



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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

## Evolutionary Relatedness of the Predicted Gene Product of RNA Segment 2 of the Tick-Borne Dhori Virus and the PB1 Polymerase Gene of Influenza Viruses

DEBBY A. LIN,\* SONALI ROYCHOUDHURY,\* PETER PALESE,\* WILLIAM C. CLAY,† AND FREDERICK J. FULLER†<sup>1</sup>

\*Department of Microbiology, Mount Sinai School of Medicine, 5th Avenue & 100th Street, New York, New York 10029, and †Department of Microbiology, Pathology & Parasitology, North Carolina State University, College of Veterinary Medicine, Box 8401, Pylon Laboratories, Raleigh, North Carolina 27606

Received November 6, 1990; accepted January 9, 1991

The complete nucleotide sequence of the second largest RNA segment of Dhori/India/1313/61 virus was determined and the deduced amino acid sequence was compared with the polymerase (P) proteins of influenza A, B, and C viruses. RNA segment 2 (2224 nucleotides) of Dhori virus contains a single long open reading frame that can encode a 716-amino acid polypeptide (81.3 kDa). The predicted polypeptide shares between 27 and 31% sequence identities with the PB1 polypeptides of influenza A, B, and C viruses. Among the regions most highly conserved are the sequences around the Asp-Asp motif common to many RNA polymerases. In spite of the high level of sequence identity between the Dhori RNA segment 2 gene product and the influenza A, B, and C virus PB1 proteins the amino acid composition of the Dhori protein indicates an acidic charge feature at pH 7.0 in contrast to the basic nature of the PB1 proteins of the influenza viruses. We suggest that the Dhori PB1-like protein be designated the P $\alpha$  protein of this virus. © 1991 Academic Press, Inc.

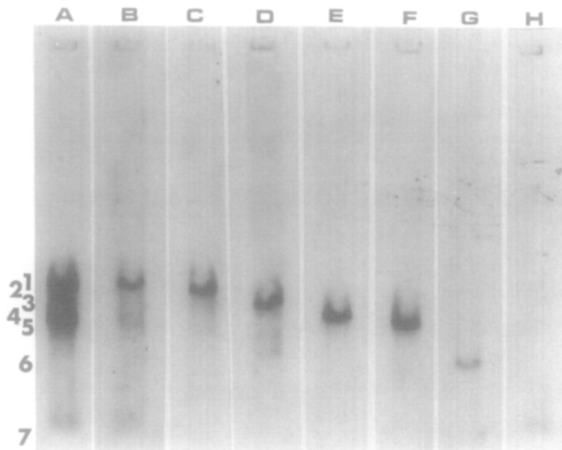
Members of the Dhori virus serogroup are as yet unclassified tick transmitted viruses that share structural and genetic properties with the Orthomyxoviridae (Clerx *et al.*, 1983). Dhori viruses have been isolated from a variety of tick and vertebrate species (Anderson and Casals, 1973; Filipe and Casals, 1979; Karabatsos, 1985; Williams *et al.*, 1973). The genome of the virus consists of seven unique segments of single-stranded RNA with a total size of approximately 11.9 kb. The viral RNAs have been shown to encode information in the negative-sense (Clerx *et al.*, 1983; Fuller *et al.*, 1987; Freedman-Faulstich and Fuller, 1990) and it was previously shown that the Dhori nucleoprotein (encoded by RNA segment 5) shares conserved amino acid sequences with the influenza A, B, and C virus nucleoproteins (Fuller *et al.*, 1987). The sequence of RNA segment 4 encodes the single envelope protein of Dhori virus (Freedman-Faulstich and Fuller, 1990). The envelope protein does not share sequence identity with any orthomyxovirus envelope protein. We have also determined the nucleotide sequence of the Dhori segment 6 RNA and this segment most likely encodes the viral matrix protein (unpublished data). The 3' and 5' nontranslated ends of RNA segments 4, 5, and 6 are very similar to the conserved ends of the genes of the recognized members of the Orthomyxoviridae (Clerx *et*

*al.*, 1983; Fuller *et al.*, 1987; Freedman-Faulstich and Fuller, 1990) and segment 3 of Thogoto virus (Staunton *et al.*, 1989).

A second group of tick-transmitted viruses, the Thogoto viruses, possess a segmented genome with six or seven segments of single stranded RNA (Clerx *et al.*, 1983; Staunton *et al.*, 1989). Thogoto virus has been shown to replicate in *Rhipicephalus appendiculatus* ticks and to be transmitted to laboratory animals (Davies *et al.*, 1986). The sequence of the third largest segment of Thogoto/SiAr/126/72 virus was recently determined (Staunton *et al.*, 1989) and shown to encode in the minus sense a protein that shares 15–20% sequence identity with the influenza A and B virus PA proteins. Recently the sequence of the influenza C virus PA-like protein (designated P3) was determined (Yamashita *et al.*, 1989) and the Thogoto PA-like protein shares a 19.9% sequence identity with the influenza C virus P3 protein. Based on this finding and the fact that the nontranslated ends of the Thogoto virus RNAs are similar to those of the Orthomyxoviridae genes, it is suggested that Thogoto virus is evolutionarily related to the Orthomyxoviridae. It is not yet known, however, to what degree Dhori and Thogoto viruses are related to each other.

We report here the nucleotide sequence of the second largest segment of the Dhori/India/1313/61 virus. The sequence of this RNA segment predicts one gene product which while likely to represent the equivalent of the influenza virus polymerase protein PB1, is

<sup>1</sup> To whom correspondence and requests for reprints should be addressed.



**FIG. 1.** Northern blot analysis of Dhori/India/1313/61 virus-specific vRNA. Purified vRNA was separated by electrophoresis on a 1.5% agarose gel containing 10 mM methylmercuric hydroxide. The viral RNAs were transferred from the agarose gel to Biotrans nylon membranes (ICN Inc., Irvine, CA). Filters were hybridized with nick-translated DNA derived from either a mixture containing pD21-39, pD50-40, pD43-22, pD27-21, pD22-39, pD26-16, and pD12a-35 (lane A) or each of the clones alone; pD21-39 (lane B), pD50-40 (lane C), pD43-22 (lane D), pD27-21 (lane E), pD22-39 (lane F), pD26-16 (lane G), and pD12a-35 (lane H). The clone, pD12a-35, contains an insert from segment 7 as well as a small portion of segment 1.

unique in having a net negative charge unlike the net positive charge of the corresponding A, B, and C virus proteins. Despite these differences the data reported here establish that the influenza viruses and Dhori virus share a common evolutionary ancestor.

## MATERIALS AND METHODS

### Virus

Dhori/India/1313/61 virus was grown in Green monkey kidney (GMK-Vero) cells as previously described (Clerx *et al.*, 1983). Virus purification and RNA extraction have also been described previously (Clerx *et al.*, 1983).

### Isolation of infected cell mRNA

Dhori virus-infected cellular RNA was isolated from Vero cells at 48 hr postinfection, phenol extracted, ethanol precipitated, and purified by oligo(dT)-cellulose chromatography as previously described (Collins *et al.*, 1982; Fuller *et al.*, 1983).

### Cloning of virus-specific DNA and identification of clones

The Dhori/India/1313/61 virus strain was used to infect GMK-Vero cells. Infected cell mRNA was reverse transcribed with an oligo(dT)<sub>12-18</sub> primer. Second-

strand cDNA synthesis was performed by the method of Gubler and Hoffman (1983). The blunt-ended double-stranded cDNAs were ligated to pBR322 which had been cut by *Pvu*II and dephosphorylated. Clones bearing Dhori virus specific inserts were identified by colony hybridization using radiolabeled cDNA which had been obtained using viral RNA as template and the oligonucleotide 5' AGCAA(A/T)AACAAAGCAGT 3' as primer. This oligonucleotide is complementary to sequences common to the 3' ends of the seven viral RNA segments (Clerx *et al.*, 1983). Virus-specific clones were organized by cross-hybridization into several pools. Viral RNA was extracted from cesium chloride gradient (12–42% (w/w)) purified Dhori virus and resolved on a 1.5% agarose gel containing 10 mM methylmercuric hydroxide. The viral RNAs were transferred from the agarose gel to Biotrans nylon membranes (ICN Inc., Irvine, CA) and hybridized with nick-translated plasmid DNA.

### Sequencing of cloned cDNA

Both strands of insert DNA of pD50-40 were directly sequenced using specific synthetic oligonucleotide primers and the modified T7 DNA polymerase Sequenase (United States Biochemical Corp.). All oligonucleotides were synthesized using an Applied BioSystems 380A synthesizer (Applied BioSystems, Foster City, CA). For plus-strand as well as for minus-sense sequencing specific primers were synthesized to permit the reading of new sequences over a distance of 150–200 nucleotides. In order to obtain sequences located at the 3' end terminus of the RNA segment 2, polymerase chain reactions were performed using purified viral RNA and the following oligonucleotides, 5'AGCAA(A/T)AACAAAGCAGT 3' and 5'ATCTCTGTG-GAAGCCAC 3'. The sequence located at the 5' end terminus of the RNA segment 2 was also obtained by polymerase chain reaction performed on purified viral RNA and the following oligonucleotides, 5'AGTAGA(C/A)ATCAAAGCA 3' and 5'ACCTCTTTTGTGAAG 3'. The resulting dsDNAs were cloned into M13mp18RF and directly sequenced using universal primer.

## RESULTS AND DISCUSSION

An analysis of the Dhori virus genome was attempted by DNA cloning of the poly(A)-containing RNA of virus-infected GMK-Vero cells. Virus-specific cDNA clones were obtained and by cross-hybridization were grouped into seven pools. Nick-translated DNAs of plasmids pD21-39, pD50-40, pD43-22, pD68-30, pD22-39, pD26-16, and pD12a-35 hybridized to RNAs 1, 2, 3, 4, 5, 6, and 7, respectively in a Northern blot analysis of Dhori virus RNA (Fig. 1). The clone pD12a-

<u>AGCAAAAACAAGCAGTTGACAGTCATGAATCTCTTCTCTCCAGAACAGGGCTGAGCCCCACAGAAACCAAGA</u>	100
<u>M N L F S P R T G L S P T E T Q E L L Y A Y T G P</u>	25
CGCCAGTAGCATACGGGACCAACAAGAGCAGTTTTAGAGAATGTGCTGAGACCGTATAAATACTTCTATAAAGAAGAAAACGTTCCCAAGCATTGCA	200
A P V A Y G T R T R A V L E N V L R P Y K Y F Y K E E N V P Q A L H	59
CATCAAAACAGGCCAAAAGGGACCAAGAAGAAATCTCTACGACATCTCCGAGTAGTGGCTTCCACAGAGATTCTGTGATCCTGCTCTAGAGAAGTGAAG	300
I K T G Q K G P E E I S T T S P S S G F H R D S V I L L S R E L K	92
GCCAAATACCCAGAAGCATTGAAACGGCTGAAAGCATGGATTGATTGTAACCTGCTAGAGATGGAGTATGCAGAGCTCAGCAAGGGCAGACAGACTTTGA	400
A K Y P E A F E R L K A W I D C E L L E M E Y A E L S K G R Q T L	125
GCTTCTTAAGGAACAGGAATCAACAGCACCATAGCATTAGAAGAACCATAGAGTATCTACAACAAAACCTAGGAAGGCCAATCGGACAATCAATGCT	500
S F L R N R N Q P A P I A L E E T I E Y L Q Q N L G R P I G Q S M L	159
AAGTTATCTCAGAGCTGTGATGGAAGTGTGGCAATGCCTAAAACCTTTTACATACGAAGTAGCAACAAACCTTGCCAGATTGACTTTGAGGAATAT	600
S Y L R A V M E V L A M P K T T F T Y E V A T N L A R F D F E E Y	192
GATGACGGAGAGACAGGGCCCTCATGATGCACCTCGAGAAAGAGAAGAAGAAAATCACCATCACCATCACTCAAGCAGAATATGGGAAAAACCTGCA	700
D D G E T G P L M M H F E K E K K K I T I T I T Q A E L W E K T C	225
CATTGGGCACCATGTGGAAGCATCTAGAAAGAGGACGGCTCAACAGGCGTACCATTGCTACTCCTTCTATGCTGGCGAGAGGGTTGTAATAAGTGA	800
T L G T M W K H L E R G R L N R R T I A T P S M L A R G F V K I V E	259
GGACGCTGCAAGGGTATTGCTTGAAGTGCCTTCAAGCTCGGTGTGCTGTAGGAGGGGAAGAGAAGTTGGCCAAGCTTCAAGCAAGTTGGAGGCTGTC	900
D A A R V L L E C L P S S G V P V G G E E K L A K L S S K L E A V	292
TCAGAAGTAACTGGGAGCTGAGTGGGATCAGGAAAATCAATGAGTGTGGACCCAGATGCAATGAGGCTAATGTGGACAGTGTTCCTGAAGACT	1000
S E V T G E L S G D Q E K F N E C L D P D A M R L M W T V F L E D	325
ATCCACAGTGGGTAAGAAGCTTCAACATCCCTTTCTTAATATTCAAGGCAAAAATAGCAGATATCGGAGAAGGACTGACTTACCAGAAAGAAGGAGT	1100
Y P Q W V K E L F N I P F L I F K A K I A D I G E G L T Y Q K E G V	359
TGTTAGGGTGTCCCTTTGGGGAGATGCCCTCAGAATTCGATGAAGTGTGCTCAATGCAATAAAGGACAAAGAGGGGAAAATGTAGGCATCAGATGC	1200
V R V F P F G E M P S E F D E L L P N A I K D K E G K I V G I R C	392
ACATTAGGCATGTTTATGGGAATGTTCAATCTTAGTCTACTCTACTAGCCTTAATAGCAGCAGATAGGCTGAAATACAGGGGACCAGCTGGAGAGTA	1300
T L G M F M G M F N L S S T L L A L I A A D R S E I T G D H V E S	425
GTGACGATTTACCACTTTTCAAAGCAAAGAGTACGACGACATGTTCAAGCAAGCAGAATTGTTGCGATGGTCTCTAAAACCTCGTTGGAATAAATCAAT	1400
S D D F I H F F K A K S Y D D M F K Q A E L L R W S L K L V G I N M	459
GTCACCGAGTAAATGCATCTGATAAGTCTGCAAGTATCGGAGAATTCAATCAAAAATATCATCACCAGGATTTGTGGGCAATGTCGCAACCGATTTA	1500
S P S K C I L I S P A G I G E F N S K Y H H R D F V G N V A T D L	492
CCTTCACTGGTCCCTGGTGGAAAAATCCTTCAAGTATCTAGCCATGGGACTAAATGTTATCAGGCATTCTATCAACACCAACCAATGAACCTTCATAA	1600
P S L V P G G K N P S S D L A M G L N V I R H S I N T N Q M N F I	525
GTGGAGACTTGGCATTGAGGATCTTACCAGCCTATCGCCATTCTTACATGGCAGAAGGAATCACTCGAAGAACAATACTCGGAGCATTAAAGAA	1700
S G D L A L R I F T K A Y R H S Y M A E G I T R R T K F L E A F K K	559
GGACCCAGTACTCTGAACCAAGGTCTCAACCGTCCACAGCGTGTCAACCCCTCCACCTCGACGAAGTTTGTCTCAGATATCAAATGCATTTGCTAGGA	1800
D P V L L N Q G A P T V H S V S T L H L D E V C L R Y Q M H L L G	592
GAGGAAGAGCTCAGAAGAATCATGAACCTAGCAATCAATCACAGCAAGAAGTGAAGAAGTGGTGAAGTCCGGCCAGAAGGCAAGCTGCCAATGATTC	1900
E E E L R R I M N P S N P I T A R T E E V V S F R P E G K L P M I	625
TGGAAGACAACAGTGTGGGACGCTGTTTCAATACACCTTTACCAGAAACCCGACAGTACTGACAAGCCACATAGGGTACTTTTAGAAAAGAACAGCA	2000
L E D N S V G S C F K Y T F T R N R T V T D K P H R V L L E K E Q Q	659
ATACCAAAAGATCACCTCTTTGTTGAAGAATGCTTCCAGAACTCACTATCGGGAACACCACAATGCCAGGCACAGTAAAACAGGCTTGAACCAAGG	2100
Y Q K I T S F V E E C F P E L T I G N T T M P G T V K Q A C K R R	692
CTGGAGTACATTATTGAACAGAGTACTTGCCTACAGAGCAAAAACGGGCTCTGCTGGAAGAGATGGATAGTTAGAGACATGTGTATACCTGAATACTCA	2200
L E Y I I E Q S D L P T E Q K R A L L E E M D S ***	716
<u>GGAAAAACTGCTTTGATATCTCT</u>	2224

Fig. 2. Complete nucleotide sequence of RNA segment 2 of Dhori/India/1313/61 virus and the deduced amino acid sequence. Sequencing of segment 2-specific cDNA from pD50-40 was accomplished using Sequenase (United States Biochemical Corp., Cleveland, OH), and specific oligonucleotides (see Materials and Methods). The 25 3'-terminal and the 16 5'-terminal sequences (underlined) of the gene were determined as described in the text.

35, contains a small insert of segment 1 as well as an insert for segment 7 (unpublished results). Plasmid pD50-40 was selected for further study since it had the largest insert among a total of nine different clones

belonging to the group that hybridized to RNA segment 2.

The insert of pD50-40 was then sequenced using specific synthetic oligonucleotide primers and the mod-

TABLE 1

AMINO ACID SEQUENCE IDENTITIES AND ALIGN SCORES OF THE DHORI P $\alpha$  PROTEIN AND THE PB1 PROTEINS OF INFLUENZA A, B, AND C VIRUSES

	Dhori <sup>a</sup>	A/PR/8/34 <sup>b</sup>	B/ Ann Arbor/ 1/66 <sup>c</sup>	C/JJ/50 <sup>d</sup>
Dhori/India/ 1313/61	—	26.3 <sup>e</sup>	25.8	25.7
A/PR/8/34	30.7% <sup>f</sup>	—	152.1	64.3
B/Ann Arbor/ 1/66	26.9%	60.9%	—	70.0
C/JJ/50	29.8%	39.2%	42.1%	—

<sup>a</sup> P $\alpha$  protein 716 amino acids (81.3 kDa) (this article).

<sup>b</sup> PB1 protein 757 amino acids (86.6 kDa) (Winter and Fields, 1982).

<sup>c</sup> PB1 protein 752 amino acids (84.3 kDa) (DeBorde *et al.*, 1987).

<sup>d</sup> P3 protein 754 amino acids (85.9 kDa) (Yamashita *et al.*, 1989).

<sup>e</sup> Alignment score (above diagonal) determined by the program PCOMPARE (Intelligenetics) using the unitary matrix, a bias of 0, a gap penalty of 3, and 100 random runs. A score of 5 or greater is considered significant (Dayhoff, 1983).

<sup>f</sup> Percent amino acid identity (below diagonal) based on an alignment of the two sequences by the method of Needleman and Wunsch (1970).

ified T7 DNA polymerase–Sequenase. Sequences of the 3'- and 5'-terminal regions of the RNA segment 2 absent from the insert in pD50-40 were determined by polymerase chain reaction amplification, cloning of the amplified DNAs into M13mp18RF, and subsequent sequencing of the inserts.

The complete sequence of segment 2 of Dhori/India/1313/61 virus is shown in the plus sense in Fig. 2. The RNA is 2224 nucleotides in length and the cDNA insert of pD50-40 only lacks 25 nucleotides at the 5' end (in the plus-sense orientation). The RNA encodes a single long open reading frame that could be initiated by an AUG codon at nucleotides 25 through 27. No other reading frame is open in the antigenome sense for longer than 64 amino acids and the longest open reading frame in the genome sense is 109 amino acids. We do not know if these smaller reading frames can be expressed but we have not detected any sub-genomic messenger RNAs derived from RNA segment 2 (unpublished data). The AUG codon at nucleotides 25 through 27 is in a strong context for protein synthesis initiation (G at -3, A at +4; Kozak, 1986). A 716-amino acid polypeptide is predicted from the largest open reading frame and it is terminated with a single UAG codon at nucleotides 2173 through 2175.

A search of the NBRF (National Biomedical Research Foundation) protein database using the FASTP program (Lipman and Pearson, 1985) indicated that the predicted gene product of Dhori segment 2 was most

closely related to the PB1 proteins of the influenza virus group. The PB1 polymerase proteins are the most highly conserved among the proteins of the influenza A, B, and C viruses (Yamashita *et al.*, 1989) and they are most likely required for nucleotide addition during viral RNA synthesis (Braam *et al.*, 1983). A pairwise comparison of the amino acid sequence identities of the Dhori virus protein and the influenza A, B, and C virus PB1 proteins are shown in Table 1. The optimized alignments were made by the method of Needleman and Wunsch (1970) and indicate that the Dhori PB1-like protein shares sequence identities of 30.7, 26.9, and 29.8% with the PB1 proteins of influenza A/PR/8/34, influenza B/Ann Arbor/1/66, and influenza C/JJ/50 virus, respectively (Table 1). In addition, we compared the sequence of the Dhori PB1-like protein to the PB1 proteins of influenza A, B, and C viruses using the PCOMPARE (Intelligenetics) program. This program determines the probability that the similarity between two sequences could occur by chance. The score is expressed as the number of standard deviation units by which the maximum score for the real sequences exceeds the average maximum score for 100 randomizations of the sequence. An alignment score greater than 5 (probability of similarity occurring by chance  $\leq 3 \times 10^{-7}$ ) implies an evolutionary relationship between proteins (Dayhoff *et al.*, 1983). Pairwise comparisons (see Table 1 above diagonal) between Dhori and influenza PB1 proteins yield alignment scores of 25–26 which clearly indicate an evolutionary relationship between the Dhori virus PB1-like protein and the influenza PB1 proteins. A similar comparison between the Dhori segment 2 gene product and the influenza A/PR/8/34 PB2 protein and PA protein gave alignment scores of 0.25 and 0.015, respectively.

Comparative analysis of the deduced amino acid sequences revealed highly conserved regions in the PB1 proteins of influenza A, B, and C viruses (Kawaoka *et al.*, 1989; Yamashita *et al.*, 1989). Although the functional domains of the PB1 proteins of influenza viruses have not been determined, the conserved regions may be important for basic functions of the PB1 proteins, such as initiation and chain elongation (Braam *et al.*, 1983). It has also been observed that there are four highly conserved motifs among RNA-dependent RNA polymerases (plus- or minus-strand virus polymerases) as described by Poch *et al.* (1989) and the Dhori P $\alpha$  protein shares all of these motifs (Fig. 3). A comparison of the Dhori virus protein with the PB1 proteins shown in Fig. 3 also reveals regions of highly conserved sequences. A 15-residue sequence motif (a double aspartic acid core flanked by hydrophobic residues) is found in many DNA and RNA polymerases (Argos, 1988; Poch *et al.*, 1989). The PB1 gene in influenza A,

DHO/INDIA/61 A/PR/8/34 B/ANN ARBOR/66 C/JJ/50	MNLFSPRTGLSPTETQELLYA--YTGPAPVAYGTRTRAVLENVLRPYKYFYK-----EENVQALHIKTGQKGPTEEISTTSPSSGFHRDSV *DVNPTLLF*KVPAQNAISTTFP***DP*YSH**G*GYTMDT*N*THQ*SE*GRWTTNT*TGA**LNP*DGPLPEDN*-----*GYAQT*C* **INPYFLFIDVPIQAAISTTFP***VP*YSH**G*GYTIDT*I*THE*SN*GK--QYISDVIGC*MVDP*NGPL**DNE---**AYAQL*C* *E1*PYLMF*NNDV*SLISTTYP****P*MSH*SS*KYT**TIK*T*D*SRTSV----EKTSKVFNIPRRKFCNCL*DKDDLKPT*NVDI*S	84
DHO/INDIA/61 A/PR/8/34 B/ANN ARBOR/66 C/JJ/50	ILLSRELKAKYPEA--FERLKAWIDCELLEMEYAEASKGRQTL SFLRNRNPAPIALEETIEYLQNLGRPIGQMSLSY-LRAVMEVLAMPKT LEAMAF*EESH*GI--**NSCIETMEVVQQRVDK*TQ****YDWTL*****AT**AN***VFRS*GLTANESGR*IDF*KD***SMKKEEM LEALDRMDEEH*GL--*QAASQNAMEA*MVTTVDK*TQ****FDWTV*****AT**NT**TSFRL*DLNGADKGG*VPFCQDIIDSLDK*EM L*GLA*MME*RMGEGF*KHCVMEAE*T*I*K*HFSR*TE****YDWTSE**M**AT**QL*VDAIKETE*PFK*TT**E*-CNKMI*M*DWKEV	174
DHO/INDIA/61 A/PR/8/34 B/ANN ARBOR/66 C/JJ/50	TFTYEAVTNLARDFEEDDGETGLMMHFEKEKKITITITQAELEWKTCTLGTMMKHLERGRINRRTIATPMSLARGFVKIVEDAARVLE GI*THFQRKRVR*NMTKK-----*ITQRTIG*RKQLNKRSLIRAL**N**T*DA***K*K**A***G*QI****YF**TL**SIC* **FSVK*IKKKLPAKNRKGFLIKRIP*-----*VKDR**RV*YIKRALS*N**T*DA***K*K**A***AGIQ****LV**NL*KNIC* K*RKVKTMVRREK*KRSKGEIK*KVPV*GI-----DS*KHD*FLIRAL*IN**A*DG***K*Q**A***G*IV*P*S****TV*QKIC*	267
DHO/INDIA/61 A/PR/8/34 B/ANN ARBOR/66 C/JJ/50	CLPSSGVPVGGEEKLAKLSSKLEAVS----- <u>domain 1</u> EVTGELSGDQEKSFNECLDPDAMRLMWTVFLEDYDQWVKELFNIPFLIFKAKIADIGEG--- K*EQ**L****N**K***ANVVRKMMTNSQDT*L*SLTIT**NT*W**NQN*RMFLA*I*YMTRNQ*E*FRNVL*S*APIM*SN*M*RL*K*--- N*EQ**L****N**K***NAVAKMLSNCPGGISMTVT**NT*W*****RIFLA*TERITR*S*I*FRDFCS*APVL*SN***RI*K*FMI K*KE**L****N**K***--*T*VT*SNCPGGISMTVT**NS*W***QQ*E*YLALLAYITK*SSDLM*D*CSVAPVL*CN*FVKL*Q*IRL	351
DHO/INDIA/61 A/PR/8/34 B/ANN ARBOR/66 C/JJ/50	-----LTYQ-KEGVVRFVFPFGEMPSEFDPELLNAIKDKEGKIVGIRCTL----- <u>domain 2</u> GMFMGMFNLSSTLLALIAADRSEITG-----D -----YMFES*SMKL*TQIPA**LASI*LYF*DSTR*--*EK**PL*IEGTASLSP**M****ML**V*GVSILNLGQKRYTKTTYW* TSKTKR*KA*-IPCPDLFNIPL*RYN*ETRAKLKLL*PFFNE-----EGTASLSP**M****ML**V*GVA*LGIKN*GNRE--YLW* SNKRT-----*VIKAEK*KYKNLMR*EYK*LFEP*LY*QKDV*F*PG-----*L****ML**V*GVSTLCYMDEELKAKGCFWT	421
DHO/INDIA/61 A/PR/8/34 B/ANN ARBOR/66 C/JJ/50	<u>domain 3</u> <u>domain 4</u> HVESDDFIHFFKAKSYDDMFQKAE LLRWLKL VGINMSPSKCILISPAGIEGFNSKYHHRDFVGNVATDPLSLVPGGKNPSSDLAMGLNVIR GLQ*****ALIVN*PNHEGIQAGVDRFYRTC**L*****K*K*SY-NRT*TF**T*FFYRYG**A*FSME***FGVS*I*E*A*MSI*VT**K GLQ*****AL*VN**DEETCMEGINDFYRTC**L*****K*K*SYNET--MF**T*MFYRDG**S*F*ME***FGVA**V*E*A*M*I*MTI*K GLQ*****VL*AV*SNWSNIHWITIRRFNAVC**I*****L*E*SYGSL*ELF--*T*MFDFGE**S*L*ME**AFFTA*V*EGV*FTAAMSI*K	514
DHO/INDIA/61 A/PR/8/34 B/ANN ARBOR/66 C/JJ/50	HSINTNQMFISGDALRIFTKAYRHSYMAEGITRRTKFLFAFK-----KDPVLLNQGAPT VHSVSTLHLDEVCLRYQMHLGEEELR NNMIN*DLGPATAQM**QL*I*D**YT*CHRGDGTIQTRRS*EIKLWEQTRS*AGL*VSD*G*NLNIRN**IP****KWE--*MD*DYQG NNMIN*G*GPATAQT*IQL*IAD**YT*KCHRGDGSKVEGKRMKIKELWENTKGR*GL*VAD*G*NINLNRN**IP*IV*K*N--LMDPEYKG TNMIN*SLSPSTALM***CLQEF*AT*RVHPWDSKV*GGRMKIINEFIKTIIES**GL*IAD*GKLMNNI****IP*LVKFEK*D---*QYRN	597
DHO/INDIA/61 A/PR/8/34 B/ANN ARBOR/66 C/JJ/50	RIMNPSNPITARTEEVVFR---PEGKLPMILEDNSVSGCFKYTFTRNRTVTDKPHRVLLEKEQYQKITSFVEECFPELTIGNTTTPGTVK *LC**L**FVSHK*IESMNAVMM*AHGPAKNM*YDA*ATTHSWIPK***SILNTSQ*GV**D**M**RCCNLF*KF**SSSYRRPVGISSMV *LLH*Q**FVGHLSIEGKEADITPAHGPIKMDYDA*SGTHSWRTK***SILNTDQ*NMILE**C*A*CCNLF*A**NSASYRKPVGQHSML *VF**K**F*NFDKTIDI**---AH*PIRVE-*NEA*V*THSFRTRA***LLNTDM*AMMAE*KR**MVCDFKSV*ESAD*N*PPIGAMSI	686
DHO/INDIA/61 A/PR/8/34 B/ANN ARBOR/66 C/JJ/50	QACKRRLEYIEQSDLPTEQKRALLEEMDS E*MVS*ARIDARIDFESGRI*KEEFT*IMKICSTIEELRRQK E*MAH**RMDARLDYESGRMSKDDFEKAMHLGEIGHI G-----*A*E*KLLERAKM**DIGATE**EYEEIKDIIRDAKKARIESR	716

Fig. 3. Comparison of the deduced amino acid sequences of the PB1 proteins of influenza A/PR/8/34 (Winter and Fields, 1982), B/Ann Arbor/1/66 (DeBorde *et al.*, 1988), and C/JJ/50 (Yamashita *et al.*, 1990) with the P $\alpha$  protein of Dhori/India/1313/61 virus. The alignment was obtained by a compilation of the pairwise comparisons by the method of Needleman and Wunsch (1970). An asterisk (\*) denotes identity with the amino acid of the Dhori P $\alpha$  protein at that position and a dash (-) indicates a gap introduced in the sequence for optimal pairing. The four domains designated by the lines above and below the sequence represent those highly conserved regions defined in RNA polymerases by Poch *et al.* (1989).

B, and C shares this motif (W/FWD/TGLQSSDDFA/VLI/FV/A), at residues 435/438 to 449/452 (domain 3, Fig. 3). In the Dhori virus protein, we found a similar sequence at residues 419 to 433 (TGDHVESDDFIHFF). All rules of this Asp-Asp polymerase motif are fulfilled except for position 7 which contains a serine instead of the traditional Gly, Met, Cys, Val, or Leu residue (Ar-

gos, 1988). Thus it is likely that the Dhori virus protein also possesses a polymerase function similar to that of the PB1 proteins of influenza viruses.

The influenza virus PB1 protein is transported to the nucleus independently of other viral proteins (Smith *et al.*, 1987; Akkina *et al.*, 1987). Nath and Nayak (1990) have recently shown that the nuclear localization sig-

nal of influenza virus (A/WSN/33) PB1 protein was located within residues 180–252 of PB1. The study suggested that two discontinuous regions of the PB1 both containing a stretch of basic amino acids were required for its nuclear localization. Homologous sequences, however, were not found in the Dhori virus protein.

A comparison of some of the physical properties of the Dhori and influenza PB1 proteins indicates that the Dhori protein is the smallest in this group (716 amino acids;  $M_r$  81,292) and unlike the influenza PB1 proteins which have a strong positive charge at neutral pH (+23-A/PR/8/34; +12-B/AA/1/66; +14-C/JJ/50) the Dhori protein is slightly acidic with a charge of  $-10$  (PEPTIDESORT; Devereux *et al.*, 1984). Since the nomenclature of the influenza virus polymerase proteins is based on their charge characteristics a different name for the related protein of Dhori virus may be appropriate and we thus suggest the designation  $P\alpha$  protein. This nomenclature is proposed for the designation of the Dhori polymerase proteins [ $P\alpha$ -(PB1),  $P\beta$ -(PB2), and  $P\gamma$ -(PA)] since the charge characteristics do not appear to be constant across virus groups. For example, the PA-like polymerase protein of influenza C/JJ/50 virus (Yamashita *et al.*, 1989) encodes a weakly basic protein designated P3 (segment 3 gene product). The third largest RNA segment of Thogoto virus encodes a PA-like protein with only a very weak negative charge ( $-2.5$  at pH 7.0; Staunton *et al.*, 1989).

Genes of other viruses also share sequence identity with influenza virus segments. It has been previously demonstrated that some members of the coronavirus group and influenza C virus have a hemagglutinin-esterase (HE) glycoprotein that recognizes the same *N*-acetyl-9-*O*-acetylneuraminic acid receptors on cells (Vlasak *et al.*, 1988). The influenza C virus HE glycoprotein shares an approximately 30% amino acid sequence identity with a coronavirus HE glycoprotein present in bovine coronavirus (Kienzle *et al.*, 1990) and mouse hepatitis virus (Luytjes *et al.*, 1988). This example of significant identity between a plus-strand RNA and a minus-strand RNA virus supports the concept of a cassette model of evolution for the HE gene and is most likely an example of a recombinational event. In contrast, we feel that the relationship of the Dhori virus group to the orthomyxovirus group is most likely an example of divergent evolution in light of the similarities in genome structure and common replication strategies and of the observation that several Dhori virus gene products share significant identity with orthomyxovirus genes.

#### ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health Grants AI-18998 (P.P.), AI-11823 (P.P.) and AI-20939 (F.F.). This work was in partial fulfillment of a Westinghouse project (D.L.).

#### REFERENCES

- AKKINA, R. K., CHAMBERS, T. M., LONDO, D. R., and NAYAK, D. P. (1987). Intracellular localization of the viral polymerase proteins in cells infected with influenza virus and cells expressing PB1 protein from cloned cDNA. *J. Virol.* **61**, 2217–2224.
- ANDERSON, C. R., and CASALS, J. (1973). Dhori virus, a new agent isolated from *Hyalomma dromedarii* in India. *Indian J. Med. Res.* **61**, 1416–1420.
- ARGOS, P. (1988). A sequence motif in many polymerases. *Nucleic Acids Res.* **16**, 9909–9916.
- BRAAM, J., ULMANEN, I., and KRUG, R. M. (1983). Molecular model of a eukaryotic transcription complex: Function and movements of influenza P proteins during capped RNA primed transcription. *Cell* **34**, 609–618.
- CLERX, J. P. M., FULLER, F. J., and BISHOP, D. H. L. (1983). Tick-borne viruses structurally similar to orthomyxoviruses. *Virology* **127**, 205–219.
- DAVIES, C. R., JONES, L. D., and NUTTALL, P. A. (1986). Experimental studies on the transmission cycle of Thogoto virus, a candidate orthomyxovirus, in *Rhipicephalus appendiculatus*. *Am. J. Trop. Med. Hyg.* **35**, 1256–1262.
- DAYHOFF, M. O., BARKER, W. C., and HUNT, L. T. (1983). Establishing homologies in protein sequences. In "Methods in Enzymology" (C. H. W. Hirs and S. N. Timasheff, Eds.), Vol. 91, pp. 524–545. Academic Press, San Diego.
- DEBORDE, D. C., DONABEDIAN, A. M., HERLOCHER, M. L., NAEVE, C. W., and MAASSAB, H. F. (1988). Sequence comparison of wild-type and cold-adapted B/Ann Arbor/1/66 influenza virus genes. *Virology* **163**, 429–443.
- DEVEREUX, J., HAEBERLI, P., and SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387–395.
- FILIFE, A. R., CALISHER, C. H., and LAZUICK, J. (1985). Antibodies to Congo-Crimean haemorrhagic fever, Dhori, Thogoto and Bhanja viruses in southern Portugal. *Acta Virol.* **29**, 324–328.
- FILIFE, A. R., and CASALS, J. (1979). Isolation of Dhori virus from *Hyalomma marginatum* ticks in Portugal. *Intervirology* **11**, 124–127.
- FREEDMAN-FAULSTICH, E. Z., and FULLER, F. J. (1990). Nucleotide sequence of the tick-borne, orthomyxo-like Dhori/India/1313/61 virus envelope gene. *Virology* **175**, 10–18.
- FULLER, F. J., FREEDMAN-FAULSTICH, E. Z., and BARNES, J. A. (1987). Complete nucleotide sequence of the tick-borne, orthomyxo-like Dhori/India/1313/61 virus nucleoprotein gene. *Virology* **160**, 81–87.
- GUBLER, U., and HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269.
- KARABATSOS, N. (1985). International Catalogue of Arboviruses Including Certain Other Viruses of Vertebrates. Am. Soc. Trop. Med. Hyg., San Antonio.
- KAWAOKA, Y., KRAUSS, S., and WEBSTER, R. G. (1989). Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. *J. Virol.* **63**, 4603–4608.
- KIENZLE, T. E., ABRAHAM, S., HOGUE, B. G., and BRIAN, D. A. (1990). Structure and orientation of expressed bovine coronavirus hemagglutinin-esterase protein. *J. Virol.* **64**, 1834–1838.
- KOZAK, M. (1986). Point mutations define a sequence flanking the AUG initiation codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.
- LIPMAN, D. J., and PEARSON, W. R. (1985). Rapid and sensitive protein similarity searches. *Science* **227**, 1435–1441.
- LUYTJES, W., BREDEBEEK, P. J., NOTEN, A. F. H., HORZINEK, M. C., and SPAAN, W. J. M. (1988). Sequence of mouse hepatitis virus A59 mRNA 2: Indications for RNA recombination between coronaviruses and influenza C virus. *Virology* **166**, 415–422.
- NATH, S. T., and NAYAK D. P. (1990). Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 Influenza virus (H1N1). *Mol. Cell. Biol.* **10**, 4139–4145.

- NEEDLEMAN, S. B., and WUNSCH, D. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–458.
- POCH, O., SAUVAGET, I., DELARUE, M., and TORDO, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8**, 3867–3874.
- SMITH, G. L., LEVIN, J. Z., PALESE, P., and MOSS, B. (1987). Synthesis and cellular location of the ten influenza polypeptides individually expressed by recombinant vaccinia viruses. *Virology* **160**, 336–345.
- STAUNTON, D., NUTTALL, P. A., and BISHOP, D. H. L. (1989). Sequence analyses of Thogoto viral RNA segment 3: Evidence for a distant relationship between an arbovirus and members of the Orthomyxoviridae. *J. Gen. Virol.* **70**, 2811–2817.
- VLASAK, R., LUYTJES, W., SPAAN, W., and PALESE, P. (1988). Human and bovine coronaviruses recognize sialic acid containing receptors similar to those of influenza C viruses. *Proc. Natl. Acad. Sci. USA* **85**, 4526–4529.
- WILLIAMS, R. E., HOOGSTRAAL, H., CASALS, J., KAISER, M. N., and MOUSSA, M. I. (1973). Isolation of Wanowrie, Thogoto, and Dhori viruses from *Hyalomma* ticks infecting camels in Egypt. *J. Med. Entomol.* **10**, 143–146.
- WINTER, G., and FIELDS, S. (1982). Nucleotide sequence of human influenza A/PR/8/34 segment 2. *Nucleic Acids Res.* **10**, 2135–2143.
- YAMASHITA, M., KRYSTAL, M., and PALESE, P. (1989). Comparison of the three large polymerase proteins of influenza A, B, and C viruses. *Virology* **171**, 458–466.