J. gen. Virol. (1987), **68**, 2245–2247. *Printed in Great Britain*

Key words: *YFV/genetic stability/structural proteins*

High Genetic Stability of the Region Coding for the Structural Proteins of Yellow Fever Virus Strain 17D

By PHILIPPE DESPRÈS, ANNIE CAHOUR,* ALFRED DUPUY,[1] VINCENT DEUBEL,[1]† MICHÈLE BOULOY, JEAN PIERRE DIGOUTTE[1] AND MARC GIRARD

Unité de Virologie Moléculaire, UA CNRS 545, Institut Pasteur, 75724 Paris Cedex 15, France and [1]*Institut Pasteur, B.P. 220, Dakar, Sénégal*

(*Accepted 28 April 1987*)

SUMMARY

The genome of the Pasteur 17D-204 vaccine strain of yellow fever virus has been cloned into pBR327. The inserts of recombinant plasmids were analysed by restriction cleavage pattern and compared with that of the genome of another substrain previously cloned and sequenced. Ten of the overlapping inserts were found to contain the sequence of the complete genome. We have sequenced approximately 3680 bases of the 5′ region which codes for the C, M and E structural proteins and the NS1 non-structural protein. This sequence is the same as that reported previously, indicating a remarkable stability of these two vaccine substrains.

Recently, Grange *et al.* (1985) have cloned and sequenced the 565 bases of the 3′ terminal end of the genome of the Pasteur 17D-204 strain of yellow fever virus (YFV). Independently, Rice *et al.* (1985) published the sequence of the entire RNA genome of another substrain of the 17D YFV. Comparison of the sequences indicated that the 3′ end of the genome of the two substrains was highly conserved. We have pursued the cloning of the Pasteur strain genome. In the present communication, we report the cloning procedures and the sequence of the 5′ region of the genome which codes for the structural viral proteins C, prM, M and E, and for the non-structural protein NS1.

Preparation and purification of the virus and extraction of the RNA have been described (Grange *et al.*, 1985). Briefly, the avian leukosis virus-free Pasteur substrain IP/F2 (235th passage) was purified twice by endpoint dilution in Vero cells, then propagated once in Vero cells and once in SW13 cells. The virus was grown in SW13 cells and purified by sucrose gradient centrifugation.

Prior to reverse transcription viral RNA was denatured with 10 mM-methylmercuric hydroxide which was then complexed with 70 mM-2-mercaptoethanol in the presence of the primer. The synthesis of the cDNA was primed by a custom-made synthetic oligodeoxynucleotide complementary to nucleotides 2458 to 2473 on the viral genome (Rice *et al.*, 1985). Double-stranded DNA was synthesized using *Escherichia coli* DNA polymerase I in the presence of RNase H and T4 DNA polymerase, according to the procedure of Okayama & Berg (1982), modified by Gubler & Hoffman (1983). The large size double-stranded cDNA molecules isolated by chromatography on a Sephacryl S-1000 column were tailed with oligo(dC), hybridized to oligo(dG)-tailed pBR327, cut at the *Pst*I site and used to transform competent *E. coli* HB101 cells. Colonies selected for their tetracycline resistance and ampicillin sensitivity were hybridized *in situ* with a probe prepared by labelling purified YFV RNA with [^{32}P]ATP in the presence of polynucleotide kinase. Seventy recombinant plasmids were analysed for size and restriction cleavage pattern. Ten of the overlapping inserts, ranging in size from 0·7 to 2·7 kbp

† Present address: Centers for Disease Control, Fort Collins, Colorado 80522, U.S.A.

0000-7722 © 1987 SGM

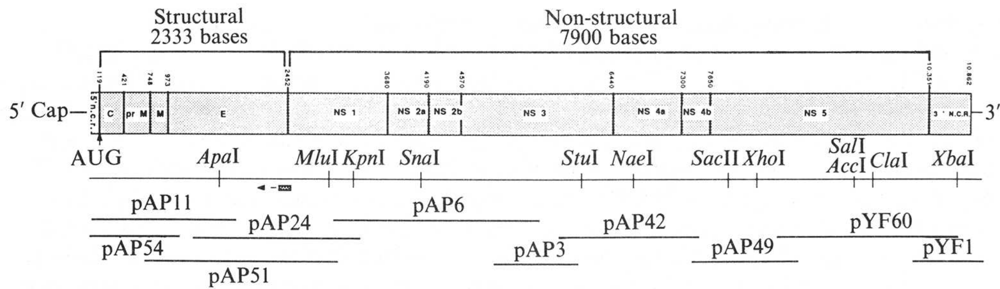

Fig. 1. Mapping of the IP/F2 cDNA inserts on the genome of the YFV 17D strain. The organization of the YFV genome is that published by Rice *et al.* (1985). Unique endonuclease restriction sites are indicated. The position of the oligonucleotide used as a primer for reverse transcription is represented by a box. Clones pYF1 and pYF60 were obtained previously using different cloning procedures (Grange *et al.*, 1985). N.C.R., non-coding region.

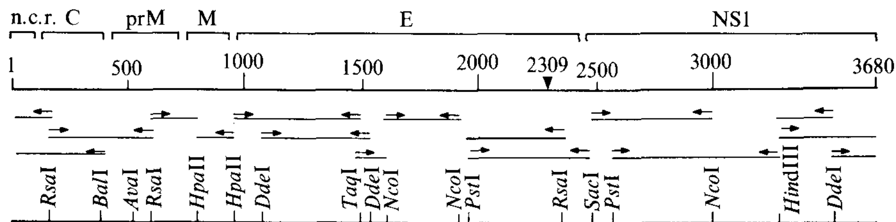

Fig. 2. Strategy for sequencing. Restriction fragments isolated from clones pAP6, 11, 24, 51 and 54 were ligated into M13 mp18 replicative form DNA cut at the *Sma*I site. Arrows indicate the orientation for sequencing. The black triangle at position 2309 indicates the U to C substitution in clone pAP51.

are shown in Fig. 1. Surprisingly, such inserts were found to originate from all regions of the viral genome. This was unexpected as the sequence of the oligodeoxynucleotide used as a primer for cDNA synthesis was located within the NS1 protein gene (Rice *et al.*, 1985) so that the synthesis of the cDNA should have occurred only upstream from the NS1 protein gene. The result obtained could most likely be explained by non-specific hybridization of the primer with genome regions of limited sequence homology, due to incubation of the primer and RNA at low temperature. Similar observations have been reported by Auperin *et al.* (1986).

We have sequenced the 5′ region of the genome which spans the 5′ non-coding region and the region coding for the structural proteins. Sequencing was also extended to the contiguous region coding for the non-structural protein NS1. Appropriate fragments from plasmids pAP11, 54, 24, 51 and 6 (Fig. 1) were isolated as indicated in Fig. 2, treated with the large Klenow fragment of DNA polymerase I and ligated to M13 mp18 replicative form DNA cleaved by *Sma*I. Sequence determination was performed by the dideoxynucleotide termination method (Sanger *et al.*, 1977). The data obtained showed that the nucleotide sequence of the stretch of 3680 nucleotides starting at nucleotide 12 is the same as that reported by Rice *et al.* (1985).

Preliminary sequence data and fine restriction mapping of the rest of the genome suggest that the identity extends to the whole genome.

Plasmids pAP54 and pAP11 contain inserts of the 5′ outermost region but are missing the first 12 and 23 nucleotides, respectively. Only one difference from the sequence reported by Rice *et al.* (1985), a U to C substitution at position 2309, was found in plasmid pAP51. However, this modification which changed a Phe residue to a Leu residue in the sequence of the E protein was not present in the insert of plasmid pAP24 which contains the same region of the genome, nor in the viral RNA as checked by direct sequencing of purified virus RNA using the oligonucleotide primer extension method and reverse transcriptase. This suggests that an error was introduced during the reverse transcription or the cloning of pAP51. Alternatively, it could be that the viral

population was heterogeneous and contained a minor amount of a variant which was not detected through the direct sequencing of RNA but was serendipitously cloned in pAP51.

The 17D strain sequenced by Rice *et al.* (1985) and provided by the American Type Culture Collection represents passage 234 *in vitro* and, although not stated, is probably the 17D 204 substrain of the vaccine produced in the U.S.A. The IP/F2 17D strain cloned in our laboratory also represents the 17D 204 substrain. The passage history of these two vaccines (WHO, 1979; Monath *et al.*, 1983) indicates that they both derive from a common seed, Colombia 88, which has undergone several additional passages in eggs. At passage 232, divergence occurred: the American strain has been carried to egg passage 234 whereas the French vaccine has undergone an additional passage. Thus, only the last two or three passages respectively have been carried out independently in separate laboratories. Our data show an extraordinarily high degree of homology between the two vaccine substrains. This is in agreement with the observation reported by Monath *et al.* (1983) who analysed the genomes of several 17D vaccines by T1-resistant oligonucleotide mapping and showed 100% homology between the French and American strains as well as between most of the currently manufactured 17D-204 vaccines, except those produced in South Africa. This stability is remarkable in view of the known variability of viral RNA genomes (Holland *et al.*, 1982), but suggests the existence of a strong selective pressure operating during the two or three passages in eggs when these viruses were propagated independently. In addition, it is known that some YFV vaccines are composed of a heterogeneous population of variants differing in plaque morphology and, in some cases, their degree of neurovirulence for mice (Liprandi, 1981). The question is thus raised of whether the two initial virus populations were actually homogeneous or whether the cloning procedures used in both laboratories in a totally independent fashion led to selection of the same variant.

It would also be interesting to know whether this sequence conservation is maintained in the vaccines which have a different passage history such as those produced in the U.K. (or in India) and derived also from the 17D 204 substrain but from a different seed of Colombia 88, or those produced in Brazil and derived from the 17DD substrain.

We thank A. Diop for skilful technical assistance and P. Vialat for his help in virus preparation. This work was supported in part by DRET research contract no. 85 34 120.

REFERENCES

AUPERIN, D., SASSO, D. R. & McCORMICK, J. B. (1986). Nucleotide sequence of the glycoprotein gene and intergenic region of the Lassa virus S genome RNA. *Virology* **154**, 155–167.

GRANGE, T., BOULOY, M. & GIRARD, M. (1985). Stable secondary structures at the 3′-end of the genome of yellow fever virus (17D vaccine strain). *FEBS Letters* **188**, 159–163.

GUBLER, U. & HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269.

HOLLAND, J., SPINDLER, K., HORODYSKI, F., GRABAU, E., NICHOL, S. & VAN DE POL, S. (1982). Rapid evolution of RNA genomes. *Science* **215**, 1577–1585.

LIPRANDI, F. (1981). Isolation of plaque variants differing in virulence from the 17D strain of yellow fever virus. *Journal of General Virology* **56**, 363–370.

MONATH, T. P., KINNEY, R. M., SCHLESINGER, J. J., BRANDRISS, M. W. & BRÈS, P. (1983). Ontogeny of yellow fever 17D vaccine: RNA oligonucleotide fingerprint and monoclonal antibody analyses of vaccines produced world-wide. *Journal of General Virology* **64**, 627–637.

OKAYAMA, H. & BERG, P. (1982). High efficiency cloning of full-length cDNA *Molecular and Cellular Biology* **2**, 161–170.

RICE, C. M., LENCHES, E. M., EDDY, S. R., SHIN, S. J., SHEETS, R. I. & STRAUSS, J. H. (1985). Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. *Science* **229**, 726–733.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain termination inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

WORLD HEALTH ORGANIZATION (1979). *Procedure for WHO Approval of Yellow Fever Vaccine for Issue of International Vaccination Certificates.* Report of a WHO Working Group. Geneva 22–24 May 1979.

(*Received 3 March 1987*)