Virology 493 (2016) 22–30

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Exploring the virome of cattle with non-suppurative encephalitis of unknown etiology by metagenomics

Daniel Wüthrich [a,b], Céline L. Boujon [c], Laura Truchet [c], Senija Selimovic-Hamza [b,c], Anna Oevermann [c], Ilias G. Bouzalas [c], Rémy Bruggmann [a], Torsten Seuberlich [c,*]

[a] Interfaculty Bioinformatics Unit, University of Bern and Swiss Institute of Bioinformatics, Bern, Switzerland
[b] Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland
[c] Neurocenter, Division of Neurological Sciences, Vetsuisse Faculty, University of Bern, Bern, Switzerland

ARTICLE INFO

Article history:
Received 20 December 2015
Returned to author for revisions
17 February 2016
Accepted 10 March 2016
Available online 18 March 2016

Keywords:
Encephalitis
Cattle
Metagenomics
Herpesvirus
Astrovirus
Polyomavirus
Parainfluenza virus
Retrovirus
Neurological disease

ABSTRACT

Non-suppurative encephalitis is one of the most frequent pathological diagnosis in cattle with neurological disease, but there is a gap in the knowledge on disease-associated pathogens. In order to identify viruses that are associated with non-suppurative encephalitis in cattle, we used a viral metagenomics approach on a sample set of 16 neurologically-diseased cows. We detected six virus candidates: parainfluenza virus 5 (PIV-5), bovine astrovirus CH13/NeuroS1 (BoAstV-CH13/NeuroS1), bovine polyomavirus 2 (BPyV-2 SF), ovine herpesvirus 2 (OvHV-2), bovine herpesvirus 6 (BHV-6) and a novel bovine betaretrovirus termed BoRV-CH15. In a case-control study using PCR, BoAstV-CH13 ($p=0.046$), BoPV-2 SF ($p=0.005$) and BoHV-6 ($p=4.3E-05$) were statistically associated with the disease. These data expand our knowledge on encephalitis-associated pathogens in cattle and point to the value of NGS in resolving complex infection scenarios in a clinical disease setting.

1. Introduction

Neuroinfectious diseases in livestock, such as Japanese Encephalitis, Rabies, Schmallenberg disease, Aujeszky's disease, and Nipah Encephalitis have implications for animal welfare, public health, and international trade. In many countries, they are subject to disease control measures. A main pillar of disease control is passive surveillance, involving notification of clinically suspicious animals and subsequent pathological as well as laboratory testing.

Since the 1960s our division has systematically investigated brains of cattle with neurological disease in the framework of neuropathological services and disease surveillance (Theil et al., 1998; Fatzer, 1971; Fatzer and Steck, 1974; Heim et al., 1997). In most animals the diagnosis was conclusive, either by the identification of typical histopathological lesions or follow-up diagnostics. Yet a considerable proportion of cattle (10–15%) was diagnosed with a non-suppurative meningoencephalitis, suggesting a viral infection, but remained etiologically unresolved despite attempts to identify various infectious agents (Theil et al., 1998; Bestetti et al., 1976). This type of encephalitis has also been referred to as European sporadic bovine encephalitis (Vandevelde et al., 2012; Fankhauser, 1961) and similar cases have been reported from other countries, although sometimes using different terminologies (Sanchez et al., 2013; Bachmann et al., 1975; Jeffrey, 1992; Bozzetta et al., 2003).

Viral metagenomics has evolved as a new technique to identify known and novel viruses using unbiased next-generation sequencing (NGS) and bioinformatics. Novel types of bovine astroviruses (BoAstV) have been identified by viral metagenomics as a possible cause of non-suppurative encephalitis in cattle (Bouzalas et al., 2014; Li et al., 2013; Seuberlich et al., In press). BoAstV-CH13, which is very similar to the previously described BoAstV-NeuroS1 in the USA (Li et al., 2013), has been detected in around 25% of the cattle with non-suppurative encephalitis of unknown etiology in Switzerland. Nevertheless, the etiology in the remaining 75% of these animals remained unknown.

The objective of the present study was to identify viruses in addition to BoAstV-CH13/NeuroS1 that are associated with neurological disease and non-suppurative encephalitis in cattle. We subjected a set of 16 cattle brain samples to NGS and refined the bioinformatics pipeline for pathogen detection and discovery. We

* Correspondence to: Neurocenter-Division of Neurological Diseases, Department of Clinical Research and Veterinary Public Health, Vetsuisse Faculty, University of Berne, Bremgartenstrasse 109a, CH 3001 Bern, Switzerland
E-mail address: torsten.seuberlich@vetsuisse.unibe.ch (T. Seuberlich).

http://dx.doi.org/10.1016/j.virol.2016.03.009

found a series of known and novel candidate viruses comprising (i) parainfluenza virus 5 (PIV-5), (ii) bovine polyomavirus 2 (BPyV-2 SF), (iii) BoAstV-CH13/NeuroS1, (iv) ovine herpesvirus 2 (OvHV-2), (v) bovine herpesvirus 6 (BHV-6), and (vi) a previously unknown betaretrovirus, tentatively named bovine retrovirus (BoRV) CH15. This study not only provides novel insights into the spectrum of viruses that are potentially involved in bovine encephalitis but also points to opportunities and challenges related to pathogen identification and discovery by viral metagenomics using clinical specimens.

2. Material and methods

2.1. Tissue samples

Brain tissues of bovines with non-suppurative encephalitis (cases, $n=29$) and absence of histopathological brain lesions (controls, $n=50$) were selected from the tissue bank of the Division of Neurological Sciences, Vetsuisse Faculty, University of Bern. All cases presented neurological signs consistent with central nervous system (CNS) disease and their brains were submitted to our neuropathological diagnostic service between the years 1995 and 2015. Tissues of control animals were collected from animals of unknown disease status that died on farms or were euthanized for reasons other than human consumption. Standard sample processing consisted of a sagittal median cut of the entire brain (cases) or of two adjacent coronal sections of the medulla oblongata (controls). One half/section was then formalin-fixed and processed for histopathology and the second half/section was stored frozen at -80 °C. For NGS and PCR assays, tissue pieces of ≤ 25 mg were cut from frozen brainstem and used for DNA and RNA extraction. DNA extraction was performed with DNeasy Blood and Tissue Kit (Qiagen) and RNA was extracted using Trizol reagent (Thermo Fisher Scientific), both according to the manufacturer's instructions.

2.2. Next-generation sequencing

DNA and RNA libraries were constructed from nucleic acid extracts of 16 animals with non-suppurative encephalitis of unknown etiology using the TruSeq DNA Sample Preparation kit (Illumina) and the TruSeq Stranded Total RNA Kit (Illumina), respectively. For RNA libraries ribosomal RNA was selectively depleted with the RiboMinus kit (Thermo Fisher Scientific). Half a lane of paired-end reads (2×100 bp) per library were collected using an Illumina HiSeq2500 instrument. The sequencing of the DNA samples yielded between 43,127,450 (DNA27020) and 96,661,539 (DNA43484) reads and for the RNA samples between 57,746,275 (RNA34510) and 112,846,881 (RNA43661) reads (Table S1).

2.3. Read mapping

Reads of each library were mapped to the RefSeq viral genome database (6th May 2015) using Bowtie2 (version 2.2.1,–sensitive) (Langmead and Salzberg, 2012). The read depth was determined using GATK (version 3.3.0,-T DepthOfCoverage) (McKenna et al., 2010). All viral genomes of which at least 5% were covered by reads were selected for further analysis. The distances between the start sites of the forward reads were tested for the expected mean distance of the read start sites, i.e. the length of the virus genome divided by the number of mapped reads, using a *t*-test. Genomes with a p-value <0.05 were excluded.

2.4. De novo assembly

Reads of RNA and DNA libraries were mapped to the bovine reference genome (version UMD 3.1 (dna_sm.toplevel), www.ensembl.org) using STAR (version 2.3.0, default parameters) (Dobin et al., 2013) and Bowtie2 (version 2.2.1, default parameters) (Langmead and Salzberg, 2012), respectively. The unmapped reads were quality selected using Trimmomatic (version 0.30, options: SLIDINGWINDOW:4:15 MINLEN:101) (Bolger et al., 2014). Selected reads were *de novo* assembled with SPAdes (version 3.1.1, options:–sc -k 21, 33, 55, 77, 91, 95, 97, 99) (Bankevich et al., 2012). The contiguous sequences (contigs) were scaffolded using SSPACE (version 3.0, default parameters) (Boetzer et al., 2011). The resulting sequences were then aligned to entries of the non-redundant nucleotide database of NCBI (30th June 2015) using BLASTN (version 2.2.29+, default parameters) (Altschul et al., 1990).

2.5. Virus discovery pipeline

The contigs from the *de novo* assembly with a minimal length of 500 bp were aligned to entries of the viral protein databases of uniprot (tremble_viruses 24th June 2015, sprot_viruses 24th June 2015) by BLASTX (version 2.2.29+, default parameters) (Altschul et al., 1990). The alignments were selected if the identity was at least 30% and if the virus proteins were covered by at least 30%. Alignments to phage proteins were excluded. Only contigs fulfilling these criteria were further analyzed and were aligned to entries of the non-redundant nucleotide database of NCBI (30th June 2015) using BLASTN (version 2.2.29+, default parameters) (Altschul et al., 1990). Contigs that showed an alignment length of 80 bp or longer to a bovine sequence were excluded. An additional filtering was performed to remove contaminations from other sources (e.g. ruminant, bacteria, parasites, etc.) by excluding sequences with a best BLASTN hit $\geq 10\%$ identity to a mammal, parasite or bacterial sequences using the non-redundant nucleotide database of NCBI (30th June 2015). Furthermore, the contigs were aligned by BLASTN (version 2.2.29+, default parameters) to all mammalian genomes in the RefSeq database (16th July 2015) and excluded when showing $\geq 10\%$ identity to a mammalian genome. The remaining contigs were assigned to the virus with the highest homology of the viral proteins with the contig sequences.

2.6. Phylogenetic analysis

The molecular phylogenetic tree was generated with MEGA6 software (Tamura et al., 2013) by the maximum likelihood method using the Tamura-Nei model. The analyzes involved the BoRV-CH15 contig of animal 28015 (DNA library) and 15 additional full genome sequences of representative members of the retroviridae (GenBank accession numbers are given in brackets): Avian leukemia virus (NC_015116.1), Rous sarcoma virus (NC_001407.1), Mouse mammary tumor virus (NC_001503.1), Mason-Pfizer monkey virus (NC_001550.1), Jaagsiekte sheep retrovirus (NC_001494.1), Feline leukemia virus (NC_001940.1), Murine leukemia virus (NC_001362.1), Bovine leukemia virus (NC_001414.1), Human T-lymphotropic virus 1 (NC_001436.1), Walleye dermal sarcoma virus (NC_001867.1), Human immunodeficiency virus 1 (NC_001802.1), Human immunodeficiency virus 2 (NC_001722.1), Equine infectious anemia virus (NC_001450.1), Visna/Maedi virus (NC_001452.1), Simian foamy virus (NC_001364.1).

2.7. PCR assays

RT-PCR was performed with the OneTaq One-step PCR Kit (New England Biolabs) according to the manufacturer's instructions using primers MA2 and MA4 (Mittelholzer et al., 2003) for BoAstV-CH13/NeuroS1 and primers PIV-5_1L and PIV-5_2R as well as PIV-5_5L and PIV-5_6R for PIV-5, which target the coding region for the PIV-5 N-Protein and L-Protein respectively. Detection of BoPyV-2 DNA was accomplished by PCR with previously published primers BPyV3 and BPyV4 and the GoTaq Green Master Mix (Promega) (Zhang et al., 2014). Similarly, BoRV-CH15 provirus DNA was detected in three different PCR reactions with primer combinations BoRV_ENV_L and BoRV_ENV_R, BoRV_GAG_L and BoRV_GAG_R and BoRV_POL_L and BoRV_POL_R, respectively. Primers BoRV_ENV_L and BoRV_ENV_R were used for the screening of encephalitis cases and control animals. For the detection of OvHV-2 and BHV-6, we adapted previously published qPCR protocols to the Path-ID qPCR Master Mix (Applied Biosystems) (Stahel et al., 2013; Kubiś et al., 2013). The cut-off value for a positive qPCR result for BHV-6 was set to cycle threshold (ct) values < 35, because higher ct have shown to be out of the linear range of the assay. For the OvHV-2 qPCR the positive cut-off was set to ct values < 40 (Fig. S1). All primers and probes as well as the PCR cycle parameters are provided in Table S2.

3. Results

3.1. Detection of known viruses

We selected frozen brain tissues of 16 neurologically diseased cattle that were affected by non-suppurative encephalitis, which suggested the presence of a viral infection and submitted them to next generation sequencing of RNA and DNA libraries (Table 1). All the collected Illumina reads were mapped against the RefSeq viral genome database. To identify candidate viruses we established a filter consisting of two criteria: (i) at least 5% of a given viral genome are covered by reads and (ii) the mapping sites of reads are equally distributed over the viral genomes (as assessed by *t*-test, $p > 0.05$). The bioinformatics pipeline for the detection of known viruses is illustrated in Fig. 1. Applying these criteria, we found reads in 10 animals that aligned to eight different viruses. Among these viruses were five, for which reads mapped with a similarity of 98% to 100% to different parts of the virus genomes (Fig. 2, Table S3): BoAstV-CH13/NeuroS1 ($n=2$ animals), BHV-6 ($n=1$), BPyV2-SF ($n=2$), OvHV-2 ($n=1$) and PIV-5 ($n=1$). Targeted molecular testing by RT-PCR, PCR and qPCR confirmed the presence of these five viruses in the respective animals (Table 1).

The situation was different for two other viruses that were identified by our pipeline: Murine osteosarcoma virus and avian Y73 sarcoma virus. In both instances, reads originating from

Table 1
Virus detection in cattle with non-suppurative encephalitis and control cattle by next-generation sequencing, conventional PCR, reverse transcription PCR and quantitative PCR.

Animal			RT-PCR		PCR		qPCR [ct]	
ID	Year of diagnosis	Age (years)	BoAstV	PIV-5	BoPyV	BoRV-CH15	BHV6[b]	OvHV-2[c]
Next-generation sequencing								
24586	1996	4	–	–	–	–	–	–
25018	1996	na	–	–	–	+[a]	–	–
26324	1997	8	–	–	–	–	–	20.1[a,d]
26731	1998	5	–	+[a]	+	–	33.4	–
27020	1998	3	–	+	+	–	24.7	–
31292	1999	10	–	–	–	–	–	–
32450	2000	7	–	–	+	–	30.6	–
33181	2001	2	–	–	–	–	–	–
34421	2002	2	–	–	–	–	–	–
34510	2002	2	–	–	+[a]	–	–	–
36166	2003	5	–	–	–	–	–	–
36768	2004	2.5	–	–	–	–	–	–
42268	2006	6	–	–	–	–	32.9	–
42797	2007	1.5	–	–	+[a]	–	–	–
43484	2008	2	+[a]	–	–	–	–	–
43661	2009	6	+[a]	–	–	–	25.6[a]	–
Retrospective study								
24250	1995	5	–	–	–	–	29.4	–
31300	1999	3	–	–	–	–	–	–
32270	2000	2	–	–	–	–	–	–
35154	2002	4	–	–	nd	nd	nd	nd
35946	2003	11	–	–	–	–	29.4	–
36108	2003	5	–	–	–	–	–	35.6
41093	2005	10	–	–	–	–	–	25.4[d]
42145	2006	6	–	–	–	–	25.1	–
42295	2006	10	–	–	–	–	–	22.0[d]
44282	2010	8	–	–	–	–	31.1	17.6[d]
45641	2012	10	–	–	–	–	–	36.8
50772	2014	4	–	–	–	–	–	–
50773	2015	3.5	+	–	–	–	–	–
Control animals (positive/total)								
	2012	≥ 3	0/50	0/50	0/49	1/49	0/49	4/49

na, not available; nd, not done; +, positive; –, negative

[a] Virus detection by NGS and confirmed by PCR.
[b] Cycle threshold (ct), ct values < 35 are considered as positive.
[c] ct values < 40 are considered as positive, ct values in the control animals ranged from 32 to 35.
[d] Animals with non-suppurative encephalitis and histopathological features of malignant catharral fever.

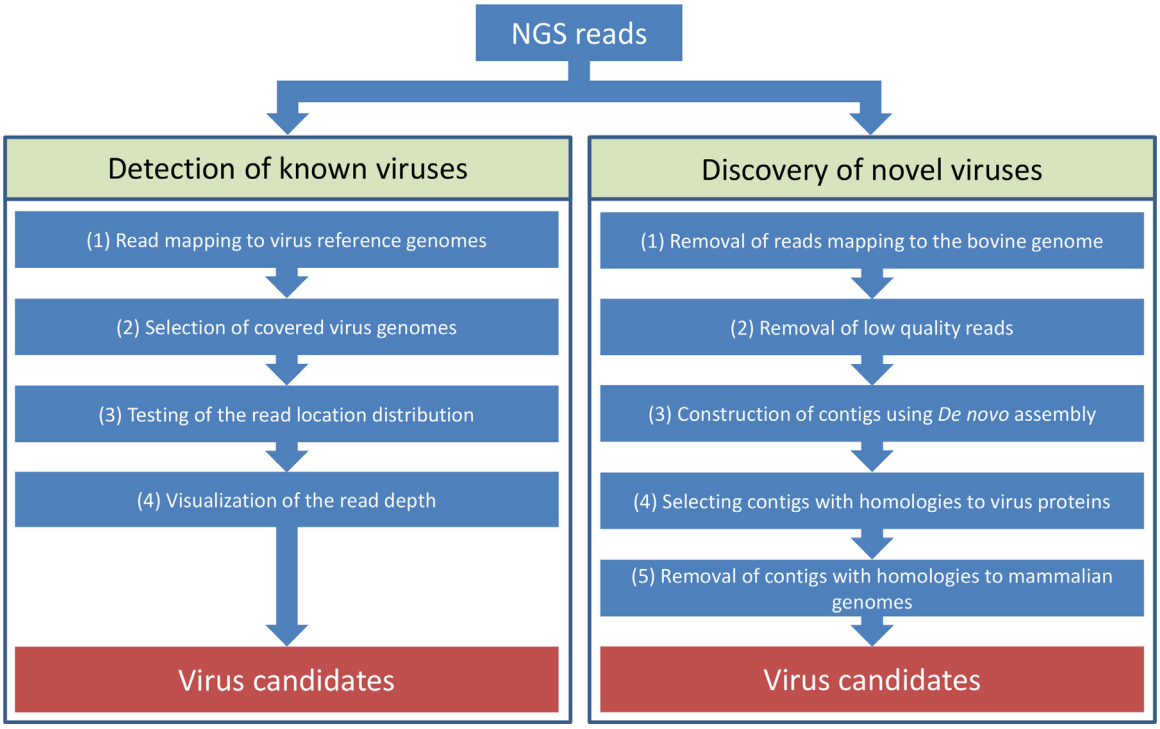

Fig. 1. *Illustration of the virus detection pipelines.* The flowchart is summarizing the two basic bioinformatics pipelines that were used to detect viruses. The NGS reads of the brain samples were analyzed in two separate pipelines. In the first pipeline (detection of known viruses) the reads were remapped to all virus reference genomes (1). The reads of each genome that was covered by reads (2), were tested for the distribution across the genome (3). If no enrichment for specific regions was found on the genomes, the read depth was visualized (4). In the second pipeline (discovery of novel viruses) the reads were mapped to the bovine reference genome to remove host DNA and RNA (1). The remaining reads were quality filtered (2) and *de novo* assembled to build contigs (3). The resulting contigs were aligned to a virus protein database to select contigs with possible viral origin (4). The selected contigs were aligned to all eukaryotic genomes to remove contigs with possible eukaryotic origin (5).

different animals mapped only to two specific regions of the virus genomes (Fig. 2) and with a similarity of only 90% to 92% (Table S3). When we analyzed these genomes by BLASTN the Murine osteosarcoma virus revealed a partial homology to a bovine oncogene (BT029837.1, sequence identity: 85%, region: 1567-2, 176 bp), and to a bovine ubiquitin-like/S30 ribosomal fusion protein (AF520959.1, sequence identity: 86%, region: 2269–2716 bp). The Y73 sarcoma virus genome showed homology to a bovine proto-oncogene (XM_010811511.1, sequence identity 83%, region: 704–2310 bp). The regions of the virus genomes with similarity to bovine sequences overlapped with the regions that were covered by reads (Fig. 2, Table S3). We therefore concluded that these reads originate from the bovine genome rather than representing virus specific reads.

In one animal (ID 43484), we found reads in the RNA library that mapped to Ryegrass mottle virus with a sequence identity of 97% (Fig. 2). BLASTN analysis of the virus genome did not reveal homologies to bovine sequences, supporting the notion that the viral RNA was indeed present in the sample. This was unexpected since this virus belongs to the genus Sobemovirus (Zhang et al., 2001), a positive stranded RNA viruses exclusively affecting plants. It is possible that the brain tissue under investigation was contaminated with plant material during the sampling procedure. Therefore, and in view of the lack of biological plausibility, we excluded Ryegrass mottle virus from further analysis.

3.2. Discovery of novel viruses

We extended our search to viruses of which the nucleotide sequence is as yet unknown. We filtered reads for bovine sequences by mapping them against the bovine reference genome. The remaining unmapped reads were then subjected to *de novo* assembly and analyzed for homologies to known viral genomes and proteins by BLASTN and BLASTX.

For a first classification, *de novo* assembled contigs were aligned to entries of the non-redundant nucleotide database from NCBI, their best blast hit was determined (minimal identity 10%) and they were assigned to the corresponding phylogenetic clades (Fig. 3). Most contigs show similarity to ruminant sequences (mean 47.8%, sd 10.8%) and sequences from other eukaryotes (mean 21.5%, sd 5.8%), however a high fraction of the contigs did not have any blast hit with the selected cutoff (mean 26.1%, sd 9.6%). Noteworthy is the low number of contigs showing similarity to bacterial sequences in the DNA samples assemblies (mean 0.2%) compared to the RNA sample assemblies (mean 8.8%). The fraction of contigs matching to virus sequences (mean $< 0.1\%$) is very small. The highest fraction of virus homologous sequences were found in sample RNA26324 (0.9%) in which OvHV-2 was detected. Taken together, these data show that the vast majority of *de novo* assembled contigs were not of viral origin but represented either ruminant, other eukaryotic or unassigned sequences.

To identify novel virus sequences, we scanned the contigs of the *de novo* assembly using our virus discovery pipeline and a filter set to a minimum identity of 30% on the amino acid level and a minimum coverage of 30% of the viral protein. We found 732 contigs with a total size of 492,336 bp in 16 of the DNA samples and 7 of the RNA samples with similarities to proteins of 446 different viruses. However, many of these contigs also showed nucleotide sequence homology to different sources (Table S4). To further select contigs that correspond to viral sequences, we first excluded all contigs that showed homology ($> 10\%$ identity) to sequences of mammal, parasitic or bacterial origin. The remaining contigs were assigned to the viruses, which showed to highest protein sequence homology to them. The bioinformatics pipeline for the discovery of novel viruses is shown in Fig. 1. This approach

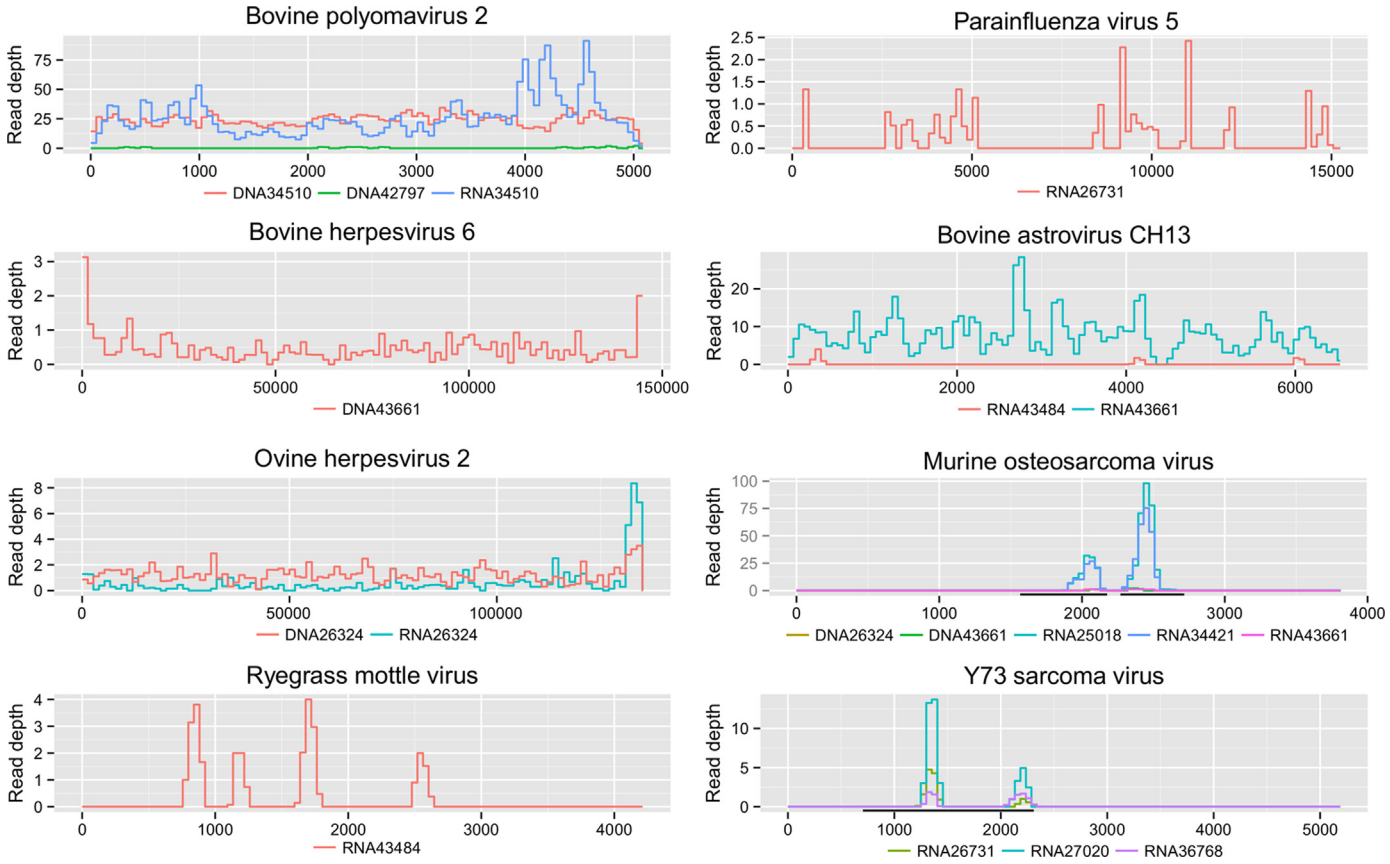

Fig. 2. *Read depth and coverage of identified virus genomes in cattle with encephalitis.* The different plots show the mean read depth of different viruses. The mean read depth was calculated with a sliding window of the length of one percent of the genome size of the corresponding virus. The *x*-axis represents the genomic location [bp]. The *y*-axis indicates the mean read depth. The colors of the lines indicate the different samples. In the plots of the Y73 sarcoma virus and the Murine osteosarcoma virus, the homologies to bovine sequences are marked with black lines.

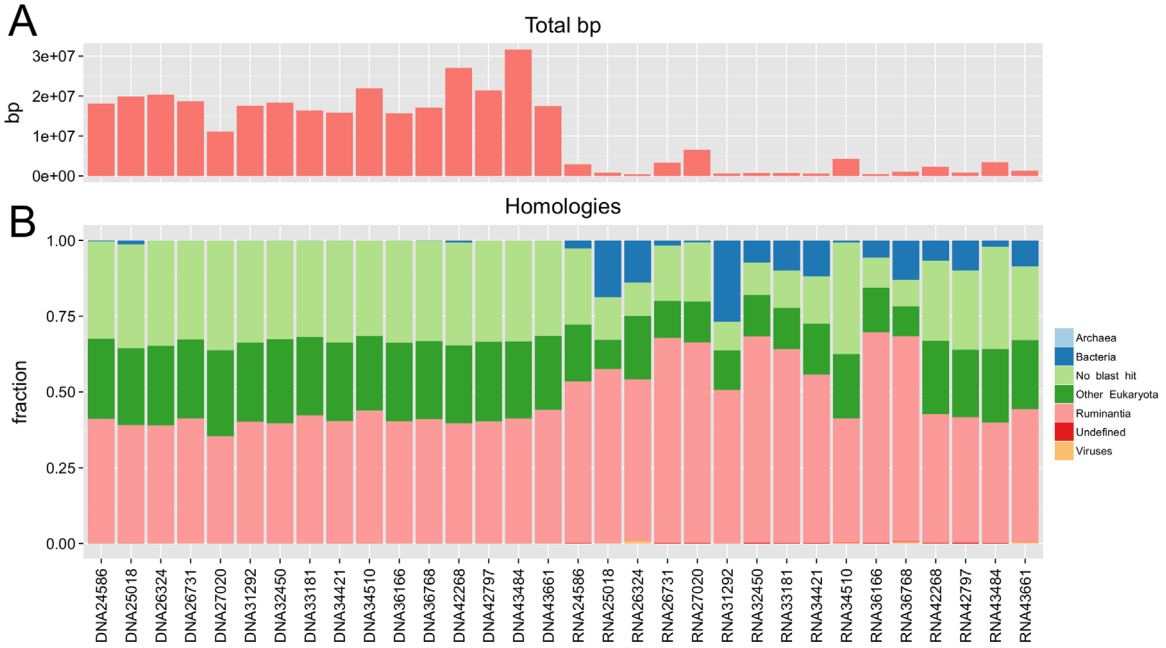

Fig. 3. *Phylogenetic assignment of the assembled contigs.* (A) The *y*-axis shows the total length of the contigs build from the different samples. (B) The bar plots indicate for each sample the origin of the best BLASTN hit of the non-redundant nucleotide database. Different colors indicate phylogenetic clades. The *y*-axis indicates the fraction of the total number of nucleotides of the contigs with homology to a specific clade.

resulted in seven virus candidates (Table S5). Three of these candidates were already identified in the virus identification procedure: BoAstV-CH13/NeuroS1 in sample RNA43661, BoPyV-2 in samples RNA34510 and DNA34510, and OvHV-2 in sample DNA26324. Additionally, we found contigs with similarity to proteins of Coxsackieviruses, Human Herpesviruses and African Swine

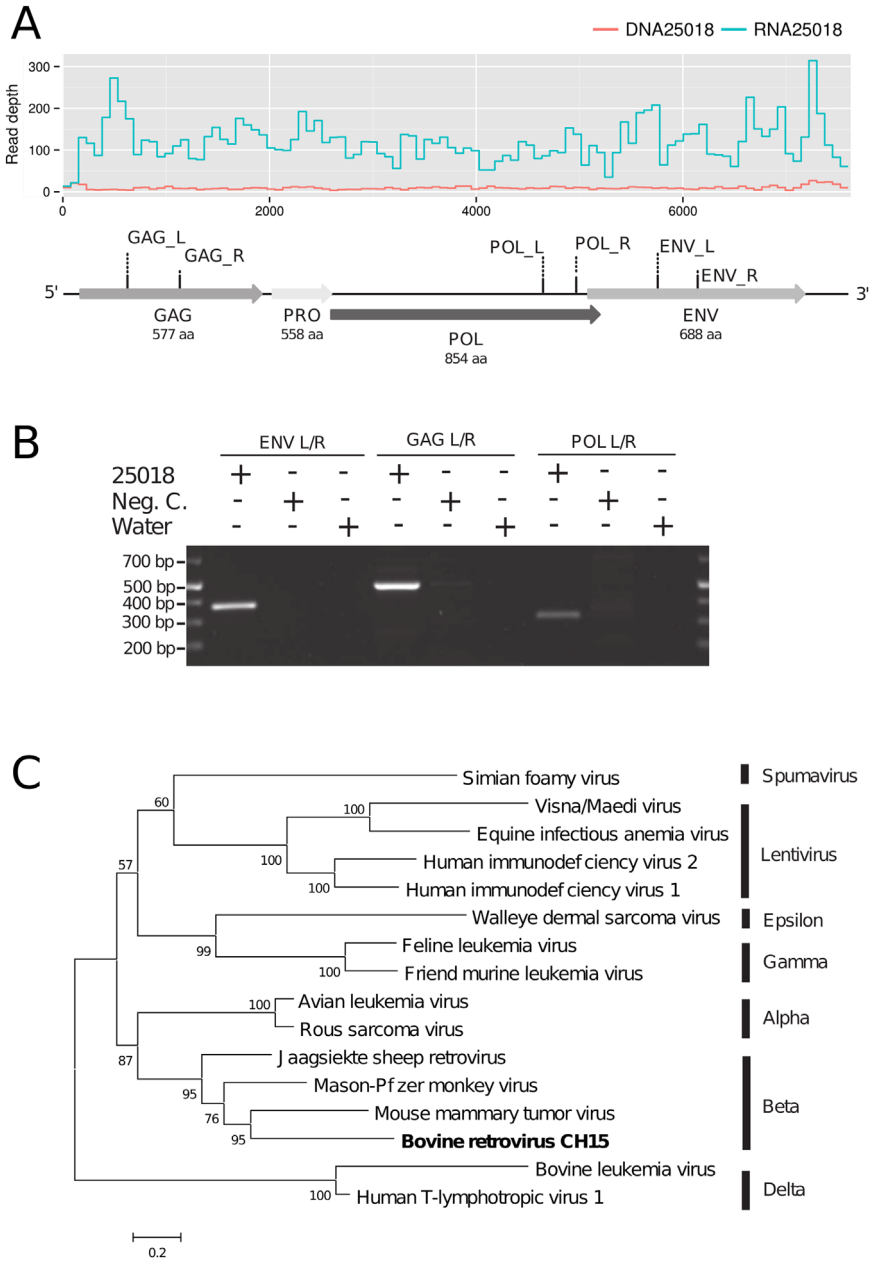

Fig. 4. *Identification of the novel bovine retrovirus (BoRV) CH15 in the brain tissue of a cow with encephalitis.* (A) Schematic representation of the putative BoRV-CH15 genome and the open reading frames encoding the group-specific antigen (GAG), the protease (PRO), the DNA polymerase (POL) and the envelope (ENV) proteins. Primer binding sites of forward (L) and reverse primers (R) for PCR are indicated. The top graph shows the mean read depth of a sliding window with the length of one percent of the genome length. The *y*-axis indicates the read depth and the *x*-axis the genomic location in bp for the DNA (red) and the RNA (green) libraries, respectively. (B) PCR confirmation of the BoRV provirus in brain tissue of animal 28015. An animal without brain lesions served as negative control (neg. c.). (C) Phylogenetic analysis of the BoRV-CH15 genome sequence and representative virus genomes of the retroviridae by the maximum-likelihood method. Genera within the retroviridae family are indicated on the right.

Fever Virus. However, these contigs primarily consisted of repetitions of short sequence motives and therefore must be considered as artifacts. Strikingly, in one animal (25018) we identified two large contigs with homology to the proteins of a simian retrovirus with a size of 7599 bp and 7470 bp in the DNA and RNA libraries of animal 25018, respectively (Table S5).

3.3. Bovine retrovirus CH15

Alignments of the simian retrovirus-like contigs in RNA25018 and DNA25018 showed that the two sequences were perfectly identical over 7458 bp. Remapping of the reads of both samples to the DNA25018 contig resulted in full sequence coverage with a mean read depth of about 8x for the DNA library and 106x for the RNA library (Fig. 4A). Sequence analysis revealed four partially overlapping open reading frames (ORFs), which is characteristic for betaretroviruses, and presumably encoding for the group-specific antigen (GAG), the protease (PRO), the DNA polymerase (POL) and the envelope protein (ENV). The presence of this sequence was confirmed by three different PCR protocols targeting the GAG, ENV and POL ORFs (Fig. 4B). Phylogenetic comparison of the DNA25018 contig with representative genomes of members of the *retroviridae* family indicated an association with betaretroviruses with closest genetic relationship to the Mouse

mammary tumor virus (Fig. 4C). Taken together these data support the presence of a previously unknown retrovirus in the brain of this cow, that we tentatively termed bovine retrovirus CH15 (GenBank accession number: KU720628).

3.4. Case-control study

In order to assess the association of the virus candidates with non-suppurative encephalitis and the sensitivity of virus detection by NGS, we investigated the nucleic acid extracts of all 16 cases submitted to NGS by targeted PCR and RT-PCR for the presence of BoAstV CH13/NeuroS1, BoPyV-2 SF, PIV-5, BoRV-CH15, BHV-6 and OvHV-6. An additional set of 13 cases of encephalitis of unknown etiology and 50 control animals without brain lesions was similarly tested. In the encephalitis cases, including those that were analyzed by NGS, we detected additional positive cases for each of the viruses, except for BoRV-CH15 (Table 1). BoRV-CH15 and OvHV-2 were also found in animals of the control group. Statistical analysis by Fisher's exact test ($p<0.05$) revealed that not only BoAstV-CH13 ($p=0.046$), but also BoPV2-SF ($p=0.005$) and BoHV-6 ($p=4.3E-05$) was associated with non-suppurative encephalitis. Noteworthy in 14 animals, only one virus was detected, and dual as well as triple infections were found in 3 and 2 animals, respectively (Table 1).

4. Discussion

We applied two strategies for virus identification to tissue samples of cattle with non-suppurative encephalitis. First, we aimed to detect known viruses by mapping NGS reads to virus database entries and second, we developed and applied a virus discovery pipeline by searching for similarities of *de novo* assembled sequences with viral genomes and proteins.

We found that the combination of mapping reads to known virus genomes and protein homology searches on *de novo* assembled non-host reads is suitable to detect viral sequences in clinical brain samples of cattle. The strength of the read mapping to known virus genomes is its relative high sensitivity. For instance, only four read-pairs were sufficient to detect the BoAstV-CH13/NeuroS1 in the RNA extract of animal 43484. Moreover, it allows the simultaneous screening of thousands of viral genomes unlike the more sensitive PCR-based assays. Also, only highly similar viruses with an identity of $>85\%$ can be detected. Viruses have very diverse genomes (Koonin and Dolja, 2006), and novel viruses might be too different from one another to be detectable by the mapping approach. Therefore, we also applied a *de novo* assembly based method to detect novel viruses that are distinct from known ones. This method, however, requires a sufficient number of reads of the virus genome to be represented in a sample and a minimum similarity to known viruses in the database. The lower sensitivity of the *de novo* assembly approach is reflected in the finding that we were not able to assemble contigs of viral sequences for all candidate viruses and animals that were detected in the read mapping.

A crucial step in viral metagenomics using clinical specimen is the depletion of host sequences. We depleted the host sequences *in silico* by mapping the reads against the reference host genome. However, we found that the contigs assembled from the remaining reads mostly originated from ruminant as well as other eukaryotic sources and only a small fraction was of viral origin (Fig. 3), indicating that a large proportion of host sequences were not removed. This incomplete depletion can be partly attributed to polymorphisms (e.g. single nucleotide polymorphism and indels) between the examined animals and the published reference sequence as well as the incomplete assembly status of the bovine reference genome. Furthermore, DNA and RNA contaminations from other sources can be introduced during the sample extraction and library preparation. This was likely the case for the Ryegrass mottle virus. Finally, removal of mammalian, bacterial and parasitic sequences to a high level were critical to find biologically plausible virus candidates. We are currently working on a k-mer based method that allows the subtraction of host reads from infected individuals using reads from healthy animals. We have preliminary results that show an almost complete removal of host reads. This more complete host read depletion can be explained by the fact that we do not rely on an incomplete host genome assembly where parts of the genome sequence are still missing.

The use of total RNA and DNA extracts of brain tissues as a starting material for NGS has advantages and disadvantages. On the one hand, this approach results in massive amounts of host-derived sequences, which compete with viral sequences in the sequencing procedure and reduce the overall sensitivity of virus detection. As a result, viruses may remain undetected in NGS, although being readily identified in conventional and quantitative PCR assays (Table 1). This downside could be overcome by filtration and nuclease treatment of tissues, which depletes free host nucleic acids and enriches for viral capsid protected nucleic acids. Indeed, we observed a strong increase in read-depth by at least two orders of magnitude in BoAstV-CH13 positive tissue samples that were nuclease treated when compared to the read depths in untreated samples (I.G. Bouzalas, unpublished data). On the other hand, viruses with episomal and proviral latency as well as unpackaged viral RNA genomes or transcripts will be degraded by these procedures. For instance, when we analyzed an RNA extract of the BoRV-CH15 positive tissue sample after filtration and nuclease treatment by NGS, the virus remained undetected in the discovery pipeline (I.G. Bouzalas, unpublished data). In the present study we have opted for a procedure that detects a broad spectrum of virus candidates. For some viruses the sensitivity of detection may be increased by alternative data processing and tissue pretreatment procedures which could then facilitate virus identification in those cases of bovine non-suppurative encephalitis that still remain unresolved.

Viruses known to be associated with encephalitis in cattle as previously described are BoAstV-CH13/NeuroS1 (Bouzalas et al., 2014; Li et al., 2013), and OvHV-2 (Wiyono et al., 1994). Our aim was not to specifically target BoAstV-CH13/NeuroS1, and animals with a known positive status were not included. Yet, we identified three BoAstV-CH13/NeuroS1 positive animals by NGS and RT-PCR. None of the control animals was BoAstV-CH13/NeuroS1 positive. Taken together, this further supports the association of the virus with the disease. OvHV-2 has consistently been isolated from cattle with so-called sheep-associated malignant catarrhal fever (MCF) (Wiyono et al., 1994). We found several animals with encephalitis positive for OvHV-2. However, four animals without lesions indicative of encephalitis were similarly OvHV-2 positive. Possible explanations are that these animals were affected by a form of MCF, which did not present lesions in the brain, or alternatively, as described for cattle, bison and water buffaloes, the infection may have been subclinical (Stahel et al., 2013; O'Toole et al., 2002; Sausker and Dyer, 2002; Powers et al., 2005).

PIV-5, BPyV-2 SF and BHV-6 are not known to be pathogenic in cattle, however, all of them have been detected in cattle tissues on different occasions: (i) the paramyxovirus PIV-5, formerly known as simian virus 5 (SV5), has very recently been reported as a possible cause of severe respiratory disease in calves in China (Liu et al., 2015), (ii) BPyV-2 SF was discovered in two independent metagenomics studies analyzing commercial beef samples (Zhang et al., 2014; Peretti et al., 2015) and (iii) BHV-6, also termed bovine lymphotropic herpesvirus, has been identified in bovine peripheral blood mononuclear cells in symptomless adult cattle with very high frequency (91%) (Rovnak et al., 1998).

Interestingly, there has been antigenic and morphological evidence for the involvement of paramyxoviruses in cases of European sporadic

non-suppurative bovine encephalitis. In brain tissues of some cows there was immunoreactivity with antibodies against canine distemper virus and rinderpest virus (Theil et al., 1998), both members of the genus morbillivirus within the family paramyxoviridae. Ultrastructural findings in infected cells supported the presence of a morbillivirus (Bachmann et al., 1975). However, specific molecular testing for paramyxoviruses was not undertaken at this point in time. Whether these findings can be explained by an infection with PIV-5 or another paramyxovirus remains to be investigated.

Members of the polyomaviridae such as JC-virus and human polyomavirus 6 have been reported in brains of human patients with progressive multifocal leukencephalopathy (Brooks and Walker, 1984; Delbue et al., 2015). Thus, there is precedent for this family of viruses causing neurological infection and disease. On the other hand, polyomaviruses are commonly occurring without causing a disease. Further research is needed to clarify the role of BPyV-2 SF in the pathogenesis of non-suppurative in cattle.

In the virus discovery pipeline, we identified in one animal contigs in the RNA and DNA libraries that likely represent the genome of a previously unknown betaretrovirus, BoRV-CH15. We found in total 129 read pairs in which one read aligned to the terminal regions of BoRV-CH15 sequence and the corresponding mate read aligned to the Bos taurus genome. Mate reads were assigned to 18 different bovine chromosomes (D. Wüthrich, unpublished data). This argues for the integration of the viral genome into the host genome as a provirus at multiple sites and transcription of host-sequence-flanked provirus sequences to mRNA, which suggests an active infection. Yet, there is evidence for BoRV-CH15 infection in one animal of the control group. Therefore, the virus may not be strictly associated with encephalitis.

Collectively, these results expand our knowledge on the spectrum of virus candidates causing non-suppurative encephalitis in cattle. In many animals under investigation more than one virus was identified, adding to the complexity in determining the causative relationship between the presence of a virus and the disease. This aspect needs further investigation by *in situ* detection of viral nucleic acids and antigens, virus isolation, and finally, transmission experiments.

Funding information

This study was funded by the Swiss Federal Food Safety and Veterinary Office.

Acknowledgments

We thank the Next Generation Sequencing Platform of the University of Bern, in particular Cord Drögemüller and Tosso Leeb, for performing the high-throughput sequencing experiments as well as Michèle Ackermann and Muriel Fragnière for excellent technical support.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2016.03.009.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W.L.D., 1990. Basic local alignment search tool. J Mol. Biol. 5, 403–410.

Bachmann, P.A., ter Meulen, V., Jentsch, G., Appel, M., Iwasaki, Y., Meyermann, R., Koprowski, H., Mayr, A., 1975. Sporadic bovine meningo-encephalitis-isolation of a paramyxovirus. Arch. Virol. 48, 107–120.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. a, Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin V, A., Sirotkin V, A., Vyahhi, N., Tesler, G., Alekseyev, M. a, Pevzner, P. a, 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput. Biol. 19, 455–477.

Bestetti, G., Fatzer, R., Fankhauser, R., 1976. Ultrastructural investigations concerning a sporadically occurring meningo-encephalomyelitis of cattle in Switzerland. Schweiz. Arch. Tierheilkd. 118, 351–357 (In German.).

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., Pirovano, W., 2011. Scaffolding preassembled contigs using SSPACE. Bioinformatics 27, 578–579.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**, 2114–2120.

Bouzalas, I.G., Wüthrich, D., Walland, J., Drögemüller, C., Zurbriggen, A., Vandevelde, M., Oevermann, A., Bruggmann, R., Seuberlich, T., 2014. Neurotropic astrovirus in cattle with nonsuppurative encephalitis in europe. J Clin. Microbiol. 52, 3318–3324.

Bozzetta, E., Caramelli, M., Casalone, C., Acutis, P.L., Ru, G., 2003. BSE surveillance in Italy: neuropathological findings in cattle in the frame of the passive surveillance programme. J Vet. Med. A Physiol. Pathol. Clin. Med. 50, 48–49.

Brooks, B.R., Walker, D.L., 1984. Progressive multifocal leukoencephalopathy. Neurol. Clin. 2, 299–313.

Delbue, S., Elia, F., Signorini, L., Bella, R., Villani, S., Marchioni, E., Ferrante, P., Phan, T.G., Delwart, E., 2015. Human polyomavirus 6 DNA in the cerebrospinal fluid of an HIV-positive patient with leukoencephalopathy. J Clin. Virol. 68, 24–27.

Dobin, A., Davis, C. a, Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Fankhauser, R., 1961. Sporadic meningo-encephalomyelitis in cattle. Schweiz. Arch. Tierheilkd. 103, 225–235 (In German).

Fatzer, R., 1971. Brains of rabies negative animals. Schweiz Arch. Tierheilkd. 112, 59–67 (In German).

Fatzer, R., Steck, F., 1974. Histological differential diagnosis in cattle suspected of rabies. Schweiz. Arch. Tierheilkd. 116, 347–356 (In German).

Heim, D., Fatzer, R., Hornlimann, B., Vandevelde, M., 1997. Frequency of neurologic diseases in cattle. Schweiz. Arch. Tierheilkd. 139, 354–362 (In German).

Jeffrey, M., 1992. A neuropathological survey of brains submitted under the Bovine Spongiform Encephalopathy Orders in Scotland. Vet. Rec. 131, 332–337.

Koonin, E.V., Dolja, V.V., 2006. Evolution of complexity in the viral world: the dawn of a new vision. Virus Res. **117**, 1–4.

Kubiś, P., Materniak, M., Kuźmak, J., 2013. Comparison of nested PCR and qPCR for the detection and quantitation of BoHV6 DNA. J Virol. Methods 194, 94–101.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Li, L., Diab, S., McGraw, S., Barr, B., Traslavina, R., Higgins, R., Talbot, T., Blanchard, P., Rimoldi, G., Fahsbender, E., Page, B., Phan, T.G., Wang, C., Deng, X., Pesavento, P., Delwart, E., 2013. Divergent astrovirus associated with neurologic disease in cattle. Emerg. Infect. Dis. 19, 1385–1392.

Liu, Y., Li, N., Zhang, S., Zhang, F., Lian, H., Hu, R., 2015. Parainfluenza virus 5 as possible causative agent of severe respiratory disease in calves, China [letter]. Emerg. Infect. Dis., 14–1111. http://dx.doi.org/10.3201/eid2112.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M. a, 2010. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.

Mittelholzer, C., Hedlund, K.O., Englund, L., Dietz, H.H., Svensson, L., 2003. Molecular characterization of a novel astrovirus associated with disease in mink. J Gen. Virol. 84, 3087–3094.

O'Toole, D., Li, H., Sourk, C., Montgomery, D.L., Crawford, T.B., 2002. Malignant catarrhal fever in a bison (Bison bison) feedlot, 1993–2000. J Vet. Diagn. Investig. 14, 183–193.

Peretti, A., FitzGerald, P.C., Bliskovsky, V., Buck, C.B., Pastrana, D.V., 2015. Hamburger polyomaviruses. J Gen. Virol. 96, 833–839.

Powers, J.G., VanMetre, D.C., Collins, J.K., Dinsmore, R.P., Carman, J., Patterson, G., Brahmbhatt, D., Callan, R.J., 2005. Evaluation of ovine herpesvirus type 2 infections, as detected by competitive inhibition ELISA and polymerase chain reaction assay, in dairy cattle without clinical signs of malignant catarrhal fever. J Am. Vet. Med. Assoc. 227, 606–611.

Rovnak, J., Quackenbush, S.L., Reyes, R.A., Baines, J.D., Parrish, C.R., Casey, J.W., 1998. Detection of a novel bovine lymphotropic herpesvirus. J Virol. 72, 4237–4242.

Sanchez, S., Clark, E.G., Wobeser, G.A., Janzen, E.D., Philibert, H., 2013. A retrospective study of non-suppurative encephalitis in beef cattle from western Canada. Can. Vet. J 54, 1127–1132.

Sausker, E.A., Dyer, N.W., 2002. Polymerase chain reaction and DNA sequencing for detection of ovine herpesvirus 2 in American Bison (Bison). J Vet. Diagn. Investig. 14, 40–46.

Seuberlich, T., Wüthrich, D., Selimovic-Hamza, S., Drögemüller, C., Oevermann, A., Bruggmann, R., Bouzalas, I., 2016. Identification of a second encephalitis-associated astrovirus in cattle. Emerging Microbes & Infections (2016) 5, e5, http://dx.doi.org/10.1038/emi.2016.5.

Stahel, A.B.J., Baggenstos, R., Engels, M., Friess, M., Ackermann, M., 2013. Two different macaviruses, ovine herpesvirus-2 and caprine herpesvirus-2, behave differently in water buffaloes than in cattle or in their respective reservoir species. Plos One 8, e83695.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.
Theil, D., Fatzer, R., Schiller, I., Caplazi, P., Zurbriggen, A., Vandevelde, M., 1998. Neuropathological and aetiological studies of sporadic non-suppurative meningoencephalomyelitis of cattle. Vet. Rec. 143, 244–249.
Vandevelde, M., Higgins, R., Oevermann, A., 2012. Veterinary Neuropathology. John Wiley & Sons, Chichester, United Kingdom, p. 57.
Wiyono, a, Baxter, S.I.F., Saepulloh, M., Damayanti, R., Daniels, P., Reid, H.W., 1994. PCR detection of ovine herpesvirus-2 DNA in Indonesian ruminants – normal sheep and clinical cases of malignant catarrhal fever. Vet. Microbiol. 42, 45–52.
Zhang, W., Li, L., Deng, X., Kapusinszky, B., Delwart, E., 2014. What is for dinner? Viral metagenomics of US store bought beef, pork, and chicken. Virology 468–470, 303–310.
ZHANG, F., TORIYAMA, S., TAKAHASHI, M., 2001. Complete nucleotide sequence of ryegrass mottle virus : a new species of the genus sobemovirus. J. Gen. Plant Pathol. 67, 63–68.