Sequence analysis of *Vicia faba* repeated DNA, the FokI repeat element

Atsushi Kato, Kimitaka Yakura\(^1\) and Shigeyuki Tanifuji

Department of Botany, Faculty of Science, Hokkaido University, Sapporo 060, Japan

Received 7 June 1984; Accepted 5 July 1984

**ABSTRACT**

A type of highly repeated DNA sequences present in the genome of *Vicia faba* was detected by digestion its nuclear DNA with FokI endonuclease and fractionating the digests on polyacrylamide gels. Four fragments of 59, 108, 177 and 246 bp of the FokI repeated sequences were collected from the gels and their primary structures were determined by the method of Maxam and Gilbert. These repeated DNA sequences were shown to be a multiple tandem array of a 59 bp sequence element. And its nucleotide sequence was almost completely conserved among all the sequence members of each the size class and also among these classes. This sequence element consists of a duplet of an about 20 bp short sequence and another unrelated 20 bp sequence. Each repeat unit of the duplet has an incomplete dyad symmetrical structure.

**INTRODUCTION**

Highly repeated DNA sequences comprise a large fraction of the genomic DNA in most eukaryotic cells. Although their functions are not clear, widespread occurrence in eukaryotes, coupled with their high concentration in heterochromatin, has led to speculation on their roles in several cellular functions (1,2,3,4). *Drosophila* genome contains a simple repeated DNA sequence which consists of a 20 bp unit sequence tandemly repeating in a homogeneous array (5,6). More highly organized repeated sequences consisting of some subrepeats, each of which had accumulated different base mutations, were detected in bovine (7,8,9), mouse (10,11,12) and some primates (13,14,15). Several types of organization pattern of repeated DNA sequences have been reported also in plants. In *Secale cereale*, tandemly repeating units, which consist of original repetitive elements and inserted unrelated sequences, were found by analysis using various restriction enzymes (16). In *Zea mays*, a 185 bp repeated sequence was reported that was localized exclusively to knob heterochromatin and its nucleotide sequence determined indicated the absence of...
any internal subrepeats (17). However, there is no clear evidence of sequence conservation among members of the 185 bp sequence element (17). Several classes of the repeated sequences found in the *Scilla siberica* genome were all shown to be very highly conserved in nucleotide sequence among members of each class and also among the classes, except the region of multiple successive insertion of the tetranucleotide G-T-C-C. However, there is no information on the relationship in the arrangement of the variants of these repeated sequences within the genome (18). Some of the various families of repeated DNA sequences found within the genome of an organism may be related to one another and useful to study the evolution of DNA sequences. The subject of this communication is the molecular organization of one type of repeated DNA sequences found in *Vicia faba*. A hypothetical model of the evolution of this repeated sequence is presented.

**MATERIALS AND METHODS**

**Extraction of DNA**

Embryonic axes used for preparation of DNA were obtained from *Vicia faba* seeds after soaking for 24h in the dark at 27°C. A nuclear fraction was obtained (19,20), from which DNA was prepared by CsCl-equilibrium centrifugation according to the method of Bedbrook (16).

**Restriction endonuclease digestion, gel electrophoresis and sequence analysis**

In most cases DNA was digested with 2 units of a restriction enzyme per μg DNA in the buffer suggested by the suppliers. DNA fragments were fractionated by electrophoresis on 1-2% agarose or 5% polyacrylamide gels. DNA fragments stained with ethidium bromide were visualized by a UV lamp and eluted electrophoretically in dialysis bags. After labeling 5'-ends of the DNA fragments with (γ-32P)-ATP, DNA was denatured in 0.3% NaOH and the strands were separated by electrophoresis on 5 or 8% polyacrylamide gels. Fast and slow strands were sequenced according to the method of Maxam and Gilbert (21).

**32P-end labeling and DNA/DNA hybridization**

The treatment of DNA fragments with alkali phosphatase and their 5'-end labeling were carried out according to Maizels (22).
DNA fragments recovered from agarose gels were denatured by alkali (23) or glyoxal and dimethyl sulfoxide before subsequent electrophoresis (24). Hybridization of DNA blotted on nitrocellulose filters was performed after preincubation of the filters at 37°C for 4h in a solution containing 50% formamide, 5 x SSC and 10 x Denhardt's solution (25,26). The filters were hybridized with heat-denatured probes in the above-mentioned solution containing additionally 100 µg/ml heat-denatured sonicated E. coli DNA for 24h at 37°C. Then the filters were washed for 6 times with the preincubation solution containing 0.1% SDS at 37°C, twice with 2 x SSC and then exposed to Fuji-RX film.

RESULTS

Isolation and purification of FokI repeated sequence

The presence of satellite DNA in the V. faba genome and its unusual non-localized distribution on all chromosomes have been reported (27). The amount of V. faba satellite DNA has been shown to be very small and the use of Ag⁺-Cs₂SO₄ gradients, instead of CsCl gradients, did not allow to increase the detectable amount of satellite (data not shown). However, when nuclear DNA was digested with FokI and produced fragments were separated on 5% polyacrylamide gels, several discrete bands were visualized, which were supposed to be a basic sequence element of a certain repeated sequence and its multimers. These fragments were resistant to digestion with EcoRI, HindIII, BamHI, Hinfl, HpaII, HaeIII or Sau3Al. Therefore, it seemed that the repeated DNA sequences revealed by FokI cut were distributed with high frequencies in some "relic DNA fractions" that were characterized by relative lacks of the target sequences for these enzymes. In the present study, HaeIII was used to prepare a relic DNA fraction. After digestion of the nuclear DNA with HaeIII, its products were layered on a 10-40% linear sucrose gradient (1M Na-acetate, 10mM Tris-HCl, pH 8.0, 1mM EDTA) in a HITACHI RPS-27 centrifugation tube, then followed by centrifugation at 24000 rpm for 15h at 20°C. After fractionation, small portions were removed from each fraction and each was digested with FokI, then subjected to electrophoresis on agarose or polyacrylamide gels to search the distribution of the repeated sequences defined by FokI cut among these fractions of
Fig. 1 DNA fragments obtained by digestion of *V. faba* nuclear DNA with FokI were separated by electrophoresis on 5% polyacrylamide gels and stained with ethidium bromide. A) The effect of enrichment of the FokI repeated sequences by preparing a HaeIII-relic DNA fraction. Twenty µg of FokI-digested *V. faba* nuclear DNA (lane 1). One µg of the HaeIII-relic DNA digested with FokI (lane 2). Repeated DNA sequences defined with FokI were enriched in the relic DNA fraction more than 20-fold. Molecular sizes of the major bands are as follows: I: 60 bp, II: 120 bp, III: 180 bp, IV: 240 bp. B) Weak or extensive digestion of the HaeIII-relic DNA with FokI. Conditions of digestion were as follows: Lane 1: 0.5 units/µg DNA for 5 min. Lane 2: 1.5 units/µg DNA for 5 min. Lane 3: 3.0 units/µg DNA for 60 min. Lane 4: 20 units/µg DNA for 240 min.

Various sizes. Next, fractions containing the repeated DNA sequences were pooled. Fig. 1A shows the degree of enrichment of the FokI repeated sequences in the HaeIII-relic DNA fraction and also their sizes. Digestion of the HaeIII-relic DNA fraction with FokI gave a typical ladder of fragments which were exact multiples in length of a basic repeat unit of approximately 60 bp. Fig. 1B shows the patterns of electrophoresis with partial di-
Fig. 2 Autoradiographs of parts of the sequencing gels of the FokI repeat fragments. Target sequences recognized by FokI are indicated by brackets, the cleavage points are indicated by arrow heads, and positions of methylated cytosines are indicated by M. F-2-1: One strand of alkali-denatured 118 bp fragment. F-2-2: Another strand of alkali-denatured 118 bp fragment. F-3-2: One strand of alkali-denatured 177 bp fragment.

Upon weak digestions, smaller fragments decreased, whereas larger fragments increased without additional occurrence of new sizes of fragments. Therefore, it
Fig. 3 Nucleotide sequences of the FokI cleaved fragments, I, II, III, IV. Indicated symbols are the same with Fig. 2.

was concluded that this repeated DNA comprised only 60 bp basic sequence elements.

Sequence analysis

The 60, 120, 180 and 240 bp fragments were eluted from the gels and their separated strands were subjected to sequence analysis. Some examples of the autoradiography are shown in Fig. 2. Although there was a slight cross contamination between two separated strands, it was quite possible to identify one nucleotide as the representative at each position of the sequence ladders. Therefore, the sequence of these fragment classes should be highly homologous among most members of each the repeating sequence. Fig. 3A shows nucleotide sequences of these 60, 120, 180 and 240 bp fragments. The results indicated that exact numbers of nucleotides of these fragments were 59, 118, 177 and 236 bp and also that the 118 and 177 bp fragments were exact dimer and trimer molecules of the 59 bp sequence element. Furthermore, there was no variation in nucleotide sequence among the 59 bp sequence elements constituting these fragment classes of a multiple series, thus again indicating a very high degree of sequence conservation. A part of the sequence of the 236 bp fragment class remained to be determined, but probably this fragment should be an exact tetramer.

Only a band corresponding to one cytosine residue in C-G sequence per the shortest 59 bp basic sequence was absent in the
sequence ladders, whereas the band of cytosines and other bases were quite clear with high intensities of radioactivity. Then, it was concluded that only two cytosine residues in both DNA strands were methylated in this 59 bp sequence element.

Reiteration of the FokI repeated sequences

To estimate the copy number of the FokI repeated sequences within the genome of *V. faba*, FokI digested nuclear DNA was electrophoresed on 2% agarose gels, in parallel with known quantities of the repeat units. Then, Southern blot hybridization was carried out using the $^{32}$P-labeled 59 bp repeat units. In the case of this experiment, as shown in Fig.4, DNA fragments 59 bp in length did not appear and a faint radioactive band of 118 bp was detected. We determined the degree of reiteration of this repeat units by scanning the autoradiograms and negative films of the stained agarose gels using a densitometer. The relative amounts of the 59, 118 and 177 bp fragments were estimated from the negative films and the whole amount of fragments of the related sequence longer than 118 bp was estimated from the autoradiograms.
As the result of summation of the relative values with all the bands, we estimated the total copy number of the repeated sequence elements to be in the range of $5 \times 10^6$ to $5 \times 10^7$ per diploid genome.

**DISCUSSION**

We could concentrate the FokI repeated sequences about 20-fold in a certain DNA fraction of high molecular weight fragments that was characterized by a relative lack of HaeIII sites. Most of contaminating DNAs in the relic DNA fraction were DNA fragments with molecular weights higher than those of the multiplied repeated sequences. Therefore, as to the DNA fractions of the bands themselves of the FokI-cleaved fragments, this type of repeated sequences was enriched more than 20 times. Preparation of relic DNA fractions is generally a beneficial method, when the size of a repeat unit is rather small. This type of repeated sequences revealed by FokI digestion consisted of only one sequence element of 59 bp with a perfectly conserved nucleotide sequence. Such the organization pattern is similar to that observed with satellite DNA I, II and IV of *Drosophila melanogaster* (5,6). In the cases of repeated sequences of bovine (7,8,9), mouse (10,11,12) and some primates (13,14,15), their nucleotide sequences exhibited more complex variation. In plants, a class of repeated sequences consisting of only basic sequence element was found in *Secale cereale* (18), *Cucumis sativus* and *Cucumis melo* (28), but these elements have not yet been sequenced. A highly repeated sequence limited to knob heterochromatin of *Zea mays* (17) was shown to consist of a repeating unit of 185 bp. Although the nucleotide sequence of one cloned member of the repeating unit had been determined, there was no evidence for sequence conservation among members of the element, nor evidence for their organization within the genome. In the case of *Scilla siberica* (16), primary structures of several classes of repeated sequences were considerably conserved among their members of each class and also among the classes, but clearly exhibited some extents of variation with several base substitutions or insertions of subsets of tetranucleotide. However, no information was given on the arrangement of these classes of *Scilla siberica* repeated sequences.
In *V. faba*, there are other types of repeated sequences showing about 20% sequence variation in each sequence family besides the FokI repeated sequences described here (29,30). Then, the facts that this repeated DNA detected by FokI cut consisted of only one sequence element and its nucleotide sequence was extremely conserved are particularly important features among several types of repeated sequences found in the *V. faba* genome.

When total nuclear DNA was digested with FokI, several fragments as the products of incomplete digestion appeared rather consistently. The cause of their occurrence is unknown. Some of the FokI sites in the repeated DNA were not digested, even when 1μg DNA was digested with 20 units of FokI endonuclease of a large amount for 6h of a longer period (Fig.1B). And these uncleaved FokI sites amounted to about 50% of total FokI sites in the repeated DNA fraction, which value was calculated from the ratio of the sum of estimated numbers of the 59 bp basic sequence unit present in each multimer to its total copy number in the whole genome. However, the bases of the recognition sequence of the uncleaved FokI sites present in these multimer fragments were supposed to be not methylated, because the bands corresponding to the bases of the target sequence for FokI were very clear in the autoradiographs of sequencing gels. This imperfect digestion might be due to a certain secondary structure of the DNA.

Generally, plant DNAs have high contents of 5'-methylcytosine and particularly its methylation is found in the cytosine residues in C-G dinucleotide sequence. For example, in *Scilla siberica* and *Cucumis sativus*, about 55 or 30% cytosines of satellite DNAs are methylated (18) and in wheat total DNA (31) cytosines are methylated in about 82% C-G sequences. The fact that *V. faba* DNA was not digested mostly with *MluI*, *HhaI* or *HpaII* suggests a high level of methylation in most C-G sequences. There is one C-G sequence in the 59 bp repeated sequence element and in fact the cytosine residues at this position were mostly methylated. This result was similar to the case observed with the repeated DNA of *Scilla siberica*.

As shown in Fig.5A, the 59 bp sequence element consists of a duplet of an about 20 bp short sequence and another unrelated 20 bp sequence. The sequence homology between two subsets (19 and
The sequences and nucleotides showing homology between the two tandem subsets are boxed. Bar underlined indicates the target sequence recognized with FokI and an arrow head indicates the cleaved site by FokI. B) Dyad symmetrical structures of each of the duplet sequence. C) Intrastrand symmetrical structures of each subsets of the duplet sequence.

20 bp) of the duplet was more than 80%. Then, it seems likely that this 59 bp repeat unit had been formed by a tandem duplication of an ancestral 19-20 bp sequence and subsequent association with an unrelated 20 bp sequence. This hypothesis for generation of this 59 bp repeat unit resembles that for a sub-repeating unit of about 325 bp found in the rDNA of _V. faba_ (32). Although we can not image its detail at present, there might be a somewhat common mechanism to yield such the structure. Fig.5B shows secondary structures of each half of the duplex. These 19 and 20 bp sequence both have imperfect dyad symmetry, which resembles that found with _Scilla siberica_ satellite DNA (18). On the other hand, as shown in Fig.5C, each half of the duplex, the 19 and 20 bp short sequences, have imperfect intra strand symmetry, which resembles that found with an ancestral sequence of bovine satellite DNA (9). Although, we have no evidence to determine which model of dyad symmetry or intra strand symmetry re-
fects its ancestral structure, such the molecular structures, either or both, might be related with establishment and development of a certain repeated sequence.

We could not detect any sequences being homologous to this \( V. \) \( \text{faba} \) FokI repeats in the genomes of other some species of \textit{Leguminosae} by Southern blot hybridization (data not shown). Furthermore, any DNA sequences contained in the \( V. \) \( \text{faba} \) genome except those FokI repeated sequences were not hybridized with this 59 bp FokI repeat unit, when digests of \( V. \) \( \text{faba} \) DNA with various restriction enzymes were analyzed (data not shown). On the basis of these results, we speculate that, although tandem duplication of the 19-20 bp short sequence had occurred at a certain very old time, reiteration of the 59 bp repeat unit has occurred rather recently during the course of evolution.

ACKNOWLEDGEMENTS

We thank Professor M. Sugiura and Dr. K. Shinozaki of Nagoya University and Dr. J. Kusuda and Mrs. M. Kusuda of National Institute of Genetics, Mishima for their kind help and discussion. This study was supported by the grants from the Ministry of Education, Science and Culture of Japan (No. 58119002, 58212001, 58480009) and the Mishima Kaiun Science Foundation.

\(^1\)Present address: Department of Biology, Faculty of Education, Kanazawa University, Kanazawa 920, Japan

REFERENCES