinach chloroplast genes for the D2 and 44 kd reaction-centre proteins of photosystem II and for $tRNA^{Ser}$ (UGA). Nucleic Acids Res 12:8819–8834

Li W-H, Luo C-C, Wu C-I (1985) Evolution of DNA sequences. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York

Moon E, Kao T-H, Wu R (1988) Rice mitochondrial genome contains a rearranged chloroplast gene cluster. Mol Gen Genet 213:247–253

Margulis L (1970) Origin of eukaryotic cells. Yale University Press, New Haven

Neale AD, Wahleithner JA, Lund M, Bonnett HT, Kelly A, Ry Meeks-Wagner D, Peacock WJ, Dennis ES (1990) Chitinase, beta-1,3-glucanase, osmotin, and extensin are expressed in tobacco explants during flower formation. Plant Cell 2:673–684

Ngernprasirtsiri J, Chollet R, Kobayashi H, Sugiyama T, Akazawa T (1989) DNA methylation and the differential expression of C_4 photosynthesis genes in mesophyll and bundle sheath cells of greening maize leaves. J Biol Chem 264:8241–8248

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

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

Pace NR, Olsen GJ, Woese CR (1986) Ribosomal RNA phylogeny and the primary lines of evolutionary descent. Cell 45:325–326

Palmer JD, Thompson WF (1981) Clone banks of the mung bean, pea and spinach chloroplast genomes. Gene 15:21–26

Pichersky E (1990) Nomad DNA – a model for movement and duplication of DNA sequences in plant genomes. Plant Mol Biol 15:437–448

Pichersky E, Tanksley SD (1988) Chloroplast DNA sequences integrated into an intron of a tomato nuclear gene. Mol Gen Genet 215:65–68

Pichersky E, Logsdon JM, McGrath JM, Stasys RA (1991) Fragments of plastid DNA in the nuclear genome of tomato: prevalence, chromosomal location, and possible mechanism of integration. Mol Gen Genet 225:453–458

Raleigh EA, Murray NE, Revel H, Blumenthal RM, Westaway D, Reith AD, Rigby PWJ, Elhai J, Hanahan D (1988) McrA and McrB restriction phenotypes of some *E. coli* strains and implications for gene cloning. Nucleic Acids Res 16:1563–1575

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain termination inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Scott NS, Possingham JV (1980) Chloroplast DNA in expanding spinach leaves. J Exp Bot 123:1081–1092

Scott NS, Timmis JN (1984) Homologies between spinach nuclear and plastid DNA. Theor Appl Genet 67:279–288

Scott NS, Timmis JN, Ayliffe M (1991) Chloroplast genes in higher plants. Aust J Plant Physiol 18:533–541

Schuster W, Brennicke A (1988) Interorganellar sequence transfer: plant mitochondrial DNA is nuclear, is plastid, is mitochondrial. Plant Sci 54:1–10

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

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Stern DB, Lonsdale DM (1982) Mitochondrial and chloroplast genomes of maize have a 12-kb DNA sequence in common. Nature 299:698–702

Tanksley SD, Pichersky E (1988) Organisation and evolution of sequences in the plant nuclear genome. In: Gottleib LD, Jain S (eds) Plant evolutionary biology. Chapman and Hill, London, pp 55–83

Timmis JN, Scott NS (1983) Spinach nuclear and chloroplast DNAs have homologous sequences. Nature 305:65–67

Timmis JN, Scott NS (1984) Promiscuous DNA: sequence homologies between DNA of separate organelles. Trends Biochem Sci 9:271–273

van Loon LC, Trewavas A, Chapman KSR (1975) Phosphorylation of chromatin-associated proteins *Lemna* and *Hordeum*. Plant Physiol 55:288–292

Weeden NF (1981) Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. J Mol Evol 17:133–139

Westhoff P, Alt J, Nelson N, Bottomley W, Bunemann H, Herrman RG (1983) Genes and transcripts for the P_{700} chlorophyll a apoprotein and subunit 2 of photosystem I reaction centre complex from spinach thylacoid membranes. Plant Mol Biol 2:95–107

Woodcock DM, Crowther PJ, Doherty J, Jefferson S, DeCruz E, Noyer-Weidner M, Smith SS, Michael MZ, Graham MW (1989) Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. Nucleic Acids Res 17:3469–3478

Zurawski G, Perrot B, Bottomly W, Whitfeld PR (1981) The structure of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase from spinach chloroplast DNA. Nucleic Acids Res 14:3251–3270