Theor Appl Genet (1985) 71:417–429

The *Cucumis* plastome: physical map, intrageneric variation and phylogenetic relationships

R. Perl-Treves and E. Galun
Department of Plant Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

Accepted June 10, 1984
Communicated by D. von Wettstein

Summary. A restriction map of the *Cucumis melo* L. (melon) plastome was constructed by using several mapping approaches: single and double digestions of the chloroplast DNA (chlDNA) with endonucleases (XhoI, SmaI, SacI and PvuII) and hybridization to heterologous chlDNA probes and to isolated melon chlDNA fragments. Four plastome-coded genes were located using heterologous probes. The overall organization and gene position of the melon plastome was found to be similar to that of tobacco and other angiosperm species. Restriction patterns based on digestion of the chlDNA with nine endonucleases were obtained in over 20 wild species and cultivated varieties of *Cucumis*. These led to mutational analysis of the restiction sites yielding the most parsimonious phylogenetic tree of the *Cucumis* plastome. Most African species from a compact group ("Anguria group") which is distant from the melon, the cucumber and a few other species (*C. sagittatus, C. metuliferus* and *C. humifructus*). All of these are also far apart from each other. The distribution of polymorphic restriction sites along the *Cucumis* plastome is described and conservative regions as well as "hot spots" are suggested.

Key words: Chloroplast DNA – *Cucumis* – Restriction-patterns – Phylogeny – Plastome – Parsimonious tree

Introduction

The genus *Cucumis* includes two distinct sets of species, differing in their origin and basic chromosome number (see Frankel and Galun 1977). The African group has 2n=24 (or its polyploid versions), and includes the melon (*C. melo*) and most other species in this study. Cucumber (*C. sativus*) and *C. hardwickii* represent in our study the South-Asian group, and have 2n=14 chromosomes.

Wild *C. melo* varieties are found in Africa and South Asia, and belong to 2 races: *C. melo agrestis* and *C. melo melo*. The evolution within the genus, and in particular the relationship between species having different chromosome numbers are not clear. Except for cucumber and *C. hardwickii,* barriers between the cultivated species and their wild relatives in the genus are high and crosses between them commonly failed to produce fertile F1.

Physical maps of the plastome, based on chlDNA restriction patterns, were constructed in recent years for numerous species of monocot and dicot plant families (see Vedel and Mathieu 1983). Palmer (1982) derived a cucumber chlDNA map using PvuII and SalI as restriction endonucleases.

ChlDNA restriction patterns were also used to investigate taxonomic relations and the evolution of angiosperm species. These phylogenetic studies, based on chlDNA variation, were published for *Nicotiana* (Rhodes et al. 1981), *Brassica* (Erickson et al. 1983; Palmer et al. 1983b), *Lycopersicon* (Palmer and Zamir 1982), *Triticum* and *Aegilops* (Bowman et al. 1983; Terachi et al. 1984), *Coffea* (Berthou et al. 1983), *Pennisetum* (Clegg et al. 1984) and *Solanum* (Hosaka et al. 1984). These studies varied in the number of restriction endonucleases employed and the method of analyzing the patterns.

The physical plastome map and the chloroplast phylogenetic data obtained in the present study should serve as tools in the genetic investigation of *Cucumis* at the molecular level. A practical aim would be to utilize the knowledge on plastome relatedness between cultivated and wild species, in order to enable the increase the genetic resources of cultivated melon and cucumber by introgression of wild genes, possibly by a somatic fusion approach (Galun 1984). The restriction patterns could also serve as plastome markers in the hybrids, and for evaluating genetic distance between fusion partners.

In a subsequent article we shall present a *Cucumis* phylogeny based on isozyme analysis. From the evolutionary and systematic point of view, it is interesting to compare the phylogenies, based on chlDNA and on nuclear-coded isozymes, for a given taxon. The implications of such a bifurcate study on plastome and nuclear evolution shall be discussed in the second article.

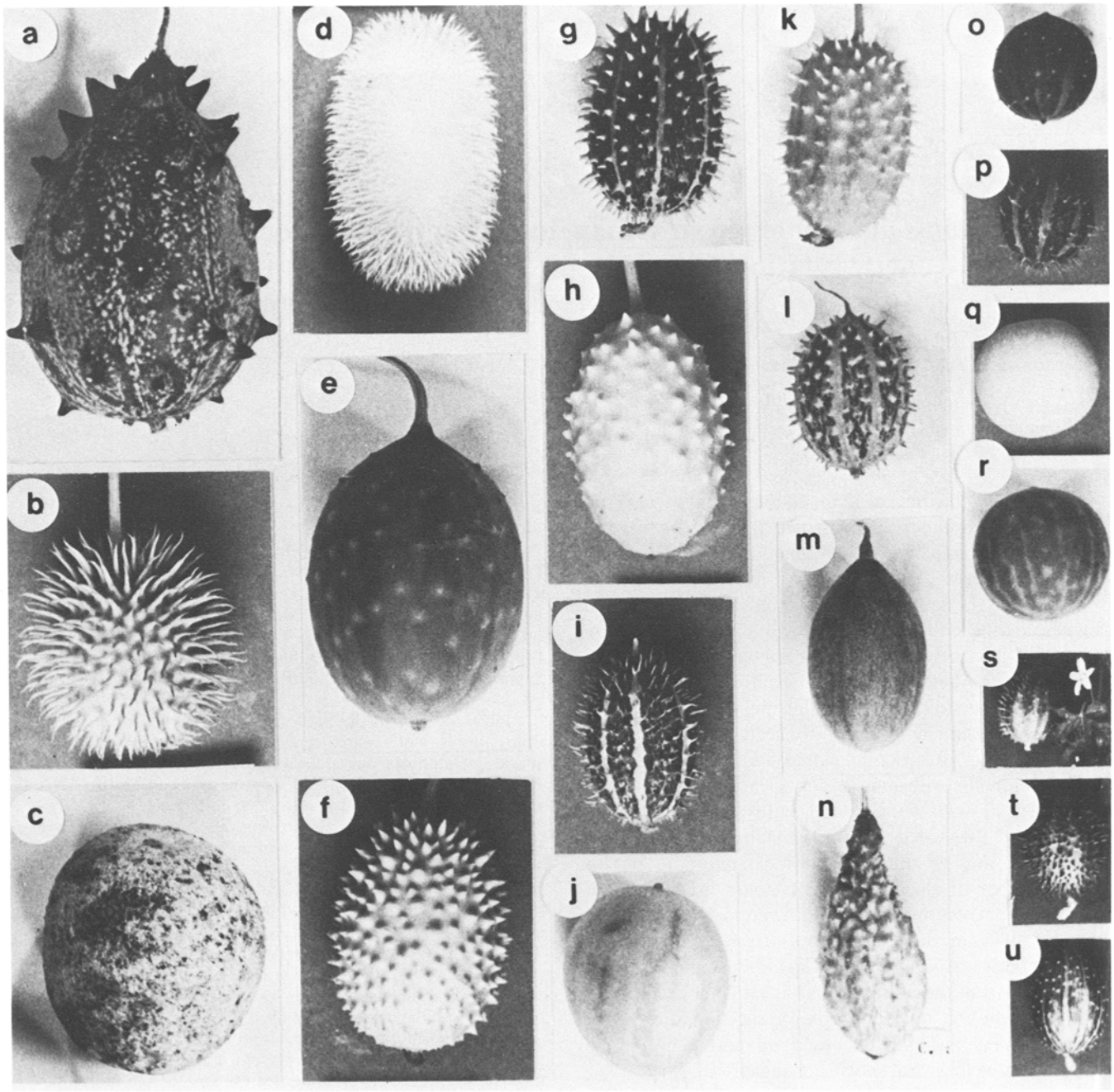

Fig. 1. Fruits of wild *Cucumis* species used in this study. **a** *C. metuliferus* (code number 3); **b** *C. longipes* (33); **c** *C. humifructus* (32); **d** *C. dipsaceus* (27); **e** *C. figarei* (9); **f** *C. meeusei* (7); **g** *C. hookeri* (43); **h** *C. anguria* (4); **i** *C. prophetarum* (11); **j** *C. ficifolius* (6); **k** *C. pustulatus* (24); **l** *C. africanus* (14); **m** *C. melo* var 'agrestis'(8); **n** *C. anguria* (5); **o** *C. leptodermis* (41); **p** *C. myriocarpus* (10); **q** *C. sagittatus* (35); **r** *C. dinteri* (28); **s** *C. zeyheri* (40); **t** *C. zeyheri* (12); **u** *C. heptadactylus* (34). Fruit size in the photograph is about 1 : 1 in respect to natural size, except for the fruits in s, t, u which were reduced in the photograph to about half-natural size (1 : 2 size reduction)

Materials and methods

Plant material

Table 1 lists the cultivated varieties and wild *Cucumis* species used in this study, and Fig. 1 illustrates the fruits of the wild species. Plants for propagation and for chlDNA extraction were grown in the greenhouse throughout the year (24 ± 4 °C). To obtain seed setting and to avoid cross fertilization, hand self-pollinations were performed as required.

Chloroplast DNA extraction

ChlDNA was extracted from fresh, fully-expanded leaves from plants of various ages. In the case of cultivated melon,

Table 1. *Cucumis* species and varieties in the study

A Wild species Species	Assigned code	Source
C. leptodermis	41	1[a]
C. sagittatus	35	1
C. dipsaceus (Ecuador)	27	2
C. melo var 'agrestis'	8	3
C. prophetarum	11	3
C. ficifolius	6	3
C. metuliferus	3	3
C. heptadactylus	34	1
C. humifructus	32	1
C. longipes	33	1
C. pustulatus	24	2
C. dinteri	28	2
C. meeusei	7	3
C. africanus	14	3
C. figarei	9	3
C. anguria	4	3
C. myriocarpus	10	3
C. zeyheri	40	1
C. hookeri	43	1
C. hardwickii	CH	4
B Cultivars Name	Assigned code	Source
C. melo cv 'Yokneam'	MY	4
C. melo cv 'Tzahov-Yetzu' (Cassaba)	MC	4
C. melo cv 'Hales-Best' (Cantaloupe)	MH	4
C. melo var 'Flexuosus'	MF	4
C. melo cv 'Monoecious Rondo'	MR	4
C. sativus cv 'Shimshon'	CS	4

[a] Sources of seeds were as follows:
1 = A. E. Swanepoel, Dept. of Agriculture South Africa.
2 = T. W. Whitaker, Dept. of Agriculture United States.
3 = A. P. M. Den Nijs, Wageningen, The Netherlands.
4 = Weizmann Institute of Science, Plant Genetics Dept., Israel

cotyledons (two-week-old seedlings) were used. Plants were darkened for 2–4 days prior to the extraction.

All operations, until lysis, were performed in the cold (4 °C). Chloroplast isolation and chlDNA purification was modified from Fluhr and Edelman (1981a) and Saltz and Beckman (1981). A sample of 100 g of leaves was chopped (3×5 s bursts) in a Waring Blender in 500 ml of buffer A (see below). The material was then filtered through four layers of gauze and two layers of Miracloth. The filtrate was centrifuged for 2 min at 250×g (1,000 RPM, Sorval HG4L); the supernatant was centrifuged for 5 min at 1,000×g (2,500 RPM). The pellet was suspended in 360 ml buffer B (see below), and centrifuged as above. The pellet was resuspended in 12 ml buffer B and loaded on six SW-28 tubes of discontinuous sucrose gradients (20/45/60% sucrose in buffer B), and centrifuged for 1 h at 20,000 RPM (Beckman SW-28).

Two chloroplast bands usually appeared. Only the upper one (yielding cleaner DNA) was collected with a large-mouthed Pasteur-pipette into 30 ml Corex tubes. Fifty μg/ml Proteinase-K, as well as NaCl and SDS were added up to 0.2 M and 0.5%, respectively and lysis went on for 40 min at 37 °C. The lysate was extracted twice with an equal volume of distilled phenol (freshly neutralized with an equal volume of 0.1 M Tris-base), then extracted once with chloroform and isoamyl-alcohol (24:1 v/v). The DNA was ethanol precipitated from the acqueous phase. Yields of successful extractions were about 3 μg DNA/g fresh weight.

Buffer A: Sorbitol 0.35 M, MES 10 mM, EDTA 2 mM, $MgCl_2$ 1 mM, $MnCl_2$ 1 mM, K_2HPO_4 0.5 mM, NaCl 50 mM, PVP (360) 2 mg/ml; pH=6.1. Buffer B: Sorbitol 0.35 M, HEPES 25 mM, EDTA 8 mM, $MgCl_2$ 1 mM, $MnCl_2$ 1 mM, K_2HPO_4 0.5 mM, NaCl 50 mM, PVP (360) 2 mg/ml; pH=7.6. Before use, 7 mM mercapto-ethanol was added to buffer A, BSA 0.1% to buffers A and B, 1 mM spermine and 1 mM spermidine (hydrochloric salts) to buffer B.

Digestion, fractionation and blotting of DNA

DNA was digested with restriction endonucleases (Biolabs), according to Maniatis et al. (1982). Agarose gel electrophoresis was according to Fluhr and Edelman (1981a). Commonly 0.8% agarose gels were used; 0.5% gels were run for better separation of large (>20 Kb) fragments at 4 °C. Gels with higher agarose concentration (1.2–1.5%) were used to better separate fragments of 0.5–3 Kb. Different digests of lambda-phage DNA were used as molecular weight markers.

Southern transfer and hybridization were performed basically according to Wahl et al. (1979), and Fluhr and Edelman (1981a). Nick-translation was according to Maniatis et al. (1982).

Elution of DNA fragments from gels

The procedure used was based on the method of Otto and Snejdarkova (1981) for protein elution. Preparative-tube gels, 0.6% agarose, were run, stained and specific bands were cut under U.V. light. Excised pieces were loaded on agarose tube gels and overlayed with molten 1% agarose in 0.5×TAE buffer (Fluhr and Edelman 1981a) and 0.1% bromophenol blue. This was overlayed with 300 μl of 50% glycerol in 0.5×TAE buffer; then a layer of 2 M NaCl was added to the top of the tube. The DNA was run upwards for 10–15 min (100 Volts), and collected with the blue stain from the interface. The DNA was then ethanol-precipitated, dried and redissolved. Good recovery was achieved also for large fragments.

Results

Construction of C. melo chloroplast DNA restriction map

Restriction fragments of melon chlDNA were run on agarose gels. Table 2 lists the sizes of these fragments (means from several gels). Figure 2 shows the restriction patterns with the four endonucleases used for the mapping. Some fragments clearly appeared to be in bimolar dose, but often the dose could be determined only during the mapping procedure.

The first approach to align the above fragments into a map was heterologous hybridization. Plasmids containing tobacco PstI fragments were used as radioactive probes. Fourteen such plasmids covering the total tobacco plastome (prepared by Dr. R. Fluhr) were hybridized to blots of melon chlDNA digested with PvuII, SmaI and XhoI. They were designated as PS-, according to the PstI fragment carried. The homology between tobacco and melon chlDNA was high enough to give strong and specific hybridization. Moreover, the overall organization of the two molecules was found to be similar: both contain the inverted repeat (IR)

Table 2. Chloroplast DNA fragments of *C. melo*

Pvu II		Sma I		Xho I		Sac I	
Fragment	kb dose	Fragment	kb dose	Fragment	kb dose	Fragment	kb dose
P1	42.0	S1	28.7	X1	25.3	Sac1	30.8
P2	22.5	S2	21.7	X2	13.1	Sac2	23.2
P3	16.8	S3	17.9	X3	12.0×4	Sac3	19.3
P4	14.0	S4	15.2	X4	8.3	Sac4	16.5
P5	12.1	S5	13.0×2	X5	7.7	Sac5	14.4
P6	10.3	S6	6.5	X6	6.7	Sac6	9.4
P7	9.1	S7	4.3	X7	6.3	Sac7	6.7
P8	8.1	S8	4.1×2	X8	4.8	Sac8	5.2
P9	6.5×2	S9	3.7×2	X9	3.6	Sac9	4.6
P10	2.5	S10	3.2	X10	3.3	Sac10	3.4×2
		S11	2.5×2	X11	3.1×4	Sac11	3.1
		S12	1.9×2	X12	2.4	Sac12	2.0×2
		S13	1.3	X13	2.1×3	Sac13	1.9
				X14	1.0	Sac14	1.8
						Sac15	1.0
Total	150.4		149.3		149.2		148.8

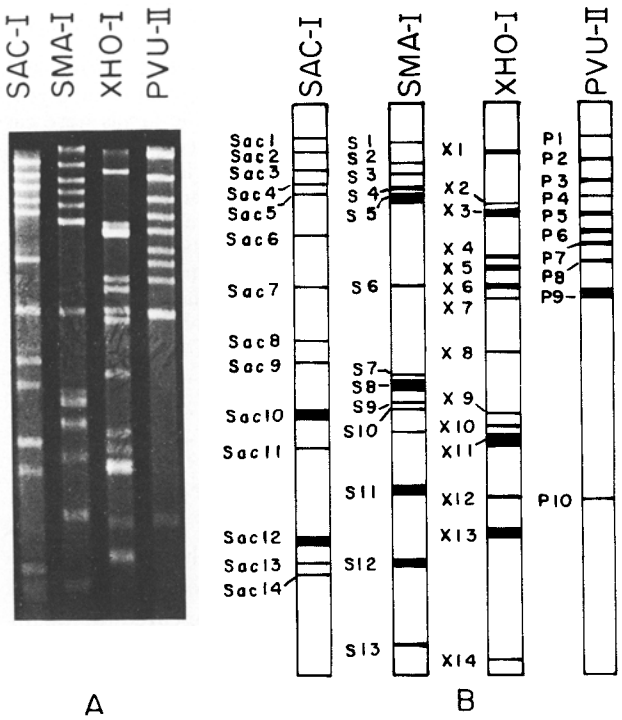

Fig. 2. *C. melo* chloroplast DNA restriction patterns. DNA was digested with four endonucleases and run on a 0.8% agarose gel. In the scheme, fragment designations are marked next to the bands. Fragment sizes are presented in Fig. 5

element, and the melon fragments could be arranged in a co-linear fashion with the tobacco fragments; no gross rearrangements could be detected. The two molecules differ in size, (about 160 kb and 150 kb in tobacco and melon, respectively) but no localized deletion was found. Figure 3 summarizes all the hybridizations between the tobacco probes and the melon PvuII and SmaI fragments. This provided us with an unambiguous PvuII and SmaI fragment alignment, taking into account virtually all data. The pattern of XhoI hybridization was more difficult to interpret, because of the apparent multiple dose of some fragments, hybridizing to different regions of the tobacco map.

As a second approach, isolated melon PvuII fragments (P1 to P10 and also XhoI fragments X1 and X3) eluted from preparative gels were used as radioactive probes. These were hybridized to melon chlDNA cut with XhoI, SmaI, SacI and PvuII; the latter were used as controls for probe identity. Figure 4 is an example of such an hybridization. The results confirmed the above PvuII and SmaI maps, and helped in mapping XhoI and SacI sites.

Heterologous probes available in our lab were used to locate four melon plastome genes. The clones (prepared by H. Fromm) contained internal sequences from the 32-kd protein gene (*psb* A), the large subunit of the RuBP-carboxylase gene (*rbc* L), and the genes for the beta and alpha subunits (*atpA atpB* of the ATP-synthase. In Fig. 5 the respective gene positions appear as black boxes, delimited by hybridization data and by probe size.

A double-digestion approach was taken as well in order to refine the map, especially for some problematic parts. Also a better evaluation of the largest fragments' size was required. We therefore hybridized the PvuII (and XhoI) eluted fragments mentioned above also to double digests of PvuII + XhoI, PvuII + SmaI, and XhoI + SmaI. In such hybridizations the probe is expected to "turn on" fragments which together are of the same sizes as the probe (except for IR sequences).

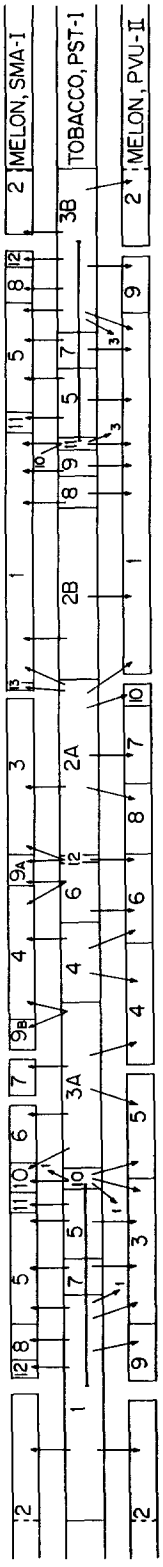

Fig. 3. Hybridization of cloned chloroplast DNA probes from tobacco to melon chloroplast DNA digested with SmaI and PvuII. The *arrows* show which probe hybridized to which fragment. Co-linearity of both plastomes and the presence of IR (inverse-repeat) can be observed

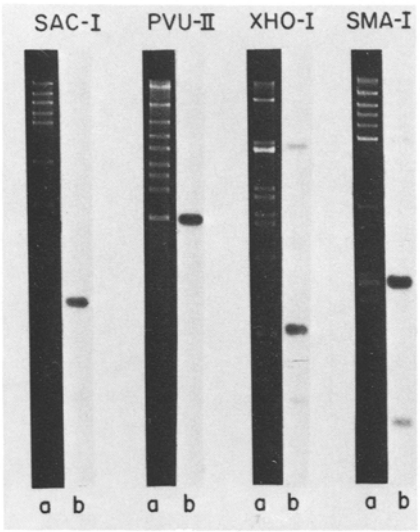

Fig. 4. Hybridization of melon chloroplast DNA probe P9 to melon chlDNA digests with SacI, PvuII, XhoI and SmaI. The probe DNA was isolated from preparative gels by electroelution. Identity of hybridizing fragments is presented in Fig. 5. *a* Ethidium bromide stained gel; *b* Same gel after transfer to nitrocellulose and hybridization

In some cases a second "double-digestion" procedure was applied: isolated fragments were cut by a second restriction endonuclease and run on a gel; hybridization to a total chlDNA probe helped to visualize the weaker bands. An example of such an experiment is shown in Fig. 6. The redigestion of the X3 band (containing a multiple-dose fragment of about 12 kb) with PvuII gave fragments of the following sizes: 9.5, 7.6, 6.6, 4.5, 4 and 1.8 kb. None of these products is likely to be a partial cut (a sum of two adjacent fragments). Therefore, the fact that they sum up to 34 kb (close to 36 kb) proved that we were dealing with three different X3 fragments, and the dose is 4 because X3A is in the IR.

The restriction map shown in Fig. 5 is a compilation of all the above mentioned data.

Restriction pattern analysis of Cucumis species and varieties

ChlDNA from different *Cucumis* species was cut with several 6-base recognition-site restriction endonucleases and run on gels. Nine enzymes were used. Eight of these, PvuII, SmaI, SacI, XhoI, SalI, PvuI, PstI and BglI, yielded relatively few bands, while the EcoRI restriction pattern was too complex to be properly interpreted. The gels revealed a considerable pattern variability between species, but also some overall similarity, rendering the phylogenetic analysis amenable.

Table 3 and Fig. 7 examplify restriction patterns with one enzyme; similar tables were prepared for the

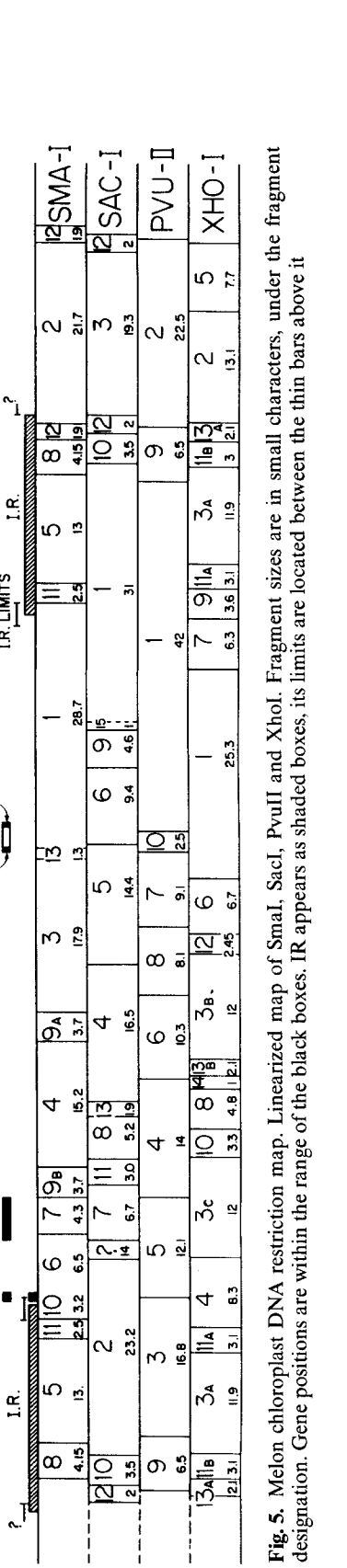

Fig. 5. Melon chloroplast DNA restriction map. Linearized map of SmaI, SacI, PvuII and XhoI. Fragment sizes are in small characters, under the fragment designation. Gene positions are within the range of the black boxes. IR appears as shaded boxes, its limits are located between the thin bars above it

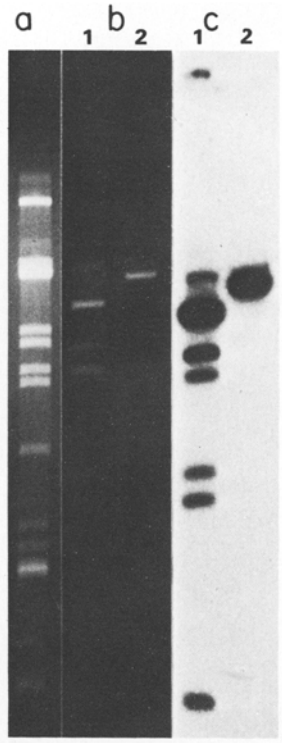

Fig. 6. Second digestion with PvuII of the isolated XhoI fragment X3. The multiple fragment X3 was electroeluted, cut with PvuII and run on a gel. *a* Total XhoI digest run as a control; *b* Ethidium bromide stain: *1* X3 digested with PvuII, *2* uncut isolated X3 DNA; *c* same gel as in b, transferred to nitrocellulose and hybridized to a total chloroplast DNA tobacco probe. Identity of fragments is given in the text. Note the contamination with X2 (faint band above X3) which does not contain PvuII sites and remains uncut

other enzyme patterns (not shown). Part of the differences between species consist of missing or extra bands. Many other changes can beviewed as small size-differences between two presumably homologous bands.

A phylogeny may be based on measuring the distances between taxa; by counting or weighing the differences between them we may derive such a measure of divergence. In our case this could be based either on counting fragment changes between species (proportions of equal or different restriction fragments), or by comparing restriction sites. The latter method is more accurate and was therefore used; it implies an interpretation of the fragment changes observed in terms of mutations. A single base substitution between two species can create a new restriction site, or cause the disappearance of an existing one. As a result, one of the two species compared will have a large band (defined as "state 0, no site"), while the other will have two smaller bands instead, the sum of their sizes should equal the first band ("state 1"). We do not know the direction of the mutation, whether it was $0 \rightarrow 1$ or $1 \rightarrow 0$.

Table 3. XhoI restriction patterns of the different *Cucumis* species. Restriction fragments are denoted by their size in kb. Species designation is according to Table 1. Fragment dose is given as 1,2 etc; (–) is absence of fragment. 1″ (or 2″, etc) means a slightly larger fragment; 1′ means a slightly smaller one. The total size in kb was calculated for species having different patterns

Fragment	Species																					
kb	MY	8	CS	CH	35	28	32	3	11	4	33	27	7	34	41	40	10	43	14	9	24	6
31.0	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
25.3	1	1	1	1	1	1	–	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16.0	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	1	–	1	–
14.5	–	–	–	–	–	–	–	–	1	1	1	1	–	1	1	1	1	1	–	1	–	1
13.1	1	1	1	1	1	1	1′	1′	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12.2	–	–	3	3	–	–	3	3	–	–	–	–	–	–	–	–	–	–	–	–	–	–
11.9	4	4	1	1	4	4	–	–	2	3	3	3	3	3	3	3	3	3	3	3	3	3
11.5	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–
11.1	–	–	–	–	–	–	1	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
10.5	–	–	1	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
9.0	–	–	–	–	1	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
8.3	1	1	–	–	–	–	1″	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
7.6	1	1	1	1	–	–	1″	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7.3	–	–	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6.7	1	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
6.3	1	1	1	1	1	1	–	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5.9	–	–	1	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
5.3	–	–	–	–	–	–	–	–	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4.8	1	1	1	1	–	–	1	?	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4.1	–	–	–	–	2	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
3.6	1	1	–	–	1	1	–	?	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3.4	–	–	–	–	–	–	1	?	–	–	–	–	–	–	–	–	–	–	–	–	–	–
3.3	1	1	1	1	2	2	1	?	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3.1	4	4	2″	2″	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
2.5	1	1	1	1	1	1	1	?	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2.1	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Total	147.8		147.6		148.3		147.8			149.5		151.0										

When all other bands are similar, it is easy to infer the homology between the three bands involved in the change, and characterize the mutation. Although a site change could also result from a deletion/addition event, we referred to it as a "point mutation"; most deletion/additions do not involve sites and were treated separately.

When the sum of the two fragments involved is not exactly equal to the big one, the reason could be an error (10%) in size estimation. Alternatively, on top of the point mutation, an addition/deletion such as those listed in Table 4 may have occurred. Often the two patterns compared differed in many bands, and the interpretation became difficult, as indicated by the notes to Table 4.

Interpretations were achieved in different ways. For BglI the pattern was simple and the few differences easy to relate by direct inspection. For SalI and PvuII, the physical map of cucumber (Palmer 1982) gave sufficient clues to locate site mutations between cucumber and melon and between these and the rest of the species. For PstI, PvuI, SacI, SmaI and XhoI, gels with several species were hybridized to a few tobacco probes covering different regions of the plastome. The results helped to establish the homology-relationships between fragments and reach a consistent interpretation of most fragment differences. Also, the use of gels of low and high agarose content helped to focus pattern differences where big and small fragments were, respectively, involved. At last, all the observed differences could be classified; those interpreted as point mutations (polymorphic sites) are summarized in Table 4. There are 39 such mutations, and all species can be assigned a 0 or 1 state. All the other differences are grouped in Table 5, and can have different interpretations. Part of the latter are probably point mutations which could not be characterized properly because they involve small fragments (e.g. no. 73, 42, 53), or – for species 3 and 32 – because of the low quality of their chlDNA. However, most of the mutations in Table 5 are likely to be small deletions/additions. When the change is small, it could not be small deletions/additions. When the change is small, it could not be recognized when superimposed on a more prominent point mutations; thus many mutations in the list could not be fully characterized in respect to distribution among species. Therefore some

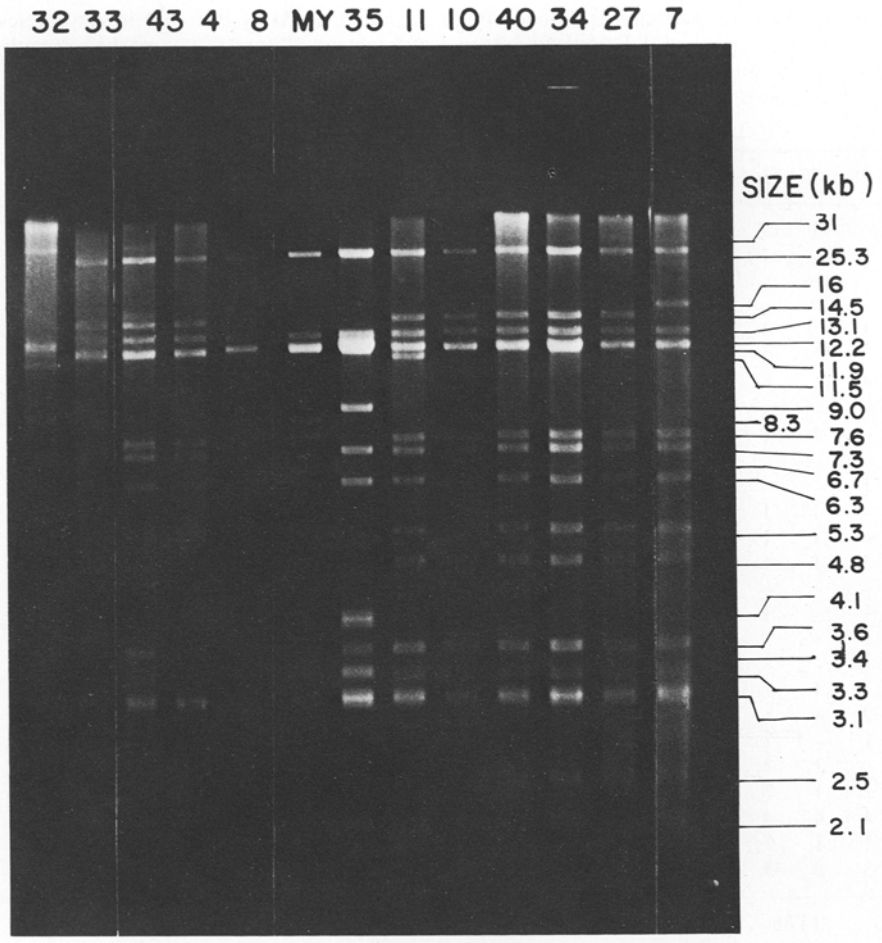

Fig. 7. XhoI restriction patterns of different *Cucumis* species. Species are denoted according to Table 1; fragments are designated by their size in kb

species were listed as having an unknown ("?") state for specific mutations. Grouped under the same number are mutations having the same "pattern" with different enzymes, e.g. – 49a and 49b, where a presumptive deletion is revealed both by XhoI and SacI. By counting them as one we avoided counting the same deletion several times and did not risk over-estimating the number of mutations – although it cannot be excluded that 49a and 49b are two independent mutations. Only mutation 40 in the IR (involving always double-bands) which was detected consistently with five different enzymes, can be regarded with confidence as a deletion/addition.

The method used to derive a dendrogram from the above data was parsimony analysis, discussed by Felsenstein (1983). This method assumes that evolution has taken the shortest pathway to account for the observed differences between taxonomic units – hence, closer species resemble because they branched more recently and not because of parallel evolution. The "most parsimonious tree" is the one involving the least of such events (parallel or back-mutations).

Figure 8 presents the most parsimonious tree derived from our data. It was constructred according to the characterized mutations and includes only two back-or parallel mutations, numbers 20 and 12. A slightly less-parsimonius tree would group MY and 35 closer than MY and CS, but in that case both 16 and 39, instead of 12, will apear as parallel mutations. The "less characterized mutations" were added only afterwards; they turned out to yield less phylogenetic information anyway.

C. hardwickii (CH) and *C. sativus* (CS) had identical patterns, as did *C. sagittatus* (35) and *C. dinteri* (28). Five varieties and cultivars covering the range of variability within cultivated *C. melo* (Table 1) were also checked with eight enzymes and found to be identical. The wild variety *C. melo* var 'agrestis' (no. 8) differs from the cultivated ones by virtue of a small deletion in the IR (Table 5, 40).

A considerable number of mutations separate the species designated as "group A" or "Anguria group" from all others. Within this group, species are separated by a maximum of three characterized mutations, some species being identical.

The other taxonomic units are rather isolated from each other, being on the distal ends of long branches. Very few mutations can yield information on the rela-

Table 4. Characterized polymorphic sites ("point mutations") in the different *Cucumis* species. Only differences characterized as site mutations are described, To every species a state is attributed, whether 0 (not having a particular restriction site) or 1 (having it). Fragments are identified by their size, in kb. Species designation is by code or number (Table 1): MY represents also 8; 4 represents also 33; 35 represents 28, CS is like CH, and A includes: 4, 33, 11, 27, 7, 9, 6, 41, 40, 43, 34, 10, 24 and 14 – unless any of them is specified

No. Enzyme	Description		Distribution	
	State 0: no site frag.size, kb =	State 1: site frag.size + frag.size	State 0	State 1
1. BglI	26 kb	21 kb + 5.9 or 19 + 5.9[a]	group A	all others
2. PvuI	22.5	19 + 2.2	all others	34
3. PvuI	37	11.2 + 26.5 or 10.8 + 26.5	A, 32, 3	CS, MY, 35
4. PvuI	26.5	8.8 + 19 or 8.4 + 19 or 7.7 + 19	MY	all others
5. PvuI	4.1, or 4.0	2.5 + 1.7	all others	MY
6. PvuI	4.1 or 4.0 or 5/1[b]	3.3 + 0.7 or 3.2 + 0.7	all others	A
7. PstI	27 or 8/0	14 + 13	all others	35
8. PstI	> 50	32 + 26.5	CS	all others
9. PstI	6.2 + 10.3 or 6.2 + 9.6	8.9 + 7.5	all others	CS
10. PstI	8.9 + 7.5[c]	10.3 + 6.2 or 9.6 + 6.2, or 11/0	CS	all others
11. PstI	7.5	6.2 + 1.3 or 6.2 + 1.7 or 9/1	27	all others
12. SalI	19.2 or 15/0	11.2 + 6.5 or 14/1	A, CS, 3, 32	35, MY
13. SalI	10.2 + 20.5	11.2 + 16.8 or 12/0 or 14II/0	35	all others
14. SalI	11.2 + 16.8 or 12/0 or 14II/0	10.2 + 20.5	all others	35
15. SalI	I: 15[d] II: 21.5	13.2 + 2.1 19.2 + 2.1 or 12/1, or 14/1	A	all others
16. SalI	23	12.2 + 12.2	35, 3, 32, A	CS, MY
17. SalI	32	20.5 + 12.2 or 18/1	11, 4, 33	all others
18. SalI	20.5 or 17/0	10.8 + 9.5	all others	40
19. SalI	23 or 16/1	20 + 3.0	all others	10
20. PvuII	8.1	1.8 + 6.2	3, all others	41, 34, 10, 32
21. PvuII	45, or 22/1 or 23II/1	31 + 16.8	all others	A
22. PvuII	45, or 22/1 or 23II/1	42 + 2.5	all others	MY
23. PvuII	I: 16.8 II: 45, or 21/1 or 22/1	9.5 + 7.3 37 + 7.3	all others	35
24. PvuII	29	16.8 + 12.1 or 16.8 + 11.5 or 23I/1	CS	all others
25. SacI	7.7 or 8.2	5.2 + 3.0	all others	MY
26. SacI	25, or 23.5	19.2 + 6.0	all others	A
27. SacI	31.5	19.4 + 14.4, or 28/1	MY	all others
28. SacI	19.4 or 27/0	16.4 + 3.2	all others	35
29. XhoI	7.6	4.1 + 3.3	all others	35
30. XhoI	31.6	25.3 + 6.3	32	3, all others
31. XhoI	14.5 or 16	12.2 + 2.1 or 11.9 + 2.1	A	all others
32. XhoI	8.3 or 9.0 or 33II/0	5.3 + 3.6	all others	A
33. XhoI	I: 5.9 II: 10.5	3.6 + 3.1 or 3.4 + 3.1 8.3 + 3.1 or 9.0 + 3.1	CS	all others
34. SmaI	13.1 or 13.7 or 35/1	10.8 + 2.4	all others	A
35. SmaI	13.1 or 13.7 or 34/1	11.0 + 1.7	all others	CS
36. SmaI	23.5	21.5 + 2 × 1.9	CS	all others
37. SmaI	28.5	15.3 + 15.3	3, all others	32
38. SmaI	21.5	8.5 + 13.7	all others	35
39. SmaI	21.5	18.3 + 3.7	all others	CS, MY

[a] More than one possibility in either state 0 or 1 means that more than one difference in the fragment pattern was observed. Here, part of the species having state 1 for mutation 1, have a 21 kb and the others – a 19 kb fragment; this difference is described as mutation 41a in Table 5

[b] When the fragment pattern has 2 point mutations superimposed, the alternative fragments observed are indicated by the number and state (e.g. 5/1) in Table 4. MY has state 0 for mutation 6, but instead of the 4.1 or 4.0 kb ones, fragments of 2.5 + 1.7 kb are observed

[c] Mutations 9 and 10 were interpreted as 2 site mutations specific to CS; no species is actually showing state 0 because the 2 mutations are superimposed

[d] A mutation in a site within the IR, causing the disappearance of 2 single fragments and appearance of 1 double and 2 single ones

Table 5. List of mutations not characterized as point mutations. All differences that could not be interpreted unambiguously as polymorphic sites are listed. They could be point mutations, or deletions and additions. Because they involve small differences, not every species could have a state assigned, and some appear under "?" (unknown state). When the same "pattern of change" between species appears more than once with different enzymes, the mutations were given the same number, to indicate a probable deletion/addition. Indication of fragments and taxonomic units is as in Table 4

No. Enzyme	Description				Distribution between species			
	State 1	2	3	4	1	2	3	4
40a BglI	6.8	6.4	6.3	6.1	32, CS	8, 35	MY	A
40b PvuI	7.7	7.3	7.0	–	32, CS	8, 35	MY, A	–
40c PstI	4.8	4.5	4.4	4.3	32, 3, CS	8, 35	MY	A
40d XhoI	12.2	11.9	–	–	32, 3, CS	rest	–	–
40e SmaI	13.7	13.1	–	–	32, 3	My, 35	–	–
	State 1	2	3		1	2	3	?
41a BglI	21	19	–		MY, CS, 35	32, 3	–	A
41b PstI	10.3+9.6 or 2×10.3	2×9.6	–		rest	32, 3	–	–
41c XhoI	11.9	11.1	–		rest	32, 3	–	–
42 PstI	38	15.5	–		rest	32, 3	–	–
43 PstI	25	–	–		32	3, rest	–	–
44a PvuI	3.9	3.7	–		32	rest	–	3
44b PstI	3.4	3.3	–		32	rest	–	3
44c PvuII	17	16.8	–		32	A, MY	–	3, CS, 35
45 PvuI	3.0	2.9	–		rest	32	–	3
46 SmaI	4.4	4.3	4.2		32	MY, 35	CS, A	3
47 XhoI	9.0	8.5	8.3		35	32	MY	3, rest
48 PstI	2×10.3	10.3+9.6	–		A	MY, 35 11, 3, 32	–	CS
49a SacI	6.7	6.3	–		rest	11	–	–
49b XhoI	11.9	11.5	–		rest	11	–	3, 32
50a SacI	4.5	4.4	–		rest	4, 11	–	–
50b PvuI	3.3	3.2	–		A	4, 11	–	rest
51 EcoRI	2.4	–	–		11, 4	A	–	rest
52 EcoRI	5.9	4.4	–		40	A	–	rest
53 EcoRI	3.1	–	–		A	34	–	rest
54 SacI	8.2	7.7	–		10, 41, 14, 43	rest	–	MY
55 XhoI	16	14.5	–		7, 14, 24	A	–	rest
56a PvuI	8.8	8.4	7.7		A	32, 35	CS	3, MY
56b SmaI	3.5 or 3.2	4.0	–		rest	A	–	3
57 PvuII	12.1	11.6	11.5		rest	3, 32	11	CS
58 SmaI	6.7	6.5	6.3		A, 35	MY, CS	32, 11	3
59 PvuI	13.2	12.5, 5.4	–		rest	CS	–	–
60 SacI	9.4	9.2	7.7		rest	35	CS	–
61 SmaI	3.7	3.6	–		MY	CS	–	rest
62 SmaI	2.8	2.5	–		CS	rest	–	–
63a PvuI	11.2	10.8	–		MY, CS	35	–	rest
63b XhoI	4.8	4.1	–		rest	35	–	–
64 PvuI	4.1	4.0	–		32	35	–	rest
65 PstI	33	32	–		35	rest	–	3, 32
66 SmaI	3.5	3.2	–		35	MY, CS, 32	–	3, A
67 SacI	25	23.5	–		CS, 35, 32, 3	MY	–	A
68a XhoI	7.3	6.7	–		rest	MY	–	–
68b PvuII	9.5	9.1	–		rest	MY	–	–
69 SalI	2.2	2.1	–		CS, 35, 32	MY	–	3, rest
70 SmaI	3.8	3.7	–		32	rest	–	3
71 EcoRI	23	16.5, 5.1	–		43, 14	A	–	rest
72 PstI	1.7	1.3	–		MY	rest	–	–
73 PstI	2.6	–	–		rest	35	–	CS

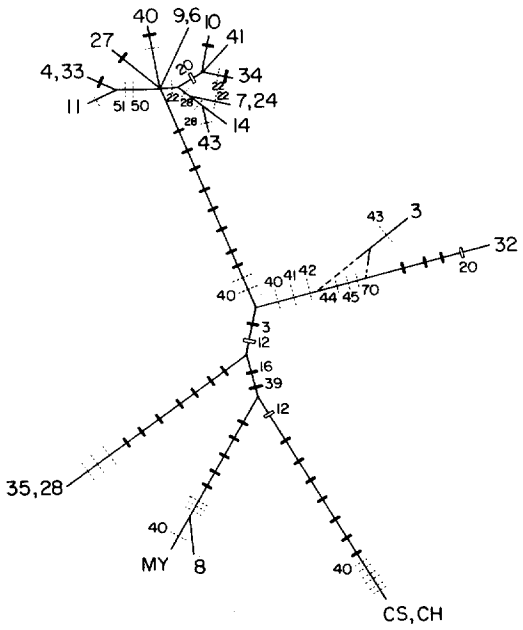

Fig. 8. Most parsimonious rootless tree of *Cucumis* chloroplast DNA mutations. Point mutations (Table 4) and "less characterized" ones (Table 5) are represented as solid and dashed bars, respectively. Part of the mutations are indicated by their (small) number as examples. Species located at the distal ends of the branches, are indicated by their code or (large) number. Point mutations appearing more than once (as back or parallel mutations) are shown as white bars; some of the less characterized mutations appear more than once as well. Two possibilities are given for the branching point of species 3

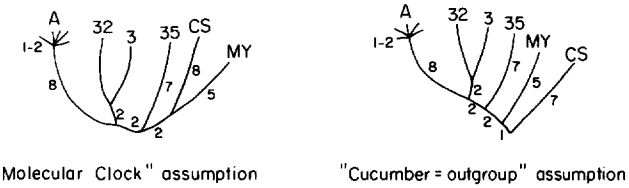

Fig. 9. Two alternatives for rooting the chloroplast DNA phylogenetic tree. On the left is an example of a tree where the root was placed so as to create branches of approximately equal length (length of the different internodes is written as the number of point mutations from Fig. 8). On the right, cucumber is considered an outgroup due to previous taxonomic information and the root is on the branch between it and the rest of the tree

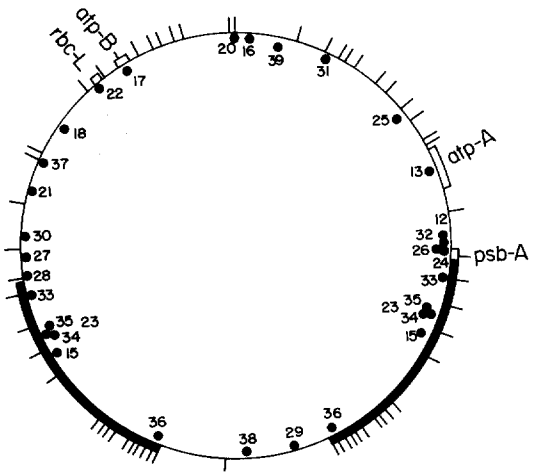

Fig. 10. Distribution of polymorphic and conserved sites along the plastome of *Cucumis*. Conserved restriction sites, existing in all the species checked, are indicated by bars; polymorphic sites, existing only in part of the species, are marked by circles, and the mutation number from Table 4 is indicated. IRs are indicated by solid bars and four genes are positioned according to their localization in *C. melo*.

tionship between CS, MY, 35, 3 and 32 by defining inner internodes of the dendrogram. One mutation groups together A, 32 and 3 versus CS, MY and 35; one joins MY and 35, two- CS and MY and three mutations (no 40, 42 and 41) join species 3 and 32, thus showing some affinity (or a common ancestor); but mainly this points towards a long history of divergence.

Providing a root to the above network (or "rootless tree") is a matter of speculating: a somewhat arbitrary decision has to be taken as to what was the first branching event between the members of the phylogeny. The attributed ancestral state of the sites (whether 0 or 1) and direction of changes – are the consequence of this decision. The tree may be "rooted" by assuming that the mutation rate is constant ("Molecular Clock" hypothesis, see Wilson et al. 1977); the distances between root and different branch-edges should therefore be approximately equal. Another possibility is to include in the phylogeny a species, known to be relatively distant, as an outgroup; the root is then placed on the branch separating the outgroup from the rest. In our case, taking cucumber as an outgroup is quite reasonable. Figure 9 exemplifies two possible roots for the tree.

Distribution of mutations along the Cucumis plastome

Most restriction sites of PvuII, XhoI, SacI, SmaI and SalI were mapped in the different species. Sites commonly present in all the analysed species are denoted as "conserved sites"; the others differ in at least one species and are called "mutated" or "polymorphic sites". Figure 10 shows map positions of most sites.

The Small Single Copy (SSC) region of the *Cucumis* plastome contains very few sites. Thus its conservation cannot be evaluated accurately. In the IR two distinct regions are seen: the region close to SSC is very conserved (10 conserved, 1 mutated sites). The region close to the Large Single Copy (LSC) contains many mutations (5 mutated, 4 conserved sites). In the LSC a dense concentration of polymorphic sites is found near one of the IR borders, adjacent to the *psb* A gene (4 polymorphic, 1 conserved sites). In the rest of the LSC region, 16 mutated vs 27 conserved sites show a rather even distribution.

Discussion

Plastome organization in Cucumis

The plastomes of *C. melo* and *N. tabacum* were found to be colinear and to share similar gene positions for the four genes checked. Also, within the different *Cucumis* species no gross rearrangements could be detected.

Cucumber was already classified by Palmer and Thompson (1982) as being colinear with spinach and petunia; the latter were shown to be colinear with tobacco (Fluhr and Edelman 1981b). Palmer and Thompson (1982) traced some major phylogenetic events according to a few plastome reorganization events detected by heterologous mapping: only one large inversion was shown to separate maize from the above dicot plastomes; another one separates all the above from the legumes; but, surprisingly, in a number of legumes very extensive reorganization has "scrambled" the plastome, along with the loss of one IR.

Except for the latter case, the plastome seems very conserved in terms of basic organization. This rendered possible the use of alien chlDNA probes to map *Cucumis.* Also the gene-probes from *Spirodela,* tobacco and maize gave specific signals.

The IR is referred by some authors as a very conserved part of the plastome (Rhodes et al. 1983; Palmer and Zamir 1982; Bowman et al. 1983). In our study, this property was found only for part of the IR, shown in other species to code for the extremely conservative rRNA operon (a 10 kb region near the SSC). This is in agreement with Palmer et al. (1983a) and Clegg et al. (1984). The coding function of the rest of the IR is unknown; in *Cucumis* it is a relatively unconserved region. Our findings (Fig. 10) agree with a "hot spot" near the *psb* A gene, (and possibly also within IR). In wheat (Bowman et al. 1983), the variable regions, also in the LSC, are different. It would be interesting to know whether such regions contain structures that could favor rearrangements; some evidence for this is presented by Zurawski et al. (1984). In our case, 20 of the presumed 'deletions' detected (Table 5) can be assigned to mapped fragments: 3–5 are in the IR, 6–10 in the 15 kb covering the main "hot spot" region and 7–9 in the remaining 65 kb of the LSC. Thus there seems to exist a "hot spot" also for deletions/additions in the same region as for point mutations.

Chloroplast DNA phylogeny – theoretical basis and method of derivation

The first step in determining a phylogenetic tree is to select the appropriate data-base. Restriction-pattern data can by analyzed by comparing either the fragments or the restriction sites between the species. The latter is a more direct and accurate way to measure distances between species. This is because a change in a site is likely to represent directly a mutation of one nucleotide, but a fragment difference can be interpreted in more ways. When sites are compared, a fairly good distinction can be made between point mutations and deletions/additions. Rearrangements are not very suitable for mutational analysis, because: (a) they could be counted several times with different enzymes; (b) they are less specific events, more likely to occur in parallel than point-mutations; c) being mostly small changes, they cannot be properly recognized in all the species, and turn out to be less informative. Indeed the characterized site mutations gave, in our case, almost all of the phylogenetic information, with only two cases of possible back-or parallel mutation (Fig. 8).

Previously analyzed genera differed in the nature and amount of the changes. *Triticum* and *Aegilops* (Bowman et al. 1983; Terachi et al. 1984), differed almost exclusively by small deletions and additions, while *Lycopersicon* showed only site mutations (Palmer and Zamir 1982). In our case, both differences were numerous showing that *Cucumis* is a well diverged genus.

A second issue to consider is the model on which the phylogenetic study is based. The parsimonious tree can be derived in different ways; we did it by a stepwise arrangement of individual characters (the site mutations) in the most parsimonious fashion. The other possibility, as performed for the isozyme dendrogram presented in a subsequent publication, is to compute total-distance values between species over all the characters, and use these to build the tree.

If, in the true evolutionary process, many convergent mutations occurred, the parismony assumption will lead to an erroneous tree. When the amount of evolutionary change is too large, either because the rate of change is high, or the taxons are distant, the error contributed by parallel- and back-mutations grows and the model is less suitable. In the latter case the comparison can be made by using only the most conservative portions of the chlDNA. Also, a relatively constant rate of evolution along different branches is important in the model; but if the rates are small enough, some rate differencies can be tolerated (Felsenstein 1983). In our case the rates are small enough and the proportion of parallel mutations seems low.

Parsimonious models also assume that no "flow of information" or "horizontal gene transfer" between already diverged branches took place. Such transfer is, however, possible; introgressive hybridization is an important phenomenon in plant evolution, but cannot be differentiated by a parsimony model. We might recognize such events only due to previous knowledge,

by finding a branch "out of place" (e.g., in Palmer et al. 1983 b).

Rates of molecular evolution in this study

Estimates of p, the number of chlDNA substitutions per nucleotide, were calculated for pairs of species, according to the model of Engels (1981) for restriction sites. The total number of sites assayed is about 90, which constitute a 0.4% sample from the 150 kb plastome. The p values were 0.001–0.005 among pairs from group A; all other combinations involving melon, cucumber, group A, 32 and 35 were between 0.01 and 0.02.

Values of p between millet species (Clegg et al. 1984) were small, 0.003–0.007, and those between *Lycopersicon* species (Palmer and Zamir 1982) were about 0.007. In *Brassica* the estimate was about 0.024 (Palmer et al. 1983 b), which seems typical to a more diverged genus like *Cucumis*. Among genera in the Solanaceae 0.03 divergence was calculated (Fluhr and Edelman 1981 b) but in the legumes the intergeneric values were up to 0.13 (Palmer et al. 1983 a). Maize and barley have about 0.06 and 0.09 sequence divergence in coding and non-coding regions, respectively (Zurawski et al. 1984). This could reflect the fact that some families are more ancient or had a faster rate of evolution. The divergence time between cucumber and melon could be based on the separation of India and Africa, about 90 million years ago (Raven et al. 1981). This leads to an estimate of $0.02/90 \times 10^6 = 0.2$ nucleotide-substitution/nucleotide/10^9 years.

Acknowledgements. We are grateful to Dr. Dvora Aviv for her advice throughout this work, to Shlomit Bleichman for her help in laboratory procedures, to Dr. R. Fluhr and to H. Fromm for chlDNA clones, and to A. P. M. Den Nijs, T. W. Whitaker and A. E. Swanepoel for providing *Cucumis* spp. seed samples.

References

Berthou F, Mathieu C, Vedel F (1983) Chloroplast and mitochondrial DNA variation as indication of phylogenetic relationships in the genus *Coffea* L. Theor Appl Genet 65: 77–84

Bowman CM, Bonnard G, Dyer TA (1983) Chloroplast DNA variation between *Triticum* and *Aegilops*. Location of the variation on the cp genome and its relevance to the inheritance and classification of the cytoplasm. Theor Appl Genet 65:247–262

Clegg MT, Rawson JRY, Thomas K (1984) Chloroplast DNA variation in pearl millet and related species. Genetics 106:449–461

Engels WR (1981) Estimating genetic divergence and genetic variability with restriction endonucleases. Proc Natl Acad Sci USA 78:6329–6333

Erickson LR, Straus NA, Beversdorf WD (1983) Restriction patterns reveal origins of chloroplast genomes in *Brassica* amphiploids. Theor Appl Genet 65:201–206

Felsenstein J (1983) Parsimony in systematics: biological and statistical issues. Annu Rev Ecol Syst 14:313–333

Fluhr R, Edelman M (1981 a) Physical mapping of *Nicotiana tabacum* chloroplast DNA. Mol Gen Genet 181:484–490

Fluhr R, Edelman M (1981 b) Conservation of sequence arrangement among higher plant chloroplast DNA: molecular cross hybridization among the *Solanaceae* and between *Nicotiana* and *Spinacia*. Nucleic Acids Res 9: 6841–6853

Frankel R, Galun E (1977) Pollination mechanisms, reproduction and plant breeding. Springer, Berlin Heidelberg New York, 281 pp

Galun E (1984) Plant organelle transfer by cybridization: experimental results and application in crop improvement. In: Kyoto Symposium on Bioscience, pp 31–39

Hosaka K, Oghiara Y, Matsubayashi M, Tsunewaki K (1984) Phylogenetic relationship between the tuberous *Solanum* species as revealed by restriction endonuclease analysis of chloroplast DNA. Jpn J Genet 59:349–369

Kroon GH, Custers JBM, Kho YO, Den Nijs APM, Varekamp HQ (1979) Interspecific hybridization in *Cucumis* (L.). 1. Need for genetic variation, biosystematic relations and possibilities to overcome crossability barriers. Euphytica 28:723–728

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. Cold Spring Harbor NY, pp 100–112

Palmer JD (1982) Physical and gene mapping of chloroplast DNA from *Agriplex triagularis* and *Cucumis sativa*. Nucleic Acids Res 10:1593–1605

Palmer JD, Thompson WF (1982) Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell 29:537–550

Palmer JD, Zamir D (1982) Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon*. Proc Natl Acad Sci USA 79:5006–5010

Palmer JD, Singh GP, Pillay DTN (1983 a) Structure and sequence evolution of three legume chloroplast DNAs. Mol Gen Genet 190:13–19

Palmer JD, Shields CR, Cohen DB, Orton TJ (1983 b) Chloroplast DNA evolution and the origin of amphiploid *Brassica* species. Theor Appl Genet 65:181–189

Raven PH, Evert RF, Curtis H (1981) The biology of plants, ch 19. Evolution of the flowering plants. Worth Publishers, NY, pp 372–375

Rhodes PR, Zhu YS, Kung SD (1981) *Nicotiana* cp genome. 1. Chloroplast DNA diversity. Mol Gen Genet 182:106–111

Saltz Y, Beckman J (1981) Chloroplast DNA preparation from *Petunia* and *Nicotiana*. Plant Mol Biol Newslett 2:73–74

Terachi T, Ogihara Y, Tsunewaki K (1984) The molecular basis of genetic diversity among cytoplasms of *Triticum* and *Aegilops*. 3. Chloroplast genomes of the M and modified M genome carrying species. Genetics 108: 681–695

Vedel F, Mathieu C (1983) Physical and gene mapping of chloroplast DNA from normal and cytoplasmic male sterile (radish cytoplasm) lines of *Brassica napus*. Curr Genet 7:13–20

Wahl GM, Stern M, Stark GR (1979) Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization using dextran sulphate. Proc Natl Acad Sci USA 79:2264–2267

Wilson AC, Carlson SS, White TJ (1977) Biochemical evolution. Annu Rev Biochem 46:573–639

Zurawski G, Clegg MT, Brown AHD (1984) The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics 106:735–749