



OPEN

Inheritance of chloroplast and mitochondrial genomes in cucumber revealed by four reciprocal F₁ hybrid combinations

Hyun-Seung Park^{1,5}, Won Kyung Lee^{1,5}, Sang-Choon Lee², Hyun Oh Lee², Ho Jun Joh¹, Jee Young Park¹, Sunggil Kim³, Kihwan Song⁴✉ & Tae-Jin Yang¹✉

Both genomes in chloroplasts and mitochondria of plant cell are usually inherited from maternal parent, with rare exceptions. To characterize the inheritance patterns of the organelle genomes in cucumber (*Cucumis sativus* var. *sativus*), two inbred lines and their reciprocal F₁ hybrids were analyzed using a next generation whole genome sequencing data. Their complete chloroplast genome sequences were de novo assembled, and a single SNP was identified between the parental lines. Two reciprocal F₁ hybrids have the same chloroplast genomes with their maternal parents. Meanwhile, 292 polymorphic sites were identified between mitochondrial genomes of the two parental lines, which showed the same genotypes with their paternal parents in the two reciprocal F₁ hybrids, without any recombination. The inheritance patterns of the chloroplast and mitochondria genomes were also confirmed in four additional cucumber accessions and their six reciprocal F₁ hybrids using molecular markers derived from the identified polymorphic sites. Taken together, our results indicate that the cucumber chloroplast genome is maternally inherited, as is typically observed in other plant species, whereas the large cucumber mitochondrial genome is paternally inherited. The combination of DNA markers derived from the chloroplast and mitochondrial genomes will provide a convenient system for purity test of F₁ hybrid seeds in cucumber breeding.

Chloroplasts and mitochondria are essential plant organelles that perform photosynthesis and ATP generation, respectively, in addition to their involvement in the biosynthesis of compounds such as starch, oils, and amino acids¹. While the nuclear genes show biparental Mendelian inheritance patterns, the chloroplast and mitochondrial genomes show non-Mendelian inheritance patterns, predominantly maternal inheritance². This maternal inheritance pattern is probably a result of the high mutational load of the male gametes². Alternative chloroplast and mitochondria inheritance patterns have been reported in some plant species however, making it difficult to understand the inheritance patterns of their organelles².

Both the chloroplasts and the mitochondria in plants are believed to have evolved from ancient endosymbionts, and as such contain their own genomic material. While the chloroplast genome structure and size are maintained around 120–160 kb^{3–13}, plant mitochondrial genome structure and size are highly dynamic, including several forms and sizes in the range of 220–3000 kb^{14–16}. This variation contrasts with the animal mitochondrial genome, which is a highly compact, simple, and intact structure, comprising approximately 16 kb of circular DNA¹⁷.

Cucumber (*Cucumis sativus* var. *sativus*) is one of the most widely cultivated and economically important fruit crops in the world^{18,19}. In addition, cucumber is a model plant for the study of sex determination and vascular biology^{20,21}; therefore, many cucumber cultivars and wild cucumber species have been subjected to analyses of their nuclear and organellar genomes^{22–30}. These studies have shown that cucumber has a narrow genetic diversity, which makes the breeding of this species more difficult¹⁹. To date, eight cucumber chloroplast genome sequences have been reported³¹. By contrast, the complete mitochondrial genome sequence has only been reported for the

¹Department of Agriculture, Forestry and Bioresources, Plant Genomics and Breeding Institute, College of Agriculture and Life Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea. ²Phyzen Genomics Institute, Seongnam, Gyeonggi-do 13558, Republic of Korea. ³Department of Horticulture, Chonnam National University, Gwangju 61186, Republic of Korea. ⁴Department of Bioresources Engineering, College of Life Sciences, Sejong University, Seoul 05006, Republic of Korea. ⁵These authors contributed equally: Hyun-Seung Park and Won Kyung Lee. ✉email: khsong@sejong.ac.kr; tjyang@snu.ac.kr

| Cucumber samples | Generation | Raw data | | Trimmed data | | NABIC accession nos |
|--------------------------|-----------------------|-----------|---------------|-------------------|---------------------|---------------------|
| | | Reads | Bases | Reads | Bases | |
| MGL ^a | Parent | 2,861,968 | 856,903,647 | 2,404,700 (84.0%) | 568,007,755 (66.3%) | NN-4028 |
| CFL ^b | Parent | 3,110,572 | 931,526,838 | 2,655,587 (85.4%) | 628,169,399 (67.4%) | NN-4031 |
| F ₁ (MGL×CFL) | F ₁ hybrid | 3,388,454 | 1,014,964,305 | 2,873,832 (84.8%) | 679,680,316 (67.0%) | NN-4030 |
| F ₁ (CFL×MGL) | F ₁ hybrid | 3,191,966 | 955,434,149 | 2,758,084 (86.4%) | 654,654,520 (68.5%) | NN-4032 |

Table 1. Cucumber samples and NGS data used in this study. ^aKorean solid green type inbred line. ^bChinese long green type inbred line.

cucumber cultivar Calypso³² and was found to comprise three chromosomes, including a major larger chromosome (~1.6 Mb) and two smaller chromosomes (84 kb and 45 kb each). The cucumber mitochondrial genome is almost 10 times longer than that of its chloroplast genome (~160 kb).

Analyses of the polymorphism patterns present in small regions of the cucumber organelle genomes have revealed that the chloroplast and mitochondrial genomes of this species are inherited maternally and paternally, respectively^{33–36}. Understanding these inheritance patterns is particularly important because some key cucumber traits are closely associated with the organelles and their inheritance; for example, the cucumber chilling temperature response was reported to be maternally or paternally inherited, depending on the genetic background, implying that it could be under the control of a factor encoded by an organelle genome²³. In addition, cytoplasmic male sterility and variegated phenotypes have been linked to the mitochondrial genome in plant species such as radish (*Raphanus sativus*), pepper (*Capsicum annuum*), *Arabidopsis thaliana*, and tobacco (*Nicotiana tabacum*)^{16,37–39}.

Next-generation sequencing (NGS) techniques enable the simultaneous assembly of the chloroplast and mitochondrial genome sequences at a low cost, which has facilitated the sequencing of a large number of plant organelle genomes, enabling their in-depth study^{3,4,8}. In this study, we tried to clarify the inheritance pattern of chloroplast and mitochondrial genomes based on NGS whole-genome sequencing data using several cucumber parental lines and their reciprocal F₁ hybrids.

Results

Complete chloroplast genomes of two parental inbred lines and their F₁ hybrids. A total of 3.8 Gb PE reads (about 0.9–1.0 Gb for each of the four cucumber samples) were obtained and quality-trimmed (Table 1). The resulting high-quality PE reads (670–800 Mb) from each sample were independently de novo assembled to generate the complete chloroplast genome sequences of two cucumber parental lines, MGL and CFL, and their two reciprocal F₁ hybrids, MGL×CFL and CFL×MGL. All the chloroplast genomes had the same size, 155,525 bp, and possessed a typical quadripartite structure, consisting of a large single-copy (LSC) region of 86,877 bp, a small single-copy (SSC) region of 18,274 bp, and a pair of inverted repeats (IRa and IRb) comprising 25,187 bp (Fig. 1, Supplementary Figs. S1 and S2, and Table S1). All of the chloroplast genomes contained a total of 120 genes, including 79 protein-coding genes, 37 tRNA genes, and four rRNA genes. The average depths of the trimmed NGS data mapped to the complete chloroplast genome sequences ranged from 540× to 690×. The complete chloroplast genome sequences of the two parental inbred lines and their two reciprocal F₁ hybrids were deposited in GenBank under the accession numbers KX231327, KX231328, KX231329, and KX231330, respectively.

The only one SNP was identified between the chloroplast genomes of the two parental inbred lines MGL and CFL, which was located at position 53,610 bp in the intergenic region between *ndhC* and *trnV-UAC* (Fig. 1a). A comparison of our chloroplast genome sequences with those reported for other cucumber cultivars [GY14 (DQ865975), CHIPPER (DQ865976), Baekmibaekdadagi (DQ119058), Borszczagowski (AJ970307), and hardwickii isolate PI183967 (MF536709)] (Supplementary Table S2) revealed very low-level diversity (1 to 39 SNPs and 3 InDels) among Korean breeding lines and a few dozen SNPs and InDels (50 to 288 SNPs and 19 to 150 InDels) between the wild collections or the North-European cucumber (Borszczagowski, AJ970307).

Polymorphisms in the mitochondrial sequences of the two parental inbred lines. Since the cucumber mitochondrial genome is huge (~1.6 Mb in NC_016005), the complete mitochondrial genome could not be assembled using the short NGS reads generated in this study. Instead, the polymorphic sites, SNPs, and InDels were investigated using NGS read mapping. A total of 292 polymorphic sites, including 246 SNPs and 46 InDels, were identified between the mitochondrial sequences of the two parental inbred lines (Fig. 1b, Supplementary Table S3). The NGS mitochondrial reads mapped onto the polymorphic sites had an average coverage of 22.7×. Among the 272 polymorphic sites, 240, 46, and six were identified in intergenic regions, introns, and exons, respectively. The six exonic SNPs were identified in three genes; *rpsI*, encoding a ribosomal protein; *ccmB*, encoding an ABC transporter subunit; and *nad7*, encoding a NADH dehydrogenase subunit. All six exonic SNPs were non-synonymous substitutions that resulted in amino acid changes.

Validation of chloroplast and mitochondrial genome inheritance using DNA markers. Molecular markers were developed based on the polymorphic sites identified in the chloroplast and mitochondrial sequences. For the chloroplast sequence, a pair of dCAPS markers were designed based on the single SNP identi-

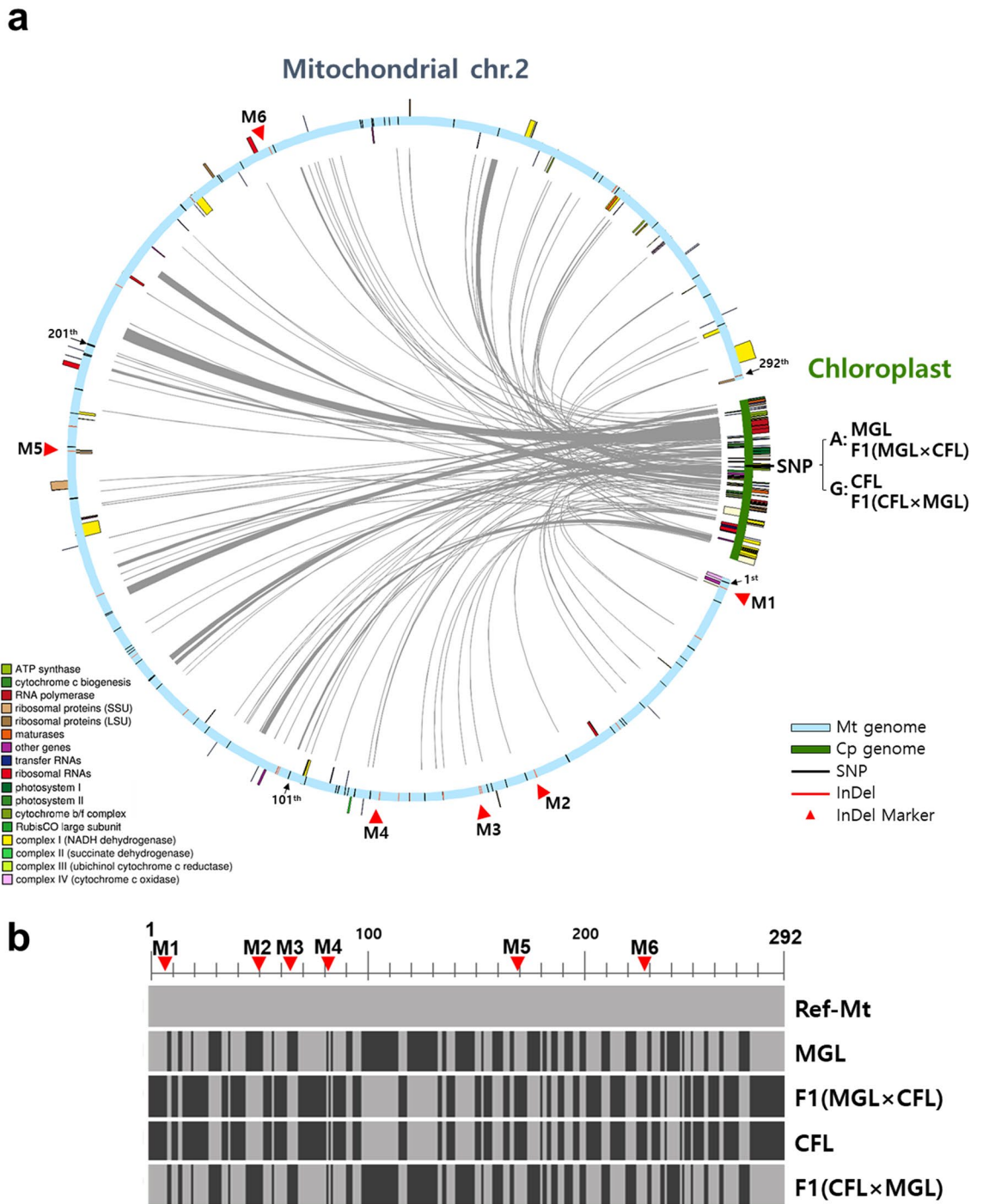


Figure 1. Schematic representation of mitochondrial and chloroplast genome of *Cucumis sativus*. (a) Each colored bar indicates the mitochondrial chromosome 2 retrieved from NCBI (NC_016005.1) and chloroplast genome of the inbred line MGL in this study. One of the pair of the inverted repeat regions in chloroplast genome was removed. Genes transcribed clockwise and counterclockwise are located on the outside and inside of the bar, respectively. The 292 mitochondrial and 1 chloroplast polymorphisms between two parental lines, MGL and CFL, are represented as black lines for SNP and red lines for InDels. The mitochondrial plastid DNAs and their plastid origins are linked with gray line. The positions of six InDel markers (M1 to M6) are labeled with red triangles. (b) Genotype comparison of the 292 polymorphic sites in mitochondrial genome of parental lines and F₁ hybrids of reciprocal cross. The same color indicates the same genotype for polymorphic sites among parental lines and F₁ hybrids. Gray-mitochondrial indicates the genotype of the reference mitochondrial genome (NC_016005.1).

| Position ^a | Intergenic region location ^b | Consensus | Reference ^c | MGL | CFL | F1 (MGL × CFL) | F1 (CFL × MGL) | Developed marker ID |
|-----------------------|---|----------------------|------------------------|-----|----------------|----------------|----------------|---------------------|
| 7,613 | <i>ccmFc-nad5</i> | GTCCTACTTCCATATGGATA | – | – | InDel1 (20 bp) | InDel1 (20 bp) | – | Mt-InDel-01 |
| 226,850 | <i>rrmS-rps7</i> | CTACGGATAAATCCTCCGGA | TR20 (20 bp) | – | TR20 (20 bp) | TR20 (20 bp) | – | Mt-InDel-02 |
| 274,021 | <i>rps7-trnH-GUG</i> | GTAGGACTATCCATTAGAA | InDel3 (20 bp) | – | InDel3 (20 bp) | InDel3 (20 bp) | – | Mt-InDel-03 |
| 355,356 | <i>rps7-trnH-GUG</i> | CTACTTCTAATGGATAGTC | InDel4 (19 bp) | – | InDel4 (19 bp) | InDel4 (19 bp) | – | Mt-InDel-04 |
| 766,753 | <i>rps4-nad6</i> | TTATCCGTAGTCCGGAGGAT | – | – | InDel5 (21 bp) | InDel5 (21 bp) | – | Mt-InDel-05 |
| 1,074,192 | <i>rrnL-trnP</i> | CTACGGATAAATCCTCCGGA | – | – | TR20 (20 bp) | TR20 (20 bp) | – | Mt-InDel-06 |

Table 2. InDel sites and sequences with length difference of more than 20 bp in mt sequences between two parental lines. ^{a,b,c}Based on mt genome sequence (NC_016005.1, 1,555,935 bp) previously reported in cucumber cultivar Calypso³². “–”, is representing non-detected indels or tandem repeats.

fied between the chloroplast genomes of the two parental inbred lines (Fig. 1a). For the mitochondrial sequence, six targets with more than a 20-bp InDel between the two parental lines were selected among the 292 polymorphic sites (Fig. 1b, Table 2). Overall, seven polymorphic sites, one from the chloroplast genome and six from the mitochondrial genome, were used to design molecular markers for the investigation of the inheritance patterns of the chloroplasts and mitochondria (Table 3).

The molecular markers were used for genotyping of the two parental inbred lines and their two reciprocal F₁ hybrids. The chloroplast-derived markers revealed that the F₁ hybrids shared an identical genotype with their maternal parent in each cross (Fig. 2). By contrast, all six of the mitochondria-derived markers were used to reveal identical genotypes between the F₁ hybrids and their paternal parent (Fig. 2).

The inheritance patterns of the chloroplast and mitochondrial markers were also validated in other cucumber parental inbred lines (BP15, YHB, HHG, and KWS) and their F₁ hybrid plants (Fig. 2, Supplementary Fig. S3). The maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome were confirmed from all six reciprocal crosses of the four parental lines. The heterozygous genotypes of the nuclear genomes in the F₁ hybrids were confirmed by genotyping the two nuclear InDel markers, Nu-InDel-01 and Nu-InDel-02 (Table 3, Fig. 3, Supplementary Fig. S3), which were designed based on the InDel polymorphisms identified in the intron of the cucumber gene *Csa1G042170* on chromosome 1 and the 3' UTR of the cucumber gene *Csa3G127780* on chromosome 3.

Discussion

Reconfiguration of the organellar genome inheritance pattern in cucumber using NGS sequencing. Here, we generated approximately 1 Gb of NGS data, which is about three-fold the haploid genome coverage of *Cucumis sativus*²⁴. These data were used to assess the genetic diversity and inheritance of the organellar genomes. This quantity of data was sufficient to facilitate the assembly of the chloroplast genomes of the four assessed cucumber lines, which confirmed their maternal inheritance based on the inheritance of one SNP identified between the parental lines. We also identified 246 SNPs and 46 InDels across the mitochondrial genome by mapping NGS reads directly to the reference cucumber mitochondrial genome.

Here, we used complete genome sequences to validate the previously reported inheritance patterns of the organellar genomes, which were identified using only a few Restriction Fragment Length Polymorphism (RFLP) markers^{33,34}. This inheritance pattern was further inspected in the present study using another four cucumber lines and their reciprocal F₁ progenies. Our method is based on advanced NGS technologies and could be widely used not only for the study of the inheritance patterns of plant organellar genomes, but also for the study of mitochondrial genomes in general, which are usually more difficult to fully assemble than the chloroplast genomes.

Chloroplast genome diversity is low in cucumber, while its mitochondrial genomes are more diverse. In this study, the chloroplast genomes of the four cucumber lines (MGL, CFL, their reciprocal F₁ hybrids) could be fully assembled; however, the mitochondria genomes could not. This is not only because the chloroplast copy number is generally much higher than that of the mitochondria, but also because the cucumber mitochondrial genome is unusually large and complex³².

The mitochondrial gene sequences are more conserved than the chloroplast genes^{40,41}; however, here, we identified 52 polymorphic sites in the mitochondrial gene regions between two cucumber lines but none in the chloroplast gene regions. The six exonic polymorphic sites in the mitochondrial genome were non-synonymous mutations, which caused amino acid changes in the corresponding protein sequence. Genes containing the non-synonymous mutations included *rps1*, encoding a ribosomal protein; *ccmB*, encoding an ABC transporter subunit; and *nad7*, encoding a NADH dehydrogenase subunit. In Arabidopsis, the chloroplast *rps1* gene is involved in heat stress tolerance⁴², and may help to optimize chloroplast integrity under heat stress. It is therefore possible that the cucumber *rps1* gene in the mitochondrial genome might also be involved in the heat stress response. The mitochondrial *ccmB* is involved in cytochrome c and c1 biogenesis in wheat⁴³. *Nad7* encodes a

| Marker ID | Marker type | Direction | Primer sequences (5'-3') | Estimated amplicon sizes (bp) | | | | Genotyping ^c |
|-------------|------------------------|-----------|--|-------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| | | | | MGL | CFL | F1(MGLx CFL) | F1(CFLx MGL) | |
| Cp-SNP-01 | dCAPS (<i>EcoRV</i>) | F | GGTGGA GTTTT AACCGG TTTGAT ^a | 269 (245 ^b) | 269 (269 ^b) | 269 (245 ^b) | 269 (269 ^b) | Cp |
| | | R | TTCGAG TCCGTA TAGCCCTAT ATTA | | | | | |
| Cp-SNP-02 | dCAPS (<i>RsaI</i>) | F | TGAAGA AGGAAA GAGAAG ACGACTA | 256 (256 ^d) | 256 (231 ^d) | 256 (256 ^d) | 256 (231 ^d) | Cp |
| | | R | GTAATTTT GGGAGT CAAAATCA GTA ^c | | | | | |
| Mt-InDel-01 | InDel | F | GCGGAC TGCTTC GGTTAAT | 258 | 265 | 265 | 258 | Mt |
| | | R | ACCTGTCCT ACGTGC CAAAG | | | | | |
| Mt-InDel-02 | InDel | F | AGAAAG GTCAA AGCCCTTC | 329 | 354 | 354 | 329 | Mt |
| | | R | GCAGCG AGAAAG TCAGAGGA | | | | | |
| Mt-InDel-03 | InDel | F | TTCTCT CCGTAC CGCGTAGT | 347 | 371 | 371 | 347 | Mt |
| | | R | TCACCT GGACCT TTCTTTGG | | | | | |
| Mt-InDel-04 | InDel | F | ACCTTA GGCCGG ACCTCTTA | 365 | 389 | 389 | 365 | Mt |
| | | R | CATCCATAT GGGCCA AAGAA | | | | | |
| Mt-InDel-05 | InDel | F | TTAGAA GCAGTC CGCAGGAT | 303 | 328 | 328 | 303 | Mt |
| | | R | ACCGTC CGTAGG AGTCCTTT | | | | | |
| Mt-InDel-06 | InDel | F | CCGGGA AATAGCAAT GAAAG | 555 | 580 | 580 | 555 | Mt |
| | | R | ACCTTC CTGCTT CAGGGAAT | | | | | |
| Nu-InDel-01 | InDel | F | GCTTGA ATTTGG GCAACA GGT | 332 | 373 | 332/373 | 332/373 | Nucleus |
| | | R | TCTTGA AGTGGG GGTGTCCG | | | | | |
| Nu-InDel-02 | InDel | F | CCACTG ATGACA ATGCCTGT | 309 | 264 | 309/264 | 309/264 | Nucleus |
| | | R | ACGAAA TTTTCTGTA ACTCAT GTAA | | | | | |

Table 3. Molecular markers designed in this study. ^{a,c}Bold letters indicate modified bases for cleavage of PCR amplicon by restriction enzymes (*EcoRV* and *RsaI*). ^{b,d}Fragment size after cleavage of PCR amplicon by restriction enzymes for each dCAPS marker. ^eCp indicating chloroplast and Mt indicating mitochondria.

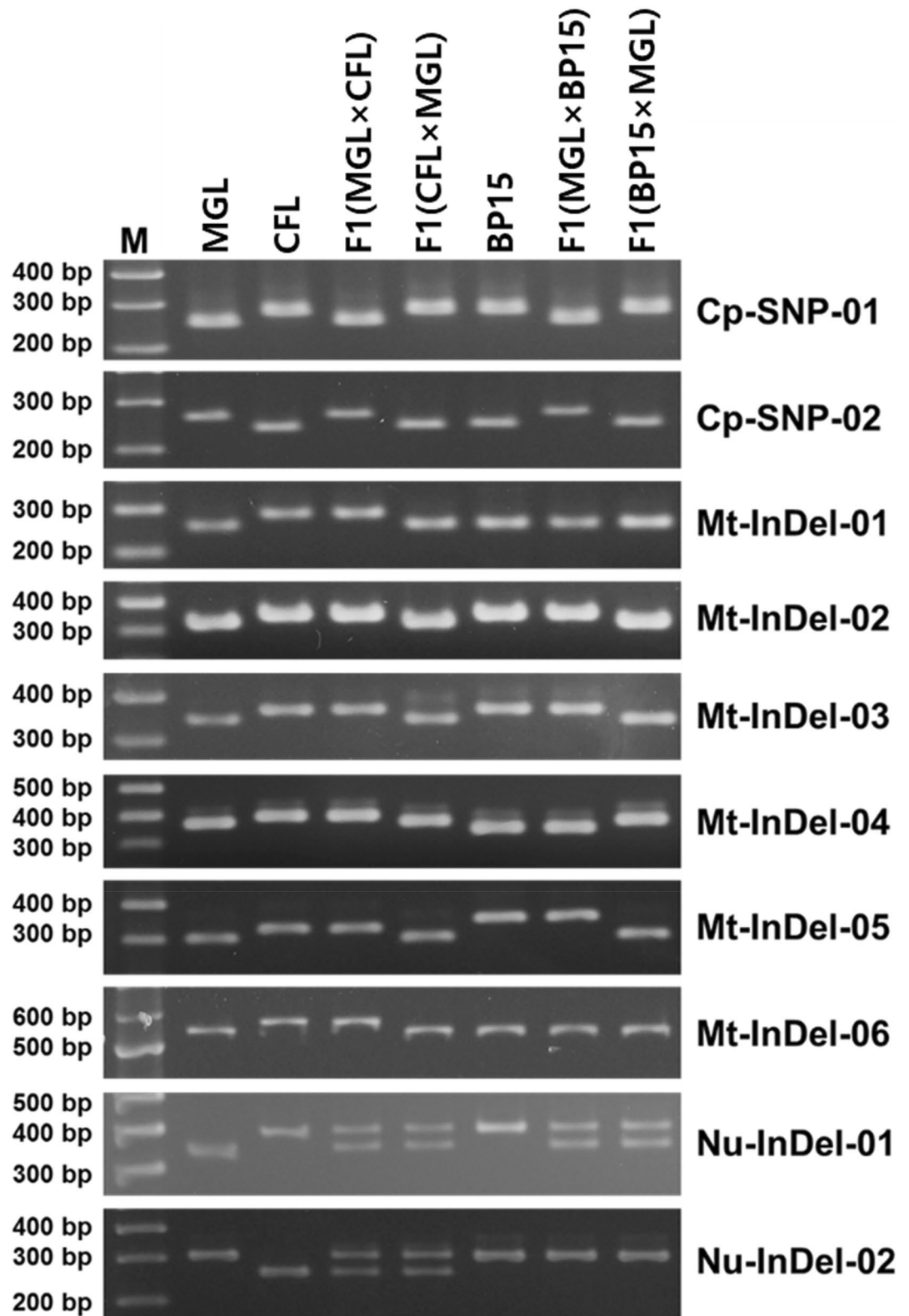


Figure 2. Validation of molecular markers to confirm inheritance pattern of organelles in cucumber. Eight molecular markers (Cp-SNP-01 to Mt-InDel-06) were designed based on chloroplast and mitochondrial sequence polymorphisms and validated using genomic DNA PCR analyses with seven cucumber samples. Two nuclear InDel markers, Nu-InDel-01 and Nu-InDel-02, were used to confirm heterozygous genotypes of F_1 hybrid nuclear genomes. M, 100-bp size marker; MGL, a Korean solid green-type inbred line; CFL, a Chinese long green-type inbred line; F_1 (MGL \times CFL), an F_1 hybrid between MGL (maternal) and CFL (paternal); F_1 (CFL \times MGL), an F_1 hybrid between CFL (maternal) and MGL (paternal); BP15, a Beith alpha-type inbred line; F_1 (MGL \times BP15), an F_1 hybrid between MGL (maternal) and BP15 (paternal); F_1 (BP15 \times MGL), an F_1 hybrid between BP15 (maternal) and MGL (paternal).

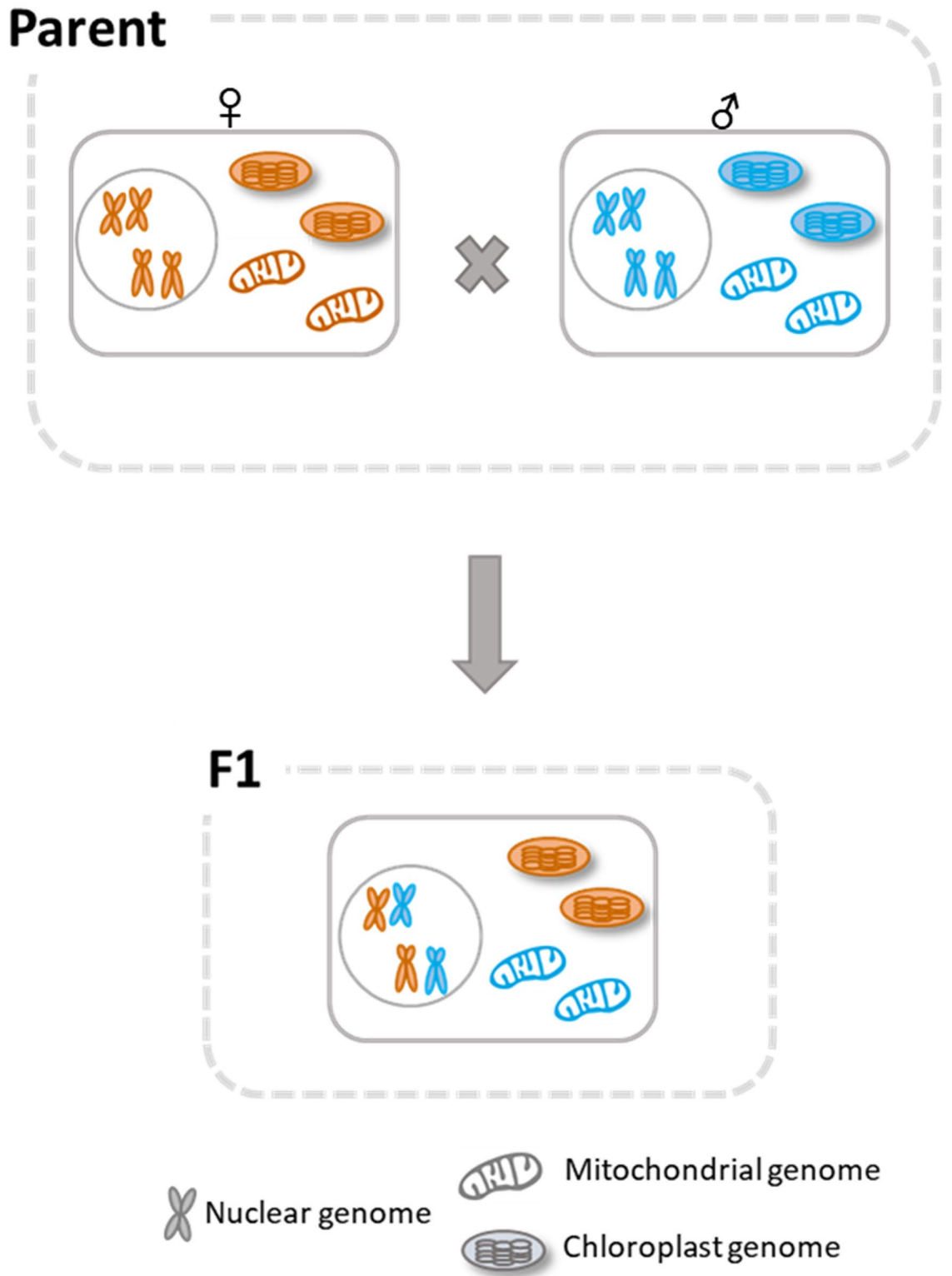


Figure 3. Summarized inheritance patterns of chloroplast and mitochondria in cucumber. Organelle inheritance from parental inbred lines to their two reciprocal F₁ progeny was elucidated through whole-organellar genome- and polymorphic marker-based genotyping, confirming that cucumber chloroplasts were inherited maternally while mitochondria were inherited paternally. Colors indicate genotypes.

NADH dehydrogenase, which is involved in the essential respiratory chain. The inhibition of the respiration is a main cause of the production of reactive oxygen species, which can damage cells and tissues⁴⁴. The important roles these genes play in plants suggests that future studies should explore how their identified mutations in cucumber affect the phenotypes of these plants.

The rich mitochondrial genome diversity may be caused by its paternal inheritance pattern in cucumber. Many plastid-derived DNA fragments have been identified in plant mitochondrial genomes^{45,46}. The total length of this mitochondrial plastid DNA (MTPT) in cucumber is 69 kb, which is the highest amount of MTPT among the mitochondrial genome sequences reported to date⁴⁶ (Supplementary Fig. 3). At 1.6 Mbp, the cucumber mitochondrial genome is one of the biggest in plants and is almost six times larger than the smallest plant mitochondrial genome, which is around 220 kbp in *Brassica* species¹⁴. As discussed above, the cucumber mitochondrial genes are particularly diverse relative to the diversity observed in other plants⁴¹, and MTPT is also abundant in the cucumber mitochondrial genome. We can therefore assume that the abundant gene diversity in the cucumber mitochondrial genome is related to its paternal inheritance trait and its high content of MTPT fragments. Similar cases were reported in the organelle genomes of gymnosperms including the Pinaceae and Taxaceae, which are often paternally inherited⁴⁷. The synonymous substitution rates of both chloroplast and mitochondrial genomes were reported to be higher in species displaying paternal inheritance rather than maternal inheritance⁴⁸.

In animals, the organellar genomes are rarely inherited paternally or biparentally⁴⁹; however, plants with unique organelle inheritance patterns are not as rare as their animal counterparts. Most conifers display paternal chloroplast inheritance patterns^{50,51}, while the chloroplast genomes of alfalfa (*Medicago sativa*) and *Oenothera* spp. are biparentally inherited². Like cucumber, wild bananas (*Musa acuminata*) display a maternal chloroplast inheritance and a paternal mitochondrial inheritance⁵².

DNA markers derived from the chloroplast and mitochondria genomes can be used to assess the genotypes of F₁ hybrid seeds. Ten markers designed to target polymorphisms in the mitochondrial, chloroplast, and nuclear genomes were successfully applied to validate the genomic inheritance patterns in the F₁ progenies derived from crosses between several cucumber lines.

Plant cytoplasmic organelles are not only involved in photosynthesis and respiration; recent studies have revealed their diverse roles, including in agriculturally important traits such as male sterility^{16,53,54}. The cucumber mitochondrial genes may also affect male sterility properties, making the paternal inheritance of the cucumber mitochondrial genome especially important for agriculture. The independent inheritance of the cucumber chloroplasts and mitochondria must be considered in breeding programs. The genome sequences and markers developed here are therefore expected to be of great value in the cucumber breeding industry. In addition, we expect that our results could be used as a foundation to further the research into plants with unusual organelle inheritance traits, particularly in cucumber itself. The molecular markers developed in this research could be practically applied to the genotyping of other cucumber inbred lines or samples; for example, combinations of the chloroplast and mitochondrial markers could be used to check the purity of F₁ hybrid cucumber seeds as they are easily detected high-copy targets derived from the maternal and paternal parents, respectively.

Materials and methods

Plant materials. Two cucumber parental inbred lines, MGL bred from Korean solid green-type cucumber and CFL bred from Chinese long green-type cucumber, and their two reciprocal F₁ hybrids, MGL × CFL and CFL × MGL, were subjected to whole-genome sequencing using NGS technology. The additional inbred lines BP15, HHG, YHB, and KWS, as well as their F₁ hybrids, were also used for the molecular marker tests. All the breeding lines and reciprocal F₁ hybrids were developed in this study.

Whole-genome sequencing of cucumber. Total genomic DNAs were extracted from fresh leaves using a modified cetyltrimethyl ammonium bromide (CTAB) method⁵⁵ and their quality was examined using agarose electrophoresis and a spectrometer. Paired-end (PE) libraries with a 300-bp insert size were constructed according to the standard Illumina PE protocol, and the pooled PE libraries were sequenced by LabGenomics (<http://www.labgenomics.co.kr>, Seongnam, Republic of Korea) using the HiSeq2000 platform (Illumina, USA). The sequencing data were deposited in the National Agricultural Biotechnology Information Center (NABIC, <http://nabic.rda.go.kr>)⁵⁶.

Assembly and comparison of the complete chloroplast genome sequences. The PE reads were quality-trimmed using the CLC quality trim tool with default parameters, which is included in the CLC ASSEMBLY CELL package (ver. 4.6 beta; CLC Inc., Denmark; <http://www.clcbio.com/products/clc-assembly-cell/>). Afterward, high-quality PE reads (Phred scores > 20) were de novo assembled using the CLC genome assembler included in the package, as described previously^{3,4}. The chloroplast genome contigs were extracted, ordered, and merged to generate a single draft sequence, based on the reported reference chloroplast genome sequence of *C. sativus* (DQ865975; Gy14 cultivar; 155,525 bp)²³. The draft chloroplast sequences were manually combined, corrected, and gap-filled using a series of PE read mapping. A single ambiguous sequence caused by low coverage of PE read mapping was found only in F₁ (MGL × CFL) and confirmed using genomic PCR amplification and nucleotide sequencing. The chloroplast genome was annotated using GESEQ (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>), BLAST searches, and a comparison with the reference cucumber chloroplast genomes. The chloroplast genome sequence of each sample was independently assembled.

Sequence polymorphisms were identified by a comparison of the chloroplast genomes in the two parental inbred lines and their F₁ hybrids using multiple sequence alignment tools such as ClustalW (<http://www.genome.jp/tools/clustalw/>), MAFFT ver. 7 (<http://mafft.cbrc.jp/alignment/server/index.html>), and a BLAST-based alignment (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Investigation of polymorphic sites in the mitochondrial genome sequences. High-quality PE reads of each of the four cucumber samples were mapped to the previously reported reference mitochondrial genome sequence (NC_016005; 1,555,935 bp) from the cucumber cultivar Calypso³² using BWA ver. 0.7.10 (<http://bio-bwa.sourceforge.net>) with default parameters. Positions of properly mapped reads were selected using SAMtools ver. 1.1 (<http://samtools.sourceforge.net/>) and in-house scripts. Polymorphic sites [SNPs and insertion and deletion (InDel) mutations] were identified using Picard ver. 1.112 (<http://broadinstitute.github.io/picard/>) and GATK ver. 3.1 (<https://www.broadinstitute.org/gatk/>) with default parameters. All procedures were performed by Phyzen (<http://www.phyzen.com/>; Seongnam, Republic of Korea).

Development and validation of molecular markers. Polymorphic sites identified in the chloroplast and mitochondrial sequences were used to design molecular markers for the analysis of the chloroplast and mitochondrial genotypes in other cucumber lines. PCR primers were designed to target the polymorphic regions using the NCBI Primer-BLAST tool (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) for InDel markers and the dCAPS Finder 2.0 (<http://helix.wustl.edu/dcaps/dcaps.html>) for derived cleaved amplified polymorphic sequence (dCAPS) markers.

A PCR amplification with 2–50 ng genomic DNA templates was used to validate the markers. The PCR consisted of 25–30 cycles of 95 °C for 1 min, 52–65 °C for 1 min, and 72 °C for 1 min. For the optimal amplification of each specific PCR product, the cycles and annealing temperatures of the PCR reactions were adjusted for each of the markers. A final concentration of 1.0–1.5 M betaine was also used to enhance the specificity of the amplification. For the dCAPS marker, the amplified PCR products were digested with the corresponding restriction enzymes for 24 h. The PCR products or cleaved PCR products were separated on a 1.5–3.0% agarose gel containing ethidium bromide and visualized using an UV illuminator. All of the uncropped gel image was included in Supplementary information (Supplementary Fig. S4 and S5).

In addition, InDel polymorphic sites in the nuclear genome sequences of the two parental lines were identified by mapping the NGS reads to the cucumber reference nuclear genome sequence (v2.0; <http://www.icugi.org/cgi-bin/ICuGl/index.cgi>)²⁴ using the same method described for the mitochondrial genome sequence. These nuclear InDel sites were then used to design markers to validate the heterozygous nuclear genotype of the F₁ hybrids.

Data availability

The sequencing data were deposited in the National Agricultural Biotechnology Information Center (NABIC, <http://nabic.rda.go.kr>) with accession numbers of NN-4028, NN-4031, NN-4030, NN-4032. The four of chloroplast genomes of *C. sativus* were deposited in NCBI Nucleotide Database with accession numbers of KX231327, KX231328, KX231329, and KX231330.

Received: 17 March 2020; Accepted: 13 January 2021

Published online: 28 January 2021

References

- Buchanan, B. B., Grissem, W. & Jones, R. L. *Biochemistry and Molecular Biology of Plants* (Wiley, New York, 2015).
- Greiner, S., Sobanski, J. & Bock, R. Why are most organelle genomes transmitted maternally?. *BioEssays* **37**, 80–94. <https://doi.org/10.1002/bies.201400110> (2015).
- Kim, K. *et al.* Comprehensive survey of genetic diversity in chloroplast genomes and 45S rDNAs within panax ginseng species. *PLoS ONE* **10**, e0117159. <https://doi.org/10.1371/journal.pone.0117159> (2015).
- Kim, K. *et al.* Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Sci. Rep.* **5**, 15655. <https://doi.org/10.1038/srep15655> (2015).
- Daniell, H., Lin, C.-S., Yu, M. & Chang, W.-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* **17**, 134 (2016).
- Jang, W. *et al.* The complete chloroplast genome sequence of *Cynanchum auriculatum* Royle ex Wight (Apocynaceae). *Mitochondrial DNA Part A* **27**, 4549–4550 (2016).
- Park, H.-S. *et al.* The complete chloroplast genome sequence of an important medicinal plant *Cynanchum wilfordii* (Maxim.) Hemsl. (Apocynaceae). *Mitochondrial DNA Part A* **27**, 3747–3748 (2016).
- Joh, H. J. *et al.* Authentication of golden-berry *P. ginseng* cultivar ‘Gumpoong’ from a landrace ‘Hwangsook’ based on pooling method using chloroplast-derived markers. *J. Plant Breed. Biotech.* **5**, 16–24 (2017).
- Kim, K. *et al.* Evolution of the Araliaceae family inferred from complete chloroplast genomes and 45S rDNAs of 10 *Panax*-related species. *Sci. Rep.* **7**, 4917 (2017).
- Nguyen, V. B. *et al.* Authentication markers for five major *Panax* species developed via comparative analysis of complete chloroplast genome sequences. *J. Agric. Food Chem.* **65**, 6298–6306 (2017).
- Kim, I. *et al.* Discrimination and authentication of *Eclipta prostrata* and *E. alba* based on the complete chloroplast genomes. *Plant Breed. Biotechnol.* **5**, 334–343 (2017).
- Nguyen, V. B. *et al.* Comprehensive comparative analysis of chloroplast genomes from seven *Panax* species and development of an authentication system based on species-unique SNP markers. *J. Ginseng Res.* **44**, 135–144 (2020).
- Lee, H. O. *et al.* Dynamic chloroplast genome rearrangement and DNA barcoding for three apiaceae species known as the medicinal herb “Bang-Poong”. *Int. J. Mol. Sci.* **20**, 2196 (2019).
- Chang, S. *et al.* Mitochondrial genome sequencing helps show the evolutionary mechanism of mitochondrial genome formation in *Brassica*. *BMC Genom.* **12**, 497 (2011).

15. Kim, S., Park, J. Y. & Yang, T. Comparative analysis of the complete chloroplast genome sequences of a normal male-fertile cytoplasm and two different cytoplasm conferring cytoplasmic male sterility in onion (*Allium cepa* L.). *J. Hortic. Sci. Biotechnol.* **90**, 459–468 (2015).
16. Park, J. Y. *et al.* Complete mitochondrial genome sequence and identification of a candidate gene responsible for cytoplasmic male sterility in radish (*Raphanus sativus* L.) containing DCGMS cytoplasm. *Theor. Appl. Genet.* **126**, 1763–1774. <https://doi.org/10.1007/s00122-013-2090-0> (2013).
17. Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res.* **27**, 1767–1780 (1999).
18. Pitrat, M., Chauvet, M. & Foury, C. in *I International Symposium on Cucurbits* 492. 21–28 (1997).
19. Acquah, G. *Principles of Plant Genetics and Breeding: Breeding Cucumber* 2nd edn, 676–681 (Wiley, New York, 2012).
20. Tanurdzic, M. & Banks, J. A. Sex-determining mechanisms in land plants. *Plant Cell* **16**(Suppl), S61–71. <https://doi.org/10.1105/tpc.016667> (2004).
21. Lough, T. J. & Lucas, W. J. Integrative plant biology: role of phloem long-distance macromolecular trafficking. *J. Annu. Rev. Plant Biol.* **57**, 203–232 (2006).
22. Kim, J. S. *et al.* Complete sequence and organization of the cucumber (*Cucumis sativus* L. cv. Baekmibaekdadagi) chloroplast genome. *Plant Cell Rep.* **25**, 334–340. <https://doi.org/10.1007/s00299-005-0097-y> (2006).
23. Chung, S.-M., Gordon, V. S. & Staub, J. E. J. G. Sequencing cucumber (*Cucumis sativus* L.) chloroplast genomes identifies differences between chilling-tolerant and-susceptible cucumber lines. *Genome* **50**, 215–225 (2007).
24. Huang, S. *et al.* The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* **41**, 1275–1281. <https://doi.org/10.1038/ng.475> (2009).
25. Cavagnaro, P. F. *et al.* Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genom.* **11**, 569. <https://doi.org/10.1186/1471-2164-11-569> (2010).
26. Woycicki, R. *et al.* The genome sequence of the North-European cucumber (*Cucumis sativus* L.) unravels evolutionary adaptation mechanisms in plants. *PLoS ONE* **6**, e22728. <https://doi.org/10.1371/journal.pone.0022728> (2011).
27. Liu, B., Zhang, D. & Gao, L. Z. The complete chloroplast genome sequence of *Cucumis sativus* var. *Hardwickii*, the wild progenitor of cultivated cucumber. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* **27**, 4627–4628. <https://doi.org/10.3109/19401736.2015.1101588> (2016).
28. Gao, J., Liu, B. & Jiang, H. The complete chloroplast genome sequence of wild cucumber (*Cucumis sativus* var. *Hardwickii*). *Mitochondrial DNA A DNA Mapp. Seq. Anal.* **27**, 4524–4525. <https://doi.org/10.3109/19401736.2015.1101545> (2016).
29. Wu, Z. *et al.* The complete chloroplast genome sequence of the wild cucumber *Cucumis hystrix* Chakr (*Cucumis*, *cucurbitaceae*). *Mitochondrial DNA A DNA Mapp. Seq. Anal.* **27**, 142–144. <https://doi.org/10.3109/19401736.2013.878915> (2016).
30. Płader, W., Yukawa, Y., Sugiura, M. & Maleszky, S. The complete structure of the cucumber (*Cucumis sativus* L.) chloroplast genome: its composition and comparative analysis. *J. Cell. Mol. Biol. Lett.* **12**, 584 (2007).
31. Lee, S.-C. *et al.* The complete chloroplast genome sequence with a novel 24-bp deletion of a Korean solid green-type cucumber variety (*Cucumis sativus* var. *sativus*). *Mitochondrial DNA Part B* **2**, 755–756. <https://doi.org/10.1080/23802359.2017.1398604> (2017).
32. Alverson, A. J., Rice, D. W., Dickinson, S., Barry, K. & Palmer, J. D. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell* **23**, 2499–2513. <https://doi.org/10.1105/tpc.111.087189> (2011).
33. Havey, M. Predominant paternal transmission of the mitochondrial genome in cucumber. *J. Heredity* **88**, 232–235 (1997).
34. Havey, M. J., McCreight, J. D., Rhodes, B. & Taurick, G. Differential transmission of the *Cucumis* organellar genomes. *Theor. Appl. Genet.* **97**, 122–128. <https://doi.org/10.1007/s001220050875> (1998).
35. Bartoszewski, G., Gawronski, P., Szklarczyk, M., Verbakel, H. & Havey, M. J. A one-megabase physical map provides insights on gene organization in the enormous mitochondrial genome of cucumber. *Genome* **52**, 299–307. <https://doi.org/10.1139/G09-006> (2009).
36. Shen, J., Kere, M. G. & Chen, J. F. Mitochondrial genome is paternally inherited in *Cucumis* allotetraploid (*C. x hystrix*) derived by interspecific hybridization. *Sci. Hortic-Amsterdam* **155**, 39–42. <https://doi.org/10.1016/j.scienta.2013.03.009> (2013).
37. Jo, Y. D., Choi, Y., Kim, D.-H., Kim, B.-D. & Kang, B.-C. Extensive structural variations between mitochondrial genomes of CMS and normal peppers (*Capsicum annuum* L.) revealed by complete nucleotide sequencing. *BMC Genom.* **15**, 561 (2014).
38. Bonnett, H., Djurberg, I., Fajardo, M. & Glimelius, K. A mutation causing variegation and abnormal development in tobacco is associated with an altered mitochondrial DNA. *Plant J.* **3**, 519–525 (1993).
39. Carol, P. *et al.* Mutations in the Arabidopsis gene IMMUTANS cause a variegated phenotype by inactivating a chloroplast terminal oxidase associated with phytoene desaturation. *Plant Cell* **11**, 57–68 (1999).
40. Drouin, G., Daoud, H. & Xia, J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* **49**, 827–831 (2008).
41. Palmer, J. D. & Herbon, L. A. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. *J. Mol. Evol.* **28**, 87–97 (1988).
42. Yu, H.-D. *et al.* Downregulation of chloroplast RPS1 negatively modulates nuclear heat-responsive expression of HsfA2 and its target genes in Arabidopsis. *PLoS Genet.* **8**, e1002669 (2012).
43. Fairve-Nitschke, S. E., Nazoa, P., Gualberto, J. M., Grienenberger, J. M. & Bonnard, G. Wheat mitochondria *ccmB* encodes the membrane domain of a putative ABC transporter involved in cytochrome *c* biogenesis. *Biochimica et Biophysica Acta (BBA) Gene Struct. Express.* **1519**, 199–208 (2001).
44. Larosa, V. & Remacle, C. Insights into the respiratory chain and oxidative stress. *Biosci. Rep.* **38**, BSR20171492 (2018).
45. Wang, X.-C., Chen, H., Yang, D. & Liu, C. Diversity of mitochondrial plastid DNAs (MTPTs) in seed plants. *Mitochondrial DNA Part A* **29**, 635–642 (2018).
46. Park, H.-S. *et al.* Mitochondrial plastid DNA can cause DNA barcoding paradox in plants. *Sci. Rep.* **10**, 1–12 (2020).
47. Mogensen, H. L. INVITED SPECIAL PAPER: The hows and whys of cytoplasmic inheritance in seed plants. *Am. J. Bot.* **83**, 383–404 (1996).
48. Whittle, C.-A. & Johnston, M. O. Male-driven evolution of mitochondrial and chloroplast DNA sequences in plants. *Mol. Biol. Evol.* **19**, 938–949 (2002).
49. Luo, S. *et al.* Biparental Inheritance of Mitochondrial DNA in Humans. *Proc. Natl. Acad. Sci.* **115**, 13039. <https://doi.org/10.1073/pnas.1810946115> (2018).
50. Adams, R. P. Inheritance of chloroplasts and mitochondria in Conifers: a review of paternal, maternal, leakage and facultative inheritance. *Phytologia* **101**, 134–138 (2019).
51. Wagner, D. B. Nuclear, chloroplast, and mitochondrial DNA polymorphisms as biochemical markers in population genetic analyses of forest trees. *New Forest.* **6**, 373–390 (1992).
52. Fauré, S. *et al.* Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas (*Musa acuminata*). *Curr. Genet.* **25**, 265–269. <https://doi.org/10.1007/BF00357172> (1994).
53. Hanson, M. R. Plant mitochondrial mutations and male-sterility. *Annu. Rev. Genet.* **25**, 461–486. <https://doi.org/10.1146/annurev.ge.25.120191.002333> (1991).
54. Kim, N. H. *et al.* Genome and evolution of the shade-requiring medicinal herb *Panax ginseng*. *Plant Biotechnol. J.* **16**, 1904–1917 (2018).

55. Allen, G. C., Flores-Vergara, M. A., Krasnyanski, S., Kumar, S. & Thompson, W. F. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat Protoc.* **1**, 2320–2325. <https://doi.org/10.1038/nprot.2006.384> (2006).
56. Seol, Y. J., Lee, T. H., Park, D. S. & Kim, C. K. NABIC: a new access portal to search, visualize, and share agricultural genomics data. *Evol. Bioinform.* **12**, 51–58. <https://doi.org/10.4137/Ebo.S34493> (2016).

Acknowledgments

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through the Agri-Bio Industry Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (grant number 116076-03-2-HD0b0).

Author contributions

W.K.L., H.-S.P., S.-C.L., K.S., and T.-J.Y. organized and coordinated this project. S.-C.L., H.J.J., and H.O.L. prepared the plant materials and the genomic DNA. S.-C.L. and H.J.J. assembled the chloroplast genomes, designed the molecular markers. S.-C.L., H.J.J., W.K.L. performed the molecular marker validations and genotyping. H.O.L. identified the polymorphic sites in the mitochondria genome and designed the corresponding molecular markers. J.Y.P., W.K.L., H.-S.P. performed the sequence analysis. T.-J.Y., S.K., K.S., S.-C.L., W.K.L. and H.-S.P. wrote and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-81988-w>.

Correspondence and requests for materials should be addressed to K.S. or T.-J.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021