

Jin-Seog Kim · Jong Duk Jung · Jung-Ae Lee ·
Hyun-Woo Park · Kwang-Hoon Oh ·
Won-Joong Jeong · Dong-Woog Choi · Jang Ryol Liu ·
Kwang Yun Cho

Complete sequence and organization of the cucumber (*Cucumis sativus* L. cv. Baekmibaekdadagi) chloroplast genome

Received: 29 September 2005 / Revised: 13 November 2005 / Accepted: 19 November 2005 / Published online: 9 December 2005
© Springer-Verlag 2005

Abstract The nucleotide sequence of the cucumber (*Cucumis sativus* L. cv. Baekmibaekdadagi) chloroplast genome was completed (DQ119058). The circular double-stranded DNA, consisting of 155,527 bp, contained a pair of inverted repeat regions (IRa and IRb) of 25,187 bp each, which were separated by small and large single copy regions of 86,879 and 18,274 bp, respectively. The presence and relative positions of 113 genes (76 peptide-encoding genes, 30 tRNA genes, four rRNA genes, and three conserved open reading frames) were identified. The major portion (55.76%) of the *C. sativus* chloroplast genome consisted of gene-coding regions (49.13% protein coding and 6.63% RNA regions; 27.81% LSC, 9.46% SSC and 18.49% IR regions), while intergenic spacers (including 20 introns)

made up 44.24%. The overall G-C content of *C. sativus* chloroplast genome was 36.95%. Sixteen genes contained one intron, while two genes had two introns. The expansion/contraction manner of IR at IRb/LSC and IR/SSC border in *Cucumis* was similar to that of *Lotus* and *Arabidopsis*, and the manner at IRa/LSC was similar to *Lotus* and *Nicotiana*. In total, 56 simple sequence repeats (more than 10 bases) were identified in the *C. sativus* chloroplast genome.

Keywords Cucumber (*Cucumis sativus*) · Chloroplast genome sequence · Fosmid library · Shotgun library

Communicated by I. S. Chung

J.-S. Kim (✉) · J.-A. Lee · K.-H. Oh · K. Y. Cho
Biological Function Research Team, Korea Research Institute
of Chemical Technology,
P.O Box 107, Yuseong, Daejeon, 305-600 Korea
e-mail: jskim@kriict.re.kr
Tel.: +82-42-860-7026
Fax: +82-42-861-4913
e-mail: kkabi97@hanmail.net
e-mail: khoh@kriict.re.kr
e-mail: kycho@kriict.re.kr

J. D. Jung · H.-W. Park
Eugentech Inc,
BVC 112, KRIBB, 52 Eoun-dong, Yuseong, Daejeon, 305-806
Korea
e-mail: seq1@eugentech.com
e-mail: dnahyun@eugentech.com

W.-J. Jeong · J. R. Liu
Plant Cell Biotechnology Laboratory, Korea Research Institute
of Bioscience and Biotechnology,
52 Eoun-dong, Yuseong, Daejeon, 305-806 Korea
e-mail: ultary@dreamwiz.com
e-mail: jrliu@kribb.re.kr

D.-W. Choi
Department of Biology Education, Chonnam National
University,
Kwangju, 500-757 Korea
e-mail: dwchoi63@chonnam.ac.kr

Introduction

Chloroplasts are plant-specific organelles that contain their own genetic system as well as the entire machinery necessary for the process of photosynthesis. Chloroplasts account for >50% of the total soluble protein in leaves, and these proteins are encoded by both the nuclear and chloroplast genomes. A number of reports have indicated that the expression of a set of nuclear genes that encode chloroplast-localized proteins is dependent on the functional state of the plastid (Surpin et al. 2002). Over the past decades, the plastid genome, its structure, expression, and evolution have been intensely studied using molecular methods (Adachi et al. 2000; Cato and Richardson 1996; Chu et al. 2004; Cosner et al. 2004; Goremykin et al. 2003, 2004; Jansen et al. 2005; Kim and Lee 2004; Martin et al. 2002; Matsuoka et al. 2002; Palmer et al. 1988; Palmer 1991; Powell et al. 1995; Raubeson and Jansen 2005; Wakasugi et al. 1998, 2001). There has been a rapid increase in our understanding of plastid genome organization and evolution, owing to the availability of many new completely sequenced genomes. There are 45 complete nucleotide sequences of chloroplast genomes published (Jansen et al. 2005; http://megasun.bch.umontreal.ca/ogmp/projects/other/cp_list.html).

Available sequence data have revealed the relatively conserved nature of chloroplast genomes with regard to both structure and gene content. Plastid genomes vary in size from 35 to 217 kb, but the vast majority from

photosynthetic organisms are between 115 and 165 kb. The 45 completely sequenced genomes encode from 63 to 209 genes with most containing 110–130 genes (Jansen et al. 2005). The presence of a large inverted repeat (IR), which ranges from 5 to 76 kb in length (Palmer 1991), is one of the conserved structural features. Much of the variation in length of the genomes can be attributed to the expansion/contraction/loss of the IR in addition to variation in length of intergenic spacers (Palmer et al. 1988; Wakasugi et al. 2001; Raubeson and Jansen 2005). The gene content and the polycistronic transcription units of the chloroplast genome are also largely conserved among most vascular plant species (Kim and Lee 2004), except in the case of the non-photosynthetic parasitic plant, *Epifagus virginiana*. The gene order of the chloroplast genome is also relatively conserved but is frequently reversed by inversion mutations that can be mediated by intra-molecular recombination events. Base substitution of chloroplast genes often occurs according to the position of the gene on the genome, and it is concentrated on intergenic spacers that lack polycistronic transcription units and genes in the IR regions (Kim and Lee 2004; Raubeson and Jansen 2005).

Cucumber is one of the most widely used plants for research into photosynthesis and greening physiology. In addition, cucumber is a good model plant for studying mitochondrial genetics because of the relatively large size of the genome (1500–2300 kb), paternal transmission, and the existence of unique mitochondrial mutations conditioning strongly mosaic phenotypes (Havey et al. 2002). It is possible to develop a cucumber transformed by an *Agrobacterium*-mediated (Szwacka et al. 1996; Yin et al. 2004) or biolistic method (Schulze et al. 1995). Considering recent improvements in transplastomic techniques, cucumber chloroplast transformation may also be possible in the near future, and nucleotide sequence information for the cucumber plastid would be very helpful in a number of research areas, including chloroplast genesis, the construction of expression cassettes for stable expression of heterologous proteins in chloroplast transformation (Maliga 2002), the development of crops having good agricultural traits (Bock and Khan 2004; Daniell et al. 2004), crosstalk in organelles (Surpin et al. 2002), and phylogenetic comparisons among relatively distantly related genera and families (Chu et al. 2004; Goremykin et al. 2003, 2004; Martin et al. 2002; Palmer et al. 1988).

In this study, we present the complete sequence and organization of the chloroplast genome identified through a series of processes: isolation of the chloroplast genome, cloning of several large size pieces of genome into fosmid vectors, construction of a shotgun library, mass sequencing, and annotation.

Materials and methods

Plant materials

Cucumber (*C. sativus* cv. Baekmibaekdadagi, Dongbu Hannong Chemical Co. Ltd. Seeds Division, Seoul, Korea;

<http://www.dongbuchem.com>) seeds were planted in a pot containing a synthetic soil (Bunong Horticulture Mix no. 5, Bio-Media Co. Ltd., Kyungju, South Korea; <http://www.bio-media.co.kr>). They were grown to the 8-leaf stage in a greenhouse at 25–35°C during the day and at 15–25°C during the night for 4 weeks.

Isolation of chloroplast DNA (cpDNA) from intact chloroplasts

Isolation of the chloroplast and cpDNA was conducted according to a modification of the method described by Triboush et al. (1998).

Cucumber plants were kept in the dark for 2 days at 25°C to reduce the starch level. For chloroplast extraction, three to five leaves from the shoot apex were chopped and ground in a Polytron PT2000 homogenizer (Kinematica Inc., Cincinnati, OH; <http://www.kinematica.ch>). All the subsequent operations were conducted at 4°C unless otherwise specified. The homogenate was filtered and centrifuged (Beckman J2-MC, Palo Alto, California, USA; <http://www.beckman.com>) at 300×g for 20 min. The supernatant was then re-centrifuged two times at 300×g for 20 min, and then at 5200×g for 20 min. DNase-I (Sigma; <http://www.sigmaaldrich.com>) was added to the chloroplast pellet, which was then re-suspended. The mixture was finally centrifuged at 5200×g for 20 min and the pellet kept at –20°C or lysed immediately.

The pellets were re-suspended and lysed as described by Triboush et al. (1998). Phenol/chloroform/isoamylalcohol deproteinization was then carried out. The DNA pellet was washed repeatedly with ethanol (70 and 98%), air dried, and re-dissolved in distilled water. RNase-A was dissolved in distilled water and added to the extracted DNA solution. The RNase-treated DNA solution was quantified for DNA content. Nuclear and mitochondrial DNA contamination was investigated by PCR using the primer sets *psbA-matK* (forward, acgggaaggggaaccgcgca; reverse, cc-gactagtccgggttcga) for cpDNA, the mitochondrial *coxI* gene (forward, ggtggaagtattaaatgacg; reverse, gagtcccttatggtgaaca) for mitochondrial DNA, and the ribosomal DNA spacer (forward, ctttcaagtgtcatggaggt; reverse, aa-catagtaggcattagcg) for nuclear DNA, respectively.

Fosmid library construction

The cpDNA was sheared into approximately 40-kb fragments by passage through a small-bore pipette tip. To estimate the size of the DNA, the sheared DNA was run on a 0.7% agarose gel. The sheared DNA was blunted and phosphorylated using a DNA End-Repair kit (Epicentre, Madison, WI, <http://www.epibio.com>) according to manufacturer's instructions. The end repaired DNA was fractionated by agarose gel electrophoresis, and DNAs of approximately 35 to 45 kb were excised from the gel and recovered using GELase. The recovered DNA was ligated to a pEpiFOS-5 vector at room temperature for 2 h

and packaged using MaxPlax Lambda Packaging Extracts (Epicentre, Madison, WI). Packaged fosmid clones were infected into EPI100-T1 cells and plated on a LB chloramphenicol plate. Colonies were randomly selected and cultured and fosmid DNAs purified using the alkaline lysis method. The titer for the fosmid library was 1×10^3 cfu ml⁻¹ and recombination percentage was 90%. To obtain both 5' and 3' sequences, DNA inserts were labeled with ET terminator dye (Amersham, Uppsala, Sweden; <http://www.amersham.com>) by PCR using pEpiFOS-5 forward and reverse sequencing primers. Sequencing reactions were separated on an ABI 377 sequencer (Applied Biosystems, CA, USA; <http://www.appliedbiosystems.com>). Both end sequences were aligned to the chloroplast full genome sequences of tobacco in the GenBank database. Five positive clones were selected covering the full genome sequence of the chloroplast.

Shotgun library construction

The above five positive fosmid clones were used to construct the shotgun library. Random DNA fragments were generated using the HydroShear process (GeneMachines, San Carlos, CA, <http://www.genemachines.com>). The fragmented DNA was fractionated with MicroSpin S-400 HR columns (Amersham, UK) and DNAs of approximately 0.5–1.5 kb were isolated from the sheared DNA. The fragments were ligated into a pUC118 plasmid vector that had previously been digested with *HincII* and treated with bacterial alkaline phosphatase. The ligated DNA samples were introduced into *Escherichia coli* DH5 α by electroporation.

Sequencing, gene annotation and comparative genome analysis

Clones obtained from each shotgun library were sequenced by the dideoxy chain termination method. Individual clones were transferred into deep well blocks containing 760 l of TB with 8% glycerol and 50 μ g ml⁻¹ ampicillin and grown overnight at 37°C, shaking at 600 rpm. Plasmid DNA was isolated using a plasmid preparation robot (HT prep machine, Bioneer, Korea; <http://www.bioneer.co.kr>). The 5' and 3' DNA sequences were determined using a capillary DNA sequencer (RISA 384, Shimadzu, Tokyo, Japan; <http://www.shimadzu.com>) and ABI 377 automatic sequencer (Applied Biosystems, USA), using the DYEnamic ET Terminator cycle sequencing kit (Amersham Pharmacia Biotech, USA; <http://www4.amershambiosciences.com>) and Big Dye 2.0TM Terminator cycle sequencing kit (Applied Biosystems, USA). The average read-length per shotgun library was 447 bp. The average sequenced length in each shotgun library was 169.6 kb (4.3 times). As a result of shotgun sequencing and assemblage, contigs covered approximately 97% of each fosmid. The remaining small gap and weak regions (21 regions) were amplified by PCR and sequenced directly from the amplification products. The sequence fragments were assembled using the Phrap

program (<http://www.phrap.com>). Eighteen complete plastid sequences of angiosperms were obtained from GenBank at the National Center for Biotechnology Information (NCBI). For gene annotations, open reading frames, ribosomal RNAs and transfer RNAs were predicted using TIGRscan program (<http://www.tigr.org>). Then, the extracted sequences were compared with reported chloroplast genes by BLAST homology search. Comparative genome analyses were performed using Stand-alone BLAST 2.2.8 (NCBI).

Results and discussion

Organization of the cucumber chloroplast genome

The nucleotide sequence of the cucumber (*Cucumis sativus* L. cv. Baekmibaekdadagi) chloroplast genome was completed (DQ119058). The total genome size was 155,527 bp. The circular double-stranded DNA had a pair of inverted repeat regions (IRa and IRb) of 25,187 bp each. The two IR regions were separated by a large single copy (LSC) region of 86,879 bp and a small single copy (SSC) region of 18,274 bp. One-hundred and thirteen genes were identified in the *C. sativus* chloroplast genome and their positions were shown in Fig. 1.

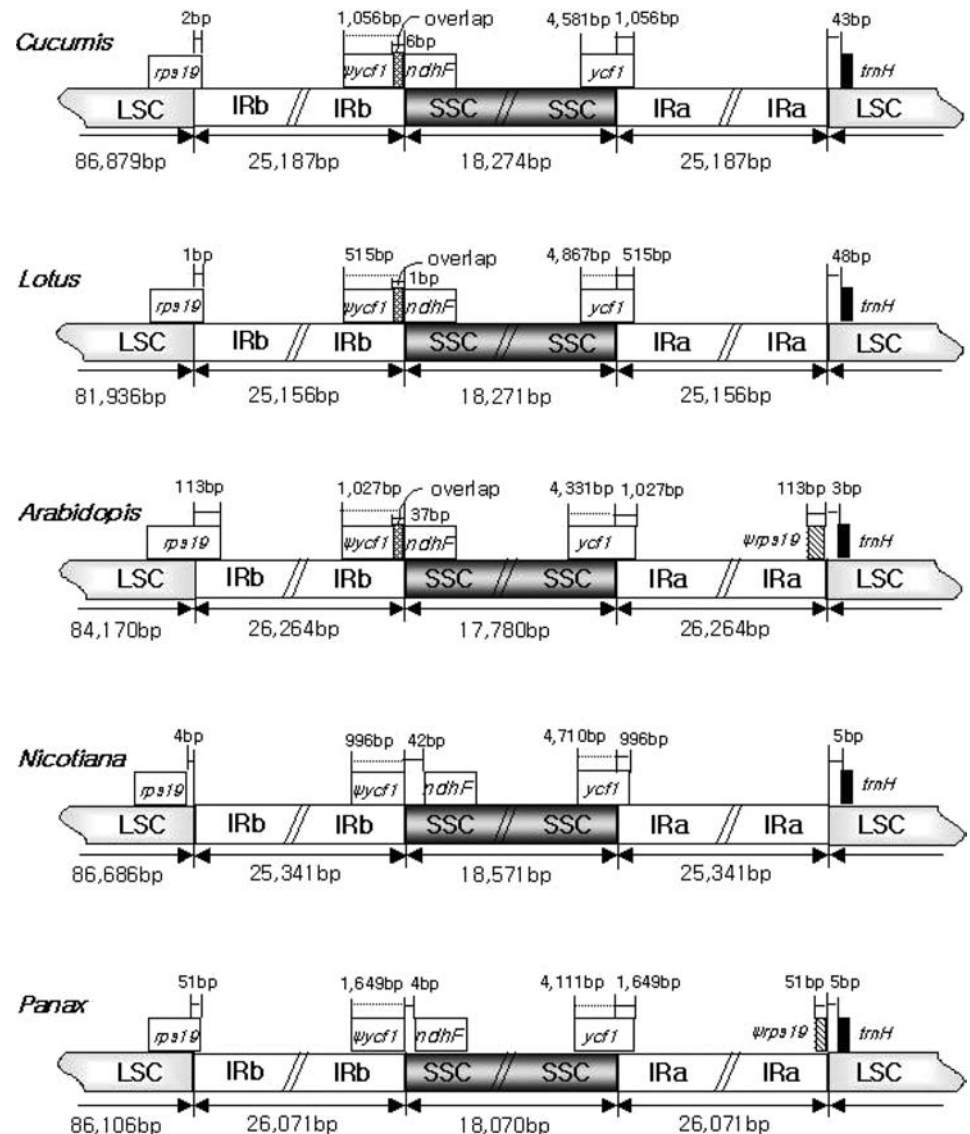
The major portion (55.76%) of the *C. sativus* chloroplast genome consisted of gene-coding regions (49.13% protein coding and 6.63% RNA regions; 27.81% LSC, 9.46% SSC, and 18.49% IR regions), whereas the intergenic spacers (including 20 introns) made up 44.24% (intron, 13.12%; spacer, 31.12%). The overall A-T and G-C contents of *C. sativus* chloroplast genome were 63.05 and 36.95%, respectively. The A-T contents of the IR region amounted to 57.21%, whereas the A-T contents in the LSC and SSC regions were 65.30 and 68.42%, respectively.

The sizes of exons and introns for each gene in the *C. sativus* chloroplast genome were summarized in Table 1. Eighteen genes contained one or two introns. Five of these introns, *ndhB*, *rpl2*, *rps12*, *trnA-UGC*, and *trnI-GAU*, were located within the IR regions. Only the *trnL-UAA* gene intron belonged to the self-splicing group I, while all others belonged to group II. Three genes, *clpP*, *rps12*, and *ycf3*, had two introns. The *rps12* gene was a trans-spliced gene with the 5' end exon located in the LSC region far away from its second and third exons.

The border between the two IR/LSC and the two IR/SSC usually differs among various species. Large expansions and contractions of IR regions often create large length variation in chloroplast genomes in different groups of plants (Palmer et al. 1988; Raubeson and Jansen 2005; Wakasugi et al. 2001). In Fig. 2, the detailed IR-SC border positions, with respect to adjacent genes, were compared among *Cucumis*, *Lotus*, *Arabidopsis*, *Nicotiana*, and *Panax* chloroplast genomes. In *Cucumis*, IR extended deep into the *ycf1* gene, and was inserted into the *ycf1* pseudogene with 1056 bp at the IRa/SSC border. The IRb/SSC borders in *Cucumis* were located within the coding regions of the *ycf1* and *ndhF* genes, respectively. Thus, a portion (6 bp) of

the 3' end of the *ndhF* gene overlapped with the internal portion of *ycf1*. In contrast to the IR/SSC borders, only minor shifts were observed in the IR/LSC borders. The IRa/LSC of *Cucumis* were located downstream of the non-coding region of *trnH-GUG*. Thus, no pseudogene was created at the border. However, the IRb/LSC borders extended into the 5' portion of the *rps19* gene in *Cucumis*, *Lotus*, *Arabidopsis*, and *Panax*. As a result, the various lengths of *rps19* pseudogenes were located on the IRa/LSC border in *Cucumis* (2 bp), *Lotus* (1 bp), *Panax* (51 bp), and *Arabidopsis* (113 bp). The expansions/contractions of IR, as observed in the IR/SSC borders, are probably mediated by intra-molecular recombination between two short direct repeat sequences that frequently occur within the genes located at the borders (Kim and Lee 2004). Taken together, the expansion/contraction manner of IR at LSC and SSC in *Cucumis* was similar to that of *Lotus*. On the other hand, *Arabidopsis* and *Nicotiana* had a similar expansion/contraction manner of IR at the IRb/LSC and IR/SSC borders, and at IRa/LSC in *Cucumis*, respectively (Fig. 2).

Fig. 2 Comparison of the border positions of SSC, LSC, and IR regions among five chloroplast genomes. Figures of *Nicotiana*, *Panax* and *Arabidopsis* are from Kim and Lee (2004)



During the preparation of this manuscript, a cucumber chloroplast genome sequence in the GenBank data base was deposited on June 10 (AJ970307, gi: 67511377). However, this genome sequence differed from ours in a number of ways. For example, the full length of the *C. sativus* chloroplast genome was 155,527 bp in this study, whereas 155,293 bp in AJ97037. A Blast 2 comparison between the two genome sequences showed 580 bp of gaps and identity of 99% (154,857/155,700 bp). Furthermore, in the AJ970307, the sequence of IRa did not accord with that of IRb.

Genes in the cucumber chloroplast genome

The total number of genes in a chloroplast genome varies by species, e.g., 113 genes in tobacco, 110 genes in rice, and 42 genes in *Epifagus virginiana*. It has been suggested that most of the size variation among land plant cpDNAs can be accounted for by changes in the length of the IR

Table 2 Distribution of simple sequence repeats (SSRs) loci from the *Cucumis sativus* chloroplast genome

Length	Number and position of SSR			
	As	Cs	Ts	Gs
10	8 (1,664, 4,268, 8,434, 28,431, 30,098, 84,194, 112,341, 118,964)	0	12 (10,868, 16,287, 27,141, 32,927, 45,517, 51,380, 57,229, 59,427, 69,136, 73,991, 84,354, 115,247)	0
11	6 (5,769, 30,018, 49,865, 59,502, 83,598, 129,137)	1 (51,996)	4 (29,394, 33,805, 54,224, 61,326)	0
12	4 (3,208, 44,660, 73,215, 115,335)	0	0	0
13	3 (7,880, 8,125, 141,653)	0	4 (19,393, 100,742, 116,174, 128,517)	0
14	3 (6,674, 83,625, 155,502)	0	4 (67,472, 82,576, 86,892, 125,481)	0
15	2 (73,118, 122,461)	0	1 (127,344)	0
16	1 (23,576)	0	2 (72,931, 85,009)	0
17	1 (9,240)	0	0	0
Total	28	1	27	0

The loci were indicated in parentheses as starting nucleotide number of each SSR

(Palmer et al. 1988; Raubeson and Jansen 2005; Wakasugi et al. 2001).

In the *C. sativus* chloroplast genome, 113 genes were identified, which consisted of 76 peptide-encoding genes, 30 tRNA genes, four rRNA genes, and three conserved open reading frames (*ycf*). Positions and sizes of protein- and RNA-coding genes in the *C. sativus* chloroplast genome were presented in GeneBank (DQ119058). In terms of the protein components of transcription and translation, 12 genes for small subunits of ribosomes, nine genes for large subunits of ribosomes, and four genes for DNA-dependent RNA polymerase were identified from the *C. sativus* chloroplast genome. The numbers and kinds of tRNA and rRNA genes from *C. sativus* were identical to well characterized vascular plant chloroplasts (Kim and Lee 2004; Raubeson and Jansen 2005; Wakasugi et al. 2001). The information associated with the positions of tRNA genes and the sizes of spacers is important in constructing plastid transformation vectors specialized to a crop. Since recombination is targeted by homologous sequences, site-directed gene incorporation can be possible without any negative traits (Maliga 2002).

It has been known that land plant chloroplasts contain six genes for photosystem I (PS I) components, 15 genes for photosystem II (PS II) components, six genes for cytochrome b/f complex components, and six genes for ATP synthase subunits. In addition, they contain two genes contributing to assembly and/or stability of PS I complexes. Among them, *psaM* is absent in angiosperms (Wakasugi et al. 2001). *Ycf3* and *ycf4* products have been found to be involved in the assembly/stability of PS I in *Chlamydomonas* (Boudreau et al. 1997), as has *ycf3* in tobacco (Ruf et al. 1997). *Ycf3* and *ycf4* sequences were also identified in *Cucumis* and their positions were indicated in Fig. 1. The gene lengths and the sequence divergence of the *ycf1* and *ycf2* genes vary among plant species. The length variation of *ycf1* is largely due to the large indels in the middle of the gene and to the extensions/contractions of the IR into the SSC regions. The length variation of *ycf2* is largely responsible for the internal indel mutations associated with direct repeat sequences (Kim and Lee 2004). In the *Cu-*

cumis chloroplast genome, the length of *ycf1* and *ycf2* ($\times 2$) was 5640 bp and 5967 ($\times 2$) bp. The *infA* gene was absent in the *Arabidopsis*, *Lotus*, and *Medicago* chloroplast genomes but was present as a truncated pseudogene (239 bp) in the *Cucumis* chloroplast genome. Recently, 10 additional *ycfs* (*ycf66* and *ycf68-76*) have been proposed in land plant cpDNAs (Stoebe et al. 1998). Identification of these genes was not attempted in this study. Eleven *ndh* genes known to be absent in most algal chloroplasts (Stoebe et al. 1998) were identified from the *Cucumis* chloroplast genome, as in other vascular plants except black pine.

Simple sequence repeats

Chloroplast simple sequence repeats (SSR), which repeat the single nucleotide bases A or T more than 10 times, were reported in *Pinus radiata* (Cato and Richardson 1996; Powell et al. 1995) and *Oryza sativa* (Shimada and Sugiura 1989). Some of the *Panax* SSR loci could have systematic utilities in identifying species or cultivars (Kim and Lee 2004). We therefore investigated the status of the SSR in the *C. sativa* chloroplast genome and identified a total of 56 SSR loci (Table 2). There were 28 multiple As (10–17 bases), 27 multiple Ts (10–17 bases) and one multiple C (11 bases), but no multiple Cs or Gs in the *C. sativa* chloroplast genome. Kim and Lee (2004) reported that there were 12 multiple As (10–14 bases) and 27 multiple Ts (10–17 bases), but no multiple Cs or Gs in the *Nicotiana* chloroplast genome. However, in the *Panax* chloroplast genome, there were four multiple As (10–14 bases), 10 multiple Ts (10–11 bases), three multiple Cs (10–11 bases), and one multiple G (10 bases). These results show that SSRs vary significantly among plant species (Cato and Richardson 1996; Powell et al. 1995). In addition, Kim and Lee (2004) have proposed that the *16S*, *23S*, *ndhB*, *psbA*, *psbD*, *psaB*, *psaA*, *psbC*, *psbB*, and *rbcl* genes are probably good candidate genes for the phylogenetic study of higher plants. In contrast, *ycf1*, *ycf2*, *accD*, *matK*, *rpoC2*, and *ndhF* genes (if present) are good candidate genes for phylogenetic study among closely related families or infra-family levels

of vascular plants. Therefore, with respect to phylogenetic study, sequence divergences of these genes in the Cucurbitaceae, which contains white gourd, oriental melon, watermelon, muskmelon, pumpkin, squash, and bottle gourd, should be further analyzed based on the information obtained on the cucumber chloroplast genome. This will provide numerous insights into the evolution of these genomes and much new data for assessing relationships at deep nodes in plants and other photosynthetic organisms.

References

- Adachi J, Waddell PJ, Martin W, Hasegawa M (2000) Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. *J Mol Evol* 50:348–358
- Boudreau E, Takahashi Y, Lemieux C, Turmel M, Rochaix JD (1997) The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the Photosystem I complex. *EMBO J* 16:6095–6104
- Bock R, Khan MS (2004) Taming plastids for a green future. *Trend Biotechnol* 22(6):311–318
- Cato SA, Richardson TE (1996) Inter- and intraspecific polymorphism at chloroplast SSR loci and the inheritance of plastids in *Pinus radiata* D. Don. *Theor Appl Genet* 93:587–592
- Chu KH, Qi J, Yu ZG, Anh V (2004) Origin and phylogeny of chloroplasts revealed by a simple correlation analysis of complete genomes. *Mol Biol Evol* 21:200–206
- Cosner ME, Raubeson LA, Jansen RK (2004) Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evol Biol* 4:27
- Daniell H, Ruiz ON, Dhingra A (2004) Chloroplast genetic engineering to improve agronomic traits. *Methods Mol Biol* 286:111–138
- Goremykin V, Hirsch-Ernst KI, Wölfl S, Hellwig FH (2003) The chloroplast genome of the “basal” angiosperm *Calycanthus fertilis*—structural and phylogenetic analyses. *Plant Syst Evol* 242:119–135
- Goremykin VV, Hirsch-Ernst KI, Wölfl S, Hellwig FH (2004) The chloroplast genome of *Nymphaea alba*: Whole-genome analyses and the problem of identifying the most basal angiosperm. *Mol Biol Evol* 21:1445–1454
- Havey MJ, Lilly JW, Bohanec B, Bartoszewski G, Malepszy S (2002) Cucumber: a model angiosperm for mitochondrial transformation? *J Appl Genet* 43(1):1–17
- Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC, Wyman SK, Alverson AJ, Peery R, Herman SJ, Fourcade HM, Kuehl JV, McNeal JR, Leenens-Mack J, Cui L (2005) Methods for obtaining and analyzing whole chloroplast genome sequences. *Method Enzymol* 395:348–384
- Kim KJ, Lee HL (2004) Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res* 11:247–261
- Maliga P (2002) Engineering the plastid genome of higher plants. *Curr Opin Plant Biol* 5:164–172
- Martin W, Rujan T, Richly E., Hansen A, Cornelsen S., Lins T, Leister D, Stobe B, Hasegawa M, Penny D (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci USA* 99:12246–12251
- Matsuoka Y, Yamazaki Y, Ogihara Y, Tsunewaki K (2002) Whole chloroplast genome comparison of rice, maize, and wheat: Implications for chloroplast gene diversification and phylogeny of cereals. *Mol Biol Evol* 19:2084–2091
- Palmer JD (1982) Physical and gene mapping of chloroplast DNA from *Atriplex triangularis* and *Cucumis sativus*. *Nucl Acids Res* 10:1593–1605
- Palmer J, Jansen R, Michaels H, Chase M, Manhart J (1988) Chloroplast DNA variation and plant phylogeny. *Ann Missouri Bot Gard* 75:1180–1206
- Palmer JD (1991) Plastid chromosome: structure and evolution. In: Hermann RG (ed) *The molecular biology of plastids. Cell culture and somatic cell genetics of plants*, Springer-Verlag, Vienna, pp 5–53
- Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proc Natl Acad Sci USA* 92:7759–7763
- Raubeson LA, Jansen RK (2005) Chloroplast genomes of plants. In: Henry RJ (ed) *Plant diversity and evolution: Genotypic and phenotypic variation in higher plants*, CAB International, Wallingford, UK, pp 45–68
- Ruf S, Kössel H, Bock R (1997) Targeted inactivation of a tobacco intron-containing open reading frame reveals a novel chloroplast-encoded Photosystem I-related gene. *J Cell Biol* 139:95–102
- Schulze J, Balko C, Zellner B, Koprek T, Hänsch R, Nerlich A, Mendel RR (1995) Biolistic transformation of cucumber using embryogenic suspension cultures: long-term expression of reporter genes. *Plant Sci* 112:197–206
- Shimada H, Sugiura M (1989) Pseudogenes and short repeated sequences in the rice chloroplast genome. *Curr Genet* 16:293–301
- Stoebe B, Martin W, Kowallik KV (1998) Distribution and nomenclature of protein-coding genes in 12 sequenced chloroplast genomes. *Plant Mol Biol Rep* 16:243–255
- Surpin M, Larkin RM, Chory J (2002) Signal transduction between the chloroplast and the nucleus. *Plant Cell* S327–S338
- Szwacka M, Morawski M, Bueza W (1996) *Agrobacterium tumefaciens*-mediated cucumber transformation with thaumatin II cDNA. *J Appl Genet* 37A:126–129
- Triboush SO, Danilenko NG, Davydenko OG (1998) A method for isolation of chloroplast DNA and mitochondrial DNA from sunflower. *Plant Mol Biol Rep* 16:183–189
- Wakasugi T, Sugita M, Tsudzuki T, Sugiura M (1998) Updated gene map of tobacco chloroplast DNA. *Plant Mol Biol Rep* 16:231–241
- Wakasugi T, Sugita M, Tsudzuki T, Sugiura M (2001) The genomics of land plant chloroplasts: gene content and alteration of genomic information by RNA editing. *Photosyn Res* 70:107–118
- Yin Z, Hennig J, Szwacka M, Malepszy S (2004) Tobacco PR-2d promoter is induced in transgenic cucumber in response to biotic and abiotic stimuli. *J Plant Physiol* 161:621–629