Plant Molecular Biology 9: 485–496 (1987)

Physical map and gene localization on sunflower (*Helianthus annuus*) chloroplast DNA: evidence for an inversion of a 23.5-kbp segment in the large single copy region

Françoise Heyraud,[1] Pascale Serror,[1] Marcel Kuntz,[2] André Steinmetz[2] and Philippe Heizmann[1,*]
[1]*Laboratoire de Biologie Cellulaire, Université Claude Bernard, Lyon-I, F-69622 Villeurbanne, France (*author for correspondence);* [2]*Laboratoire de Biochimie, IBMC-CNRS, Université Louis Pasteur Strasbourg, F-67000 Strasbourg, France*

Received 24 April 1987; accepted in revised form 14 July 1987

Key words: chloroplast DNA, gene localization, *Helianthus*, physical map, restriction endonuclease analysis, sunflower

Abstract

As a first step in the study of chloroplast genome variability in the genus *Helianthus*, a physical restriction map of sunflower (*Helianthus annuus*) chloroplast DNA (cpDNA) has been constructed using restriction endonucleases BamH I, Hind III, Pst I, Pvu II and Sac I. Sunflower circular DNA contains an inverted repeat structure with the two copies (23 kbp each) separated by a large (86 kbp) and a small (20 kbp) single copy region. Its total length is therefore about 152 kbp. Sunflower cpDNA is essentially colinear with that of tobacco with the exception of an inversion of a 23.5-kbp segment in the large single copy region. Gene localization on the sunflower cpDNA and comparison of the gene map with that from tobacco chloroplasts have revealed that the endpoints of the inversion are located between the trnT and trnE genes on the one hand, and between the trnG and trnS genes on the other hand.

Analysis of BamH I restriction fragment patterns of *H. annuus, H. occidentalis* ssp. *plantagineus, H. grosssesseratus, H. decapetalus, H. giganteus, H. maximiliani* and *H. tuberosus* cpDNAs suggests that structural variations are present in the genus *Helianthus*.

Introduction

The growing importance of the cultivated sunflower (*Helianthus annuus*) as a crop for vegetable oil during the last two decades is largely due to the production of a cytoplasmic male sterile line (CMS) by Leclercq in 1969, through interspecific crosses between *H. annuus* as the male progenitor and *H. petiolaris* as the female cytoplasmic donor parent [16]. Future improvements of this culture through the production of new varieties of CMS by interspecific crosses, and through introgression of characters of agricultural interest (resistance to frost, to drought, to natural predators...) from wild species of the genus *Helianthus* into cultivated sunflower lines, will require extensive knowledge of the genetic properties of these species. Although an infrageneric classification of the genus *Helianthus* has already been proposed using complementary methods of biosystematics [31], uncertainties about the phylogenetic relationships of some species remain unresolved, preventing their rigorous classification. In addition, the genetic variability within these species, particularly within those used for the

production of CMS, is an important factor for the diversity of the CMS obtained with the various individuals of a given species.

The present paper reports the structural arrangement of the chloroplast DNA (cpDNA) of sunflower, as an initial step to the analysis of the cytoplasmic variability in the genus *Helianthus*. Its structure has been deduced by comparison with physical maps of cpDNA from other species. The structure of chloroplast genomes from higher plants is indeed highly conserved [25]; cpDNAs consist of covalently closed circular molecules sized between 120 kbp (in *Pisum sativum*, [25]) and 160 kbp (in *Atropa belladonna*, [7]). Genome sizes of *Nicotiana accuminata* (171 kbp, [33]), *Spirodela oligorhiza* (180 kbp, [38]) and *Pelargonium hortorum* (217 kbp, [23]) are exceptionally large. Most cpDNAs contain an inverted repeat structure (IR) in which each of the two copies (20 kbp) carries a set of the four ribosomal RNA genes; some legumes like pea or broad bean, having only one copy of the ribosomal genes, are exceptions to this rule. These IR separate the large single copy (LSC) and the small single copy (SSC) regions.

Using heterologous DNA cross-hybridizations, Palmer and Thompson [25] showed that the cpDNA of spinach, petunia and cucumber are essentially colinear; they differ from the cpDNA of certain legumes by an inversion of about 50 kbp located in the LSC region and from the cpDNA of monocots like maize and wheat by a second inversion of about 20 kbp located within the first one. Other dicot plastomes, for instance those of tobacco [7], cabbage [24] and tomato [26], show the same gene arrangement as those of spinach, petunia or cucumber.

Taking as a working hypothesis that sunflower cpDNA is colinear with that from tobacco, we used cloned tobacco cpDNA fragments (as well as wheat cpDNA fragments and spinach gene probes) to construct a structural map of *H. annuus* cpDNA. Twenty-five genes were located on this physical map, which was found to be indeed essentially colinear with that of cpDNAs from other dicots such as tobacco or spinach, with the exception of an inversion of about 23.5 kbp located in the large single copy region.

A variable region of the chloroplast genome in the genus *Helianthus* was identified by comparison of the restriction patterns of the cpDNA from a few species, providing the basis for a molecular study of the phylogeny in that genus and for the analysis of cytoplasmic variability within species.

Materials and methods

Material

Wild species and sunflower cultivars were grown under greenhouse conditions. *H. annuus* male fertile line HA89 or the F1 hybrid (HA89 ♀ × RHA274 ♂), and *H. tuberosus* (Jerusalem artichoke) variety ID19 were obtained respectively as seeds (from Dr Bervillé, INRA Dijon) or tubers (from Pr. Courduroux, University of Clermont-Ferrand). The wild perennial species *H. occidentalis* ssp. *plantagineus* (INRA reference 231), *H. decapetalus* (ref. 551), *H. maximiliani* (ref. 568), *H. grossesseratus* (ref. 237) and *H. giganteus* (ref. 554) were obtained as young plants from Dr Serieys (INRA Montpellier).

Chloroplast DNA isolation

Chloroplasts were extracted from young (2 to 4 weeks old) leaves after homogenization with a Waring blendor in 5 volumes of 25 mM Tris, pH 8.0, 25 mM EDTA, 12% sucrose (w/w), filtration through 50-μm mesh nylon cloth, sedimentation at 2000 *g* for 15 min, and flotation in 50% sucrose (w/w) at 20000 *g* for 60 min according to Manning and Richards [20]. After final washing and sedimentation in 12% sucrose, the chloroplast pellets were resuspended in 20 mM Tris, pH 7.5, 20 mM EDTA, 3% (v/v) SDS and deproteinized by chloroform/phenol treatments.

Restriction enzyme analysis

The cpDNAs were digested with the restriction enzymes as recommended by the supplier (Boehringer Mannheim). The restriction fragments were separated by horizontal agarose gel electrophoresis, photographed under UV illumination after ethidium

bromide staining, and transferred to Schleicher and Schuell BA85 nitrocellulose filters according to Southern [36].

Hybridizations

Hybridizations with labelled probes were performed according to Jeffreys and Flavell [13], in Denhardt's medium containing 3 × SSC, 40% formamide, at 43 °C; washings were made in 2 × to 0.1 × SSC containing 0.1% SDS, at 63 °C.

cpDNA-specific probes

The following cloned chloroplast DNA probes were used in this work:

- *Nicotiana tabacum* fragments Sal S4 to S11 were kindly provided by Dr P. Seyer [32];
- *Triticum aestivum* fragments Sal S3b to S11 except for S5 [5] were made available by Dr A. Rode;
- *Euglena gracilis* restriction fragments EcoN, EcoP and EcoL containing respectively genes rps7 and rps12, 16S-rDNA and 23S-rDNA [14, 21] had been cloned in a previous work in our laboratory;
- *Spinacia oleracea* specific probes for genes psaA1, petA, psbE/F, psbA, psbB, atpA, atpE, atpH, psbC and psbD were provided by Prof. R. G. Herrmann (see references in Table 3);
- rbcL gene [34] was subcloned from *N. tabacum* Sal 6 fragment obtained from Dr P. Seyer;
- *Vicia faba* probes for trnT-trnE-trnY [15], trnL-UAA [4], trnQ [37] and ndhF (Herdenberger *et al.*, unpublished).

The various double stranded probes of cloned cpDNA fragments from tobacco, wheat, spinach, *Euglena* or broad bean were labelled by nick-translation according to Rigby *et al.* [30]; excision of the inserts from the plasmid vectors were found unnecessary since neither pBR322 nor pUC9 plasmids gave significant hybridization with cpDNA.

The M13 single stranded ndhF radioactive probe was synthesized using the pentadecamer 5′ CGAATCGTGTAGGGC 3′ (internal to the gene) as primer in the presence of dNTPs (including ^{32}P α-dATP) and Klenow enzyme. Free radionucleotides were eliminated by gel filtration on Sephadex G50.

The trnL-UAA probe was labelled as riboprobe by transcription of the gene cloned in the vector pT7-2. The reaction mixture (final volume 50 μl) was 400 mM Tris-HCl, pH 8.0, 15 mM $MgCl_2$, 5 mM DTT, 1 mg/ml BSA, and contained 50 nmol of each of the three (CTP, UTP and GTP) unlabelled ribonucleotides (1 mM final concentration), 20 μCi ^{32}P α-ATP (400 Ci/nmol), 0.5 μg pT7-Leu(UAA) linearized with Eco RI, 30 units RNasin and 4 units T7 RNA polymerase. Incubation was for 45 minutes at 37 °C. To complete the chains, 50 nmol unlabelled ATP (2.5 μl of a 20 mM solution) was added and the mixture was incubated for 75 min at 37 °C. After phenol extraction and ethanol precipitation the labelled RNA was redissolved in 50 μl H_2O.

The following synthetic oligonucleotides were used for the detection of three tRNA genes:

trnR (UCU): 5′ AGGTTTAGAAGACCTCTGTCCTA 3′
trnS (GCU): 5′ CAACGGATTAGCAATCCGCCGC 3′
trnG (UCC): 5′ GCATCGTTAGCTTGGAAGGC 3′

The oligonucleotides (3 μg of each) were 5′-labelled using polynucleotide kinase in the presence of 50 μCi ^{32}P γ-ATP. Unincorporated radioactivity was eliminated by gel filtration (Sephadex G25).

Cloning of H. annuus *chloroplast DNA*

A Hind III cpDNA fragment library was constructed using phage λL47 [17], since this vector can accommodate the largest fragment (15.5 kbp) produced by Hind III cleavage [19]. Clones of appropriate tobacco fragments have been used to screen out of this library sunflower fragments Hind H1, H2, H3, H4a, H4b, H4′, H6, H6′ and H13. The fragments have been subcloned in plasmid vectors pBR322 [3] or pEMBL8− [6].

In the following lines, we will use the gene nomen-

clature proposed by Hallick [8] and we will refer to nucleotide positions deduced from the cpDNA tobacco sequence published by Shinozaki *et al.* [35].

Results and discussion

Restriction mapping of H. annuus *chloroplast DNA*

Five restriction enzymes (BamH I, Hind III, Pst I, Pvu II and Sac I), chosen for the low or moderate number of restriction fragments they produce, have been used for the construction of the physical map of the sunflower plastome. The restriction patterns obtained with these enzymes are shown in Fig. 1, and the nomenclature and size of the fragments are presented in Table 1. The sum of the fragment sizes generated by the restriction enzymes shows that the total length of the cpDNA is about 152 kbp. The lower value obtained with BamH I (147.1 kbp) is due to the large number of small BamH I restriction fragments that go undetected on 0.7% agarose gels; some of these fragments have been demonstrated by fractionation on overloaded 1.2% gels (Fig. 1b) and by BamH I digestion of cloned Hind III fragments (for instance fragments Hind H3 and H4a, Fig. 2b). These small fragments were not numbered and have not been taken into account in the proposed molecular weight of cpDNA when digested with BamH I.

We had at our disposal a continuous array of cloned cpDNA fragments corresponding to the sequences spanning from nucleotides 50000 to 156000 on the physical map of tobacco plastome [35]. They include tobacco fragments Sal S4, S6, S7, S8, S9, S10 and S11 and wheat fragments Sal S8 and S3b (the latter corresponding to the SSC region). These probes allowed us to identify and to order the restriction fragments of sunflower cpDNA in the corresponding region (Table 2A and Fig. 2). No

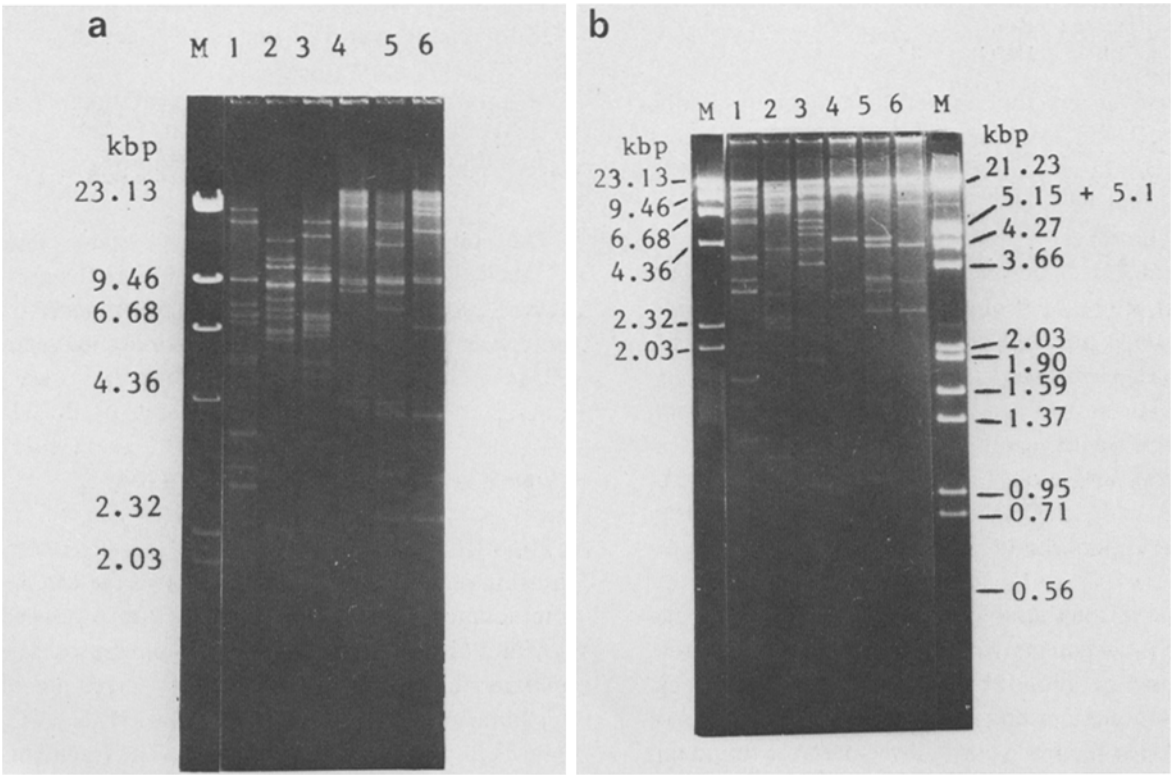

Fig. 1. Restriction patterns of sunflower cpDNA generated by digestion with BamH I (1), Hind III (2), Sac I (3), Pst I (4), Pst I + Pvu II (5) and Pvu II (6). Fragments were separated by electrophoresis for 15 h at 2 V/cm in a 0.7% (a) or a 1.2% (b) agarose gel.
M: lambda DNA digested with Hind III (on the left) or Hind III + Eco RI (on the right), used as molecular weight markers.

Table 1. Sizes (in kbp) and stoichiometries (in brackets) of sunflower cpDNA restriction fragments produced by BamH I, Hind III, Sac I, Pst I, Pvu II and Pst I + Pvu II digestions

BamH I		Hind III		Sac I		Pst I*		Pvu II*		Pst I + Pvu II	
	kbp		kbp		kbp		kbp		kbp		kbp
B1	25.5	H1	15.5	S1	19.5	Ps1	40.5	Pv1	37.5	Ps3	20.0
B2	19.4	H2	12.0	S2	16.0	Ps2	28.0	Pv2	28.0	A	18.5
B3(2)	9.4	H3(2)	10.5	S3	10.5	Ps3	20.0	Pv2′	23.5	B	14.5
B4	7.9	H4(3)	9.0	S4	9.8	Ps3′	19.0	Pv3	17.0	Pv4	9.6
B5	7.4	H4′	8.9	S4′(3)	9.2	Ps4	9.5	Pv4	9.6	Pv4′ + C	9.5
B6	7.0	H5	8.2	S5	7.4	Ps4′	9.2	Pv4′	9.5	C′	9.4
B7	5.6	H6	7.7	S5′	7.3	Ps5	8.8	Pv5	7.0	Ps5	8.8
B7′	5.5	H6′	7.3	S6	7.0	Ps6(2)	4.3	Pv6a(2)	4.1	D	7.8
B8(3)	3.7	H7	6.7	S6′	6.8	Ps6′	4.2	Pv6b	4.05	Ps6(2)	4.3
B9(2)	2.95	H7′	6.6	S7	6.5	Ps7	2.7	Pv7	3.0	Ps6′	4.2
B10(4)	2.85	H8	6.3	S8	5.5	Ps8	<1	Pv8(2)	2.55	Pv6a(2)	4.1
B11	2.45	H9	3.95	S9	5.0					Pv6b	4.05
B12	2.15	H10	2.75	S10	4.55					E (2)	3.1
B13	2.05	H11a	2.4	S11	3.9					Pv7	3.0
B14	1.9	H11b	2.4	S12(2)	3.5					Ps7	2.7
B15(3)	1.75	H12(2)	2.3	S13	2.4					Pv8(2)	2.55
B16	1.55	H13	1.7	S14(2)	1.95					F	1.5
B17a(2)	1.35	H14a(2)	1.25							Ps8	<1
B17b	1.3	H14b	1.2								
B18	1.15	H14c	1.15								
B19	1.0	H15(2)	0.6								
Sum: (kbp)	147.1		151.0		150.65		151.0		152.5		152.0

* The molecular weights of restriction fragments Ps1, Ps2, Pv1, Pv2 and Pv2′ have been deduced from the double digestion with Pst I + Pvu II.

hybridization signal was obtained that could be considered as inconsistent with the hypothesis of colinearity between the tobacco and sunflower cpDNAs.

For the region spanning from nucleotide positions 1 to 50000 on the sunflower cpDNA, we did not have tobacco clones at our disposal. We used instead wheat fragments Sal S4, S11, S9, S10, S6 and S7. But since there is an inversion between cpDNA from wheat and from most other dicots in this area [27], these probes could not provide the rigorous arrangement of sunflower fragments in that region: they were used only to determine the sequence homologies between the sunflower cpDNA fragments produced by the various restriction enzymes (Table 2B).

A complete map was proposed for the large Pst I and Pvu II fragments, taking into account (a) the relative positions of Pst I and Pvu II sites demonstrated by double digestion and (b) the homologies existing between these fragments and the wheat probes. Most of the restriction fragments produced by BamH I, Hind III and Sac I could subsequently be placed on this Pst I + Pvu II map (Fig. 2).

In addition, sunflower cpDNA fragments Hind H1, H2, H3, H4a, H4b, H4′, H6, H6′ and H13 were screened out of a L47 library; the use of these clones allowed us to precisely localize the restriction site for all five enzymes in the corresponding regions.

Mapping of defined genes on the sunflower chloroplast genome

The genes coding for 15 protiens, the 16S- and 23S-

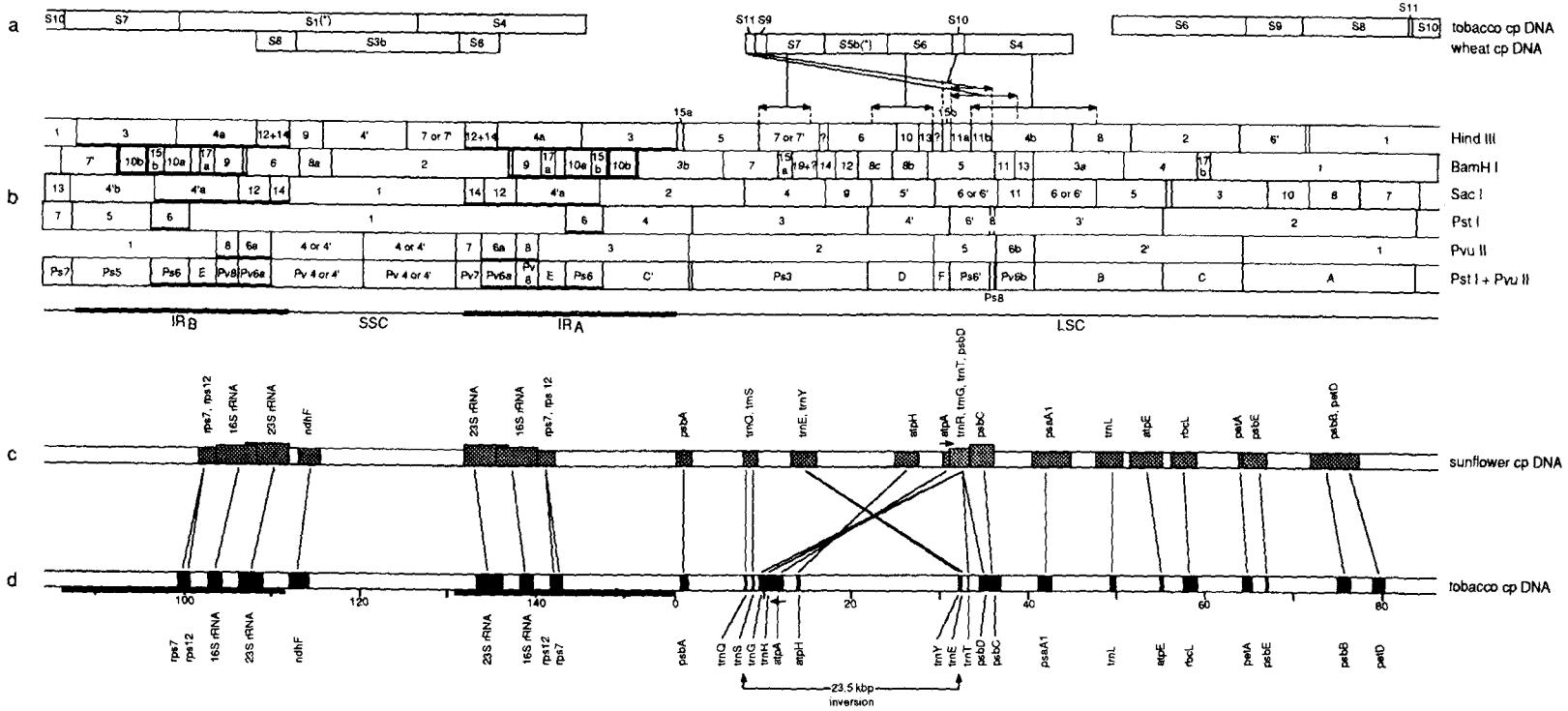

Fig. 2. Physical restriction map of sunflower (*H. annuus*) cpDNA showing the location of genes coding for chloroplast proteins or RNAs.

(a) Map of the tobacco or of the wheat restriction fragments used as probes [32, 5]. Vertical lines join homologous regions wheat cpDNA and sunflower cpDNA (– –), according to results from Table 2B.

*: Tobacco (T) Sal I or Wheat (W) Sal I probes not available.

(b) Restriction map of sunflower cpDNA. The fragment nomenclature is the same as that used in Table 1. The thick black lines represent the minimal extent of the inverted repeats (IR_A and IR_B).

(c) Gene map of the sunflower cpDNA. Gene location is deduced from Table 3. The stippled regions represent the most accurate positions for these genes.

(d) Gene map of tobacco cpDNA (from Shinozaki *et al.* [35]).

(c and d) Comparison of gene maps of sunflower and tobacco cpDNAs: evidence for an inversion of 23.5 kbp and location of its endpoints. Vertical lines join homologous genes between the two maps; the inversion is evidenced by lines crossing each other.

Table 2. Summary of hybridizations with Tobacco (T) or Wheat (W) cloned Sal I restriction fragments

Table 2A:

Probes		Filter-bound fragments hybridized				
	kbp	BamH I	Hind III	Pst I	Pvu II	Sac I
T. Sal6	15.2	B1-B3a(w)-B4-B17b	H2-H6'-H8	Ps2-Ps3'	Pv2'	S3-S5
T. Sal9	5.6	B1	H6'	Ps2	Pv1	S3-S10
T. Sal8	11.4	B1	H1-H6'(w)	Ps2	Pv1	S7-S8-S10(w)
T. Sal11	0.65	B1	H1	Ps2	Pv1	S7
T. Sal10	2.9	B1-B7'	H1	Ps2(w)-Ps7	Pv1	S7-S13
T. Sal7	13.2	B3b-B7'-B10-B15b-*	H1(w)-H3-H15a(w)	Ps4-Ps5-Ps6-Ps7(w)	Pv1-Pv3	S2-S4'a and/or b-S13(w)
W. Sal8	4.5	B2(w)-B6	H7 or 7'(w)-H9-(w)-H12-H14	Ps1	Pv4 or 4'-Pv6a-Pv7	S12-S14
W. Sal3b	13.6	B2-B6(w)-B8a	H4'-H7 or 7'-H9	Ps1	Pv4 and/or 4'	S1
T. Sal4	19.7	B2-B6-B9-B10-B17a-*	H4 and/or 4'-H7 or 7' (w)-H9(w)-H12-H14-H3(w)	Ps1-Ps6	Pv1(w)-Pv3-Pv4 or 4'-Pv6a-Pv7(w)-Pv8	S1(w)-S4'-S12-S14

Tabel 2B:

Probes		Filter-bound fragments hybridized				
	kbp	Pst I	Pvu II	BamH I	Hind III	Sac I
W. Sal4	11.8	Ps3'-Ps6'(w)	Pv2'-Pv5(w)-Pv6b	B3a-B5(w)-B11(w)-B13	H4b-H8-H11(w)	S6 and/or 6'-S11
W. Sal11	0.8	Ps3'-Ps6'	Pv5-Pv6b	B5-B11	H4b-H11	S6 or 6'-S11
W. Sal9($)	1.2	Ps1(w)-Ps6'	Pv5-Pv6a(w)-Pv8(w)	B5-B9(w)	H4a(w)-H11 or 12	S4'a(w)-S6 or 6'-S12(w)-S14(w)
W. Sal10	1.1	Ps4'-Ps6'(w)	Pv5	B5	H15b	S6 or 6'
W. Sal6	6.8	Ps4'	Pv2-Pv5(w)	B5(w)-B8b	H6-H10-H13(w)	S5'
W. Sal7	6.1	Ps3	Pv2	B7-B15a-*	H7 or 7'	S4

* Hybridization to low MW (<1 kbp) fragments.
(w): weak hybridization.
($): W. Sal9 probe gives low hybridization signals in the IR regions.

In Table 2A, probes are arranged in the same way as they are on the restriction map to tobacco cpDNA (see Seyer *et al.* [22]). W. Sal8 and W. Sal3b probes correspond to the IR-SSC junctions and to the SSC region respectively (see Bowman et al. [5]). It was assumed that tobacco and wheat cpDNAs are colinear in the SSC and IR regions.
In Table 2B, wheat cpDNA probes are arranged in order to reveal the relative organization of Pst I and Pvu II restriction fragments on the sunflower cpDNA map.

rRNAs and 8 tRNAs have been localized on the sunflower cpDNA physical map presented above, using heterologous hybridization with various cloned gene probes from spinach, tobacco, broad bean, and *Euglena*, as well as with synthetic oligonucleotide probes (see Materials and methods, and Table 3). Some of these gene probes (atpA, atpH, psbC, psbD, trnQ, trnS, trnG, trnR, trnT and trnT-trnE-trnY) allowed us to ascertain the position of small BamH I, Hind III and Sac I fragments, and to demonstrate the occurrence of an inversion between plastomes of sunflower and of other dicots, in the region spanning between nucleotide positions 1 to 50000.

Homologies between *H. annuus* and the various probes used in this work are listed in Table 3 and reported on the physical map of sunflower cpDNA

Table 3. Summary of gene mapping hybridizations

Gene probes				Filter-bound fragments hybridized				
Gene	Region cloned	Source species	Ref. (a)	BamH I	Hind III	Pst II	Pvu II	Sac I
atpA	end of the gene	Spinach	(1)	B5	H11a	Ps4'(w)-Ps6'	Pv5	S6 or 6'
atpA	beginning of gene	Spinach	(1)	B5	H15b	Ps4'	Pv5	S6 or 6'
atpE		Spinach	(1)	B4	H2	Ps3'	Pv2'	S3(w)-S5
atpH	AA2 to AA62	Spinach	(1)	B8b	H10	Ps4'	Pv2	S5'
ndhF	AA180 to AA360	Broad bean	(2)	B8a	H9	Ps1	Pv4 or 4'	S1
petA	AA86 to AA244	Spinach	1	B1	H6'	Ps2	Pv2'	S3
petD	AA101 to AA192	Spinach	9	B1	H1	Ps2	Pv1	S8
psaA1	AA17 to AA382	Spinach	(1)	B3a	H4b	Ps3'	Pv2'	S6 or 6'
psbA	AA87 to AA341	Spinach	(1)	B3b	H5-H15a	Ps4	Pv2(w)-Pv3	S2
psbB	AA87 to AA471	Spinach	(1)	B1	H1	Ps2	Pv1	S8
psbC	AA52 to AA330	Spinach	2	B5	H4b(w)-H11b	Ps6'-Ps8	Pv5	S6 or 6'
psbD	-68 to AA306	Spinach	2	B5	H11b	Ps6'	Pv5	S6 or 6'
psbE/F	AA5 to AA39 + 28 nucleotides	Spinach	10	B1	H6'	Ps2	Pv1	S3
rbcL	-34 to AA298	Tobacco	34	B4	H2	Ps2	Pv2'	S3
rps7-rps12	Restriction fragment EcoN	*Euglena gracilis*	21	B9(w)-B17a	H4a	Ps1	Pv1-Pv3	S4'a
16S-rRNA	Restriction fragment EcoP	*Euglena gracilis*	14	B2(w)-B6(w)-B9	H4a	Ps1	Pv6a-Pv8-Pv5?	S4'a-S12(w)
23S-rRNA	Restriction fragment EcoL	*Euglena gracilis*	14	B2-B6	H4a(w)-H12-H14	Ps1	?	S12-S14
trnE-UUC + trnY-GUA ($)	Part of restr.frag. Bam19: 875 bp	Broad bean	15	B19	H7 or 7'	Ps3	Pv2	S4
trnG-UCC		Synth. oligonuc.		B5	H11a or 11b	Ps6'	Pv5	S6 or 6'
trnL-UAA		Broad bean	4	B3a	H8	Ps3'	Pv2'	S5
trnQ-UUG	Restr.frag. Bam16: 2.8 kbp	Broad bean	37	B7	H5	Ps3	Pv2	S4
trnR-UCU		Synth. oligonuc.		B5	H11a or 11b	Ps6'	Pv5	S6 or 6'
trnS-GCU		Synth. oligonuc.		B7	H5	Ps3	Pv2	S4
trnT-GGU ($)	Part of restr.frag. Bam19: 1139 bp	Broad bean	15	B5	H11a	Ps6'	Pv5	S6 or 6'

(a): (1) Hermann *et al.*, unpublished. (2) Herdenberger *et al.*, unpublished. 1, 2, 4, 9, 10, 14, 15, 21, 34, 37: see References.
(w): weak hybridization
($): trnE-trnY probes gave hybridization signals in the IR regions: Bam B2, B6 and B9, Hind H4a, H12 and H14, Pst Ps1, Pvu Pv6a and Pv8, Sac S4'a, S12 and S14 were hybridized.
Synth. oligonuc.: Synthetic oligonucleotide

(Fig. 2). As expected, sunflower and tobacco cpDNA appear to be essentially colinear for the regions extending from $psaA_1$ (nucleotide position 42000 on tobacco map) through IR_B, SSC and IR_A to psbA (nucleotide position 900).

However, the arrangement of the genes for trnT and trnE-trnY on the sunflower chloroplast genome differs from that observed in all other higher plants studied: in tobacco [35], broad bean [15], spinach [11], pea [28, 29] and liverwort [22], the genes trnT

and trnE-trnY are located in close proximity, separated by only 800 to 1000 bp, on the two opposite strands of DNA. In wheat chloroplasts, the three genes are clustered on the same DNA strand [27]. The cloned DNA probe containing these clustered genes from broad bean hybridized to sunflower restriction fragments Bam B5 and B19, located far apart from each other on our physical restriction map (i.e. about 20–25 kbp). Hybridizations with subcloned probes of broad bean trnE-trnY on the one hand and trnT on the other hand confirmed that the two groups of genes are not contiguous in the sunflower plastome, but hybridize respectively to Bam B19 (trnE-trnY) and to Bam B5 (trnT).

In the same way, tobacco genes for trnQ, trnS, trnG and trnR form a cluster of less than 4 kbp around position 10000 [20]. They hybridize to two distinct distant regions in sunflower cpDNA, represented by fragments Bam B7 and Bam B5 for instance (see Fig. 2).

Inversely, spinach and tobacco genes atpA on the one hand and psbC-psbD on the other hand, are separated from one another by about 24 kbp; they hybridize to the same unique region of sunflower cpDNA contained in Bam B5 (7.4 kbp) or Pvu Pv5 (7 kbp) fragments. Figure 3 shows autoradiographs of hybridizations of sunflower cpDNA digests with probes of trnS, trnG and trnR. This new gene arrangement in sunflower can be explained by an inversion of the region extending from trnG(UCC) to trnE (these two genes representing the endpoints of the inversion) of the tobacco genome (Fig. 2). In fact, one endpoint is located in the approximately 850-bp-long region separating trnT(GGU) and trnE(UUC) (about nucleotide position 32750 on the tobacco map) and the second endpoint lies in the 870-bp-long region separating trnS(GCU) and trnG(UCC) (i.e. *upstream* from trnG) (position 9150 on the tobacco map). This inversion therefore spans about 23.5 kbp. A further indication for the existence of this inversion is given by the fact that the atpA gene orientation could be determined using the

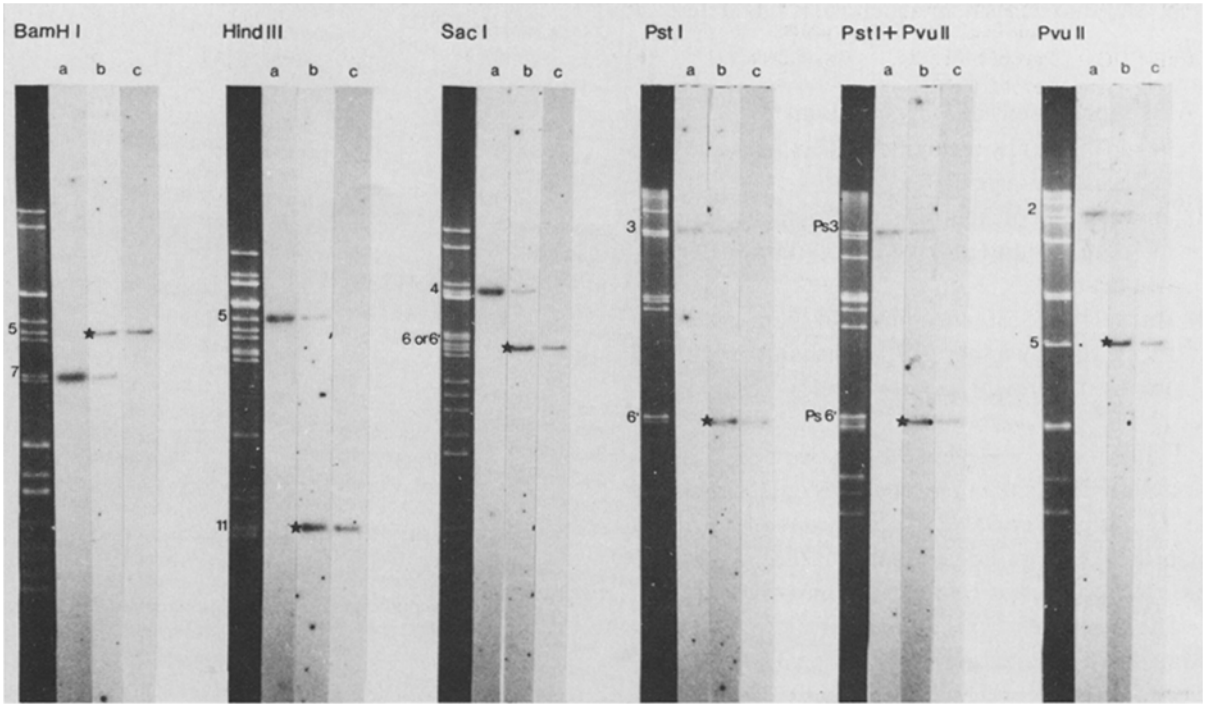

Fig. 3. Hybridization of specific probes for trnS (a), trnG (b) and trnR (c) genes (see Materials and methods) to restriction fragments sunflower cpDNA generated by digestions with BamH I, Hind III, Sac I, Pst I, Pst I + Pvu II and Pvu II. Restriction patterns are the left and the corresponding autoradiographs on the right.

* Specific signal with trnG probe; other signals are due to residual hybridization previously performed with trnS probe on the same filter.

two available probes, one corresponding to the beginning of the gene ($atpA_2$) and the other to the end of the gene ($atpA_1$). An inversion involving trnG was also observed in wheat chloroplast genome; however the endpoint in this case lies between trnG and trnR (i.e. *downstream* from trnG) [27].

Inversions have also been described or suggested in two other composites. In lettuce cpDNA, Jansen and Palmer [12] reported a large inversion with the same endpoints as those demonstrated here for sunflower cpDNA. In safflower cpDNA, the atpA gene is located around nucleotide position 30000 [18], whereas its position is 10625 in tobacco cpDNA. Therefore, the Compositae family could be characterized by an inversion involving the atpA gene.

Comparative analysis of the structure of H. annuus *and* H. tuberosus *chloroplast DNAs*

The comparison of BamH I cpDNA restriction patterns from *H. annuus* and *H. tuberosus* shows (Fig. 4A) that Jerusalem artichoke is characterized by:

1) the replacement of fragment Bam B1 (25.5 kbp in sunflower) by a fragment Bam B1* (21.5 kbp in Jerusalem artichoke);
2) an increase of the size of fragment Bam B7 (5.6 kbp in sunflower versus 5.7 kbp in Jerusalem artichoke);
3) the existence of an additional Bam B8* fragment, slightly larger (3.7 kbp) than the Bam B8 fragments present in both species.

Points 1 and 3 suggest the existence of an additional BamH I site in *H. tuberosus* cpDNA relative to *H. annuus* cpDNA, since the added sizes of Bam B1* + Bam B8* fragments (21.5 + 3.8 kbp) are close to that of Bam B1 fragment (25.5 kbp).

The existence of this additional site was demonstrated by hybridization of the two cpDNAs with the cloned sunflower fragment H6′. Only Bam B1 fragment hybridizes in *H. annuus* cpDNA, while fragments Bam B1* and Bam B8* hybridize in *H. tuberosus* (Fig. 4B).

The latter structural variation is found in several species of the genus *Helianthus* (results not shown). While *H. occidentalis* ssp. *plantagineus* presents a BamH I restriction pattern identical to that of *H. annuus*, other species like *H. grossesseratus, H. decapetalus, H. giganteus* and *H. maximiliani* give the same BamH I patterns as *H. tuberosus*.

Clones of this variable region, as well as others un-

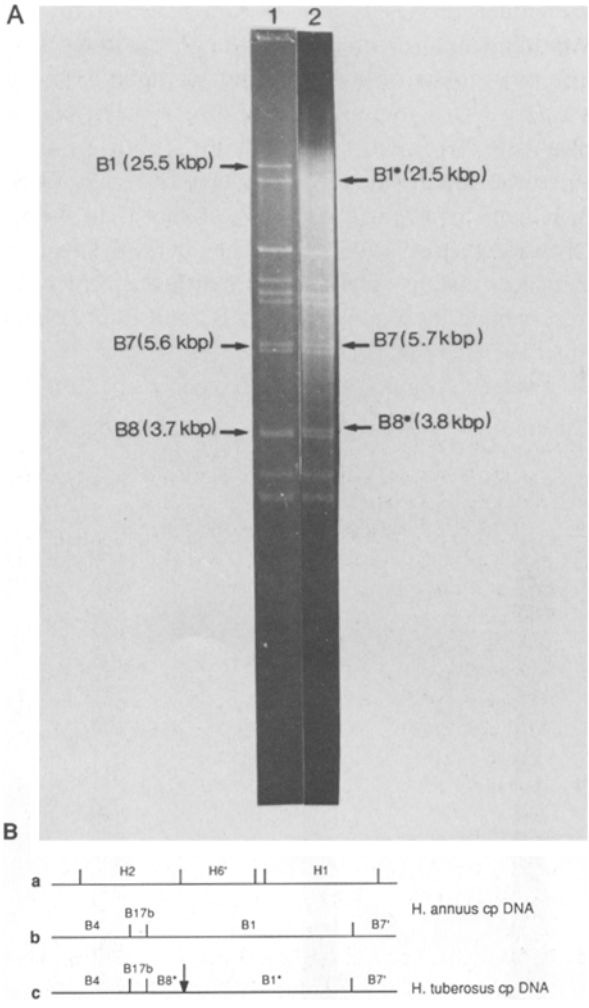

Fig. 4. Comparative analysis of *H. annuus* and *H. tuberosus* cpDNAs.

(A) BamH I restriction patterns of *H. annuus* (1) and *H. tuberosus* (2) cpDNAs. Fragment nomenclature as described in the text.

(B) BamH I restriction maps of *H. annuus* (b) and *H. tuberosus* (c) cpDNAs in the region containing restriction fragments Hind H2, H6′ and H1 (a): presence of an additional BamH I site (↓) on *H. tuberosus* cpDNA.

der investigation, will be used to propose a molecular phylogeny of the cytoplasms in the genus *Helianthus*. They should also allow a molecular discrimination of various cytoplasms within species or subspecies, already shown to differ by their nucleo-cytoplasmic behaviors.

Acknowledgements

We thank Pr. Courduroux for his gift of Jerusalem artichoke tubers, Drs A. Bervillé, H. Serieys and P. Vincourt for supplying *Helianthus* seeds or wild plants. We are grateful to Drs A. Rode, P. Seyer and Pr. R. G. Herrmann for wheat, tobacco and spinach probes. This work was supported by grants from the CNRS and the INRA.

References

1. Alt J, Herrmann RG: Nucleotide sequence of the gene for preapocytochrome *f* in the spinach plastid chromosome. Current Genetics 8: 551–557 (1984).
2. Alt J, Morris J, Westhoff P, Herrmann RG: Nucleotide sequence of the clustered genes for the 44 kd chlorophyll *a* apoprotein and the "32-kd"-like protein of the photosystem II reaction center in the spinach plastid chromosome. Current Genetics 8: 597–606 (1984).
3. Bolivar F, Rodriguez RL, Greene PJ, Betlach MC, Heynecker HL, Boyer HW, Crosa JH, Falkow S: Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 2: 95–113 (1977).
4. Bonnard G, Michel F, Weil JH, Steinmetz A: Nucleotide sequence of the split $tRNA^{Leu}$(UAA) gene from *Vicia faba* chloroplasts: evidence for structural homologies of the chloroplast $tRNA^{Leu}$ intron with the intron from the *Tetrahymena* ribosomal RNA precursor. Mol Gen Genet 194: 330–336 (1984).
5. Bowman CM, Koller B, Delius H, Dyer TA: A physical map of wheat chloroplast DNA showing the location of the structural genes for the ribosomal RNAs and the large subunit of ribulose-1,5-bisphosphate carboxylase. Mol Gen Genet 183: 93–101 (1981).
6. Dente L, Cesareni G, Cortese R: pEMBL: a new family of single stranded plasmids. Nucleic Acids Res 11: 1645–1655 (1983).
7. Fluhr R, Edelman M: Conservation of sequence arrangements among higher plant chloroplast DNAs: Molecular cross-hybridization among the Solanaceae and between *Nicotiana* and *Spinacia*. Nucleic Acids Res 9: 6841–6855 (1981).
8. Hallick RB, Bottomley W: Proposal for the naming of chloroplast genes. Plant Mol Biol Reporter 1: 38–43 (1983).
9. Heinemeyer W, Alt J, Herrmann RG: Nucleotide sequence of the clustered genes for apocytochrome *b*-6 and subunit 4 of the cytochrome *b/f* complex in the spinach plastid chromosome. Current Genetics 8: 543–549 (1984).
10. Herrmann RG, Alt J, Schiller B, Widger WR, Cramer WA: Nucleotide sequence of the gene for apocytochrome *b*-559 on the spinach plastid chromosome: Implications for the structure of the membrane protein. FEBS Letters 176: 239–244 (1984).
11. Holschuh K, Bottomley W, Whitfeld PR: Organization and nucleotide sequence of the genes for spinach chloroplast $tRNA^{Glu}$ and $tRNA^{Tyr}$. Plant Mol Biol 3: 313–317 (1984).
12. Jansen RK, Palmer JD: Chloroplast DNA from lettuce and *Barnadesia* (Asteraceae): structure, gene localization and characterization of a large inversion. Current Genetics 11: 553–564 (1987).
13. Jeffreys AJ, Flavell RA: A physical map of the DNA region flanking the rabbit globin gene. Cell 12: 429–439 (1977).
14. Jenni B, Stutz E: Physical mapping of the ribosomal DNA region of *Euglena gracilis* chloroplast DNA. Eur J Biochem 88: 127–134 (1978).
15. Kuntz M, Weil JH, Steinmetz A: Nucleotide sequence of a 2-kbp BamH I fragment of *Vicia faba* chloroplast DNA containing the genes for threonine, glutamic acid and tyrosine transfer RNAs. Nucleic Acids Res 12: 5037–5047 (1984).
16. Leclercq P: Une stérilité mâle cytoplasmique chez le tournesol. Ann Amélior Plantes (Paris) 19: 99–106 (1969).
17. Loenen WA, Brammar WJ: A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. Gene 10: 249–259 (1980).
18. Ma C, Smith MA: Construction and mapping of safflower chloroplast DNA recombinants and location of selected gene markers. Theor Appl Genet 70: 620–627 (1985).
19. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: a Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory (1982).
20. Manning JE, Richards OC: Isolation and molecular weight of circular chloroplast DNA from *Euglena gracilis*. Biochim Biophys Acta 259: 285–296 (1972).
21. Montandon PE, Stutz E: The genes for the ribosomal proteins S12 and S7 are clustered with the gene for the EF-Tu protein on the chloroplast genome of *Euglena gracilis*. Nucleic Acids Res 12: 2851–2859 (1984).
22. Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeushi M, Chang Z, Aota S, Inokushi H, Ozeki H: Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322: 572–574 (1986).
23. Palmer JD, Nugent JM, Herbon LA: Unusual structure of Geranium chloroplast DNA: a triple-sized inverted repeat, extensive gene duplication, multiple inversions and two repeat families. Proc Natl Acad Sci USA 84: 769–773 (1987).
24. Palmer JD, Shields CR, Cohen DB, Orton TJ: Chloroplast DNA evolution and the origin of amphiploid *Brassica* species. Theor Appl Genet 65: 181–189 (1983).

25. Palmer JD, Thompson WF: Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell 29: 537–550 (1982).
26. Phillips AL: Restriction map and clone bank of tomato plastid DNA. Current Genetics 12: 147–152 (1985).
27. Quigley F, Weil JH: Organization and sequence of five tRNA genes and of an unidentified reading frame in the wheat chloroplast genome: evidence for gene rearrangements during the evolution of chloroplast genomes. Current Genetics 9: 495–503 (1985).
28. Rasmussen O, Jepsen B, Stummann B, Henningsen KW: Nucleotide sequence of the pea chloroplast $tRNA^{Thr}$(GGU) gene and flanking regions. Comparison to the *Vicia faba* sequence. Nucleic Acids Res 15: 854 (1987).
29. Rasmussen OF, Stummann B, Henningsen KW: Nucleotide sequence of a 1.1 kb fragment of the pea chloroplast genome containing three tRNA genes, one of which is located within an open reading frame of 91 codons. Nucleic Acids Res 12: 9143–9153 (1984).
30. Rigby PWJ, Dieckmann M, Rhodes C, Berg P: Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick-translation with DNA polymerase I. J Mol Biol 113: 237–251 (1977).
31. Schilling EE, Heiser CB: Infrageneric classification of *Helianthus* (Compositae). Taxon 30: 393–403 (1981).
32. Seyer P, Kowallik KV, Herrmann RG: A physical map of *Nicotiana tabacum* plastid including the location of structural genes for ribosomal RNAs and the large subunit of ribulose bisphosphate carboxylase oxygenase. Current Genetics 3: 189–204 (1981).
33. Shen GF, Chen K, Wu M, Kung SD: *Nicotiana* chloroplast genome. IV. *N. accuminata* has larger inverted repeats and genome size. Mol Gen Genet 187: 12–18 (1982).
34. Shinozaki K, Sugiura M: The nucleotide sequence of the tobacco chloroplast gene for the large subunit of ribulose-1,5-bisphosphate carboxylase oxygenase. Gene 20: 91–102 (1982).
35. Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamagushi-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: The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO Journal 5: 2043–2049 (1986).
36. Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98: 503–518 (1975).
37. Steinmetz A, Bonnard G, Kuntz M, Green GA, Mubumbila M, Crouse EJ, Weil JH: Organization and sequence of transfer RNA genes in the broad bean chloroplast genome. In: Steinback KE, Bonitz S, Arntzen CJ, Bogorad L (eds) Molecular Biology of the Photosynthetic Apparatus. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory (1985) pp. 279–284.
38. Van Ee JH, Vos YJ, Bohnert HJ, Planta RJ: Mapping of genes on the chloroplast DNA of *Spirodela oligorhiza*. Plant Mol Biol 1: 117–131 (1982).