Virology 471-473 (2014) 117–125

Contents lists available at ScienceDirect

Virology

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

# Metagenomic analysis of viromes of dromedary camel fecal samples reveals large number and high diversity of circoviruses and picobirnaviruses

Patrick C.Y. Woo [a,b,c,d,*,1], Susanna K.P. Lau [a,b,c,d,1], Jade L.L. Teng [b,1], Alan K.L. Tsang [b,1], Marina Joseph [e], Emily Y.M. Wong [b], Ying Tang [b], Saritha Sivakumar [e], Ru Bai [b], Renate Wernery [e], Ulrich Wernery [e,**], Kwok-Yung Yuen [a,b,c,d]

[a] State Key Laboratory of Emerging Infectious Diseases, The University of Hong Kong, Hong Kong, China
[b] Department of Microbiology, The University of Hong Kong, Hong Kong, China
[c] Research Centre of Infection and Immunology, The University of Hong Kong, Hong Kong, China
[d] Carol Yu Centre for Infection, The University of Hong Kong, Hong Kong, China
[e] Central Veterinary Research Laboratory, Dubai, United Arab Emirates

ARTICLE INFO

*Article history:*
Received 23 April 2014
Returned to author for revisions
15 August 2014
Accepted 23 September 2014

*Keywords:*
Metagenomic analysis
Dromedary camel
Novel virus

ABSTRACT

The recent discovery of Middle East Respiratory Coronavirus and another novel dromedary camel coronavirus UAE-HKU23 in dromedaries has boosted interest in search of novel viruses in dromedaries. In this study, fecal samples of 203 dromedaries in Dubai were pooled and deep sequenced. Among the 7330 assembled viral contigs, 1970 were assigned to mammalian viruses. The largest groups of these contigs matched to *Picobirnaviridae*, *Circoviridae*, *Picornaviridae*, *Parvoviridae*, *Astroviridae* and *Hepeviridae*. Many of these viral families were previously unknown to dromedaries. In addition to the high abundance of contigs from *Circoviridae* ($n=598$ with 14 complete genomes) and *Picobirnaviridae* ($n=1236$), a high diversity of contigs from these two families was found, with the 14 *Circoviridae* complete genomes forming at least five clusters and contigs from both genogroup I and genogroup II potentially novel picobirnaviruses. Further studies comparing the incidence of these viral families in healthy and sick dromedaries will reveal their pathogenic potential.



## Introduction

Camels are one of the most unique mammals on earth. In particular, they have shown perfect adaptation to desert life where the daytime temperature is very high, diurnal temperature range is large and the supply of food and water is scarce. Such adaptations are made through their distinct anatomical and physiological properties, such as short but thick fur, long legs, water conservation and unique fat metabolism. Therefore, camels were used for transportation of people and goods as well as for military uses in the past. In addition, they also provide a good source of meat, milk and wool. They are also important recreational animals in the Middle East and are used for camel racing. Having been associated with humans for at least 5000 years, camels usually pose little physical danger to humans. Occasionally, infectious diseases, such as brucellosis, can be transmitted from camels to humans. Dromedary camels are one of the two surviving old world camel species, namely *Camelus dromedarius* (dromedary or one-humped camel), which inhabits the Middle East and North and Northeast Africa; and *Camelus bactrianus* (Bactrian or two-humped camel), which inhabits Central Asia. Among the 20 million camels on earth, 90% are dromedaries.

The recent emergence of Middle East Respiratory Coronavirus (MERS-CoV) from the Middle East and the presence of neutralizing antibodies against MERS-CoV from dromedaries in the Middle East have boosted interest in the search of novel viruses in dromedaries (de Groot et al., 2013; Lau et al., 2013; Perera et al., 2013; Reusken et al., 2013; Zaki et al., 2012). Viruses of at least eight families, including *Paramyxoviridae*, *Flaviviridae*, *Herpesviridae*, *Papillomaviridae*, *Picornaviridae*, *Poxviridae*, *Reoviridae* and *Rhabdoviridae*, have been found to infect camels (Al-Ruwaili et al., 2012; Intisar et al., 2009; Khalafalla

* Corresponding author at: State Key Laboratory of Emerging Infectious Diseases, Department of Microbiology, The University of Hong Kong, University Pathology Building, Queen Mary Hospital, Hong Kong, China. Tel.: +852 22554892; fax: 852 28551241.
** Corresponding author.
*E-mail addresses:* pcywoo@hkucc.hku.hk (P.C.Y. Woo), cvrl@cvrl.ae (U. Wernery).
[1] These authors contributed the same to the manuscript.

http://dx.doi.org/10.1016/j.virol.2014.09.020
0042-6822/

et al., 2010; Ure et al., 2011; Wernery et al., 2014; Wernery et al., 2008; Wernery and Zachariah, 1999; Yousif et al., 2004). Recently, we have discovered a novel coronavirus, named dromedary camel coronavirus UAE-HKU23 (DcCoV UAE-HKU23), in dromedaries (Woo et al., 2014b). As camels are closely associated with humans, knowledge on the variety of viruses present in this hardy group of animals is important to understand their potential for emergence. In this study, we analyzed the viromes of fecal samples of dromedaries in the Middle East, which is the first metagenomic study on animals of the family *Camelidae*. The interestingly large number and high diversity of contigs from the *Circoviridae* and *Picobirnaviridae* families were also discussed.

## Results

### *Metagenomic analysis of dromedary fecal samples*

Fecal samples of 203 dromedaries were pooled and deep sequenced using the Illumina HiSeq 2500 instrument, generating 29,247,514 paired-end 151-bp sequence reads. *De novo* assembly of the metagenome was performed using IDBA-UD to confirm isolation of viral genomes using default parameter with minimum read length of 200. There were 159,388 contigs ranging in size from 200 to 14,611 bp with a mean contig length of 540 bp.

Among 159,388 contigs, 7330 were viral sequences. The most abundant fraction of viral contigs matched to the bacteriophages, including those of the order *Caudovirales* ($n=3805$), family *Microviridae* ($n=509$) and unclassified phages ($n=319$) (Fig. 1). Viral contigs related to plant viruses included those of the families *Geminiviridae* ($n=17$), *Betaflexiviridae* ($n=15$), *Totiviridae* ($n=2$), *Nanoviridae* ($n=1$) and *Partitiviridae* ($n=4$); and those related to insect viruses included those of the families *Iflaviridae* ($n=5$), *Dicistroviridae* ($n=3$), *Poxviridae* ($n=3$), and *Nodaviridae* ($n=2$) and subfamily *Densovirinae* ($n=2$) (Fig. 1).

One thousand nine hundred and seventy (26.9%) of the 7330 viral contigs were assigned to mammalian viruses (Fig. 1). The largest group of the contigs matched to double-stranded RNA viruses in the family *Picobirnaviridae* ($n=1236$), followed by single-stranded DNA viruses in the family *Circoviridae* ($n=598$). The remaining contigs with homology to the most represented families of mammalian viruses were, in order of decreasing abundance, *Picornaviridae* [kobuviruses ($n=17$), enteroviruses ($n=26$), hunnivirus ($n=14$), encephalomyocarditis virus ($n=4$)]; *Parvoviridae* [porcine bocavirus ($n=22$), human bocavirus ($n=5$), feline bocavirus ($n=1$), gorilla bocavirus ($n=1$)]; *Astroviridae* [porcine astrovirus ($n=8$), feline astrovirus ($n=7$)]; *Hepeviridae* [HEV ($n=3$)]; *Reoviridae* [rotavirus ($n=3$)] and *Caliciviridae* [feline norovirus ($n=2$)]. These viral contigs showed a wide range of sequence identity to known viruses, suggesting some of these sequences might be derived from novel viruses.

### *Alignment of sequences of Circoviridae and Picobirnaviridae*

The 598 and 1236 contigs that belonged to the families *Circoviridae* and *Picobirnaviridae* respectively were analyzed by BLASTx. In both families *Circoviridae* and *Picobirnaviridae*, contigs that encoded the corresponding RdRp, capsid proteins and hypothetical proteins were observed (Supplementary Fig. 1).

### *Phylogenetic analysis of complete genomes in Circoviridae*

Repeated terminal sequences in the contigs indicated a circular genome. Fourteen contigs containing complete circular genomes of the *Circoviridae* family, ranging in size from 2516 to 2977 bp (Fig. 2). Overall, nucleotide identities to known members of the *Circoviridae* family were less than 75% for all the genomes. Therefore, according to the ICTV criteria (www.ictvdb.org) which state that circoviruses of the same species should share $>75\%$ and $>70\%$ nucleotide identity in their complete genome and capsid protein sequences respectively, these viruses found in dromedaries should be novel species in the *Circoviridae* family.

Phylogenetic tree of these 14 complete genomes were constructed with representative complete genomes of circovirus, cyclovirus and circo-like virus sequences, starting at Rep ATG. These 14 complete genomes formed at least five clusters, including one related to porcine circovirus-like virus, five related to bovine stool-associated circular DNA virus (BoSCV), two related to fur seal feces-associated circular DNA virus (FSfaCV), two related to rodent stool associated circular genome virus (RodSCV) and four related to other cirovirus-like virus (Fig. 3).

### *Phylogenetic analysis of sequences in Picobirnaviridae*

The 12 and 21 contigs that belonged to the family *Picobirnaviridae* with complete RdRp and capsid genes respectively were further aligned with all available complete RdRp and capsid genes of picobirnaviruses. Phylogenetic tree of the complete RdRp and capsid genes of the picobirnavirus genome is shown in Fig. 4. The contigs were highly diverse. Contigs that belonged to both genogroup I and genogroup II picobirnaviruses were observed. In addition, distinct branches that were not clustered with either genogroup I or genogroup II picobirnaviruses were also observed, suggesting that there may be one or more additional genogroups in picobirnaviruses.

## Discussion

In this first metagenomic study on viromes in animals of the family *Camelidae*, more than 500 contigs (including 14 complete genomes) and around 25% of mammalian virus contigs observed in dromedary fecal samples belonged to the *Circoviridae* family. Members of the *Circoviridae* family are small non-enveloped circular single-stranded DNA viruses found in a wide variety of mammals and birds. Circovirus infections are very common and geographically widely distributed. Although subclinical infections are common, circovirus infections have been suggested to be associated with psittacine beak and feather disease, infectious chicken anemia, circovirus disease of pigeons, and the postweaning multisystemic wasting syndrome of pigs (Biagini et al., 2011). Among the metagenomic studies on fecal samples of other mammals, only one study on fecal samples of pigs showed a comparable high number of sequences from the *Circoviridae* family (Table 1) (Sachsenroder et al., 2014). At least five metagenomic studies did not show any circovirus sequence (Lager et al., 2012; Li et al., 2011a; Li et al., 2011b; Smits et al., 2013; van den Brand et al., 2012). As for sequence diversity, the 14 complete genomes formed at least five clusters related to different known members of the *Circoviridae* family, including porcine circovirus-like virus, BoSCV, FSfaCV, RodSCV and other cirovirus-like virus, were observed in this study (Fig. 3). This high diversity of sequences from the *Circoviridae* family was rarely seen in other metagenomic studies.

The dromedary fecal samples also contain a large number (more than 1000 contigs and more than half of all mammalian virus contigs) and high diversity of picobirnavirus sequences. Picobirnaviruses are small non-enveloped bisegmented double-stranded RNA viruses found in human and a wide variety of mammals and birds. Since its first discovery in fecal samples of humans and rats in 1988 (Pereira et al., 1988a; Pereira et al., 1988b), picobirnaviruses have been reported in other mammals and birds (Browning et al., 1991; Gallimore et al., 1993; Ludert et al., 1991; Masachessi et al., 2007; Nates et al., 2011) and environmental water samples (Hamza et al., 2011). The