VERSITA

CELLULAR & MOLECULAR BIOLOGY LETTERS
Volume 12 (2007) pp 584 - 594
http://www.cmbl.org.pl

DOI: 10.2478/s11658-007-0029-7

Received: 12 February 2007
Revised form accepted: 30 April 2007
Published online: 03 July 2007

Research article

THE COMPLETE STRUCTURE OF THE CUCUMBER (*Cucumis sativus* L.) CHLOROPLAST GENOME: ITS COMPOSITION AND COMPARATIVE ANALYSIS

WOJCIECH PLĄDER[1]*, YASUSHI YUKAWA[2], MASAHIRO SUGIURA[2] and STEFAN MALEPSZY[1]

[1]Warsaw Agricultural University, Faculty of Horticulture and Landscape Architecture, Department of Plant Genetics, Breeding and Biotechnology, Nowoursynowska 159, 02-776 Warsaw, Poland, [2]Graduate School of Natural Sciences, Nagoya City University, Mizuho, Nagoya 467-8501, Japan

Abstract: The complete nucleotide sequence of the cucumber (*C. sativus* L. var. Borszczagowski) chloroplast genome has been determined. The genome is composed of 155,293 bp containing a pair of inverted repeats of 25,191 bp, which are separated by two single-copy regions, a small 18,222-bp one and a large 86,688-bp one. The chloroplast genome of cucumber contains 130 known genes, including 89 protein-coding genes, 8 ribosomal RNA genes (4 rRNA species), and 37 tRNA genes (30 tRNA species), with 18 of them located in the inverted repeat region. Of these genes, 16 contain one intron, and two genes and one *ycf* contain 2 introns. Twenty-one small inversions that form stem-loop structures, ranging from 18 to 49 bp, have been identified. Eight of them show similarity to those of other species, while eight seem to be cucumber specific. Detailed comparisons of *ycf2* and *ycf15*, and the overall structure to other chloroplast genomes were performed.

Key words: Organelle, Gene order

*Author for correspondence; e-mail: wojciech_plader@sggw.pl, tel/fax: +48-22-59-321-52

Abbreviations used: cpDNA – chloroplast DNA; IR – inverted repeat; JLA – junction IRA/LSC; JLB – junction IRB/LSC; JSA – junction IRA/SSC; JSB – junction IRB/SSC; LSC – large single copy; PCR – polymerase chain reaction; rRNA – ribosomal RNA; SSC – small single copy; tRNA – transport RNA

INTRODUCTION

Plastids are organelles present in almost every plant cell, and depending on the species, they are inherited strictly maternally, exclusively paternally, or bi-parentally [1]. They are highly specialised and contain an entire complex of proteins necessary for photosynthesis. They have their own genome DNA (cpDNA), present in many copies, each 110-200 kbp. cpDNAs are believed to be among the best-known parts of the plant genome. For over 30 years, it has been believed that the basic form of cpDNA is a circular molecule [2, 3], and while this historical form is still valid as a genetic map, it most likely does not correctly represent the much more variable *in vivo* structure that was recently described [4, 5].

The *Cucurbitaceae* family is an important group of cultivated species, mostly known as popular vegetables. They are also of significant interest due to their secondary metabolites, mostly glycoalkaloids, which are valuable for the pharmaceutical industry. The following species have the greatest economic importance: cucumber, pumpkin, summer squash (zucchini), melon and watermelon. In Poland, cucumber is the fourth most popular vegetable, and it is cultivated in both open field agriculture and under cover.

Cucumber is a well-described species in terms of its cytology [6], biochemistry [7] and physiology [8], but not at the genetics and molecular biology levels. There are good regeneration systems for its *in vitro* direct regeneration [9], somatic embryogenesis [10], regeneration from protoplasts [11], and the conditions for nuclear transformation using *Agrobacterium tumefaciens* have been determined [12, 13]. The mitochondrial genome of cucumber has unique properties; like melon, it has one of the largest mitochondrial genomes, it exhibits paternal inheritance, and it contains large non-coding regions [14].

Studies of the chloroplast genome of cucumber started over 20 years ago. At that time, a physical map was described based on restriction enzyme cleavages, and the molecule size was determined as 155 kbp, containing a large inverted repeat (IR) of 22-25 kbp [15]. While this paper was in preparation, another *C. sativus* var. Baekmibaekdadagi chloroplast genome sequence was published [16].

The aim of this paper is to publish the entire sequence of the cucumber chloroplast chromosome. Analyses of the genome and its comparison to other species are discussed.

MATERIALS AND METHODS

DNA isolation

The plant materials were cotyledons of cucumber (*C. sativus* L.), Borszczagowski line. The seeds were placed on wet absorbent paper in a dish, and covered by foil. The dish was kept in a phytotrone, at a light intensity of 52 ěmol·m^{-2}·s^{-1} with a photoperiod of 16 hours/day. After one week of growth, the cotyledons were taken for DNA isolation.

DNA was isolated using the method described previously for *S. oleracea* [17], with minor modifications. The cotyledons collected were ground in liquid nitrogen, then mixed with extraction buffer containing 50 mM HEPES (pH 8.0), 1 mM $MgCl_2$, 1 mM $MnCl_2$, 2 mM EDTA, 330 mM sorbitol, and 5 mM sodium ascorbate. After filtration through 4 layers of gauze, the filtrate was centrifuged for 5 minutes at 4500x g. The pellet was suspended in a small volume of extraction buffer with DNaseI (80 U/ml) and 10 mM $MgCl_2$ (final concentration), and incubated for 15 min on ice. The reaction was stopped by adding EDTA to a final concentration of 30 mM. The pre-treated sample was placed on Percoll gradients (40%/85%) and centrifuged for 15 min at 4000x rpm. The green strip at the phase border (unbroken chloroplasts) was retrieved, and DNA was isolated using a Gibco DNazol kit according to the manufacturer's protocol. The resulting DNA from the enriched chloroplast fraction was tested by restriction enzyme cleavage and analysis on agarose gel.

Amplification of long DNA fragments and sequencing

Based on the known sequence of the tobacco chloroplast genome [18], 18 primer pairs were designed and used for PCR amplification using a Long PCR kit (TAKARA Ltd.) The bands obtained were sent for shotgun-sequencing to Shimadzu Biotech Corporation (Japan). The received sequences were evaluated and assembled using the Phred/Prap/Consed program (www.phrap.org), then analysed using the BioEdit program (North Carolina State University, http://www.mbio.ncsu.edu/BioEdit/page2.html). To combine the contigs obtained, primers were successively designed for the missing fragments, and sequencing was performed using the method of chromosome walking. Ninety-seven primers were designed, including those for amplification directly from the clones from the library, for amplification of longer fragments joining the contigs, and for chromosome walking. The analysis of the whole chloroplast genome was performed using the BLAST algorithm (National Centre for Biotechnology Information – NCBI, USA, http://www.ncbi.nih.gov/) and the ClustalW program [19]. A detailed description of the genes in terms of the nucleotides and amino acids, along with the entire sequence written in the EMBL form, was performed using the Artemis program (www.sanger.ac.uk). Similarity analyses were performed using the PipMaker program [20].

RESULTS AND DISCUSSION

Amplification and shot gun sequencing

The isolation of the enriched fraction of cpDNA using a modified version of the procedure published by Cheng *et al.* [13] allowed a good-quality template for the PCR reaction to be obtained. The quality of cpDNA obtained was tested by Eco RI restriction cleavage resulting in clear bands, unlike the cleavage of total genomic DNA (not shown). The long fragment amplification reaction mostly allowed products of the expected length to be obtained, from about 7 kbp to 25 kbp (not shown). The product length calculations were based on tobacco

cp DNA, so the amplification products from cucumber slightly differed from the expected values.

The resulting nine DNA fragments were sequenced using the "shotgun" method, which involves mechanical division of each long fragment into shorter segments of about 2-2.5 kbp, subcloning to a vector, then sequencing from both ends. In this way, 1002 clones (2004 sequences) were obtained with a total length of 1,377,752 bp. The first stage was an analysis of the resulting sequences to identify contamination with vector terminal sequences. The sequence was combined using the Phred/Prap/Consed program. Twenty-five contigs were obtained with a mean length of 2-3 kbp, excluding four long ones: 27.7 kbp, 15.1 kbp, 14.5 kbp and 10.4 kbp. In total, the contigs contained 111.6 kbp, accounting for 71.9% of the total genome length. The remaining part of the genome was sequenced using the technique of chromosome walking. To this end, primers specific for the ends of a known sequence were designed, fragments from 0.6 kbp to 6 kbp were amplified, and sequencing and further alignment was performed. As a consequence, 9.25-fold coverage of the cucumber chloroplast genome was achieved.

Architecture of the cucumber chloroplast genome

The chloroplast genome of *C. sativus* is double-stranded DNA containing 155,293 bp (GenBank accession no. AJ970307). Its structure is consistent with most of the described chloroplast genomes from the higher plants, and is composed of two inverted repeated regions (IRA and IRB), each 25,191 bp long, separated by a small single-copy (SSC) region and a large single-copy (LSC) region, respectively 18,222 bp and 86,688 bp. The CG content is 37.08%; this value is similar to those reported for most of the chloroplast genomes of other vascular plant species (e.g. 37.8% in *N. tabacum*). The chloroplast genome of *C. sativus* contains 130 genes, of which 18 are in the IR (Tab. 1).

Most genes occur in a single copy, while all the rRNA genes and some of the tRNA and protein-coding genes in the IR occur as double copies. Of the genes identified, 19 contained one or two introns. *C. sativus*, similarly to *Nicotiana* or S*pinacia*, contains 30 different tRNA genes, and seven are present in the IR, giving a total of 37 tRNA genes per genome. Both the number and the tRNA types are identical to those presented for other species of vascular plants [e.g. 18, 21, 22]. Six of them, *trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA* and *trnV-UAC*, contain introns of 549 bp to 2497 bp in length.

A comparison between our sequence and that presented by Kim *et al.* [16] showed only minor differences. In our results, there is a deletion of 40 bp between *ycf5* and *ndhD* within the intergenic region, a deletion of 265 bp within the second intron of *clpP* and an insertion of 61 bp within the intron of *rps16*. The reason for this phenomenon could be polymorphism due to the use of a different variety for the chloroplast genome sequencing.

Tab. 1. The gene groups present in the chloroplast genome of *C. sativus*.

tRNA genes *A-UGC*[#*], *C-GCA*, *D-GUC*, *E-UUU*, *F-GAA*, *G-UCC*[*], *G-GCC*, *H-GUG*, *I-GAU*[#*], *I-CAU*[#], *K-UUU*[*], *L-UAG*, *L-CAA*[#], *L-UAA*[*], *M-CAU*, *fM-CAU*, *N-GUU*[#], *P-UGG*, *Q-UUG*, *R-UCU*, *R-ACG*[#], *S-GGA*, *S-GCU*, *S-UGA*, *T-UGU*, *T-GGU*, *V-UAC*[*], *V-GAC*[#], *W-CCA*, *Y-GUA*
Ribosomal RNA genes *rrn16*[#], *rrn23*[#], *rrn4.5*[#], *rrn5*[#]
Ribosomal protein genes *rps2*, *rps3*, *rps4*, *rps7*[#], *rps8*, *rps11*, *rps12*[#**], *rps14*, *rps15*, *rps16*[*], *rps 18*, *rps19* *rpl2*[#*], *rpl14*, *rpl16*[*], *rpl20*, *rpl22*, *rpl23*[#], *rpl32*, *rpl33*[*], *rpl36*
Photosynthetic apparatus genes *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *atpA*, *atpB*, *atpE*, *atpF*[*], *atpH*, *atpI*, *psaA*, *psaB*, *psaC*, *petA*, *petB*[*], *petD*[*], *petG*, *petL*, *rbcL*
Transcription/translation apparatus genes *rpoA*, *rpoB*, *rpoC1*[*], *rpoC2*, *infA*
NADH dehydrogenase genes *ndhA*[*], *ndhB*[#*], *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
Other protein genes *accD*, *clpP*[**], *matK*
Conserved open reading frames *ycf1*, *ycf2*[#], *ycf3*[**], *ycf4*, *ycf5*, *ycf6*, *ycf9*, *ycf10*, *ycf15*[#]

– genes present in two copies due to their location in the IR, * – genes containing one or two (**) introns

A comparison of the chloroplast genome sequence of *C. sativus* to *Nicotiana*, the first sequenced plant chloroplast genome, which serves as the reference genome for most chloroplast DNA work, and to *Panax*, which shows a high similarity to *C. sativus* in its base composition, indicated a close relationship between them. Only the IR, in which re-arrangement may be found, is an exception. The *ycf2* gene contains four main deletions. The largest, a 723-bp fragment (interrupted by 5 bp), is missing compared to *Nicotiana*; this phenomenon is similar to a deletion of 596 bp which occurs in the same region in *Panax*. The second 158-bp deletion (interrupted by 7 bp) is near the first. Deletions are often observed in *ycf2*, which encodes chloroplast proteins of unknown function (DUF825) for which no conserved domains have been found so far. It could therefore be assumed that the lack of the above sequences will not affect the functioning of the resulting protein, because similar deletions exist in other described genomes (*Spinacia*, *Dianthus*, *Pelargonium*, *Panax*).

The second gene with modified structure is *ycf15*. Compared to the other genomes, it contains a 284-bp insertion at position +105, relative to the start codon, and in frame TGA, a stop codon occurs, unlike in other species, where a GGA codon is present .
The sequencing and analysis of the *ycf15* DNA and cDNA revealed that there is no difference between the templates, which indicates that, as expected for spermatophytes, there is no U-to-C editing in *C. sativus* (not shown).
The borders between IR and the single-copy regions in *Cucumis* are not exceptional, and differ slightly from those described for other species (Fig. 1).

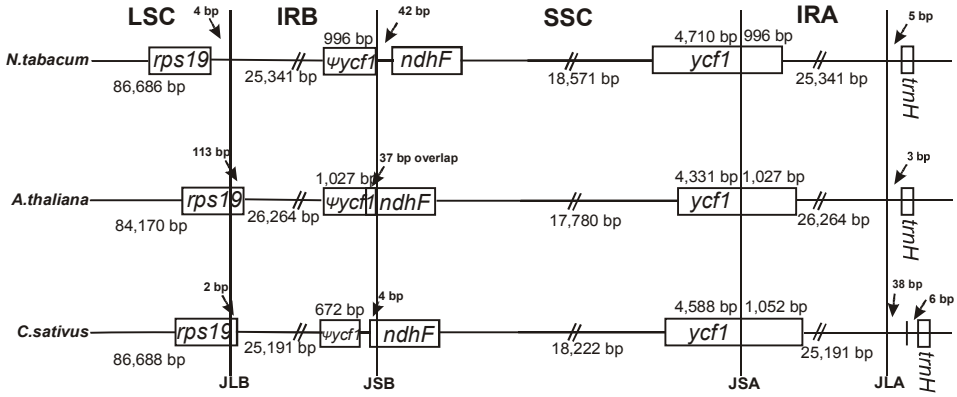

Fig. 1. A comparison of the border positions of the SSC, LSC, and IR regions between the *Nicotiana*, *Arabidopsis* and *C. sativus* chloroplast genomes. The various lengths of *ycf1* pseudogenes (5' portions) are created at the JSB. A portion (37 bp) of the 3' end of the *ndhF* gene overlaps the internal position of the *ycf1* pseudogene in *Arabidopsis*, while 4 bp of *ndhF* extends into the IRB in *C. sativus*. The JLB, JSB, JSA and JLA states for the junctions between the single-copy gene regions and the inverted repeat regions.

The *rps 19* gene on JLB expands, as in *Arabidopsis* (113 bp), 2 bp into the IRB, while in *Nicotiana*, there is 4 bp gap between the gene and the JLB. On the JSB, the *ycf1* pseudogene (with a length of 672 bp) is present in *Cucumis*, but it is smaller than in other species (e.g. 996 bp in *Nicotiana* and 1027 bp in *Arabidopsis*). A small portion (4 bp) of the 3' end of the *ndhF* gene extends to the IRB, which resembles the case of the 37-bp extension in *Arabidopsis*, but this does not overlap with the truncated *ycf1*. The JSA is located within the *ycf1* gene with only minor shifts when compared to other species. The JLA is located downstream of the *trnH* gene, and there is no pseudogene of *rps19* as in *Arabidopsis* or *Panax*.

Comparative analyses of the chloroplast genome of *Cucumber*

Comparative analyses of the substitutions, deletions and insertions within the coding and non-coding regions between *C. sativus* and *Nicotiana* gave clear

evidence not only about the conservative nature of the coding regions, but also about the differences due to the location of the analyzed sequence within a particular part of the cpDNA (Tab. 2).

Tab. 2. Substitution, insertion and deletion within the coding and non-coding regions of *C. sativus* in comparison to *Nicotiana*.

Region	Substitutions		Insertions		Deletions	
	Coding	Non coding	Coding	Non coding	Coding	Non coding
LSC	9.6%	27.4%	2.6%	6.3%	1.6%	3.4%
IR	1.2%	8.0%	0.03%	1.4%	0.03%	3.3%
SSC	12.1%	32.4%	2.5%	10.4%	2.4%	1.0%

The lowest frequency not only for substitutions but also for deletions and insertions is for the sequences located in the IR, except *ycf's*. This can be explained by the fact that seven tRNA and four rRNA genes, the most conservative genes, are located in this region. Of them, seven do not have any changes. The exceptions are: *trnV* (2 substitutions; 2.7%), *rrn 4.5* (2 substitutions; 1.9%), *rrn23* (46 substitutions; 1.6%) and *rrn16* (8 substitutions; 0.5%). The genes present in the SC regions have much higher substitution rates, ranging from 9.6% for LSC to 12.1% for SSC. The existence of substitutions for non-coding regions is 6.6 times higher for IR, 2.8 times higher for LSC, and 2.7 times higher for SSC.

The results of this analysis were confirmed by comparative analysis of the entire chloroplast genome using the PipMaker program [20]. The analysis showed that along the entire length of the coding regions for *C. sativus* and three other species: *Nicotiana*, *Panax* and *Arabidopsis*, there is high similarity between the analysed species. An analysis suggests that *rps16* and *rpl16* genes exhibit the lowest inter-species similarity. However, a more detailed analysis shows that this is not true because the two genes contain introns respectively of 913 bp and 1129 bp, which represents over 78% of the entire sequence for *rps16* (an intron and two exons), and over 73% for *rpl16*.

The analysis of *ycf1* shows that it has a rather low level of similarity to those from other species. The highest similarity occurs in the 5' region, which may be due to the initial 224 codons in *C. sativus* (350 in *Nicotiana*) being located within the IR. On the other hand, a comparison of this conservative frame to that from *Chlamydomonas reinhardtii*, in relation to both nucleotide and amino acid compositions, did not show significant similarities.

An additional element presented in the chloroplast genome of *C. sativus* is the occurrence of short inverted repeats, similarly to the other described genomes. Using the REPuter program (http://bibiserv.techfak.uni-bielefeld.de/reputer/), 55 repeats of 18 to 44 bp, without medium (from about 200 to 1000 bp) or long (inverted repeat area) fragments, were identified. Of them, the potential transcripts of twenty-one can form a hairpin structure. A similar number

(twenty) was described for *Panax* [22, 23]. The comparison of *C. sativus* and *Panax* showed that only two of them are identical: *rrn*4.5-*rrn*5 and *trnL-ndhB*. This is probably related to the location of these two repeats in the highly conserved IR. The lack of inversion located between *ycf15* and *trnL* in *Panax*, also located in IR, can be explained by a small deletion of 62 bp in *C. sativus* similar to the 82 bp deletion in *Arabidopsis*, which includes this inversion (not shown). The inversions presented in the *C. sativus* chloroplast genome (Tab. 3) are similar to those in other chloroplast genomes (*Nicotiana*, *Panax*, *Oenothera*) or have no similarity to other organisms (numbers 3, 7-9, 11, 12, 14, 15 in Tab. 3). These observations indicate that some short inversions are conservative between species, especially those located in the IRs, while others seems to be species specific.

Tab. 3. The location and nucleotide composition of 16 small inversions in *C. sativus*. The predicted loop sequence is underlined; cs, nt, pg and oe stand for *C. sativus*, *Nicotiana*, *Panax* and *Oenothera*, respectively.

Gene location	Taxa	Position in genome	Sequence alignment
1. *psbT* - *psbN*(c)	cs	76501-76545	ATTGAAGTAATGAGCCTCACAATATTGTGAGGCTCATTACTTCAA
	nt		GTTGAAGTACTGAGCCTCCCGATACCGGGAGGCTCAGTACTTCAA
	pg		ATTGAAGTAATGAGCCCCCCAATATTGGGGGGCTCATTACTTCAA
2. *trnK* - *rps16*(c)	cs	4483-4525	AATAACTATATTTATTTTATATATATAAAATAAATATAGTTAT
	nt		AAAGACTATTTTTATTTTATTCCTCCGAATAGAACATGGCCAT
3. *ndhG(c)* - *ndhI*(c)	cs	120533-120580	GTAGAGTAGATTGAAACTATATATTTCGAAATATATAGTTTCAATCTA
4. *psaI* - *ycf4*	cs	62267-62306	TTTTAGATCTAAGTAATTCATCGAATTACT--TAGATCTAAA
	nt		TTTTAGATTTAACCAATTTGATGAATTACTCCTAAAGGTTGA
	pg		TTTTAAATCTAACCAATTTGATGAATTACTCTTAAAGGTTCAC
5. *petA* - *psbJ*(c)	cs	65540-65570	GTCTATTTTTATTCTAATAGAATAAAAATAG
	oe		-TCTATTTTTATTCTAATTGAATAAAAA
6. *trnE*(c) - *trnT*	cs	32731-32755	CATCATACTATGATCATAGTATGAT
	nt		CATCATACTATGATCATAGTATGAT
	pg		CATCATACTATGTTCATAGTATGAA
7. *clpP*(c) within intron	cs	72973-72998	TTTTTTTTTTTTCAAAAAAAAAAA
8. *trnG* - *trnR*	cs	10271-10298	ATTCTAAATATAAATTAATTAATTTATA
9. *trnG* - *trnfMet*	cs	38788-38817	ATAGTAATTAATTAATATAGTAATTAATTA
10. *ndhF*(c)	cs	113278-113298	CGCTGCTACCATGGTAGCAGC
	nt		CGCCGCTACCATAGTAGCAGC
	pg		CGCTGCTACCATAGTAGCAGC
11. *trnT*(c) - *trnL*	cs	49230-49253	TATTATATAATAATATTATATAAT
12. *rpoB*(c) - *trnC*	cs	28497-28521	CTTATATTATATATATATAAGATAA
13. *rpoC2*(c)	cs	18358-18376	TTCACAATAATTATTGTGA
	nt		TTCACAATAATTGTGATGG
14. *ycf3*(c) within intron	cs	45326-45344	TCTTATATCATGATATAAG
15. *ndhC*(c) - *trnV*(c)	cs	53731-53749	AATATGAAAATTTTCATAT
16. *petB* within intron	cs	77198-77216	GTCATCTTGATCAAGATGA
	nt		GTCATCTTGATAGAGATGG
	pg		GTCATCTTGATAGAGGTGG

The occurrence of short inverted repeats in chloroplast genomes is described as the main factor determining the likely inversion sites in the genome [24]. However, until today, there was only data on short repeats for *Panax*, and single cases of short repeats for other species had been described [25]. Due to the occurrence of a hairpin sequence in close proximity to the translation termination codons, it is believed that these hairpin structures can play a role in stabilising mRNA molecules [26].

A comparative analysis between complete *C. sativus* chloroplast DNA and other accessions presented for *C. sativus* in data bases at the nucleotide level showed only small differences, not bigger than 1% of substitutions (AY274260 – 0.5%; AF206755 – 0.6%; AF209572 – 0.6%; L21937 – 0.6%; AF441853 – 0.9%; L43907 – 0.4%; X75799 – 0.5%). This could be a reason for the existence of polymorphism between the genetic material used for analyses as well as the length of the sequence (fragments not longer than 1450 bp). The longest sequence (AY521596) of 16,653 bp showed only 0.08% of substitutions between the analysed material, even that located in the LSC. This can be explained by the probable close phylogenetic distance between the genotypes.

C. sativus is a very important crop plant in Poland, with a production of 278 thousand tons per annum (the third highest in Europe). The determination of the complete sequence of its chloroplast genome opens new opportunities for both basic and application studies in this species. It can allow for the improvement of existing varieties mainly by chloroplast transformation with regards to both biotic and abiotic stresses.

Acknowledgements. We would like to thank Maki Yukawa for helping to sequence the material preparation. This study was partially supported by the Polish Ministry of Education and Science, grant 2P06A 022 28.

REFERENCES

1. Havey, M.J., Lilly, J.W., Bohanec, B., Bartoszewski, G. and Malepszy, S. Cucumber: A model angiosperm for mitochondrial transformation? **J. Appl. Genet.** 43 (2002) 1-17.
2. Kolodner, R. and Tewari, K. Molecular size and conformation of chloroplast deoxyrybonucleic acid from pea leaves. **J. Biol. Chem.** 247 (1972) 6355-6364.
3. Kolodner, R. and Tewari, K. Inverted repeats in chloroplast DNA from higher plants. **Proc. Natl. Acad. Sci. USA** 76 (1979) 41-45.
4. Deng, X.W., Wing, R.A. and Gruissem, A. The chloroplast genome exists in multimeric forms. **Proc. Natl. Acad. Sci. USA** 86 (1989) 4156-4160.
5. Lilly, J.W., Havey, M.J., Jackson, S.A. and Jiang, J. Cytogenomic analyses reveal the structural plasticity of the chloroplast genome in higher plants. **Plant Cell** 13 (2001) 245-254.

6. Hoshi, Y., Plader, W. and Malepszy, S. New C-banding pattern for chromosome identification in cucumber (*Cucumis sativus* L.). **Plant Breed.** 117 (1998) 77-82.
7. De Nisi, P. and Zocchi, G. Phosphoenolpyruvate carboxylase in cucumber (*Cucumis sativus* L.) roots under iron deficiency: activity and kinetic characterization. **J. Exp. Biol.** 51 (2000) 1903-1909.
8. Hirano, T., Kiyota, M. and Aiga I. Physical effects of dust on leaf physiology of cucumber and kidney bean plants. **Environ. Pollut.** 89 (1995) 255-261.
9. Burza, W. and Malepszy, S. Direct plant regeneration from leaf explants in cucumber (*C. sativus sativus* L.) is free of stable genetic variation. **Plant Breed.** 114 (1995a) 341-345.
10. Wróblewski, T., Filipecki, M.K. and Malepszy, S. Factors influencing cucumber (*C. sativus sativus* L.) somatic embryogenesis. I. The crucial role of pH and nitrogen in suspension culture. **Acta Soc. Bot. Pol.** 64 (1995) 223-231.
11. Burza, W. and Malepszy, S. *In vitro* culture of *C.sativus sativus* L. XVIII. Plants from protoplasts through direct somatic embryogenesis. **Plant Cell Tissue Organ Cult.** 41 (1995b) 259-266.
12. Yin, Z. and Malepszy, S. The transgenes are expressed with different level in plants. **Biotechnologia** 2 (2003) 236-260.
13. Yin, Z., Pląder, W. and Malepszy, S. Transgene inheritance in plants. **J. Appl. Genet.** 45 (2004) 127-144.
14. Havey, M.J., Lilly, J.W., Bohanec, B., Bartoszewski, G. and Malepszy, S. Cucumber: a model angiosperm for mitochondrial transformation? **J. Appl. Genet.** 43 (2002) 1-17.
15. Palmer, J.D. Physical and gene mapping of chloroplast DNA from Atriplex triangularis and *C. sativus sativa*. **Nucleic Acid Res.** 10 (1982) 1593-1605.
16. Kim, J.S., Jung, J.D., Lee, J.A., Park, H.W., Oh, K.H., Jeong, W.J., Choi, D.W., Liu, J.R. and Cho, K.Y. Complete sequence and organization of the cucumber (*C. sativus* L. cv. Baekmibaekdadagi) chloroplast genome. **Plant Cell. Rep.** 25 (2006) 334-340.
17. Cheng, M.C., Wu, S.P., Chen, L.F. and Chen, S.C. Identification and purification of a spinach chloroplast DNA-binding protein that interacts specifically with the plastid psaA-psaB-rps14 promoter region. **Planta** 203 (1997) 373-380.
18. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chungwonse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H and Sugiura, M. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. **EMBO J.** 5 (1986) 2043-2049.

19. Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G. and Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Res.** 22 (1994) 4673-4680.
20. Schwartz, S., Zhang, Z., Frazer, K.A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R. and Miller, W. PipMaker – a web server for aligning two genomic DNA sequences. **Genome Res.** 10 (2000) 577-586.
21. Maier, R.M., Neckermann, K., Igloi, G.L. and Kossel, H. Complete sequence of the maize chloroplast genome: gene content. hotspots of divergence and fine tuning of genetic information by transcript editing. **J. Mol. Biol.** 251 (1995) 614-628.
22. Kim, K.J. and Lee, H.L. Complete chloroplast genome sequence from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. **DNA Res.** 11 (2004) 247-261.
23. Kim, K.J. and Lee, H.L. Widespread occurance of small inversions in the chloroplast genomes of land plants. **Mol. Cells** 19 (2005) 104-113.
24. Palmer, J.D. Plastid chromosomes: structure and evolution In: **Cell Culture and Somatic Cell Genetics in Plants**, Vol. 7A, The Molecular Biology of Plastids (Vasil, I.K. and Bogorad, L. Eds.), Academic Press, San Diego, 1991, 5-53.
25. Kelchner, S.A. and Wende, J.F. Hairpins create minute inversions in non-coding regions of chloroplast DNA. **Curr. Genet.** 30 (1996) 259-262.
26. Shinozaki, K., Hayashida, N. and Sugiura, M. *Nicotiana* chloroplast genes for components of the photosynthetic apparatus. **Photosynthesis Res.** 18 (1988) 7-31.