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).