Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome

Lina Katul,1 Tania Timchenko,2 Bruno Gronenborn2 and H. Josef Vetten1

1 Biologische Bundesanstalt für Land- und Forstwirtschaft, Institut für Biochemie und Pflanzenvirologie, Messeweg 11–12, 38104 Braunschweig, Germany
2 CNRS, Institut des Sciences Végétales, 91198 Gif-sur-Yvette, France

Introduction

Faba bean necrotic yellows virus (FBNYV) is associated with virus diseases affecting several important food and fodder legumes (Katul et al., 1993). The known experimental and natural host range of FBNYV includes about 60 species, the majority of which belong to the Fabaceae (Franz et al., 1995, 1997). Its geographical distribution is confined to west Asian and north African countries, in some of which it occurs at a high incidence and causes destructive diseases (Makkouk et al., 1992, 1994; Franz et al., 1997). FBNYV is persistently transmitted by various aphid species, of which Aphis craccivora is the most significant natural vector (Franz et al., 1998). It has unusually small isometric particles (18 nm in diameter) made up of a single capsid protein of about 20 kDa (Katul et al., 1993). The FBNYV genome has been reported to consist of at least six circular ssDNA components, each of which is about 1 kb (Katul et al., 1995, 1997) and is probably encapsidated in a separate particle.

FBNYV shares vector-transmission and particle properties (Inouye et al., 1968; Chu & Helms, 1988; Harding et al., 1991; Thomas & Dietzgen, 1991; Franz et al., 1998) as well as genome composition and organization (Harding et al., 1993; Boevink et al., 1995; Burns et al., 1995; Sano et al., 1993) with banana bunchy top virus (BBTV), milk vetch dwarf virus (MDV) and subterranean clover stunt virus (SCSV). These four viruses have been proposed to form a new taxonomic group of ssDNA viruses of plants (Katul et al., 1997) and are referred to as nanoviruses. Coconut foliar decay virus (CFDV) has tentatively been assigned to the recently proposed genus, Nanoviruses. Only one CFDV genome component, of about 1·3 kb, which is proposed to be transcribed bi-directionally, has
been identified, and the virus is transmitted not by an aphid but by a cixiid bug, *Myndus taffini* (Rohde et al., 1990).

Whereas the genomes of BBTV (Burns et al., 1995) and SCSV (Boevink et al., 1995) have been shown to have at least six and seven ssDNA components, respectively, Sano et al. (1998) describe in the following paper that the MDV genome may consist of as many as ten ssDNA components. Therefore, the genetic information of nanoviruses seems to be distributed over six to ten molecules of circular ssDNA. However, neither purified virions nor cloned DNA components have been shown to be infective in whole plants for any of these viruses; therefore, the components constituting the complete genomes are not known.

Each DNA component of the nanoviruses contains a major ORF and appears to be transcribed unidirectionally. Each coding region is preceded by a TATA box and followed by a polyadenylation signal. At least one of the genome components of each virus encodes a replication-associated (Rep) protein (32–4–33-6 kDa). Nicking and joining activities have recently been demonstrated for the Rep protein encoded by BBTV-C1 (Hafner et al., 1997). A second virion-sense ORF, completely within the Rep protein ORF and encoding a putative 5 kDa protein, was also identified (Beetham et al., 1997). Five to six genome components of each nanovirus encode other putative viral proteins, and are referred to here as non-Rep components. Some non-Rep components have been definitively or tentatively identified as capsid proteins (18–7–20-1 kDa) and movement proteins (12–7–13-7 kDa) (Chu et al., 1993; Boevink et al., 1995; Burns et al., 1995; Katul et al., 1997; Wanitchakorn et al., 1997; Sano et al., 1998). Moreover, complementary-strand synthesis from BBTV DNA has been attributed to a population of endogenous primers, derived predominantly from BBTV-C5 (Hafner et al., 1997). All nanovirus DNA molecules have a putative stem–loop (SL) structure in their non-coding region. This structure is highly conserved among the individual components of each virus, as well as among the four nanoviruses, and is very similar to that of geminiviruses.

In addition to the six previously identified ssDNA components of FBNYV (Katul et al., 1995, 1997), we here present sequence information on four further genome components of the same isolate from Syria (FBNYV-Sy). By analysing ten genome components of an Egyptian isolate of FBNYV (FBNYV-Eg), we provide evidence for the presence of four Rep and six non-Rep components associated with the FBNYV genome. Moreover, we present information on the potential function of FBNYV-C10, as well as on the conservation and variability of individual components within the genome of FBNYV and other known nanoviruses.

## Methods

### Viral DNA preparation.

DNA used for sequencing was from the previously described DNA library of the Syrian isolate SY292-88 (referred to here as FBNYV-Sy; Katul et al., 1993). DNA from another FBNYV isolate, EV1-93 from Egypt (referred to as FBNYV-Eg), was also analysed in this study. For cloning genome components of FBNYV-Eg, total DNA was extracted from EV1-93-infected *Vicia faba* plants according to Noris et al. (1996), and fractionated on a 1% agarose gel. DNA fragments ranging in mobility from 0.5 to 2 kb were excised from the gel and eluted electrophoretically.

### Sequencing, PCR amplification and cloning.

Sequencing of DNA, PCR amplification of the circular components using purified viral ssDNA as the template, and their cloning into a dTTP-tailed EcoRV-cut pBluescript SK(−) (Stratagene), were performed as described previously (Katul et al., 1995, 1997). For PCR amplification of viral DNA, either an aliquot from a total DNA preparation (Noris et al., 1996) was used as template (FBNYV-Eg) or a simplified ‘one-tube-reaction’ (omitting the DNA extraction steps following trapping of virions) of the previously described immunocapture PCR (Katul et al., 1995) was carried out (FBNYV-Sy and -Eg). Total DNA extracted from *Vicia faba* leaves infected with FBNYV-Eg was digested singly with several restriction enzymes. One clone, containing an insert of 1 kbp resulting from a digestion with *PsI* and subsequent ligation into *PsI*-cut pUC19 (Yanisch-Perron et al., 1985), was further analysed in this study.

### Sequence analysis, alignment and phylogeny.

Sequences were analysed with the Genetics Computer Group software (Devereux et al., 1984). Multiple sequence alignments were done with CLUSTAL W (Higgins et al., 1992; Thompson et al., 1994). Trees were constructed by the neighbour-joining method of Saitou & Nei (1987) contained in the CLUSTAL W software, and the reliability of the trees was evaluated statistically by using 100 bootstrap replicates (Felsenstein, 1985). Database searches were performed as described earlier (Katul et al., 1997). Accession numbers of sequences used for comparison were those listed by Katul et al. (1997), as well as Y11405–Y11409 (FBNYV-C2 to -C6), AB000920–AB000927 (MDV-C1 to -C8), AB000946 (MDV-C9) and AB000947 (MDV-C10).

## Results

### Identification of four further components of the FBNYV-Sy genome

Upon screening of the FBNYV-Sy DNA library for further unassigned fragments of viral DNA, four clones with inserts of 200–400 bp were identified, sequences of which were distinct from one another and from the six previously identified components (C1–C6) of the FBNYV-Sy genome. Using primer pairs specific to these sequences, the missing segment of each of the new components was amplified and sequenced. Sequence analysis revealed the presence of four further covalently closed circular components, each about 1 kb in size (C7, 1014 bases; C8, 986 bases; C9, 1004 bases; and C10, 996 bases). Similar to the six previously reported FBNYV-Sy components (Katul et al., 1995, 1997), each of the four components contained one major virion-sense ORF potentially encoding a protein larger than 10 kDa, flanked by a putative transcription initiation signal (TATA box) and a polyadenylation signal. No ORFs larger than 10 kDa were located on the complementary strands. In the non-coding region, a stretch of 31–39 bases was present, of which 10–13 bases were directly inverted repeats potentially forming an SL structure. The SL also contained the conserved nonanucleotide sequence (AGTATTACC) that may
be the origin of a rolling-circle replication mechanism (Koonin & Ilyina, 1992; Laufs et al., 1995c). Further analysis of the major ORFs revealed that two components, C7 and C9, encoded putative proteins with predicted molecular masses of 33±2 and 32±7 kDa, respectively, whereas the other two components, C8 and C10, encoded smaller putative proteins of 17±4 and 19±7 kDa, respectively.

Putative functions of C7–C10 of FBNYV-Sy

Sequence comparisons showed that the major ORFs of C7 and C9, similar to those of C1 and C2 of FBNYV-Sy (Katul et al., 1995, 1997), potentially encode Rep proteins. This was evident from the presence of a conserved tyrosine (Y) at amino acid position 78 or 79 and of a highly conserved P-loop motif [GXXGXXGK(S/T)], both of which are also conserved in the Rep proteins of geminiviruses (Laufs et al., 1995a) and other nanoviruses (Katul et al., 1997).

The putative protein encoded by C8 was strikingly similar to the proteins encoded by MDV-C6, SCSV-C4 and BBTV-C6 (Fig. 1a). The presence of a conserved LXCXE motif in the C-terminal one-third of the FBNYV-C10 protein, as well as in its homologues from the other three nanoviruses (Fig. 1b), suggests that this protein might interact with a retinoblastoma-like plant protein (Xie et al., 1995).

Comparison of the non-coding regions of the ten components of FBNYV-Sy

Pairwise comparisons of the entire non-coding regions (482–645 bases) of the six non-Rep components (C3–C6, C8 and C10) revealed two major clusters, differing by about 50%. One cluster consisted of C4 and C6 (83±4% identity) and the other cluster comprised the rest, of which C3 and C10 shared an identity of 95±7%, C5 had identities of 72±4 and 74±6% to C3 and C10, respectively, and C8 had identities to the three others of about 62%.

Alignments of the non-coding regions of all ten components of FBNYV-Sy revealed two stretches that were highly conserved among at least five components. One common region (CR), consisting of 73 bases and encompassing the SL, was highly conserved among seven of the ten components, the six non-Rep and one (C2) Rep component. This CR (CR-SL) consisted of 25–28 bases 5′-proximal of the SL, 29–33 bases constituting the SL, followed by 15–16 bases 3′-proximal of the SL (Fig. 2a).

A second conserved region (CR-II) of 81–85 bases was identified 3′–24 bases 5′-proximal of the CR-SL, and was most conserved among five (excluding C5) of the six non-Rep components (Fig. 2b). Pairwise identities between these five non-Rep components within this region ranged from 65·4 to
Fig. 2. Alignments of the stem–loop common regions (CR-SL) (a) and of a second common region (CR-II) 5′-proximal of the stem–loops (b) of FBNYV-C2 to -C6, -C8 and -C10. In (a), the sequences potentially forming a stem–loop structure in each component are in bold. Nucleotides identical in at least four components are boxed; dots denote gaps.

Fig. 3. Amino acid sequence alignment of the putative Rep protein encoded by FBNYV-C1-Eg with those encoded by MDV-C2, SCSV-C2 and FBNYV-C1-Sy. Amino acid residues identical to those of FBNYV-C1-Eg are indicated by dashes; dots denote gaps. The numbering of amino acids is given on the right. For virus acronyms see text.

Genome analysis of FBNYV-Eg

Using the same primer pairs designed for PCR amplification of each of the ten FBNYV-Sy components, it was also possible to obtain specific PCR products of C2–C10 of FBNYV-Eg. Analysis of the complete (C2, C7–C9 and C10) or partial (C5, 760 bases) sequences of six components of FBNYV-Eg showed that their sequences were 96% (C8 and C10) to 99% (C2) identical to the corresponding components of FBNYV-Sy (data not shown). Sequences of C3, C4 and C6 have not yet been determined, but PCR products of the expected sizes were amplified from FBNYV-Eg, indicating that these components are also present in this isolate.

When a total DNA preparation of FBNYV-Eg was used for
cloning, the complete sequence of an apparently new DNA component (C1-Eg, 1004 bases) was obtained which was only partially related to C1 of FBNYV-Sy (Fig. 3). C1-Eg contained a major ORF in the virion sense, which potentially encoded a protein of 32.6 kDa and was flanked by a TATA box and a polyadenylation signal. In addition to the characteristic motifs present in the Rep protein encoded by this component, nucleotide and amino acid sequence alignments of C1-Sy and C1-Eg revealed identities of 66 and 63.5%, respectively. Hence, C1-Sy and C1-Eg are much more distinct from each other than the three other Rep-encoding components (C2, C7 and C9) are from their counterparts (>96% identity) in the FBNYV-Sy and -Eg genomes. However, C1-Eg and C1-Sy were strikingly similar (91% identity) in the C-terminal 120 amino acids of their putative products (Fig. 3), and had similar SLs, suggesting that the encoded Rep proteins may have similar functions.

To examine whether each of the components C1-Sy and C1-Eg was also present in the genome of the heterologous isolate, a primer pair derived from an almost completely conserved sequence (with one base difference) in their coding regions was used for PCR amplification of a product of about 850 bp from each isolate. RFLP analysis of the resulting products revealed that the FBNYV-Eg and -Sy genomes each contain only one C1-like component.

**Phylogeny of proteins encoded by nanoviruses**

Comparison of 15 potential Rep proteins identified to date from BBTV, CFDV, FBNYV, MDV and SCSV showed that there is generally a high variability among Rep proteins, not only from distinct nanovirus species but also from a given virus (Fig. 4). High levels of identity (>80%) were observed only between the Rep protein of FBNYV-C7 and that of MDV-C10 (89.8%), and between the C1 Rep protein of FBNYV-Eg and the Rep proteins of MDV-C2 (88.3%) and SCSV-C2 (81.4%).

The newly identified C9 Rep protein present in FBNYV-Sy and -Eg was <56.2% identical to any of the other 14 Rep proteins. The least related Rep protein of the nanoviruses is that of CFDV, which was <45.6% identical to any of the other Rep proteins (Fig. 4).

Comparison of 21 non-Rep proteins identified so far from BBTV, FBNYV, MDV and SCSV revealed various levels of identity (Fig. 5) between proteins that have been shown to, or, unknown
on the basis of sequence homology, suspected to have similar functions (Chu et al., 1993; Burns et al., 1995; Katul et al., 1997; Wanitchakorn et al., 1997; Sano et al., 1998). Highest identities were observed between the putative protein of FBNYV-C8 and those of MDV-C6 (90.8%), SCSV-C4 (65.8%) and BBTV-C6 (47.7%) (Fig. 5a). The similarities in the capsid proteins, putative movement proteins, proteins of unknown function(s) and putative Rb-binding proteins encoded by FBNYV-C5, -C4, -C3 and -C10, respectively, and by their corresponding MDV and SCSV components, were also highly significant, ranging from 70.8 to 83.6% between FBNYV and MDV and from 43.8 to 55.1% between FBNYV (or MDV) and SCSV (Fig. 5b–e). Therefore, FBNYV appears most closely related to MDV, immediately related to SCSV and distantly related to BBTV (identities of < 24.6%). The least closely related proteins (52% identity) of FBNYV and MDV were the apparently homologous proteins encoded by their C6 and C7 components (Sano et al., 1998), respectively, for which no counterparts have been identified in the SCSV and BBTV genomes (Fig. 5f).

Discussion

We have determined the sequences of four further components of the FBNYV-Sy genome and presented evidence that the FBNYV-Sy genome consists of ten circular ssDNA components (each about 1 kb), which potentially encode four distinct Rep and six non-Rep proteins (see also Katul et al., 1995, 1997). The presence of four distinct Rep and six non-Rep components in the FBNYV-Sy genome is further supported by sequence data and PCR results from another FBNYV isolate, FBNYV-Eg, and by data presented by Sano et al. (1998) in the following paper on MDV, a close relative of FBNYV. Thus, the FBNYV and MDV genomes appear strikingly similar, not only in the total number of identified components but also in the number and types of related non-Rep components present in each genome. The FBNYV non-Rep components C8 and C10 identified in this study are clearly distinct from one another and from the previously described non-Rep components, and both have counterparts in the MDV, SCSV and BBTV genomes. The fact that a non-Rep component related to FBNYV-C3 has not been identified in the BBTV genome, and that neither the BBTV nor SCSV genomes have known counterparts to FBNYV-C6 (Fig. 5), suggests either that some non-Rep components play specific roles in the biology and/or replication of particular viruses or that one and two non-Rep components have yet to be identified in the SCSV and BBTV genomes, respectively.

In addition to the four previously described FBNYV non-Rep components, of which C3 and C6 have unknown functions, C5 encodes the capsid protein, and C4 encodes a putative movement protein (Katul et al., 1997), the two newly identified components C8 and C10 also seem to have different functions. Whereas the function of the putative protein encoded by FBNYV-C8 and its counterparts in the other nanoviruses is unknown, the putative proteins encoded by FBNYV-C10 and its three homologues contain the amino acid motif LXCXE, which is present in the RepA proteins of subgroup I geminiviruses (mastroviruses) and has been shown to be required for efficient viral DNA replication (Xie et al., 1995). Although it remains to be demonstrated that the FBNYV-C10 protein and its homologues from the other nanoviruses share a common function with the RepA proteins of mastreviruses, the presence in nanoviruses of a protein potentially interacting with an Rb-like plant protein is not entirely unexpected, given the similarity to geminiviruses in their mode of DNA replication (Laufs et al., 1995a; Hafner et al., 1997). In BBTV, the C5 protein contains an LXCXE motif and may hence interact with an Rb-like plant protein. In addition, it has been demonstrated recently that complementary-strand synthesis of BBTV DNA is primed by non-coding sequences derived from its component C5 (Hafner et al., 1997). The origin of the BBTV primers, and the observation that Rb-binding proteins are involved in cell-cycle regulation (Weinberg, 1995), indicate that the putative Rb-binding proteins of nanoviruses may be early virus gene products which might trigger induction of the DNA replication machinery and provide a favourable environment for efficient DNA replication (Bisaro, 1996).

In addition to the two Rep components (C1 and C2) identified previously (Katul et al., 1995, 1997), we have determined the sequences of two further Rep protein-encoding components (C7 and C9) in two FBNYV isolates. Moreover, we confirmed the presence of a component homologous to FBNYV-C2 in FBNYV-Eg and detected a variant of C1 in the FBNYV-Eg genome. Of the four FBNYV-Sy Rep proteins, that encoded by C7 had 89.8% identity to one of the MDV Rep proteins (C10), whereas the Rep proteins of C1, C2 and C9 were less similar (< 62.4% identity) to those of the four other nanoviruses (Fig. 4). In contrast, the C1 Rep protein of FBNYV-Eg, which differed (63.5% identity) from that of FBNYV-Sy, was closely related to the Rep proteins of MDV-C2 (88.3%) and SCSV-C2 (81.4%), and more similar to those of BBTV-T1 and -T2 (about 53%) than to that of BBTV-C1 and those of all other Rep protein-encoding components (< 44.2%). C1-Eg, MDV-C2, SCSV-C2 and BBTV-T1 and -T2 seem to be the only Rep components the interrelationships of which largely reflect the overall levels of similarity of the non-Rep components of FBNYV, MDV, SCSV and BBTV. This, together with the observation that FBNYV-C1, MDV-C2, SCSV-C2 and BBTV-T1 and -T2 share similar patterns of identities with all other Rep proteins (Fig. 4), may indicate that they are counterparts of one particular Rep protein.

The Rep protein encoded by FBNYV-C2 may play a pivotal role in FBNYV replication, on the basis of the following observations: (i) C2 is the only Rep component that shares a CR-SL and, to a lesser extent, a CR-II with the six non-Rep components (Fig. 2); and (ii) a C2-like component was detected in all (about 30) samples collected from FBNYV-infected plants.
in eight different countries, by RFLP and single-strand conformation polymorphism analysis of PCR products (Franz, 1997; unpublished observations). In these respects, FBNYV-C2 resembles BTV-C1, which is regarded as an integral part of the BTV genome (Karan et al., 1994; Horser et al., 1996) and whose Rep protein shares the highest identity (54.7%) with that of FBNYV-C2 (Fig. 4).

The presence of four genome components potentially encoding distinct Rep proteins in two FBNYV isolates and in MDV is puzzling. This contrasts with the geminivirus genome, which contains only one (begomoviruses) or two (mastreviruses) Rep genes (Lazarowitz, 1992). Two Rep components have also been described for SCSV (Boevink et al., 1995), and there is evidence for the presence in BTV-infected bananas of at least two further Rep components (Wu et al., 1994; Horser et al., 1996) in addition to BTV-C1 (Burns et al., 1995). Since we did not identify components potentially encoding proteins with similar functions other than the four Rep components, we believe that the presence of four Rep components is not an indication of a mixed infection by four co-existing FBNYV strains. The Rep proteins of geminiviruses are multifunctional proteins with distinct functional domains, some of which are able to exert their biochemical activities independently of the rest of the protein (Heyraud-Nitschke et al., 1995; Laufs et al., 1995b; Jupin et al., 1995; Choi & Stenger, 1995). A similar ‘functional subdivision’ may exist among the smaller Rep proteins of FBNYV and its relatives, not among protein domains, however, but rather among individual proteins. Since biological functions for conserved amino acid sequence motifs have been demonstrated in geminiviruses (Laufs et al., 1995c; Desbiez et al., 1995; Xie et al., 1995) and in BTV (Haffner et al., 1997), similar functions may be distributed among distinct proteins, e.g. the Rep proteins and the FBNYV-C10 protein and its homologues in the other nanoviruses.

Analysis of the 21 known non-Rep proteins from the four nanoviruses (Fig. 5) suggests the following: (i) all six non-Rep proteins of FBNYV are most closely related to those of MDV (52–90.8%); (ii) the FBNYV and MDV non-Rep proteins are intermediately (43.8–66.7%) and distantly (13.9–47.7%) related to those of SCSV and BTV; respectively; (iii) some of the nanovirus non-Rep proteins are more conserved than others. The most conserved non-Rep protein of the nanoviruses was that of FBNYV-C8 and its related components (45.7–90.8%), followed by the capsid proteins (19.6–83.6%), the putative movement proteins (13.9–76.6%), protein homologues of FBNYV-C3 (13–74%) and the putative Rb-binding proteins (19.1–70.8%). The apparently related proteins encoded by FBNYV-C6 (Katuil et al., 1997) and MDV-C7 (Sano et al., 1998), for which no counterparts have been identified in the SCSV and BTV genomes, were only 52% identical. This suggests that, if SCSV and BTV components related to FBNYV-C6 and MDV-C7 are present at all, they may be only remotely related to the latter two.

A serological study with a range of monoclonal antibodies to FBNYV showed that an MDV isolate from Japan shared four of the six identified FBNYV epitopes with typical FBNYV isolates from Syria and Egypt, suggesting that FBNYV and MDV might be strains of the same virus (Franz et al., 1996). Although FBNYV and MDV are most similar in one of their non-Rep proteins (90.8% identity) and have two apparently closely related Rep proteins (88.3 and 89.8% identical), their coat protein amino acid sequences were 83.6% identical (Sano et al., 1998), and they differed in their four other non-Rep proteins by > 23%. Based on the available information and on the criteria proposed for the classification of strains and species of geminiviruses (Rybicki, 1994; Padidam et al., 1995), FBNYV and MDV are regarded as two distinct virus species.

The four Rep and six non-Rep components described so far may represent the complete FBNYV genome. However, conclusive evidence for the completeness of the FBNYV genome will only be obtained when various combinations or all ten cloned components are used in infectivity tests to reproduce a disease whose causal agent is indistinguishable in all its biological properties from field isolates of FBNYV.

We thank Yoshitaka Sano (Kyoto, Japan) for sharing sequence information on the MDV genome and for exchanging manuscripts with us prior to publication. We are indebted to Françoise de Kouchkovsky for her excellent technical assistance. This work was supported by a grant from the European Commission under the INCO-DC programme (contract no. ERBIC18-CT96-0121).

References


Wu, R. Y., You, L. R. & Soong, T. S. (1994). Nucleotide sequences of


Received 24 April 1998; Accepted 10 August 1998