bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Characterisation of the faecal virome of captive and wild Tasmanian devils using virus-like particles metagenomics and meta-transcriptomics

Rowena Chong[1], Mang Shi[2,3], Catherine E Grueber[1,4], Edward C Holmes[2,3], Carolyn Hogg[1], Katherine Belov[1] and Vanessa R Barrs[2,5*]

[1]School of Life and Environmental Sciences, University of Sydney, NSW 2006, Australia.

[2]Marie Bashir Institute for Infectious Diseases and Biosecurity, Sydney Medical School, University of Sydney, NSW 2006, Australia.

[3]School of Life and Environmental Sciences and Sydney Medical School, Charles Perkins Centre, University of Sydney, NSW 2006, Australia.

[4]San Diego Zoo Global, PO Box 120551, San Diego, CA 92112, USA.

[5]Sydney School of Veterinary Science, University of Sydney, NSW 2006, Australia.

*Correspondence: vanessa.barrs@sydney.edu.au

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Background: The Tasmanian devil is an endangered carnivorous marsupial threatened by devil facial tumour disease (DFTD). While research on DFTD has been extensive, little is known about the viruses present in devils, and whether any of these are of potential conservation relevance for this endangered species.

Methods: Using both metagenomics based on virus-like particle (VLP) enrichment and sequence-independent amplification (VLP metagenomics), and meta-transcriptomics based on bulk RNA sequencing, we characterised and compared the faecal viromes of captive and wild Tasmanian devils.

Results: A total of 54 devil faecal samples collected from captive (n = 2) and wild (n = 4) populations were processed for virome characterisation using both approaches. We detected many novel, highly divergent viruses, including vertebrate viruses, bacteriophage and other dietary associated plant and insect viruses. In total, 18 new vertebrate viruses, including novel sapelovirus, astroviruses, bocaviruses, papillomaviruses and gammaherpesvirus were identified, as well as known mammalian pathogens including rabbit haemorrhagic disease virus 2 (RHDV2). Captive devils showed significantly lower levels of viral diversity than wild devils. Comparison of the two methodological approaches revealed substantial differences in the number and types of viruses detected, with meta-transcriptomics mainly identifying RNA viruses, and VLP metagenomics largely identifying DNA viruses.

Conclusion: This study has greatly expanded our knowledge of eukaryotic viruses in the Tasmanian devil and provides important baseline information that will contribute to the conservation and captive management of this endangered species. In addition, our results showed that a combination of VLP metagenomics and meta-transcriptomics may be a more comprehensive approach to virome characterisation than either method alone.

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

45 **Key words:** Tasmanian devil, Marsupial, Carnivore, Gut microbiome, Endangered species,

46 Virus

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Background

The Tasmanian devil (*Sarcophilus harrisii*) is the world's largest extant carnivorous marsupial found only on the island state of Tasmania, Australia. Being predominantly a scavenger, the diet of devils largely comprises of carrion of mammals, such as wallabies, possums and kangaroos, although they are also known to consume and digest fish, insects, fruit and vegetation [1, 2]. Listed as endangered, the Tasmanian devil is facing the threat of extinction due to a contagious cancer, devil facial tumour disease (DFTD), that has caused drastic declines in wild devil populations by 77% since its discovery in 1996 [3]. There are currently two forms of DFTD affecting devils; DFT1 and DFT2 [4]. In an attempt to save the species from extinction, an insurance population was established in 2006 to serve as an important source of animals for supplementing wild populations at risk of population crashes due to DFTD [5, 6]. While extensive research has focused on DFTD itself as well as devil genetic diversity and susceptibility to the disease over the past decade [7-11], our understanding of other disease threats to devils remains limited. Specifically, virological studies are scarce and limited to the identification of a single gammaherpesvirus (DaHV-2), for which captivity was identified as a significant risk factor [12]. A comprehensive characterisation of the viral communities inhabiting the Tasmanian devil is an essential step to improving our understanding of host-microbe relationships, and maximising health and conservation management of the species. Comparative analysis of marsupial viruses with those from diverse vertebrate hosts, including eutherian mammals, birds and other vertebrates will also provide a deeper understanding of the phylogenetic history of the viruses infecting this evolutionary unique group of mammals [13, 14].

To date, the most widely used method for studying viral metagenomics relies on the enrichment of virus-like particles (VLP) and subsequent sequence-independent amplification prior to sequencing [15-17]. Removal of non-viral genomic host and bacterial nucleic acids and

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

enrichment of VLP is often necessary for the detection of low-titre viruses [17, 18]. More recently, the use of RNA sequencing of total non-ribosomal RNA from environmental samples gave rise to viral meta-transcriptomics, which has been successfully applied to characterise the viromes of diverse invertebrate and vertebrate species [13, 19, 20]. To our knowledge, no studies have been conducted to compare these two approaches, although doing so would allow us to understand the detection capabilities and biases associated with these different nucleic acid extraction and sample treatment methods.

We characterised the faecal virome of wild and captive Tasmanian devil using both the metagenomics approach based on VLP enrichment and sequence-independent amplification (hereafter “VLP metagenomics”) and the meta-transcriptomics approach based on RNA sequencing of total non-ribosomal RNA (hereafter “meta-transcriptomics”). Our objectives were threefold: (1) to provide a comprehensive characterisation of the faecal virome of Tasmanian devils; (2) to compare the two virome characterisation approaches, highlighting their advantages and potential challenges; and (3) to compare the faecal viromes of wild and captive Tasmanian devils.

Materials and Methods

Sample collection

Faecal samples were collected from wild Tasmanian devils between September 2016 and June 2017 from four locations in Tasmania (Figure 1); **A** Stony head (SH), **B** Buckby’s Road (BR), **C** Maria Island (MI), and **D** wukalina/Mt William National Park (wMW), as well as from captive devils at two Australian mainland zoos in June and July 2017 (Zoo A and Zoo B). Wild devils were trapped overnight during routine monitoring trips by Save the Tasmanian Devil Program staff, using baited PVC-pipe traps [21]. Upon capture, each animal was subjected to a thorough

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

health check, including body weight measurement, estimation of body condition score and observation for signs of disease. Fresh faecal samples were collected from either the traps, or from the hessian bag, during processing of the devils. For captive devils, faeces were collected from the enclosures shortly after defecation. All samples were stored in either liquid nitrogen or a portable -80^{o}C freezer (Stirling Ultracold, Global Cooling Inc.) immediately after collection. After arriving at the laboratory, samples were separated into two aliquots to be used in the extraction of total RNA for meta-transcriptomics and the enrichment of VLP for VLP metagenomics.

Meta-transcriptomics

Total RNA extraction, library preparation and sequencing

Samples were disrupted and homogenized in 600 µl of lysis buffer with 1.44 mm ceramic beads using a Bead Ruptor Homogenizer (Omni International) at 5 m.s^{-1} for 5 min. Total RNA was isolated using the Qiagen RNeasy Plus Mini Kit (Qiagen) following the manufacturer's instructions. The extracted total RNAs were pooled based on their source locations at equal mass amount, with each pool containing 5-10 samples. The RNA pools were depleted of host and bacteria rRNA using a Ribo-Zero-Gold (Epidemiology) kit (Illumina) before constructing sequencing libraries using a TruSeq total RNA Library Preparation Kit (Illumina). Paired-end (75 bp) sequencing of each library was performed on a NextSeq500 HO platform (Illumina) at the Ramaciotti Centre for Genomics (Sydney, Australia).

VLP Metagenomics

VLP enrichment and nucleic acid extraction

A second aliquot from each faecal sample was processed for the VLP metagenomics approach, as described previously with minor modifications [17].

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

122 VLPs were enriched as follows: Faecal suspensions consisting of 50 mg faeces in 700 µl of
123 were homogenized for 1 min using the Bead Ruptor Homogenizer (Omni International) at 5 $m.s^{-1}$
124 and centrifuged at 15000 × g for 3 min at 4 °C. The resulting supernatants were filtered
125 through 0.45 µm membrane filters (Corning) for 1 min at 15000 × g. The filtrates were treated
126 with 0.8 µl benzonase (Sigma Aldrich), 7 µl Turbo DNase (Ambion), 1 µl micrococcal nuclease
127 (New England Biolab) and 14.88 µl 1 x Turbo DNase buffer (Ambion) at 37 °C for 2 h to digest
128 free-floating nucleic acids. Viral DNA and RNA were then simultaneously extracted using the
129 QIAamp Viral RNA Mini Kit (Qiagen) according to the manufacturer's instructions, except
130 without addition of carrier RNA to the lysis buffer [17] .
131
132 *Random amplification*
133 Extracted viral nucleic acids were pooled based on their source locations at equal mass
134 amounts, using the same representative samples as the total RNA preparations for meta-
135 transcriptomics described above. The pooled extractions were subjected to first and second
136 strand synthesis and random PCR amplification for 22 cycles using the Complete Whole
137 Transcriptome Amplification (WTA2) kit (Sigma-Aldrich). The denaturation temperature was
138 increased from 70°C to 95°C to allow for the denaturation and amplification of both dsDNA and
139 RNA [17]. WTA2 PCR products were then purified using Agencourt AMPure XP beads
140 (Beckman Coulter) prior to library preparation and sequencing.
141
142 *Library preparation and Illumina Sequencing*
143 Sequencing libraries were constructed using the Nextera XT DNA Library Preparation Kit
144 (Illumina) according to the manufacturer's instructions with the following modifications as shown
145 in [17] : (i) the tagmentation time was reduced to 4 min to increase fragment size, (ii) input DNA
146 was increased to 1.2 ng/ μl^{-1} and (iii) reagent quantities were halved to further increase the

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

147 fragment size. Paired-end (100 bp) sequencing of each library was performed on a Hiseq2500
148 platform (Illumina) at the Ramaciotti Centre for Genomics (Sydney).
149
150 **Assembly and annotation**
151 Sequencing reads were de-multiplexed and quality trimmed with Trimmomatic [22] and
152 assembled *de novo* using Trinity [23]. Resulting contigs were compared against the non-
153 redundant nucleotide (nt) and protein (nr) databases on GenBank using Blastn and Blastx,
154 respectively, with an e-value cut-off at 1E-5. Blast search using Blastx was also conducted
155 against a bespoke database containing all viral RNA-dependent RNA polymerase (RdRp)
156 protein reference sequences downloaded from GenBank. Taxonomic information at the domain
157 level (i.e. Eukarya, Bacteria, or Archaea), as well as for viruses, was first assigned based on the
158 blastn results and then based on blastx results. Potential virus-related sequences were further
159 categorized into families and orders based on their genetic similarity to their closest relatives
160 and/or their phylogenetic position. Similarly, the assignment of viruses to the broader groups of
161 their hosts (i.e. Eukarya, Bacteria, or Archaea) was based on their phylogenetic relationship to
162 viruses with reliable host information obtained either using experimental or phylogenetic
163 approaches. The genetic identity cut-off for host assignment was set to be 40% based on the
164 most conserved proteins such as RdRp or DNA polymerases [24]. The threshold was set based
165 on the intra-family diversity of most vertebrate-specific virus families/genera [25]. The
166 assignment of vertebrate host was based on phylogenetic analyses, in which a potential devil
167 associated virus is expected to either cluster within, or form a sister group to, an existing
168 mammalian virus group.
169
170 To compare the abundance of each transcript/contig, we calculated the percentage of total
171 reads in the library. The abundance of host transcripts/contigs was estimated by mapping reads
172 against Tasmanian devil genome using Bowtie2 [26], whereas those of other organisms,

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

173 namely, viruses, bacteria, archaea, and non-host eukaryotes, were estimated using the RSEM
174 approach [27] implemented in Trinity.
175
176 For each virus, the genome sequence was further extended by merging related contigs. Gaps in
177 the genome were filled either by mapping the reads against the existing scaffold or by RT-PCR
178 and Sanger sequencing. Putative ORFs in viral genomes were predicted by the Geneious 8.1
179 software [28] or NCBI ORF finder, and annotated based on their similarity to previous published
180 virus genomes.
181
182 **Phylogenetic analysis**
183 Nucleotide sequences of complete or partial genomes and amino acid sequences from the
184 conserved domain (e.g. the RNA-dependent RNA polymerase; RdRp) of the newly
185 characterised viral sequences were aligned with those of reference viruses representative of the
186 diversity of the corresponding virus family or species. The alignment was performed using the
187 E-INS-I algorithm implemented in MAFFT (version 7) [29]. The quality of the alignments was
188 subsequently assessed, and all ambiguously aligned regions were removed using TrimAl
189 (version 1.2) [30]. Phylogenetic trees of aligned amino acid (all data sets with the exception of
190 RHDV) or nucleotide (RHDV) sequences were then inferred using the maximum likelihood
191 method implemented in PhyML version (3.0) [31], utilising the best-fit substitution model and the
192 Subtree Pruning and Regrafting (SPR) branch-swapping algorithm.
193
194 **Analyses of virome ecology**
195 Viral abundance tables (Additional file 2: Table S2 and Additional file 3: Table S3) were
196 generated based on complete or near complete viral contigs and the percentage of reads
197 mapped to them using Bowtie 2 [26] in each library. QIIME (v 1.9) [32] was used to perform
198 ecological and statistical analysis to compare viromes of different populations. Within-library

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

199 virotype richness (alpha diversity measured using the *Chao1* metric) and dissimilarity between
200 libraries (beta diversity measured using the Euclidean metric) for both VLP metagenomics and
201 meta-transcriptomics were calculated based on levels of viral abundance. The statistical
202 significance of differences in alpha diversity was evaluated by the Monte Carlo method (999
203 permutations), with a null hypothesis that diversity is equal in all libraries with a significance
204 threshold of $\alpha = 0.05$. Levels of viral abundance were also used to produce heatmaps and
205 dendrograms from hierarchical clustering. Principal coordinates analysis (PCoA) was performed
206 on the Euclidean distance matrix as calculated in QIIME, and additional cluster analysis was
207 conducted using K-Means clustering in R [33].
208
209 **PCR confirmation and Sanger sequencing of rabbit haemorrhagic disease virus (RHDV)**
210 Contigs with high percentage of similarity (>97%) to rabbit haemorrhagic disease virus (RHDV)
211 were detected in one of the meta-transcriptomics libraries. RHDV is used as a biocontrol for
212 rabbits in Australia and causes fatal hepatitis in European rabbits (*Oryctolagus cuniculus*) and
213 some hare species [34]. To confirm the detection of RHDV, RT-PCR was performed on
214 individual faecal RNA extractions using the Qiagen OneStep Ahead RT-PCR kit (Qiagen) and
215 primers from a previously validated Australian RHDV strain-specific PCR [35] as well as two
216 additional primer sets manually designed based on the current meta-transcriptomics assembled
217 contigs (Additional file 4: Table S4). PCR products were separated on 1.5% agarose gel (Bio-
218 Rad Laboratories, Hercules, CA, USA) in 1x tris-acetate EDTA, and visualized using SYBR Safe
219 DNA gel stain (Life Technologies, Carlsbad, CA, USA). Sanger sequencing of positive PCR
220 products was performed at the Australian Genome Research Facility (Sydney, Australia). In
221 addition, DNA was extracted from faecal samples using the ISOLATE Fecal DNA kit (Bioline,
222 London, UK) and presence of rabbit DNA was tested using primers targeting a 110 bp region of
223 the *Oryctolagus cuniculus* 12S mitochondrial rRNA gene (Fwd:

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

5′□CAAAAGTAAGCTCAATTACCACCGTA□3′; Rev: 5′□ATAAGGGCTTTCGTATATTCGGAA□3′) [36]. Rabbit DNA extracted from rabbit liver obtained from collaborators at CSIRO using the Bioline Isolate II Genomic DNA kit (Bioline) was used as a positive control. Cleanup, primer trimming and sequence analysis of Sanger data was performed using DNASTAR Lasergene [37] and Geneious [28].

Results

Overview of the devil virome

Location	Captive/Wild	Meta-transcriptomics	No. reads	No. contigs	VLP metagenomics	No.reads	No. contigs

We characterised the faecal virome of six pools consisting of a total of 54 faecal samples collected from Tasmanian devils from four wild and two captive sites using both meta-transcriptomics and VLP metagenomics approaches. Meta-transcriptomic libraries resulted in 128 - 140 million reads per pool (793,038,436 reads in total), which were assembled *de novo* into 196,919 – 358,327 contigs (Table 1). Blast analyses of sequence reads from the meta-transcriptomic protocol revealed large proportions of reads from Bacteria (55.11 – 67.77%), and only 4.32 – 7.88% from Eukarya. Mapping reads to the Tasmanian devil genome revealed that 10.75 – 18.99 % of reads originated from the host. The percentage of reads related to Archaea was less than 0.02%, and for viruses between 0.68 – 1.16% (Figure 2a).

VLP metagenomics resulted in 26 – 49 million reads per pool (237,174,236 reads in total), which were assembled *de novo* into 49,813 – 313,354 contigs (Table 1). In comparison to meta-transcriptomics, the VLP metagenomics protocol resulted in smaller proportions of non-viral nucleic acids (host and bacterial) and hence enriched viral nucleic acids (Figure 2a). The proportion of viral reads using VLP metagenomics varied from 14.69 – 60.02%, while the proportion of reads mapped to non-viral components were 17 – 49.46% for Bacteria, 0.17–2.42% for Eukarya, 1.29 –18.67% for host and less than 0.01% for Archaea (Figure 2a). For

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Zoo A	Captive	Zoo A	139,827,866	219,496	Zoo A	26,704,842	281,857
Zoo B	Captive	Zoo B	128,149,408	196,919	Zoo B	33,002,790	267,536
Maria Island	Wild	MI	129,783,390	261,483	MI	42,856,850	153,804
Buckbys Road	Wild	BR	128,589,556	358,327	BR	42,602,752	84,646
Wukalina/Mt William National Park	Wild	wMW	129,981,856	264,205	wMW	48,911,728	313,154
Stony Head	Wild	SH	136,706,360	325,769	SH	43,095,274	49,812
Total			793,038,436	1,626,199		237,174,236	1,150,809

both approaches, a substantial proportion of reads had no significant similarity to other sequences in the databases on GenBank (10.22 – 29.20%).

Table 1 Library information of meta-transcriptomics and VLP metagenomics in the present study

Despite detecting a smaller proportion of virus-related sequences with meta-transcriptomics than with VLP metagenomics, viruses detected with the meta-transcriptomics approach fell into a wide range of viral groups, of which 49.87 – 97.51% had the closest hits to RNA viruses and 2.49 – 50.13% to DNA viruses. Conversely, for VLP metagenomics, over 95.54% of the virus-related sequences had closest hits to DNA viruses, and less than 5% were identified as RNA viruses (Figure 2b). Meta-transcriptomics revealed high levels of viral diversity across all libraries, with the most abundant viral groups detected being the *Caudovirales, Luteo-Sobemo, Narna-Levi, Partiti-Picobirna, Picorna-Calici* and *Tombus-Noda* (Figure 2c). Conversely, VLP metagenomics revealed relatively lower viral diversity across the same libraries, with *Caudovirales* dominating significant proportions of the viral reads (69.89-99.49%). Other viral groups identified by VLP metagenomics, at much lower abundances, include the *Microviridae, Circoviridae, Genomoviridae, Parvoviridae, Herpesviridae, Polyomaviridae and Papillomaviridae* (Figure 2c). Overall, VLP metagenomics and meta-transcriptomics differed in the viruses detected as well as the expected counts (transcript abundance) as measured by the RSEM analysis (Figure 2d).

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The percentage of vertebrate viruses detected by meta-transcriptomics ranged between 0 – 9.41% of the total viral reads, while large proportions of viral reads belonged to either non-vertebrate eukaryotic viruses (45.08 – 97.51%), which included plant viruses, insect viruses and mycoviruses, or bacteriophage (2.48 – 48.91%) from the families *Siphoviridae, Podoviridae, Myoviridae and Microviridae*. In the VLP metagenomics data set, the percentage of reads associated with vertebrate viruses was also small, ranging between 0.04 – 0.84%, while the percentages of bacteriophage and other eukaryotic viruses ranged between 79.17 – 99.91% and 0.04 – 19.99%, respectively. Detailed information on all vertebrate viruses identified is presented in Additional file 1: Table S1.

Detection of viruses previously identified in other mammalian hosts (Rabbit haemorrhagic disease virus and Torovirus)

Rabbit haemorrhagic disease virus (RHDV) is a calicivirus in the genus *Lagovirus* [34]. All lagoviruses have a single-stranded, positive-sense RNA genome of approximately 7.5 kb and share a similar genome structure comprising two open reading frames (ORFs). Using meta-transcriptomics, we detected genomes with high nucleotide and amino acid similarity (>98 %) to RHDV in one of the wild devil meta-transcriptomic libraries (BR), with genome coverage of 98.1%. Phylogenetic analysis based on the nucleotide sequences of the major capsid and non-structural protein genes revealed that the RHDV detected here clustered with RHDV strain GI.2 (also called RHDV2) (Figure 3a and b), which was first detected in Australia in May 2015 and has since became the dominant circulating strain nationwide. RHDV-specific RT-PCR and sequencing confirmed the presence of RHDV2 in 4 of the 9 devil faecal samples from the BR meta-transcriptomics pool. In addition, no rabbit associated genes were detected during the initial sequence analysis. Additional PCR targeting a short fragment of rabbit mtDNA (<300bp) also did not detect any rabbit DNA in the original faecal samples from BR. Further RT-PCR and sequencing performed on the faecal RNA extractions from the remaining pools confirmed the

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

294 presence of RHDV in 1 of 10 devils from wMW, 2 of 9 devils from Zoo A and 1 of 9 from Zoo B.
295 One of the 4 additional RHDV positive samples, from Zoo A contained detectable levels of
296 rabbit mtDNA as confirmed by PCR.
297
298 Torovirus is a genus of viruses in the viral family *Coronaviridae* with a linear, positive-sense
299 RNA genome of about 28 – 28.5kb [38]. We identified the complete viral genome (28,463 bp) of
300 a novel torovirus most closely related to Bovine torovirus (Bredavirus) in one of the meta-
301 transcriptomic libraries (wMW). BLAST search revealed that the torovirus detected here shared
302 96% nucleotide similarity and 97% amino acid similarity with Bovine torovirus, a respiratory and
303 enteric pathogen of cattle that causes gastroenteritis and severe diarrhoea, particularly in young
304 calves. We determined the full genome structure of the novel torovirus strain, which included
305 ORF 1a and ORF 1b encoding the two polyproteins (pp1a and pp1ab), ORF 2 encoding for the
306 spike protein (S), ORF 3 encoding the membrane protein (M), ORF 4 encoding the
307 hemagglutinin-esterase protein (HE), and ORF 5 encoding the nucleocapsid protein (N) [39].
308 Based on the phylogenetic analysis of the spike protein amino acid (aa) sequence (4762 aa),
309 clustering of the novel torovirus strain with other toroviruses isolated from cows in the USA,
310 Japan and Europe indicated a bovine origin, although this will need to be confirmed with wider
311 sampling (Figure 3c).
312
313 **Detection and characterisation of novel marsupial-associated viruses (*Picornaviridae,***
314 ***Astroviridae, Reoviridae, Picobirnaviridae, Parvoviridae, Papillomaviridae, Herpesviridae,***
315 ***Polyomaviridae* and *Circoviridae*)**

316 The order *Picornavirales* includes a diverse group of vertebrate-infecting RNA viruses, including
317 enteroviruses and rhinoviruses, that can cause a wide range of diseases, such as poliomyelitis,
318 hand, foot and mouth disease, encephalitis, respiratory tract infections and the common cold

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

319 [40]. Members from the order are small, non-enveloped, positive-sense, single-stranded RNA
320 viruses about 7-8.8 kb in size [41]. There are currently five recognized families: *Picornaviridae,*
321 *Secoviridae, Iflaviridae, Marnaviridae* and *Dicistroviridae* [39]. We identified a novel member of
322 the *Picornaviridae* in one of the meta-transcriptomic libraries (wMW) and obtained the complete
323 genome of 8015 bp. According to the ICTV, members of a *Picornavirus* genus should share at
324 least 40% amino acid sequence identity in the polyprotein region [39]. The encoded 2396 aa
325 polyprotein exhibited 45.5% amino acid similarity to Simian sapelovirus, placing it in the genus
326 *Sapelovirus*. We have provisionally named this newly identified virus Devil sapelovirus (DeSV).
327 Phylogenetic analysis based on the amino acid sequence of the RdRp domains showed that
328 DeSV formed a sister lineage to sapeloviruses identified from eutherian mammals (i.e. Porcine
329 and Simian sapelovirus) (Figure 4a).
330
331 Astroviruses are single-stranded positive-sense RNA viruses from the family *Astroviridae*, that
332 can cause diarrhoea and gastroenteritis in infected hosts [42, 43]. The family has a genome size
333 of 6.4 to 7.3kb and has been identified in a broad range of vertebrate hosts including humans,
334 other mammals, reptiles, amphibians, fish and birds [44]. We detected astrovirus-related
335 sequences in 5 of the 6 meta-transcriptomic libraries (SH, wMW, BR, Zoo A and Zoo B). We
336 identified one complete and one near-complete genome sequence with 81.4% pairwise
337 nucleotide identity, denoting 2 separate strains of a single astrovirus species, which we have
338 tentatively named Devil astrovirus 1 (DeAstV1) [45]. The DeAstV1 identified here has a genome
339 structure typical of other astroviruses, with 3 putative open reading frames (ORF1a, ORF1b and
340 ORF2) each encoding for the protease, RdRp and capsid, respectively. In addition, we found a
341 ribosomal frameshift motif (AAAAAAC) within the ORF1a/1b overlap region. Phylogenetic
342 analysis based on the conserved RdRp domain showed that DeAstV1 formed a distinct cluster
343 that is more closely related to astroviruses of mammalian hosts (Mamastroviruses) than those of
344 avian hosts (Avastroviruses) (Figure 4b).

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Rotaviruses of the family *Reoviridae* are non-enveloped, double-stranded RNA viruses that infect only vertebrates through faecal-oral transmission [46]. They are the most common cause of acute viral gastroenteritis and can be found in a wide range of vertebrate hosts, predominately terrestrial mammals and birds. We detected rotavirus sequences in 3 of 6 meta-transcriptomic libraries (SH, BR and wMW, all are wild sites). Among them, we were able to identify 2 distinct segments (3481 bp and 3479 bp in length) encoding rotavirus VP1 (i.e. RdRp) from the two meta-transcriptomics libraries (BR and wMW, respectively). The two RdRp sequences shared >90 % nucleotide identity with each other, indicative of two different strains belonging to the same species. In addition, a contig encoding a partial rotavirus VP1 of 282 aa sharing 51 % sequence similarity with rotavirus H was also detected in meta-transcriptomics library wMW. We provisionally named the two viruses Devil rotavirus 1 and Devil rotavirus 2 (DeRoV1 and -2). The VP1 of DeRoV1 shared the highest amino acid similarity of ~51% with Rotavirus G, while DeRoV2 shared the highest amino acid similarity of 44% with Rotavirus H. Phylogenetic analysis with other previously characterised rotavirus species based on VP1 suggested that DeRoV1 and 2 form a distinct cluster that is most closely related to the cluster that contains Rotavirus G chicken, Human rotavirus B, Rotavirus I and Adult diarrheal rotavirus strain, which are associated with avian and mammalian hosts (Figure 4c).

Picobirnaviruses (PBV) are small, positive-stranded non-enveloped RNA viruses with abroad host ranges found in the faeces of many mammal, bird and reptile species [47]. We detected picobirnavirus sequences that encoded complete and partial viral RdRp (330 – 557 aa) in 4 of the 6 meta-transcriptomics libraries, all of which comprised of faecal samples collected from wild devils (BR, wMW, SH and MI). However, picobirnavirus sequences detected in library MI were too short to be phylogenetically informative and were discarded in the phylogenetic analysis. We provisionally named the novel picobirnaviruses detected here Devil picobirnavirus 1, 2, 3, 4, 5,

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

371 and 6 (DePBV1 -6) with two separate strains occurring in DePBV1 and DePBV5. Phylogenetic
372 analysis based on the RdRp domain of these viruses with other representative members from
373 the family showed that the 6 novel Devil picobirnaviruses are highly diverse and widely
374 distributed across the phylogenetic tree (Figure 4d).
375
376 *Parvoviridae* are a family of small, non-enveloped, ssDNA viruses [39]. Members of
377 *Parvoviridae* are associated with a number of diseases, including erythema infectiosum in
378 humans, hepatic and enteric diseases across a broad host range [48]. We identified two new
379 members of the vertebrate-associated sub-family *Parvovirinae* in the faeces of Tasmanian
380 devils from two of the VLP metagenomics libraries (MI and BR). We recovered partial and near-
381 complete protein sequences sharing ~50% identity to California sea lion bocavirus and Porcine
382 bocavirus, respectively. Two bocavirus-related sequences detected here shared >97% amino
383 acid sequence similarity, denoting two separate strains of the same species, which we have
384 provisionally named Devil bocavirus 1 (DeBoV1). A third bocavirus-related sequence was also
385 identified in VLP metagenomics library BR, sharing 71.13% amino acid similarity with DeBoV1.
386 We provisionally named this second virus Devil bocavirus 2 (DeBoV2). Phylogenetic analysis of
387 DeBoV1 and 2 in the context of other representative viruses from the *Parvoviridae* family further
388 confirmed the clustering of DeBoVs within the diversity of mammalian bocaviruses, although the
389 branching order involved DeBoVs and other members of bocavirus remain unresolved with
390 current topology (Figure 5a).
391
392 Papillomaviruses from the family *Papillomaviridae* are small, nonenveloped, icosahedral viruses
393 with a circular double-stranded DNA genome of about 8 kbp in size [39], and are associated
394 with the development of benign and malignant tumours [49, 50]. We identified fragmented
395 genome of two novel species of papillomavirus from a single VLP metagenomics library (MI),
396 among which we were able to retrieve two longer fragments of 1225 and 1335 bp in length,

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

respectively, both of which encode partial E1 protein, an ATP-dependent DNA helicase required for viral replication [51]. Since the two fragments shared approximately 64% similarity to each other, it suggests the presence of two distinct papillomavirus species, which we have tentatively designated as Devil papillomavirus 1 and 2 (DePV1 and -2). Phylogenetic analysis based on the E1 protein showed that DePV1 and 2 form a distinct cluster with *Bettongia penicillata* papillomavirus type 1 (BpPV1) isolated from the woylie (Figure 5b). While the marsupial papillomaviruses viruses are clustered together in the phylogenetic tree, their relationship with viruses identified from eutherian mammals remain unresolved.

Polyomaviruses from the family *Polyomaviridae* are small, nonenveloped, circular DNA viruses about 5 kb in size that have been identified in a wide variety of mammalian and avian hosts [39, 52]. We detected two novel polymaviruses from the faeces of Tasmanian devils in 3 VLP metagenomics libraries (MI, wMW and Zoo B), which we have tentatively designated Devil polyomavirus 1 and 2 (DePyv1 and -2), respectively. We recovered the complete circular genome of 4894 bp for DePyv1 (Figure 5c) and partial gene sequence (2251 bp) of the large T antigen (LTAg) protein for DePyv2. Phylogenetic analyses revealed strikingly different evolutionary histories for the structural and non-structural parts of the genome (Figure 5c), indicative of recombination [53]. In the LTAg phylogeny, DePyV1 and -2 formed a distinct lineage with another marsupial virus Bandicoot papillomatosis carcinomatosis virus type 2 (BPCV-2), which in turn is clustered with polyomaviruses of avian hosts (Figure 5c). In contrast, in the VP1 phylogeny, the marsupial viruses showed no close relationship with the avian viruses. Interestingly, the Bandicoot papillomatosis carcinomatosis viruses (BPCV1 and -2) only showed close relationship to DePyV1 at the LTAg region but not the VP1 region. Indeed, both BPCV1 and -2 are hybrid viruses between the family *Papillomaviridae* and *Polyomaviridae* [54, 55]. While its non-structural genes (i.e. large T and small T antigen genes) are related to

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

polyomavirus, its structural genes (i.e. L1 and L2 protein genes) are related to papillomavirus [54].

Herpesviridae are a large family of DNA viruses with large linear, dsDNA genome of about 120-240 kb, and that comprise 3 subfamilies; *Alphaherpesvirinae*, *Betaherpesvirinae* and *Gammaherpesvirinae* [39]. In one of the captive VLP metagenomics libraries (Zoo A), we identified more than 70 contigs matching different regions of a herpesvirus genome, which totaled 62,821 bp in length and included partial gene sequences of the DNA polymerase (575 aa), major DNA binding protein (465 aa), helicase (396 aa), glycoprotein M (378 aa) and H (427 aa), and major capsid protein (358 aa), amongst others. We tentatively named the novel herpesvirus Dasyurid herpesvirus 3 (DaHV-3). Phylogenetic analysis based on these non-structural and structural proteins suggested that DaHV-3 clustered with other gammaherpesviruses from the *Gammaherpesvirinae* subfamily (Additional file 5: Figure S1 and Additional file 6: Figure S2). Further analysis of the *Gammaherpesvirinae* phylogenetic structure based on the DNA polymerase showed that DaHV-3 forms a distinct lineage that is most closely related to Bovine gammaherpesvirus 6 and Common bottlenose dolphin gammaherpesvirus 1 strain Sarasota (Figure 5d). The previously characterised Dasyurid herpesvirus 2 (DaHV-2) isolated from Tasmanian devils [12] could not be included in the phylogenetic analysis because there are no available sequences from the same genomic regions. A BLASTx search of the DNA polymerase showed that DaHV-3 exhibited the greatest amino acid similarity (93.3%) with the previously characterised Macropodid herpesvirus 3(MaHV-3) [12], whose DNA polymerase amino acid sequence was also too short (<50% of the other representative herpesviruses) to be included in the phylogenetic analysis.

Circoviruses from the family *Circoviridae* have a circular ssDNA genome which range from 1.7 to 2.3 kb in size [56]. Here, we identified circovirus-related sequences and recovered the partial

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

448 replicase gene sequences (899 bp) in one of the wild devil metagenomes (SH) and tentatively
449 named it Devil circovirus (DeCV). Phylogenetic analysis based on the Rep proteins of the novel
450 DeCV and representative strains of circoviruses and cycloviruses suggested that DeCV is
451 clustered with circoviruses previously isolated from bats and pigs, sharing the highest sequence
452 identity (62%) with bat circovirus (AIF76281.1) (Figure 5e).
453
454 **Other viruses: plant and insect viruses, and bacteriophage**
455 In both the VLP metagenomics and meta-transcriptomics analyses, large proportions of viral
456 sequence reads could be attributed to viruses that infect plants, insects and bacteria. Indeed,
457 bacteriophage sequences from the order *Caudovirale*s were detected in all libraries and made
458 up over 90% of all viral VLP metagenomic reads and up to 48.91% of the meta-transcriptomic
459 reads. Sequences related to newly identified arthropod viruses, such as Wuhan fly virus and
460 Wuhan mosquito virus were also detected. Most of the insect viruses detected belong to RNA
461 virus groups the *Bunyavirales*, the *Mononegavirales*, and the *Chuviridae*, as well as DNA virus
462 subfamily *Densovirinae* from the *Parvoviridae* family. The detection of these arthropod virus
463 related sequences indicates the possible ingestion of arthropods and/or the environmental
464 contamination of faeces. Despite being a carnivorous species, sequences related to various
465 plant and fungal viruses were also observed in all libraries, including various sobemoviruses,
466 tombusviruses and mitoviruses. Herbivorous species such as wallabies often make up a large
467 proportion of devil's diet, and the presence of plant viruses in these prey items might be easily
468 detected in the faeces of devil through deep sequencing, especially if the intestines of the prey
469 were consumed [1]. Nonetheless, devils have also been observed to consume vegetation, and
470 the abundance of highly diverse eukaryotic and vertebrate viruses that are thought to be
471 associated with diverse host taxa are reflective of devils' overall generalist diet [1].
472
473 **Virome ecology: comparison between devil populations**

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Within-library viral diversity (alpha diversity) as characterised by our meta-transcriptomics approach was significantly different between captive and wild populations ($p < 0.05$). In general, captive populations had lower diversity in their faecal viromes compared to wild populations, as measured using the chao1 metric. However, when considering bacteriophage alone, we did not find any significant associations between devil populations (captive or wild) in either diversity or abundance.

The meta-transcriptomics analysis of the virome of Maria Island (MI) devils displayed a similar level of overall viral diversity to the two captive populations (Zoo A and Zoo B), and was lower than that found in the three other wild populations (wMW, BR and SH) (Figure 6a). Conversely, alpha diversity as characterized from VLP metagenomics data did not differ significantly between libraries (Figure 6b). Cluster analysis indicated that in the meta-transcriptomics data, wild and captive devils fell into two distinct clusters, while in VLP metagenomics data, BR formed its own cluster and the remaining populations formed a second cluster (Figure 6a and b).

Discussion

We provide the first comprehensive characterisation of the faecal virome of an endangered marsupial species, the Tasmanian devil. Using both VLP metagenomics and meta-transcriptomics, we identified a huge diversity of viruses in the faeces of devils, including vertebrate viruses, bacteriophage and other eukaryotic viruses. Novel viruses identified as potentially marsupial-associated – including, sapelovirus, rotaviruses, picobirnaviruses, astrovirus, bocaviruses, papillomaviruses, herpesvirus and polyomaviruses – were detected and have greatly expanded our current knowledge of viruses that are found within this unique group of mammals.

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our comparison of VLP metagenomics and meta-transcriptomics approaches revealed marked differences in terms of the viruses detected. In general, VLP metagenomics mainly detected DNA viruses, while meta-transcriptomics detected both DNA and RNA viruses, although the DNA viruses detected with meta-transcriptomics were limited to those with relatively high abundance (Fig 2d). A high abundance level is often indicative of an active viral infection, during which DNA viruses are transcribed into RNA intermediates that can be readily detected by RNA sequencing [57]. Conversely, RNA viruses identified in meta-transcriptomic were rarely detected in VLP metagenomics, even if they were highly abundant based on the RSEM estimated counts.

The use of VLP enrichment and sequence-independent amplification in VLP metagenomics increased the number of viral reads in each library compared to the meta-transcriptomic approach. However, the viral compositions of all six VLP metagenomes were highly skewed towards DNA viruses, particularly DNA bacteriophage from the order *Caudovirales*. This is consistent with a previous study comparing various enrichment methods, in which bacteriophage accounted for >80% of all reads in all of the enrichment methods tested, but <5% when no enrichment steps were incorporated [58]. Despite being able to substantially increase the total number of viral reads in the metagenomes, sequence-independent amplification is bias prone, resulting in fewer viruses being detected and lower genome coverage due to preferential amplification of certain sequences [59-61]. However, the over-representation of bacteriophage found here could also be attributed to the fact that they make up the bulk of the virobiota in the gut, which is dominated by bacteria [62, 63]. Regardless of its known bias [58, 59], VLP metagenomics still holds merit for use in virome characterisation due to its ability to identify low abundance DNA viruses, which is especially relevant for dormant or non-active viruses.

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

524 Compared to VLP metagenomics, meta-transcriptomics is non-viral specific, requires less
525 sample processing, and reveals the entire transcriptome within a sample [20, 64]. Omitting the
526 need for VLP enrichment and additional sample processing, the likelihood of biased detection is
527 plausibly reduced in meta-transcriptomics. In this study, the proportions of viral reads
528 sequenced by meta-transcriptomics were less than 2% per library, but the numbers of viral
529 groups detected were significantly higher than those detected in VLP metagenomics, which
530 included both RNA and DNA viruses.
531
532 Importantly, then, our results show that the taxonomic compositions of viral communities as
533 revealed by VLP metagenomics and meta-transcriptomics were not interchangeable and neither
534 of the approaches was able to detect all viruses present. However, these two approaches were
535 complementary to one-another, and an integrated approach using both VLP metagenomics and
536 meta-transcriptomics will prove to be a powerful tool for obtaining a complete overview of both
537 the taxonomic and functional profiles of viral communities in a sample.
538
539 Ecological analysis of virome composition and diversity revealed significant differences between
540 captive and wild devil populations, especially when characterised using meta-transcriptomics.
541 The two captive populations displayed lower levels of viral diversity compared to the four wild
542 populations. This loss of diversity within the captive populations is consistent with trends
543 previously observed in the gut bacteriome, where captive devils also exhibited lower bacterial
544 diversity compared to wild devils [65]. The microbiome is dynamic and sensitive to changes in
545 the environment. Hence, the extreme lifestyle and diet changes that often occur in captivity are
546 likely to affect the viral communities of devils. Interestingly, Maria Island devils had viromes that
547 are more similar to the two captive populations. Maria Island is a 115 km^2 island off the east
548 coast of Tasmania (Fig 1) and is home to approximately 100 free ranging devils. There were no
549 Tasmanian devils living on Maria Island until 27 devils were introduced to the island in 2012 and

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2013 from several captive populations [66]. Non-native vertebrate species, such as cats, black rats and house mice, are present on the island, whilst other species such as rabbits and livestock are absent, although all these species are common in the wild on mainland Tasmania. Two factors may have potentially contributed to the lower viral diversity observed on Maria Island. First, due to its isolation from mainland Tasmania, animal movements or immigrations are limited to only marine or bird species. This restriction to animal movements (particularly vertebrate species) between the mainland and Maria Island may in turn limit the introduction of viruses to the island. Second, captive-born devils are likely to have a “captive-type” virome, that is, lower viral diversity compared to wild devils. Indeed, some of the devils included in this study are captive-born animals recently translocated to the island. As samples were pooled prior to sequencing, it was not possible to distinguish between viromes of captive-born and wild-born devils. To understand whether captive devils can reacquire a “wild-type” virome following translocations, future studies should focus on comparing captive born translocated devils with the incumbent devils in the wild.

Overall, we detected sequences related to 20 vertebrate viruses, including 18 novel marsupial related viruses and two known mammalian viruses (RHDV2 and torovirus). While some of the viruses identified in this study come from families that include important pathogens, their pathogenic potential in devils is unclear. It is also important to note that some of these viruses may in fact be dietary viruses or occur naturally as part of the normal gut flora and therefore do not necessarily lead to disease. In the case of RHDV, we showed that RHDV2 was present in both captive and wild devils. In some areas of Tasmania, rabbits are common and may be a dietary component of wild devils. They are also regularly fed to devils in captivity, along with other meat sources such as wallaby and chicken. Feeding records provided by the two zoos in this study confirmed the feeding of rabbits to some of the sampled devils (zoo A) around the time of faecal collection. While we did not detect rabbit genes by meta-transcriptomics, VLP

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

576 metagenomics or PCR in any of the wild devil samples that tested positive for RHDV, we did
577 detect rabbit mtDNA in the stools of one captive devil from zoo A that tested positive for RHDV
578 by PCR. Thus, more specific diagnostic techniques such as targeted PCR of blood or internal
579 tissues (i.e. liver), *in situ* hybridization and serological assays are required to determine whether
580 these viruses can actively replicate in devils and cause disease or are simply gut contaminants.
581 Nevertheless, exposure to host-adapted viruses could pose significant health threats, especially
582 for devils that are immunocompromised due to DFTD, old age or other concurrent diseases [67,
583 68]. Furthermore, even commensal or latent viral infections can be exacerbated or reactivated in
584 immunocompromised hosts [69, 70]. For example, in the giant panda, while a number of viruses
585 (e.g. papillomaviruses, picornaviruses and anelloviruses) were detected in both sick and healthy
586 animals, the virus titres were much higher in the diseased individuals, indicating a compromised
587 balance between host immune response and virus replication [71].
588
589 Tasmanian devils have low genetic diversity both across their genomes and at functionally
590 important loci such as the major histocompatibility complex (MHC) [72-74]. This lack of diversity
591 renders them particularly vulnerable to changes in the environment, including the emergence of
592 new pathogens [75]. For instance, populations of Italian agile frogs with lower microsatellite
593 diversity have been shown to be more susceptible to an emerging strain of Ranavirus (FV3- frog
594 virus 3) and experience higher mortality rates than populations with higher diversity [76].
595 Similarly in cheetahs, a coronavirus-associated feline infectious peritonitis outbreak causing
596 mass mortality in a captive breeding colony was linked to the species' extreme genetic
597 monomorphism, particularly at the MHC [77]. While none of the devils included in this study
598 exhibited overt signs of disease at the time of sample collection, the viruses described here will
599 provide a fundamental baseline of the normal Tasmanian devil faecal virome, which can be
600 used as a reference for comparing healthy and diseased animals.
601

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Phylogenetic analyses of the newly identified viruses, including divergent members of their respective viral families, provided us with insights into the evolutionary history of marsupial-associated viruses relative to viruses of eutherian mammals and other host taxa. Generally, a long-term relationship between viruses and hosts are expected for mammalian viruses [25]. Strong evidence for this lies in the observation that devil viruses are usually clustered with other marsupial viruses, as a marsupial-specific lineage that is distinct from the eutherian viruses, as observed in papillomavirus [78, 79], polyomavirus, and herpesvirus (Fig. 5). Furthermore, in several cases the branching order of viruses broadly reflects that of their hosts such that a general co-divergence can be inferred. For instance, in the phylogenies of *Picornaviridae* and *Astroviridae*, the devil (marsupial) viruses formed sister clade to eutherian viruses, which in turn are sister to avian viruses, consistent with the evolutionary history of the host. Although such a relationship is not observed in every virus phylogeny, a deep divergence between eutherian and marsupial viruses is typical of our dataset. This observation indicates that the timescale of virus evolution is very likely to reflect that of the hosts [25].

The gut virome is increasingly recognized as an integral component of the gut microbiome, and studies of the devil virome will continue to shine light on the biology and health of this iconic endangered species. For example, bacteriophage, which we have shown to dominate the devil's faecal virome, can contribute to host health by maintaining the diversity and structure of the gut bacteriome through direct interactions with the bacterial communities. While the functions of bacteriophage on devil health remain to be determined, future studies will be able to exploit the extensive microbiomic data that is now available to answer important questions about the host-microbe relationship between devils and their microbiome [65].

Conclusion

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

627 We provided the first comprehensive characterisation of the faecal virome of the Tasmanian
628 devil, greatly expanding our knowledge of the diverse groups of vertebrate viruses observed in
629 this endangered species. Similar to their bacterial microbiome, captive devils have significantly
630 lower diversity in their faecal viromes compared to wild devils, likely reflective of their captive
631 diets and lifestyle. Identification of vertebrate and marsupial-specific viruses in devils provides
632 potential candidate viruses for future disease surveillance and routine screening as part of the
633 broader conservation management of devils. However, future work will first need to focus on
634 elucidating the pathogenic impact of these viruses on devil health. Finally, we showed that a
635 combination of VLP metagenomics and meta-transcriptomics may be a more comprehensive
636 virome characterisation approach that will encompass both DNA and RNA viruses.
637

638 ## Abbreviations

639 SH: Stony Head; MI: Maria Island; BR: Buckbys Road; wMW: wukalina/Mt William National
640 Park; VLP: Virus-like particles; RHDV: Rabbit haemorrhagic disease virus; DeSV: Devil
641 sapelovirus; DeAstV1: Devil astrovirus 1; DeRoV1 and -2; Devil rotavirus 1 and 2; DePBV1-6:
642 Devil picobirnavirus 1-6; DeBoV1 and -2: Devil bocavirus 1 and 2; DePV1 and -2: Devil
643 papillomavirus 1 and 2; DePyV1 and -2: Devil polyomavirus 1 and 2; DaHV-3: Dasyurid
644 herpesvirus 3; DeCV: Devil circovirus

645 ## Acknowledgements

646 Thank you to the field teams of the Save the Tasmanian Devil Program and the University of
647 Sydney who collected faecal samples. Thanks also to N. Mooney for his assistance at the BR
648 site, the two zoos (and their keeping staff) who provided access to faecal samples from the
649 captive devils, J. Mahar for providing rabbit liver tissue and A.V. Lee for the map in Figure 1.

650 ## Funding

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

This work was funded by an Australian Research Council (ARC; LP140100508) grant to KB/CEG/CH/VB.

ECH is funded by an ARC Australian Laureate Fellowship (FL170100022).

Availability of data and material

Raw sequence reads generated in this study are available in NCBI SRA database under the BioProject ID PRJNA495667. Viral genomes described in detail here are available in the figshare repository https://doi.org/10.6084/m9.figshare.7185146.v3 and GenBank (submission currently under process).

Authors' contributions

RC, VB, MS, CH, CEG and KB designed the experiments. CH coordinated sample collection from all sites, with RC, CEG, CH participating in collection from the zoos, BR, SH (respectively). RC carried out the extractions and PCRs. RC, MS analysed the data. RC wrote the manuscript and all authors critically reviewed and edited the manuscript.

Ethics approval and consent to participate

Wild Tasmanian devils were trapped by the Save the Tasmanian Devil Program staff under the Tasmanian Government Department of Primary Industries, Parks, Water and Environment Standard Operating Procedure "Trapping and handling wild Tasmanian devils".

Consent for publication

Not applicable

Competing interests

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

672 The authors declare no competing interests.

673 ## References

674 1. Pemberton, D., S. Gales, B. Bauer, R. Gales, B. Lazenby, and K. Medlock, *The diet of the*
675 *Tasmanian devil, Sarcophilus harrisii, as determined from analysis of scat and stomach contents.*
676 Papers and Proceedings of the Royal Society of Tasmania, 2008. **142**(2): p. 13-22.
677 2. Owen, D. and D. Pemberton, *Made for travelling rough': devil ecology*, in *Tasmanian Devil: A*
678 *Unique and Threatened Animal*. 2005, Allen & Unwin.
679 3. Lazenby, B.T., M.W. Tobler, W.E. Brown, C.E. Hawkins, G.J. Hocking, F. Hume, S. Huxtable, P. Iles,
680 M.E. Jones, and C. Lawrence, *Density trends and demographic signals uncover the long-term*
681 *impact of transmissible cancer in Tasmanian devils.* Journal of Applied Ecology, 2018. **55**(3): p.
682 1368-1379.
683 4. Pye, R.J., D. Pemberton, C. Tovar, J.M. Tubio, K.A. Dun, S. Fox, J. Darby, D. Hayes, G.W. Knowles,
684 and A. Kreiss, *A second transmissible cancer in Tasmanian devils.* Proceedings of the National
685 Academy of Sciences, 2016. **113**(2): p. 374-379.
686 5. Hogg, C., A. Lee, C. Srb, and C. Hibbard, *Metapopulation management of an endangered species*
687 *with limited genetic diversity in the presence of disease: the Tasmanian devil Sarcophilus harrisii.*
688 International Zoo Yearbook, 2017. **51**(1): p. 137-153.
689 6. Jones, M.E., P.J. Jarman, C.M. Lees, H. Hesterman, R.K. Hamede, N.J. Mooney, D. Mann, C.E.
690 Pukk, J. Bergfeld, and H. McCallum, *Conservation management of Tasmanian devils in the*
691 *context of an emerging, extinction-threatening disease: devil facial tumor disease.* EcoHealth,
692 2007. **4**(3): p. 326-337.
693 7. Loh, R., J. Bergfeld, D. Hayes, A. O'hara, S. Pyecroft, S. Raidal, and R. Sharpe, *The pathology of*
694 *devil facial tumor disease (DFTD) in Tasmanian devils (Sarcophilus harrisii).* Veterinary Pathology,
695 2006. **43**(6): p. 890-895.
696 8. McCallum, H., *Tasmanian devil facial tumour disease: lessons for conservation biology.* Trends in
697 Ecology & Evolution, 2008. **23**(11): p. 631-637.
698 9. Kreiss, A., Y. Cheng, F. Kimble, B. Wells, S. Donovan, K. Belov, and G.M. Woods, *Allorecognition*
699 *in the Tasmanian devil (Sarcophilus harrisii), an endangered marsupial species with limited*
700 *genetic diversity.* PLOS one, 2011. **6**(7): p. e22402.
701 10. Pearse, A.-M. and K. Swift, *Allograft theory: transmission of devil facial-tumour disease.* Nature,
702 2006. **439**(7076): p. 549-549.
703 11. Grueber, C.E., E. Peel, R. Gooley, and K. Belov, *Genomic insights into a contagious cancer in*
704 *Tasmanian devils.* Trends in Genetics, 2015. **31**(9): p. 528-535.
705 12. Stalder, K., P.K. Vaz, J.R. Gilkerson, R. Baker, P. Whiteley, N. Ficorilli, L. Tatarczuch, T. Portas, K.
706 Skogvold, and G.A. Anderson, *Prevalence and clinical significance of herpesvirus infection in*
707 *populations of Australian marsupials*. PLOS one, 2015. **10**(7): p. e0133807.
708 13. Shi, M., X.-D. Lin, X. Chen, J.-H. Tian, L.-J. Chen, K. Li, W. Wang, J.-S. Eden, J.-J. Shen, and L. Liu,
709 *The evolutionary history of vertebrate RNA viruses.* Nature, 2018: p. 1.
710 14. Weisbecker, V. and R. Beck, *Marsupial and Monotreme Evolution And Biogeography*. 2015. 1-32.
711 15. Djikeng, A., R. Halpin, R. Kuzmickas, J. DePasse, J. Feldblyum, N. Sengamalay, C. Afonso, X.
712 Zhang, N.G. Anderson, and E. Ghedin, *Viral genome sequencing by random priming methods.*
713 BMC genomics, 2008. **9**(1): p. 5.
714 16. Thurber, R.V., *Methods in Viral Metagenomics.* Handbook of Molecular Microbial Ecology II:
715 Metagenomics in different habitats: Wiley-Blackwell, 2011.

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

17. Conceicao-Neto, N., M. Zeller, H. Lefrere, P. De Bruyn, L. Beller, W. Deboutte, C.K. Yinda, R. Lavigne, P. Maes, M. Van Ranst, E. Heylen, and J. Matthijnssens, *Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis.* Scientific Report, 2015. **5**: p. 16532.
18. Hall, R.J., J. Wang, A.K. Todd, A.B. Bissielo, S. Yen, H. Strydom, N.E. Moore, X. Ren, Q.S. Huang, and P.E. Carter, *Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery.* Journal of virological methods, 2014. **195**: p. 194-204.
19. Shi, M., X.-D. Lin, J.-H. Tian, L.-J. Chen, X. Chen, C.-X. Li, X.-C. Qin, J. Li, J.-P. Cao, and J.-S. Eden, *Redefining the invertebrate RNA virosphere.* Nature, 2016. **540**(7634): p. 539.
20. Shi, M., P. Neville, J. Nicholson, J.-S. Eden, A. Imrie, and E.C. Holmes, *High-Resolution Metatranscriptomics Reveals the Ecological Dynamics of Mosquito-Associated RNA Viruses in Western Australia.* Journal of Virology, 2017. **91**(17): p. e00680-17.
21. Hawkins, C.E., C. Baars, H. Hesterman, G.J. Hocking, M.E. Jones, B. Lazenby, D. Mann, N. Mooney, D. Pemberton, S. Pyecroft, M. Restani, and J. Wiersma, *Emerging disease and population decline of an island endemic, the Tasmanian devil Sarcophilus harrisii.* Biological Conservation, 2006. **131**(2): p. 307-324.
22. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-2120.
23. Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, and Q. Zeng, *Full-length transcriptome assembly from RNA-Seq data without a reference genome.* Nature biotechnology, 2011. **29**(7): p. 644.
24. Rost, B., *Twilight zone of protein sequence alignments.* Protein Engineering, Design and Selection, 1999. **12**(2): p. 85-94.
25. Shi, M., X.-D. Lin, X. Chen, J.-H. Tian, L.-J. Chen, K. Li, W. Wang, J.-S. Eden, J.-J. Shen, and L. Liu, *The evolutionary history of vertebrate RNA viruses.* Nature, 2018. **556**(7700): p. 197.
26. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature methods, 2012. **9**(4): p. 357.
27. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC bioinformatics, 2011. **12**(1): p. 323.
28. Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, and C. Duran, *Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.* Bioinformatics, 2012. **28**(12): p. 1647-1649.
29. Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7: improvements in performance and usability.* Molecular biology and evolution, 2013. **30**(4): p. 772-780.
30. Capella-Gutiérrez, S., J.M. Silla-Martínez, and T. Gabaldón, *trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses.* Bioinformatics, 2009. **25**(15): p. 1972-1973.
31. Guindon, S. and O. Gascuel, *A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.* Systematic biology, 2003. **52**(5): p. 696-704.
32. Caporaso, J.G., J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, N. Fierer, A.G. Pena, J.K. Goodrich, J.I. Gordon, G.A. Huttley, S.T. Kelley, D. Knights, J.E. Koenig, R.E. Ley, C.A. Lozupone, D. McDonald, B.D. Muegge, M. Pirrung, J. Reeder, J.R. Sevinsky, P.J. Turnbaugh, W.A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight, *QIIME allows analysis of high-throughput community sequencing data.* Nat Methods, 2010. **7**(5): p. 335-6.
33. R Core Team, *R: A language and environment for statistical computing*. 2018: R Foundation for Statistical Computing, Vienna, Austria.

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

34. Abrantes, J., W. Van Der Loo, J. Le Pendu, and P.J. Esteves, *Rabbit haemorrhagic disease (RHD) and rabbit haemorrhagic disease virus (RHDV): a review.* Veterinary research, 2012. **43**(1): p. 12.
35. Hall, R.N., J.E. Mahar, A.J. Read, R. Mourant, M. Piper, N. Huang, and T. Strive, *A strain-specific multiplex RT-PCR for Australian rabbit haemorrhagic disease viruses uncovers a new recombinant virus variant in rabbits and hares.* Transboundary and emerging diseases, 2018. **65**(2): p. e444-e456.
36. Martín, I., T. García, V. Fajardo, M. Rojas, N. Pegels, P. Hernández, and I. González, *Polymerase chain reaction detection of rabbit DNA in food and animal feed.* World Rabbit Science, 2010. **17**(1): p. 27-36.
37. Burland, T.G., *DNASTAR's Lasergene sequence analysis software*, in *Bioinformatics methods and protocols*. 2000, Springer. p. 71-91.
38. Snijder, E.J. and M.C. Horzinek, *Toroviruses: replication, evolution and comparison with other members of the coronavirus-like superfamily.* Journal of general virology, 1993. **74**(11): p. 2305-2316.
39. King, A.M.Q., M.J. Adams, E.B. Carstens, and E.J. Lefkowitz, *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses.* 2012, Elsevier: San Diego.
40. Koonin, E.V., Y.I. Wolf, K. Nagasaki, and V.V. Dolja, *The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups.* Nature Reviews Microbiology, 2008. **6**(12): p. 925.
41. Knowles, N., T. Hovi, T. Hyypiä, A. King, A.M. Lindberg, M. Pallansch, A. Palmenberg, P. Simmonds, T. Skern, and G. Stanway, *Family-Picornaviridae.* 2012.
42. Herrmann, J.E., D.N. Taylor, P. Echeverri, and N.R. Blacklow, *Astroviruses as a cause of gastroenteritis in children.* New England Journal of Medicine, 1991. **324**(25): p. 1757-1760.
43. Finkbeiner, S.R., Y. Li, S. Ruone, C. Conrardy, N. Gregoricus, D. Toney, H.W. Virgin, L.J. Anderson, J. Vinjé, and D. Wang, *Identification of a novel astrovirus (astrovirus VA1) associated with an outbreak of acute gastroenteritis.* Journal of virology, 2009. **83**(20): p. 10836-10839.
44. Méndez, E. and C.F. Arias, *Astrovirus*, in *Clinical Virology, Third Edition*. 2009, American Society of Microbiology. p. 1145-1153.
45. Bosch A., G.S., Krishna N.K., Méndez E., Monroe S.S., Pantin-Jackwood M., Schultz-Cherry S. , *Family - Astroviridae*, in *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses.* A.M.Q. King, et al., Editors. 2012, Elsevier: San Diego. p. 953-959.
46. Estes, M.K. and J. Cohen, *Rotavirus gene structure and function.* Microbiological reviews, 1989. **53**(4): p. 410-449.
47. Ganesh, B., G. Masachessi, and Z. Mladenova, *Animal picobirnavirus.* Virusdisease, 2014. **25**(2): p. 223-238.
48. Siegl, G., R.C. Bates, K.I. Berns, B.J. Carter, D.C. Kelly, E. Kurstak, and P. Tattersall, *Characteristics and taxonomy of Parvoviridae.* Intervirology, 1985. **23**(2): p. 61-73.
49. Walboomers, J.M., M.V. Jacobs, M.M. Manos, F.X. Bosch, J.A. Kummer, K.V. Shah, P.J. Snijders, J. Peto, C.J. Meijer, and N. Muñoz, *Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.* The Journal of pathology, 1999. **189**(1): p. 12-19.
50. Chaturvedi, A.K., E.A. Engels, R.M. Pfeiffer, B.Y. Hernandez, W. Xiao, E. Kim, B. Jiang, M.T. Goodman, M. Sibug-Saber, and W. Cozen, *Human papillomavirus and rising oropharyngeal cancer incidence in the United States.* Journal of clinical oncology, 2011. **29**(32): p. 4294.
51. Wilson, V.G., M. West, K. Woytek, and D. Rangasamy, *Papillomavirus E1 proteins: form, function, and features.* Virus Genes, 2002. **24**(3): p. 275-90.

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

52. Johne, R., C.B. Buck, T. Allander, W.J. Atwood, R.L. Garcea, M.J. Imperiale, E.O. Major, T. Ramqvist, and L.C. Norkin, *Taxonomical developments in the family Polyomaviridae.* Archives of virology, 2011. **156**(9): p. 1627-1634.
53. Buck, C.B., K. Van Doorslaer, A. Peretti, E.M. Geoghegan, M.J. Tisza, P. An, J.P. Katz, J.M. Pipas, A.A. McBride, and A.C. Camus, *The ancient evolutionary history of polyomaviruses.* PLoS pathogens, 2016. **12**(4): p. e1005574.
54. Woolford, L., A. Rector, M. Van Ranst, A. Ducki, M.D. Bennett, P.K. Nicholls, K.S. Warren, R.A. Swan, G.E. Wilcox, and A.J. O'Hara, *A novel virus detected in papillomas and carcinomas of the endangered western barred bandicoot (Perameles bougainville) exhibits genomic features of both the Papillomaviridae and Polyomaviridae.* Journal of virology, 2007. **81**(24): p. 13280-13290.
55. Bennett, M.D., L. Woolford, H. Stevens, M. Van Ranst, T. Oldfield, M. Slaven, A.J. O'Hara, K.S. Warren, and P.K. Nicholls, *Genomic characterization of a novel virus found in papillomatous lesions from a southern brown bandicoot (Isoodon obesulus) in Western Australia.* Virology, 2008. **376**(1): p. 173-182.
56. Lukert, P., G. De Boer, J. Dale, P.K. Keese, M. McNulty, J.W. Randles, and I. Tischer, *Family Circoviridae*. 1995, Springer-Verlag.
57. Shi, M., Y.-Z. Zhang, and E.C. Holmes, *Meta-transcriptomics and The Evolutionary Biology of RNA Viruses.* Virus research, 2017.
58. Kleiner, M., L.V. Hooper, and B.A. Duerkop, *Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes.* BMC genomics, 2015. **16**(1): p. 7.
59. Li, L., X. Deng, E.T. Mee, S. Collot-Teixeira, R. Anderson, S. Schepelmann, P.D. Minor, and E. Delwart, *Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent.* Journal of virological methods, 2015. **213**: p. 139-146.
60. Karlsson, O.E., S. Belák, and F. Granberg, *The effect of preprocessing by sequence-independent, single-primer amplification (SISPA) on metagenomic detection of viruses.* Biosecurity and bioterrorism: biodefense strategy, practice, and science, 2013. **11**(S1): p. S227-S234.
61. Rosseel, T., O. Ozhelvaci, G. Freimanis, and S. Van Borm, *Evaluation of convenient pretreatment protocols for RNA virus metagenomics in serum and tissue samples.* Journal of Virological Methods, 2015. **222**: p. 72-80.
62. Łusiak-Szelachowska, M., B. Weber-Dąbrowska, E. Jończyk-Matysiak, R. Wojciechowska, and A. Górski, *Bacteriophages in the gastrointestinal tract and their implications.* Gut Pathogens, 2017. **9**: p. 44.
63. Carding, S.R., N. Davis, and L. Hoyles, *Review article: the human intestinal virome in health and disease.* Alimentary Pharmacology & Therapeutics, 2017. **46**(9): p. 800-815.
64. Marzano, S.-Y.L., B.D. Nelson, O. Ajayi-Oyetunde, C.A. Bradley, T.J. Hughes, G.L. Hartman, D.M. Eastburn, and L.L. Domier, *Identification of diverse mycoviruses through metatranscriptomics characterization of the viromes of five major fungal plant pathogens.* Journal of virology, 2016. **90**(15): p. 6846-6863.
65. Cheng, Y., S. Fox, D. Pemberton, C. Hogg, A.T. Papenfuss, and K. Belov, *The Tasmanian devil microbiome—implications for conservation and management.* Microbiome, 2015. **3**(1): p. 1.
66. Wise, P., D. Lee, S. Peck, J. Clarke, S. Thalmann, J. Hockley, D. Schaap, and D. Pemberton, *The conservation introduction of Tasmanian devils to Maria Island National Park: a response to devil facial tumor disease (DFTD).* Global Re-introduction Perspectives: 2016. Case-studies from around the globe, 2016: p. 166.
67. Cheng, Y., K. Heasman, S. Peck, E. Peel, R.M. Gooley, A.T. Papenfuss, C.J. Hogg, and K. Belov, *Significant decline in anticancer immune capacity during puberty in the Tasmanian devil.* Scientific Reports, 2017. **7**: p. 44716.

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

68. Kreiss, A., C. Tovar, D. Obendorf, K. Dun, and G. Woods, *A murine xenograft model for a transmissible cancer in Tasmanian devils.* Veterinary pathology, 2011. **48**(2): p. 475-481.
69. Wiedinger, K., C. Bitsaktsis, and S. Chang, *Reactivation of human polyomaviruses in immunocompromised states.* Journal of neurovirology, 2014. **20**(1): p. 1-8.
70. Li, L., X. Deng, P. Linsuwanon, D. Bangsberg, M.B. Bwana, P. Hunt, J.N. Martin, S.G. Deeks, and E. Delwart, *AIDS Alters the Commensal Plasma Virome.* Journal of Virology, 2013. **87**(19): p. 10912.
71. Zhang, W., S. Yang, T. Shan, R. Hou, Z. Liu, W. Li, L. Guo, Y. Wang, P. Chen, X. Wang, F. Feng, H. Wang, C. Chen, Q. Shen, C. Zhou, X. Hua, L. Cui, X. Deng, Z. Zhang, D. Qi, and E. Delwart, *Virome comparisons in wild-diseased and healthy captive giant pandas.* Microbiome, 2017. **5**(1): p. 90.
72. Cheng, Y., C. Sanderson, M. Jones, and K. Belov, *Low MHC class II diversity in the Tasmanian devil (Sarcophilus harrisii).* Immunogenetics, 2012. **64**(7): p. 525-533.
73. Morris, K., J.J. Austin, and K. Belov, *Low major histocompatibility complex diversity in the Tasmanian devil predates European settlement and may explain susceptibility to disease epidemics.* Biology letters, 2013. **9**(1): p. 20120900.
74. Miller, W., V.M. Hayes, A. Ratan, D.C. Petersen, N.E. Wittekindt, J. Miller, B. Walenz, J. Knight, J. Qi, and F. Zhao, *Genetic diversity and population structure of the endangered marsupial Sarcophilus harrisii (Tasmanian devil).* Proceedings of the National Academy of Sciences, 2011. **108**(30): p. 12348-12353.
75. Willi, Y., J. Van Buskirk, and A.A. Hoffmann, *Limits to the adaptive potential of small populations.* Annu. Rev. Ecol. Evol. Syst., 2006. **37**: p. 433-458.
76. Pearman, P.B. and T.W. Garner, *Susceptibility of Italian agile frog populations to an emerging strain of Ranavirus parallels population genetic diversity.* Ecology Letters, 2005. **8**(4): p. 401-408.
77. O'Brien, S.J., M.E. Roelke, L. Marker, A. Newman, C. Winkler, D. Meltzer, L. Colly, J. Evermann, M. Bush, and D.E. Wildt, *Genetic basis for species vulnerability in the cheetah.* Science, 1985. **227**(4693): p. 1428-1434.
78. Van Ranst, M., A. Fuse, P. Fiten, E. Beuken, H. Pfister, R.D. Burk, and G. Opdenakker, *Human papillomavirus type 13 and pygmy chimpanzee papillomavirus type 1: comparison of the genome organizations.* Virology, 1992. **190**(2): p. 587-596.
79. Tachezy, R., G. Duson, A. Rector, A.B. Jenson, J.P. Sundberg, and M. Van Ranst, *Cloning and genomic characterization of Felis domesticus papillomavirus type 1.* Virology, 2002. **301**(2): p. 313-321.

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

Figure 1. Map of Tasmania, Australia showing the four wild sampling sites**.**

Figure 2. Overview of the Tasmanian devil faecal virome characterized by meta-transcriptomics and VLP metagenomics. Sequencing library/sampling site are represented at the bottom of the bar charts by SH for Stony Head, BR for Buckbys Road, wMW for wukalina/Mt William National Park, MI for Maria Island and Zoo A and B for the two captive populations. **a)** Proportions of all sequence reads in meta-transcriptomics (top) and VLP metagenomics (bottom) showing the proportions of reads belong to bacteria, eukaryote, archaea, viruses, host and unidentified. **b)** Proportions of RNA and DNA viruses detected by meta-transcriptomics (top) and VLP metagenomics (bottom). **c)** Virome composition and the proportions of viral groups in meta-transcriptomics (top) and VLP metagenomics (bottom). **d)** Estimated counts as calculated by RSEM (all six sequencing libraries from both VLP metagenomics and meta-transcriptomics combined and log transformed) to a selection of viruses, showing the differences in viruses detected by VLP metagenomics (blue) and meta-transcriptomics (orange).

Figure 3. Maximum likelihood phylogenetic trees of viruses detected that were previously identified in other mammalian hosts. **a)** phylogenetic analysis of representative RHDV strains based on the 1414 nt nucleotide sequence of major capsid protein. **b)** phylogenetic analysis of representative RHDV strains based on the 5890 nt nucleotide sequence of the non-structural protein. **c)** phylogenetic analysis based on the 4762 aa spike protein of the novel torovirus strain detected here with other previously identified toroviruses. All trees are mid-point rooted and scaled to either the number of amino acid substitution or nucleotide substitutions per site based on the nature of alignment. The viral sequences detected in this study are shown in red in each tree.

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

914 **Figure 4**. Phylogenetic analyses and genomic structures of the RNA viruses identified in the
915 faeces of Tasmanian devils. All phylogenetic analyses were performed based on the amino acid
916 sequence of the RdRp. **a)** Devil sapelovirus, **b)** Devil astrovirus, **c)** Devil rotavirus 1 and 2 and
917 **d)** Devil picobirnavirus 1-6. For Devil astrovirus where the whole genome sequence was
918 obtained, the genomic structure was also included and are shown below the corresponding
919 phylogenetic tree. Predicted ORFs of these genomes are labelled with information of the
920 potential protein or protein domain they encode. All trees are mid-point rooted and scaled to the
921 number amino acid substitutions per site. The newly discovered viruses are shown in red in
922 each phylogenetic tree.

923 **Figure 5**. Phylogenetic analyses and genomic structures of the DNA viruses identified in the
924 faeces of Tasmanian devils. **a)** Devil bocavirus 1 and 2 based on the amino acid sequences of
925 the NS1 protein. **b)** Devil papillomavirus 1 and 2 based on the amino acids of the E1 protein. **c)**
926 Devil polyomavirus 1 and 2 based on the amino acid sequences of the LTAg and VP1 proteins.
927 For devil polyomavirus 1, the whole genome sequence was obtained and the genomic structure
928 is shown next to the VP1 tree. Predicted ORFs of these genomes are labelled with information
929 of the potential protein or protein domain they encode. **d)** Dasyurid herpesvirus 3 based on the
930 amino acid sequence of the DNA polymerase. **e)** Devil circovirus 1 based on the amino acid
931 sequence of the replicase protein. All trees are mid-point rooted and scaled to the number of
932 amino acid substitutions per site. The newly discovered viruses are shown in red.

933 **Figure 6**. Heatmaps (left) showing the hierarchical clustering and percentage of reads from
934 each library mapping to various viral groups, and principal coordinate analysis (PCoA) plots
935 (right) showing the similarity relations among libraries based on Euclidean distances as seen in
936 **a)** meta-transcriptomics and **b)** VLP metagenomics.

937

bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Additional files

Additional file 1: Table S1 List of vertebrate viruses identified by meta-transcriptomics and VLP metagenomics, including information on the sequence library, identification, classification, virus and strain name, sample type, genome length and percentage reads, Blastx hits to known viruses. (.xls)

Additional file 2: Table S2. Virus abundance table showing the proportion of reads mapped to viral contigs in each meta-transcriptomic library. (.xls)

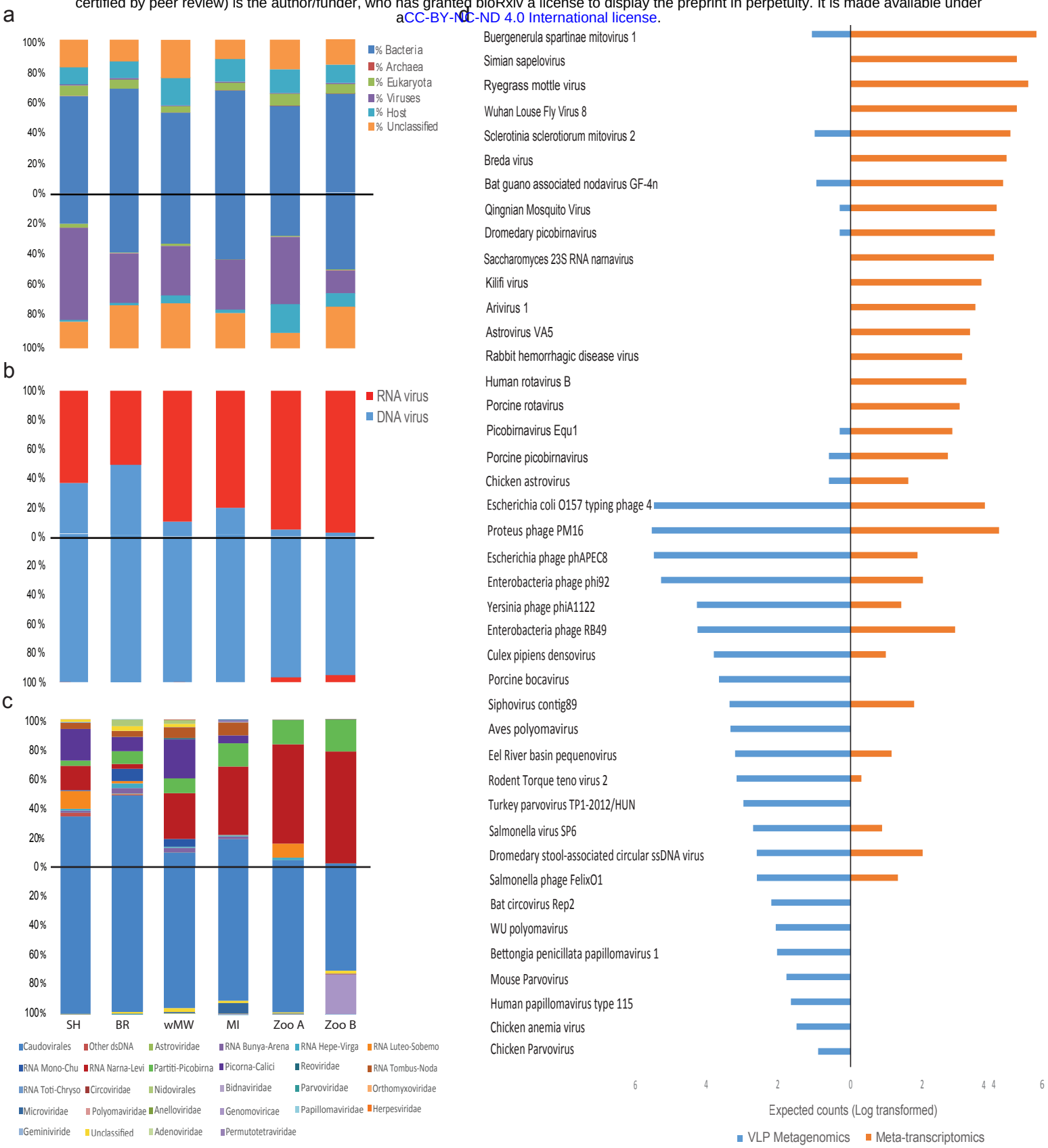
Additional file 3: Table S3. Virus abundance table showing the proportion of reads mapped to viral contigs in each VLP metagenomics library. (.xls)

Additional file 4: Table S4. List of primer sets used for PCR confirmation of RHDV2. (.xls)

Additional file 5: Figure S1. Maximum likelihood phylogenetic tree of Dasyurid herpesvirus 3 in the context of representatives of the *Herpesviridae* family based on the amino acid sequences of glycoprotein M. The tree is mid-point rooted and scaled to the number of amino acid substitutions per site. The newly discovered Dasyurid herpesvirus 3 is shown in red. (.pdf)

Additional file 6: Figure S2. Maximum likelihood phylogenetic tree of Dasyurid herpesvirus 3 in the context of representatives of the *Herpesviridae* family based on the amino acid sequences of major capsid. The tree is mid-point rooted and scaled to the number of amino acid substitutions per site. The newly discovered DaHV-3 is shown in red. (.pdf)

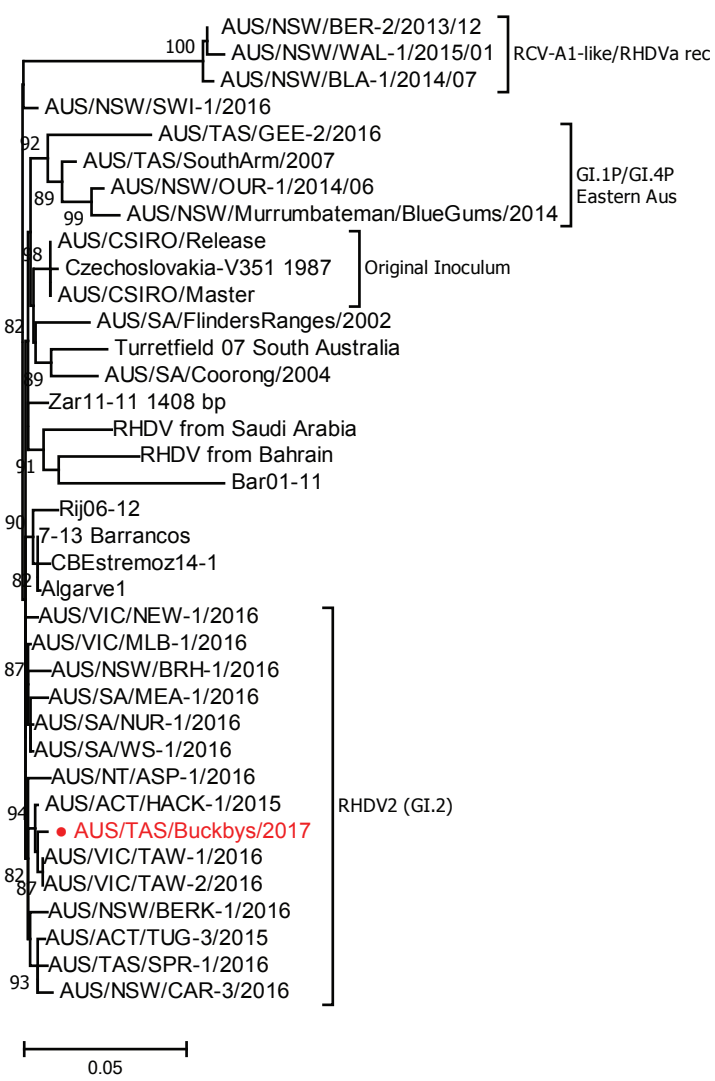
bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

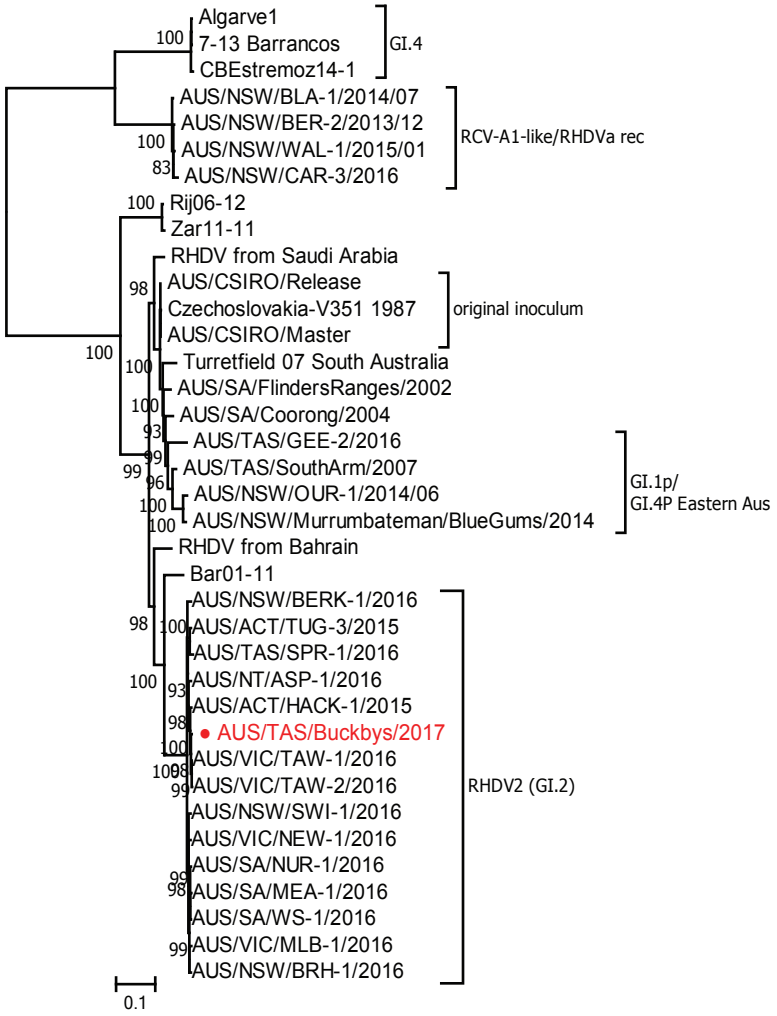
a

Caliciviridae: RHDV (Major capsid)



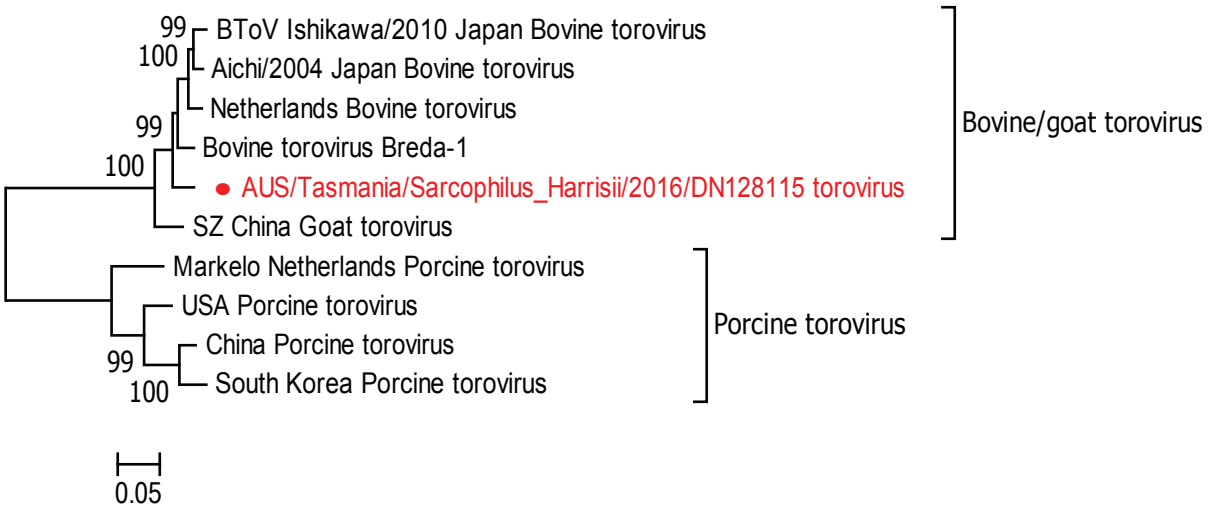
b

Caliciviridae: RHDV (Non-structural gene)

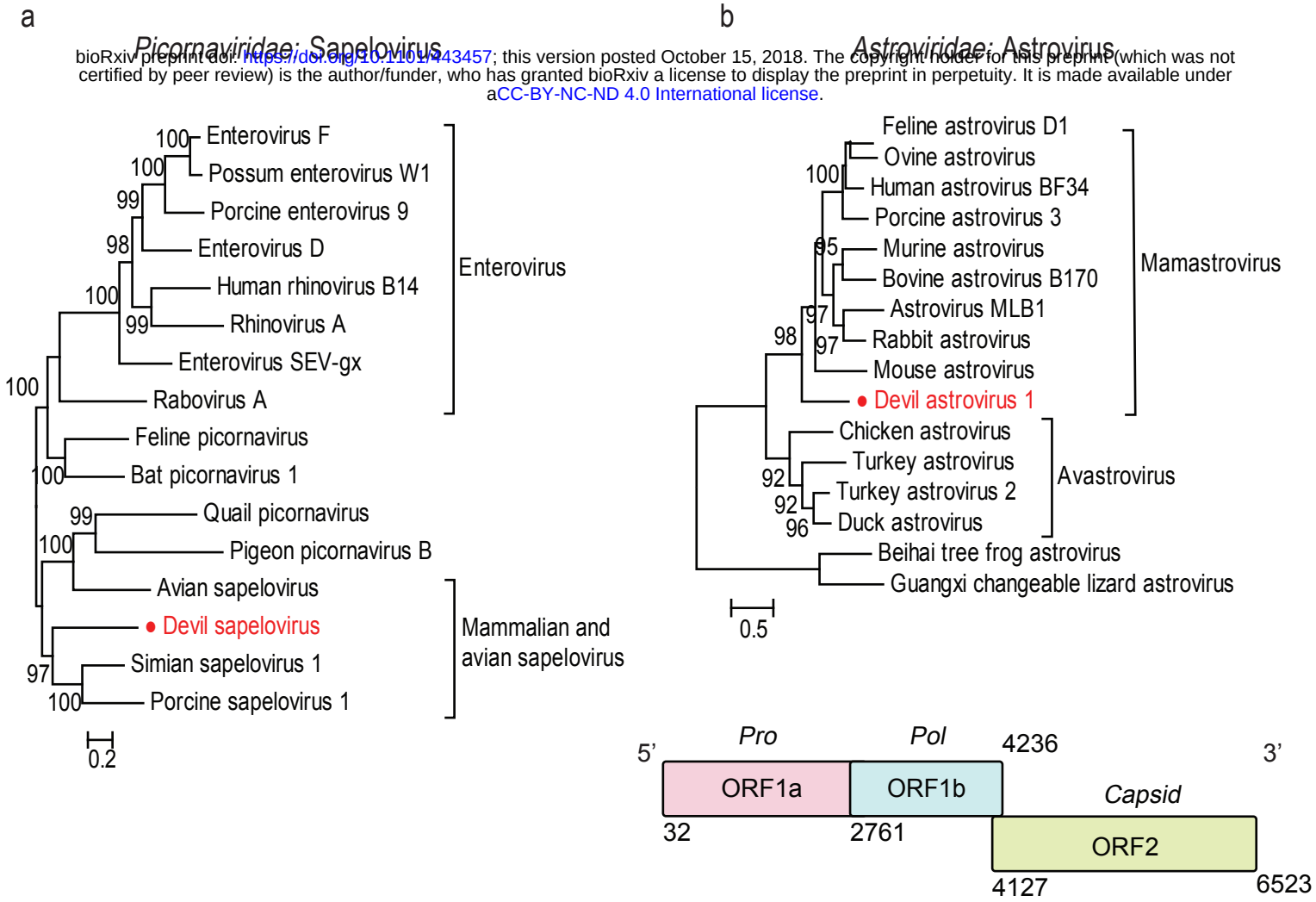


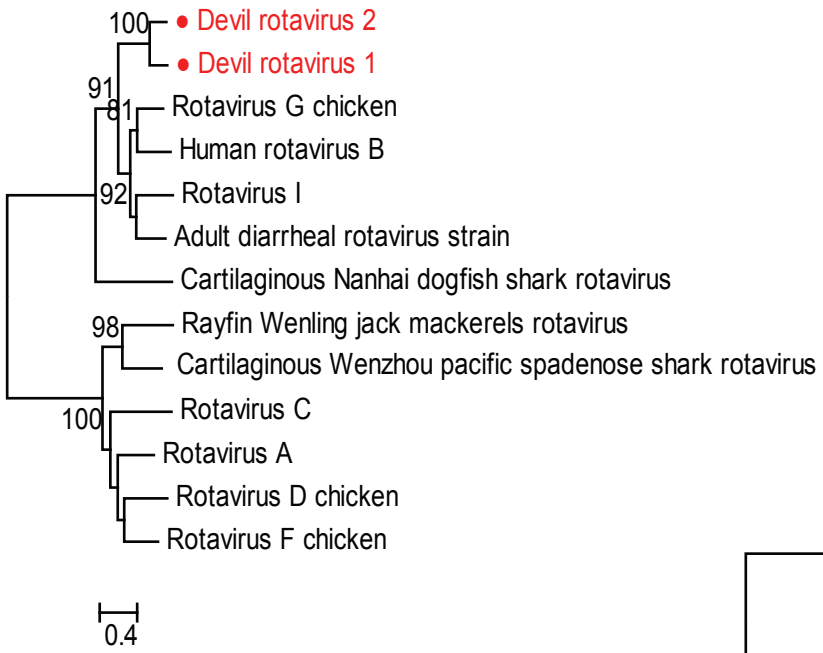
c

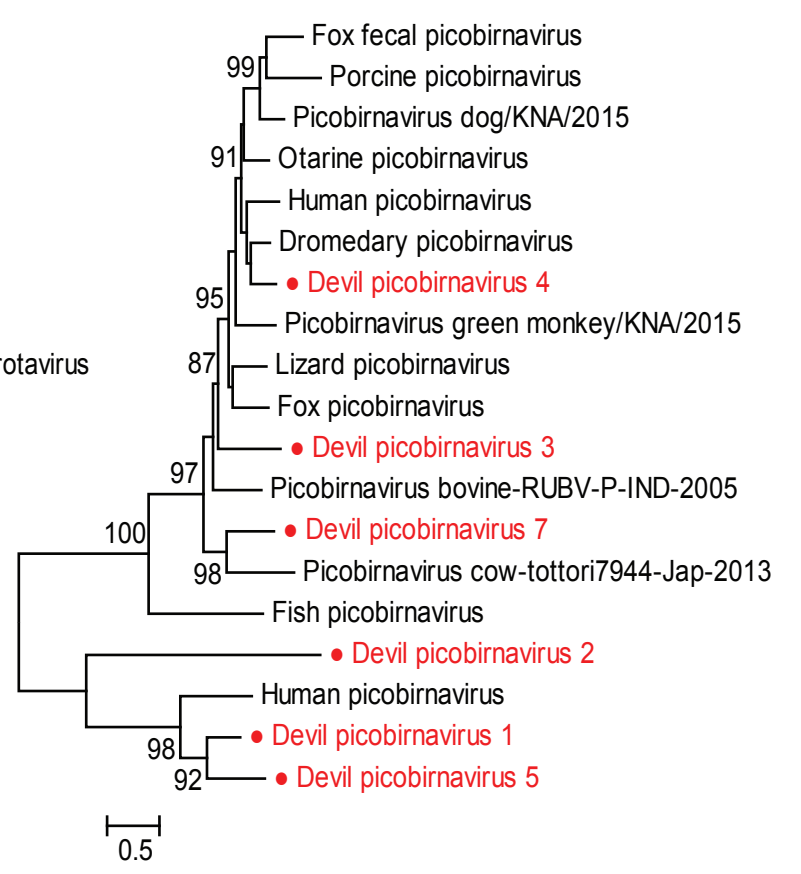
Coronaviridae: Torovirus/Bredavirus



bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



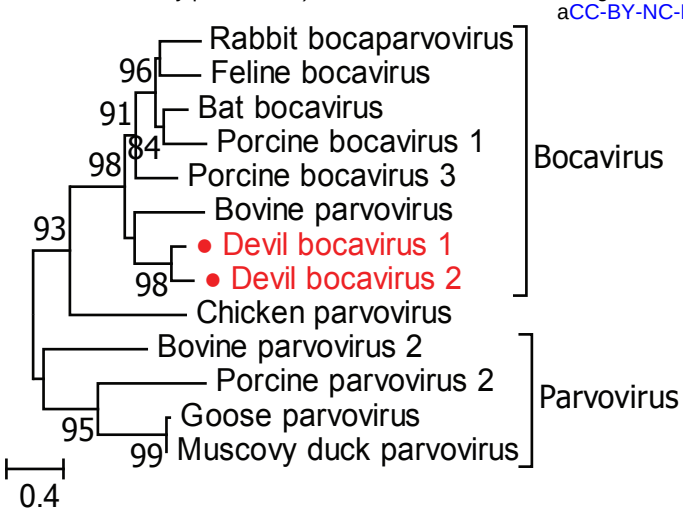




bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

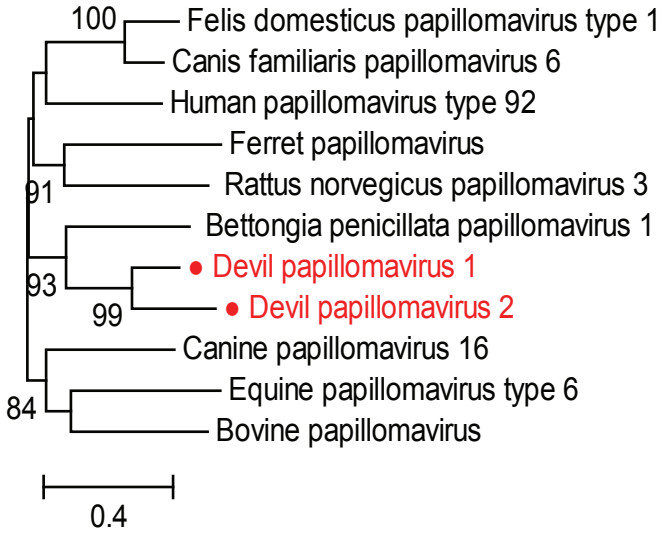
a

Parvoviridae: Bocavirus



b

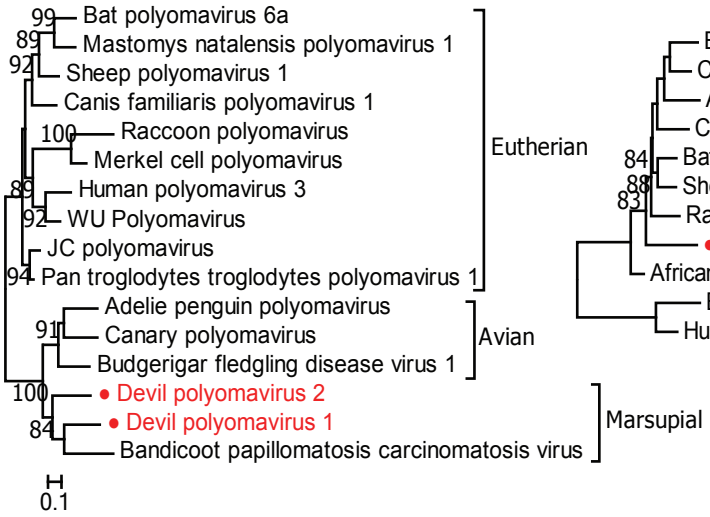
Papillomaviridae: Papillomavirus



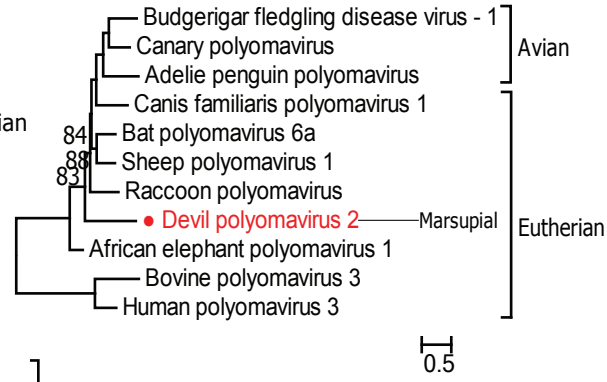
c

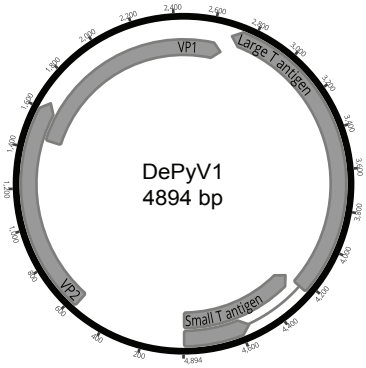
Polyomaviridae: Polyomavirus

LTAg



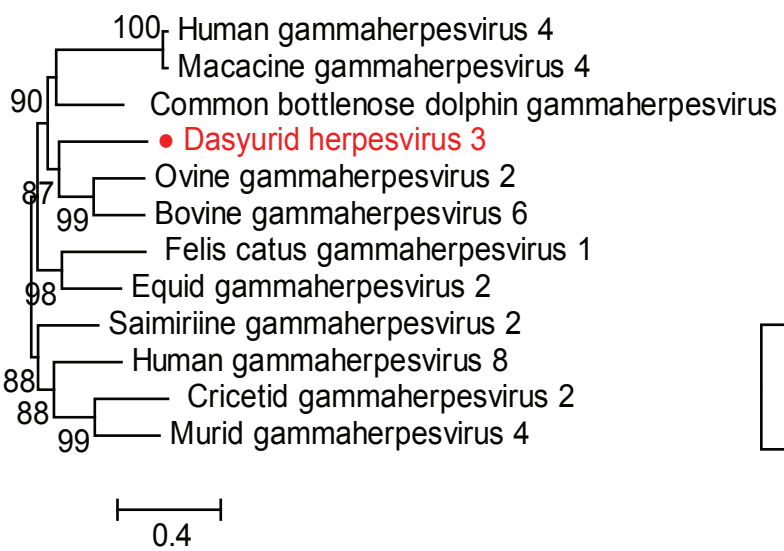
VP1





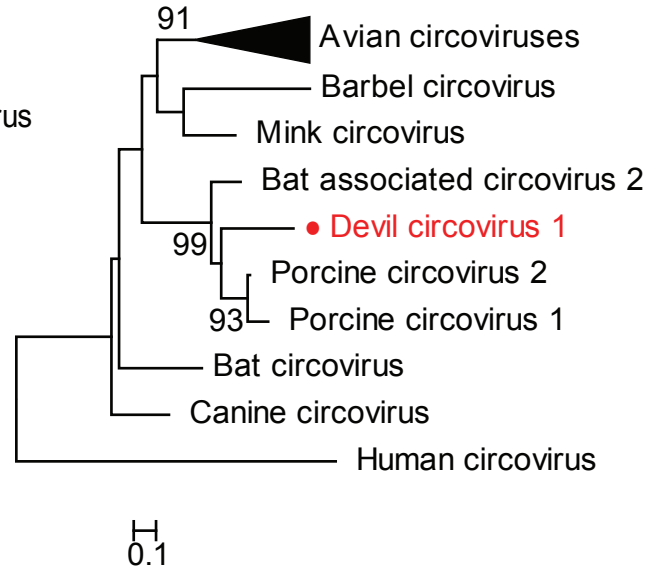
d

Herpesviridae: Gammaherpesvirus



e

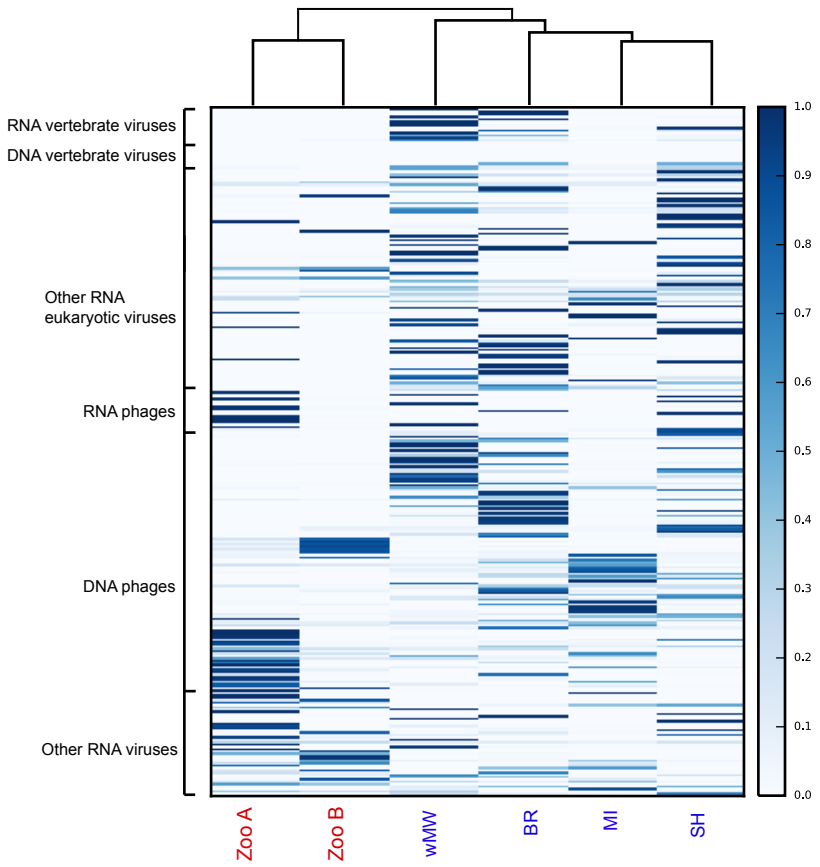
Circoviridae: Circovirus

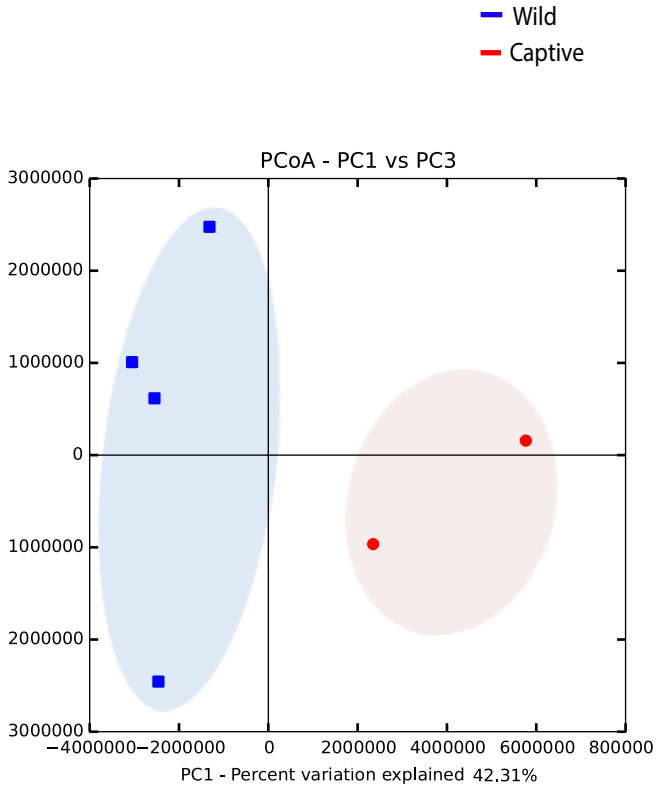


a

Meta-transcriptomics

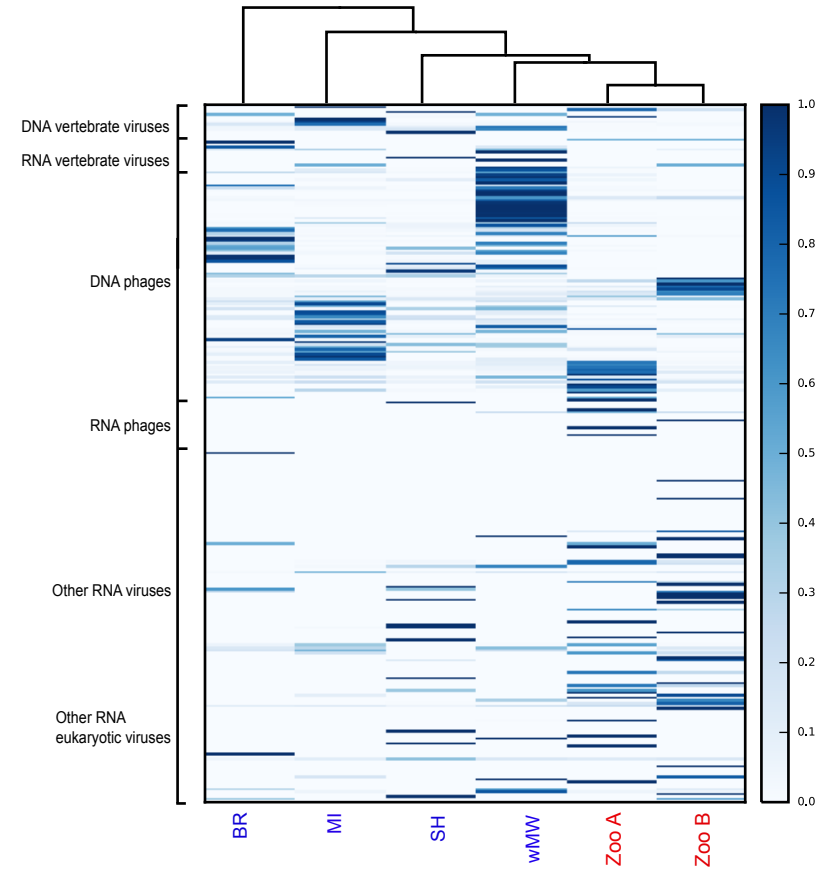
bioRxiv preprint doi: https://doi.org/10.1101/443457; this version posted October 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

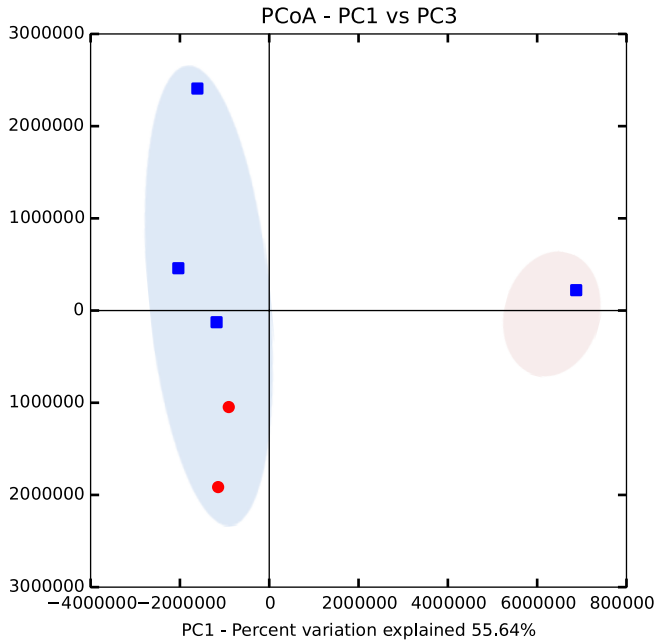




b

VLP Metagenomics






Buckby's Road
wukalina/ Mt William
National Park
Stony Head
Maria Island
N
S
0
100 km