Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

VIROLOGY 180, 448–452 (1991)

Comparison of the Genome Organization of Toro- and Coronaviruses: Evidence for Two Nonhomologous RNA Recombination Events during Berne Virus Evolution

ERIC J. SNIJDER[1] JOHAN A. DEN BOON, MARIAN C. HORZINEK,* AND WILLY J. M. SPAAN

*Department of Virology, Institute of Medical Microbiology, State University of Leiden, Postbus 320, 2300 AH Leiden, The Netherlands; and *Institute of Virology, Veterinary Faculty, State University of Utrecht, Yalelaan 1, 3584 CL Utrecht, The Netherlands*

Received June 21, 1990; accepted September 13, 1990

Recently, toroviruses and coronaviruses have been found to be ancestrally related by divergence of their polymerase and envelope proteins from common ancestors. In addition, their genome organization and expression strategy, which involves the synthesis of a 3′-coterminal nested set of mRNAs, are comparable. Nucleotide sequence analysis of the genome of the torovirus prototype, Berne virus (BEV), has now revealed the results of two independent nonhomologous RNA recombinations during torovirus evolution. Berne virus open reading frame (ORF) 4 encodes a protein with significant sequence similarity (30—35% identical residues) to a part of the hemagglutinin esterase proteins of coronaviruses and influenza virus C. The sequence of the C-terminal part of the predicted BEV polymerase ORF1a product contains 31–36% identical amino acids when compared with the sequence of a nonstructural 30/32K coronavirus protein. The cluster of coronaviruses which contains this nonstructural gene expresses it not as a part of their polymerase, but by synthesizing an additional subgenomic mRNA.

In 1982 and 1983 the characterization of two morphologically similar viruses in fecal material from cattle (Breda virus; BRV;(*1*)) and horse (Berne virus; BEV; (*2*)) was reported. BRV and BEV are antigenically related to each other but no cross-reactivity with antisera against other animal viruses could be detected (*2*). Although the peplomers on the envelope of the new viruses resembled those of coronaviruses, the unique nucleocapsid morphology and morphogenesis of BRV (*3*) and BEV (*4, 5*) justified their classification as representatives of a new group of animal RNA viruses, the toroviruses (*6, 7*).

During the past four years, we have studied the replication strategy and genome organization of BEV, the prototype torovirus. The BEV genome consists of a single RNA molecule of positive polarity (*8*) with an estimated length of 25–30 kb (*8, 9*). In infected cells four 3′-coterminal mRNAs are transcribed from the 3′ end of the BEV genome (*8, 9*). *In vitro* translation of subgenomic BEV RNAs and nucleotide sequence analysis of BEV cDNA have revealed that the subgenomic RNA species are employed to express the structural genes ((*10, 11*) J. A. den Boon *et al.*, submitted).

Coronaviruses also express their genetic information from a 3′-coterminal nested set of mRNAs (reviewed in (*12*)). In addition, the corona- and toroviral genomes are of similar size (25–30 kb) and display the same basic gene order: 5′-polymerase–spike protein–membrane protein–nucleocapsid protein-3′ (Fig. 2A). Nucleotide sequence analysis of (parts of) the polymerase genes of the coronaviruses infectious bronchitis virus (IBV; (*13*)) and mouse hepatitis virus (MHV; (*14*)), and the torovirus BEV (*15*) has revealed that their predicted polymerase proteins contain several homologous domains (*15*). Furthermore, their polymerase genes consist of two large open reading frames (ORFs) of which the more downstream one (ORF1b) is expressed through ribsomal frameshifting (*14–16*).

Both similarities at the level of genome organization and the presence of homologous replicase protein sequences are taken as indications for common ancestry (*17, 18*). Moreover, two of the three structural BEV proteins are thought to be related to coronaviral structural proteins. The BEV peplomer (P) and the coronaviral spike (S) protein are post-translationally cleaved, N-glycosylated proteins of similar size. Although no linear protein sequence similarity was detected, their disposition in the viral membrane and their tertiary structure are predicted to be analogous; for both proteins dimerization, probably leading to the formation of the distinctive club-shaped spikes, has been demonstrated (*11*). The structural characteristics of the BEV envelope (E) and coronaviral membrane (M) proteins are also strikingly similar: they are triple-spanning 25K–30K membrane proteins with comparable membrane topologies (J. A. den Boon *et al.*, submitted). The small BEV nucleocapsid (N) protein (18.3K; (*10*)), on the other hand, seems to have little in common with its much larger (45K–50K; (*12*)) coronaviral counterpart.

The nucleotide sequence data reported in this paper have been submitted to the EMBL nucleotide sequence database and have been assigned the Accession Number X52375.

[1] To whom requests for reprints should be addressed.

0042-6822/91 $3.00

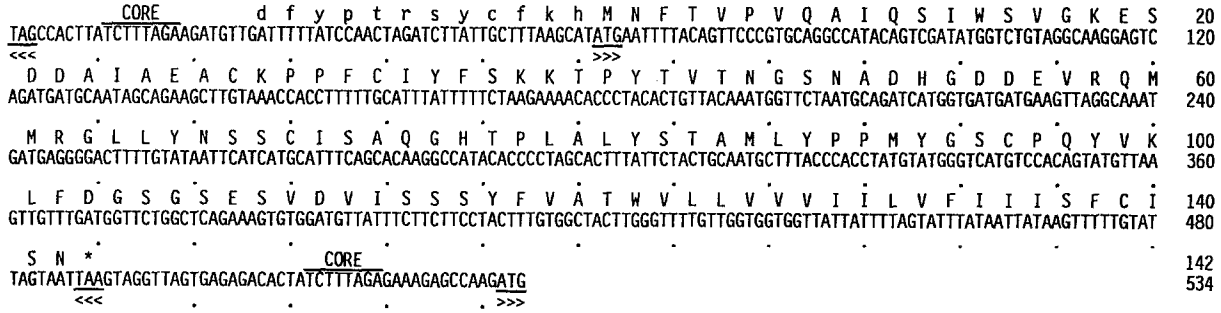


FIG. 1. cDNA sequence and translation of BEV ORF 4. The preparation, cloning, and sequence analysis of BEV cDNA was described previously (*9*). The termination codon (⋘) of the upstream E protein gene (ORF 3) and the initiation codon (⋙) of the downstream N protein gene (ORF 5) are also included in the figure. The conserved putative "core promoter" sequences for RNA 4 and RNA 5 transcription are indicated. The translation of the region upstream of the ORF 4 initiation codon (used in Fig. 2B) is shown in lowercase letters. The nucleotide sequence data in this figure have been submitted to the EMBL nucleotide sequence database and have been assigned the Accession Number X52375.

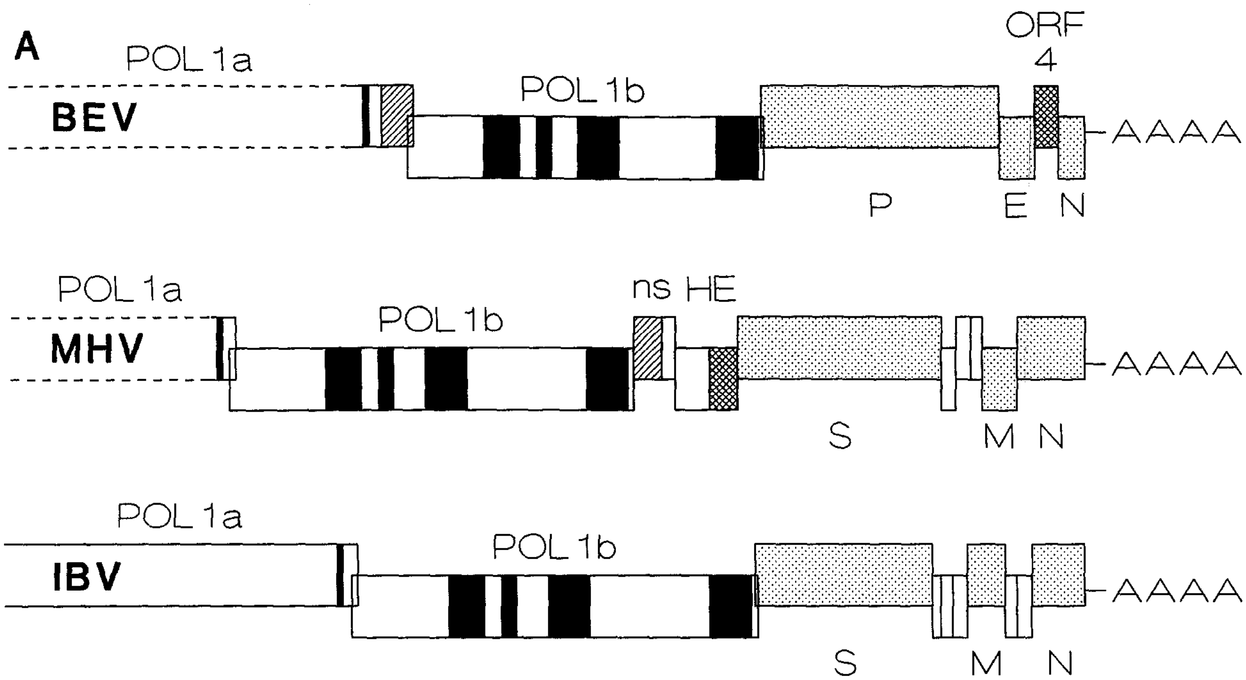
In addition to divergence from a common ancestor, RNA recombination is considered an important factor in RNA virus evolution (*17, 18*). Homologous recombination between highly similar RNA sequences has been found to occur during the multiplication of a number of plant and animal RNA viruses (19–26). Nonhomologous RNA recombination events (i.e., the incorporation of heterologous RNA sequences) have been advocated, e.g., to explain the presence of tRNA sequences in alphaviral defective interfering RNAs (*27*). Undisputed examples of nonhomologous recombination in infectious (nonretroviral) RNA virus genomes have been described only recently (*28–30*). One of these recombinations (*28*) involves the gene which encodes the hemagglutinin esterase (HE) protein of influenza virus C (IVC). Proteins with remarkable sequence similarity to the IVC hemagglutinin HE1 subunit are encoded by genes of murine (MHV) and bovine (BCV) coronaviruses (*28, 31, 32*). Because such a gene is lacking in the genomes of coronaviruses from other antigenic clusters (e.g., IBV; Fig. 2A), a heterologous recombination event involving and IVC-like virus and an ancestral coronavirus was postulated to explain the presence of an HE gene in MHV and BCV (*28*).

In this report we present evidence for two independent nonhomologous RNA recombination events during BEV evolution. It is remarkable that, in addition to the evidence for common ancestry presented above, also these recombinations associate toroviruses with coronaviruses.

Figure 1 shows the previously unreported nucleotide sequence of BEV ORF 4, which is located between the E and N protein genes ((9); see also Fig. 2A). The protein encoded by this ORF (Fig. 1) shows sequence similarity to the C-terminal parts of the coronaviral HE protein and the IVC HE1 subunit (Fig. 2B). However, the ORF 4 product consists of only 142 amino acids (aa), whereas both the coronaviral HE protein and the IVC HE1 subunit are more than 400 aa in length. The sequence of the ORF 4 product shares 30–35% identical amino acid residues with both the IVC and the MHV/BCV HE sequences. The predicted BEV product contains a hydrophobic C-terminus, but lacks the catalytic center of the acetylesterase which is located in the N-terminal part of the protein (*33*). Five cysteine residues in the C-terminus of the HE protein which are conserved between IVC and coronaviruses (*34*) are also found in the BEV sequence (Fig. 2B). Possibly, the 5′ part of BEV ORF 4 has been removed by a recent deletion event which did not inactivate the RNA 4 transcription initiation site (*9*). The first ORF 4 AUG codon would in this case not be the "original" translation initiation codon. This hypothesis is supported by the fact that the similarity with the IVC sequence and, to a lesser extent, the coronaviral sequence continues upstream of the present ORF 4 starting methionine residue (Fig. 2B).

The ORF 4 sequence similarities do not indicate a closer relationship to the homologous gene of either coronaviruses or IVC. The IVC HE1 subunit derives from cleavage of a HE1–HE2 precursor at an internal stretch of hydrophobic amino acid residues (*35*). Neither the BEV ORF 4 product nor the coronaviral HE protein contains sequences which are homologous to the IVC HE2 subunit. Instead they possess a very hydrophobic C-terminus which may represent the result of an adaptation of the hydrophobic HE2 N-terminus to become a membrane anchor. Though independent recombination events cannot be excluded, the presence of the same C-terminal adaptation in the proteins of both MHV/BCV and BEV lends some credibility to a recombination involving the ancestors of these viruses. Either a coronavirus or a torovirus may have been involved in the initial recombination with IVC.

The second nonhomologous recombination event in the BEV genome is quite similar to the case of ORF 4.



B

```
MHV HE     265  LslPsKAICLhKtKrFmPVQVVDSRWssiRQSDNMTAaAC.QlPYCfFRNtsaNYs.GgthDahHGDfhFrqlLSGLLYnvsCiaQQGaFlYnNVSSs
BCV HE     268  LtvPtKAICLnKrKdFtPVQVVDSRWnnaRQSDNMTAvAC.QpPYCyFRNsttNYv.G.vyDinHGDagFtsiLSGLLYdspCfsQQGvFrYdNVSSv
BEV ORF4   -12  dfyPtrsyCfkhmnftvPVQaiqSiWsvgkeSDdaiAeAC.kpPfCiyfskktpYtvtngsnadHGDdevrqmmrGLLYnssCisaQG.htplalySt
BEV ORF4   -12  dfyPtRSYCFkhMnftvPVqAiQSIWsvGkeSDdAiaeAC.kpPfCiyfsKktPYtvtngsnadHGDdEvRqmmrGLlYnssCISaqG.htplalyst
IVC HE     307  llmPeRSYCF.dMkekgPVtAvQSIWgkGrkSDyAvdqAClstPgCmliqKqkPY.ig.eaddhHGDqEmRellsGLdYearCISqsGwvnetspfte

MHV HE     361  WPaYg..YGhCPTAAnIgy.maPvCiYDPLPvILLGv.......LLGiAVlII..VfLnvlFydg*          413
BCV HE     363  WPlYs..YGrCPTAAdIntpdvPiCvYDPLPlILLGi.......LLGvAViII..VvLllyFmvdngtrlhda*  424
BEV ORF4    85  amlYppmYGsCPqyvklfdgsgsesvdvisssyfvatwv.....LL..vVvII.lVfiiisFcisn*         142
BEV ORF4    85  amLyPPmyGsCPqyvKlfdgsgsesvdvIssSyfvaTwv.....ll..vvvII.LvFiiIsfcisn*         142
IVC HE     402  eyLlPPkfGrCPlaaKeesipkipdgllIptSgtdtTvtkpksrifgiddlIIgLlFvaIveagiggyllgsrk  475
```

C

```
BEV POL1a         qsivyADdPtHFlslPvvn.knFlaafydLQ....e.GfpgKkQvAPHiSltmLklsdediekVe....dilDemvlpnsw.vtitNPHmmGkhyVcDVeG
MHV ns30K    1    ..mAfADKPNHFINFPLaQFsGFmgkYlkLQsQLvemGlDCKlQkAPHvSitlLDIkadqYkqVefAIQEiIDDlaayEG.dIvFDNPHmLGRClVLDVrG
BCV ns32K    1    mavAyADKPNHFINFPLtQFeGFvlnYkgLQfQLldeGvDCKiQtAPHiSlamLDIqpedYrsVdvAIQEvIDDmhwgEGfqIkFDNPHiLGRCiVLDVkG

BEV POL1a         ldsLHdevVsvlRehGiacDQkRlWkpHlTigelndv..sfdkfkdFaisckledc..........dfVKlGapKanarYefittlplGdlnc*
MHV ns30K   99    fEELHeDiVeilRrrGCtADQSRhWIpHCTvAQfdee..retkgmqFyhkepF.YlkhnNlltdAgLELVKiGssKiDGFYcselSvWcGeRlcY   190
BCV ns32K  102    vEELHdDlVnyiRdkGCvADQSRkWIgHCTiAQltdaalsikenvdFinsmqFnYkitiNpsspArLEIVKlGaeKkDGFYetivShWmGsRfeY   195
```

D

```
BEV POL1a     ?  Tvgyttwvsst VC...............rDnthk.hpwFVQiPvneKDPewfhmntql..kdnqW <66 aa> <165 aa insert> RFS <1 aa> *
MHV POL1a     ?  TtnQDSYGGASVCiYCRsrveHPd....vDGlCklrGkFVQvPlgiKDPVsyvLthdiCqVCgfW <23 aa>                  RFS <6 aa> *
IBV POL1a  3845  TpdQDSYGGASVClYCRahiaHPgsvgnlDGrCqfkGsFVQiPtteKDPVgfcLrnkvCtVCqcW <32 aa>                  RFS <9 aa> *
```

FIG. 2. Comparison of the genome organizations of the torovirus BEV and the coronaviruses MHV and IBV. A, Schematic representation of the open reading frames in the 3′ half of the genomes of BEV, MHV, and IBV. The three basic structural genes P, E, and N (BEV) and S, M, and N (MHV and IBV) are represented by dotted boxes. Filled boxes indicate homologous domains in the polymerase proteins of toro- and coronaviruses. The hatched (ns) and cross-hatched (ORF 4 / HE) areas indicate the position of "recombinant" genes in the genomes of BEV and MHV. B, Alignment of the deduced amino acid sequence of the BEV ORF 4 product with the C-terminus of the coronaviral HE protein (upper three rows) and the IVC HE1 protein (lower two rows). Identical amino acid residues are shown in capitals; –, amino acid identity or conservative substitution between BEV and MHV/BCV or BEV and IVC; +, conserved cysteine residue; *, termination codon for translation; the arrow points towards the (present) starting methionine residue of the BEV ORF 4 product. C, Alignment of the C-terminal part of the amino acid sequence of the BEV ORF1a product with the N-terminal parts of the MHV ns30K and BCV ns32K sequences. Legend as for B. D, Alignment of a possible conserved amino acid sequence motif from the C-terminus of the BEV, MHV, and IBV ORF1a products (see Fig. 2A). Legend as for B. The distances (in amino acids) to the ribosomal frameshifting site (RFS) and the ORF1a termination codon are indicated.

Amino acid sequence comparison revealed similarity between a previously reported part of the BEV polymerase (*15*) and a coronaviral nonstructural (ns) protein: the C-terminus of the predicted BEV ORF1a product contains 31-36% identical amino acid residues when compared to the N-terminal 190 aa of the MHV ns30K (*28, 31*) and the BCV ns32K (*36*) protein (Fig. 2C). Like the HE gene, the ns30/32K gene, which is located between the polymerase and HE genes in MHV and BCV (Fig. 2A), is absent in coronaviruses from other antigenic clusters (e.g., IBV; Fig. 2A). Apparently, a sequence related to the 5′ two-thirds of this coronaviral ns gene, which is expressed from a separate subgenomic mRNA in MHV- and BCV-infected cells, has been integrated into BEV ORF1a and is now expressed as a part of the BEV polymerase. The expression of the MHV ns30K protein in infected cells has recently been studied (*37*), but no information about its role in viral replication has been obtained. The suggestion that the ns30K protein contains a nucleotide binding motif (*28*) is opposed by the lack of conservation of this postulated MHV domain in BCV and BEV (Fig. 2C).

The BEV sequence which is homologous to the coronaviral ns protein gene is located just upstream of the ribosomal frameshifting site (*15*). In the coronaviruses IBV and MHV this frameshift region (at the nucleotide level) and the downstream ORF1b (at the amino acid level) are highly conserved (*14*). The ORF1a sequence of IBV has been determined completely (*13*), but from the C-terminal region of the MHV ORF1a product only about 100 aa are known (*14*). Also these C-terminal ORF1a polymerase sequences of IBV and MHV are highly similar (Fig. 2D). In addition, a small domain of sequence similarity with the C-terminal part of the BEV ORF1a product was identified (Figs. 2A and 2D). This similarity is reminiscent of the homologous polymerase domains which were identified in the ORF1b products of toro- and coronaviruses (*15*). Although the motif is very short, its position, immediately upstream of the presumed recombination site, indicates that a recombination–insertion event between this region and the frameshift area in the BEV genome may have taken place.

Information on the genome structure of corona- and especially toroviruses is still quite fragmentary. Although it is difficult to reconstruct the sequence of events which resulted in the present genome organization of viruses like BEV, IBV, and MHV (Fig. 2A), it is clear that nonhomologous RNA recombination has played an important role in their evolution. Apparently, both an ancestor of MHV/BCV and an ancestral torovirus have acquired homologous protein sequences as the result of independent recombination events; the HE and ns30/32K genes are lacking in other coronaviruses (which excludes divergent evolution) and the corresponding BEV sequences are located at different positions in the genome (Fig. 2A). Considering the fact that several present-day representatives of both virus groups cause enteric infections, direct recombination between toro- and coronaviruses during coinfection of the same cell seems feasible. However, the involvement of "a third party" of viral or cellular origin cannot be excluded.

ACKNOWLEDGMENTS

The authors thank Peter Bredenbeek and Willem Luytjes for many enjoyable discussions and assistance with computer analysis. We acknowledge the help of Georges Verjans during a part of the sequence analysis. Part of this project was supported by the Division for Health Research TNO in cooperation with the Organization for Scientific Research (NWO) (Project 900-502-081).

REFERENCES

1. Woode, G. N., Reed, D. E., Runnels, P. L., Herrig, M. A., and Hill, H. T., *Vet. Microbiol.* **7,** 221–240 (1982).
2. Weiss, M., Steck, F., and Horzinek, M. C., *J. Gen. Virol.* **64,** 1849–1858 (1983).
3. Fagerland, J. A., Pohlenz, J. F. L., and Woode, G. N., *J. Gen. Virol.* **67,** 1293–1304 (1986).
4. Horzinek, M. C., Ederveen, J., and Weiss, M., *J. Gen. Virol.* **66,** 1287–1296 (1985).
5. Weiss, M., and Horzinek, M. C., *J. Gen. Virol.* **67,** 1305–1314 (1986).
6. Horzinek, M. C., and Weiss, M., *Zentralbl. Veterinaermed. B* **31,** 649–659 (1984).
7. Horzinek, M. C., Flewett, T. H., Saif, L. J., Spaan, W. J. M., Weiss, M., and Woode, G. N., *Intervirology* **27,** 17–24 (1987).
8. Snijder, E. J., Ederveen, J., Spaan, W. J. M., Weiss, M., and Horzinek, M. C., *J. Gen. Virol.* **69,** 2135–2144 (1988).
9. Snijder, E. J., Horzinek, M. C., and Spaan, W. J. M., *J. Virol.* **64,** 331–338 (1990).
10. Snijder, E. J., Boon, J. A. den, Verjans, G. M. G. M., Spaan, W. J. M., and Horzinek, M. C., *J. Gen. Virol.* **70,** 3363–3370 (1989).
11. Snijder, E. J., Boon, J. A. den, Spaan, W. J. M., Weiss, M., and Horzinek M. C., *Virology* **178,** 355–363 (1990).
12. Spaan, W. J. M., Cavanagh, D., and Horzinek, M. C., *J. Gen. Virol.* **69,** 2939–2952 (1988).
13. Boursnell, M. E. G., Brown, T. D., Foulds, I. J., Green, P. H., Tomley, F. M., and Binns, M. M., *J. Gen. Virol.* **68,** 57–77 (1987).
14. Bredenbeek, P. J., Pachuk, C. J., Noten, J. F. H., Charité, J., Luytjes, W., Weiss, S. R., and Spaan, W. J. M., *Nucleic Acids Res.* **18,** 1825–1832 (1990).
15. Snijder, E. J., Boon, J. A. den, Bredenbeek, P. J., Horzinek, M. C., Rijnbrand, R., and Spaan, W. J. M., *Nucleic Acids Res.* **18,** 4535–4542.
16. Brierley, I., Diggard, P., and Inglis, S. C., *Cell* **57,** 537–547 (1989).
17. Strauss, J. H., and Strauss, E. G., *Annu. Rev. Microbiol.* **42,** 657–683 (1988).
18. Goldbach, R., and Wellink, J., *Intervirology* **29,** 260–267 (1988).
19. King, A. M. Q., *Trends Genet.* **3,** 60–61 (1987).

20. HAHN, C. S., LUSTIG, S., STRAUSS, E. G., and STRAUSS, J. H., *Proc. Natl. Acad. Sci. USA* **85,** 5997–6001 (1988).
21. KING, A. M. Q., *Nucleic Acids Res.* **16,** 11,705–11,723 (1988).
22. KIRKEGAARD, K., and BALTIMORE, D., *Cell* **47,** 433–443 (1986).
23. LAI, M. M. C., BARIC, R. S., MAKINO, S., KECK, J. G., EGBERT, J., LEIBOWITZ, J. L., and STOHLMAN, S. A., *J. Virol.* **56,** 449–456 (1985).
24. MAKINO, S., KECK, J. G., STOHLMAN, S. A., and LAI, M. M. C., *J. Virol.* **57,** 729–737 (1986).
25. KECK, J. G., SOE, L. H., MAKINO, S., STOHLMAN, S. A., and LAI, M. M. C., *J. Virol.* **62,** 1989–1998 (1988).
26. KECK, J. G., MATSUSHIMA, G. K., MAKINO, S., FLEMING, J. O., VANNIER, D. M., STOHLMAN, S. A., and LAI, M. M. C., *J. Virol.* **62,** 1810–1813 (1988).
27. MONROE, S., and SCHLESINGER, S., *Proc. Natl. Acad. Sci. USA* **80,** 3279–3283 (1983).
28. LUYTJES, W., BREDENBEEK, P. J., NOTEN, J. F. H., HORZINEK, M. C., and SPAAN, W. J. M., *Virology* **164,** 415–422 (1988).
29. KHATCHIKIAN, D., ORLICH, M., and ROTT, R., *Nature (London)* **340,** 156–157 (1989).
30. MEYERS, G., RÜMENAPF, T., and THIEL, H. J., *Nature (London)* **341,** 491 (1989).
31. SHIEH, C. K., LEE, H. J., YOKOMORI, K. MONICA, N. LA, MAKINO, S., and LAI, M. M. C., *J. Virol.* **63,** 3729–3736 (1989).
32. PARKER, M. D., COX, G. J., DEREGT, D., FITZPATRICK, D. R., and BABIUK, L. A., *J. Gen. Virol.* **70,** 155–164 (1989).
33. VLASAK, R., MUSTER, T., LAURO, A. M., POWERS, J., and PALESE, P., *J. Virol.* **63,** 2056–2062 (1989).
34. KIENZLE, T. E., ABRAHAM, S., HOGUE, B., and BRIAN, D. A., *J. Virol.* **64,** 1834–1838 (1990).
35. NAKADA, S., CREAGER, R. S., KRYSTAL, M., AARONSON, R. P., and PALESE, P., *J. Virol.* **50,** 118–124 (1984).
36. COX, G. J., PARKER, M. D., and BABIUK, L. A., *Nucleic Acids Res.* **17,** 5847 (1989).
37. BREDENBEEK, P. J., NOTEN, J. F. H., HORZINEK, M. C., and SPAAN, W. J. M., *Virology* **175,** 303–306 (1990).