Unique Sequence Characteristics of Genes in the Leftmost Region of Unique Long Region in Duck Enteritis Virus

Xiaoli Liu Shengwang Liu Huixin Li Zongxi Han Yuhao Shao
Xiangang Kong

Division of Avian Infectious Diseases, National Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin, PR China

Key Words

Unique long region • Duck enteritis virus • Infected-cell polypeptide 4

Abstract

Objective: Our objective was to further understand the evolutionary position of duck enteritis virus (DEV) in the family Herpesviridae and to determine the genomic structure in the leftmost region of the unique long region (UL) of DEV. **Methods:** 'Targeted gene walking PCR' was used to amplify the unknown gene of DEV clone-03 strain adjacent to known sequences in the UL region. Open reading frames (ORFs) were determined by Gene Runner and an online transcriptional element search engine. The phylogenetic tree was constructed by DNASTar based on the deduced amino acid sequences. **Results:** A fragment of 13,947 bp, containing 5 ORFs (designated *ICP4*, *ORF1*, *ORF2*, *LORF2* and *LORF3*, respectively), was amplified from the DEV clone-03 strain genome. In addition, a tandem repeat region of α -type-like sequences was detected between *ICP4* and *LORF2*. Phylogenetic analysis of *ICP4* indicated that DEV *ICP4* should be clustered into Varicellovirus. DEV *ORF1* and *ORF2* were unique among herpesviruses. Additionally, the existence and characteristics of *LORF2* and *LORF3* indicated a close relationship with Mardivirus members. **Conclusion:** The characteristics and

gene content in the leftmost end of the UL region of the DEV genome showed that the DEV genome is different from those of other herpesviruses. Copyright © 2009 S. Karger AG, Basel

Introduction

The family Herpesviridae contains a number of large DNA viruses that infect humans and other mammals, as well as birds, amphibians and reptiles [1]. In nature, most infect only a single species [2]. The Herpesviridae have been divided into 3 subfamilies, α -, β - and γ -Herpesvirinae, initially based on their distinct biological properties and more recently on the basis of more precise genomic attributes [3]. Avian herpesviruses fit readily into the subfamily of α -Herpesvirinae [4], with the exception of duck enteritis virus (DEV) which is an unassigned Herpesviridae virus [5]. DEV or anatid herpesvirus 1, is the causative agent of duck viral enteritis or duck plague, an acute and contagious disease with high morbidity and mortality in birds of all ages from the order Anseriformes, which contains ducks, geese and swans [6]. Duck viral enteritis is difficult to monitor and control since DEV establishes an asymptomatic carrier state in waterfowl that is detectable only during the intermittent shedding pe-

Fig. 1. Arrangement of genes at the left of the DEV UL region and PCR amplification strategy. Arrows indicate genes and white bars indicate interval sequences between genes. The asterisk indicates a putative splice site. The strategy for PCR amplification is shown below. Six overlapping fragments were obtained to form a continuous DNA fragment. Primer positions and amplification directions are shown with black arrows.

riod of the virus [7]. The genome sequence and genomic structure information on the virus is limited, so molecular diagnosis and genetic vaccines against duck viral enteritis are not yet available.

All herpesvirus genomes are linear, double-stranded DNA molecules, which vary in size, sequence arrangement and base composition [1]. The herpes simplex virus (HSV) genome consists of 2 covalently linked components, designated long (L) and short (S), with each component consisting of unique sequences bracketed by inverted repeats [8]. DEV has a genome of approximately 180 kb [9]. The complete gene sequences of the unique long (UL) region have been sequenced and determined in our laboratory [10–14], and several other genes in the unique short (US) region have also been reported [15–17]. The results demonstrate that DEV has a similar, but not the same, genomic organization as other herpesviruses. In general, DEV has a closer evolutionary relationship to members of Mardivirus than to other herpesviruses [10–14]. Like other herpesviruses, DEV has a large DNA genome, but little is known about the genomic structure and the sequences in the terminus of the unique region.

Here we report on a 13,947-bp sequence located in the leftmost end of the DEV UL region that includes 5 open reading frames (ORFs) designated infected-cell polypeptide 4 (*ICP4*), *ORF1*, *ORF2*, *LORF2* and *LORF3*. These results further enrich our knowledge of the DEV genome and lay the foundation for the taxonomic classification of DEV within Herpesviridae. Furthermore, this gives us additional information about the genomic structure of DEV and aids our understanding of viral replication and infection. These results could also provide the basis for the development of sequence-based prophylactic, therapeutic and diagnostic techniques for the control of DEV.

Materials and Methods

Virus Propagation

The DEV clone-03 strain was purified from a commercial Chinese DEV vaccine using a plaque assay as described previously [13]. The virus was propagated in chicken embryo fibroblasts growing in Dulbecco's minimum essential medium (Invitrogen, Carlsbad, Calif., USA) with 8% FBS, and were harvested when the cytopathic effect reached 80%. After 3 freeze-thaw cycles, cell lysates were confirmed by electron microscopy and PCR as described previously [11]. The viral DNA was extracted from cell lysates as previously described [13].

DNA Extraction, PCR and Sequencing

'Targeted gene walking PCR' was used to amplify the DEV genome fragment, as previously described [7, 18] and summarized in figure 1. Four nonspecific walking primers, N1, N2, N3 and N4 [14], were used. Briefly, a pair of primers, L1 and L2, were designed based on the partial DEV *UL1* sequence from GenBank (GenBank accession No. EF449516) [12]. The 1st fragment (F1) was amplified in the direction opposite to the *UL1* gene. The targeted primer L3 and internal primer L4 were designed based on the newly generated fragment and used to amplify the neighboring gene fragment with 1 of the 4 nonspecific primers. Similarly, primers L5-L6, L7-L8, L9-L10 and L11-L12 were designed and used in subsequent PCR amplifications. The primers used are listed in table 1.

PCR was in 25- μ l reaction volumes, as previously described [13]. Reactions were 95° for 5 min, followed by 30 cycles of 94° for 1 min, 50° for 1 min and 72° for 2 min, ending with an elongation at 72° for 10 min. PCR products were analyzed on 0.8% agarose gels and sequenced directly or cloned into a pMD18-T vector (TaKaRa, Dalian, China) according to the manufacturer's instructions, before sequencing. Each fragment was sequenced at least 3 times from different PCR products.

ORF Determination and Sequence Analysis

The obtained sequences were compiled using Gene Runner (Version 3.00, Hastings Software Inc., Hudson, N.Y., USA) and compared to the sequence counterparts of Marek's disease virus (MDV)-1, HSV-1 and varicella-zoster virus (VZV). The same pro-

gram was used to detect ORFs encoding proteins of 75 amino acids or more that had methionine start codons. The predicted ORFs and flanking sequences were evaluated for coding potential by detecting promoters (http://www.fruitfly.org/seq_tools/promoter.html) [19], TATA boxes (<http://motif.genome.jp/>) and transcription terminal signals (http://www.ch.embnet.org/software/tmpred_form.html). Gene splicing (http://www.fruitfly.org/seq_tools/splice.html) was predicted using an online service. Signal peptides (<http://www.cbs.dtu.dk/services/SignalP/>), transmembrane regions (<http://bioinfo.si.hirosaki-u.ac.jp/~ConPred2/>), N-linked glycosylation sites (<http://www.cbi.dtu.dk/services/NetNGlyc/>), and serine, threonine and tyrosine phosphorylation sites (<http://www.cbs.dtu.dk/services/NetPhos/>) were predicted using online searches. Homologue searches were conducted using BLAST [20], and 21 reference strains in the subfamily α -Herpesvirinae were selected for phylogenetic analysis using the MEGALIGN program in Lasergene (DNASar, Madison, Wisc., USA) with CLUSTAL W multiple alignment and weight matrix Gonnet 250 [21].

GenBank Accession Numbers

The DNA sequence of the 13,947 bp from the DEV clone-03 genome has been deposited in GenBank under the accession No. FJ932653.

Results

ORF Determination and Gene Arrangement of Partial DEV Clone-03

We obtained 6 overlapping fragments, for a total of 14,376 bp, from the DEV clone-03 genome anchored by the *UL1* gene. Five ORFs were predicted in the sequence, and designated *ICP4*, *ORF1*, *ORF2*, *LORF2* and *LORF3*, with gene arrangement 5'-*ICP4-ORF1-ORF2-LORF2-LORF3*-3' (fig. 1). This is different from the arrangement of homologues in MDV-1, HSV-1, ILTV and VZV (fig. 2). *ICP4*, *LORF2* and *LORF3* have the same transcription orientation as *UL1*, while *ORF1* and *ORF2* are opposite to the *UL1* orientation, and *LORF3* is separated from *UL1* by a 513-bp interval. While *ICP4* is conserved among homologues of α -herpesviruses, *LORF2* and *LORF3* homologues were found only in the Mardivirus, including MDV-1 (fig. 2), MDV-2, and turkey herpesvirus (HVT). *ORF1* and *ORF2* are unique in DEV. The relative position of the motifs is given in table 2, based on the 14,376-bp sequence.

Molecular Characteristics of DEV LORF3

The DEV *LORF3* is separated from *UL1* by a 513-bp interval and is predicted to encode a protein of 480 amino acids with no homologous proteins in α -herpesviruses, other than the products of MDV-1 *LORF3*, MDV-2 *ORF4* and HVT *LORF2*. The protein encoded by DEV

Table 1. Sequence of oligonucleotide primers used for PCR amplification

Primer	Sequence (5'-3')	Position ^a
L1(-)	CCCAGAAATCGCAAGGTTA	14358-14376
L2(-)	CCATTACGACTAGGATCACG	14097-14116
L3(-)	GTACGTGTTATGAGCTCTGC	13117-13136
L4(-)	GTTTCATCAAACTATCCGAG	13011-13030
L5(-)	GCGAAGCATCGGTGTTGTTC	9672-9691
L6(-)	TCTGTCTGTCCTGTAACCGT	9584-9603
L7(-)	TGGCAACATGCCAACTCC	7152-7171
L8(-)	CTTGGGCTGTTGCCAACC	6986-7003
L9(-)	AGTTCCCAGGTTATTGTAGG	4761-4780
L10(-)	GAATGTCATCTGGGGTCTC	4638-4656
L11(-)	CGAACCGTCACAGTCTGCAG	2421-2440
L12(-)	GCGGGGAAAAGCATGTAAGC	2301-2320
N1	TATAGGTTT(C/A)TGTT	NP
N2	CTTTTGAGCTG	NP
N3	GAATGTGA(A/G)AA	NP
N4	CATGTCTGCCGA	NP

NP = No prediction.

^a Position is according to the sequence of the whole fragment of 14,376 bp.

LORF3 is predicted to belong to the DUF1509 superfamily, which contains several uncharacterized viral proteins from Mardivirus (fig. 3). Two potential N-linked glycosylation sites, ⁸⁵NRT⁸⁷ and ³⁹⁶NES³⁹⁸, were detected in the DEV *LORF3* protein, as well as a nuclear localization signal, ⁴¹²RRRVDRSTT⁴²⁰, at its N terminus.

A Lipase-Like Protein Encoded by DEV LORF2

Similar to examples in Mardivirus, the DEV *LORF2* is predicted to be a spliced gene [22]. It encodes a putative protein of 725 amino acids that is similar to other viral proteins and known eukaryotic lipases. Interestingly, the *LORF2*-encoded proteins show similarity to a gene product in avian adenovirus 8 (fig. 4a). A possible signal peptide was detected in the N-terminal domain that would anchor in the membrane without cleavage, and a possible transmembrane domain was detected from the amino acids 29-49 (fig. 4b). In addition, DEV *LORF2* contains predicted Trp-Asp (WD) repeats between amino acids 658 and 672, which have not been reported in other herpesviruses, although its function in the *LORF2* protein is unclear.

Our results showed that the putative protein of DEV *LORF2* and those of Mardivirus examples showed similarity to proteins involved in lipid metabolism and transport (fig. 4a). The common motif Gly-X₁-Ser-X₂-Gly (X₁

Fig. 2. Sequence arrangements of MDV-1, ILTV, HSV-1, VZV and DEV. The schematic representation shows layouts of several α -herpesviruses with unique DEV DNA sequences (shown as single lines) and repeat elements (shown as black boxes). The corresponding regions of the fragments isolated in this study are shown below the genomic structures. Genes are shown with bold arrows and genes from this study are in red (grey in the printed version). White bars indicate interval sequences between genes. Asterisks indicate possible splice sites. The grey box in the DEV genome indicates the regions whose existence is uncertain.

Table 2. Core promoter searched in the neural network and polyadenylation signals predicted by POLYADQ

Gene	Promoter location	Promoter score	TATA sequence	TATA location	TSS location	Poly(A) sequence	Poly(A) location	Poly(A) score
<i>ICP4</i>	55–104	0.88	TTTGTAATAAT	63–72	95	AATAAAA	5160–5165	0.317475
<i>ORF1</i>	5419–5468r	0.93	ATATAAAGCGGTAGT	5443–5457r	5428	NP	NP	NP
<i>ORF2</i>	NP	NP	NP	NP	NP	NP	NP	NP
<i>LORF2</i>	7440–7489	0.98	TTATAAAGACGCGGT	7449–7463	7480	AATAAAA	9820–9825	0.287518
<i>LORF3</i>	11682–11731	0.98	NP	NP	11722	AATAAAA	13747–13752	0.304017

r = Reverse direction; NP = no prediction.

and X_2 are variable residues) in lipase was detected at ²⁶⁹GHSIG²⁷³ in DEV LORF2, with the S²⁷¹ being the 1st amino acid of the catalytic triad responsible for hydrolysis of a lipid [23]. The 2nd amino acid of the catalytic triad is D²⁹⁹ in the N terminus of DEV LORF2, and the 3rd is an H in other lipases (data not shown) [23]. From the above analysis, we speculate that DEV LORF2 might encode a transmembrane protein that could either bind or metabolize lipids, similar to the role of LORF1 in HVT [22].

A Tandem Repeat Region in the DEV Genome

A region of 320-bp sequences containing 8 tandem copies of a 40-bp repeat (fig. 5) was detected downstream

of DEV *ICP4*. The structure of the repeat sequences is similar to α -type sequences in MDV-1 [24]. Hence, we designated the tandem repeats in the DEV genome to be α -type-like sequences.

Two Unique ORFs in DEV

Two unique ORFs, *ORF1* and *ORF2*, which do not have counterparts in other herpesviruses, were observed between *ICP4* and *LORF2*. *ORF1* is predicted to encode an 81-amino-acid protein, with 1 predicted transmembrane region at the N-terminal 14–34 amino acids, and 1 potential N-glycosylation site, ⁵⁵NST⁵⁷. In addition, 2 possible Ser phosphorylation sites were predicted at positions 40 and 69. *ORF2* encodes a putative protein of 121 amino

Fig. 3. Multiple homology alignments based on DEV LORF3 and typical strains from Mardivirus in the subfamily α -Herpesvirinae.

255	FLNLDVEQTKLHCIGHSIGYLCISAICRDFSSTTGAKCKRLVAIDPS	DEV
293	SAMP--SQLRIHCMGHSLSYVCGSICRQYHSLMSGICKGILGINP-	MDV-1
295	TAVP--PHSQVRMCHSLGMYTCGSICRQYNSVGTTRCKSILSIDP-	MDV-2
274	YAVP--SNMGIQCIHSLGSYVCGSICRQYYYTGKGRCKRILGINA-	HVT
153	RTQLGYSPSNVHVIHSLGSHAAGEAGRRTN---GAIGRITGLDP-	FA8
143	EKFNYPNPNVHLLGYSLGAHAAGIAGSLTK---KKVNRITGLDP-	CHICK
137	QSSFDYSPSNVHIIHSLGSHAAGEAGRRTN---GAVGRITGLDP-	1HPL_A
137	QSAGFYSPSNVHVIHSLGAHAAGEAGRRTN---GTIGRITGLDP-	1LPB_B
137	STSLNYPENVHIIHSLGAHTAGEAGKRLN---GLVGRITGLDP-	1GPL
138	STEMGYSPENVHLIHSLGAHVVEAGRRLE---GHVGRITGLDP-	1BU8_A
137	KSSLGYSPSNVHVIHSLGSHAAGEAGRRTN---GTIERITGLDP-	PIG
a		

Fig. 4. Structure prediction for the DEV LORF2 protein. **a** DEV LORF2 homologues. Sequences in red (grey in the printed version) indicate the common lipase motif. Stars indicate the residues which form the catalytic triad in functional lipases. **b** Transmembrane region in the LORF2 protein. The ConPredII server predicted 1 N-terminal transmembrane region at 29–49 (i-o). The catalytic domain may be on the outer side of the membrane; functional structure is in **a**.

Fig. 5. Schematic representation of the repeat region between ICP4 and LORF2 in the DEV genome. A set of tandem repeats was detected from 6,567 to 6,886 bp, containing 8 copies of a 40-bp sequence. Underlining indicates the sequences flanking the repeat region, which is shown with black bars. The model of the repeat region is shown below. Pink bar (grey in the printed version) represents the 40-bp repeat.

Fig. 6. Analysis of DEV ICP4 with those of other viruses within the subfamily α -Herpesvirinae. **a** Conserved domains in the N-terminus of DEV ICP4.

acids. The gene overlaps the 39-bp sequence in the tandem repeat region. One potential N-glycosylation site, ¹¹NIS¹³, and 1 potential Ser⁴⁷ phosphorylation site were observed in ORF2.

Molecular Characteristics of DEV ICP4

DEV ICP4 is predicted to encode a 1,626-amino-acid protein that is highly conserved among different α -herpesviruses (fig. 6a, b). DEV ICP4 had a G + C content of 53.4%, which was lower than that of HSV-1 [25], but slightly higher than MDV-1 ICP4 [26]. A number of potential transcriptional regulatory sites occur within this region. A 16-bp sequence, ATCGTCGTCCTCCTCC, in DEV ICP4 has striking homology to the consensus-binding sequence in ICP4 of the HSV (ATCGTCN4YCGRC, where 'Y' is pyrimidine and 'R' is purine), located in the transcription initiation site of ICP4 in HSV-1 [27]. An 8-bp sequence, ATGCAATC, in ICP4 of DEV showed high homology to the octamer ATGCAAAT [26]. An inverted octamer, ATTTGCTT, located in the 5' coding region of

DEV ICP4, showed high similarity to ATTTGTCAT [26]. Other motifs, including 2 SP1-binding sites (GRG-GCRGGGW) and 2 CCAAT/enhancer protein α -sites were also found in DEV ICP4. In addition, 2 zinc finger interaction domains and 7 potential N-linked glycosylation sites were found in DEV ICP4.

Similar to HSV-1, 5 conserved intramolecular domains were observed in DEV ICP4 (fig. 6c) [25]. Region 1 in ICP4 of DEV comprises residues 1–561 and is 147 residues longer than the corresponding region in HSV-1, but with a shared, conserved serine-rich tract. However, the serine-rich tract was followed by acidic residues in HSV-1 [25] and flanked by proline- and basic-amino acid-enriched flanking regions in MDV-1 [26], neither of which was found in DEV ICP4. Region 2 of DEV ICP4 consists of residues 562–738 and is one of the most conserved regions among the 5 domains. Region 3 (residues 739–1124) is not conserved in the herpesviruses. Region 4 (residues 1125–1538) is highly conserved in ICP4 of viruses in subfamily α -Herpesvirinae. Region 5 is the smallest, com-

Fig. 6. Analysis of DEV ICP4 with those of other viruses within the subfamily α -Herpesvirinae. **b** Conserved domains in the C-terminus of DEV ICP4.

prising residues 1539–1626. It has little homology to ICP4 from other viruses in subfamily α -Herpesvirinae. A phylogenetic-rooted tree constructed from alignment of the putative ICP4 protein showed DEV ICP4 had a close evolutionary relationship to members of Varicellovirus (fig. 6d).

Discussion

The whole genome sequence and organization of DEV is unknown, with genome information limited to individual ORFs. Generally, the arrangement of partial genes in the UL region of the DEV genome is collinear with homologues in HSV-1, and phylogenetic analysis of most

genes in the DEV UL region has shown that DEV has a close relationship to members of Mardivirus [10–14]. In the present study, we report sequences in the leftmost region of DEV UL and the molecular characteristics of predicted genes in this region. The DEV gene arrangement and content in this region (fig. 2) showed differences from other genes in previously reported UL regions [10–14].

In MDV-1, the UL is flanked by terminal UL repeats (TRL) and internal UL repeats (IRL), respectively [24]. The MDV-1 *LORF3* and *LORF2* are located at the leftmost region of UL, with terminal UL repeats adjacent to it. However, neither *LORF2* nor *LORF3* have been detected in other herpesviruses, except for members of Mardivirus. In this study, we found that DEV also has 2 LORFs,

Region 1	
MAKTQYPDFDLDIEAAASADCFSMSNISSIDIQHGIQGPNTPLIDESQYLLDDSRCLIDNMNAAAMAATTKMETVNRGDGRILYLDYMGPLSDAAFTKPLQPNPGPTEIDSSDEEPTT	120
-----MASEN-----EQRPGSPGP--TD-----	16
CGNLNPIKEATPPKMQSPSQGMRQKGDTPPKGVVRAEGGPKTGSDLQKKRKSEYKLRGGKTKMAKPGVNKTQVAAAGGKNYAGIADVIDIGESSEECSDNGGAAVGVKKPRSATAGRPSA	240
---G---PPPTP---SPD---RDERGALGWVETEEGGDDPDHDPDHPHDLDDAR-----RDG--R--APAAG-----TDAGEDAGDAVSFRQLALLASMV	91
PDKPIASPARSPATAAPPKNTARRHRSPSSASPPPARCDSGRMTPTKSSPTSVTSAGWYRSPARSPTQGVVERSGPPSTSPRSPRGRSPPALPIVSCCSPAASPRRSVDRGRSRSSS	360
EEAVRTIPTPDPAASPPPTPAFRAD-----DDG-----DEYDDAADAGGDRAPAGREREAPLRGAYPDPTDRLSPR---PPAQPPR-----RRRHGRRRPSASS	179
SSSTSSSSSSSDGSSVSRGPSRDPPTPKRPPSPFIPSAFKRRPAMQRKDNAEQRSFKTDLSKLFSQQLTHPPTSNTAASKATGTVKRKDKQFVRRAVQQEETEQQFPQTMQLAN	480
TSSDSGSSSSSSASSSSSSDEDEDDGDNDADAREARAVGRGPSS--AAPAAPGRTPPPPGPPPLSEAPKP--RAAARTPAASAGRIERRRARAAVAGRDATGR-----FTAG	286
Region 1 <--> Region 2	
SPRESTQYQHPSSDDTIDLVSCEPPPAANNAFYQRRRGRGRRRPSAMKRVSSSSAGAESCMSIGVPEPMTMNGPLMTNGDPWPGSEPPPPGKVRYGGTGDYRQGLWDEPEIRYARERYA	600
QPR--R-----VELDADATSG--AFYARYRDG-----YVSGEPWPGAGPPPPGRVLYGGLGDSRPLWGAPEAEERRRFE	353
ESRGPVAVFVGEEMGSSQYDALVKALYSNSFDAMGWIQCNKITGEDLLLEQVCQKQCNLRGGTSGSFFITGSVARPLPHIGDAMAAGNALWGLPHVAAAVAMSRRYDKVQKQFLQMSLR	720
ASGAPAAVWAPELGDAAQYALITRLLYTPDAEAMGNLQNPVVPGDVALDQACFRIS--GAARNSSSFFITGSVARPLPHIGYAMAAGRFGWGLAHAAAVAMSRRYDRAQKGFLLTSLR	471
Region 2 <--> Region 3	
RAYAYMLFPAKAVDASLTDDISSIYRRMNAVRRSMRETEGLQTVTVRGD---SQIEPQNTQSTDHLATEQSAADPDFCDPEHASDYSTEELADACVLACNAVIEVLLCEFGG----	833
RAYAYPLARENAALTGAAGSPGAGADDEGVAVVAAAAAPGERAVPAGYGAAGTILALGRLSAAAPSPAGDDPDARHADADDAGRRAGQAGRAVECLAACRGILEALAEFGDGLAA	591
---AGNIPGISSELEDAACPS--AQLHRHHVQLLSWIKDLKLLPICINAMRLRGELSSSQGIESCVMALRAVGTIVSTVRPFVNFVSDRARRYSAWNLTALFAPSNGNSAKINGAK	947
VPGLAGARPASPPRPEGAGPASPFPHPADAPRLRAWLRELFVRDALVLMRLRGDLRVAGGSEAAVAARAVSLVAGALGP-----ALPR-----D-----PRLPSSAAAAAAD	691
LSWVLSHLTPKDKQTSPLVWPIASKTIDKMTSGSGIQDGGSKRKDQTRSVACQTNCHKQDREKIDGANRVNKQCCQIPGSEERANGKEQPRRFASKRKSMPVQALPRDGDNGADDV	1067
LLFDNQSLR-----PLLAAS-----APDAADALAAASAAPREGKRKSPGARP-----PG-----GGGPR-----	746
Region 3 <--> Region 4	
HGPPQTKRHSVPSNMSGAYATDDTAPKDRDRPDGLPKHQNIQNSKKDNQTELPWTLPSVKALGMPQGPHFRGGYRRVPTGITHTPLPTESARSTYCPDIVAEILLDDPMFPNWKPA	1187
--PKTKKSGADAPGSDARAPLAPAP-----PSTPPGPEPAPAQPAAPRAAAQARPRPVALSRRTAEGPDLGGWRRQPPGPSHTAAPAAALEAYCSFRAVAELTDHPLFFVPRWA	859
LMFDPEALARVAARRGG-----SHVEGRLYGLRVTDFLRKQCAWMNQIPDPEDVKVIMCYCLVGEELSAAPVASCMPMWSDEKGLSFLLAALCDRLCSYGS	1286
LMFDPRALASIAARCAAPAAQAACGGDDDDNPHPHGAAGGRLFGPLRASGLRMAAWMRQIPDPEDVRVVLYSPLPGLDLADGGASGGPPEWSAERGGSLCLLAALANRLCGPDT	979
HAWAGNWTGVPDISALNSRGVLLSVRDIFGAGAVEYLCRLGSAHKRLIFLDTIDTVWDPADGPAISSYHSYIKATITPQAHGAHRWPDCEAMSRVMTSRGIVGPGVFARVEAAFAQL	1406
AAGNWTGAPVVSALGAQGVLLSLTRDLAFAGAVEFLGLASAGDRRLIVVTVRACDWPADGPAISRQHAYLACELLPAVQCAVRWPAARDLRLTVLASGRVFGPGVFARVEAAHARL	1099
YFNEPPLQLMRGDNVYVETKAGRTTRVMTPREYRGLVLPQYD-RKDMAFQADGLSIRGPDFLEGAAYNHRAANRWGLGAPLRPLYLTESRKHPFETDDIPEVVRYSCAAALLEPQM	1525
YDAPPLRLCRGGNVRYRVTRFGPDTPVMSPREYRRAVLPAIDGRAAASGTTDAMAPGAPDFCEEAHSHRACARWGLGAPLRVYVALGREAVRAGPARWRGPRRDFCARALLEPDD	1219
Region 4 <--> Region 5	
EGNVCVNPNTIQRGLPTITWELCHGSSRATSIIEAGDMKQAIIPPEQNSNEDDVTITAVQPSITPMPKAPHKIKIEILSDGDDSGDEGNPYAGACRDF	DEV
D-----A-PPLVLRGD-----DDDDGPGALPPAKPGIRWASATGRSGTVLAAAGAVEVLGAEGALATPPRREVVDWEGAWDE--DDGGAFFGEGDVL	HSV-1

c

Fig. 6. Analysis of DEV ICP4 with those of other viruses within the subfamily α -Herpesvirinae. **c** Alignment between DEV ICP4 and HSV-1 IE175 amino acid sequences. DEV ICP4 is the upper line and HSV-1 is the lower, with identical aligned residues in red (grey in the printed version). Introduced gaps are shown as hyphens. Residue numbering for ICP4 is on the right, and the boundaries of the proposed regions are shown by arrowheads. This alignment was produced by the MEGALIGN (DNASTar) program, which is associated with BLAST [20].

LORF3 and *LORF2*, which are similar to those in *Mardi-virus*, indicating a close relationship between avian herpesviruses.

A repeat region, named α -type sequences, is located at the genomic termini and at the IRL/IRS (internal US repeats) of MDV-1 [24]. These sequences contain 7 tandem copies of a 60-bp repeat associated with the long direct repeat, and 64 tandem copies of a 6-bp repeat associated with the short direct repeat [28–30]. In DEV, a similar tandem repeat (α -type-like sequences) was found associated only with the long direct repeat, and was similar to that found in MDV-1 to the left of *LORF2*. The existence of α -type-like sequences might imply that *LORF2* is at the end of the UL region in the DEV genome.

It is reported that the *ICP4* gene is present in 2 copies in HSV-1, in the repeated sequences bordering the US region of the genome [31]. The same structure has been found in other α -herpesviruses. We found that DEV *ICP4* is linked to *LORF2* in the UL region. Recently, Li and Huang [32] submitted sequences of this region to the GenBank database (GenBank accession No. EU082088.1). HSV DNA circularizes quickly, very early after infection, and in the absence of viral protein synthesis [33, 34]. Hence, HSV ICP4 has been located at the terminus of terminal US repeats (TRS) linked to the L region. This may also be the case for DEV ICP4. The DNA we used as the PCR template might have contained part of the circularized DEV genome. If this was the case, the DEV genome

Fig. 6. Analysis of DEV ICP4 with those of other viruses within the subfamily α -Herpesvirinae. **d** Evolutionary relationships of DEV ICP4 within the subfamily α -Herpesvirinae. Sequence distances are indicated with a scale calculated using the Gonnet 250 matrix in LASERGENE [21].

might have no obvious TRL region at the end of the UL region, similar to the VZV [35] and ILTV genomes (fig. 2). Thus, the DEV *ORF2* that overlaps the tandem repeat region might be the junction of the terminal US repeats TRS and UL regions. Alternatively, this structure of the DEV genome may be unique among herpesviruses.

G + C content analysis of herpesvirus strains indicated that the avian herpesvirus *ICP4* has a lower G + C content than nonavian *ICP4* homologues; DEV *ICP4* has a similar G + C content, and shows sequence similarity to *ICP4* from other avian herpesvirus. The length of *ICP4* in nonavian herpesviruses selected for this study (fig. 6d) was from 1,100 to 1,500 amino acids, but *ICP4* from Mardivirus is over 2,000 amino acids [24, 36]. DEV *ICP4* is predicted to be 1,626 amino acids, similar to that of ILTV. An N-terminal extension of 800–900 amino acids is reported for MDV-1 [26], MDV-2 and HVT *ICP4* [22], which was not found in DEV *ICP4*. Interestingly, the phylogenetic tree based on the putative *ICP4* protein sequence clustered DEV with Varicellovirus. In addition, one of the predicted N-linked glycosylation sites at the C terminus of DEV *ICP4* was conserved in HSV-1, HSV-2, VZV,

pseudorabies virus (PRV), Equine herpesviruses (EHV)-1 and -4. In summary, we conclude that the composition of DEV *ICP4* puts it between the avian and nonavian herpesviruses, which might indicate that DEV *ICP4* is an intermediary, but as these postulations are from sequence analysis only, further investigations are required.

References

- 1 Weir JP: Genomic organization and evolution of the human herpesviruses. *Virus Genes* 1998;16:85–93.
- 2 Davison AJ: Herpesvirus genes. *Rev Med Virol* 1993;3:237–244.
- 3 Minson AC, Davison AJ, Desrosiers RC, Fleckenstein B, McGeoch DJ, Pellett PE, Roizman B, Studdert DMJ: Herpesviridae; in Van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR, Wickner RB (eds): *Virus Taxonomy*. New York, Academic Press, 2000, pp 203–225.
- 4 Davison AJ: The genome of salmonid herpesvirus 1. *J Virol* 1998;72:1974–1982.

- 5 Fauquet CW, Mayo MA, Maniloff J, Desselberger U, Ball LA: Virus taxonomy: Eighth report of the international committee on taxonomy of viruses. San Diego, Elsevier Academic Press, 2005.
- 6 Davison S, Converse KA, Hamir AN, Eckroade RJ: Duck viral enteritis in domestic muscovy ducks in Pennsylvania. *Avian Dis* 1993;37:1142–1146.
- 7 Burgess EC, Ossa J, Yuill TM: Duck plague: a carrier state in waterfowl. *Avian Dis* 1979;23:940–949.
- 8 Wadsworth S, Jacob RJ, Roizman B: Anatomy of herpes simplex virus DNA. II. Size, composition, and arrangement of inverted terminal repetitions. *J Virol* 1975;15:1487–1497.
- 9 Gardner R, Wilkerson J, Johnson JC: Molecular characterization of the DNA of anatis herpesvirus 1. *Intervirology* 1993;36:99–112.
- 10 An R, Li H, Han Z, Shao Y, Liu S, Kong X: The UL31 to UL35 gene sequences of duck enteritis virus correspond to their homologs in herpes simplex virus 1. *Acta Virol* 2008;52:23–30.
- 11 Chen S, Han Z, Shao Y, Liu S, Kong X: Molecular characterization of UL6 gene of duck enteritis virus. *Bing Du Xue Bao* 2006;22:391–396.
- 12 Li H, Liu S, Han Z, Shao Y, Chen S, Kong X: Comparative analysis of the genes UL1 through UL7 of the duck enteritis virus and other herpesviruses of the subfamily alphaherpesvirinae. *Genet Mol Biol* 2009;32:121–128.
- 13 Li H, Liu S, Kong X: Characterization of the genes encoding UL24, TK and gH proteins from duck enteritis virus (DEV): a proof for the classification of DEV. *Virus Genes* 2006;33:221–227.
- 14 Liu S, Chen S, Li H, Kong X: Molecular characterization of the herpes simplex virus 1 (HSV-1) homologues, UL25 to UL30, in duck enteritis virus (DEV). *Gene* 2007;401:88–96.
- 15 Hu Y, Zhou H, Yu Z, Chen H, Jin M: Characterization of the genes encoding complete US10, SORF3, and US2 proteins from duck enteritis virus. *Virus Genes* 2009;38:295–301.
- 16 Zhao Y, Wang JW, Liu F, Ma B: Molecular analysis of US10, S3, and US2 in duck enteritis virus. *Virus Genes* 2009;38:243–248.
- 17 Zhao Y, Wang JW, Ma B, Liu F: Molecular analysis of duck enteritis virus US3, US4, and US5 gene. *Virus Genes* 2009;38:289–294.
- 18 Parker JD, Rabinovitch PS, Burmer GC: Targeted gene walking polymerase chain reaction. *Nucleic Acids Res* 1991;19:3055–3060.
- 19 Reese MG, Harris NL, Eeckman FH: Large scale sequencing specific neural networks for promoter and splice site recognition; in Hunter L, Klein TE (eds): *Biocomputing: Proceedings of the 1996 Pacific Symposium Co.*, Singapore, World Scientific Publishing, 1996; pp 2–7.
- 20 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Bio* 1990;215:3389–3402.
- 21 Liu S, Chen J, Chen J, Kong X, Shao Y, Han Z, Feng L, Cai X, Gu S, Liu M: Isolation of avian infectious bronchitis coronavirus from domestic peafowl (*Pavo cristatus*) and teal (*Anas*). *J Gen Virol* 2005;86:719–725.
- 22 Kingham BF, Zelnik V, Kopacek J, Majerciak V, Ney E, Schmidt CJ: The genome of herpesvirus of turkeys: comparative analysis with Marek's disease viruses. *J Gen Virol* 2001;82:1123–1135.
- 23 Dodson GG, Lawson DM, Winkler FK: Structural and evolutionary relationships in lipase mechanism and activation. *Faraday Discuss* 1992;9:95–105.
- 24 Tulman ER, Afonso CL, Lu Z, Zsak L, Rock DL, Kutish GF: The genome of a very virulent Marek's disease virus. *J Virol* 2000;74:7980–7988.
- 25 McGeoch DJ, Dolan A, Donald S, Brauer DH: Complete DNA sequence of the short repeat region in the genome of herpes simplex virus type 1. *Nucleic Acids Res* 1986;14:1727–1745.
- 26 Anderson AS, Francesconi A, Morgan RW: Complete nucleotide sequence of the Marek's disease virus ICP4 gene. *Virology* 1992;189:657–667.
- 27 Faber SW, Wilcox KW: Association of the herpes simplex virus regulatory protein ICP4 with specific nucleotide sequences in DNA. *Nucleic Acids Res* 1986;14:6067–6083.
- 28 Dolan A, Jamieson FE, Cunningham C, Barnett BC, McGeoch DJ: The genome sequence of herpes simplex virus type 2. *J Virol* 1998;72:2010–2021.
- 29 Kishi M, Bradley G, Jessip J, Tanaka A, Nonoyama M: Inverted repeat regions of Marek's disease virus DNA possess a structure similar to that of the α sequence of herpes simplex virus DNA and contain host cell telomere sequences. *J Virol* 1991;65:2791–2797.
- 30 Telford EA, Watson MS, McBride K, Davison AJ: The DNA sequence of equine herpesvirus-1. *Virology* 1992;189:304–316.
- 31 Rixon FJ, Campbell ME, Clements JB: The immediate-early mRNA that encodes the regulatory polypeptide Vmw 175 of herpes simplex virus type 1 is unsplliced. *EMBO J* 1982;1:1273–1277.
- 32 Li YF, Huang B, Ma XL, Wu J, Li F, Ai W, Song MX, Yang HC: Molecular characterization of the genome of duck enteritis virus. *Virology* 2009; doi:10.106/j.virol.200906018.
- 33 Garber DA, Beverley SM, Coen DM: Demonstration of circularization of herpes simplex virus DNA following infection using pulsed field gel electrophoresis. *Virology* 1993;197:459–462.
- 34 Poffenberger KL, Roizman B: A noninverting genome of a viable herpes simplex virus 1: presence of head-to-tail linkages in packaged genomes and requirements for circularization after infection. *J Virol* 1985;53:587–595.
- 35 Davison AJ: Structure of the genome termini of varicella-zoster virus. *J Virol* 1984;65:1969–1977.
- 36 Afonso CL, Tulman ER, Lu Z, Zsak L, Rock DL, Kutish GF: The genome of turkey herpesvirus. *J Virol* 2001;75:971–978.