5522–5523 *Nucleic Acids Research, 1993, Vol. 21, No. 23*

Nucleotide sequence variations surrounding the standard recombination junction of pseudorabies viruses

Andrew Cheung and Roger Maes[1]
Virology Swine Research Unit, National Animal Disease Center, USDA, Agricultural Research Service, Ames, IA 50010 and [1]College of Veterinary Medicine, Michigan State University, East Lansing, MI 48824, USA

Received August 25, 1993; Accepted October 7, 1993

GenBank accession nos M81222 and L24958

The inverted repeats, internal repeat (IR) and terminal repeat (TR), of pseudorabies virus (PRV) are responsible for the inversion of the unique short (US) region that resulted in two isomeric forms of the PRV genome (1) and constitute part of the cleavage-encapsulation signal during maturation of the viral genome (2). We have obtained the complete *Bam*HI-J DNA nucleotide (nt) sequence (4968 bases) of the virulent Indiana Funkhauser (InFh) strain and a partial sequence of the virulent Becker strain (a dog isolate) (3) that spans the standard recombination junction between the unique long (UL) and IR sequences. Other sequences available for comparison included a laboratory strain (Ka), a field strain (PRV-USA7) and an attenuated vaccine strain (PRV-Norden) (Figure 1) (4, 5). Two versions of the Ka sequences have been reported and they differ in many locations. The sequence by Vleck *et al.* (4) was designated Kv and that by DeMarchi *et al.* (5) was designated Kd. Nucleotide 1 to 2451 (InFh and Kv): This portion is the left end of *Bam*HI-J. other than a deletion of 67 bases at nt 987–1054, there are only occasional base changes, insertions and deletions. Nucleotide 2542 to 3280 (InFh, Becker, Kv and Kd): This sequence represents the right end of the UL region. As indicated, there are base changes, small deletions and insertions between the two Ka sequences. Both the Kv and Kd sequences showed an insertion of an A residue in the TATA sequence present at nt 2947 and become TAATA. A 30-base tandem repeat is noted at nt 3079. This 30-nt sequence occurs twice in the InFh and Ka strains and 3 times in the Becker strain. The nucleotides surrounding the UL–IR junction also vary among different strains. The last dinucleotide of UL in InFh is CC while the last dinucleotide for Becker and Ka are AT. Similarly, the beginning IR nucleotide of Norden are different from those of the other strains compared. Nucleotide 3281 to 3706 (InFh, Becker, Kv, Kd, Norden and USA7): This portion represents the left end of the IR region abutting the UL sequence and is best characterized by the presence of a 22 nt direct repeat sequence designated here as DR22 (TAACCCCTCCCCCTAATCTGCA) or R2+R3 by DeMarchi *et al.* (5) which occurs tandemly in various frequencies among different strains. In InFh, DR22 is repeated 8 times. However, the last 8 nt (AATCTGCA) of the third tandem was replaced by an 18 nt sequence, CATTTGCATGCATTTGCA. Interestingly, this 18-nt sequence is also present at nt 3415 to 3432. In Becker, DR22 is repeated 10 times. The same 8 nt in the second and third tandems were replaced by CATTTGCA and CATTTGCATGCATTTGCA, respectively. In Kv and Kd, DR22 is repeated 7 and 6 times, respectively. In both sequences, the latter portion of DR22 in the second tandem was replaced by a 50-nt sequence, CATTTGCATACATTTGCATAGCCCCGCCCCTTCATTTGCATAAATTAGCA. The last 48 nt of this 50-nt sequence is also present at nt 3353 to 3400. In Norden, DR22 is repeated 3 times. In USA7, DR22 is repeated 5 times, and the last 8 nt of the second tandem were replaced by the CATTTGCA sequence. Nucleotide 3707 to 4002 (InFh, Becker, Kv and Kd): There are some differences between the Kv and Kd sequences. Two unrelated repeat sequences were observed in this portion. A 23-nt sequence (CATTTGCATATGACCGCTTCCCG) at nucleotide 3716–3739 was duplicated, with a base change, at 3774–3797. A 63-base sequence at nucleotide 3812 to 3874 is present, as a single copy, in the InFh and Becker strains; but occurs as a tandem repeat in the Kv and Kd sequences. Nucleotide 4003 to 4986 (InFh, Becker and Kv): The sequence in this region is conserved among all three strains with only occasional base changes, deletions and insertions. The most significant difference is an additional 28-base sequence at nt 4509 in the Kv sequence and not in the InFh or Becker sequences. This addition occurs downstream of the 3′ splice junction (nt 4324) of the PRV large latency transcript and it is present in the coding sequences (6). As reported previously, the polyadenylation signal for the IE180 gene is located in the opposite orientation at nt 4778 to 4773, and the AATAAA sequence at 4792 to 4797 in InFh and Becker has mutated to become AACAAA in the Ka virus (7, 8). Our findings support the observations (5) that nucleotide specificity is not absolutely required at the actual site of cleavage or at the standard junction and the frequency of DR22 is not pivotal for the survival of the virus. However, it is clear that the most variable area in *Bam*HI-J is around DR22. In all cases examined, interruptions occurred within DR22 and at a relatively specific location. The AATCTGCA sequence was replaced by a neighboring sequence containing the CATTTGCA sequence. The high degree of variability and relatively precise interruptions within DR22 suggest that DR22 may play an essential role during virus replication. DR22 may have other functions, for the SP1 binding site present in DR22 was recognized by nuclear proteins in *in vitro* binding assays (4).

ACKNOWLEDGMENTS

We thank Ms L.Patrilla for technical assistance and Ms L.Hornung and Ms S.Ohlendorf for typing the manuscript.

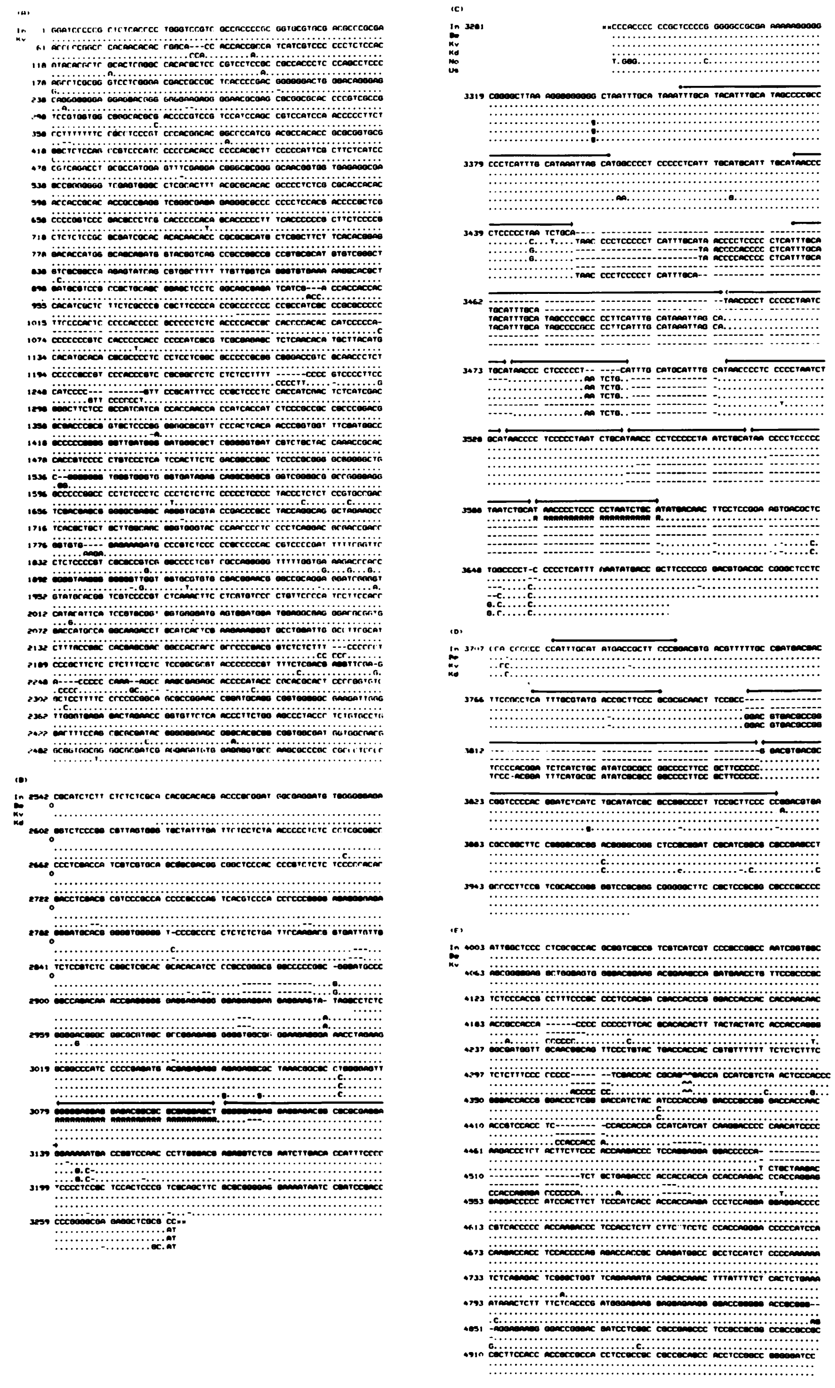

Figure 1. Pseudorabies virus *Bam*HI-J fragment DNA nucleotide sequence comparison. The Indiana–Funkhauser (InFh), Becker, Norden and USA7 viruses are denoted as In, Be, No and US, respectively. The two DNA sequences reported for the Ka strain are denoted as Kv and Kd. For convenience, the comparison was carried out in five sections (A to E), and the sequences analyzed are indicated at the beginning of each section. The InFh sequence, punctuated with numbers, is used as reference. Bases identical to that of InFh are indicated by dots under the InFh sequence, deletions are indicated by dashes, base changes are indicated by the appropriate nucleotides in capital letters and insertions are indicated by small letters. (RRRR) indicates tandem repeats in the Becker sequence. The symbols (* --- *) and (< --- >) denote sequence duplication at different locations and direct tandem repeats, respectively, within the same section. (xx) denotes the US-IR junction and (^^) denotes the 3′ spliced junction of the Pseudorabies Virus large latency transcript.

REFERENCES

1. Ben-Porat,T. and Kaplan,A.S. (1985) In Roizman,B. (ed.), *The Herpesviruses 3*. Plenum, New York, pp. 105–173.
2. Wu,C.A., *et al.* (1986) *J. Virol.* **59**, 318–327.
3. Platt,K.B., *et al.* (1979) *Arch. Virol.* **60**, 13–23.
4. Vlcek,C., *et al.* (1990) *Virology* **179**, 365–377.
5. DeMarchi,J.M., *et al.* (1990) *J. Virol.* **64**, 4968–4977.
6. Cheung,A.K. (1991) *J. Virol.* **65**, 5260–5271.
7. Cheung,A.K. (1988) *J. Virol.* **62**, 4763–4766.
8. Cheung,A.K., *et al.* (1990) *Virus Genes* **4**, 261–265.