Theor Appl Genet (1988) 75:357–361

Insertion and amplification of a DNA sequence in satellite DNA of *Cucumis sativus* L. (cucumber)

M. Ganal* and V. Hemleben**

Lehrstuhl für Genetik, Institut für Biologie II, Universität Tübingen, Auf der Morgenstelle 28, D-7400 Tübingen, Federal Republic of Germany

Received August 12, 1987; Accepted August 19, 1987
Communicated by F. Mechelke

Summary. Another satellite DNA repeat (type IV) in the genome of *Cucumis sativus* (cucumber) was found and investigated with respect to DNA sequence, methylation, and evolution. This satellite shows a repeat length of 360 bp and a GC-content of 47%. The repeats of type IV are highly conserved among each other. Evidence for CG and CNG methylation is presented. By comparison to the previously described satellites (type I/II and type III) from cucumber, it is evident that this repeat is created by an insertion of a 180 bp DNA sequence similar to type I–III into another DNA sequence (or vice versa), and subsequent amplification forming a new satellite repeat. The different satellites of the type I/II, type III, and the 180 bp insert of type IV show a sequence homology of 60%–70%, indicating that the complex satellite DNA of cucumber is originated from a common progenitor by mutation, additional insertion, and amplification events. Copies of a sequence similar to a part of type IV are present in the genome of the related species *Cucumis melo* (melon).

Key words: *Cucumis sativus* – Satellite DNA – Insertion – Amplification – Sequence homology – *Cucumis melo*

Introduction

Highly repeated DNA sequences often occur as tandemly arranged DNA sequences in almost all eukaryotic genomes (Singer 1982; Beridze 1986; Miklos 1986). A common feature of these sequences is that the single repeating units evolve in a specific manner, also called "concerted evolution" (Dover 1982, 1986; Dover and Tautz 1986). The mechanism of unequal crossing over is mostly responsible for this kind of evolution and variation (Smith 1976). This process is also involved in the creation of new variants from a single repeat type, often resulting in either two related satellite repeats or one repeat family replacing another in a population (Dover 1986). In addition, there are further possibilities of generating new variants by processes such as reverse transcription and transposition (Dover 1982; Flavell 1982, 1985).

Plants in the *Cucurbitaceae* family are often characterized by an enormous amount of such repeated DNA sequences (Ingle et al. 1975) therefore offering a system to study the generation and spreading of different satellite repeats. Species from the genera *Cucumis* and *Cucurbita* show several satellite repeats of either 169–182 bp or about 352 bp in length (Shmookler-Reis et al. 1981; Hemleben et al. 1982; Brennicke and Hemleben 1983; Ganal et al. 1986; Ganal and Hemleben 1986a).

An extreme example in this family of dicotyledonous plants is *Cucumis sativus*. The proportion of heterochromatin in this organism is about 50% (Ramachandran et al. 1985). An analysis of the nuclear DNA by gradient centrifugation revealed that approximately 30% of the total nuclear DNA is satellite DNA and an additional 10% is represented by the genes coding for the large ribosomal rRNAs (Timmis and Ingle 1977; Hemleben et al. 1982; Ganal and Hemleben 1986b).

We have recently characterized two tandemly repeated DNA sequences out of this satellite fraction of cucumber (type I/II and III), and shown that they are related in their DNA sequence and not interspersed into

* Present address: Department of Plant Breeding and Biometry, 416 Bradfield Hall, Cornell University, Ithaca, NY 14853, USA
** To whom correspondence should be addressed

each other (Ganal et al. 1986). Here we have analyzed a further tandemly arranged DNA sequence, with interesting features concerning its structure and evolutionary relationships to the other satellite-type sequences in the genome of cucumber.

Material and methods

Plant material

Seeds of *Cucumis sativus* cv Sensation Typ Neckarruhm and of *Cucumis melo* cv Zuckerkugel were purchased commercially (Hild, Marbach am Neckar, FRG), and cultivated for 7 days in the dark before harvesting.

DNA isolation

DNA from seedlings was isolated as described by Hemleben et al. (1982). After a first purification step on a CsCl gradient, satellite DNA was separated from the bulk of nuclear DNA by CsCl-actinomycin D gradients (Hemleben et al. 1977). Standard methods for analyzing DNA were applied according to Maniatis et al. (1982).

Hybridization

The eluted inserts were used for hybridization experiments. Type IV repeats were cut either from plasmids or from the replicative form of M13. Hybridization and labelling of DNA was performed as previously described (Ganal et al. 1986). Stringent hybridization temperature was 68 °C, and relaxed conditions included a hybridization temperature of 60 °C.

Cloning and sequencing

All cloning experiments were carried out using the plasmid pUC 8 (Vieira and Messing 1982), or the phages M13mp8 or M13mp9 (Messing and Vieira 1982), and DNA fragments from enriched satellite DNA which were eluted from preparative polyacrylamide gels using the method of Maxam and Gilbert (1977). Recombinant plasmids and M13 phages were selected by colony and plaque hybridization (Grunstein and Hogness 1975), and DNA sequencing was done according to Sanger et al. (1977). The programs of Larson and Messing (1982) were used for computer analysis.

Results

Two different, but related, satellite DNA repeat types of either 182 bp (type I/II) or 177 bp (type III) have been characterized in *Cucumis sativus* by cloning, DNA sequencing, and hybridization (Ganal et al. 1986). However, after digestion with HaeIII some additional bands of 212, 80 and 70 bp were observed, which could not be explained by the DNA sequences obtained from a number of type I–III repeats. The observation of these additional bands suggests the presence of at least one further element in the genome of *Cucumis sativus*.

Cloning of the type IV satellite DNA of Cucumis sativus

The region containing the 212 bp HaeIII fragment was eluted from a preparative polyacrylamide gel and the DNA directly cloned into M13mp8. By sequencing a number of clones, we found that almost all clones of the expected size contained an identical insert (as indicated in Fig. 1). Although the fragments were cloned after a complete digestion with HaeIII, the clones contained an internal HaeIII site. This indicates that this site is masked in genomic DNA by methylation. Since HaeIII does not cut the sequence GG^{m}CC (Kessler et al. 1985), and the second cytosine residue of the cutting site is followed by a guanosine residue (Fig. 1), it is direct evidence that plant DNA can be methylated in CNG (here CCG) sequences (Gruenbaum et al. 1981). To reveal the genomic organization of this DNA sequence, the 212 bp HaeIII fragment was used to probe genomic DNA of *Cucumis sativus*. The hybridization pattern resulted in a ladder which is typical for a tandemly arranged satellite DNA sequence (see Fig. 2). The digestion with Sau3A revealed a repeat length of 360 bp, which is in contrast to the previously analyzed satellite sequences with a repeat length of either 177 or 182 bp. Therefore, the 212 bp sequence is part of another satellite repeat in cucumber.

To characterize the entire 360 bp repeat (type IV), enriched genomic satellite DNA was then cut with Sau3A and the 360 bp region eluted and cloned. Clones

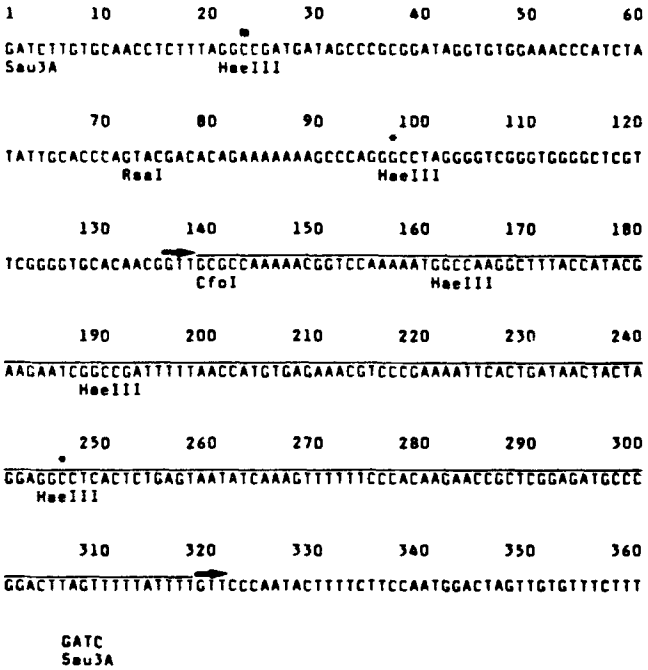

Fig. 1. DNA sequence of a representative type IV repeat of *Cucumis sativus* (cucumber). Region showing the homology to type I/II and type III satellites is indicated by a continuous line. The direct repeats flanking this region are marked by *two arrows*. The beginning (245) and end (97) of the 212 bp HaeIII fragment are marked by *an asterisk*. The methylated cytosine residue of the methylated HaeIII site (23) is indicated by a ^{m}

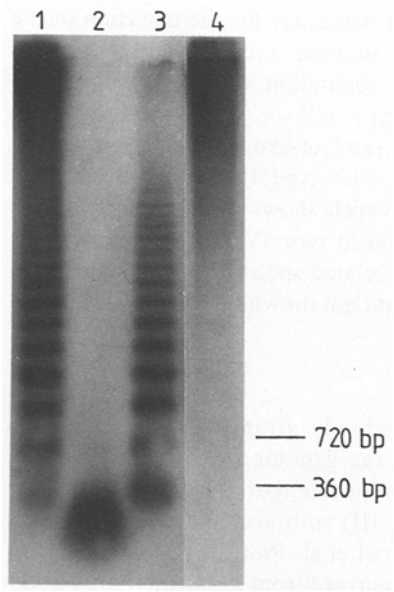

Fig. 2. Hybridization analysis of type IV satellite DNA. Total nuclear DNA was digested with (*1*) RsaI, (*2*) HaeIII, (*3*) Sau3A, and (*4*) CfoI, separated on an 1.5% agarose gel, and after transfer hybridized with a cloned ^{32}P-labelled type IV repeat

Fig. 3. Sequence homologies of type I/II and type III satellites to the type IV repeat. Representative type I and type III repeats were aligned to that part of type IV showing homology to these repeats. From type IV only the homologous region and the flanking direct repeats are shown. The *asterisks* indicate identical bases

Table 1. Different satellite repeats in the genome of *Cucumis sativus* (cucumber). For each repeat type at least three independently cloned repeats were sequenced and analyzed (Ganal et al. 1986; unpublished results)

Satellite DNA repeat type	Repeat length in bp	GC-content %	Maximum variation	% of satellite DNA (approx.)
I/II	182	47	4%	50
III	177	53	6%	35
IV	360	47	3%	15

were identified by hybridization to the 212 bp HaeIII fragment at high stringency, since the type III dimer (354 bp) is migrating at approximately the same site.

Characterization of the type IV repeat by DNA sequencing and hybridization

Some clones containing the entire type IV repeat were sequenced (one representative is shown in Fig. 1); all show a repeat length of exactly 360 bp with a GC-content of 47%. The heterogeneity within individual type IV repeats was very limited (3% sequence divergence), as it was observed for the type I–III repeats (Table 1).

Furthermore, the genomic organization of the type IV repeats was investigated by hybridization to Southern blots of digested cucumber DNA. There was no significant difference between the hybridization pattern with the complete 360 bp repeat (Fig. 2), and the 212 bp HaeIII fragment (not shown). Sau3A is creating a typical ladder characteristic for a tandemly repeated sequence. RsaI (and to a lower extent TaqI) are cutting only part of the type IV repeats, indicating a limited heterogeneity. HaeIII is degrading the repeat into small fragments since it contains several sites as revealed by the DNA sequence. There is no evidence that the type IV repeat is associated with the other repeat types (I–III) to a major extent since due to their differing repeat length there should occur additional bands at 360 ± 180 bp in the hybridization pattern. Finally, CfoI, which is sensitive for C-methylation (Kessler et al. 1985), does not degrade type IV satellite DNA, although there was a single CfoI site in all sequenced repeats (Fig. 1) indicating CG-methylation.

Sequence homologies between the different cucumber satellite DNA repeats

Previous data and the results presented here show that the satellite DNA of cucumber contains at least three different repeat types. In Table 1 these repeat classes are compared with respect to their repeat length, GC-content, and sequence heterogeneity within each single type (Ganal et al. 1986). The data suggest that each type is highly conserved in DNA sequence, since the heterogeneity within each type is between 3% and 6%. The detailed sequence comparison indicates a region of sequence homology in type IV to type I/II of ca. 70%, and to type III of ca. 65% (Fig. 3). The homology, however, is limited to a 180 bp stretch within the 360 bp type IV repeat. The other parts of the type IV repeat

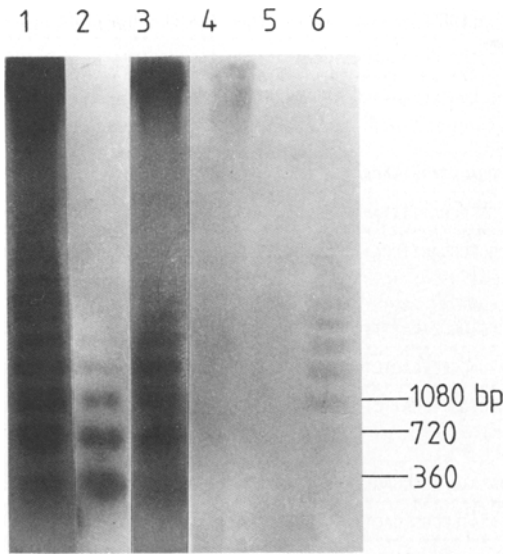

Fig. 4. Hybridization of the type IV repeat to total nuclear DNA from *Cucumis melo*. Total DNA from *Cucumis sativus* (about 50 ng, *lanes 1–3*) and *Cucumis melo* (about 10 µg, *lanes 4–6*) was separated on a 1.5% agarose gel and hybridized at reduced stringency (60 °C) to a type IV repeat. Lanes are as follows: (*1, 4*) RsaI, (*2, 5*) Sau3A, and (*3, 6*) TaqI

show no significant homology to any of the type I–III satellite repeats. This suggests that the type IV repeat is created by an insertion. An indication of this insertional event occurred when we found a 3 bp direct repeat (GTT) immediately flanking the region of homology to the types I–III (see Figs. 1 and 3). Since the difference in sequence homology of type I/II is nearly identical to the one of type III, it is not possible to determine the ancestor of the homologous type IV sequence. It is, however, evident that all repeat types are derived from the same ancestor.

A type IV homologous sequence is present in Cucumis melo

Since the satellite repeat of *Cucumis melo* is present in a low copy number in cucumber (Ganal et al. 1986), we have tested the presence of the type IV repeat of cucumber in melon. At reduced stringency there is a weak signal from the melon DNA indicating the presence of a few copies of a type IV related sequence (see Fig. 4). Compared to the pattern of cucumber there are significant differences. In cucumber only a small proportion of the type IV repeats are cut by TaqI and most of the repeats are cut by RsaI. In melon the type IV homologous sequence is almost completely cut by TaqI and only rarely cut by RsaI.

Sau3A, which cuts the type IV repeats of cucumber mostly into monomers, degrades the melon sequences into very small fragments. This could be hardly detected in the hybridization which indicates that the type IV-like repeats of melon are cut more than once by this enzyme. This, and the reduced stringency for the detection of the respective sequences, indicate a considerable sequence divergence since the separation of these two species. Hybridization with type I/II or III repeats to melon DNA does not give a satellite-like pattern. Therefore the hybridization pattern with type IV is probably due to that part of type IV which shows no homology to type I–III. No hybridization to type IV repeats was detected in the more distantly related species *Cucurbita pepo* and *Cucurbita maxima* (data not shown).

Discussion

Several different, tandemly arranged DNA sequences are now detected in the genome of *C. sativus*. Two of them show a similar repeat length of either 182 bp (type I/II) or 177 bp (type III) with a sequence homology of about 60%–63% (Ganal et al. 1986). Therefore, we concluded that they are derived from a common progenitor by mutation and subsequent amplification as it is observed for many other satellites (Roizes and Pages 1982; Flavell 1982; Singer 1982; Dover 1986; Miklos 1986). The type IV satellite repeat described here is clearly related to the previously described repeats as it is revealed by a sequence homology of ca. 65%–70% to type I–III in one half of the repeating unit, and therefore is generated from the same common ancestor as the type I–III repeats. However, it is created by a different mechanism. A 180 bp DNA sequence is probably inserted into a type I–III related sequence, or vice versa, and subsequently amplified. As a relict of this insertional event the region showing homology to the other satellites is flanked by direct repeats of three base pairs (here GTT), a phenomenon which is observed for many insertional events caused by transposition (Freeling 1984; Saedler and Nevers 1985).

Insertional events in the evolution of satellite DNA were only observed in a few species until now. In plants, the highly repeated telomeric sequences of *Secale* species are created by rearrangements, variation, insertion, and amplification (Bedbrook et al. 1980; Flavell 1982, 1986), resulting in a clustered appearance of specific elements on individual telomers. In animals, comparable events were observed in the satellite DNA of the bovine genome. From an ancestral sequence several distinct satellites are formed by similar mechanisms (Pech et al. 1979; Pöschl and Streeck 1980; Streeck 1982; Taporowsky and Gerbi 1982). For the insertion of DNA sequences transposition and RNA-mediated DNA transfer are discussed (Dover 1982, 1986).

Once such a new variant has occurred by any of these processes they can be spread over the whole genome by mechanisms like unequal crossing-over or gene conversion (Smith 1976; Dover 1982; Arnheim 1983; Dover and Tautz 1986).

The repeat was found in the related species *Cucumis melo* at a very low copy number and only at low stringency hybridization, indicating a considerable sequence heterogeneity between the repeats of these plants. Similar results were received with the melon satellite in cucumber (Ganal et al. 1986; unpublished results). Since the species *Cucumis melo* and *Cucumis sativus* are only distantly related in the *Cucumis* genus (Perl-Treves and Galun 1985; Perl-Treves et al. 1985) there is the possibility to trace the variation of these repeats by comparing more closely related species, and to get some indications on the sequence variation and rearrangements in this family during speciation and cultivation.

Acknowledgement. The support by a grant from the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

References

Arnheim N (1983) Concerted evolution of multigene families. In: Koehn R, Nei M (eds) Evolution of genes and proteins. Sinauer, Sunderland, pp 38–61

Bedbrook JR, Jones J, O'Dell M, Thompson R, Flavell RB (1980) A molecular description of telomeric heterochromatin in *Secale* species. Cell 19:545–560

Beridze T (1986) Satellite DNA. Springer, Berlin Heidelberg

Brennicke A, Hemleben V (1983) Sequence analysis of the cloned *Cucumis melo* highly repetitive satellite DNA. Z Naturforsch 38c:1062–1065

Dover G (1982) Molecular drive: A cohesive mode of species evolution. Nature 299:111–117

Dover GA (1986) Molecular drive in multigene families: How biological novelties arise, spread and are assimilated. Trends Genet 2:159–165

Dover GA, Tautz D (1986) Conversion and divergence in multigene families: alternatives to selection and drift. Philos Trans R Soc London, Ser B 312:275–289

Flavell RB (1982) Sequence amplification, deletion, and rearrangement: Major sources of variation during genome evolution. In: Dover GA, Flavell RB (eds) Genome evolution. Academic Press, London, pp 301–323

Flavell R (1985) Repeated sequences and genome change. In: Hohn B, Dennis E (eds) Genetic flux in plants. Springer, Wien New York, pp 139–156

Flavell RB (1986) Repetitive DNA and chromosome evolution in plants. Philos Trans R Soc London, Ser B 312:227–242

Freeling M (1984) Plant transposable elements and insertion sequences. Annu Rev Plant Physiol 35:277–298

Ganal M, Hemleben V (1986a) Different AT-rich satellite DNAs in *Cucurbita pepo* and *Cucurbita maxima*. Theor Appl Genet 73:129–135

Ganal M, Hemleben V (1986b) Comparison of the ribosomal RNA genes in four closely related *Cucurbitaceae*. Plant Syst Evol 154:63–77

Ganal M, Riede I, Hemleben V (1986) Organization and sequence analysis of two related satellite DNAs in cucumber (*Cucumis sativus* L.). J Mol Evol 23:23–30

Gruenbaum Y, Narveh-Many T, Cedar H, Razin A (1981) Sequence specifity of methylation in higher plant DNA. Nature 292:860–862

Grunstein M, Hogness D (1975) Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. Proc Natl Acad Sci USA 72:3961–3965

Hemleben V, Grierson D, Dertmann H (1977) The use of equilibrium centrifugation in actinomycin D-caesium chloride for the purification of ribosomal DNA. Plant Sci Lett 9:129–135

Hemleben V, Leweke B, Roth A, Stadler I (1982) Organization of highly repetitive satellite DNA of two *Cucurbitaceae* species (*Cucumis melo* and *Cucumis sativus*). Nucleic Acids Res 10:631–644

Ingle J, Timmis JN, Sinclair (1975) The relationship between satellite deoxyribonucleic acid, ribosomal ribonucleic acid gene redundancy, and genome size in plants. Plant Physiol 55:496–501

Kessler C, Neumaier PS, Wolf W (1985) Recognition sequences of restriction endonucleases and methylases – a review. Gene 33:1–102

Larson R, Messing J (1982) Apple II software for M13 shotgun sequencing. Nucleic Acids Res 10:39–49

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. Cold Spring Harbor Laboratory

Maxam A, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci USA 71:2461–2465

Messing J, Vieira J (1982) A new pair of M13 vectors for selecting either DNA strand or double-digest restriction fragments. Gene 19:269–276

Miklos GLG (1986) Localized highly repetitive DNA sequences in vertebrate and invertebrate genomes. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum Press, New York London, pp 241–322

Pech M, Streeck RE, Zachau HG (1979) Patchwork structure of a bovine satellite DNA. Cell 18:883–893

Perl-Treves R, Galun E (1985) The *Cucumis* plastome: physical map, intrageneric variation and phylogenetic relationships. Theor Appl Genet 71:417–429

Perl-Treves R, Zamir D, Navot N, Galun E (1985) Phylogeny of *Cucumis* based on isozyme variability and its comparison with plastome phylogeny. Theor Appl Genet 71:430–436

Pöschl E, Streeck RE (1980) Prototype sequence of bovine 1.720 satellite DNA. J Mol Biol 143:147–153

Ramachandran C, Brandenburg WA, den Nijs APM (1985) Intraspecific variation in C-banded karyotype and chiasma frequency in *Cucumis sativus* (*Cucurbitaceae*). Plant Syst Evol 151:31–41

Roizes GP, Pages M (1982) Conserved and divergent sequences of bovine satellite DNAs. In: Dover GA, Flavell RB (eds) Genome evolution. Academic Press, London, pp 301–323

Saedler H, Nevers P (1985) Transposition in plants: a molecular model. EMBO J 4:585–590

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5476

Shmookler-Reis R, Timmis JN, Ingle J (1981) Divergence, differential methylation, and interspersion of melon satellite DNA sequences. Biochem J 195:723–734

Singer MF (1982) Highly repeated sequences in mammalian genomes. Int Rev Cytol 76:67–112

Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. Science 191:528–535

Streeck RE (1982) A multicopy insertion sequence in the bovine genome with structural homology to the long terminal repeats of retroviruses. Nature 298:767–769

Taparowsky EJ, Gerbi SA (1982) Structure of 1.711 g/cm³ bovine satellite DNA: evolutionary relationship to satellite I. Nucleic Acids Res 10:5503–5515

Timmis JN, Ingle J (1977) Variation in satellite DNA from some higher plants. Biochem Genet 15:1159–1173

Vieira J, Messing J (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with universal primers. Gene 19:259–268