Theor Appl Genet (1989) 77:625–633

A contiguous sequence in spinach nuclear DNA is homologous to three separated sequences in chloroplast DNA

W.Y. Cheung* and N.S. Scott

CSIRO, Division of Horticulture, P.O. Box 350, Adelaide S.A. 5001, Australia

Received August 8, 1988; Accepted December 30, 1988
Communicated by A.L. Kahler

Summary. A 3.4-kbp nuclear (n) DNA sequence has greater than 99% sequence homology to three segments of the chloroplast (cp) genes *rps*2, *psb*D/C, and *psa*A respectively. Each of these cpDNA segments is less than 3 kbp in length and appears to be integrated, at least in part, into several (>5) different sites flanked by unique sequences in the nuclear genome. Some of these sites contain longer homologies to the particular genes, while others are only homologous to smaller parts of the cp genes. Both the cpDNA fragments found in the nuclear genome and their flanking nDNA sequences are invested with short repeated A-T rich sequences but, apart from a hexanucleotide sequence and a palindromic sequence identified near each recombination point, there is no obvious structure that can suggest a mechanism of DNA transfer from the chloroplast to the nucleus in spinach.

Key words: Nuclear DNA – Chloroplast DNA – Sequence homology – Spinach

Introduction

There have been several reports of "promiscuous" DNA in a wide range of phyla in recent years. The term was coined (Ellis 1982) to describe DNA sequences found in more than one of the three genetic compartments, nucleus (n), chloroplast (cp) and mitochondrion (mt) of eukaryotes. Southern hybridization experiments showed extensive sequence homology between cp- and mtDNA in several plant species (Stern and Palmer 1984), and a sequence homologous to the gene for the large subunit of ribulose bisphosphate carboxylase/oxygenase, a chloroplast coded enzyme, was found in maize mtDNA. This mitochondrial sequence could be translated in vitro to yield a protein recognized by an antibody specific to the native chloroplast protein (Lonsdale et al. 1983) Sequence homologies have also been reported between mtDNA and nDNA in sea urchin (Jacobs et al. 1983).

In spinach virtually the entire chloroplast genome appeared to be integrated at specific sites in the nuclear genome and the homology, as characterized by Southern hybridization, was the equivalent of about five copies of cpDNA integrated into the nuclear genome (Timmis and Scott 1983; Scott and Timmis 1984). This homology also extends to include sequences common to chloroplast, mitochondria and nucleus in spinach (Whisson and Scott 1985), and to chloroplast and nuclei in other species of Chenopodiaceae (Ayliffe et al. 1988).

In this report we demonstrate that a 3.4-kbp segment of nDNA contains a continuous sequence homologous to three separate pieces of cpDNA. There are only 32 base differences between the cp- and nDNA sequences amounting to a greater than 99% sequence homology between the two. Changes to reading frames indicate that this nuclear sequence is unlikely to encode proteins.

Materials and methods

Isolation of total spinach DNA and nuclear DNA

Total DNA was prepared as previously described (Scott and Possingham 1980) from spinach leaf tissues, and nuclear DNA from nuclei of spinach (Scott and Timmis 1984).

Molecular cloning into λ phage

nDNA from leaves was digested with HpaII (5 units/μg) of DNA at 37 °C for 24 h, and the buffer was adjusted to conditions suitable for EcoRI digestion as recommended by the supplier. The sample was divided into three parts and digested with 0.03,

* Present address: Plant Breeding Institute, Maris Lane, Trumpington, Cambridge CB2 2LQ, UK

0.1 and 0.6 units of EcoRI per μg of DNA, respectively, for 30 min at 37 °C to generate a partial EcoRI digest, and the enzyme was heat inactivated at 65 °C for 10 min. The partial DNA digest was fractionated in a 10–35% sucrose gradient as described in Maniatis et al. (1982). Fractions containing DNA fragments of 15–25 kbp were collected and precipitated by ethanol.

The partially digested nDNA was ligated into the EcoRI site of λ phage vector EMBL4 (kindly donated by Dr A.M. Frischauf), as described by Frischauf et al. (1983). Preparation of phage packaging proteins and the packaging of ligated λ DNA were performed as described by Maniatis et al. (1982). *E. coli* strain NM539 was infected as the selective host for the recombinant λ DNA.

Screening of the library

The yield of recombinant clones in the library was about 10^3 pfu/μg of DNA, and consequently the plaque density on plates was too low to allow direct plaque selection by hybridization. Two libraries were prepared (23,000 and 18,000 clones, respectively), pooled and amplified once, before selection by colony hybridization. The phage DNA was transferred to nylon filters (Pall – Australia), and colonies hybridizing to the radioactively labelled 7.7-kbp PstI cpDNA fragment were selected (Maniatis et al. 1982). Two related nuclear clones showing hybridization to the cpDNA fragment were purified.

Molecular cloning into plasmids

Various DNA sequences from the recombinant clone were cloned into the plasmids pUC18 and pUC19 using the procedures described by Maniatis et al. (1982). The recombinant plasmids were purified from 100 ml liquid cultures using the alkaline lysis method (Maniatis et al. 1982), followed by a final centrifugation step in cesium chloride gradients ($\sigma = 1.55$ g/ml) containing ethidium chloride (100 μg/ml) at 100,000 rpm for 4 h at 25 °C in a TLA 100.2 rotor using the Beckman TL100 centrifuge.

Gel electrophoresis and Southern hybridizations

DNA fragments were fractionated by electrophoresis on 0.8% agarose gels in TAE buffer (20 m*M* TRIS, 10 m*M* Acetate, 1 m*M* EDTA; pH 7.8) and transferred to nylon membrane (Pall) according to Southern (1975). The radioactive probes used in the hybridizations were labelled with [α-^{32}P]-dATP (1,800 Ci/m*M*, BRESA, South Australia) by nick-translation of double stranded DNA, as described by Schachat and Hogness (1974). Nucleic acid hybridizations were performed under conditions already described (Timmis and Scott 1983). HindIII and HindIII – EcoRI double digests of bacteriophage lambda DNA were used as size markers.

DNA sequence analysis

Cloning for sequence analysis was carried out using M13 bacteriophage vectors mp18 and mp19 (Yanisch-Perron et al. 1985). DNA sequences were determined by the chain termination method of Sanger et al. (1977) with the 17-mer universal primer (BRESA) followed by electrophoresis in 0.25 mm thick, 6% acrylamide gels containing 7 *M* urea in TBE buffer (0.09 *M* TRIS, 0.09 *M* boric acid, 0.0025 *M* EDTA; pH 8.3). The sequencing strategies are shown in Fig. 3a.

Computer analysis of nucleotide sequences was performed on the MBIS system (Bucholz and Reisner 1986) furnished by CSIRO Division of Molecular Biology.

Results

Isolation of a nuclear sequence homologous to chloroplast DNA

The cloning and subsequent isolation of nDNA sequences free of contaminating cpDNA, in a λ phage library, took advantage of the methylation of most of the HpaII-sensitive sites in the nDNA of spinach and the accompanying lack of methylation of cpDNA, to separate all of the cpDNA fragments from most of the nDNA (Timmis and Scott 1983). The extensive digestion of total leaf DNA preparations, which contain some 1,000–2,000 copies of cpDNA for every copy of nDNA (Scott and Possingham 1980), with HpaII, reduced the cpDNA to fragments which had both the wrong ends and were too small to be cloned in the EcoRI sites of λ EMBL4. In addition, the partial digestion of the leaf DNA with EcoRI and sizing of the resultant DNA fragments ensured that there was no detectable contamination of the 15–25 kbp nDNA fragments with cpDNA. No clones of cpDNA have been found in the library constructed from this material. Most of the nuclear sequences which have homology to cpDNA are in the methylated fraction of the genome (Scott and Timmis 1984) and should be represented in this library. There are relatively few sites in native spinach nDNA for HpaII (Timmis and Scott 1983).

The spinach nuclear genome is about 1.0 pg/C (0.9×10^6 kbp; Bennett et al. 1982) and, thus, one complete genome could be contained in 45,000 20-kbp fragments. The nDNA sequence reported here was selected from a library of about 40,000 clones and would represent only a part of the spinach genome. Attempts to obtain a genomic library from purified nDNA were not successful, despite successful ligation of nDNA fragments into EMBL4 DNA. It seemed likely that the formaldehyde treatment used during the purification of nuclei modified the nDNA to the extent that it could not be replicated in the phage system.

Using the 7.7-kbp PstI cpDNA fragment (Palmer and Thomson 1981) containing the cp gene *psa*A (Kirsch et al. 1986) as a probe, a genomic clone λ18/2 with a 14.1-kbp nDNA insert was selected from the library. λ18/2 contains 6 EcoRI nDNA fragments (Fig. 1a). By restriction digestion and hybridization analyses using the 7.7-kbp PstI cpDNA as a probe, the sequences in λ 18/2 with homology to cpDNA were located within the 2.7-kbp and the 1.9-kbp EcoRI nDNA fragments (Fig. 1b). These two EcoRI fragments were subsequently analysed in detail by DNA sequencing. Figure 1c and d show the hybridization of λ18/2 DNA (including the arms of the clone) to EcoRI digests of approximately equimolar amounts of λ18/2 clone can be seen in both autoradiograms, indicating that this DNA clone is not an artefactual rearrangement of DNA sequences during the

cloning procedure, but corresponds to the arrangement of sequences in the nDNA. The two smaller nDNA fragments are not visible in this autoradiogram. λ18/2 also hybridized to other sequences in the nDNA (Fig. 1 d) and some of these will be discussed further below.

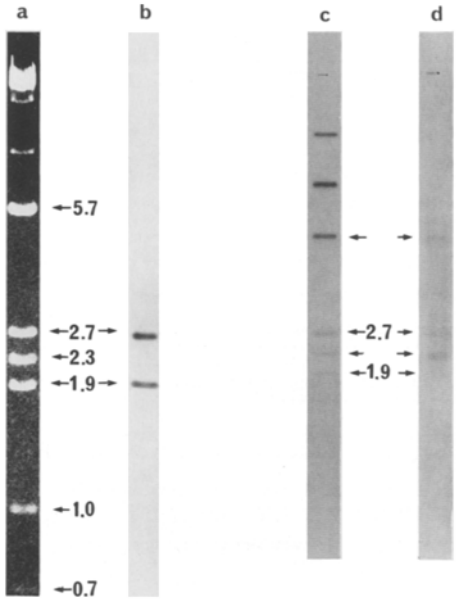
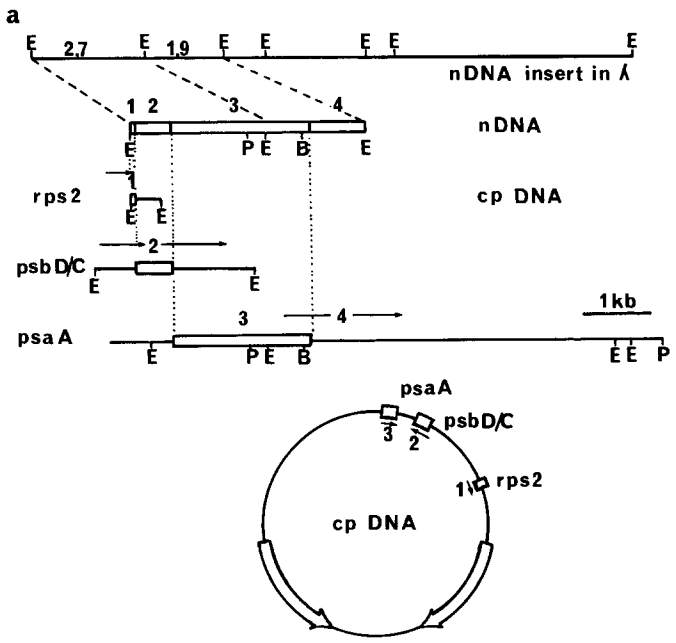


Fig. 1 a–d. Homology of a genomic clone, λ18/2, of spinach nuclear DNA to chloroplast DNA. An EcoRI digest of the clone stained with ethidium bromide with the 6 EcoRI nDNA fragment marked **a**; and hybridized to a 7.7-kbp PstI cpDNA fragment, **b**. Approximately equimolar amounts of EcoRI digests of the genomic clone λ18/2 (0.1 ng), **c**; and spinach nuclear DNA (7 μg), **d**; hybridized with the genomic clone. In **a** the λDNA cos ends are annealed together, while in **c** they are not

A map of the six EcoRI fragments comprising λ18/2 is shown in Fig. 2a. The homologies of the first two fragments (2.7 and 1.9 kbp, respectively) to chloroplast DNA sequences are also shown. The derivation of this map from the results in Fig. 1, 2b and 3b will be described below.

The 2.7-kbp EcoRI nDNA fragment hybridizes to three regions of the cp genome (Fig. 2b, track 1): a 0.6-kbp EcoRI cpDNA fragment which is internal to the chloroplast gene coding for the ribosomal small subunit protein 2 (*rps*2) (W. Bottomley, personal communication), a 3.1-kbp EcoRI cpDNA fragment containing the genes for the photosystem II proteins, D2 and the 44 kD reaction-centre, chlorophyll a-binding protein (psbD/C) (Holschuh et al. 1984), and a 2.3-kbp EcoRI cpDNA fragment adjacent to the 6.6-kbp EcoRI cpDNA containing the gene for the P700 chlorophyll a-binding protein (*psa*A) (Kirsch et al. 1986). The position of these sequences on the cpDNA genome is shown in Fig. 2a. The 1.9-kbp EcoRI nDNA fragment adjacent to the 2.7-kbp EcoRI nDNA fragment was divided into two pieces: a 0.6-kbp EcoRI-BamHI fragment and a 1.3-kbp BamHI-EcoRI fragment. Both of these subfragments hybridized to the 6.6-kbp EcoRI cpDNA fragment (Fig. 2b, lanes 2 and 3), which is internal to the 7.7-kbp *Pst*I cpDNA fragment used to select the nuclear clone, and contains the *psa*A gene (Fig. 2a). Combining the results of the hybridization analyses of Fig. 1 and 2b with the sequence data in Fig. 3b discussed below, the relation between the 2.7-kbp and the 1.9-kbp EcoRI fragments of λ18/2 and the three separate areas of the chloroplast genome is summarized in Fig. 2a.



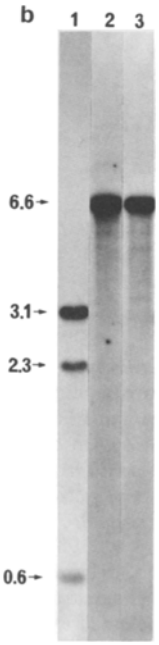


Fig. 2 a and b. A nuclear DNA sequence homologous to cpDNA. **a** The map of a genomic clone of nDNA (λ18/2) is shown in relation to the three fragments of cpDNA marked 1, 2 and 3, respectively, with which it has homology. The region marked 4, flanking the sequence 3, does not have homology to cpDNA. The directions of transcription of the respective chloroplast genes involved in the homology (*rps2*, *psb*D/C, *psa*A), and their extent, are marked by arrows. The location of these chloroplast genes on the spinach cpDNA circular genome is also shown. The *open arrows* on the spinach chloroplast genome represent the long inverted repeats of cpDNA. The abbreviations for restriction sites are: E – EcoRI; B – BamHI; P – PstI. **b** EcoRI digests of cpDNA hybridized with nuclear DNA fragments: *1* the 2.7-kbp EcoRI fragment; *2* 0.6-kbp EcoRI-BamHI fragment (from the ExoRI 1.9-kbp fragment); and *3* the 1.3-kbp BAMHI-EcoRI fragment (from the EcoRI 1.9 kbp-fragment)

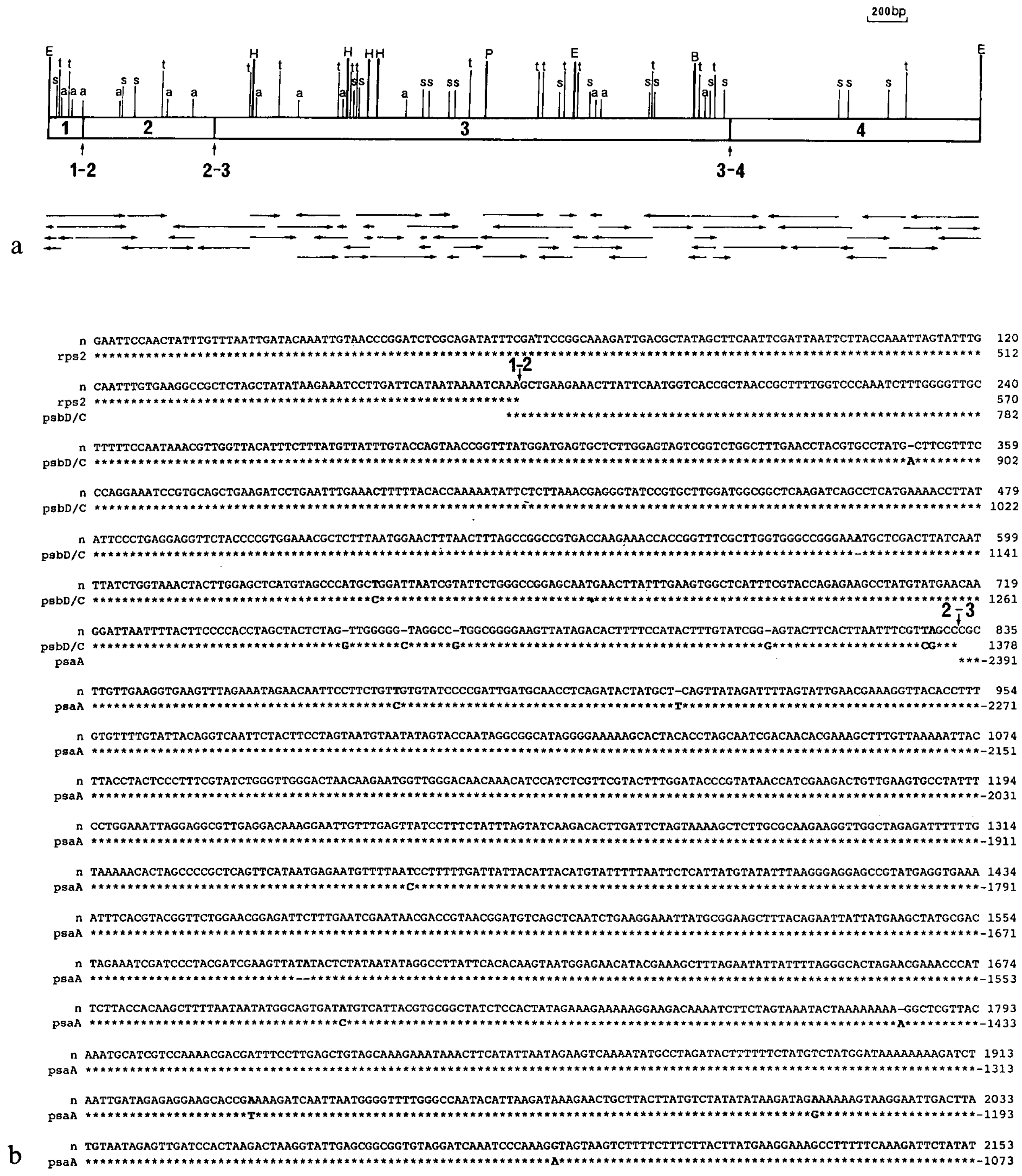
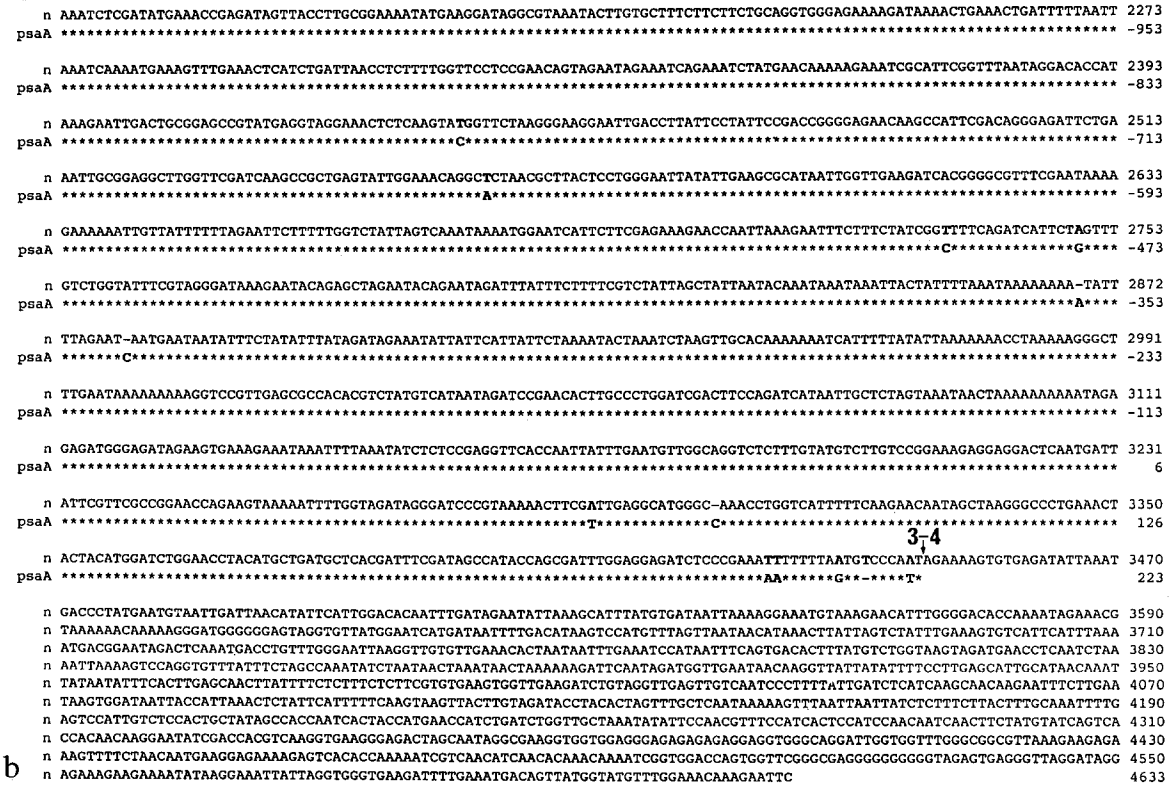


Fig. 3. **a** Sequencing of the two contiguous EcoRI fragments of λ18/2 showing homologies to three separate regions of cpDNA. The M13 clones sequenced are aligned with the restriction map and the *arrrows* below the map show the direction and the length of the sequence analysed. Restriction site abbreviations are as follows: E – EcoRI; B – BamHI; H – HindIII; P – PstI; s – Sau3A; a – AluI; t – TaqI. 1, 2 and 3 represent areas of homology in nDNA to separate segments of the cp genome and 4 is the adjacent, non-homologous part of the nuclear genome as described in Fig. 2. **b** Nucleotide sequence of the two contiguous EcoRI nDNA fragments with homologies to cpDNA. Sequences of portions of chloroplast genes homologous to nDNA are represented by *asterisks*. Sequence numbers of the *rps*2 are in relation to the *rps*2 putative start codon (W. Bottomley, personal communication). Sequence numbers of the *psb*D/C genes are as published (Holschuh et al. 1984). Sequence numbers of the *psa*A gene are as published (Kirsch et al. 1986). The junctions between sequences are marked with *arrow* and designated as 1–2, 2–3 and 3–4 for the junction of the *rps*2 portion with *psb*D/C, *psb*D/C with *psa*A and *psa*A with nuclear DNA, respectively. The nucleotide differences between the nDNA and the cpDNA along the region of homology are indicated in *bold type*

Fig. 3 (continued)



The sequences of nuclear DNA homologous to chloroplast DNA

Figure 3a shows the strategy used to sequence the two contiguous EcoRI fragments of λ18/2 shown in Fig. 2a which are homologous to cpDNA. Figure 3b shows the sequence of 4,633 bp of nDNA which includes 3,448 bp of nDNA sequence homologous to cpDNA and extends 1,185 bp into nDNA with no homology to cpDNA. The other end of the sequence terminates in the nuclear clone at an EcoRI site common to both nuclear and chloroplast genomes. There is, presumably, further homology to cpDNA in the adjoining fragment of nDNA which is not present in this genomic clone.

The homology to cpDNA begins at the left-hand EcoRI terminus of the genomic clone (Fig. 2a), at a corresponding EcoRI site in the *rps*2 gene of cpDNA. The homology (region 1, Fig. 2a and 3a) is to 178 bp, including 158 bp of the C-terminal coding region of the *rps*2 gene, and the homology is exact. The sequence of *rps*2 has been kindly provided by Dr. W. Bottomley prior to publication. Overlapping the *rps*2 homology by two bases is a sequence of 656 bp of nDNA, which is homologous to the end of the gene *psb*D and the beginning of *psb*C (Holschuh et al. 1984), (region 2, Figs. 2a and 3a). The nDNA differs from the cpDNA in this case by 1 base insertion, 5 deletions and 3 substitutions.

Immediately adjacent to the end of region 2 is a sequence of 2,616 bp which is homologous to 2,615 bp of cpDNA. This cpDNA region consists of 2,391 bp of noncoding DNA flanking the *psa*A gene, and the first 224 bp of *psa*A coding sequences (region 3, Figs. 2a and 3a). The majority of the sequence of this cpDNA has been made available before publication by Dr. W. Bottomley and that of the *psa*A gene by Dr. R. G. Herrmann (Kirsch et al. 1986). Compared to the cpDNA at region 3 (Figs. 2a and 3a), the nuclear sequence contains 15 base changes, 3 insertions and 5 deletions. The remaining nuclear sequence (Fig. 3b, region 4) shows no homology with cpDNA by Southern blotting or by comparison directly with recorded cpDNA sequences.

Earlier observations (Timmis and Scott 1983; Scott and Timmis 1984) indicated that each cpDNA sequence studied had more than one site of homology in the nuclear genome. Such a result is seen in Fig. 1d when the genomic clone λ18/2 hybridized to other EcoRI fragments in the nuclear DNA apart from its parent sequence. Three separate fragments of this nuclear clone, each containing homology to cpDNA, were hybridized to genomic blots of nuclear DNA (Fig. 4). The 2.7-kbp EcoRI nDNA fragment hybridizes to its parent band as well as to other EcoRI fragments of 3.1, 2.3, 1.8 and 0.6 kbp (Fig. 4a) The same nDNA also hybridizes to five BamHI fragments (Fig. 4b), of which the 5.2-kbp frag-

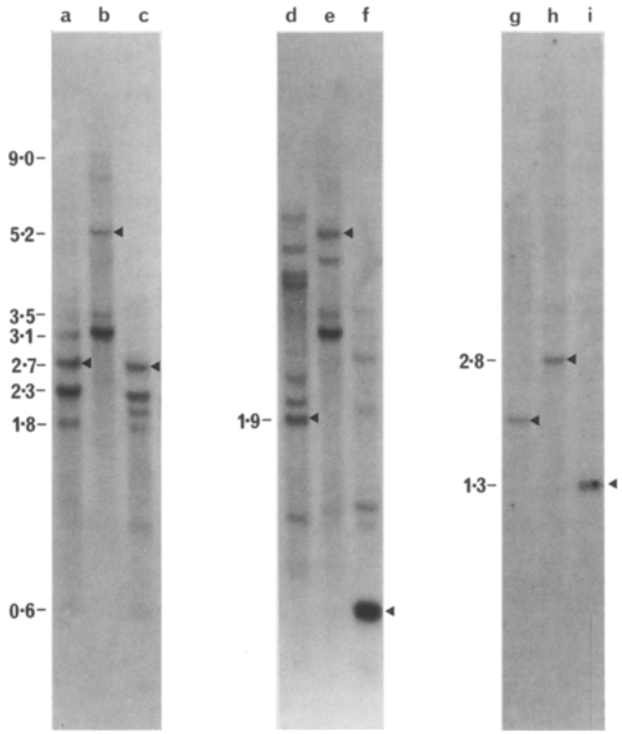


Fig. 4a–i. Hybridizations of fragments of the nDNA clone, λ18/2, to nDNA digests. nDNA was digested with EcoRI (**a, d** and **g**), BamHI (**b, e** and **h**) and EcoRI-BamHI (**c, f** and **i**) and hybridized with the 2.7-kbp EcoRI fragment of nDNA from the nuclear genomic clone λ18/2 (**a–c**), the 0.6-kbp EcoRI-BamHi nDNA fragments (**d–f**) and the 1.3 kbp-BamHI-ExoRI fragment (**g–i**), respectively. Fragments marked with the solid *arrow heads* are those containing the sequences in the nuclear DNA clone used as a probe in the respective experiments. The other fragments represent further sequence homologies in the nuclear DNA to the individual probes

ment common to Fig. 4b and e is deduced to be the fragment containing the entire 2.7-kb EcoRI fragment (Fig. 4a). The others (Fig. 4b) are of sizes 9.0, 7.5, 3.5 and 3.1 kbp. All these BamHI fragments contain internal EcoRI sites, as on double digestion of nDNA by EcoRI and BamHI a population of six fragments of smaller sizes are generated (Fig. 4c).

The 0.6-kbp EcoRI-BamHI fragment of nuclear DNA, which is homologous to a similar 0.6-kbp cpDNA fragment (Fig. 2a and 3b), hybridizes to several EcoRI nDNA fragments (Fig. 4d), of which the 1.9-kbp EcoRI fragment contains the 0.6-kbp EcoRI-BamHI fragment used as the probe. With the BamHI digest (Fig. 4e), apart from the 5.2-kbp fragment containing the 0.6-kbp probe, a population of other fragments with various sizes is also observed. Three fragments (5.2 kbp, 3.5 kbp and 3.1 kbp) hybridize to both the 2.7-kbp EcoRI fragment and the 0.6-kbp BamHI-EcoRI (compare Fig. 4b and e), indicating that other than the nDNA sequence described here, there are at least two other locations in nuclear DNA similar to region 3 (Fig. 2a), which include the same EcoRI site and flanking sequences. Several of the BamHI fragments (Fig. 4e) can be digested by EcoRI to excise the 0.6-kbp EcoRI-BamHI fragment of nuclear DNA (Fig. 4f). The absence of the 6.6-kbp EcoRI cpDNA fragment in Fig. 4d also provides a positive control for the absence of contaminating cpDNA in the nDNA reparation used in this experiment. Thus, the many homologous fragments observed in these hybridizations are located in different regions of the nuclear genome.

Dot blotting of nDNA with the 0.6-kbp EcoRI-BamHI fragment (results not shown) indicates that there are about six copies of this sequence in the nuclear DNA – a figure consistent with the multiple hybridization sites observed in Fig. 4a, and the earlier estimates of four to five copies of cpDNA integrated into the nuclear genome (Scott and Timmis 1984).

The 1.3-kbp BamHI-EcoRI nDNA fragment, which contains a 224-bp homology to cpDNA (Fig. 2a and 3b), hybridizes to the 1.9 kbp-EcoRI fragment (Fig. 4g), a BamHI fragment of 2.8 kbp (Fig. 4h), and to itself in the EcoRI, BamHI and BamHI-EcoRI digests, respectively. This indicates that despite the multiple homologies to the adjoining 0.6-kbp fragment, the major part of this particular nuclear homology to cpDNA is almost unique in the nuclear genome, and that one end of the insertion involving the 0.6-kbp fragment in multiple copies in the nDNA is always very close to its EcoRI site.

Discussion

A large number of nuclear DNA sequences sharing sequence homology to the chloroplast genome have been shown to occur at specific sites by Southern hybridizations in spinach (Timmis and Scott 1983; Scott and Timmis 1984). In this report, with the isolation and analysis of such a nuclear sequence we provide conclusive evidence of a high degree of sequence homology between corresponding nDNA and cpDNA regions. In this particular area three such sequences are clustered at a specific site in the nuclear genome. The *rps*2 gene is 14 kbp away from the *psb*D/*psb*C genes which are also separated by a further 4 kbp from the *psa*A gene. These three fragments are derived from segments which are almost certainly transcribed as separate operons in the circular chloroplast genome. In addition, in the nuclear genome, the relative polarity of the *psb*D/C gene fragment is reversed with respect to the arrangement of the *psa*A and the *rps*2 genes in the chloroplast genome (Fig. 2a). If this arrangement was achieved by recombination (and possibly by deletion as well), then there must have been at least three events involved, one for each of the junctions presumably at the points marked in Figs. 2a and 3a.

Recombination points 1–2 and 2–3 involve the joining of two pieces of cpDNA while 3–4 involves cpDNA and nDNA interactions. It is not possible to distinguish whether the two recombinations between cpDNA are of intramolecular or intermolecular origin.

The substantial sequence homology (99%) between the nDNA sequence cloned here and the corresponding cpDNA sequences with a low frequency of point mutations in the nDNA indicates that any transfer and integration event may have been relatively recent. Alternatively there may exist strong selective pressure to maintain such a structure in the nucleus, although there are no obvious functions for these sequences. The joining of the three cpDNA domains does not result in any long open reading frames, nor does the recombination between the N-terminus coding region of *psa*A with the unique nDNA sequence appear related to a coding sequence in the nucleus. Rather, changes in these sequences ensure that the reading frames are terminated frequently. These insertion/deletion events, each involving a single nucleotide and their effects on the coding capacity of the cp genes involved, are most likely to have occurred during or after any transfer of the cpDNA sequences into the nuclear genome, since chloroplast functions would be greatly impaired by such changes.

Comparison of the sequences at the junction points does not reveal an obvious mechanism to explain how these three pieces of cpDNA are brought together in the nDNA. The best alignment of the three junction sequences (Fig. 5) is a hexanucleotide region of sequence AA(G/A) (C/G)TG close to, but not exactly at, the recombination points and this could also be seen as a slightly longer and less specific 10-base sequence pRAApuNTGNpupR. It is not clear whether this sequence has a role in the recombination events or is a coincidence. There is also a palindromic structure close to the two "chloroplast/chloroplast" junctions (Fig. 5). There is no obvious duplication of short sequences precisely at the recombination points such as are found in the case of transposable elements like *Mu, Spm,* and *Ac* in maize (Saedler and Nevers 1985; Döring and Starlinger 1984).

There are numerous short domains 9–15 nucleotides long, and mostly A-T rich, repeating once along the length of this nuclear fragment of DNA. Their distribution is summarized in Fig. 6. The majority of these repeats occur in the regions of homology to cpDNA, particularly in the 5' flanking sequences of the *psa*A homology. It has been suggested that distant short repeated sequences can cause deletion mutations due to misalignments during DNA replication (Drake et al. 1983). These repeats may generate short deleted sequences which, during DNA replication in either the chloroplast prior to transposition to the nucleus or in the nucleus itself, could be the source of the family of homologies found in the nuclear genome. Whatever the mechanism, it is clear that the sequences described in this report also occur in part, or in other combinations in other nuclear backgrounds, in the spinach genome.

Another possible mechanism for transposition was suggested by Schuster and Brennicke (1987), who found a reverse transcriptase-like open reading frame in the mtDNA of *Oenothera,* linked to regions of mtDNA containing homology to intergenic regions of both a cpDNA and nDNA fragment. They suggested that if this reverse transcriptase were expressed, it might be involved in transfers of transcribed DNA sequences. Not all our results would support such a mechanism. For example, most of the longest stretch of homology to cpDNA described here is probably non-transcribed sequences, and our earlier work (Scott and Timmis 1984) suggests that nearly all cpDNA sequences – transcribed or not – are involved in the transfer.

The very high levels of homology found in this study between cp- and nDNA sequences have also been found where some cpDNA sequences have been reported in maize mtDNA. In particular, a mt sequence homologous to the cp gene *rbc*L (Lonsdale et al. 1984) could be transcribed and translated into the protein (as identified by an antibody) in an in vitro system. In a male sterile maize line, a distinct 420-bp region of the cp *psb*A gene was found in the mtDNA and in the 6.4-kbp S1 plasmid (Sederoff et al. 1986). It was about 90% homologous to the cpDNA and differed only in base substitutions and length mutations. In a tobacco male-sterile cytoplasm, a 99-bp fragment of a cp gene *trn*A was found in mtDNA (Dewey et al. 1986).

It is most likely that "promiscuous" cpDNA sequences in nDNA arise from transfer of cpDNA from

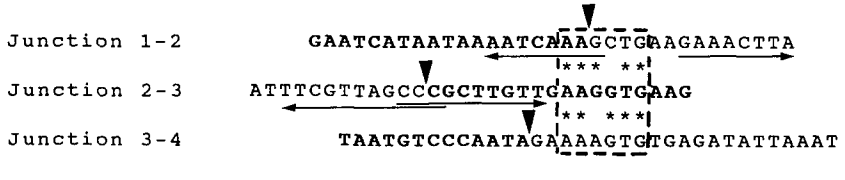


Fig. 5. Alignment of the sequences at the junctions 1–2, 2–3 and 3–4 (Figs. 2–3). The *vertical arrows* above the sequence mark the positions of the junctions. The *horizontal arrows* below the sequence mark the two reverse repeats at the two junctions of the cpDNA homologies (1–2 and 2–3). The hexanucleotide sequence "AA(G/A) (C/G)TG" found near the junctions is *boxed*

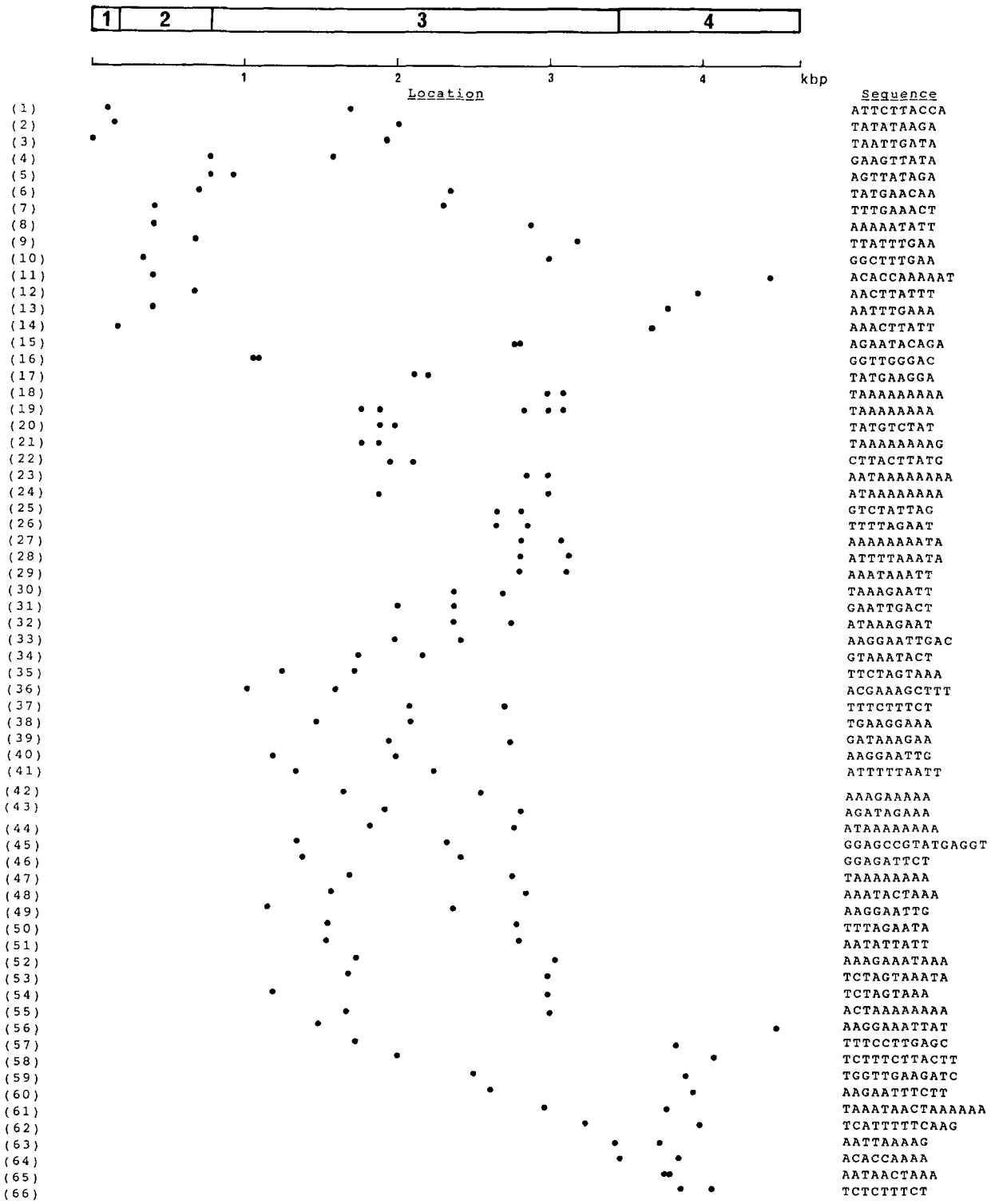


Fig. 6. Distribution of short direct repeats in the nDNA with sequence homology to cpDNA. The *open bar* represents the entire nDNA sequence (4,633 kbp) shown in Fig. 3 b, subdivided into four regions: *1* – with sequence homology to *rps*2; *2* – with sequence homology to *psb*D/C; *3* – with sequence homology to *psa*A; and *4* with no sequence homology to cpDNA. The position of each repeating domain is indicated by a *dot* with respect to the scale drawn underneath the sequence, with the context of the domain summarized in the right-hand column

the chloroplast to the nucleus, rather than the reverse – i.e. cpDNA containing "promiscuous" nDNA sequences – for a number of reasons. In particular, the evidence used to support the symbiotic hypothesis of chloroplast and mitochondrial origin suggests that there has been a transfer of most of the genes from the organelle precursors to the nucleus, such that organelle function is now largely dependent on the nuclear genome and its interaction with a small number of remaining organellar genes (Dyer 1984). Chloroplast DNA in particular is also characterized by a very economic and highly conserved arrangement of genes (Dyer 1984; Shinozaki et al. 1986) and there is little scope for incorporation of foreign DNA fragments in the genome. Insertion into intergenic regions would inactivate vital genes for chloroplast function and there are few intragenic spaces in the chloroplast genome. These observations, combined with the high copy number of cpDNA molecules in chloroplasts and cells (Scott and Possingham 1980), may explain why promiscuous DNA sequences have not been found in cpDNA. Our results are most simply explained by the ongoing transfer of cpDNA fragments to the nDNA, with the possibility of subsequent nuclear rearrangement.

Acknowledgements. The contribution of Dr. J. N. Timmis to the construction of the genomic library is gratefully acknowledged.

References

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

Bennett MD, Smith JB, Heslop-Harrison JS (1982) Nuclear DNA amounts in angiosperms. Proc R Soc London Ser B 216:179–199

Bucholz CA, Reisner AH (1986) An integrated system for the retrieval and analysis of sequence data from nucleic acids and proteins Nucleic Acids Res 14:265–272

Dewey RE, Levings CS, Timothy DH (1986) Novel recombinations in the maize mitochondrial genome produce a unique transcriptional unit in the Tesios male – sterile cytoplasm. Cell 44:439–449

Döring H-P, Starlinger P (1984) Barbara McLintock's controlling elements: now at the DNA level. Cell 39:253–259

Drake JW, Glickman BW, Ripley LS (1983) Updating the theories of mutation. Am Sci 71:621–630

Dyer TA (1984) The chloroplast genome: its nature and role in development. In: Baker NR, Barber J (eds) The chloroplast genome: its nature and role in development in chloroplast biogenesis. Elsevier, Amsterdam, pp 23–69

Ellis RJ (1982) Mobile genes of chloroplasts and the promiscuity of DNA. Nature 299:678–679

Frischauf A-M, Lehrach H, Poustka A, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. J Mol Biol 170:827–842

Holschuh K, Bottomley W, Whitfeld PR (1984) Structure of the spinach chloroplast genes for the D2 and 44 kb reaction centre proteins of photosystem 11 and for +RNA ser (UGA). Nucleic Acids Res 12:8819–8834

Jacobs HT, Posakony JW, Grula JW, Roberts JN, Xin J-H, Britten RJ, Davidson EH (1983) Mitochondrial DNA sequences in the nuclear genome of strongylocentrotus purpuratus. J Mol Biol 165:609–632

Kirsch W, Seyer P, Herrmann RG (1986) Nucleotide sequence of the clustered genes for two P-700 chlorophyll A apoproteins of the photosystem I reaction centre and the ribosomal protein S-14 of the spinach plastid chromosome. Curr Genet 10:843–856

Lonsdale DM, Hodge TP, Howe CJ, Stern DB (1983) Maize mitochondrial DNA contains a sequence homologous to the ribulose-1,5-bisphosphate carboxylase large sub unit gene of chloroplast DNA. Cell 34:1007–1014

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor/NY

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

Saedler H, Nevers P (1985) Transposition in plants: a molecular model. EMBL J 4:585–590

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

Schachat FH, Hogness DS (1974) Repetitive sequences in isolated Thomas circles from Drosophila melanogaster. Cold Spring Harbor Symp Quant Biol 38:371–481

Schuster W, Brennicke A (1987) Plastid, nuclear and reverse transcriptase sequences in the mitochondrial genome of Oenothera: is genetic information transferred between organelles via RNA. EMBL J 6:2857–2863

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

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

Sederoff PR, Ronald P, Bedinger P, Rivin C, Walbot V, Bland M, Levings CS (1986) Maize mitochondrial plasmid S-1 sequences share homology with chloroplast gen psbA. Genetics 113:469–482

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organiszation 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, Palmer JD (1984) Mitochondrial and chloroplast genomes of maize have a 12 kb DNA sequence in common. Proc Natl Acad Sci USA 81:1946–1950

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

Whisson DL, Scott NS (1985) Nuclear and mitochondrial DNA have sequence homology with a chloroplast gene. Plant Mol Biol 4:267–273

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119