Journal of General Virology (2007), **88**, 1080–1096 **DOI** 10.1099/vir.0.82600-0

Comparative full-length sequence analysis of oncogenic and vaccine (Rispens) strains of Marek's disease virus

Stephen J. Spatz,[1] Lawrence Petherbridge,[2] Yuguang Zhao[2] and Venugopal Nair[2]

[1]Southeast Poultry Research Laboratory, Agricultural Research Service, United States Department of Agriculture, Athens, GA 30605, USA

[2]Institute for Animal Health, Compton, Berkshire RG20 7NN, UK

Correspondence
Stephen J. Spatz
sspatz@seprl.usda.gov

Received 27 September 2006
Accepted 4 December 2006

The complete DNA sequence of the Marek's disease virus serotype 1 vaccine strain CVI988 was determined and consists of 178 311 bp with an overall gene organization identical to that of the oncogenic strains. In examining open reading frames (ORFs), nine differ between vaccine and oncogenic strains. A 177 bp insertion was identified in the overlapping genes encoding the Meq, RLORF6 and 23 kDa proteins of CVI988. Three ORFs are predicted to encode truncated proteins. One, designated 49.1, overlaps the gene encoding the large tegument protein UL36 and encodes a severely truncated protein of 34 aa. The others, ORF5.5/ORF75.91 and ORF3.0/78.0, located in the repeat regions (diploid), encode a previously unidentified ORF of 52 aa and a truncated version of the virus-encoded chemokine (vIL-8), respectively. Subtle genetic changes were identified in the two ORFs encoding tegument proteins UL36 and UL49. Only one diploid ORF (ORF6.2/ORF75.6) present in the genomes of the three virulent strains is absent in the CVI988-BAC genome. Seventy non-synonymous amino acid substitutions were identified that could differentiate CVI988-BAC from all three oncogenic strains collectively. Estimates of the non-synonymous to synonymous substitution ratio (ω) indicate that CVI988 ORFs are generally under purifying selection ($\omega < 1$), whereas UL39, UL49, UL50, RLORF6 and RLORF7 (Meq) appear to evolve under relaxed selective constraints. No CVI988 ORF was found to be under positive evolutionary selection ($\omega \gg 1$).

INTRODUCTION

Marek's disease (MD) is a highly contagious neoplastic and neuropathic disease of chickens responsible for great economic losses to the poultry industry (Calnek, 2001). The causative agent of the disease is a mardivirus, Marek's disease virus (MDV-1), a member of the subfamily *Alphaherpesvirinae* in the family *Herpesviridae*. MDV is by far the most oncogenic herpesvirus known and represents the first of three neoplastic diseases (including hepatocellular carcinoma, caused by *Hepatitis B virus*, and cervical carcinoma, caused by *Human papillomavirus*) for which an effective vaccine has been developed (Churchill *et al.*, 1969; Pagliusi & Teresa Aguado, 2004; Prince, 1996). The disease is largely controlled through mass vaccination of chick embryos or day-old chicks with live-attenuated strains of MDV-1 (Witter, 2001b). Since their introduction in the 1970s, MDV vaccines have reduced economic losses from MD dramatically, by more than 99 % (Witter, 2001a). However, over the last 40 years of intense vaccination, virulent strains have reappeared, necessitating the continued introduction of new vaccines at fairly regular intervals (Witter, 1997). In the 1970s, MD was controlled by a vaccine derived from herpesvirus of turkeys (HVT) (Witter *et al.*, 1970), only to be replaced in the 1980s with a bivalent vaccine consisting of SB-1 (serotype 2 MDV) and HVT (Schat & Calnek, 1978). Since the 1990s, the disease has been largely controlled by vaccination with the attenuated serotype 1 strain CVI988 (Rispens *et al.*, 1972). Despite its global usage and exceptional protective efficacy, there are signs of further increases in MDV virulence, as shown by the isolation of strains with higher pathogenicity even in birds vaccinated with the CVI988 strain (Davison & Nair, 2005; Nair, 2005). Because of this, there is great concern that CVI988-based vaccination programmes may eventually fail to protect poultry in the same way that previous generations of MD vaccines have lost their usefulness. The emergence of so-called breakthrough variants may have devastating effects on the poultry industry in the future.

In order to prevent this scenario, a better understanding of the genes involved in virulence is needed. Although many unique MDV genes, such as the viral lipase homologue

The GenBank/EMBL/DDBJ accession number for the CVI988-BAC genome described in this study is DQ530348.

A supplementary table and figure are available in JGV Online.

(vLIP), virus-encoded telomerase RNA (vTR), pp38 and Meq, have been demonstrated to play a role in virulence or oncogenicity through studies on knockout mutants, it still remains to be determined whether vaccine strains have alterations in the genes encoding these factors or possibly in other genes (Fragnet *et al.*, 2003; Gimeno *et al.*, 2005; Kamil *et al.*, 2005; Lupiani *et al.*, 2004). To this end, we have embarked on a comparative genome analysis of vaccine and oncogenic strains in the hope that it will provide some clue to the multigenic nature of MDV virulence and identify CVI988 ORFs that have evolved under positive Darwinian selection as a result of immune evasion.

The genome organization of linear MDV has been reviewed elsewhere (Osterrieder & Vautherot, 2004; Silva *et al.*, 2001). In this paper, we present the sequence of the full-length bacterial artificial chromosome (BAC) clone of the CVI988 genome, designated pCVI988-BAC (Petherbridge *et al.*, 2003). The genomic structure and gene content will be compared with those of three oncogenic strains of MDV: GA, Md5 and Md11 (Lee *et al.*, 2000a; Niikura *et al.*, 2006; Tulman *et al.*, 2000).

METHODS

Construction of pCVI988. The construction of pCVI988-BAC was carried out essentially as described previously by the insertion of the F plasmid into the US2 gene (Petherbridge *et al.*, 2003). Briefly, CVI988 genomic DNA and pDS-pHA1 were cotransfected into secondary chicken embryo fibroblasts by the calcium phosphate precipitation procedure. After selection in medium containing mycophenolic acid, xanthine and hypoxanthine, genomic DNA from infected cells was electroporated into *Escherichia coli* (DH10B) cells (Invitrogen) and plated onto agar dishes containing chloramphenicol.

DNA sequencing. DNA from *E. coli* DH10B cells harbouring pCVI988-BAC was isolated by Midiprep alkaline lysis according to the manufacturer's instruction (Qiagen). Sequencing of pCVI988-BAC was initially carried out commercially. This involved the construction of a random shotgun library of pCVI988-BAC DNA in the pGEM-T vector (Promega) and sequencing of sufficient clones to obtain a sixfold coverage of the genome. To obtain the sequence of the US region of the CVI988 genome replaced by the pDS-PHA1 vector, the corresponding region was amplified by PCR using *Pfu* DNA polymerase (Promega) and wild-type CVI988 DNA. PCR products from the pCVI988-BAC DNA were also used for gap closure of the initial contigs using primers derived from the sequence. DNA sequence data were obtained from recombinant plasmids by using BigDye Terminator cycle sequencing and a model ABI-3730 XL DNA Analyzer (Applied Biosystems). Problematic regions were sequenced from clones containing inserts generated by PCR using Platinum *Taq* DNA polymerase (Invitrogen) and numerous custom primers. Regions in the genome containing highly repetitive elements (a-like sequences) were sequenced from clones by using random transposon-mediated mutagenesis (Epicentre Technologies) and transposon-specific primers. The diverse regions of the UL36 and UL48 genes were sequenced from PCR products generated in amplification reactions containing whole-genome DNA isolated from two additional CVI988 strains, Intervet p27 and ADOL BP5.

Copy-number determination of a-like sequences. Anchor PCR and restriction analysis were used to determine the number of a-like sequences within the IRL/IRS and TRS/TRL junctions. Amplification reactions contained 100 ng pCVI988-BAC, a primer specific for repeat short sequences that flanked the a-like sequence and a primer specific for either the 5′ or 3′ end of the unique long sequence (Fig. 1). PCR products (~16 kb) were digested with *Sma*I and resolved on 1× TAE agarose gels.

DNA sequence analysis. DNA sequences were assembled by using the Sequencher program (Gene Codes) and edited manually. The final sequence represents, on average, a sixfold redundancy at each base pair. Open reading frames (ORFs) and DNA regulatory sequences were identified by using various software packages that included DNASTAR, GeneMarkS (GeneProbe) (Besemer & Borodovsky, 2005), NCBI Entrez and other web-based tools. Novel ORFs were identified based on codon usage and investigated by using Spin (Staden) and CUSP (EMBOSS) programs (Staden *et al.*,

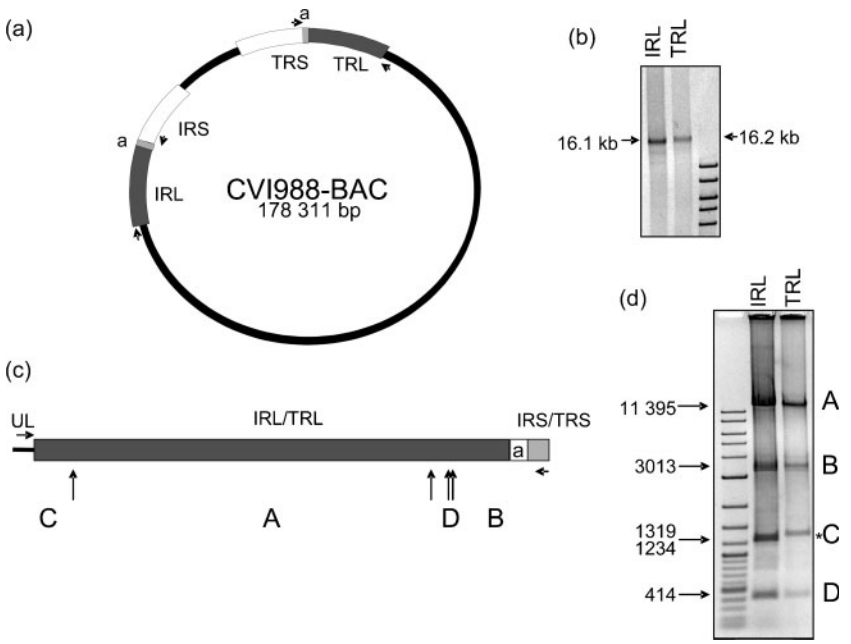

Fig. 1. Determination of a-like sequence copy number. (a) Anchor PCR scheme. Arrows denote locations of primers. (b) Ethidium bromide-stained agarose gel resolving anchor PCR products. (c) Schematic diagram of the anchor PCR products showing the positions of the *Sma*I sites with unique long (UL), internal repeat long (IRL) or terminal repeat long (TRL) sequences, a-like sequence (a) and internal repeat short (IRS) or terminal repeat short (TRS) sequences. (d) Ethidium bromide-stained agarose gel of *Sma*I-digested anchor PCR products. The asterisk indicates the length polymorphism of the C band (C), which contained unique long and repeat long (either internal or terminal) sequences. The unmarked lane to the left contains a 1 kb molecular mass marker.

2000). Homology searches were conducted by using the NCBI programs BLAST (BLASTP) and PSI-BLAST (Schaffer *et al.*, 2001) with default settings, as well as Pairwise FLAG (BMEC/ITRI). Published mRNA and cDNA data were compared with the CVI988 genome by using PROT_MAP (SoftBerry). Multiple protein sequences were aligned by using MultAlin, CLUSTAL_W and MAFFT (Corpet, 1988; Katoh *et al.*, 2005). Repeat elements were investigated by using the Tandyman program (Los Alamos National Laboratory, NM, USA). Phylogenetic trees were constructed by using MAFFT and MEGA version 3 (Kumar *et al.*, 2004) with full-length genomic sequences corresponding to the available Md11 sequencing data. Evolutionary adaptation estimations based on synonymous and non-synonymous substitutions were calculated by using a model that includes allelic mixtures and the Jukes–Cantor correction for back mutations (SYNSCAN) (Gonzales *et al.*, 2002; Jukes & Cantor, 1969).

RESULTS AND DISCUSSION

Genomic organization

The CVI988-BAC genome (GenBank accession no. DQ530348) was found to have a G+C content of 44 mol% and a length of 178 311 bp, a size similar to those of the other sequenced strains: Md5 (177 kb), Md11 (178 kb) and GA (174 kb). The genome is organized into six regions characteristic of class E alphaherpesviruses. The unique long (UL) region is 113 490 bp in length and extends from positions 14 477 to 127 966. The short (US) region is 11 651 bp in length and extends from positions 154 606 to 166 256. In order to determine the length of the repeat regions (TRL, IRL, IRS and TRS), the nucleotide sequences and copy numbers of the adjacent a-like sequences needed to be determined. PCR primers (flanking the a-like sequences) were designed based on multiple alignments of the published genomic sequences, and used in amplification reactions. It was hypothesized that two copies of the a-like sequence would be present at the TRS/TRL junction in the circular CVI988-BAC construct, as the linear form of MDV contains a-like sequences at both termini. Surprisingly, only a single PCR product of 2.3 kb was generated, suggesting that only one copy of the a-like sequence was present within both the IRL/IRS and TRS/TRL junctions. To confirm this, additional PCR experiments involving anchor PCR and restriction-endonuclease analysis were implemented. As shown in Fig. 1, large PCR products with the predicted sizes of 16.1 and 16.2 kb were produced with primers specific for the IRL and TRL regions, respectively. A *Sma*I restriction digestion of these products generated a nearly identical restriction profile. Differences were only noticeable for the bands (C) that contained unique long and repeat long sequences. The B bands, which contained the a-like sequences, were the same size for both digestions. These results and the inability to generate PCR products containing two a-like copies indicated that only one copy was present at the IRL/IRS and TRS/TRL junctions. Single copies of the a-like sequence within these junctions have also been reported for the Md11-BAC construct (Niikura *et al.*, 2006).

As it is impossible to determine the boundaries of the a-like sequence in a circular construct containing only one copy at each junction, the combined lengths of the internal repeat regions (long and short) of CVI988-BAC, including the a-like sequences, are 26 639 bp. The lengths of the terminal long (TRL) and short (TRS) repeats (14 476 and 12 055 bp, respectively) were determined based on the assumption that cleavage of concatameric replicative intermediates occurs within telomeric repeats bracketed by *pac*1 and *pac*2 sites.

In comparison to other MDV isolates, the genomic regions of CVI988-BAC are most similar to those of Md11. Overall, the sizes of the unique regions (UL and US) were comparable among all of the strains. Size differences among the genomes can be attributed to the junctions between the unique short and the repeat short regions, as well as tandem reiterations found in the repeat long regions. An evolutionary tree of the four strains of MDV-1 is presented in Fig. 2, based on the nucleotide alignment of the TRL, UL, IRL, IRS regions and part of the US region, and illustrates that both CVI988 and GA are related more distantly to a common ancestor than are Md5 and Md11.

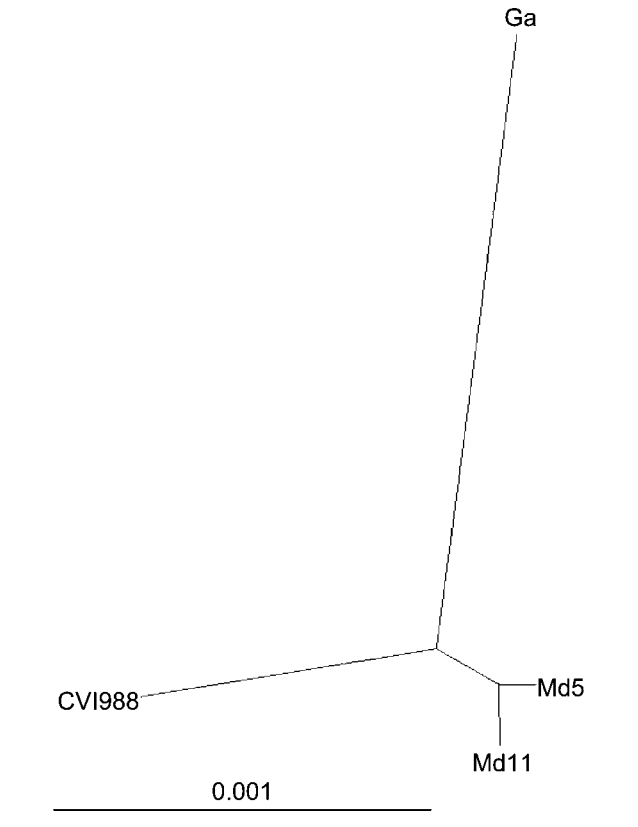

Fig. 2. Unrooted phylogenetic tree of genomes of MDV-1 strains CVI988, Md5, Md11 and GA based on the genomic segments TRL, UL, IRL and IRS and partial US sequences. Alignments were made by using the MAFFT program and a dendrogram was drawn by using MEGA3 programs.

ORFs

The genome of CVI988-BAC contains over 478 ORFs encoding proteins with $\geqslant$50 aa. Those of significance for a genomic comparison between pathotypes are presented in Table 1 and are numbered according to the decimal system (Tulman *et al.*, 2000). Novel ORFs, predicted based on the codon usage of well-characterized mardivirus genes, were named according to their first 4 aa (Table 1b). A homology comparison of each ORF from the four strains is also presented in Table 1, along with the putative functions of many of these ORFs based on protein binding partners and similarities to homologues belonging to other alphaherpesviruses (Niikura *et al.*, 2004; Nishiyama, 2004). Overall, identity is typically >98 % at the amino acid level for non-novel ORFs (Table 1a). Because of frameshift mutations, a large number of genes (LORF11, RLORF1, UL6, UL7, UL12, UL15, UL16, UL17, UL18, UL19, UL24, UL26, UL29, UL37 and UL52) within the GA genome have a lower degree of identity than homologues found in other mardiviral genomes.

Single-nucleotide polymorphisms (SNPs)

In order to identify amino acid changes important in attenuation, the nucleotide sequence of CVI988-BAC was compared with sequences from the oncogenic strains (Md5, Md11 and GA) by using the web-based programs MultAlin and MAFFT. As illustrated in Supplementary Fig. S1 (available in JGV Online), SNPs are present in ORFs throughout the genomes of MDV-1. Relative to CVI988, 109 and 99 non-synonymous substitutions were found in the ORFs of Md11 and Md5, respectively. Nearly three times as many non-synonymous substitutions, a total of 258, were found in the ORFs of strain GA. This number does not include the substitutions due to the 23 frameshift mutations (or possibly sequencing errors) found in 17 ORFs within the genome of GA. Most of the substitutions (n=188) were exclusive to strain GA and were not shared with either Md11 or Md5. Only 15 non-synonymous substitutions were exclusive to Md11 ORFs and five were exclusive to Md5 ORFs. Twenty-four of these substitutions were common to both Md11 and Md5 strains. The rest (n=70) were common to all three strains and could differentiate CVI988-BAC from the oncogenic strains of MDV (Table 2). These mutations were exclusive to CVI988 and corresponding substitutions were identical in all three of the virulent strains.

SNPs within coding regions without amino acid changes (i.e. synonymous substitutions) were also found among the strains, although the differences were not as dramatic as those noted for non-synonymous substitutions. The genes of Md11, Md5 and GA contain 54, 50 and 86 synonymous substitutions, respectively. The elevated level of non-synonymous and synonymous substitutions as well as frameshift mutations largely contributed to the phylogenetic partitioning of the GA strain (Fig. 2). To investigate this further and to determine whether evolutionary pressures are acting upon individual ORFs, the non-synonymous to synonymous substitution ratio ($\omega = d_N/d_S$) was determined for ORFs containing non-synonymous mutations (see Supplementary Table S1 in JGV Online). Most of the genes examined had a mean ω score of <1.0, indicating purifying (negative) selection or conservation of functional amino acids. For example, the UL49.5 gene in the CVI988 genome, encoding a 95 aa protein, is under strong selection for the synonymous mutations Ile^{6}, Val^{10} and Thr^{36}, with a ω score of 0.10. If GA ORFs that contain frameshift mutations are disregarded (e.g. LORF11, UL26, 29, 37 and 52), then only two genes, UL50 (dUTPase) and RLORF7 (Meq), can be considered to have neutral evolution, with mean ω scores of >0.95. Interestingly, UL49.5 has been implicated in the repression of major histocompatibility complex (MHC) class I presentation by inhibiting TAP, and both dUTPase and Meq have been demonstrated to encode virulence factors in studies involving knockout mutants (Jons *et al.*, 1997; Koppers-Lalic *et al.*, 2005; Lipinska *et al.*, 2006; Lupiani *et al.*, 2004; Tischer *et al.*, 2002). With our limited dataset, no ORF was discovered to be under positive selection, i.e. with a d_N/d_S ratio $\gg$1.0.

Differences in ORFs within the unique regions

Most CVI988-BAC ORFs from the unique regions were found to be virtually identical (>98 % amino acid identity) to homologous ORFs from other MDV-1 strains. These genes are generally most similar to those in the genomes of Md5 and Md11 and show a reduced level of homology to those of the GA strain. Based largely on its size, the majority (>85 %) of non-synonymous amino acid substitutions occur within ORFs mapping to the unique long region and occur frequently within ORFs encoding structural proteins, mainly tegument proteins. ORF49 (UL36), which encodes the large tegument protein (McNabb & Courtney, 1992) of CVI988-BAC, is particularly interesting due to the fact that it contains small deletions and the largest number of nucleotide polymorphisms. Its herpes simplex virus type 1 (HSV-1) orthologue has been shown to be involved in protein–protein interactions with other tegument proteins (Klupp *et al.*, 2002; Vittone *et al.*, 2005) and, recently, a ubiquitin (Ub)-specific cysteine protease function has been mapped to the hightly conserved amino terminus of both HSV-1 and MDV-1 orthologues (Kattenhorn *et al.*, 2005; Schlieker *et al.*, 2005). Most of the non-synonymous substitutions in UL36 of MDV-1, however, occur in the carboxyl terminus, which is extremely heterogeneous both between isolates of MDV-1 and within other members of the subfamily *Alphaherpesvirinae* (Gomi *et al.*, 2002). As shown in Fig. 4, we have determined the sequence of this region for two additional isolates of the vaccine strain CVI988 and, in comparison to CVI988-BAC and other strains, have found this region to contain reiterations of two sequences: KPPPPDPDFKS/TPAPKP and KPPPA/TPDSKPSPAPKP. Similar reiterations of proline-rich motifs have been found to be important in protein–protein interactions via SH3,

Table 1. ORFs present in the genome of CVI988-BAC

(a) Annotated ORFs based on the nomenclature of Tulman *et al.* (2000) and Lee *et al.* (2000a). (b) Novel ORFs based on a codon-usage table generated by well-characterized ORFs presented in (a). Start and termination positions and length of each ORF are listed. Relatedness of the Md5, Md11 and GA homologues is given as percentage identity.

Name	ORF	Position		Length (codons) in strain:				Identity (%)			Function
		Start	End	CVI988	Md5	Md11	GA	Md5	Md11	GA	
(a)											
RLORF 1	2.0	1359	1955	198	198	198	239	99	99	56	ICP0-like protein
RLORF 2*	3.0	2395	2560	121	134	134	142	100	98	98	vIL-8, chemokine
		2659	2794								
		2970	3033								
RLORF 3	3.2	2617	2925	102	102	102	102	99	100	100	
RLORF 4	3.4	3229	3657	142	142	142	125	99	100	87	
RLORF 5	3.6	3632	3805	57	57	57	115	100	100	96	
L1	3.8	4043	4366	107	107	107	82	100	100	98	Transcript identified in MD lymphoblastoid chicken cells
23 kDa	4.0	5355	5942	195	136	136	136	68	68	68	23 kDa nuclear protein, binds α-enolase
RLORF 7	5.0	5211	6407	398	339	339	339	81	81	81	MEQ protein
RLORF 6	5.1	5141	5935	264	205	205	205	86	86	87	
RLORF 8	5.4	6732	7139	135	135	135	137	100	100	97	Binds C1q-binding protein
RLORF 9	5.7	9381	9704	107	107	107	107	100	100	100	14 kDa lytic protein C
RLORF 10	5.8	10108	10395	95	95	95	100	100	100	100	Binds MHC class II invariant chain
14 kDa A*	6.0	10158	10372	85	85	85	54	100	100	73	14 kDa lytic protein A
		13553	13595								
14 kDa B*	6.1	10158	10372	93	93	93	62	98	100	80	14 kDa lytic protein B
		11412	11478								
RLORF 11	6.4	11257	11568	103	103	103	103	100	100	99	
B68	6.6	13218	13439	73	63	63	63	100	100	98	VZV *trans*-inducing protein
RLORF 12	7.0	13493	13840	115	115	115	115	99	99	99	Binds growth-related TCTP
LORF 1	9.0	13785	14786	333	333	333	333	99	99	96	
RLORF 13	9.5	14172	14486	104	104	104	104	100	99	99	Binds C1q-binding protein
RLORF 14	8.0	14281	14748	155	155	155	155	100	100	99	Early 24 kDa phosphoprotein, pp24
LORF 2	8.4	15269	17323	684	684	684	684	100	100	100	
LIPASE*	10.0	14983	15079	756	756	756	756	100	100	100	Lipase
		15150	17324								
LORF 3	12.0	18276	19430	384	384	384	398	100	100	98	
UL1	13.0	19620	20207	195	195	195	195	100	100	100	Virion surface glycoprotein L
LORF 4	13.5	19628	20056	142	142	142	142	100	100	100	Binds MHC class II β chain
UL2	14.0	20089	21030	313	313	313	313	100	100	99	Uracil–DNA glycosylase
UL3	15.0	21055	21741	228	228	228	228	100	100	100	Nuclear phosphoprotein
UL3.5	15.5	21763	21939	58	58	58	58	100	100	100	Tegument protein, virus egress

Table 1. cont.

Name	ORF	Position		Length (codons) in strain:				Identity (%)			Function
		Start	End	CVI988	Md5	Md11	GA	Md5	Md11	GA	
UL4	16.0	22257	23063	268	268	268	268	100	100	100	Nuclear protein
UL5	17.0	23117	25693	858	858	858	858	100	100	97	DNA helicase–primase-associated protein
UL6	18.0	25761	27929	722	722	722	718	99	100	99	Minor capsid protein, DNA packaging
UL7	19.0	27766	28683	305	305	305	305	100	100	96	Capsid protein, DNA packaging/cleavage
UL8	20.0	28717	31026	769	769	769	768	100	99	99	DNA helicase–primase-associated protein
UL8.5	20.5	31040	32506	488	488	488	488	100	100	100	Origin-binding protein
UL9	21.0	31040	33565	841	841	841	841	100	100	99	Origin-binding protein
UL10	22.0	33664	34938	424	424	424	424	100	100	99	Virion membrane glycoprotein M
UL11	23.0	34992	35246	84	84	84	84	100	100	100	Myristylated tegument protein
UL12	24.0	35222	36796	524	524	524	524	99	99	91	DNase
UL13	25.0	36787	38328	513	513	513	513	100	100	99	Serine/threonine protein kinase
UL14	26.0	38046	38777	243	243	243	248	100	100	100	Minor tegument protein
UL15*	27.0	38808	39851	737	737	737	737	100	100	97	DNA-packaging protein (terminase)
		43333	44502								
UL16	28.0	39891	40973	360	360	360	360	100	100	91	Tegument protein
UL17	29.0	41000	43189	729	729	729	743	100	100	95	Tegument protein, DNA packaging
UL18	30.0	44621	45580	319	319	319	319	100	100	94	Capsid protein
UL19	31.0	45706	49887	1393	1393	1393	1391	99	99	98	Major capsid protein
UL20	32.0	50191	50895	234	234	234	234	100	100	100	Membrane protein, virus egress
UL21	33.0	51156	52796	546	546	546	546	99	99	99	Tegument protein
UL22	34.0	52943	55384	813	813	813	813	100	100	99	Envelope glycoprotein H
UL23	36.0	55567	56625	352	352	352	352	100	100	98	Thymidine kinase
UL24	35.0	56580	57494	304	304	304	324	100	100	98	Non-glycosylated membrane-associated protein
UL25	37.0	57589	59340	583	583	583	583	100	100	99	DNA packaging
UL26	38.0	59381	61372	663	663	663	638	99	99	91	Protease, minor capsid scaffold protein
UL26.5	39.0	60335	61372	345	345	345	320	100	100	84	Minor capsid scaffold protein
LORF 5	39.5	61444	61803	119	119	119	119	100	100	100	
UL27	40.0	61503	64100	865	865	865	865	100	100	100	Virion membrane glycoprotein B
UL28	41.0	64175	66556	793	793	793	793	100	100	100	DNA-packaging protein
UL29	42.0	66748	70323	1191	1191	1191	1191	99	99	98	Single-stranded DNA-binding protein
UL30	43.0	70592	74254	1220	1220	1220	1218	100	99	98	DNA polymerase
UL31	44.0	74178	75080	300	300	300	300	100	100	99	Nuclear phosphoprotein
UL32	46.0	75097	77022	641	641	641	641	100	100	99	DNA packaging
UL33	45.0	77021	77425	134	134	134	120	100	100	99	DNA packaging
UL34	47.0	77518	78351	277	277	277	276	100	100	98	Membrane phosphoprotein
UL35	48.0	78435	78830	131	131	131	130	100	100	99	Capsid protein, VP26
UL36	49.0	78891	88862	3323	3342	3324	3325	97	98	97	Large tegument protein, putative deubiquitinase
LORF 6	49.5	88510	88977	155	155	155	155	100	100	100	

Table 1. cont.

Name	ORF	Position		Length (codons) in strain:				Identity (%)			Function
		Start	End	CVI988	Md5	Md11	GA	Md5	Md11	GA	
UL37	50.0	89077	92217	1046	1046	1046	1046	99	99	99	Tegument protein
LORF 7	50.5	91928	92290	120	120	120	120	100	99	98	
UL38	51.0	92588	94000	470	470	470	470	99	99	99	Capsid protein
UL39	52.0	94225	96693	822	822	822	822	99	99	99	Ribonucleotide reductase, large subunit
UL40	53.0	96746	97777	343	343	343	343	100	99	100	Ribonucleotide reductase, small subunit
UL41	54.0	97825	99150	441	441	441	441	100	99	99	Tegument protein, virion shutoff protein
UL42	55.0	99807	100916	369	369	369	369	100	100	100	DNA polymerase
UL43	56.0	101076	102338	420	420	420	420	99	99	99	Probable membrane protein
UL44	57.0	102558	104063	501	501	501	501	99	99	99	Virion glycoprotein C
LORF 8	57.8	104760	105386	208	208	208	208	99	99	99	23 kDa protein
UL45	58.0	104926	105561	211	211	211	211	99	99	99	Cell-fusion protein
UL46	59.0	105694	107400	568	568	568	568	99	99	99	Tegument phosphoprotein, VP11/12
UL47	60.0	107542	109968	808	808	808	808	99	99	99	Tegument phosphoprotein, VP13/14
UL48	61.0	110207	111490	427	427	427	427	100	100	100	Tegument immediate-early protein, VP16
UL49	62.0	111598	112329	243	249	249	249	95	95	95	Tegument phosphoprotein, VP22
UL49.5	64.0	112475	112762	95	95	95	95	98	98	98	Envelope/tegument protein
UL50	63.0	112742	114052	436	436	436	436	99	99	98	dUTPase
UL51	65.0	114140	114889	249	249	249	249	99	99	99	Virion phosphoprotein
UL52	66.0	114891	118115	1074	1074	1074	1075	99	99	97	DNA helicase–primase-associated protein
UL53	67.0	118094	119158	354	354	354	354	99	99	100	Glycoprotein K
UL54	68.0	119304	120725	473	473	473	473	100	100	100	ICP27-like protein, post-translational gene regulation
LORF 9	69.0	120855	121664	269	269	269	269	100	100	100	
UL55	70.0	121832	122332	166	166	166	166	100	100	100	Nuclear matrix-associated protein
LORF 10	71.0	122688	123272	194	194	194	193	99	99	98	VZV ORF 2 homologue
LORF 11	72.0	123905	126616	903	903	903	903	99	99	97	
LORF 12	72.8	126818	127198	126	126	126	148	100	100	98	
RLORF 14a	73.0	127290	128162	290	290	290	290	99	99	99	Early 38 kDa phosphoprotein, pp38
RLORF 13A	73.4	127897	128271	124	124	124	124	100	99	99	
RLORF 12	74.0	128603	128806	67	115	115	115	95	95	95	Binds growth-related TCTP
14 kDa A*	75.0	128843	128885	85	85	85	54	100	100	73	14 kDa lytic protein A
		132066	132280								
B68	75.1	129029	129220	63	63	63	63	100	100	98	VZV *trans*-inducing protein
14 kDa B*	75.3	130960	131026	93	93	93	62	100	100	80	14 kDa lytic protein B
		132066	132280								
RLORF 11	75.4	130870	131181	103	103	103	103	100	100	99	
RLORF 10	75.7	132043	132330	95	95	95	100	100	100	100	Binds MHC class II invariant chain
RLORF9	75.8	132734	133057	107	107	107	107	100	100	100	14 kDa lytic protein C
RLORF 8	75.92	135299	135706	135	135	135	137	100	100	97	Binds C1q-binding protein

Table 1. cont.

Name	ORF	Position		Length (codons) in strain:				Identity (%)			Function
		Start	End	CVI988	Md5	Md11	GA	Md5	Md11	GA	
RLORF 7	76.0	136031	137227	398	339	339	339	81	81	81	MEQ protein
23 kDa	77.0	136496	137083	195	136	136	136	68	68	68	23 kDa nuclear protein
RLORF 6	77.5	136503	137297	264	205	205	205	86	86	87	
RLORF 2*	78.0	139405	139468								
		139644	139779								
		139878	140043	121	134	134	142	100	98	98	vIL-8, chemokine
L1	78.1	138072	138395	107	107	107	82	100	100	98	Transcript identified in MD lymphoblastoid chicken cells
RLORF 5	78.2	138633	138806	57	57	57	115	100	100	96	
RLORF 4	78.3	138781	139209	142	142	142	125	99	100	87	
RLORF 3	78.4	139513	139821	102	102	102	102	99	100	100	
RLORF 1	79.0	140483	141079	198	198	198	239	99	99	56	ICP0-like protein
RSORF 1	82.0	144348	144674	108	108	108	29	100	100	100	
ANTISENSE	83.0	151837	152175	112	112	112	112	100	100	95	Antisense RNA protein
ICP4	84.0	145031	151996	2321	2321	2321	2323	99	99	98	Immediate-early protein, ICP4
SORF 1	86.6	154605	154835	76	43	76	89	100	100	97	
SORF 2	87.0	155025	155564	179	179	179	179	100	100	100	Binds growth hormone
US1	88.0	155731	156270	179	179	179	179	100	100	100	ICP22, binds retinoblastoma-binding protein 4
US10	89.0	156559	157200	213	213	213	213	100	100	100	Binds stem-cell antigen 2^5 lymphocyte antigen 6 complex
SORF 3	90.0	157311	158366	351	351	351	351	100	100	99	Raf kinase homologue
US2	91.0	158593	159405	270	270	270	270	100	100	99	
US3	92.0	159517	160725	402	402	402	402	99	99	99	Serine/threonine protein kinase
SORF 4	93.0	160835	161278	147	147	147	147	100	100	98	
US6	94.0	161446	162657	403	403	§	403	99		99	Membrane glycoprotein D
US7	95.0	162764	163831	355	355	355	355	100	100	98	Membrane glycoprotein I
US8	96.0	163970	165463	497	497	497	497	100	100	100	Membrane glycoprotein E
ICP4	100.0	168753	175718	2321	2321	§	2323	99		98	Immediate-early protein, ICP4
ANTISENSE	101.0	168574	168912	112	112	§	112	100		95	Antisense RNA protein
RSORF 1	102.0	176075	176401	108	108	§	29	100		100	
(b)											
MEQG	0.5	529	819	96	29	29	29	100	100	100	
MNDR	1.0	991	1653	220	219	219	¶	99	32	0	
MALT	2.3	1732	1995	87	87	87	¶	98	98	0	
MSVH	2.6	1891	2190	99	99	99	51	100	100	59	
MGGG	5.2	5827	6051	74	75	75	75	100	100	100	
MLHG	5.3	6143	6382	79	79	79	79	98	98	98	
MYFK	5.5	7633	7791	52	82	82	86	100	100	90	
MRLK	5.6	9149	9505	118	118	118	51	100	100	96	
MRFG	6.2			†	66	66	38				

Table 1. cont.

Name	ORF	Position		Length (codons) in strain:				Identity (%)			Function
		Start	End	CVI988	Md5	Md11	GA	Md5	Md11	GA	
MFAE	6.3	10581	10838	85	47	54	47	100	90	100	
MLTI	6.5	11921	12907	328	64	64	64	100	100	82	
MKNI	8.8	15336	15593	85	85	85	85	100	100	100	
MTSE	11.0	17879	18136	85	85	85	40	100	100	100	
MGCQ	11.5	18191	18367	58	58	58	78	100	100	73	
MCYW	12.4	18711	18995	94	94	94	94	100	100	100	
MIKQ	12.8	19177	19452	91	91	91	91	100	100	97	
MCNH	14.5	21039	21332	97	97	97	97	100	100	100	
MQLD	31.5	50019	50258	79	79	79	37	100	100	100	
MEEI	38.5	60990	61217	75	75	75	¶	100	100	0	
MNFH	49.1	80737	80872	34	93	93	94	1.0	100	100	
MIVP	57.4	104374	104601	75	75	75	75	100	100	100	
MKTP	63.5	114057	114293	78	78	78	78	100	100	98	
MQLA	71.4	123001	123279	92	92	92	91	98	98	96	
MIKR	71.8	123760	123999	79	79	79	79	100	100	100	
MVES	72.2	124500	125120	206	206	206	206	100	100	100	
MHSG	72.4	125511	125846	111	111	111	111	99	99	99	
MTTS	72.6	126768	127067	99	99	99	99	100	100	97	
MLTI	75.2	129531	130517	328	64	64	64	100	100	82	
MFAE	75.5	131600	131857	85	47	54	47	100	90	100	
MRFG	75.6			†	66	66	66				
MRLK	75.9	132933	133289	118	118	118	51	100	100	96	
MYFK	75.91	134647	134805	52	82	82	86	100	100	90	
MLHG	76.4	136056	136295	79	79	79	79	98	98	98	
MGGG	76.8	136387	136611	74	75	75	75	100	100	100	
MSVH	78.5	140248	140547	99	99	99	51	100	100	59	
MALT	78.6	140443	140706	87	87	87	¶	98	98	0	
MNDR	80.0	140785	141447	220	219	219	¶	99	99	0	
MEQG	80.5	141619	141909	96	29	29	29	100	100	100	
MRGH	81.0	142615	142863	82	162	62	43	100	100	0	
MRAN	81.5	143193	143336	47	47	47	97	100	100	68	
MVFK	84.5	152225	152527	94	67	67	101	95	95	98	
MFAY	85.0	152244	152624	100	132	132	101	98	97	99	
MGKY	85.3	152731	152961	76	57	77	77	100	100	100	
MFSK	85.6	152951	153160	69	69	69	67	100	100	98	
MNNT	85.9	152958	153176	72	72	72	145	100	100	92	
MSWP	86.0	153211	153375	54	87	95	91	90	88	90	
MWGR	86.1	153412	153543	43	43	43	115	100	100	100	

Table 1. cont.

Name	ORF	Position		Length (codons) in strain:				Identity (%)			Function
		Start	End	CVI988	Md5	Md11	GA	Md5	Md11	GA	
MQTR	86.2	154322	154435	37	‡	37	59	0	67	55	
MGPP1	86.3			†	‡	234	¶				
MHGE	86.4	154371	154529	52	72	85	52	97	96	98	
MGPP2	86.5			†	40	145	¶				
MSIR	89.5	156633	157019	128	128	128	128	100	100	100	
MAHG	91.5	158894	159439	181	181	181	181	100	100	98	
MLYC	92.4	159597	159827	76	76	76	76	100	100	100	
MDQV	92.8	160103	160330	75	75	75	75	100	100	98	
MLLV	94.5	161937	162365	142	60	§	142	98		98	
MSAS	95.5	163948	164238	96	96	96	96	100	100	100	
MLDC	96.5	164922	165197	91	91	91	91	100	100	100	
MQRQ	97.0			†	149	§	¶				
MQYA	97.3	166243	166554	103	‡	§	¶				
MQTR	97.6	166203	166541	112	‡	§	59			40	
MWGR	97.9			†	43	§	115				
MSWP	98.0	167372	167536	54	87	§	91	90		90	
MNNT	98.3	167571	167789	72	72	§	145	100		92	
MFSK	98.6	167587	167796	69	69	§	67	100		98	
MGKY	98.9	167841	168017	58	57	§	77	100		100	
MFAY	99.0	168124	168426	100	132	§	101	98		99	
MVFK	99.5	168221	168505	94	67	§	101	95		98	
MRAN	102.5	177413	177556	47	47	§	97	100		68	
MRGH	103.0	177886	178134	82	162	§	¶	100			

*Spliced genes.

†ORFs in bold are absent in the CVI988 genome.

‡ORFs absent in the MD5 genome.

§Not present in the BAC construct of Md11.

¶ORFs absent in the GA genome.

Table 2. SNPs in the genome of CVI988-BAC

All ORFs listed in Table 1 were examined for non-synonymous amino acid substitutions in comparison to homologous ORFs from the oncogenic strains Md5, Md11 and GA. Amino acid positions are noted for the ORFs of Md5 and CVI988, respectively.

ORF	Non-synonymous substitution
24.0 UL12	Val^{383}→Ile^{383}, G→A
31.0 UL19	Ile^{448}→Lys^{448}, T→A
33.0 UL21	Gly^{6}→Glu^{6}, G→A
38.0 UL26	Thr^{305}→Arg^{305}, C→G
49.0 UL36	Asp^{360}→Tyr^{360}, G→T
49.0 UL36	Pro^{363}→Ser^{363}, C→T
49.0 UL36	Leu^{1640}→Phe^{1640}, C→T
49.0 UL36	Cys^{2576}→Arg^{2576}, T→C
49.0 UL36	Thr^{2713}→Pro^{2714}, A→C
49.0 UL36	Ser^{2759}→Pro^{2742}, T→C
49.0 UL36	Thr^{2765}→Lys^{2748}, CT→AA
49.0 UL36	Ala^{2767}→Pro^{2750}, GGC→CCG
49.0 UL36	Ser^{2771}→Thr^{2758}, T→A
49.0 UL36	Thr^{2787}→Ser^{2774}, A→T
49.0 UL36	Ser^{2796}→Pro^{2783}, T→C
49.0 UL36	Lys^{2799}→Ser^{2786}, AAA→TCT
49.0 UL36	Pro^{2801}→Ala^{2788}, CCG→GCC
49.0 UL36	Asp^{2803}→Lys^{2790}, GAT→AAG
49.0 UL36	Asp^{2805}→Ser^{2792}, GAT→TCT
49.0 UL36	Phe^{2806}→Pro^{2793}, TTC→CCG
49.0 UL36	Lys^{2807}→Ala^{2794}, AAG→GCC
49.0 UL36	Thr^{2809}→Lys^{2796}, ACT→AAG
49.0 UL36	Ala^{2811}→Ser^{2798}, GCC→TTC
49.0 UL36	Lys^{2813}→Ala^{2800}, AAG→GCC
49.0 UL36	Pro^{2817}→Ser^{2804}, CCG→TCT
49.0 UL36	Asp^{2819}→Ala^{2806}, GAT→GCC
49.0 UL36	Ser^{2856}→Pro^{2841}, T→C
49.0 UL36	Pro^{2867}→Lys^{2848}, CCG→AAG
49.0 UL36	Ala^{2869}→Ser^{2850}, G→T
49.0 UL36	Asp^{2871}→Ala^{2852}, GAT→GCC
49.0 UL36	Ser^{2872}→Pro^{2853}, T→C
49.0 UL36	Thr^{2885}→Ala^{2866}, A→G
50.0 UL37	Val^{144}→Ile^{144}, C→T
50.0 UL37	Lys^{949}→Glu^{949}, A→G
51.0 UL38	Thr^{68}→Ala^{68}, A→G
52.0 UL39	Glu^{727}→Asp^{727}, A→C
52.0 UL39	Ala^{806}→Thr^{806}, G→A
57.0 UL44	Ala^{460}→Ser^{460}, G→T
57.8 LORF8	Thr^{41}→Met^{41}, C→T
58.0 UL45	Val^{114}→Ile^{114}, G→A
59.0 UL46	Leu^{30}→Ser^{30}, T→C
59.0 UL46	Asp^{505}→Gly^{505}, C→T
60.0 UL47	Thr^{122}→Ala^{122}, A→G
60.0 UL47	Asp^{267}→Gly^{267}, A→G
60.0 UL47	Ala^{664}→Val^{664}, C→T
62.0 UL49	His^{58}→Tyr^{58}, C→T
62.0 UL49	Ala^{152}→Pro^{152}, G→C
62.0 UL49	Ser^{155}→Cys^{155}, C→G
62.0 UL49	Ser^{245}→Gly^{239}, A→G
64.0 UL49.5	Met^{65}→Leu^{65}, A→C
63.0 UL50	Asn^{2}→Asp^{2}, A→G
63.0 UL50	Arg^{391}→Gly^{391}, G→A
65.0 UL51	Asp^{191}→Asn^{191}, G→A
66.0 UL52	Ala^{99}→Thr^{99}, G→A
66.0 UL52	Ala^{127}→Val^{127}, C→T
66.0 UL52	Val^{192}→Ala^{192}, T→C
66.0 UL52	Ala^{588}→Val^{588}, C→T
72.0 LORF11	Ala^{302}→Thr^{302}, G→A
72.0 LORF11	Pro^{780}→Leu^{780}, C→T
72.4 MHSG	Ala^{68}→Val^{68}, C→T
73.0 RLORF14a	Gln^{107}→Arg^{107}, A→G
73.0 RLORF14a	Glu^{109}→Gly^{109}, A→G
74.0 RLORF12	Thr^{47}→Arg^{47}, C→G
76.0 RLORF 7	Ala^{71}→Ser^{71}, G→T
76.0 RLORF 7	Lys^{77}→Glu^{77}, A→G
76.0 RLORF 7	Thr^{326}→Ile^{385}, C→T
76.4 MLHG	Ser^{13}→Pro^{13}, T→C
77.5 RLORF 6	Leu^{169}→Phe^{228}, C→T
78.6 MALT	Cys^{10}→Gly^{10}, T→G
86.4 MHGE	Gly^{45}→Trp^{45}, G→T

WW and EHV1 domains (Kay *et al.*, 2000; Zarrinpar *et al.*, 2003).

Overlapping the UL36 gene, one previously uncharacterized gene (ORF49.1) was discovered via an extensive computer analysis to determine the ORFs capable of encoding proteins ($\geqslant$50 aa) that are present in the genomes of oncogenic strains, but absent in the CVI988-BAC genome. In the genomes of the oncogenic strains, ORF49.1 encodes a protein of 93 aa (Md5 and Md11) or 94 aa (GA), but in the genome of CVI988-BAC, a truncated protein of 34 aa is predicted, due to a point mutation. It is important to note that this is a silent mutation with respect to the UL36 ORF. Comparative genomic studies with other MDV strains have indicated that the genomes of the CVI988 vaccine strains from Intervet (p27) and ADOL (BP5) also encode the 34 aa derivative, whereas virulent strains RB1B, JM/102W, 549a, 571, 584a, 595, 648a and 686, representing the virulent pathotypes, encode the full-length protein (unpublished results). A BLAST search for proteins similar to ORF49.1 failed to generate any believable hits. The role of this ORF in the pathogenesis of MDV remains unknown and will require additional investigation.

Another gene within the unique long region that differs among attenuated and oncogenic strains is UL49 (ORF62). This gene encodes the tegument protein VP22 and has been shown to be indispensable for virus propagation in cell culture (Dorange *et al.*, 2000, 2002). A short deletion of 18 bp was found in this gene within the CVI988-BAC genome. This deletion corresponds to a domain that is rich in serine and threonine residues in the oncogenic homologues. Close inspection of the multiple alignment (Fig. 3)

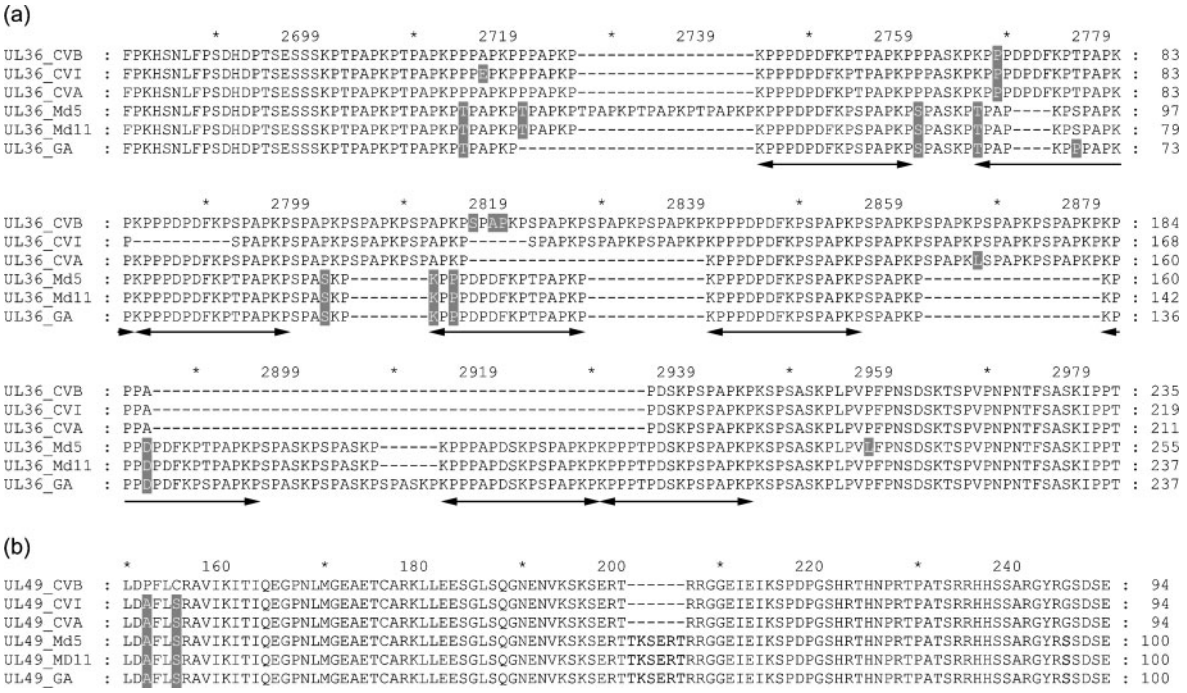

Fig. 3. Multiple alignments of the variable regions in UL36 (a) and UL49 (b). Amino acids are numbered according to their position in the complete alignments of these ORFs. Underlined residues denote regions that contain reiterations. Additional sequences for UL36 and UL49 were included in the alignments with the following abbreviations: CVB, CVI988 BAC; CVI, CVI988-p27 (Intervet); CVA, CVI988-BP5 (ADOL).

suggests a domain deletion (KSERT). In addition to CVI988-BAC, this deletion was also found in other vaccine derivatives of CVI988 from Intervet (p27) and ADOL (BP5). There is little sequence resemblance between this region and corresponding regions in the VP22 homologues of HSV and varicella-zoster virus (VZV). However, like its HSV-1 counterpart, VP22 (MDV-1) is a phosphorylated protein with DNA-binding activity and displays similar functional properties: intracellular spreading and nuclear-targeted protein transporting (Blaho *et al.*, 1994; Dorange *et al.*, 2000; Elliott *et al.*, 1999; Phelan *et al.*, 1998).

The genes within the unique short region of CVI988-BAC are unremarkable in a genomic comparison, with protein identities >98 % and few polymorphisms. Unlike Md5, which encodes two sorf2 proteins (ORF87 and ORF97) (Tulman *et al.*, 2000), CVI988-BAC is more similar to Md11 and GA and lacks ORF97. Overall, the gene organization in the US region of the four sequenced strains is almost identical, with the exception of ORF97 in the Md5 genome. What is remarkable is the lack of sequence conservation in the inverted-repeat regions, which flank the unique short region.

Differences in ORFs within the inverted-repeat regions

This region is by far the most interesting, well-characterized region in the MDV genome and many of its ORFs have been defined functionally (Lupiani *et al.*, 2001; Osterrieder & Vautherot, 2004). These ORFs are largely mardiviral-specific, some even serotype 1-specific, and homologues are not generally found in other members of the subfamily *Alphaherpesvirinae*. The protein products encoded by these ORFs include the oncoprotein MEQ (ORF5/ORF76) (Jones *et al.*, 1992), the 23 kDa protein (ORF4/ORF77) (Peng & Shirazi, 1996b) and viral interleukin-8 (vIL-8; ORF3/ORF78) (Liu *et al.*, 1999; Parcells *et al.*, 2001).

Within the repeat long region, the ORFs of CVI988-BAC show a mean of 97.2, 94.5 and 91.7 % identity to those of Md5, Md11 and GA, respectively. This subdued level of similarity is largely due to proteins that are truncated or contain insertions, most notably the gene products of ORF4/ORF77, ORF5/ORF76 and ORF5.1/ORF77.5, encoding the 23 kDa protein, MEQ and RLORF6, respectively (Jones *et al.*, 1992; Lee *et al.*, 2000a; Peng & Shirazi, 1996b). As illustrated in Fig. 4, the overlapping genes encoding the 23 kDa, MEQ and RLORF6 proteins of CVI988-BAC all contain 177 bp insertions. Other researchers have previously reported insertions in the Meq loci of CVI988 (Chang *et al.*, 2002b; Lee *et al.*, 2000b; Shamblin *et al.*, 2004). In the original paper by Lee *et al.* (2000b), a 178 bp insertion within the coding region of Meq (CVI988), specifically in the proline-rich domain, was reported to result in a frameshift. Petherbridge *et al.* (2003) reported previously, and we confirm, that the genes encoding the MEQ proteins

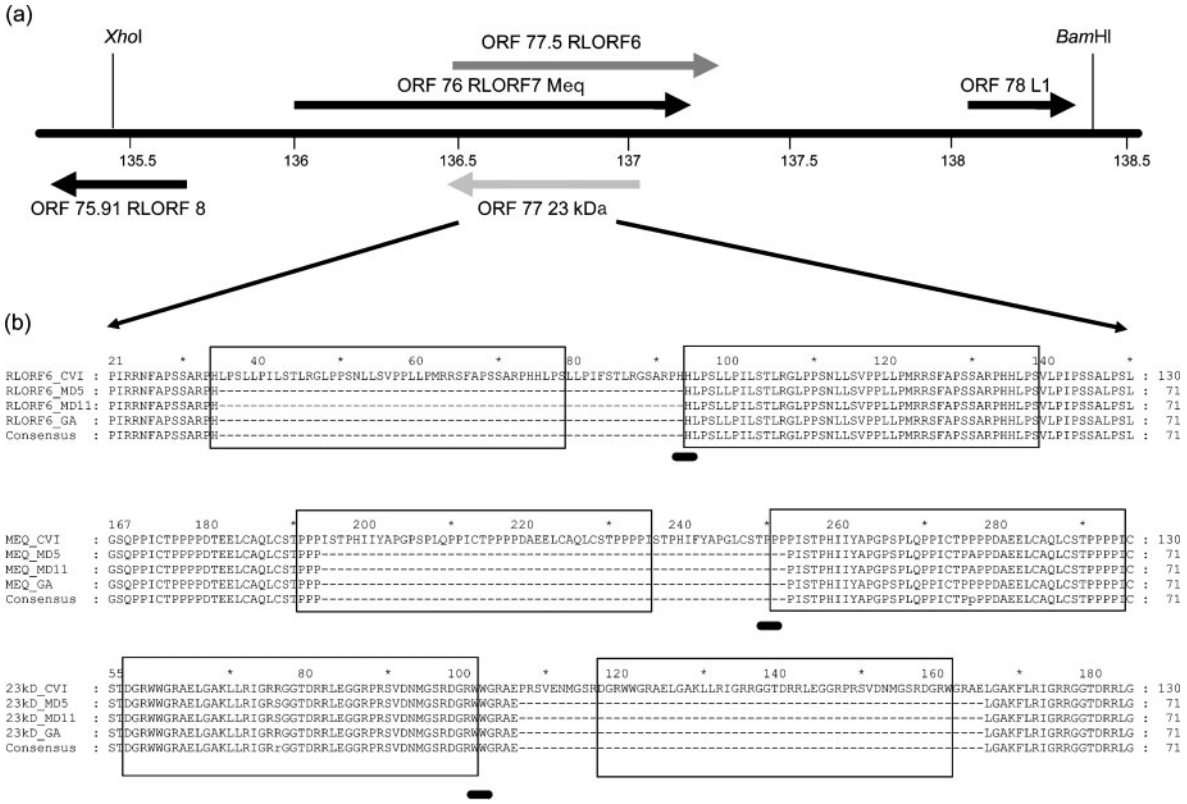

Fig. 4. Multiple alignments of ORFs within the Meq loci. (a) Schematic diagram of the Meq locus within the IRL region of CVI988 BAC. (b) Alignments of the duplication domains within the overlapping ORFs RLORF6, RLORF 7 (Meq) and 23 kDa of CVI988, Md5, Md11 and GA strains. Stretches of amino acids enclosed in boxes represent domain duplications. Solid bars denote two identical amino acids at the duplication junctions. Amino acids are numbered according to their position in the complete alignments of these ORFs.

of CVI988-BAC contain insertions of only 177 bp that are in frame. This is also in agreement with research examining Meq mutations not only in CVI988, but also in other attenuated strains of MDV-1 (Chang *et al.*, 2002a; Petherbridge *et al.*, 2003; Shamblin *et al.*, 2004). The general conclusion is that most attenuated strains of MDV-1 contain in-frame insertions (59 aa) in the Meq ORF, the probable result of domain duplication (see boxed regions in Fig. 4). One notable exception is the attenuated strain R2/23, in which an additional thymidine residue within the Meq locus causes a frameshift mutation (Spatz & Silva, 2007). Interestingly, all of the insertion mutations within the Meq genes of attenuated strains occur in the same region, the proline-rich domain, which has been shown to be essential for the *trans*-activation of MEQ-responsive promoters (Liu & Kung, 2000; Qian *et al.*, 1996). This domain, however, when used by itself in *trans*-activation analysis, represses transcription and behaves as a dominant-negative mutant. In comparing the *trans*-activation/repression abilities of MEQ variants from attenuated and virulent strains, Chang *et al.* (2002c) reported that those containing the duplicated proline-rich motifs (coined L-MEQ) exhibit a higher level of *trans*-repression than do MEQ variants containing the single proline-rich motif (S-MEQ) found in virulent strains.

In addition to insertional mutations in the Meq genes of CVI988-BAC, three point mutations were discovered at aa 71, 77 and 385 (Table 2). This places the sequence of CVI988-BAC in the IV allelic group described by Laurent *et al.* (2004). Other strains of MDV-1 classified as virulent (e.g. BC-1, JM102, 567, 617A, 637, 571, 573) have been reported to contain mutations at either position 71 or 77 (Shamblin *et al.*, 2004). Only one other strain, CU-2, which has a mild virulent phenotype, contained the double mutations at position 71 and 77, similar to CVI988-BAC. The Ile^{385} mutation is exclusive to the CVI988-BAC MEQ protein. The SNP responsible for Ile^{385}, when examined in the overlapping ORF (RLORF6), encodes either leucine or phenylalanine. RLORF6 of CVI988-BAC contains Phe^{228}; in contrast, all oncogenic strains were found to contain Leu^{169}.

Within the repeat long region of CVI988-BAC, three ORFs (ORF3/ORF78, ORF5.5/ORF75.91 and ORF74) encode truncated proteins relative to those encoded by the genomes of the three oncogenic strains. Because of an opal (TGA) mutation, 20 aa are missing in the putative proteins of ORF5.5/ORF75.91 (CVI988-BAC). An extensive homology search using BLAST programs provided few clues to the possible function of these proteins. The function of the

ORF3/ORF78 gene products, however, is far more defined. These ORFs are spliced genes with similarity to genes encoding the CxC chemokine IL-8 and hence termed vIL-8. We have discovered that, in the CVI988-BAC genome, these ORFs encode a truncated version of vIL-8 that is 11 and 13 aa shorter than those of Md5 and Md11/GA, respectively. Although the function of the carboxyl-terminal residues in the mammalian IL-8 orthologue has been reported to be non-essential for neutrophil binding, little information is available about the function of these residues in vIL-8 or avian IL-8 (Kaiser *et al.*, 1999). In fact, avian IL-8 contains only 103 aa, thus to some extent resembling CVI988-BAC vIL-8. As vIL-8-null MDV-1 mutants have been reported to be mildly virulent and to have a lower incidence of tumour formation, truncated versions of vIL-8 may also contribute to an attenuated phenotype (Cortes & Cardona, 2004; Cui *et al.*, 2004).

One ORF (ORF74 or RLORF12) within the internal repeat long region is predicted to encode a truncated protein of 67 aa. Unexpectedly, its diploid counterpart (ORF7) within the terminal repeat long region encodes a protein of 115 aa. All sequenced oncogenic strains of MDV encode RLORF12 homologues containing 115 aa; however, we have found recently that some attenuated strains (i.e. RM-1, CVI988-BP5 and JM/102W) contain deletions within this ORF and encode similar truncated proteins (Spatz & Silva, 2007). This could be significant, as these deletions occur at the origin of replication and may affect the binding of the DNA-binding proteins (UL9 and UL8.5) involved in replication, as well as other proteins demonstrated to bind the RLORF12 protein, such as the growth-related translationally controlled tumour protein (TCTP) (Niikura *et al.*, 2004).

Within the repeat long regions, only one pair of diploid ORFs (ORF6.2/ORF75.6) was absent in the CVI988 genome, due to a point mutation in the start codon. In virulent strains, these ORFs are predicted to encode short polypeptides of 66 aa. Because of their small size, no significant similarity was found to any protein in GenBank. Whether these genes actually encode proteins or represent exons of uncharacterized proteins remains to be determined. In support of the latter notion, splicing and alternative splicing have been shown to occur within this region, as evidenced by the spliced gene products vIL-8 (Liu *et al.*, 1999), 14 kDa A and 14 kDa B (Hong & Coussens, 1994), as well as alternatively spliced gene products involving exons 2 and 3 of vIL-8 RNA designated Meq/vIL-8, RLORF5a/vIL-8 and RLORF4/vIL-8 (Anobile *et al.*, 2006; Jarosinski & Schat, 2007; Peng & Shirazi, 1996a; Peng *et al.*, 1995). Given the complex transcription patterns and alternative splicing that occur in this region, it is quite possible that additional spliced gene products exist.

The repeat short regions of CVI988-BAC, like those of oncogenic strains, largely encode the major immediate-early protein ICP4 (Anderson *et al.*, 1992). Over 57 % of the repeat short regions are devoted to encoding this *trans*-activator. These proteins are >98 % identical among strains and, although there are differences, especially with ICP4 (GA), no polymorphisms were discovered that could differentiate attenuated versus oncogenic strains collectively. Thirty-four substitutions (23 non-synonymous and 11 synonymous) were discovered in the ICP4 gene of the GA strain relative to CVI988. Only three and six substitutions were found in ICP4 of Md5 and Md11, respectively, again suggesting a close evolutionary relationship between CVI988 and these two strains. Whether these substitutions are significant for attenuation remains to be determined, particularly in light of research demonstrating that ICP4 of a vaccine strain of VZV (Oka) has a lower *trans*-activation capacity than its parental counterpart and this contributes to its attenuated phenotype (Cohrs *et al.*, 2006; Gomi *et al.*, 2002). What can be hypothesized is that the *trans*-activation potential of these two alphaherpesviruses seems to play a role in attenuation. Therefore, as few mutations were noted between ICP4 proteins of Md5 and CVI988, it seems likely that the *trans*-repression capability of L-MEQ and not the ICP4 protein itself is important in attenuation of CVI988. It is interesting to note that, although MEQ is expressed predominantly in the latent state, it is also expressed early in lytic infection and is capable of repressing the transcription of the ICP4 promoter through direct binding as a MEQ homodimer (Jones *et al.*, 1992; Levy *et al.*, 2003).

Conclusions

In recent years, the use of MDV BAC recombinants and overlapping cosmid constructs has been instrumental in determining the function of various genes in the pathogenesis of MD. Five genes (RLORF4, vLIP, vTR, vIL-8 and the Meq/23 kDa locus) have been demonstrated to encode virulence factors (Cui *et al.*, 2005; Jarosinski *et al.*, 2005; Kamil *et al.*, 2005; Lupiani *et al.*, 2004; Trapp *et al.*, 2006). Of these, our data support a role for the MEQ oncoprotein in the attenuation of CVI988. No genetic changes were found in the genes encoding RLORF4, lipase (vLIP) or telomerase (vTR) of CVI988 compared with those found in Md5, Md11 and GA. It is likely that additional ORFs contribute to the attenuated phenotype of CVI988. This study has identified eight ORFs [49, 49.1, 62 and the diploid ORFs: Meq, RLORF6, 23 kDa, 5.5/75.91 and 3/78 (vIL-8)] in the CVI988-BAC genome that differ (are expanded or truncated) from those in the oncogenic strains, as well as one ORF (6.2/75.6) that is exclusively present in the genomes of the virulent strains. Only a short deletion in the carboxyl terminal of vIL-8 (CVI988) was discovered, but its significance is doubtful, due to the fact that other isolates of CVI988 contain genes encoding full-length vIL-8. Seventy non-synonymous amino acid substitutions were identified within the CVI988-BAC ORFs. These mutations were exclusive to CVI988 and corresponding substitutions were identical in all of the three virulent strains. Most of these occurred in tegument proteins, especially UL36, and will be useful in future genotyping studies. The ratios of non-synonymous to synonymous mutations were examined in order to understand the selective pressures acting on various

ORFs. Only two genes (UL50 and RLORF7) had d_N/d_S ratios >0.95, indicating a lack of selective pressure to conserve functional amino acids. All other genes examined with variations in non-synonymous and synonymous substitutions were under purifying selection.

Future comparative genomic studies examining a larger number of avirulent or less virulent strains in order to identify positively selected genes will undoubtedly contribute to our overall understanding of the genes involved in virulence. This information will be essential to engineer novel MDV-1 vaccines capable of protecting chickens against continuously evolving very virulent plus (VV+) strains.

ACKNOWLEDGEMENTS

We thank Jeremy Volkening for his excellent technical assistance and skills in bioinformatics and graphics. The authors would like to thank Dr Harry Danforth for critical evaluation of this report and Drs R. F. Silva and P. Sondermeijer for CVI988 isolates. Part of this work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) and the Department of Environment Food and Rural Affairs (DEFRA), UK, and the United States Department of Agriculture CRIS Program (project no. 6612-32000-043).

REFERENCES

Anderson, A. S., Francesconi, A. & Morgan, R. W. (1992). Complete nucleotide sequence of the Marek's disease virus ICP4 gene. *Virology* **189**, 657–667.

Anobile, J. M., Arumugaswami, V., Downs, D., Czymmek, K., Parcells, M. & Schmidt, C. J. (2006). Nuclear localization and dynamic properties of the Marek's disease virus oncogene products Meq and Meq/vIL8. *J Virol* **80**, 1160–1166.

Besemer, J. & Borodovsky, M. (2005). GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* **33**, W451–W454.

Blaho, J. A., Mitchell, C. & Roizman, B. (1994). An amino acid sequence shared by the herpes simplex virus 1 alpha regulatory proteins 0, 4, 22, and 27 predicts the nucleotidylylation of the UL21, UL31, UL47, and UL49 gene products. *J Biol Chem* **269**, 17401–17410.

Calnek, B. W. (2001). Pathogenesis of Marek's disease virus infection. *Curr Top Microbiol Immunol* **255**, 25–55.

Chang, K. S., Lee, S. I., Ohashi, K., Ibrahim, A. & Onuma, M. (2002a). The detection of the meq gene in chicken infected with Marek's disease virus serotype 1. *J Vet Med Sci* **64**, 413–417.

Chang, K. S., Ohashi, K. & Onuma, M. (2002b). Diversity (polymorphism) of the meq gene in the attenuated Marek's disease virus (MDV) serotype 1 and MDV-transformed cell lines. *J Vet Med Sci* **64**, 1097–1101.

Chang, K. S., Ohashi, K. & Onuma, M. (2002c). Suppression of transcription activity of the MEQ protein of oncogenic Marek's disease virus serotype 1 (MDV1) by L-MEQ of non-oncogenic MDV1. *J Vet Med Sci* **64**, 1091–1095.

Churchill, A. E., Payne, L. N. & Chubb, R. C. (1969). Marek's disease immunization against Marek's disease using a live attenuated virus. *Lancet* **i**, 610–611.

Cohrs, R. J., Gilden, D. H., Gomi, Y., Yamanishi, K. & Cohen, J. I. (2006). Comparison of virus transcription during lytic infection of the Oka parental and vaccine strains of varicella-zoster virus. *J Virol* **80**, 2076–2082.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881–10890.

Cortes, P. L. & Cardona, C. J. (2004). Pathogenesis of a Marek's disease virus mutant lacking vIL-8 in resistant and susceptible chickens. *Avian Dis* **48**, 50–60.

Cui, X., Lee, L. F., Reed, W. M., Kung, H. J. & Reddy, S. M. (2004). Marek's disease virus-encoded vIL-8 gene is involved in early cytolytic infection but dispensable for establishment of latency. *J Virol* **78**, 4753–4760.

Cui, X., Lee, L. F., Hunt, H. D., Reed, W. M., Lupiani, B. & Reddy, S. M. (2005). A Marek's disease virus vIL-8 deletion mutant has attenuated virulence and confers protection against challenge with a very virulent plus strain. *Avian Dis* **49**, 199–206.

Davison, F. & Nair, V. (2005). Use of Marek's disease vaccines: could they be driving the virus to increasing virulence? *Expert Rev Vaccines* **4**, 77–88.

Dorange, F., El Mehdaoui, S., Pichon, C., Coursaget, P. & Vautherot, J.-F. (2000). Marek's disease virus (MDV) homologues of herpes simplex virus type 1 UL49 (VP22) and UL48 (VP16) genes: high-level expression and characterization of MDV-1 VP22 and VP16. *J Gen Virol* **81**, 2219–2230.

Dorange, F., Tischer, B. K., Vautherot, J. F. & Osterrieder, N. (2002). Characterization of Marek's disease virus serotype 1 (MDV-1) deletion mutants that lack UL46 to UL49 genes: MDV-1 UL49, encoding VP22, is indispensable for virus growth. *J Virol* **76**, 1959–1970.

Elliott, G., O'Reilly, D. & O'Hare, P. (1999). Identification of phosphorylation sites within the herpes simplex virus tegument protein VP22. *J Virol* **73**, 6203–6206.

Fragnet, L., Blasco, M. A., Klapper, W. & Rasschaert, D. (2003). The RNA subunit of telomerase is encoded by Marek's disease virus. *J Virol* **77**, 5985–5996.

Gimeno, I. M., Witter, R. L., Hunt, H. D., Reddy, S. M., Lee, L. F. & Silva, R. F. (2005). The pp38 gene of Marek's disease virus (MDV) is necessary for cytolytic infection of B cells and maintenance of the transformed state but not for cytolytic infection of the feather follicle epithelium and horizontal spread of MDV. *J Virol* **79**, 4545–4549.

Gomi, Y., Sunamachi, H., Mori, Y., Nagaike, K., Takahashi, M. & Yamanishi, K. (2002). Comparison of the complete DNA sequences of the Oka varicella vaccine and its parental virus. *J Virol* **76**, 11447–11459.

Gonzales, M. J., Dugan, J. M. & Shafer, R. W. (2002). Synonymous-non-synonymous mutation rates between sequences containing ambiguous nucleotides (Syn-SCAN). *Bioinformatics* **18**, 886–887.

Hong, Y. & Coussens, P. M. (1994). Identification of an immediate-early gene in the Marek's disease virus long internal repeat region which encodes a unique 14-kilodalton polypeptide. *J Virol* **68**, 3593–3603.

Jarosinski, K. W. & Schat, K. A. (2007). Multiple alternative splicing to exons II and III of viral interleukin-8 (vIL-8) in the Marek's disease virus genome: the importance of vIL-8 exon I. *Virus Genes* **34**, 9–22.

Jarosinski, K. W., Osterrieder, N., Nair, V. K. & Schat, K. A. (2005). Attenuation of Marek's disease virus by deletion of open reading frame RLORF4 but not RLORF5a. *J Virol* **79**, 11647–11659.

Jones, D., Lee, L., Liu, J.-L., Kung, H.-J. & Tillotson, J. K. (1992). Marek's disease virus encodes a basic-leucine zipper gene resembling the fos/jun oncogenes that is highly expressed in lymphoblastoid tumors. *Proc Natl Acad Sci U S A* **89**, 4042–4046.

Jons, A., Gerdts, V., Lange, E., Kaden, V. & Mettenleiter, T. C. (1997). Attenuation of dUTPase-deficient pseudorabies virus for the natural host. *Vet Microbiol* **56**, 47–54.

Jukes, T. H. & Cantor, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism*, pp. 21–132. Edited by H. N. Munro. New York: Academic Press.

Kaiser, P., Hughes, S. & Bumstead, N. (1999). The chicken 9E3/CEF4 CXC chemokine is the avian orthologue of IL8 and maps to chicken chromosome 4 syntenic with genes flanking the mammalian chemokine cluster. *Immunogenetics* **49**, 673–684.

Kamil, J. P., Tischer, B. K., Trapp, S., Nair, V. K., Osterrieder, N. & Kung, H. J. (2005). vLIP, a viral lipase homologue, is a virulence factor of Marek's disease virus. *J Virol* **79**, 6984–6996.

Katoh, K., Kuma, K., Toh, H. & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* **33**, 511–518.

Kattenhorn, L. M., Korbel, G. A., Kessler, B. M., Spooner, E. & Ploegh, H. L. (2005). A deubiquitinating enzyme encoded by HSV-1 belongs to a family of cysteine proteases that is conserved across the family *Herpesviridae*. *Mol Cell* **19**, 547–557.

Kay, B. K., Williamson, M. P. & Sudol, M. (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J* **14**, 231–241.

Klupp, B. G., Fuchs, W., Granzow, H., Nixdorf, R. & Mettenleiter, T. C. (2002). Pseudorabies virus UL36 tegument protein physically interacts with the UL37 protein. *J Virol* **76**, 3065–3071.

Koppers-Lalic, D., Reits, E. A., Ressing, M. E., Lipinska, A. D., Abele, R., Koch, J., Marcondes Rezende, M., Admiraal, P., van Leeuwen, D. & other authors (2005). Varicelloviruses avoid T cell recognition by UL49.5-mediated inactivation of the transporter associated with antigen processing. *Proc Natl Acad Sci U S A* **102**, 5144–5149.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**, 150–163.

Laurent, S., Esnault, E. & Rasschaert, D. (2004). Single-nucleotide polymorphisms in two Marek's disease virus genes (*Meq* and *gD*): application to a retrospective molecular epidemiology study (1982–1999) in France. *J Gen Virol* **85**, 1387–1392.

Lee, L. F., Wu, P., Sui, D., Ren, D., Kamil, J., Kung, H. J. & Witter, R. L. (2000a). The complete unique long sequence and the overall genomic organization of the GA strain of Marek's disease virus. *Proc Natl Acad Sci U S A* **97**, 6091–6096.

Lee, S. I., Takagi, M., Ohashi, K., Sugimoto, C. & Onuma, M. (2000b). Difference in the meq gene between oncogenic and attenuated strains of Marek's disease virus serotype 1. *J Vet Med Sci* **62**, 287–292.

Levy, A. M., Izumiya, Y., Brunovskis, P., Xia, L., Parcells, M. S., Reddy, S. M., Lee, L., Chen, H. W. & Kung, H. J. (2003). Characterization of the chromosomal binding sites and dimerization partners of the viral oncoprotein Meq in Marek's disease virus-transformed T cells. *J Virol* **77**, 12841–12851.

Lipinska, A. D., Koppers-Lalic, D., Rychlowski, M., Admiraal, P., Rijsewijk, F. A., Bienkowska-Szewczyk, K. & Wiertz, E. J. (2006). Bovine herpesvirus 1 UL49.5 protein inhibits the transporter associated with antigen processing despite complex formation with glycoprotein M. *J Virol* **80**, 5822–5832.

Liu, J. L. & Kung, H. J. (2000). Marek's disease herpesvirus transforming protein MEQ: a c-Jun analogue with an alternative life style. *Virus Genes* **21**, 51–64.

Liu, J. L., Lin, S. F., Xia, L., Brunovskis, P., Li, D., Davidson, I., Lee, L. F. & Kung, H. J. (1999). MEQ and V-IL8: cellular genes in disguise? *Acta Virol* **43**, 94–101.

Lupiani, B., Lee, L. F. & Reddy, S. M. (2001). Protein-coding content of the sequence of Marek's disease virus serotype 1. *Curr Top Microbiol Immunol* **255**, 159–190.

Lupiani, B., Lee, L. F., Cui, X., Gimeno, I., Anderson, A., Morgan, R. W., Silva, R. F., Witter, R. L., Kung, H. J. & Reddy, S. M. (2004). Marek's disease virus-encoded Meq gene is involved in transformation of lymphocytes but is dispensable for replication. *Proc Natl Acad Sci U S A* **101**, 11815–11820.

McNabb, D. S. & Courtney, R. J. (1992). Analysis of the UL36 open reading frame encoding the large tegument protein (ICP1/2) of herpes simplex virus type 1. *J Virol* **66**, 7581–7584.

Nair, V. (2005). Evolution of Marek's disease – a paradigm for incessant race between the pathogen and the host. *Vet J* **170**, 175–183.

Niikura, M., Liu, H. C., Dodgson, J. B. & Cheng, H. H. (2004). A comprehensive screen for chicken proteins that interact with proteins unique to virulent strains of Marek's disease virus. *Poult Sci* **83**, 1117–1123.

Niikura, M., Dodgson, J. & Cheng, H. (2006). Direct evidence of host genome acquisition by the alphaherpesvirus Marek's disease virus. *Arch Virol* **151**, 537–549.

Nishiyama, Y. (2004). Herpes simplex virus gene products: the accessories reflect her lifestyle well. *Rev Med Virol* **14**, 33–46.

Osterrieder, K. & Vautherot, J.-F. (2004). The genome content of Marek's disease-like viruses. In *Marek's Disease: an Evolving Problem*, pp. 17–31. Edited by F. Davison & V. Nair. Oxford: Elsevier.

Pagliusi, S. R. & Teresa Aguado, M. (2004). Efficacy and other milestones for human papillomavirus vaccine introduction. *Vaccine* **23**, 569–578.

Parcells, M. S., Lin, S. F., Dienglewicz, R. L., Majerciak, V., Robinson, D. R., Chen, H. C., Wu, Z., Dubyak, G. R., Brunovskis, P. & other authors (2001). Marek's disease virus (MDV) encodes an interleukin-8 homolog (vIL-8): characterization of the vIL-8 protein and a vIL-8 deletion mutant MDV. *J Virol* **75**, 5159–5173.

Peng, Q. & Shirazi, Y. (1996a). Characterization of the protein product encoded by a splicing variant of the Marek's disease virus Eco-Q gene (Meq). *Virology* **226**, 77–82.

Peng, Q. & Shirazi, Y. (1996b). Isolation and characterization of Marek's disease virus (MDV) cDNAs from a MDV-transformed lymphoblastoid cell line: identification of an open reading frame antisense to the MDV Eco-Q protein (Meq). *Virology* **221**, 368–374.

Peng, Q., Zeng, M., Bhuiyan, Z. A., Ubukata, E., Tanaka, A., Nonoyama, M. & Shirazi, Y. (1995). Isolation and characterization of Marek's disease virus (MDV) cDNAs mapping to the BamHI-I2, BamHI-Q2, and BamHI-L fragments of the MDV genome from lymphoblastoid cells transformed and persistently infected with MDV. *Virology* **213**, 590–599.

Petherbridge, L., Howes, K., Baigent, S. J., Sacco, M. A., Evans, S., Osterrieder, N. & Nair, V. (2003). Replication-competent bacterial artificial chromosomes of Marek's disease virus: novel tools for generation of molecularly defined herpesvirus vaccines. *J Virol* **77**, 8712–8718.

Phelan, A., Elliott, G. & O'Hare, P. (1998). Intercellular delivery of functional p53 by the herpesvirus protein VP22. *Nat Biotechnol* **16**, 440–443.

Prince, A. M. (1996). Prevention of liver cancer and cirrhosis by vaccines. *Clin Lab Med* **16**, 493–505.

Qian, Z., Brunovskis, P., Lee, L., Vogt, P. K. & Kung, H. J. (1996). Novel DNA binding specificities of a putative herpesvirus bZIP oncoprotein. *J Virol* **70**, 7161–7170.

Rispens, B. H., van Vloten, H., Mastenbroek, N., Maas, H. J. & Schat, K. A. (1972). Control of Marek's disease in the Netherlands. I.

Isolation of an avirulent Marek's disease virus (strain CVI 988) and its use in laboratory vaccination trials. *Avian Dis* **16**, 108–125.

Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V. & Altschul, S. F. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res* **29**, 2994–3005.

Schat, K. A. & Calnek, B. W. (1978). Characterizations of an apparently non-oncogenic Marek's disease virus. *J Natl Cancer Inst* **60**, 1075–1082.

Schlieker, C., Korbel, G. A., Kattenhorn, L. M. & Ploegh, H. L. (2005). A deubiquitinating activity is conserved in the large tegument protein of the herpesviridae. *J Virol* **79**, 15582–15585.

Shamblin, C. E., Greene, N., Arumugaswami, V., Dienglewicz, R. L. & Parcells, M. S. (2004). Comparative analysis of Marek's disease virus (MDV) glycoprotein-, lytic antigen pp38- and transformation antigen Meq-encoding genes: association of meq mutations with MDVs of high virulence. *Vet Microbiol* **102**, 147–167.

Silva, R. F., Lee, L. F. & Kutish, G. F. (2001). The genomic structure of Marek's disease virus. *Curr Top Microbiol Immunol* **255**, 143–158.

Spatz, S. J. & Silva, R. F. (2007). Polymorphisms in the repeat long regions of oncogenic and attenuated pathotypes of Marek's disease virus 1. *Virus Genes* (in press). doi:10.1007/s11262-006-0024-5

Staden, R., Beal, K. F. & Bonfield, J. K. (2000). The Staden package, 1998. *Methods Mol Biol* **132**, 115–130.

Tischer, B. K., Schumacher, D., Messerle, M., Wagner, M. & Osterrieder, N. (2002). The products of the UL10 (gM) and the UL49.5 genes of Marek's disease virus serotype 1 are essential for virus growth in cultured cells. *J Gen Virol* **83**, 997–1003.

Trapp, S., Parcells, M. S., Kamil, J. P., Schumacher, D., Tischer, B. K., Kumar, P. M., Nair, V. K. & Osterrieder, N. (2006). A virus-encoded telomerase RNA promotes malignant T cell lymphomagenesis. *J Exp Med* **203**, 1307–1317.

Tulman, E. R., Afonso, C. L., Lu, Z., Zsak, L., Rock, D. L. & Kutish, G. F. (2000). The genome of a very virulent Marek's disease virus. *J Virol* **74**, 7980–7988.

Vittone, V., Diefenbach, E., Triffett, D., Douglas, M. W., Cunningham, A. L. & Diefenbach, R. J. (2005). Determination of interactions between tegument proteins of herpes simplex virus type 1. *J Virol* **79**, 9566–9571.

Witter, R. L. (1997). Increased virulence of Marek's disease virus field isolates. *Avian Dis* **41**, 149–163.

Witter, R. L. (2001a). Marek's disease virus vaccines – past, present and future (chicken vs. virus – a battle of the centuries). In *Current Progress on Marek's Disease Research. Proceedings of the 6th International Symposium on Marek's Disease*, pp. 1–9. Edited by K. A. Schat, R. M. Morgan, M. S. Parcells & J. L. Spencer. Kennett Square, PA: American Association of Avian Pathologists.

Witter, R. L. (2001b). Protective efficacy of Marek's disease vaccines. *Curr Top Microbiol Immunol* **255**, 57–90.

Witter, R. L., Nazerian, K., Purchase, H. G. & Burgoyne, G. H. (1970). Isolation from turkeys of a cell-associated herpesvirus antigenically related to Marek's disease virus. *Am J Vet Res* **31**, 525–538.

Zarrinpar, A., Bhattacharyya, R. P. & Lim, W. A. (2003). The structure and function of proline recognition domains. *Sci STKE* **2003**, RE8.