Sequence analysis of a faba bean necrotic yellows virus DNA component containing a putative replicase gene

Lina Katul,* Edgar Maiss and H. Josef Vetten

Biologische Bundesanstalt für Land- und Forstwirtschaft, Institut für Biochemie und Pflanzenvirologie, Messeweg 11-12, 38104 Braunschweig, Germany

Faba bean necrotic yellows virus (FBNYV) has a circular ssDNA genome possibly consisting of several components of about 1 kb each. The complete nucleotide sequence of one component of FBNYV (FBNYV DNA 1) containing a putative replicase gene is presented. This component consists of 1002 nucleotides and, in the virion orientation, contains one large open reading frame (ORF1) potentially encoding a 32.3 kDa replicase with the NTP-binding motif GGEKKS. No obvious functions could be assigned to two smaller ORFs (7.4 and 9.3 kDa) occurring in the complementary orientation. Amino acid sequence comparisons of the putative replicase of FBNYV with that of other similar ssDNA viruses yielded higher homologies to subterranean clover stunt virus than to banana bunchy top and coconut foliar decay viruses. A potential stem-loop structure and a TATA box were identified within the noncoding region. Two oligonucleotides derived from FBNYV DNA 1 were used for direct sequencing of the virion ssDNA to determine its virion polarity and for amplifying part of this component by immunocapture PCR in extracts from FBNYV-infected plants.

Introduction

Faba bean necrotic yellows is an economically important disease of crop and pasture legumes in west Asia and north Africa. Isometric virus-like particles are closely associated with the disease and have been regarded as the putative causal agent although their infectivity has not yet been shown. The putative causal agent is not transmitted mechanically, but by aphids in a persistent manner. The virus-like particles measure 18 nm in diameter, contain a capsid protein of about 22 kDa and circular ssDNA of about 1 kb and are called faba bean necrotic yellows virus (FBNYV) (Katul et al., 1993). Based on these criteria FBNYV is very similar to banana bunchy top virus (BBTV) (Thomas & Dietzgen, 1991; Harding et al., 1991, 1993), coconut foliar decay virus (CFDV) (Randles & Hanold, 1989; Rohde et al., 1990), milk vetch dwarf virus (MDV) (Sano et al., 1993) and subterranean clover stunt virus (SCSV) (Chu & Helms, 1988). In addition, MDV and SCSV particularly resemble FBNYV in not only having the same aphid vector species and a similar host range, but also in being serologically related (Katul et al., 1993). Recent studies on the genome of BBTV (Harding et al., 1993; Burns et al., 1994; Yeh et al., 1994), CFDV (Rohde et al., 1990) and SCSV (Boevink et al., 1993; Chu et al., 1993; Surin et al., 1993) confirmed the tentative grouping of these viruses since their circular ssDNA genomes share similar organization and appreciable sequence homologies. Apart from CFDV, the genome of which has not been further analysed, BBTV and SCSV were shown to have a multipartite genome, consisting of at least six and seven circular components, respectively (Burns et al., 1993; Boevink et al., 1993). This clearly differentiates them from geminiviruses, the only other group of plant viruses with a circular ssDNA genome (Lazarowitz, 1987). Based on the results from our previous cloning and hybridization experiments with FBNYV DNA (Katul et al., 1993) and by analogy with BBTV and SCSV, we expect the genome of FBNYV to be made up of several covalently closed circular ssDNA components of about 1 kb each. In this paper we present the sequence of the first FBNYV DNA component and compare it with the corresponding genome components of similar viruses. We also describe the use of the polymerase chain reaction (PCR) for the amplification and detection of viral DNA from plant extracts.

Extraction of the nucleic acid from purified virus preparations and cloning of the DNA into the phagemid vectors pGEM-3Zf(+) (Promega) and pT7T3 19U (Pharmacia LKB) was as previously described (Katul et
Fig. 1. Complete nucleotide sequence of the FBNYV DNA 1 component and the amino acid sequence of ORF1, encoding a putative replicase. The stem-loop structure is in bold and underlined, the TATA box (TATAAA) and the poly(A) signal (AATAAG) are in bold and double underlined, and the potential NTP-binding motif (GGEGKS) is in bold and shaded. The positions of primer 1 (I~ I~) and the reverse complement of primer 2 ('II .q) are indicated by arrows over the corresponding nucleotide sequence.

al., 1993). The dideoxynucleotide chain termination sequencing method (Sanger et al., 1977) was conducted using the T7 sequencing kit (Pharmacia LKB) on ssDNA templates generated with the helper phage M13K07. Sequences were analysed using the program package of the Genetics Computer Group (Devereux et al., 1984). Sequences of at least 20 templates from independent clones overlapped partially or completely with no mismatches. When connected, they formed a covalently closed circle of 1002 nucleotides, thereafter called FBNYV DNA 1 (Fig. 1). Moreover, several fragments of about 1 kb (from agarose gel analysis) were found to have 1002 bp when completely sequenced and were therefore considered to contain the full length of this circular component.

Sequence analysis of this component revealed three open reading frames (ORFs) potentially encoding proteins larger than 5 kDa (Fig. 2). One large ORF (ORF1) in the virion orientation (polarity determination described below) spans from nucleotides 163 to 996 (278 amino acids) and codes for a protein of 32.3 kDa. There are two other smaller ORFs in the complementary orientation (ORF2 from nucleotides 354 to 160, 65 amino acids, 7.4 kDa; ORF3 from nucleotides 949 to 719, 77 amino acids, 9.3 kDa). ORF1 is preceded by a TATA box (TATAAA) 33 nucleotides upstream from the ATG start codon and contains a polyadenylation [poly(A)] signal
poly(A) signals in vertebrate systems, a high degree of the possible poly(A) signal based on the findings of considered the further downstream variant, AATAAG, as the TGA stop codon (Fig. 1). Although the consensus (TA) = TA, TT or AA. consensus sequence from 15 geminiviruses. X = A, T, C or none; region of several ssDNA viruses. The geminivirus sequence is a virus poly(A) site. Moreover, ORF1 potentially encodes directed efficient processing at the cauliflower mosaic this consensus sequence were actually recognized and tolerated by plants, and that all single base variants of the consensus sequence (AATAAA) for the poly(A) is also present, it (AATAAG) which falls 34 nucleotides upstream from the TGA stop codon (Fig. 1). Although the consensus sequence (AATAAA) for the poly(A) is also present, it falls 53 nucleotides further upstream. We chose to consider the further downstream variant, AATAAG, as the possible poly(A) signal based on the findings of Rothnie et al. (1994). They suggested that, in contrast to poly(A) signals in vertebrate systems, a high degree of sequence variation of the AATAAA motif can be tolerated by plants, and that all single base variants of this consensus sequence were actually recognized and directed efficient processing at the cauliflower mosaic virus poly(A) site. Moreover, ORF1 potentially encodes a protein with replicase activity since it contains the NTP-binding motif GXGKS [G(GE)GKS] which has been shown to be present in the replicases of several viruses (Gorbalenya et al., 1990) and may therefore encode a protein with replicase activity. Boevink et al. (1993) showed that the SCSV genome has two distinct, possibly co-existing or interchangeable components (SCSV2 and SCSV6) which potentially code for a replicase. Whether a similar phenomenon exists for FBNYV isolates is not known yet. Further work aimed at studying the expression level of these potential ORFs and their function will be required.

Alignments of the amino acid sequence of the putative replicase of FBNYV DNA 1 with those of SCSV2, SCSV6 (Boevink et al., 1993, and personal communication), CFDV (Rohde et al., 1990) and BBTV (Harding et al., 1993) gave sequence identities of 58.5%, 47.7%, 35.7% and 35.0%, respectively, when using the program Align (Lipman & Pearson, 1985). This sequence comparison supports our earlier observations that FBNYV is more similar to SCSV than to BBTV and CFDV (Katul et al., 1993). Sequence data for MDV, which appeared to be serologically closely related to FBNYV (Katul et al., 1993), have not been reported yet. For ORF2 and ORF3, the two smaller ORFs, no striking amino acid sequence similarities to any of the ORFs of BBTV and CFDV were revealed. Similarly, no significant amino acid or nucleotide sequence similarities were found in the GenBank database and hence no obvious function could be assigned to them.

Further sequence analysis of the FBNYV DNA 1 component revealed a potential stem-loop structure outside the coding region consisting of a 10 bp stem (5' CCAGGGCGG 3' and 5' CCCGGCCTTTG 3') and an 11 nucleotide loop (5' TATAGTATTAC 3') (Fig. 3a). Similar to geminiviruses, the potential stem-loop structure consists of a GC-rich base-paired stem and an AT-rich loop (Lazarowitz, 1987). Lazarowitz et al. (1992) and Fontes et al. (1994) have pointed out that the stem–loop structure is highly conserved among the components of each geminivirus and have suggested that it is necessary for DNA replication and may be part of the viral replication origin. Partial comparisons of the noncoding region of FBNYV DNA 1 with that of BBTV (Harding et al., 1993; Yeh et al., 1994), CFDV (Rohde et al., 1990) and several geminiviruses (Faria et al., 1994) confirmed earlier findings that the potential loop structure is a highly conserved region among these viruses (Fig. 3b). Moreover, the noncoding region was also found to contain sequence stretches common among the different genome components of BBTV (Burns et al., 1993, 1994) and SCSV (Chu et al., 1993; Boevink et al., 1993, and personal communication). Yeh et al. (1994) referred to another domain of 12 nucleotides (5' ATTTAAATATTG 3') which is present within the coding region of BBTV component 1 (Harding et al., 1993) and of CFDV DNA (Rohde et al., 1990), as well as within the noncoding region of BBTV component 2 (Yeh et al., 1994). This domain was not found in the FBNYV DNA 1 component. However, long stretches of high levels of amino acid identity were observed in the large ORFs of FBNYV DNA 1 and SCSV2, but were chiefly restricted to their C-terminal halves (data not shown).

To determine the polarity of the virion ssDNA, two 19-mer oligonucleotides (primer 1, 5' AATTAATAT
GGCTTGTTC 3'; primer 2, 5' AAACAAATTCAACATATTGA 3') (Fig. 1) derived from the FBNYV DNA 1 were synthesized and used separately for direct sequencing of the noncloned viral ssDNA. Only primer 2 primed with the viral ssDNA producing a readable

![Fig. 3. (a) Putative stem-loop structure of the FBNYV DNA 1 component; (b) nucleotide sequence alignment of the potential loop region of several ssDNA viruses. The geminivirus sequence is a consensus sequence from 15 geminiviruses. X = A, T, C or none; (TA) = TA, TT or AA.](attachment://image.png)
Fig. 4. Agarose gel analysis of the PCR products. Primers 1 and 2 were used in an IC-PCR experiment to amplify a 852 bp fragment from FBNYV-infected pea tissue (lane 2) and from cloned viral DNA (lane 3), but not from noninoculated pea tissue (lane 1). Lane M contains a fragment of the same size.


(Received 8 August 1994; Accepted 4 October 1994)