Application Reference nr: 5806

Title: Metagenomic analysis of European wild carnivore species for novel virus discovery

Applicant: Dr. Aritz Ruiz-Gonzalez. Conservation Genetics Laboratory, National Institute for Environmental Protection and Research, ISPRA, ***Italy***.

Host Institution: Prof. Dr. Albert D.M.E. Osterhaus / Dr. Saskia L. Smits / Dr. Rogier Bodewes. Department of Viroscience, Erasmus Medical Center, Rotterdam, ***The Netherlands***.

1. PURPOSE OF THE VISIT

Emerging and re-emerging infectious diseases pose a continuous threat towards health and disease burden on humans, domestic animals and wildlife (Kuiken et al. 2005). Animals, and particularly wild animals, are thought to be the source of the majority of emerging infections, highlighting the importance of wildlife species as a reservoir of viruses that could pose a potential threat to humans and other species (Kuiken et al. 2005 Woolhouse et al. 2005; Morse et al. 2012). A thorough understanding of virus diversity in wildlife may be used for forecasting future transmission risks or eventual outbreaks of viral disease among humans and domestic animals. Additionally, the characterization of the normal viral population in healthy wildlife can provide a baseline level of viruses that are present in healthy hosts, which might help to understand the role of certain pathogens in case of an outbreak of infectious disease in wildlife before the wide spread of viruses to other species (Delwart 2012). Viral metagenomics is a powerful, fast and sensitive technique that has revolutionized the identification of (new) viruses in a wide variety of samples (Rosario & Breitbart, 2011; Mokili et al. 2012). The host institution for the research visit - Department of Viroscience. Erasmus Medical Center- is a world-renowned centre in viroscience that, in collaboration with Viroclinics Biosciences BV, several other partners of the EU-funded EMPERIE project (www.emperie.eu) and the VIRGO/FES project (www.genomics.nl/Research/GenomicsCentres/VIRGO.aspx), focuses on a systematic exploration of the presence of new viruses in humans and animals through novel metagenomic analysis. The current knowledge of viruses that infect humans and animals is incomplete and many acute and chronic diseases of unknown etiology are expected to be caused by as yet unidentified viruses (Rosario & Breitbart, 2011). In a number of recent studies, novel viral metagenomics analyses have been performed in healthy wildlife using random PCR and next-generation sequencing (NGS) (Ge et al. 2012; Li et al. 2011; van den Brand et al. 2012; Phan et al. 2012). Wild carnivore species are of special interest, since metagenomic analysis of fecal samples can provide information about the presence of known and unknown viruses specific for the host and those derived from their prey (Li et al. 2011; van den Brand et al. 2012; Bodewes et al. 2013). However, and in spite of their relevance as potential carriers of pathogens, only a few recent studies have conducted thorough metagenomic analysis of fecal material on wild carnivore species (Li et al. 2011; van den Brand et al. 2012; Bodewes et al. 2013). Knowledge of the diversity of viruses present in wildlife, and specifically in wild carnivore will help to get a comprehensive understanding of potential cross-species transmission of viruses between diverse hosts (Allison et al. 2013) and provide epidemiological baseline information about potential pathogens (Delwart 2012). A close collaboration between virologists specialized in NGS virus discovery with conservation biologists dealing with wildlife genetics/genomics studies, will constitute a proficient framework in order to provide further insights into virus diversity in conservation concern, invasive alien and widely distributed common carnivore species at European scale.

The **main objective** of the short-term research visit is to:

1) Apply a metagenomic approach to evaluate fecal viral flora of European mesocarnivore species using random PCR in combination with NGS platforms (454 GS Junior). (The virus discovery study will be focused on a small subset of carnivore fecal samples (n=1-10 of each target species): *Martes martes; Martes foina; Mustela lutreola; Mustela putorius; Genetta genetta; Neovison vison; Meles meles; Lutra lutra and Vulpes vulpes*.)

2) Promote the exchange of practical knowledge in the field of metagenomic analysis in order to initiate collaborative projects and establish efficient multidisciplinary networks between European laboratories focus in ***conservation genetics*** and specialized laboratories in ***viral metagenomic*** analysis of wildlife species.

2. DESCRIPTION OF THE WORK CARRIED OUT DURING THE VISIT

Sample collection

Rectal swabs or fecal specimens were collected for the purpose of this study from 43 carnivores in the Basque Country and La Rioja region (Spain) during 2012-2013. Samples of American mink (*Neovison vison*; n=10), Common genet (*Genetta genetta*; n=7), Eurasian badger (*Meles meles*; n=4), Eurasian otter (*Lutra lutra*; n=1), European mink (*Mustela lutreola*; n=9), European pine marten (*Martes martes*; n=2); European polecat (*Mustela putorius*; n=3), Red fox (*Vulpes vulpes*; n=4), Stone marten (*Martes foina*; n=1), and Wild cat (*Felis silvestris*; n=1). Samples were collected from animals that were either found dead, live trapped, or euthanized in the frame of an eradication program of invasive species (American mink), in Spain. Following collection, samples were directly stored at -20ºC and were stored at -70ºC within 2 months after collection until further processing. During the short visit a first sub-set of 12 samples was processed and analysed.

Sequence independent RNA and DNA virus screening of collected samples

Samples were processed for viral metagenomics as described previously (van Leeuwen et al. 2010; van den Brand et al. 2012; Bodewes et al. 2013). In brief, samples were depleted from host nucleic acids and filtered through a 0.45uM filter. Subsequently, RNA and DNA were extracted using the Nucleospin RNA XS kit (Macherey-Nagel) and the High Pure viral nucleic acids kit (Roche). First and second strand synthesis and random PCR amplification were performed. PCR products were purified and processed for next-generation sequencing with a 454 GS Junior Instrument (Roche). Obtained reads were assembled using de novo assembly in CLC Genomics Workbench 5 (CLC Bio) and contigs and individual reads were analyzed by BLASTn and BLASTx respectively. Cut off E values for significant virus hits for BLASTn and BLASTx were respectively 1.0 x 10-3 and 1.0 x 10-10. Based on the taxonomic origin of the best-hit sequence, classification of the sequences was performed in MEGAN 4.70.4 (Huson et al. 2011).

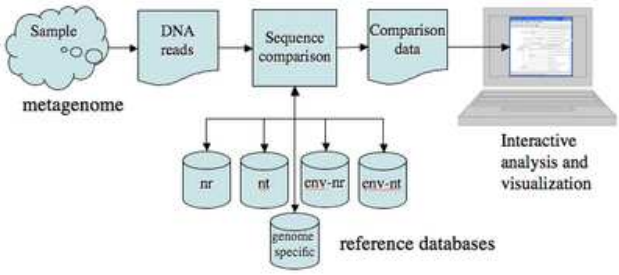

Figure 1. Metagenomic analysis workflow.

3. DESCRIPTION OF THE MAIN RESULTS OBTAINED

Metagenomic overview

Using random amplification in combination with 454 next-generation sequencing, more than 84,000 reads were obtained for the 11 out of the 12 processed samples. Sample 3 failed during PCR amplification, and was discarded for further NGS analysis. After a demultiplexing process 82,660 reads were assigned to one of the 11 MIDS used, with a mean value of 7514 reads per sample (± 3,531) (Table 1).

Barcode	Number of reads	Percentage of reads
MID 10	11934	14
MID 6	9303	11
MID 8	8245	10
MID 5	4589	5
MID 9	7001	8
MID 2	11178	13
MID 12	12135	14
MID 11	1040	1
MID 4	6570	8
MID 7	6997	8
MID 1	3668	4
Subtotal	82660	97
Not grouped	2134	3
Total	84891	100

Table 1. Reads per barcode for the 11 analysed samples before and after trimming procedure.

After removal of barcodes, primers and reads <50bp, 66104 trimmed sequence reads were obtained from fecal carnivore samples (Table 2; Figure 1)

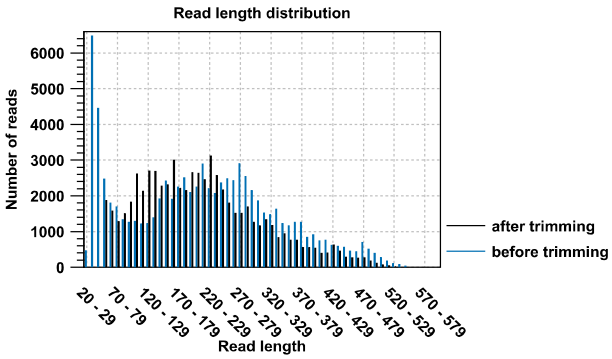

Figure 1. Read length distribution and number of reads before and after trimming procedure

Name	Number of reads	Avg.length	Number of reads after trim	Percentage trimmed	Avg.length after trim
MID 10	11934	208,7	9741	81,62	203
MID 6	9303	234,8	7327	78,76	243
MID 8	8245	219,3	6758	81,96	213,3
MID 5	4589	203,6	3692	80,45	197,7
MID 9	7001	180,2	5500	78,56	171
MID 2	11178	224,6	8828	78,98	229,1
MID 12	12135	171,2	9624	79,31	160,9
MID 11	1040	145,4	710	68,27	156,9
MID 4	6570	245,9	5163	78,58	257,7
MID 7	6997	252,6	5842	83,49	249,4
MID 1	3668	207,5	2919	79,58	204
Total	**82660,00**	**208,53**	**66104,00**	**79,05**	**207,82**

Table 2. Read length distribution and number of reads before and after trimming procedure

Reads were *De novo* assembled into 2190 contigs with a mean contig length of 410bp. Summary for *De novo* assembly results is depicted in Table 3 and Figure 2. An example of *De novo* assembly in CLC workbench is shown in Figure 3.

Contig set	Contig count	Mean contig length	Standard deviation	Total contig length	% GC
N25 contigs	189	1189,61	844,79	224836	45,92
N50 contigs	635	707,61	561,15	449332	45,72
N75 contigs	1281	526	434,71	673810	45,68
All contigs	**2190**	**410,23**	**360,4**	**898396**	**45,67**
Long contigs (>1,000bp)	79	1741,48	1083,59	137577	46,05
Short contigs (<200bp)	40	170,48	35,71	6819	48,6

Table 3. *De novo* assembly results summary.

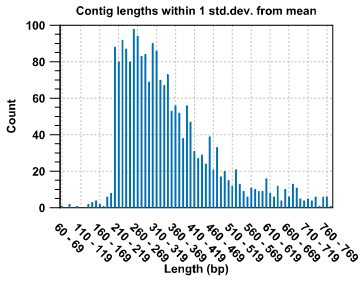

Figure 2. Contig length distribution.

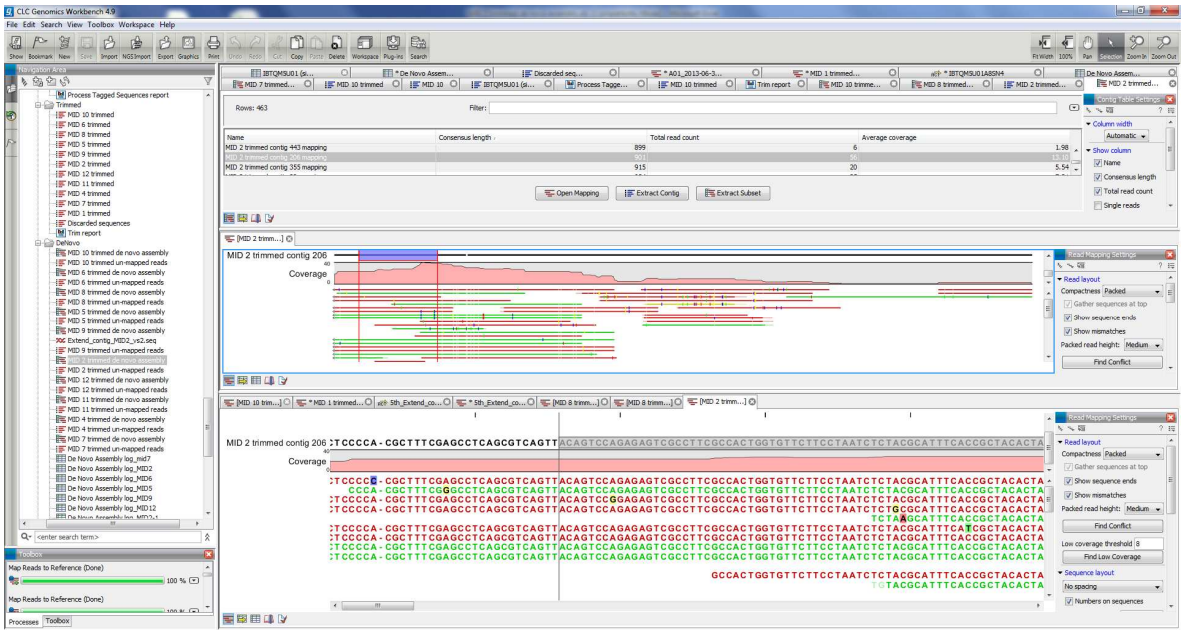

Figure 3. De novo assembly in CLC workbench of Mid-2 sample. Contig 206 is shown.

Reads and contigs were classified into viruses, bacteria and eukaryotes by processing Blast results into Megan software. As Blast and Megan analyses are currently underway for the whole dataset, only preliminary result obtained for 2 out of 11 analysed samples are shown (Mid-6: Common genet-*Gennetta genetta*- and Mid-9: Eurasian otter-*Lutra lutra*-) (Figure 4 and Figure 5). Many of the identified sequences were of bacterial or eukaryotic origin. In addition, several reads were detected that had the closest similarity to viruses. All the obtained reads will be deposited at the European Nucleotide Archive under archive number PRJEB4910. A substantial proportion of the reads did not have any significant hits for nucleotide or amino acid sequences in GenBank (Figure 4).

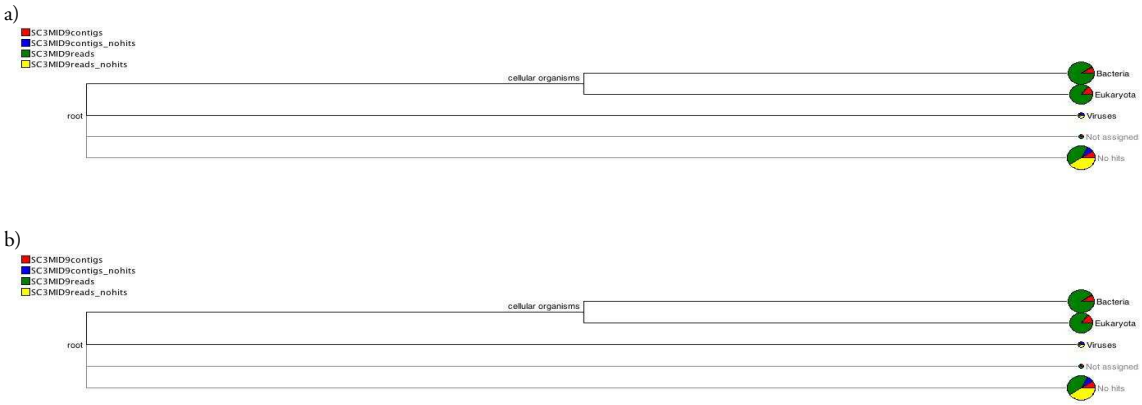

Figure 4: Megan taxonomic classification of a) Mid-6 and b) Mi-9 contigs and reads obtained from Blast-n and Blas-x (no hits) into Bacteria, Eukaryota, Viruses, Not assigned and No hits.

In the common genet sample, a greater proportion of all taxonomic groups, including viruses were detected (Figure 5).

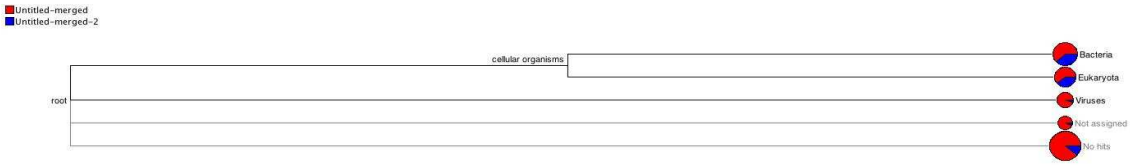

Figure 5. Comparison of Megan results for Mid-6 (red) and Mid-9 (blue).

In all 2 the samples, sequences of the order Caudovirales (Bacteriophages) were detected and, as previously reported in other recent studies, only a few proportion of the obtained sequences had the closest similarity to viruses known to infect vertebrates (van Leeuwen et al. 2010; van den Brand et al. 2012; Bodewes et al. 2013). An example of Megan taxonomic classification results for Mid-6 is shown in Figure 6.

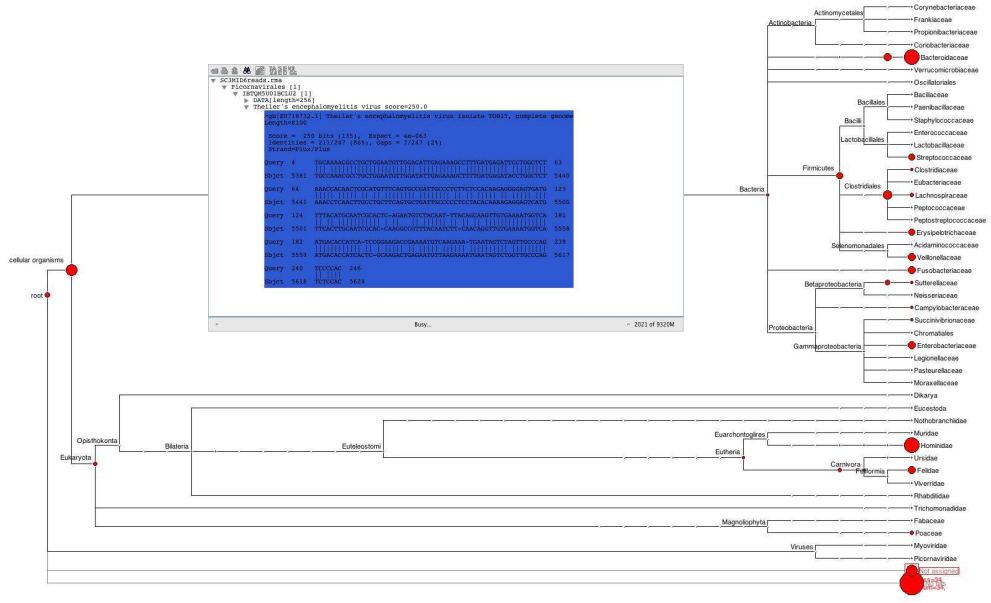

Figure 6. Megan taxonomic classification of a) Mid-6 reads obtained from Blast-n and b) assigment details of a 256 bp read asssigned to Theiler's Murine Encephalomyelitis Virus.

The main viral metagenomic results obtained for Mid-6 and Mid-9 are summarized below.

- Common genet (*Genetta genetta)* virus results (Mid-6)

Approximately, 70% of the identified virus corresponded to viruses that infects and replicates within bacteria. We identified phages from the order caudovirales (Myoviridae, Shiphoviridae and Podoviridae) that account for 95% of all the phages reported in the scientific literature, and possibly make up the majority of phages on the planet (McGrath & van Sinderen, 2007). Some other phages from the Microviridae family were also identified. Interestingly, viruses that had the closest similarity to viruses known to infect vertebrates belonging to the family of Circoviridiae, Picobirnaviridae and Picornaviridae, were detected. Members of the genus *Circovirus* in the family *Circoviridae*, are non-enveloped, icosahedral viruses with a single-stranded circular DNA (ssDNA) genome of approximately 2 kb, the smallest known autonomously replicating viral genomes (ICTV 2012). The genus Circovirus includes several recognized species that infect mammals and birds, namely, porcine circovirus 1, porcine circovirus 2, Canine circovirus 1, starling circovirus, canary circovirus, goose circovirus, pigeon circovirus, and beak and feather disease virus (Li et al. 2011). In the common genet we identified several sequences with high homology to know *Circovirus*, previously identified in birds (canary and starling circoviruses). Several new circoviruses have been recently identified in animal feces and environmental samples; however the natural hosts of most of these viruses remain unidentified (Li et al. 2011).

Picobirnaviruses are small, non-enveloped, bisegmented double-stranded RNA viruses. These viruses have been often detected in fecal samples of humans and various animal species with and without disease (van Leeuwen et al. 2010; Phan et al. 2011; Smits et al. 2012; Bodewes et al. 2013). In the present study, sequences that had the highest similarity to sequences belonging to the family of *Picobirnaviridae* were detected in the common genet samples. However, due the low homology detected from known Picobirnaviruses, this potentially novel virus will be further characterized.

Picornaviruses are small, positive stranded non-enveloped RNA viruses. This large family was subdivided into several genera (ICTV 2012). The genus *Cardiovirus* of the family of *Picornaviridae* currently consists of two species, Theilovirus and encephalomyocarditis virus (EMCV) (ICTV 2012). In the common genet rectal swab we detected sequences that had the highest similarity to viruses of the species Theilovirus. Viruses belonging to the species Theilovirus were detected mainly in rodents. In these animals, viruses cause primarily infection of the digestive tract without clinical signs, but extra-intestinal infection occurs and can cause an acute encephalomyelitis and a chronic demyelinating infection of the central nervous system (Brahic et al. 2005). Vilyuisk human encephalomyelitis virus, another strain of the species Theilovirus, was isolated from humans with encephalomyelitis after serial passage over mice brains (Casals 1963; Pritchard et al. 1992). More recently, Saffold virus was discovered in a stool sample of a child with fever of unknown origin, and additional research revealed that infection with this virus was common (Jones et al. 2007; Zoll et al. 2009; Himeda & Ohara 2012). Further sequencing and phylogenetic analysis are required to further characterize this virus.

- Eurasian otter-*Lutra lutra*- virus metagenomic results (Mid-9)

In fecal samples of the Eurasian otter, viruses belonging to the family of *Bunyaviridae*, were detected. Phlebovirus is one of five genera of the family Bunyaviridae. The genus Phlebovirus consists of a genetically diverse group of viruses, some of which were described very recently (Palacios et al. 2013a,b; Swei et al. 2013). A few members of this genus were identified as important pathogens in humans and domestic animals, including Rift valley fever virus and the Severe fever with thrombocytopenia syndrome virus or Huaiyangshan virus. In fecal material of the otter sequences were detected that had the highest similarity to viruses of the genus *Phlebovirus.* Further analyses are required in other to confidently determine the phylogenetic position of this potentially new Phlevbovirus. Additionally, sequences were detected that had the closest similarity to the recently proposed family of *Breviviridae* and the recently described hybrid DNA virus NIH-CQV (van den Brand et al. 2012; Xu et al. 2013).

A few sequences/contigs with homology to known viruses that infects plants, fungus, diatomea and alga were also identified. For example, we have identified the Watercress white vein virus, a ssRNA virus from the Thymoviridae family, that have been recently discovered in the aquatic Watercress white vein plants from Spain (Harju et al. 2012).

Main conclusions

Here, we described preliminary results from a viral metagenomic study with fecal samples collected from wild carnivores in Spain. Various sequences of known and potentially novel viruses were identified. However, the potentially novel viruses identified should be further characterized before drawn any kind of conclusion. Up to now no known zoonotic viruses were detected among the analysed samples. However, we detected some relevant viruses that might cause outbreaks of disease among animals or humans in the future (e.g. phleboviruses). Since for almost all animals only fecal material was available, and detection of viral DNA in feces may reflect either passive passage through the gut without replication or actual enteric replication in that host, it is unknown whether these viruses have caused disease in the host or if they might potentially transmit to domestic animals or humans. A proportion of the detected viral sequences had the closest similarity to viruses previously detected in birds and rodents, suggesting that these viruses most likely originate from the diet of the animals. Additional studies based on collection of blood samples and complete necropsies are necessary to elucidate the pathogenicity and epidemiology of the identified viruses. This study is an important step forward in order to improve the limited knowledge about virus diversity present in wild carnivores in Europe and will help to get a better understanding of potential cross-species transmission of viruses between diverse hosts (Allison et al. 2013) and provide epidemiological baseline information about potential pathogens (Delwart 2012).

4. FUTURE COLLABORATION WITH HOST INSTITUTION (IF APPLICABLE)

The ESF short visit grant has allowed the reinforcement of the collaboration framework with the Host institution, and we are currently exploring the possibility of conducting new metagenomic surveys on other carnivores species and/or increasing the number of individuals analysed in order to get a more comprehensive knowledge about the virus diversity in European carnivore species.

5. PROJECTED PUBLICATIONS/ARTICLES RESULTING OR TO RESULT FROM THE GRANT (ESF must be acknowledged in publications resulting from the grantee's work in relation with the grant)

Rogier Bodewes#, Aritz Ruiz-Gonzalez#, Anita C. Schürch, Claudia M.E. Schapendonk, Judith M.A. van den Brand, Bart L. Haagmans, Albert D.M.E. Osterhaus, and Saskia L. Smits. Viral metagenomic analysis of feces of wild small carnivores in Spain: identification of various novel viruses (In prep.) Target Journal: Plos One
#: both authors contributed equally

6. OTHER COMMENTS (IF ANY)

-

REFERENCES

Allison AB, Kohler D.J, Fox KA, Brown JD, Gerhold RW, Shearn-Bochsler VI, Holmes EC. 2013. Frequent cross-species transmission of parvoviruses among diverse carnivore hosts. *J Virol* 87:2342-2347.

Bodewes R, van der Giessen J, Haagmans BL, Osterhaus AD, Smits SL (2013) Identification of multiple novel viruses, including a parvovirus and a hepevirus, in feces of red foxes. J Virol, 87, 7758-7764.

Brahic M, Bureau JF, Michiels T (2005) The genetics of the persistent infection and demyelinating disease caused by Theiler's virus. Annu Rev Microbiol, 59, 279-298.

Casals J (1963) Immunological Characterization of Vilyuisk Human Encephalomyelitis Virus. Nature, 200, 339-341.

Delwart, E. 2012. Animal virus discovery: improving animal health, understanding zoonoses, and opportunities for vaccine development. *Curr Opin Virol* 2, 344-352.

Duarte MD, Henriques AM, Barros SC, Fagulha T, Mendonça P, Carvalho P, Cunha, MV. 2013. Snapshot of viral infections in wild Carnivores reveals ubiquity of parvovirus and susceptibility of Egyptian mongoose to feline panleukopenia virus. PloS one, 8(3), e59399.

Ge X, Li Y, Yang X, Zhang H, Zhou P, Zhang Y, Shi Z. 2012. Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J Virol* 86:4620-4630.

Himeda T, Ohara Y (2012) Saffold virus, a novel human Cardiovirus with unknown pathogenicity. J Virol, 86, 1292-1296.

Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. Genome Research 21: 1552-1560.

ICTV (2012) Virus Taxonomy: 2012 Release. http://wwwictvonlineorg/virusTaxonomyasp.

Jones MS, Lukashov VV, Ganac RD, Schnurr DP (2007) Discovery of a novel human picornavirus in a stool sample from a pediatric patient presenting with fever of unknown origin. J Clin Microbiol, 45, 2144-2150.

Kuiken T, Leighton FA, Fouchier RA, LeDuc JW, Peiris JS, Schudel A, Stohr K, Osterhaus AD. 2005. Public health. Pathogen surveillance in animals. *Science* 309:1680- 1681.

Li L, Shan T, Wang C, Cote C, Kolman J, Onions D, Gulland FM, Delwart E. 2011. The fecal viral flora of California sea lions. *J Virol* 85:9909-9917.

Li L 2010. Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. J. Virol. 84:1674–1682.

Mc Grath S and van Sinderen D (editors). (2007). Bacteriophage: Genetics and Molecular Biology (1st ed.). Caister Academic Press.

Mokili JL, Rohwer F, Dutilh BE. 2012. Metagenomics and future perspectives in virus discovery. Curr. Opin. Virol. 2:63–77

Morse SS, Mazet JA, Woolhouse M, Parrish CR, Carroll D, Karesh WB, Zambrana- Torrelio C, Lipkin WI, Daszak P. 2012. Prediction and prevention of the next pandemic zoonosis. *Lancet* 380:1956-1965.

Palacios G, Tesh R, Savji N, Travassos da Rosa AP, Guzman H, et al. (2013) Characterization of the Sandfly fever Naples species complex and description of a new Karimabad species complex (genus Phlebovirus, family Bunyaviridae). J Gen Virol.

Palacios G, Savji N, Travassos da Rosa A, Guzman H, Yu X, et al. (2013) Characterization of the Uukuniemi virus group (Phlebovirus: Bunyaviridae): evidence for seven distinct species. J Virol 87: 3187-3195.

Phan TG, Kapusinszky B, Wang C, Rose RK, Lipton HL, Delwart EL. 2011. The fecal viral flora of wild rodents. *PLoS Pathog* 7:e1002218.

Pritchard AE, Strom T, Lipton HL (1992) Nucleotide sequence identifies Vilyuisk virus as a divergent Theiler's virus. Virology, 191, 469-472.

Rosario, K. and Breitbart M. 2011.Exploring the viral world through metagenomics. *Curr Opin Virol* 1: 289–297

Smits SL, van Leeuwen M, Schapendonk CM, Schurch AC, Bodewes R, Haagmans BL, Osterhaus AD (2012) Picobirnaviruses in the human respiratory tract. Emerg Infect Dis, 18, 1539-1540.

Swei A, Russell BJ, Naccache SN, Kabre B, Veeraraghavan N, et al. (2013) The genome sequence of Lone Star virus, a highly divergent bunyavirus found in the Amblyomma americanum tick. PLoS ONE 8: e62083.

van den Brand JM, van Leeuwen M, Schapendonk CM, Simon JH, Haagmans BL, Osterhaus AD, Smits SL. 2012. Metagenomic analysis of the viral flora of pine marten and European badger feces. J Virol 86:2360-2365.

Woolhouse ME, Haydon DT, Antia R. 2005. Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol. Evol.* 20:238 –244.

Xu B, Zhi N, Hu G, Wan Z, Zheng X, Liu X, Wong S, Kajigaya S, Zhao K, Mao Q, Young NS (2013) Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing. Proc Natl Acad Sci U S A, 110, 10264-10269.

Zoll J, Erkens Hulshof S, Lanke K, Verduyn Lunel F, Melchers WJ, Schoondermark-van de Ven E, Roivainen M, Galama JM, van Kuppeveld FJ (2009) Saffold virus, a human Theiler's-like cardiovirus, is ubiquitous and causes infection early in life. PLoS Pathog, 5, e1000416.