particles (100 $\mu g\ ml^{-1}$) still gives a positive to negative ratio of 7, which is well above the limit.

The results show that the TR-FIA using Eu-labelled monoclonal antibodies is a simple one-step procedure for the detection of HBsAg which exceeds RIA in sensitivity, at the same time avoiding all the disadvantages of the use of isotopes. The assay could simplify the large-scale screening of blood donors, and the sensitivity of the test should decrease the number of false negative results and thus reduce the number of post-transfusion HBV infections. A more sensitive technique will also assist HBsAg detection earlier during the incubation period and in the diagnosis of chronic HBV infection, in which low levels of HBsAg are present. In addition, the development of TR-FIA offers considerable potential for improvements in the whole field of immunoassay technology.

Received 10 May; accepted 17 November 1982.

1. Soini, E. & Hemmilä, I. *Clin. Chem.* **25**, 353–361 (1979).
2. Yamada, S., Miyoshi, F., Kano, K. & Ogawa, T. *Analyt. chim. Acta* **127**, 195–198 (1981).
3. Tiollais, P., Charnay, P. & Vyas, G. N. *Science* **213**, 406–411 (1981).
4. Wright, R. *Clin. Gastroenterol.* **9**, 97–115 (1980).
5. Burrell, C. J. *Clin. Gastroenterol.* **9**, 47–63 (1980).
6. Koistinen, V. U. *J. Virol.* **35**, 20–23 (1980).

Nucleotide sequence of cassava latent virus DNA

John Stanley & Michael R. Gay

Department of Virus Research, John Innes Institute, Colney Lane, Norwich NR4 7UH, UK

Only two groups of plant viruses, the caulimoviruses[1,2] and the geminiviruses[3], are known to contain a genome of DNA. Unlike that of the caulimoviruses, the genome of the geminiviruses is composed of single-stranded, covalently-closed circles of DNA. There is evidence that the geminiviruses, specifically bean golden mosaic virus[4] and tomato golden mosaic virus[5], have a genome composed of two similar-sized circles of DNA, and the molecular cloning of both components from tomato golden mosaic virus has recently been reported[6]. No information is available, however, concerning the protein coding capacity of the genome, the possible modes of RNA transcription and DNA replication and the assembly of mature components into the characteristic geminate particles. We report here on the derivation of the nucleotide sequence of the geminivirus, cassava latent virus (CLV), as a fundamental step towards the investigation of the mode of replication of this group of viruses. Results show that CLV DNA comprises two similar-sized molecular species with a common region encompassing almost 200 nucleotides. The organization of the genome is discussed.

With the presently available rapid sequencing techniques, it was considered unnecessary to derive a detailed restriction enzyme map of the genome before initiating the investigation of the sequence. Furthermore, it was assumed that the genome of CLV would mirror those of other geminiviruses so far examined and have a bipartite genome[4,5]. This being the case, exhaustive sequencing of the viral DNA should lead to the accommodation of all derived data into two circular structures. This was found to be the case.

The strategy adopted for the elucidation of the sequence is outlined in Fig. 1 legend. As the CLV genome is single stranded, it was necessary to generate a second, complementary strand before cloning the material into M13 vectors. The sequence GGAATTCC, found within part of the genome, enabled us to use a double-stranded octanucleotide of the same sequence to prime DNA polymerase-directed second-strand synthesis on the viral DNA template. Although this octanucleotide sequence appears only once, at position 1,867 in DNA 1, cloned material from both DNAs 1 and 2 was generated by this method. Presumably, the sequence GGAATTCA at position 1,527 in DNA 2 bound the primer sufficiently well to allow primer extension even in the presence of the terminal mismatch. By the combination of the various approaches described in Fig. 1 legend, 90% and 80% of the sequences of DNAs 1 and 2, respectively, were determined in both orientations.

The sequences of the CLV DNAs are shown in Fig. 1, and indicate that the viral DNA does, in fact, contain two sequences of similar length, DNAs 1 and 2, containing 2,779 and 2,724 nucleotides, respectively (the larger of the two DNAs was arbitrarily designated DNA 1). This size difference of less than 2% will account for the failure to resolve the molecules by gel electrophoresis[7].

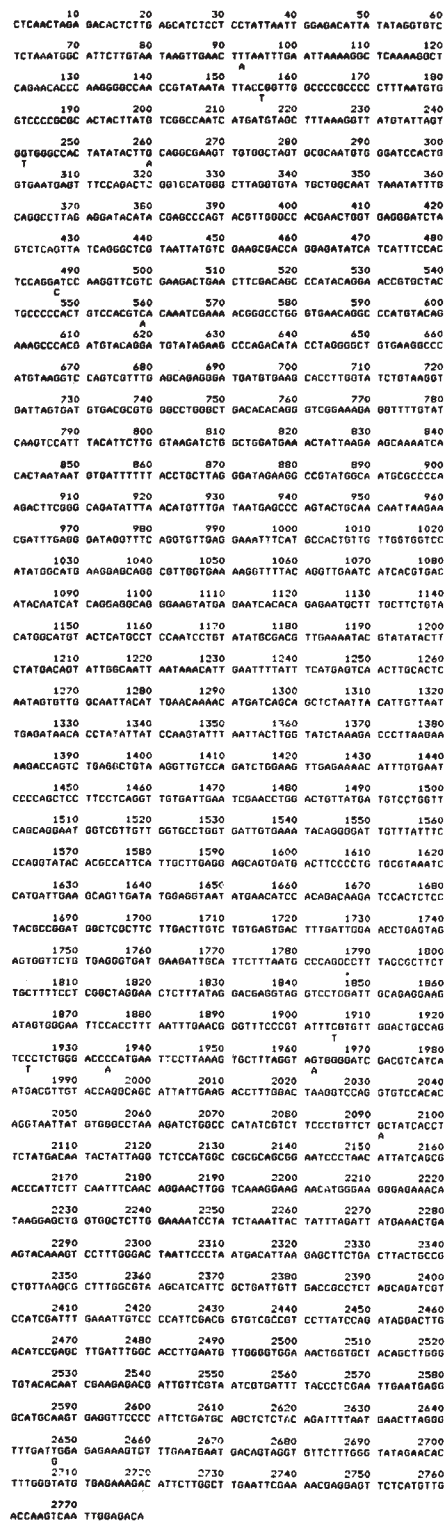
Although the entire sequence was established from a single preparation of virus, a number of nucleotide variations were found within the fragments that were cloned in M13; these are shown in Fig. 1 as an alternative nucleotide below the sequence. As the virus preparation was not derived from a local lesion isolate, it might be expected to contain a population of slightly varying molecules.

There is a region of almost 200 nucleotides common to both DNAs, the 5′ extremities of these regions having been arbitrarily designated as nucleotide 1 in Fig. 1. Figure 2 compares the homologous sequences. In addition to small repeated or near-repeated sequences, the region shows one potentially very stable hairpin structure between nucleotides 133 and 165, the loop of which is composed almost entirely of A and T residues. The purpose of a region common to both molecules is unknown but it may be speculated that it serves as a recognition site during the process of virus multiplication, possibly containing transcription promoters (see below) and the origin of DNA replication. This is the only region showing extensive homology between the two DNA molecules.

To investigate the potential coding capacity of the genome, the sequences were screened in all three reading frames for open regions. The results of this, both for the viral DNA sequences as presented in Fig. 1 and for their complement, are summarized in Fig. 3. Based on these data, and bearing in mind that no information is available concerning possible splicing and polypeptide maturation events, a tentative proposal for a number of virus-specific proteins is suggested in Fig. 4. It was assumed that the first ATG triplet in each open region would initiate protein synthesis and only those open regions with a potential coding capacity of molecular weight (M_r) >10,000 are included in the figure. Figure 4 shows that, when read in the virion DNA sense, the sequence of DNA 1 can code for overlapping proteins of M_rs 30,100 and 12,400. However, when read in the opposite sense, ~85% of the sequence may be involved in protein coding to give products of M_rs 40,200, 27,000, 15,600, 15,500, 15,100, 13,500 and 10,600. All sequences involved lie outside the region common to the two DNAs.

The situation for DNA 2 seems to be slightly less complex. DNA 2 is capable of coding for a protein of M_r 29,200 and a smaller protein of M_r 13,500 when read in the virion DNA sense, and a protein of M_r 33,600, overlapping with the putative M_r 13,500 protein, when read in the opposite sense. The open reading regions of the two large putative proteins do not overlap, being separated by a short A+T-rich sequence. Together, the two major putative proteins utilize ~60% of the available sequence and, as seen for DNA 1, all potential coding regions lie outside the region common to both viral DNAs.

The data concerning the open reading regions of DNA 2 strongly suggest that transcription of both the viral DNA and its complement takes place, implying the production of a double-stranded intermediate. A double-stranded viral DNA structure, isolated from plant material infected with bean golden mosaic virus, has been shown to be infectious, suggesting that it is competent for RNA transcription[8]. Additional evidence for the bidirectional nature of transcription of the genome is provided by an analysis of the amino acid content of each putative protein. The amino acid composition of the CLV coat protein (M. Short, personal communication) shows a close correlation with that of the putative 30,100 M_r protein of DNA



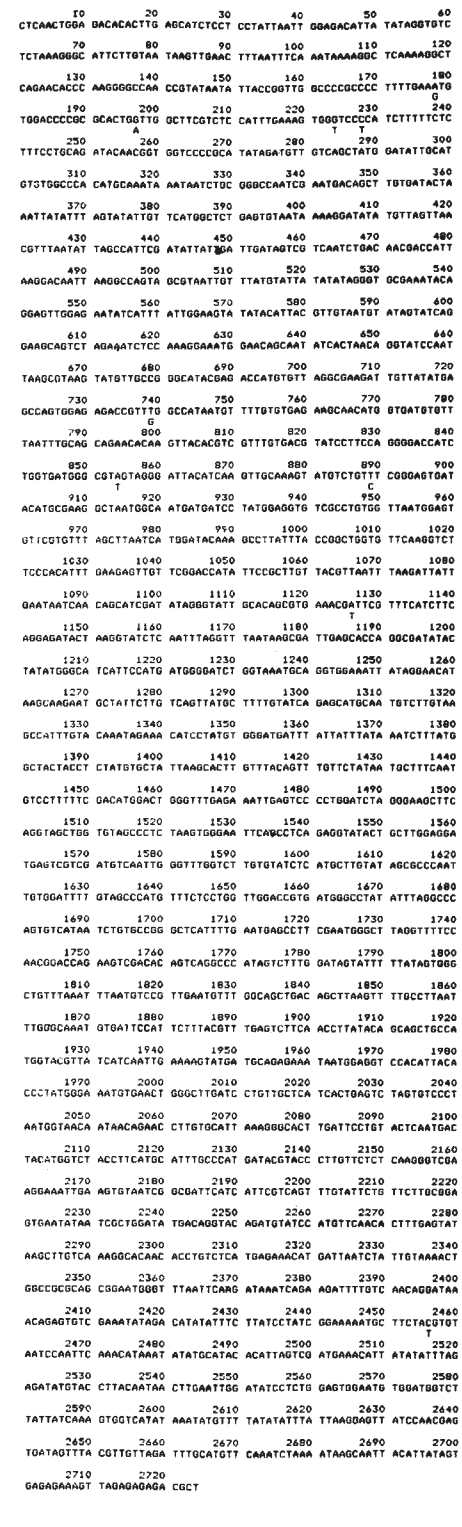


Fig. 1 The nucleotide sequence of DNA 1 (left) and DNA 2 (right). CLV (West Kenyan isolate 844) was propagated and the DNA isolated from purified virions as described previously[12]. The initial strategy adopted to accumulate sequence data was to digest the viral genome with restriction enzymes *Hae*III and *Hha*I, known to cut single-stranded DNA[13,14]. The products of digestion were terminally labelled at either the 5′ (ref. 15) or 3′ (ref. 16) termini and sequences established using the chemical degradation technique of Maxam and Gilbert[15]. The identification of the octanucleotide GGAATTCC within the derived sequences allowed the use of double-stranded DNA of the same sequence (molecular linker recognized by *Eco*RI) to prime DNA polymerase-directed second-strand synthesis using the viral DNA as the template, essentially as described by Hong[17]. The resulting double-stranded DNA was digested with restriction enzymes *Bam*HI, *Bgl*II, *Eco*RI, *Hin*dIII, *Mbo*I, *Msp*I, *Pst*I or *Taq*I and fragments cloned in the appropriate M13 vector[18]. After selection of recombinants and isolation of M13 DNA, sequences of the inserts were established using the dideoxy-termination procedure[19], using as a primer a synthetic deoxyribonucleotide 17 nucleotides in length[20]. At positions where sequence data remained ambiguous, the replicative form of appropriate M13 clones was isolated, and the inserted DNA excised with a suitable restriction enzyme and purified by sucrose gradient centrifugation. Sequence analysis of enzyme digests of the inserts was by the method of Maxam and Gilbert[15].

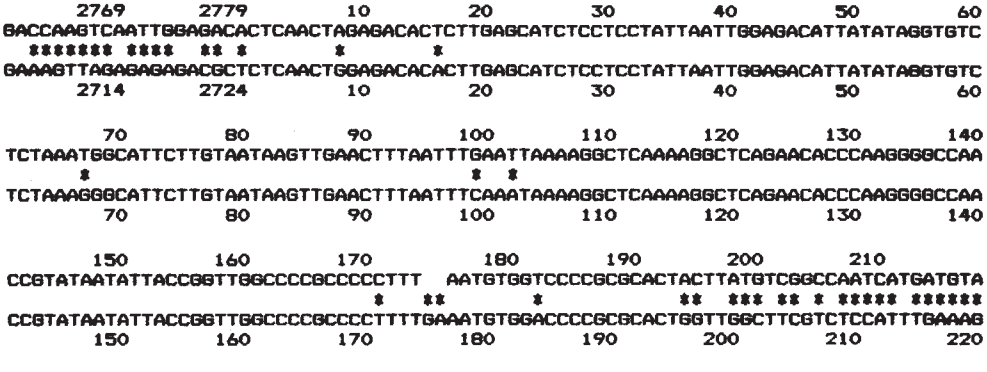


Fig. 2 Regions of the CLV genome containing sequences common to both strands. Differences between DNA 1 (upper sequence) and DNA 2 are shown by asterisks. A gap corresponding to two nucleotides has been inserted into DNA 1 after nucleotide 174 for purposes of comparison. At positions 92 and 155 in DNA 1 and position 178 in DNA 2, where variability is observed in the sequence (see Fig. 1), the homologous nucleotide is given.

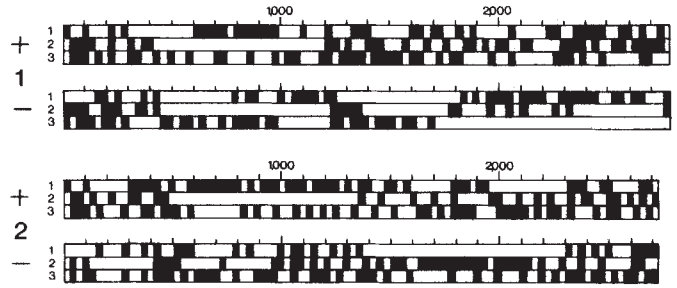


Fig. 3 Open reading regions contained within CLV DNAs 1 and 2 in both the virion DNA sense (+) and its complement (−). The nucleotide numbering in each case is directly related to that of Fig. 1, open reading frames 1, 2 and 3 beginning at nucleotide positions 1, 2 and 3, respectively. Each reading frame was divided into groups of 10 nucleotide triplets and the appropriate regions above shaded at positions where nonsense codons appeared.

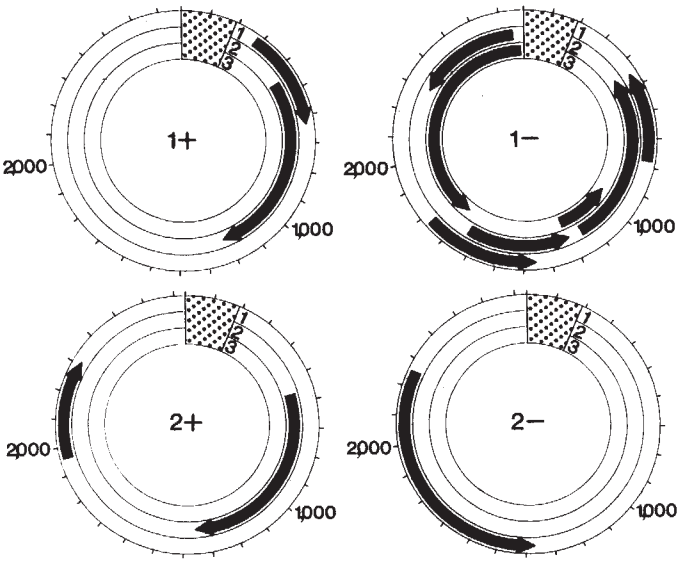


Fig. 4 Potential protein coding regions within CLV DNAs 1 and 2 in both the virion DNA sense (+) and its complement (−). Assuming that the first in-phase ATG triplet of each open reading frame of Fig. 3 initiates protein synthesis, those regions with a coding capacity of $M_r \geqslant 10{,}000$ are given. The stippled regions correspond to the homologous regions as shown in Fig. 2.

1. The compositions of all other putative products discussed above fail to show such a correlation. Although a large part of the sequence complementary to virion DNA 1 may be implicated in coding for proteins (as shown in Fig. 4), this observation suggests that the DNA in the virion sense is transcribed at least in part. This being the case, the region common to both DNAs 1 and 2 is well situated to contain the necessary signals for the initiation of bidirectional transcription of a putative double-stranded DNA intermediate. The determination of which of the above overlapping, potential coding regions are recognized and which have arisen fortuitously must wait until information becomes available concerning transcription and translation products of the virus. Comparative studies with other geminivirus sequences should also help to elucidate the overall transcription and translation strategy of the group.

It will be of interest to see if either of the DNA strands can be independently replicated as has been demonstrated for some plant viruses containing a bipartite RNA genome[9–11]. The function of such a multi-component genome is unknown but it has been suggested that it serves to separate early (for example, DNA replication) and late (for example, capsid protein, virion assembly) genes and in doing so, helps regulate their expression. The bipartite nature of the geminivirus genome may greatly aid in the identification of not only gene products but also possible control elements within the genome.

We thank Drs J. C. Sequeira, D. J. Robinson and B. D. Harrison for providing CLV DNA, Dr J. W. Davies for critical reading of the manuscript and Dr G. Lomonossoff for interesting discussion throughout. This work was carried out under the MAFF licence number PHF 49/54.

Received 11 October; accepted 19 November 1982.

1. Shepherd, R. J. *A. Rev. Pl. Physiol.* **30**, 405–423 (1979).
2. Hull, R. in *Nucleic Acids in Plants* (eds Hall, T. C. & Davies, J. W.) 3–29 (CRC Press, Boca Raton, Florida, 1979).
3. Goodman, R. M. *Handbook of Plant Virus Infections and Comparative Diagnosis* (ed. Kurstak, E.) 879–910 (Elsevier, Amsterdam, 1981).
4. Haber, S., Ikegami, M., Bajet, N. B. & Goodman, R. M. *Nature* **289**, 324–326 (1981).
5. Hamilton, W. D. O., Bisaro, D. M. & Buck, K. W. *Nucleic Acids Res.* **10**, 4901–4912 (1982).
6. Bisaro, D. M., Hamilton, W. D. O., Coutts, R. H. A. & Buck, K. W. *Nucleic Acids Res.* **10**, 4913–4922 (1982).
7. Harrison, B. D. *et al. Nature* **270**, 760–762 (1977).
8. Ikegami, M., Haber, S. & Goodman, R. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4102–4106 (1981).
9. Sanger, H. L. *J. Virol.* **3**, 304–312 (1969).
10. Goldbach, R., Rezelman, G. & van Kammen, A. *Nature* **286**, 297–300 (1980).
11. Robinson, D. J., Barker, H., Harrison, B. D. & Mayo, M. A. *J. gen. Virol.* **51**, 317–326 (1980).
12. Sequeira, J. C. & Harrison, B. D. *Ann. appl. Biol.* **101**, 33–42 (1982).
13. Horiuchi, K. & Zinder, N. D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2555–2558 (1975).
14. Blakesley, R. W. & Wells, R. D. *Nature* **257**, 421–422 (1975).
15. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
16. Tu, C-P. D. & Cohen, S. N. *Gene* **10**, 177–183 (1980).
17. Hong, G. F. *Biosci. Rep.* **1**, 243–252 (1981).
18. Messing, J., Crea, R. & Seeburg, P. H. *Nucleic Acids Res.* **9**, 309–321 (1981).
19. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. *J. molec. Biol.* **143**, 161–178 (1980).
20. Duckworth, M. L. *et al. Nucleic Acids Res.* **9**, 1691–1706 (1981).

Homology between human bladder carcinoma oncogene product and mitochondrial ATP-synthase

Nicholas J. Gay & John E. Walker

Laboratory of Molecular Biology, MRC Centre, Hills Road, Cambridge CB2 2QH, UK

More than 10 different dominant transforming genes (oncogenes) have been identified in human tumours[1,2]. A human bladder carcinoma oncogene, closely related in sequence to retroviral transforming genes, is split into four exons; the first encodes the N-terminal 37 residues of p21, a protein of unknown function[3,4]. The oncogene is activated by a single point mutation (guanine to thymine) resulting in the change glycine to valine at position 12 of p21 (refs 3, 4). We report here that the amino acid sequence surrounding this residue is highly homologous to the β-subunit of mitochondrial and bacterial ATP-synthase in the region of the polypeptide that is believed to contribute to nucleotide binding[5]. Thus, p21 may form part of an enzyme that uses purine nucleotides in catalysis. This is consistent with the finding that an equivalent murine oncogene product binds GTP[6,7].

The alteration of a single nucleotide in a normal cellular gene of a bladder cell activates an oncogene. In bladder carcinoma cells this mutation replaces glycine with valine at residue 12 of a normal cellular protein, p21, resulting in transformation of the cell[3,4]. The transforming *ras* genes of Harvey and Kirsten murine sarcoma viruses (Ha-MuSV and Ki-MuSV, respectively) also encode p21 proteins and are closely related to each other for most of their sequences[8,9] and identical to the bladder cell p21 over at least the first 37 amino acids, except at residue 12; at this position the Harvey virus has arginine and the Kirsten virus serine[3,4].

Little is known about p21 except that it has a molecular weight of 21,000 and it appears to be bound to the inner surfaces of plasma membranes[10,11]. The murine homologue transformed