Small, Repetitive DNAs Contribute Significantly to the Expanded Mitochondrial Genome of Cucumber

Jason W. Lilly* and Michael J. Havey†

**Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, New York 14853 and †Agricultural Research Service, U.S. Department of Agriculture, Vegetable Crops Unit, Department of Horticulture, University of Wisconsin, Madison, Wisconsin 53706*

Manuscript received September 5, 2000
Accepted for publication June 19, 2001

ABSTRACT

Closely related cucurbit species possess eightfold differences in the sizes of their mitochondrial genomes. We cloned mitochondrial DNA (mtDNA) fragments showing strong hybridization signals to cucumber mtDNA and little or no signal to watermelon mtDNA. The cucumber mtDNA clones carried short (30–53 bp), repetitive DNA motifs that were often degenerate, overlapping, and showed no homology to any sequences currently in the databases. On the basis of dot-blot hybridizations, seven repetitive DNA motifs accounted for >13% (194 kb) of the cucumber mitochondrial genome, equaling >50% of the size of the Arabidopsis mitochondrial genome. Sequence analysis of 136 kb of cucumber mtDNA revealed only 11.2% with significant homology to previously characterized mitochondrial sequences, 2.4% to chloroplast DNA, and 15% to the seven repetitive DNA motifs. The remaining 71.4% of the sequence was unique to the cucumber mitochondrial genome. There was <4% sequence colinearity surrounding the watermelon and cucumber *atp9* coding regions, and the much smaller watermelon mitochondrial genome possessed no significant amounts of cucumber repetitive DNAs. Our results demonstrate that the expanded cucumber mitochondrial genome is in part due to extensive duplication of short repetitive sequences, possibly by recombination and/or replication slippage.

PLANT mitochondrial genomes range in size from 200 to 2400 kb and are at least 10 to 100 times the size of typical animal mitochondrial genomes (reviewed in Gillham 1994). The largest plant mitochondrial genomes exist within the Angiosperms; yet these genomes possess fewer coding regions than the more compact mitochondrial genomes of the Bryophytes (Oda *et al.* 1992). The Cucurbitaceae possess the largest known plant mitochondrial genomes and closely related species show great size differences. The sizes of the mitochondrial genomes are 330, 800, 1500, and 2400 kb for watermelon (*Citrullus lanatus* L.), squash (*Cucurbita pepo* L.), cucumber (*Cucumis sativus* L.), and melon (*C. melo* L.), respectively (Ward *et al.* 1981). These size differences do not parallel size variations in the nuclear or chloroplast genomes among these same species (Palmer 1982; Havey *et al.* 1998). There is no evidence that the larger cucurbit mitochondrial genomes possess more coding regions, gene duplications, or accumulation of introns (Ward *et al.* 1981; Stern and Newton 1985; Havey 1997; Havey *et al.* 1998).

Accumulation of repetitive DNA is a key factor in genome expansion and a major contributor to the large nuclear genomes of many organisms (reviewed in Heslop-Harrison 2000). Replication slippage within simple sequence repeats increases the size of repetitive regions, contributes to genome expansion (Hancock 1996), and produces regions of long repeats identifiable by reassociation kinetics (Ward *et al.* 1981; Hancock 1996). Ward *et al.* (1981) demonstrated that only 10% of the cucumber mitochondrial genome showed a higher reassociation rate and that the larger melon mitochondrial genome did not exhibit the same amount of repetitive DNA as cucumber. Ward *et al.* (1981) also examined the maize and pea mitochondrial genomes and observed that neither had significant proportions of rapidly reassociating regions. Recombination within plant mitochondrial genomes around repetitive sequences is well established and may create chimeric genes (Sederoff 1987; Newton 1988; Palmer and Herbon 1988; Fauron 1995; Fauron *et al.* 1995). Bendich (1985) suggested that frequent recombination within the mitochondrial genome would lead to scrambling of mtDNA sequences on a fine scale. This type of recombination could duplicate repetitive regions, but continued rearrangements would lead to divergence of the duplicated regions and a loss of homology, thereby contributing to genome expansion (reviewed in Sederoff 1987).

Retrotransposons of many different classes contribute significantly to genome expansion, such as the large nuclear genome of maize (reviewed in Bennetzen 1996). Sequencing of the Arabidopsis mitochondrial genome revealed that 4% showed homology to retrotransposons (Unseld *et al.* 1997). Within sugar beet mtDNA, 21 retrotransposon-like sequences were identi-

Corresponding author: Michael J. Havey, USDA-ARS and Department of Horticulture, 1575 Linden Dr., Madison, WI 53706.
E-mail: mjhavey@facstaff.wisc.edu

fied with lengths of 25 to 2800 bp; one cluster was related to the *gypsy*-type of retrotransposons (Kubo *et al.* 2000). MtDNA sequence with homology to reverse transcriptase has also been identified within the mitochondrial genome of higher plants (Schuster and Brennicke 1987; Unseld *et al.* 1997). To date, retrotransposon accumulation has not been identified as a factor contributing to the huge Cucumis mitochondrial genomes. Similarly, the large mitochondrial genomes of the cucurbits do not possess large regions of transposed chloroplast DNA (Stern and Palmer 1984; Stern 1987; Havey 1997; Havey *et al.* 1998). Sequence analysis of the Arabidopsis and sugar beet mitochondrial genome revealed only 1.0 and 2.1%, respectively, with homology to chloroplast DNA (Unseld *et al.* 1997; Kubo *et al.* 2000).

We investigated mitochondrial genome expansion within the cucurbits using hybridizations to select mitochondrial sequences present at high copies in cucumber and at low levels in watermelon. We then sequenced 15 clones to identify sequences repeated throughout the cucumber mitochondrial genome. Additional sequence data were generated from other clones and >136 kb of random cucumber mtDNA sequence was analyzed. We also compared a conserved region of the genome from cucumber and watermelon to determine what significant homologies existed between these closely related species. A model for the accumulation and preservation of these repetitive sequences is presented.

MATERIALS AND METHODS

DNA isolations and mitochondrial libraries: Seeds from cucumber, melon, squash, and watermelon were planted in sterilized vermiculite and germinated in the dark at 30° for 4 days. Cotyledons and hypocotyls were harvested, surface sterilized with 5% bleach, rinsed three times with distilled water, and placed at 4°. Approximately 300 grams of fresh tissue was used and mtDNA was extracted using DNAse I treatments and Percoll-gradient centrifugation (Klein *et al.* 1994). Cucumber mtDNA was partially digested with *Sau*3A (Promega, Madison, WI) and subjected to electrophoresis through a 1.0% agarose gel, and fragments of 2.2–3.5 kb were excised from the gel. DNA was ligated into dephosphorylated *Bam*HI-cut pUC18 according to the manufacturer's directions (Pharmacia, Piscataway, NJ). Standard heat-shock transformation and blue/white selection were performed (Sambrook *et al.* 1989). An initial library of 3000 colonies was prepared for hybridization analysis. Colony lifts were made according to the manufacturer's directions (BioRad). Initial clone selection lifts were hybridized with either purified cucumber or watermelon mtDNA. Labeling of probes, hybridization, and autoradiography were performed according to Kennard *et al.* (1994). Clones were selected by aligning the autoradiograms with the corresponding bacterial plate, and 46 clones showing strong hybridization signals to cucumber mtDNA and no signal to watermelon mtDNA were selected. Hybridization signals were confirmed by Southern hybridization of mini-prepped plasmid DNA with cucumber and watermelon mtDNAs as probes.

For clone selection and contig assembly, additional libraries consisting of larger insert clones were generated from cucumber and watermelon. Libraries were prepared as described above, but with an average insert size of 15 kb. Colonies were picked into 384-well plates (Nunc) as described by Nizetic *et al.* (1991a). Colony filters were prepared using a 384-well plate replicator. Bacteria colonies were placed onto nylon membrane (BioRad) and grown overnight at 37° on solid Luria broth plus ampicillin (50 μg/ml). Bacterial colonies were removed, membranes washed, and DNA fixed to the nylon membrane according to Nizetic *et al.* (1991b).

Sequence duplication among selected cucumber mitochondrial clones: Homologies among selected cucumber mitochondrial clones were determined using dot-blot hybridizations. Plasmid DNA was isolated and dot-blots created using a vacuum apparatus (BioRad) applied to Zetaprobe (BioRad) nylon membrane. Plasmid DNA was doubly digested with one of the following combinations: *Pst*I/*Eco*RI, *Xba*I/*Sac*I, *Pst*I/*Sac*I, or *Xba*I/*Eco*RI according to the manufacturer's directions (Promega) to release the inserts from the vector. Insert DNAs were individually radiolabeled and hybridization conditions were according to Kennard *et al.* (1994). The dot-blots were washed (0.1× SSC + 0.2% SDS) for 30 min at 65°. Autoradiography of dot-blot membranes was carried out using the Cyclone phosphor imaging system (Packard Instruments, Meriden, CT). The 46 cucumber mtDNA clones were classified as unique (no duplication among selected clones), moderately repetitive (<60% duplication), or highly repetitive (>60% duplication). To evaluate for plastid DNA homology, we hybridized the same dot-blot membranes with gel-purified radiolabeled DNA from 14 Petunia clones covering the entire chloroplast genome (Sytsma and Gottlieb 1986). Isolation of insert DNA, radiolabeling, Southern hybridization, and autoradiography were as previously described.

Copy number and occurrence in other cucurbit species: Fifteen clones, 5 each from the unique, moderately repetitive, and highly repetitive classes, were randomly selected for hybridization and sequencing. Visual estimations of relative copy number for each clone in cucumber, melon, watermelon, and squash were assessed by hybridization to *Eco*RI digests of total DNA from three accessions of each species (Table 1).

Sequencing of mtDNA clones: DNA sequence analyses of the 15 selected clones were performed and similarities among clones established by computer analysis. Cycle sequencing reactions were performed according to the manufacturer's (ABI, Columbia, MD) directions and were analyzed on an ABI377 Prism automated DNA sequencer at the University of Wisconsin Biotechnology Center.

Computer analyses were conducted to identify both unique and repeated DNA sequences within and among our 15 selected clones. BLAST (Altschul *et al.* 1997) similarity searches were conducted using the BLASTN algorithm to both the nuclear and mitochondrial databases. Putative open reading frames (ORFs) were identified using FRAMES in the Genetics Computer Group (GCG) Wisconsin Package V. 10 (Oxford Computing). Repeats within clones were identified using the REPEAT command in GCG. Repeats among our clones were putatively identified by pairwise comparisons of each clone using a window of 25 with a similarity of 20 (80%). Repetitive sequences present in more than one clone were further analyzed using the SBLAST function in GCG. These repetitive sequences were analyzed against our entire clone data set and the number of homologies calculated.

Additional end sequence data was generated from the remaining 31 random cucumber mtDNA clones and from other clones for contig assembly around the *atp9* and *cob* regions (described below). These sequences were also analyzed for specific repeat motifs using both COMPARE and SBLAST in GCG.

Identification and analysis of the cucumber and watermelon *atp9* region and cucumber *cob* region: We hybridized the

TABLE 1

Origins and accessions of cucurbit species

Species	Number[a]	Cultivar or plant introduction number	Origin
Cucumis sativus	1	SMR-18	USDA
C. sativus	2	432860	USDA
C. sativus	3	Calypso	Asgrow
C. melo	4	Iroquois	Asgrow
C. melo	5	357775	USDA
C. melo	6	414723	USDA
Citrullus lanatus	7	Dixielee	Hollar Seeds
C. lanatus	8	New Hampshire Midget	Hollar Seeds
C. lanatus	9	269340	USDA
Cucurbita pepo	10	Dark Green Zucchini	Asgrow
C. pepo	11	Golden Summer Crookneck	Asgrow
C. pepo	12	Crown of Thorns	Jung's Seed

[a] Numbers correspond to positions on autoradiograms in Figure 1, C and D.

larger-insert cucumber and watermelon libraries with the *atp*9 (Salazar *et al.* 1991) and *cob* (Dawson *et al.* 1984) clones as previously described. Colonies showing strong signals were picked from the library and plasmid DNA was isolated (QIAGEN, Chatsworth, CA). After end sequencing of clones, a population of transposon insertions was generated and sequenced using the Genome Priming System (New England Biolabs, Beverly, MA). Internal primers were generated to fill remaining gaps. Sequence data from all clones were edited and contigs were aligned with Sequencer 3.0 (Gene Codes, Ann Arbor, MI). Sequence analysis of these contigs was completed as described above.

Genome-wide copy estimations: Genomic DNAs of one accession each of cucumber (Calypso), melon (Iroquois), watermelon (Dixielee), squash (Golden Summer), and pumpkin (Connecticut Field) were digested with *Eco*RI, subjected to electrophoresis through 0.8% agarose gels, and transferred to nylon membranes (Sambrook *et al.* 1989). Clones carrying repetitive DNA sequences were analyzed for unique restriction enzyme sites using Sequencer 3.0 (Gene Codes). Clones were digested with the appropriate restriction enzymes to separate unique from repetitive regions and subjected to electrophoresis through 1.0% agarose gels. Individual fragments were excised, purified, and used as probes to the cucurbit genomic DNAs as previously described. Hybridization intensities were estimated using the phosphorimager.

Genome copy numbers were estimated using dot-blot membranes constructed with serial dilutions on the basis of the conversion that 9.65×10^8 bp of DNA equals 1 pg of watermelon, squash, cucumber, and melon mtDNAs (Arumunganathan and Earle 1991). Oligonucleotides of the repetitive DNA motifs were synthesized at the University of Wisconsin Biotechnology Center. Four replicated membranes (Zetaprobe) were prepared using a vacuum apparatus (Bio-Rad) and carried serial dilutions of oligonucleotides (10^5, 10^6, 10^7, 10^8, and 10^9 copies) and cucurbit (10^3, 10^4, 10^5, 10^6, and 10^7 genome equivalents) mtDNAs. Oligonucleotides complementary to the repetitive DNA motifs were synthesized and hybridized as previously described. Membranes were exposed to phosphorimaging screens, activities were adjusted to remove background signals, and signal intensities were measured from the four replicated blots. These intensities were regressed on oligonucleotide concentrations and standard deviations were calculated. Relative copy numbers were estimated using the regression analysis of each motif compared to the signal intensity from serial dilutions of cucumber mtDNA. A maximum genome-wide copy number for each repetitive motif was calculated by dividing the estimated copy numbers from the regression analyses by the maximum degeneracy observed for each motif.

RESULTS

Mitochondrial clones possess repetitive DNA motifs: Forty-three clones showing strong hybridization signals with cucumber mtDNA, but not with watermelon mtDNA, were selected from the cucumber mitochondrial DNA library. Dot-blot hybridizations were used to estimate relative cross-hybridization among these 43 cucumber mtDNA clones. Twelve (28%) clones showed little or no hybridization signal to any of the other randomly selected clones and were classified as unique. Twenty clones (46%) showed signal to <60% of other clones in the sample and were classified as middle repetitive. Eleven clones (26%) were classified as highly repetitive (signal with >60% of the clones). Figure 1 illustrates the dot-blot classifications. None of the 43 clones showed signal greater than background when hybridized with a complete set of Petunia chloroplast clones (autoradiograms not shown). Fifteen cucumber mitochondrial clones were randomly selected (five each from the unique, middle repetitive, and highly repetitive clones) and used as probes to cucumber, melon, watermelon, and squash total genomic DNAs. All cucumber clones showed light-to-moderate hybridization signals to melon DNA. The relative copy numbers among the four cucurbit species in Table 2 were visually estimated on the basis of Southern hybridizations (Figure 1, C and D).

Fifteen selected clones ranged in size from 0.9 to 4.4 kb and were sequenced to generate ~39 kb of random

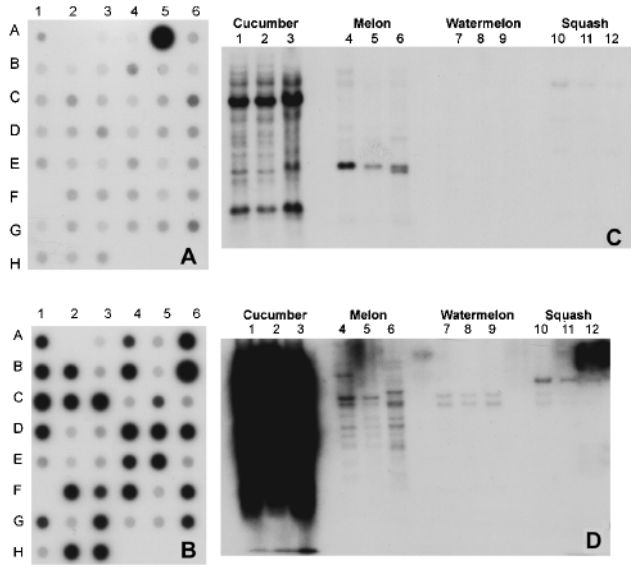

FIGURE 1.—Autoradiograms from dot-blot and Southern hybridizations of cucumber mtDNA clones. Hybridization of clone A43 (A) position A5 and C114 (B) position B6. Positions A1, A2, and A3 are mitochondrial DNAs from cucumber, watermelon, and melon, respectively. Position H1 is the pUC18 vector. Positions F1 and H3–H6 are blank. Other clone positions are listed in Table 2. Autoradiograms of Southern blots of clone A43 (C) and C114 (D) to *Eco*RI-digested genomic DNAs of three accessions of cucumber, melon, watermelon, and squash (see Table 1 for origins of accessions).

mitochondrial sequence (Table 2). The average AT content was 54%. Clone J7 possessed 1500 bp with homology to the mitochondrial *cox* I gene. Three mtDNA clones (G102, K34, and U63) possessed small (<100 bp) regions of homology to chloroplast DNA. All other sequences (37.4 kb) showed no significant homology ($P > 0.001$) to any DNA sequences in any of the databases. Excluding clone J7, no open reading frames larger than 30 amino acids (100 bp) were found among the selected clones.

Multiple types of repetitive DNA motifs were identified within individual clones: Repeats within clones were classified as direct tandem duplications (minimum of a 20-bp window at 85% stringency) with no greater than 5-bp separation or as dispersed repeats (minimum of a 25-bp window at 90% stringency) separated by >5 bp within a single clone (Table 2). Clones A43 and M102 lacked both tandem and dispersed repeats. All other clones possessed either tandem or dispersed direct repeats as listed in Table 2. Clones A10, C114, and U63 possessed multiple regions of tandem and dispersed repeats. No inverted repeats were detected in our sample.

COMPARE analyses generated dot plots of each clone to itself and were used to identify clusters of repetitive DNAs. Figure 2 illustrates both the dot plots and corresponding autoradiograms of fragments from clones C114 and U63. Clones C114 and U63 were digested (*Xba*I, *Eco*RI, and *Nde*I for C114 and *Bam*HI, *Ava*I, and *Eco*RI for U63) into three fragments each. These individual fragments did not possess any internal *Eco*RI restriction enzyme sites in cucumber. Because we hybridized to *Eco*RI-digested cucumber DNA, we can conclude that a single band on the autoradiogram represents a single copy of this region; multiple hybridization signals establish that sequences contained within these fragments occur at multiple locations in the cucumber mitochondrial genome. The repetitive regions of C114, as shown in Figure 2A, are located in the blue and yellow colored regions; the corresponding autoradiograms exhibited strong hybridization signals (smears) on cucumber and single bands in melon and squash. The green fragment, which carried no repetitive DNAs on the basis of a dot-plot comparison, revealed a signal band in cucumber mtDNA. Figure 2, F–H, shows three fragments from clone U63. The first fragment (green) possessed three repetitive clusters, including a large cluster of multiple repeats spanning 150 bp. This fragment gave intense hybridization signals to cucumber and melon DNAs, with lighter single bands in watermelon, squash, and pumpkin. The blue and yellow fragment did not possess any tandem repeats within the clone; however, both fragments revealed multiple hybridization signals to cucumber mtDNA.

Repetitive DNA sequences among all clones were identified by pairwise comparisons and similarity searches in our data set. Four (A43, S100, T106, and U38) clones did not show homologies to any of the other clones. Eleven clones (A10, B99, C114, F16, G102, I51, J7, K34, M102, R64, and U63) possessed repetitive DNA sequences in common with other clones. These latter clones exhibited moderate-to-high hybridization intensities on cucumber mtDNA. A dot-plot comparison of two clones (C114 and U63) is shown in Figure 3 and reveals that the repetitive regions in C114 (the blue and yellow fragments in Figure 2A) also exist in U63 (the green fragment in Figure 2F). These results support repetitive sequences dispersed throughout the cucumber mitochondrial genome. The yellow fragment of U63 (base pairs 1850–2500) carried some repetitive DNA (Figure 2H), but dot-plot analyses revealed no repetitive clusters with this clone.

Similarity searches among the 15 clones were used to identify the core repetitive DNA motifs that were at least 30 bp in size, existed in a minimum of 4 out of the 15 clones, and showed at least 70% identity. Seven repetitive DNA motifs were identified (Table 3). An example of the type of degeneracy within motif 5 is shown in Figure 4. None of these sequences showed any similarity to previously described repetitive sequences present within other organellar genomes (GILLHAM 1994; HILL and SINGH 1997).

An additional 64 kb of end sequence from the cucumber mitochondrial clones was generated (GenBank accession nos. AF290215–AF290301 and AF291430–

TABLE 2

Designations, copy number estimates, and homologies of cucumber mitochondrial DNA clones

Clone	Size (bp)	GenBank accession no.	% AT	Dot-blot position[a]	Homology among clones[b]	Relative copy in cucumber[c]	Homology to other cucurbits[d]	Significant BLAST score[e]	No. and size of tandem repeats[f]	Size and distance between dispersed repeats[g]
A10	4140	AF282389	56	A4	High	High	M, W, S	None	3 (15, 15, 16)	28 (480)
A43	2304	AF282390	55	A5	Low	Low	M	None	0	0
B99	2911	AF282391	55	B3	High	Low	M	None	2 (15, 15)	0
C114	1104	AF282392	59	B6	Moderate	High	M, W, S	None	2 (25, 29)	46 (196); 31 (50)
F16	1084	AF282393	57	C2	Moderate	High	M	nad2 (153)	3 (15, 20, 37)	0
G102	2935	AF282394	57	C5	High	Moderate	M, W	cpDNA (54)	0	35 (78)
I51	1626	AF282395	55	D1	Moderate	Low	M, W, S	5s rRNA (90)	0	30 (469)
J7	3700	AF282396	55	56	Low	Low	M, W, S	coxI (1635)	0	26 (963)
K34	4467	AF282397	56	D5	High	High	M, W, S	cpDNA (147)	0	29 (88); 28 (108); 25 (117)
M102	1704	AF282398	58	E1	Low	High	M	None	0	0
R64	952	AF282399	54	F3	Moderate	Moderate	M	None	1 (18)	0
S100	2929	AF282400	56	F4	Low	Low	M	None	0	26 (37); 25 (49)
T106	3153	AF282401	62	F5	Low	Low	M	None	1 (16)	0
U38	3523	AF282402	57	G1	Moderate	Moderate	M	None	1 (23)	25 (57)
U63	2459	AF282403	57	G3	Moderate	High	M, W	tRNA-Ser (232)	3 (23, 25, 43)	33 (13); 40 (460); 30 (470)

[a] Dot blots shown in Figure 1.
[b] Relative homologies among clones based on dot-blot hybridizations (Figure 1).
[c] Estimation of relative copy numbers in cucumber based on DNA gel-blot hybridizations (Figure 1).
[d] Homology to melon (M), watermelon (W), or squash (S) as determined by DNA gel-blot hybridizations (Figure 1).
[e] Significant similarities based on BLAST analysis. Score in parentheses.
[f] Presence of tandem repeats of minimum size of 15 bp, no greater than 5 bp between repeat. Numbers indicate frequencies; sizes given in parentheses.
[g] Sizes of repeats; distance between repeats given in parentheses.

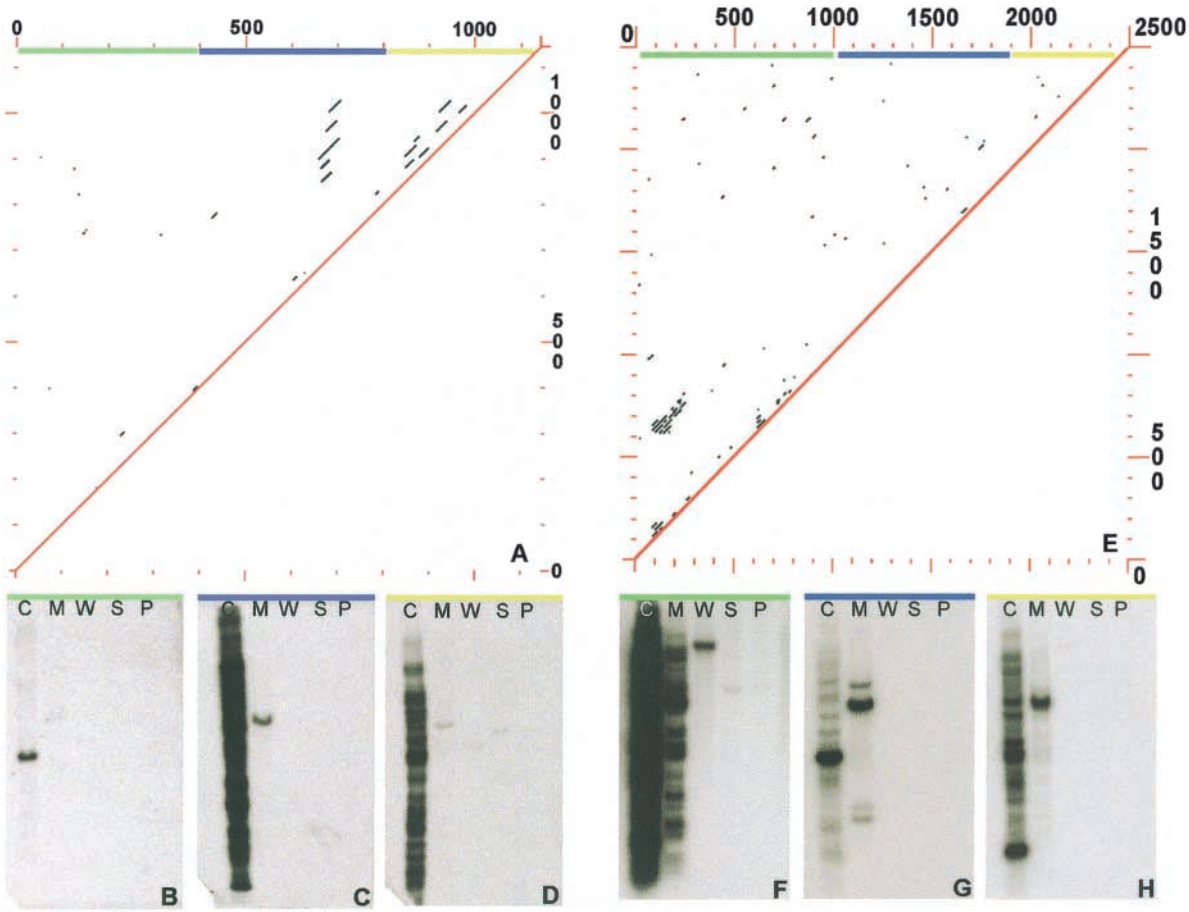

FIGURE 2.—Dot-plots of cucumber mitochondrial DNA clone C114 (A) and U63 (E), generated using a 25-bp window and 24-bp stringency. Color bars represent fragments from C114 (B–D) and U63 (F–H) that were isolated after restriction-enzyme digestions and hybridized to *Eco*RI-digested cucurbit DNAs in autoradiograms B–D. Lane orders are cucumber (C), melon (M), watermelon (W), squash (S), and pumpkin (P).

AF291435). We searched this data set for the seven repetitive DNA motifs. BLAST analyses revealed homologies of 2100 bp with cpDNA homology and 4800 bp with the mitochondrial genes *atp*α, *cox*II, or *nad*5. Dot plots revealed that 10 end sequences possessed clusters of repetitive DNA and SBLAST showed 9 of these 10 were homologous to our previously classified repetitive DNA motifs.

Dot-blot hybridizations were performed to quantify relative amounts of the seven repetitive motifs in the cucumber mitochondrial genome. Signal intensities from replicated serial dilutions of the cucumber mtDNA were regressed against signal intensities from known oligonucleotide concentrations. The seven short repetitive DNA motifs accounted for an average of 164 kb or 11% of the cucumber mitochondrial genome (Table 3). Figure 5 shows one replication of the dot-blot hybridization for repetitive DNA motif 5. These copy-number estimates are a minimum because the stringency of our washes eliminated duplexes with more than three mismatches (BRITTEN and DAVIDSON 1985). We estimated the maximum amount of each repetitive DNA motif on the basis of the degeneracies observed among our motif classes. A PILEUP of all clones possessing a particular motif was performed and the maximum number of degenerate positions determined (Figure 4). As an example, motif 5 had up to 10 possible degenerate positions (19.2%), but our hybridizations could detect only up to three (5.8%) mismatches. Therefore, under our conditions, we may not detect up to 13.4% degenerate copies of this repetitive DNA motif in the mitochondrial genome. Using high stringency washes, we estimated that 1063 copies (55.2 kb) of repetitive DNA motif 5 exist in the cucumber mitochondrial genome. Assuming an equal distribution of degenerate copies, the maximum estimate would be 13.4% larger or 1234 copies (64.1 kb). These estimates were completed for each of the repetitive DNA motifs (Table 3), increasing the maximum genome-wide amount of the seven motifs at ~194 kb or 13% of the cucumber mitochondrial genome.

Presence of repetitive sequences surrounding coding regions in the cucumber mitochondrial genome: We established the presence of the seven repetitive DNA motifs around two mitochondrial coding regions in cu-

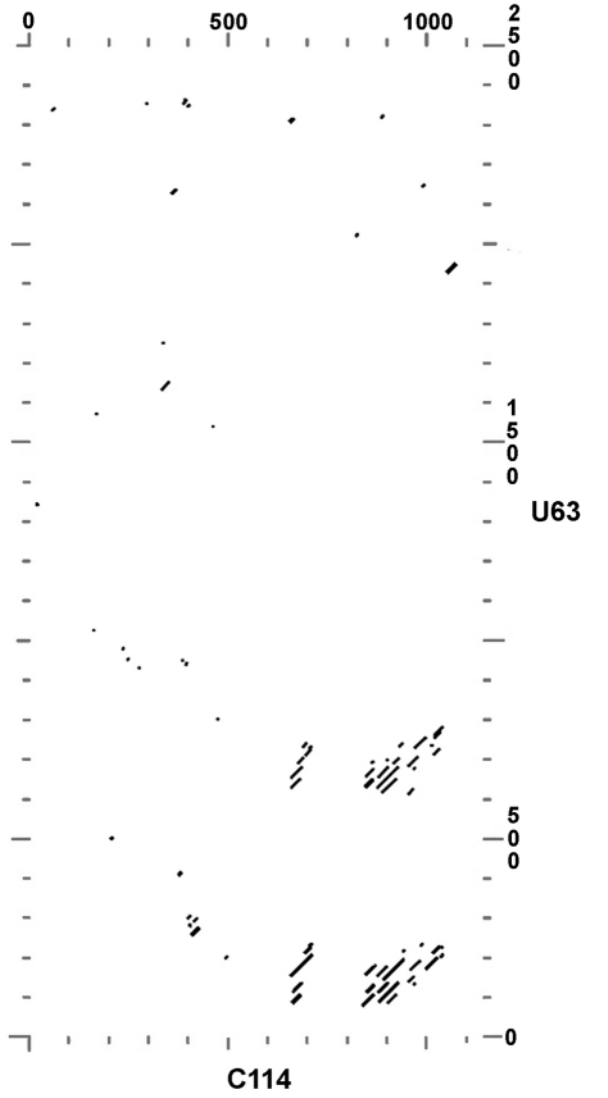

FIGURE 3.—Dot-plot comparison of clone C114 (1104 bp) and U63 (2459 bp), generated using a 25-bp window and 24-bp stringency.

cumber, a 13-kb clone (GenBank accession no. AF288043) possessing both the *atp*9 and *atp*6 genes and a 16-kb clone (GenBank accession no. AF288044) possessing the *cob* gene (Table 4). Sequencing revealed three repetitive regions within the *atp*9 clone and one repetitive region flanking the 3′ end of the *cob* gene. The *atp9* clone possessed four of the seven repetitive DNA motifs. Within the *atp*9 clone, the region from 3.1 to 3.4 kb was repeated in the region from 5.3 to 5.6 kb and partially repeated in the region from 8.3 to 8.6 kb. The *cob* clone possessed three repetitive DNA motifs in regions 0.5 to 0.6 kb and 12.1 to 12.3 kb. Across 29 kb of mitochondrial sequence from these two clones, we found 4.0 kb (13.7%) of coding sequence, 2.5 kb (8.6%) of sequence corresponding to our repetitive motifs, and 22.8 kb (77.7%) of DNA unique to the mitochondrial genome of cucumber. Dot-plot analyses of the cucumber cob

TABLE 3

Repetitive DNA motifs in the cucumber mitochondrial DNA and genome-wide copy estimations

Motif ID	Repetitive DNA sequence	Maximum degeneracy (%)	Copies in clone population	Genome-wide copies[a]	% genome[b]
1	TAGGTTTTGGGCCTATAAAGTAGCAAGAAATCATGAAAAAT	21	4	143 ± 33	0.4% (0.5)
2	CTTCGTTTTTTTCGGCCCTTTTGCGTTACTTTCTT	27	5	204 ± 26	0.4% (0.6)
3	TATGGATAGTCCTACTTCGGTTAATTACGGATG	18	5	761 ± 166	1.7% (1.9)
4	GTCCGGACTCCGGAGGATTCTTATCCGTAGGACTATCCCTATGGAAGTAG	29	6	486 ± 85	1.7% (2.1)
5	CGGACTACTTCCCTCCATTATGGATGGTCCTACGGATAAATCCTCCGACTCC	19	13	1063 ± 250	3.7% (4.0)
6	AGGACTATCCATATGGAAAGAAGGGTCCAGGTTTCATAAGACTCCTCCCTA	25	4	85 ± 4	0.3% (0.4)
7	TTCATCCTGGTAGGCATTAATTGGCTTCATCCTGAG	27	7	1154 ± 33	2.8% (3.6)

[a] Genome-wide estimates of copy numbers ± standard deviations were based on dot-blot hybridizations across four replicated blots (see MATERIALS AND METHODS).
[b] Maximum estimates based on sequence degeneracy are given in parentheses.

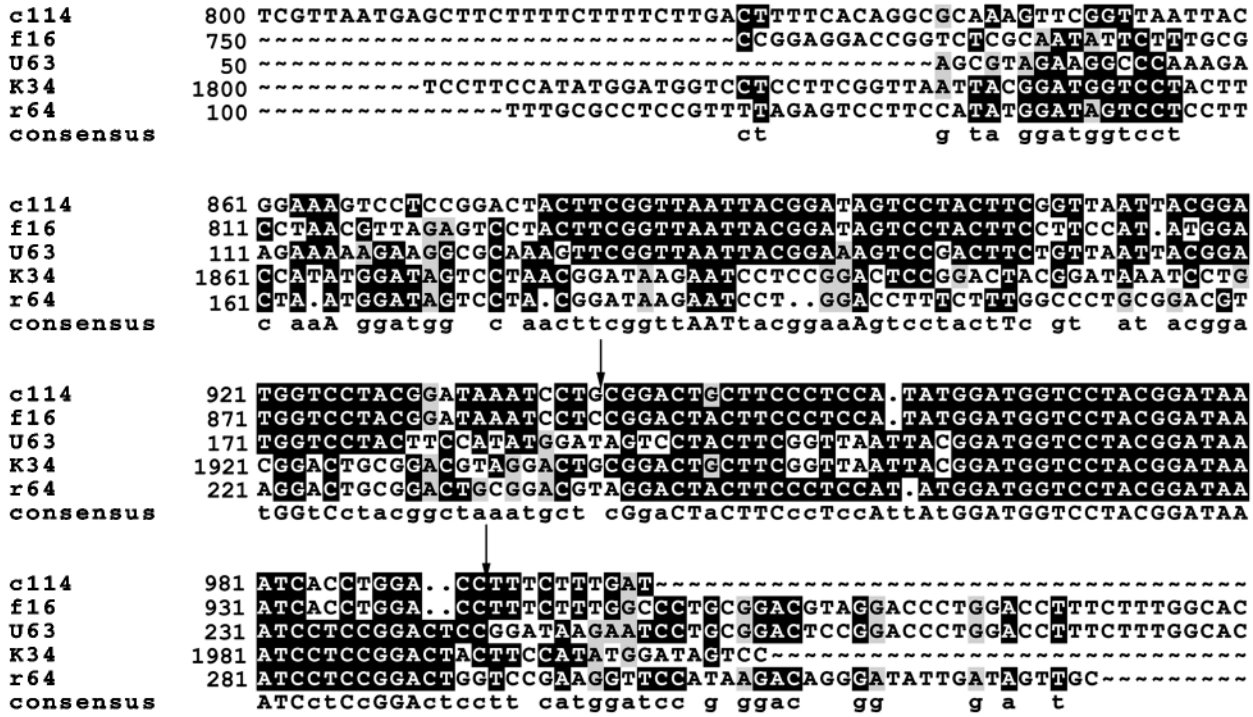

FIGURE 4.—A BOXSHADE representation of the repetitive DNA motif 5 (between arrows) from five different mitochondrial DNA clones of cucumber. Arrows delineate the region used as an oligonucleotide probe to estimate genome-wide copy number of this repeat.

and atp6-atp9 contigs revealed that these unique regions possessed no additional tandemly duplicated repetitive DNA motifs, although these regions of unique DNA could possess dispersed repetitive DNAs not identified in our sample. By sequencing >136 kb of the cucumber mitochondrial genome, we determined that 15.2 kb (11.2%) showed homology to previously sequenced mitochondrial DNA from other species, 3.3 kb (2.4%) had homology to cpDNA, and >30% of the remaining 117.5 kb of mitochondrial sequence showed homology to our repetitive DNA motifs. The remaining 82.3 kb (60%) of sequence showed no significant homology to our repetitive motifs, nor to previously sequenced mitochondrial DNA.

There is only 4% sequence similarity surrounding the cucumber and watermelon *atp*9 coding region: We compared the *atp*9 regions from cucumber and watermelon (GenBank accession no. AF288042). The watermelon clone carrying *atp*9 was 11.9 kb in size and had a higher gene content than cucumber, including 4.2 kb corresponding to coding regions for *atp*9, *nad*-5 exons a and b, *nad*-9, and tRNA-G. An additional 4.4 kb of the watermelon clone was homologous to regions of the Arabidopsis or sugar beet mitochondrial genomes (UNSELD *et al.* 1997; KUBO *et al.* 2000). The cucumber mitochondrial clone surrounding the *atp*9 region was 13.2 kb and contained only 0.9 kb (6.8%) of coding regions; the rest of the sequence had no homology to any sequences in the databases. There was one tandem repeat of 200 bp within the watermelon *atp*9 region that was not present in the cucumber clone. The overall similarity between the two regions of the cucurbit mitochondrial genome was <4% and the only significant homology ($P < 0.001$) was between the two *atp*9 coding regions (Figure 6).

DISCUSSION

The cucumber mitochondrial genome possesses unique repetitive sequences: Forty-three cucumber mitochondrial clones were selected on the basis of strong hybridization intensities to cucumber mtDNA and no relative signal to watermelon mtDNA. Less than 10% of the DNA sequence from 15 of these 43 clones showed homology to previously sequenced mitochondrial genomes (ODA *et al.* 1992; UNSELD *et al.* 1997; KUBO *et al.* 2000). TOOTHMAN *et al.* (1988) found sequences surrounding the 18s mitochondrial gene of watermelon and muskmelon to be so highly diverged that they were species specific. Our sequencing of 136 kb of cucumber mtDNA revealed only 11.2% with homology to previously sequenced mitochondrial DNAs, indicating that the cucumber mitochondrial genome possesses vast regions of species-specific sequences of noncoding or unknown function. WARD *et al.* (1981) estimated the GC content of cucumber mtDNA at 43.2%, agreeing with our estimate of 43.4% after sequencing 9% of the genome (Table 2). The relatively high AT content of the huge Cucumis mitochondrial genomes originally led to

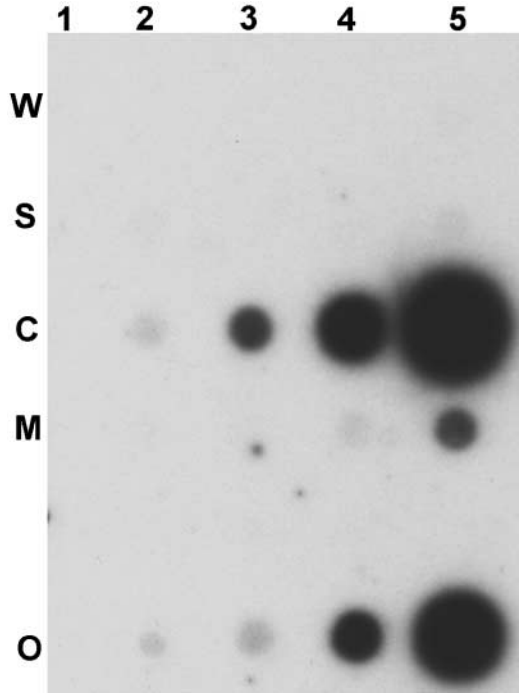

FIGURE 5.—Dot-blot autoradiogram probed with repetitive DNA motif 5 (Table 3). Watermelon (W), squash (S), cucumber (C), and melon (M) mitochondrial DNAs were diluted to 10^3 (1), 10^4 (2), 10^5 (3), 10^6 (4), and 10^7 (5) genome equivalents. O refers to the dilutions of the complementary oligonucleotide at 10^5 (1), 10^6 (2), 10^7 (3), 10^8 (4), and 10^9 (5) copies.

their discovery by density centrifugation (BENDICH and ANDERSON 1974; WARD *et al.* 1981).

We identified many tandem duplications of degenerate sequences predominating as clusters of repeats within repeats. Southern hybridizations established that these repetitive regions were reiterated throughout the cucumber mitochondrial genome (Figure 2). Greater than 90% of all fragments carrying unique (*i.e.*, no repeats revealed by dot-plot analyses) sequences hybridized to single or relatively few fragments. Among our 15 selected clones, 9 contained regions of repetitive DNA. The sequences of these repeats were difficult to align due to degeneracy and no homologies to any sequences in the databases (Tables 2 and 3). Therefore, we may have missed variants of these sequences or additional repetitive motifs. Multiple sequence alignments revealed seven repetitive DNA motifs >30 bp, with at least 70% homology, and present in at least 4 of our 15 clones (Table 3 and Figure 4). We eliminated at least six other potential repetitive motifs because they met only two of these three criteria. The seven repetitive DNA motifs accounted for 11% of the cucumber mitochondrial genome (Table 3). This is a minimal estimate because we would not detect sequences divergent beyond our wash stringency (>3 bp). WARD *et al.* (1981) noted that the larger Cucumis mitochondrial genomes contain repetitive DNA sequences; however, reassociation kinetics revealed no correlation between genome size and the amount of repetitive DNAs. Ten percent of the cucumber mitochondrial genome reassociated more rapidly, while the larger mitochondrial genome of muskmelon (2400 kb) possessed only 5% rapidly reassociating DNA (WARD *et al.* 1981). Reassociation kinetics detect highly conserved repetitive DNAs and the rate of reassociation is decreased twofold for every 10% sequence mismatch (BRITTEN and DAVIDSON 1985). Our short repetitive DNA motifs would not be classified as rapidly reassociating due to higher levels (19 to 29%) of degeneracy (Table 3 and Figure 4). Light hybridization signals to melon mtDNA were detected on dot-blot hybridizations, revealing that some repetitive sequences are present in both of these large Cucumis mitochondrial genomes. The signal intensities were always less in melon, suggesting that these sequences are not repetitive in melon or have degenerated to the point of little hybridization. The repetitive DNA motifs from cucumber were not detected by hybridizations in the smaller squash and watermelon mitochondrial genomes (Figure 5).

The cucumber mitochondrial genome is a sea of repetitive DNA with islands of coding sequence: Sequences flanking the cucumber *atp*9 and *cob* regions carried the repetitive DNA motifs (Figure 6) and showed no homology to the Arabidopsis, Beta, or Marchantia mitochondrial genomes. Although cucumber and watermelon possess similarly sized nuclear and chloroplast genomes (HAVEY *et al.* 1998), their mitochondrial ge-

TABLE 4

Presence of repetitive DNA motifs adjacent to the cucumber *atp*9 and *cob* mitochondrial coding regions

Clone name	Size (bp)	GenBank accession no.	Coding region	% AT	Tandem repeats within clone[a]	Size and distance between dispersed repeats[b]	Repeat motifs within clone[c]
Cuke*atp*9	13,291	AF288043	*atp*9 and *atp*6	55	12 (67–30 bp)	9 (62–30 bp)	1, 3, 4, 5
Cuke*cob*	16,042	AF288044	*cob*	56	2 (40, 34 bp)	7 (61–30 bp)	1, 2, 5

[a] Number of different tandem repeats no greater than 5 bp apart (unit sizes).
[b] Distant repeats are 25–100 bp sequences that are repeated within the clone at least once. Sizes of repeat clusters; distances between clusters shown in parentheses.
[c] Motif ID (see Table 3).

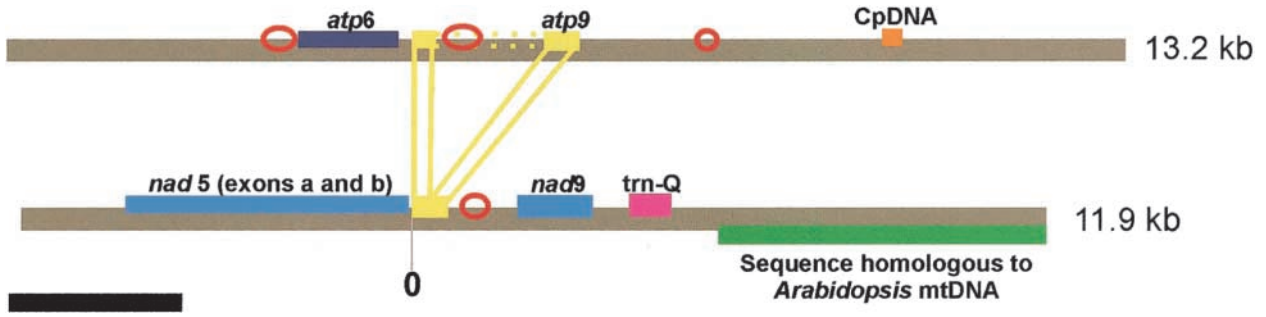

FIGURE 6.—Sequence similarity between the cucumber (top) and watermelon (bottom) clones carrying the *atp*9 coding region; position 0 represents the ATG start site of the *atp*9 coding region. Sequences with significant sequence homology represented with colored bars. Lines between cucumber and watermelon show the region of significant sequence similarity. Red circles indicate regions of repetitive DNA motifs. Bar, 2 kb.

nomes show a fivefold size difference (WARD *et al.* 1981) and share only 4% sequence homology surrounding the *atp*9 region (Figure 6). None of the seven repetitive DNA motifs identified from cucumber were observed within the watermelon *atp*9 clone. There was a much greater sequence colinearity (~75%) between the watermelon *atp*9 region and Arabidopsis or sugar beet mtDNA than between watermelon and cucumber (4%). This result supports our conclusion that the mitochondrial genomes among closely related cucurbits are changing rapidly.

The possibility exists that we compared an *atp*9 pseudogene against an actual coding region or compared two pseudogenes, especially because of the large intron discovered in the cucumber *atp*9 clone (Figure 6). Earlier work by HAVEY *et al.* (1998) used DNA gel-blot hybridizations to search for duplicated coding regions across the watermelon and cucumber mitochondrial genomes. On the basis of hybridization analyses across 10 different restriction enzymes, there was no evidence for a duplicated *atp*9 region in cucumber. However, this analysis would not reveal tandem duplications. We sequenced a 13.2-kb cucumber mtDNA clone containing the *atp*9 sequence and did not detect any duplicated sequences, eliminating the possibility of an *atp*9 pseudogene. Restriction fragment length polymorphism banding patterns from genomic DNA hybridized with *atp*6 and *atp*9 clones showed identical banding patterns for cucumber, but not for watermelon (HAVEY *et al.* 1998), and the close proximity of these coding regions in cucumber mtDNA was confirmed by our sequence analysis (Figure 6).

Mitochondrial genome expansion in cucumber: BENDICH (1985) proposed that the amplification and reshuffling of the same pieces of noncoding mtDNA could generate the huge Cucumis mitochondrial genomes. Repetitive DNA motifs may increase in copy number by duplication following unequal recombination or replication slippage. It has also been shown that repetitive DNAs containing degenerate sequences will not reassociate as quickly as highly conserved tandem or dispersed repetitive elements, thus reducing copy number estimates (BRITTEN and DAVIDSON 1985). Therefore the 10% rapidly reassociating fraction of cucumber mtDNA reported by WARD *et al.* (1981) would not include the entire complement of repetitive DNA motifs detected in our survey. The sizes of repetitive seqeuences detected by WARD *et al.* (1981) were estimated at 3 to 5 kb; our repetitive DNA motifs were much smaller (31 to 54 bp).

The role of intergenomic transfer has been discussed as a possible mechanism for mitochondrial genome expansion (HAVEY *et al.* 1998). STERN (1987) investigated the presence of five chloroplast genes (*rrn*16, *rrn*23, *rbc*L, *atp*β, and *atp*E) in the mitochondrial genome of the four cultivated cucurbits, and there was no correlation between genome size and presence of these chloroplast sequences. Our sequence analyses revealed that only 3.3 kb (<5%) of 136 kb of cucumber mtDNA showed homology to chloroplast DNA. Intergenomic transfers were not important in expansion of the cucumber mitochondrial genome because there was no evidence for significant amounts of plastid DNA sequences and no evidence of any retrotransposon-like or putative nuclear sequences based on BLAST homologies.

All mitochondrial genomes appear to be exporting sequences to the nucleus (reviewed in BLANCHARD and LYNCH 2000). It could be that the huge Cucumis mitochondrial genomes are shrinking, but at a much slower rate. This would require that the progenitor mitochondrial genome was even larger and that the cucurbit mitochondrial genomes are reducing their sizes at different rates. Another possibility is that Cucumis lacks the ability to export DNA out of the mitochondrion. However, this is unlikely because mitochondrial sequences have been detected in all plant nuclear genomes analyzed to date (THORSNESS and WEBER 1996).

The complete mitochondrial sequences of two divergent Angiosperms (*Arabidopsis thaliana* and *Beta vulgaris*) revealed that 60 and 56%, respectively, of these genomes have no obvious function (UNSELD *et al.* 1997; KUBO *et al.* 2000). If this were the same in the Cucumis

mitochondrial genomes, 900 to 1400 kb would be superfluous DNA, a massive amount of unnecessary DNA for the mitochondrion to replicate and maintain. Our sequence data from cucumber surrounding the *atp9* and *cob* coding regions, together with partial homologies in our end sequence data set, revealed high homologies ($P < 0.001$) among coding regions of cucumber and other higher plant mitochondrial genomes. If all the coding sequences in Arabidopsis or sugar beet mtDNA were also present in cucumber, genes with introns and ORFs would account for only 7% (102 kb) of the cucumber mitochondrial genome. The Arabidopsis mitochondrial genome contains 7% duplicated regions and 5% plastid and retrotransposon-like sequences (Unseld *et al.* 1997). On the basis of this and previous research (Ward *et al.* 1981; Stern and Newton 1985; Havey *et al.* 1998), we estimate that coding sequences with introns, plastid DNA transfers, and retrotransposon-like sequences could represent ~180 kb (12%) of the cucumber mitochondrial genome, leaving 1320 kb (88%) uncharacterized. Ward *et al.* (1981) proposed that the cucumber mitochondrial genome possesses 250 kb (17%) of rapidly reassociating repetitive DNA. We identified seven short repetitive DNA motifs accounting for ~11% (164 kb) of the cucumber mitochondrial genome. Assuming that degenerate repetitive DNA motifs are randomly distributed across the genome, we may have underestimated the copy number of each repetitive DNA motif by up to 26%, resulting in a maximum estimate of 194 kb (13%). This amount of repetitive DNA is equivalent to >50% of the total mtDNA of Arabidopsis or sugar beet. Together, these estimates account for up to 42% of the cucumber mitochondrial genome. Our survey of cucumber mitochondrial DNA sequence (136 kb) revealed that the majority (60%) was homologous to neither our repetitive DNA motifs nor to previously characterized sequences. On the basis of these results, we propose that the remaining unclassified sequence in the cucumber mitochondrial genome consists of both repetitive DNAs and species-specific sequences of unknown origin and prevalence.

The Cucumis mitochondrial genomes may be comparable to the massive nuclear genomes of Gymnosperms and the Liliaceae of the Angiosperms (Price 1976). These huge nuclear genomes have accumulated massive amounts of both retrotransposons and satellite DNAs (reviewed in Heslop-Harrison 2000). Like these huge nuclear genomes, the enormous Cucumis mitochondrial genomes do not appear to reduce the fitness of the organism when compared to the closely related cucurbits with much smaller mitochondrial genomes.

We thank Drs. Jiming Jiang, Philipp Simon, and David Stern for their critical reviews of the manuscript. Disclaimer: Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by USDA implies no approval of the product to the exclusion of others that may also be suitable.

LITERATURE CITED

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Arumunganathan, K., and E. D. Earle, 1991 Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. **9:** 208–219.

Bendich, A. J., 1985 Plant mitochondrial DNA: unusual variation on a common theme, pp. 111–138 in *Genetic Flux in Plants*, edited by B. Hohn and E. S. Dennis. Springer-Verlag, Vienna.

Bendich, A. J., and R. S. Anderson, 1974 Novel properties of satellite DNA from muskmelon. Proc. Natl. Acad. Sci. USA **71:** 1511–1515.

Bennetzen, J. L., 1996 The contributions of retroelements to plant genome organization, function and evolution. Trends Microbiol. **4:** 347–353.

Blanchard, J. L., and M. Lynch, 2000 Organellar genes: Why do they end up in the nucleus? Trends Genet. **7:** 315–320.

Britten, R. J., and E. H. Davidson, 1985 Hybridization strategy, pp. 3–15 in *Nucleic Acid Hybridization: A Practical Approach*, edited by B. D. Hames and S. J. Higgins. IRL Press, Oxford.

Dawson, A. J., V. P. Jones and C. J. Leaver, 1984 The apocytochrome-B gene in maize mitochondria does not contain introns and is preceded by a potential ribosome binding-site. EMBO J. **3:** 2107–2113.

Fauron, C. M., 1995 Maize as a model of higher plant mitochondrial genome plasticity. Plant Sci. **112:** 11–32.

Fauron, C., M. Casper, Y. Gao and B. Moore, 1995 The maize mitochondrial genome: dynamic, yet functional. Trends Genet. **11:** 228–235.

Gillham, N., 1994 *Organelle Genes and Genomes*. Oxford University Press, New York.

Hancock, J. M., 1996 Simple sequences and the expanding genome. Bioessays **18:** 421–425.

Havey, M. J., 1997 Predominant paternal transmission of the mitochondrial genome in cucumber. J. Hered. **88:** 232–235.

Havey, M. J., J. D. McCreight, B. Rhodes and G. Taurick, 1998 Differential transmission of the cucurbit organellar genomes. Theor. Appl. Genet **97:** 122–128.

Heslop-Harrison, J. S., 2000 Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. Plant Cell **12:** 617–636.

Hill, K. A., and S. M. Singh, 1997 The evolution of species-type specificity in the global DNA sequence organization of mitochondrial genomes. Genome **40:** 342–356.

Kennard, W. C., K. Poetter, A. Dijkhuizen, V. Meglic, J. E. Staub *et al.*, 1994 Linkages among RFLP, RAPD, isozyme, disease resistance, and morphological markers in narrow and wide crosses of cucumber. Theor. Appl. Genet. **89:** 42–48.

Klein, M., U. Eckertossenkopp, I. Schmiedeberg, P. Brandt, M. Unseld *et al.*, 1994 Physical mapping of the mitochondrial genome of Arabidopsis thaliana by cosmid and YAC clones. Plant J. **6:** 447–455.

Kubo, T., S. Nishizawa, A. Sugawara, N. Itchoda, A. Estiati *et al.*, 2000 The complete nucleotide sequence of the mitochondrial genome of sugarbeet (*Beta vulgaris* L.) reveals a novel gene for $tRNA^{CYS}$ (GCA). Nucleic Acids Res. **28:** 2571–2576.

Newton, K. J., 1988 Plant mitochondrial genomes: organization, expression and variation. Annu. Rev. Plant Physiol. Plant Mol. Biol. **39:** 503–532.

Nizetic, D., G. Zehetner, A. P. Monaco, L. Gellen, B. D. Young *et al.*, 1991a Construction, arraying, and high-density screening of large insert libraries of human chromosome-X and chromosome-21: their potential use as reference libraries. Proc. Natl. Acad. Sci. USA **88:** 3233–3237.

Nizetic, D., R. Drmanac and H. Lehrach, 1991b An improved bacterial colony lysis procedure enables direct DNA hybridisation using short (10 bases, 11 bases) oligonucleotides to cosmids. Nucleic Acids Res. **19:** 182–182.

Oda, K., K. Yamato, H. Ohta, Y. Nakamura, M. Takemura *et al.*, 1992 Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome. J. Mol. Biol. **223**: 1–7.

Palmer, J. D., 1982 Physical and gene-mapping of chloroplast DNA

from *Atriplex triangularis* and *Cucumis sativa*. Nucleic Acids Res. **10:** 1593–1605.

Palmer, J. D., and L. A. Herbon, 1988 Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J. Mol. Evol. **28:** 87–97.

Price, H. J., 1976 Evolution of DNA content in higher plants. Bot. Rev. **42:** 27–52.

Salazar, R. A., D. R. Pring and F. Kempken, 1991 Editing of mitochondrial atp9 transcripts from 2 sorghum lines. Curr. Genet. **20:** 483–486.

Sambrook, J., E. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schuster, W., and A. Brennicke, 1987 Plastid, nuclear and reverse transcriptase sequences in the mitochondrial genome of *Oenothera*: Is genetics information transferred between organelles via RNA? EMBO J. **6:** 2857–2863.

Sederoff, R. R., 1987 Molecular mechanisms of mitochondrial-genome evolution in higher plants. Am. Nat. **130:** 30–45.

Stern, D. B., 1987 DNA transposition between plant organellar genomes. J. Cell Sci. **7** (Suppl.): 145–154.

Stern, D. B., and K. J. Newton, 1985 Mitochondrial gene expression in Cucurbitaceae: conserved and variable feature. Curr. Genet. **9:** 395–405.

Stern, D. B., and J. D. Palmer, 1984 Extensive and widespread homologies between mitochondrial-DNA and chloroplast-DNA in plants. Proc. Natl. Acad. Sci. USA **81:** 1946–1950.

Sytsma, K. J., and L. D. Gottlieb, 1986 Chloroplast DNA evolution and phylogenetic relationships in Clarkia sect. *Peripetasma*. Evolution **40:** 1248–1261.

Thorsness, P. E., and E. R. Weber, 1996 Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. Int. Rev. Cytol. **165:** 207–234.

Toothman, P., S. O. Rogers, K. Stuart and A. J. Bendich, 1988 Species specificity of mitochondrial DNA flanking the 18rRNA gene in two species of cucurbitaceae. Plant Sci. **55:** 33–42.

Unseld, M., J. R. Marienfeld, P. Brandt and A. Brennicke, 1997 The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366, 924 nucleotides. Nat. Genet. **15:** 57–61.

Ward, B. L., R. S. Anderson and A. J. Bendich, 1981 The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). Cell **25:** 793–803.

Communicating editor: K. J. Newton