Viral Communities in Vampire Bats: Geographical Variation and Ecological Drivers

Laura Bergner
BSc (Hons), MSc

Submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy

Institute of Biodiversity, Animal Health and Comparative Medicine
College of Medical, Veterinary and Life Sciences
University of Glasgow

May 2018

Abstract

Microbial communities play important roles in organismal and ecosystem health. High throughput sequencing has revolutionized our understanding of host-associated microbial communities, but the viral component of these communities remains poorly characterized relative to microbes such as bacteria, particularly in non-human hosts. This knowledge gap has implications for global health, as viruses originating in wildlife are responsible for recent disease outbreaks in humans and domestic animals. Although studies have identified factors differentiating viral communities between species, we have little understanding of the variability of viral communities within species. Comparative studies of viral communities are therefore necessary to characterize novel taxa and to evaluate the ecological factors influencing intraspecific viral diversity and distribution.

Bats are recognized as "special" reservoirs for viruses because they are associated with diverse viral communities and display deep evolutionary relationships with individual viral taxa. Common vampire bats (*Desmodus rotundus*) represent a particularly interesting system in which to investigate viral communities, as they are obligate blood feeders that interact ecologically with many different host species, providing opportunities for the acquisition of diverse viruses. The overall objective of this thesis was to advance our understanding of intraspecific wildlife-associated viral communities using an established field network of common vampire bat colonies across Peru. Specifically, I developed a novel method for comparative viral community studies, characterized the viral communities of vampire bats, and examined the ecological correlates of vampire bat viral diversity across Peru.

Metagenomic sequencing is a promising technique for comparative studies of viral communities in wildlife, but there is a need to first develop standardized methods that can be applied to samples collected in the field. In Chapter 2 I developed a shotgun metagenomic sequencing approach to characterizing viral communities from non-invasive samples. Specifically, I optimized extraction and sequencing protocols using fecal and oropharyngeal swabs collected from common vampire bats in Peru. Two preliminary sequencing runs were performed, the results of which motivated four pilot studies in which I tested how different storage media, nucleic acid extraction procedures, and enrichment steps affect the viral community detected. Metagenomic sequencing revealed viral contamination of fetal bovine serum, a component of viral transport medium, suggesting that swabs should be stored in RNALater or another non-biological medium. Extraction

and qPCR tests were performed on swabs inoculated with known concentrations of virus, which revealed that nucleic acid should be directly extracted from swabs rather than from supernatant or pelleted material. Metagenomic sequencing of paired samples was used to test enrichment by ribosomal RNA depletion and light DNAse treatment, which both reduced host and bacterial nucleic acid in samples and improved virus detection. A bioinformatic pipeline was developed specifically for processing vampire bat shotgun viral metagenomic data. Finally, the optimized protocol was applied to twelve pooled samples from seven localities in Peru, and read subsampling demonstrated that the viral communities detected were consistent at commonly attained depths of sequencing. The protocol developed in this chapter enables minimally biased comparative viral community studies in non-invasive samples collected from wildlife.

Having a detailed understanding of viral diversity in key wildlife hosts is an important first step in evaluating the risk of zoonotic disease emergence, but we still lack a holistic view of viral communities in many species including vampire bats. In Chapter 3, I used the metagenomic sequencing protocol developed in Chapter 2 to thoroughly characterize viral communities in the saliva and feces of vampire bats captured across Peru. Viruses were detected from a range of natural host groups including vertebrate-associated taxa that were potentially infecting vampire bats, bacteriophages associated with gut bacteria, and plant- or insect-infecting viruses potentially acquired from the environment. There were broad differences between fecal and saliva viral communities, showing evidence of body habitat compartmentalization. Eight vertebrate-infecting viral families were selected for phylogenetic analysis to evaluate relationships with previously characterized viral taxa from bats. Novel findings included a *Hepatitis delta*-like virus in vampire bat saliva samples from three sites, representing the first detection of this virus outside of humans. Full genomes were generated for novel viruses in families that contain zoonotic taxa such as *Coronaviridae, Hepeviridae* and *Reoviridae*. Finally, widespread viral families such as *Picornaviridae* and *Adenoviridae* were identified as potential markers of vampire bat movement. Overall, these results established that vampire bat viral communities differ between body habitats and suggested that, for the vertebrate-infecting families analyzed, novel viruses mostly fall within bat-specific clades, without evidence of livestock or humans acting as a major source of viral diversity in vampire bats.

Interspecific differences in ecological and life history traits are known to impact viral richness in bats, but the factors structuring viral communities within bat species are less well understood. In Chapter 4, I examined the spatial, demographic and environmental

correlates of intraspecific viral diversity in vampire bats. Three measures of viral diversity were calculated at the colony level: richness, a novel measure of taxonomic diversity, and community composition. Generalized linear models were then used to test the effects of broad scale and local ecological variables on saliva and fecal viral diversity. Differences in saliva viral richness were positively correlated with geographic distance, and there was an association between longitude and viral richness and community composition, both of which indicated the importance of location in shaping saliva viral communities. The results also suggested the northwest region of Peru as a hotspot of vampire bat saliva viral diversity. Fecal viral communities broadly differed between ecoregions, with the Amazon exhibiting higher richness and distinct community composition, and differences in fecal community composition were positively related to geographic and genetic distance between colonies. In addition to the broad scale and spatial patterns of diversity, fecal viral richness also increased with the proportion of juveniles in a colony, and there was an effect of environmental context encompassing elevation, climate, and in some cases local livestock density. These results show for the first time that ecological variables can influence intraspecific viral diversity.

In summary, the work presented in this thesis advances our understanding of wildlife-associated viral communities in an ecologically important bat host. Future directions in comparative wildlife viral metagenomics, as discussed in Chapter 5, will include exploring the determinants of viral communities across host species, environments and time.

Table of Contents

List of Tables

List of Figures

Acknowledgements

The work presented in this thesis was a true collaborative effort so there are many people to thank. I would first like to acknowledge my amazing team of supervisors: Daniel Streicker, Richard Orton and Roman Biek. Thanks to Daniel for giving me the opportunity to work on this exciting project, constantly encouraging me out of my scientific comfort zone, and helping me to grow as a researcher over the past years. Thanks to Richard for being the bioinformatics guru of my PhD and for offering patient advice during the many hours spent sorting through confusing data. I am grateful to Roman for providing helpful perspective and for guiding this work in interesting directions. I would like to acknowledge the Wellcome Trust and Royal Society for providing the funding for this project (Sir Henry Dale Fellowship to DS; 102507/Z/13/Z).

I am grateful to my assessor Sarah Cleaveland for her positive presence throughout my PhD, and for the always constructive feedback during our yearly discussions. Thanks to the other annual review assessors Brian Shiels, Shaun Killen, and David Eckersall for providing helpful advice along the way.

Thanks to the postdoc rock stars who informally advised and encouraged me over the past few years: Julio Benavides, Maude Jacquot, Andrew Shaw, Kath Allan, and Sema Nickbakhsh. Special thanks to Julio for making statistics less scary, to Maude for bioinformatics and phylogenetics advice, and to Andrew for helping rescue my extraction protocol when it seemed like all was lost.

I would also like to thank Carlos Tello, Jorge Carrera, Anthony Almeyda and Francisco Espinoza who welcomed me into the bat team in Peru for several months; I'll never forget the sleepless nights catching vampire bats under the starry Andean sky. Special thanks to Carlos and the other members of the field crew who have tirelessly continued to collect samples for the project long since my return to Glasgow. I am also grateful to John Claxton for being the wizard of project logistics including sample transport from Peru.

I would like to acknowledge collaborators at the Universidad Peruano Cayetano-Heredia and other collaborators in Peru including Armando Hung, Nestor Falcon, Carlos Shiva, Patricia Mendoza, Ornela Inagaki and Sergio Recuenco for assistance with sample storage and research permits in Peru.

Huge thanks to Ana da Silva Filipe for helping me develop the extraction and sequencing protocols, and whose curiosity and positive attitude in the face of challenging sequencing projects is inspiring. I'm grateful to members of the CVR NGS facility for training me in the lab and helping troubleshoot library preps: Gavin Wilkie, Lily Tong, Maggie Baird, and Daniel Mair. Special thanks to Chris Davis for stepping in to train me in the low input library preps. Thanks also to Mariana Varela for contributing samples and reagents for the qPCR experiments. I'm grateful to Sarah McDonald in the Biobank and Carol Leitch in the PCR pipeline for their heroic efforts in keeping the labs running, as well as to the administrative staff of both IBAHCM and CVR for all their help over the years.

Thanks to Alice Broos for keeping everything functioning in the lab and for being the only person who is always excited to talk about PCR with me. Dan Becker's input was invaluable throughout this project from the contribution of samples to statistical advice. I'm grateful to Hannah Frank for the advice that inspired the final extraction protocol. Thanks to Danielle Dekoski for allowing me to include her work on vampire bat colony size and to Jamie Winternitz for her previous work and analysis advice on the

Author's Declaration

The material presented in this thesis is the result of research conducted between October 2014 and May 2018. Throughout this time, I was under the supervision of Daniel Streicker, Richard Orton and Roman Biek. This work has not been submitted as part of any other degree and is based on my own work unless otherwise stated. Where others have contributed to the work, this is acknowledged in the text, with key collaborators in metagenomic sample collection and processing acknowledged in Figure 1.2.

Abbreviations

3D polymerase	*Picornaviridae* RNA-dependent RNA polymerase
16S	Ribosomal RNA subunit
AdV	Adenovirus
AIC	Akaike information criterion
AIC$_c$	Akaike information criterion corrected for sample size
A_R	Allelic richness
AAC	Ayacucho-Apurímac-Cusco region, Peru
AMA	Amazonas Department, Peru
API	Apurímac Department, Peru
AYA	Ayacucho Department, Peru
AVE	Qiagen Elution Buffer
AW	Qiagen Wash Buffer
BIC	Bayesian information criterion
bp	Base pair
BVDV	Bovine viral diarrhea virus
CAJ	Cajamarca Department, Peru
cDNA	Complementary DNA
CL	Containment Level
CoV	Coronavirus
Ct	Cycle threshold in qPCR
CUS	Cusco Department, Peru
CVR	Centre for Virus Research
D	F_{ST} analogue
DAPC	Discriminant analysis of principal components
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
dNTP	Deoxynucleotide Triphosphate
DPBS	Dulbecco's phosphate buffered saline
Dr-AdV1	*Desmodus rotundus* Adenovirus 1
Dr-AdV2	*Desmodus rotundus* Adenovirus 2
DrHDV	*Desmodus rotundus* Hepatitis deltavirus
dsDNA	Double-stranded DNA
dsRNA	Double-stranded RNA
e-value	Expect value in sequence database searches
ENA	Excluding null alleles
Env	*Retroviridae* envelope gene

FBS	Fetal bovine serum
$\boldsymbol{F_{IS}}$	Inbreeding coefficient
$\boldsymbol{F_{ST}}$	Fixation index
FV	Foamy virus
$\boldsymbol{G''_{ST}}$	F_{ST} analogue
GLM	Generalized linear model
GLMM	Generalized linear mixed model
GTR	General time reversible
HBV	Hepatitis B virus
HDV	Hepatitis deltavirus
$\boldsymbol{H_E}$	Expected heterozygosity
HEV	Hepatitis E virus
$\boldsymbol{H_O}$	Observed heterozygosity
HS	High Sensitivity
HUA	Huánuco Department, Peru
HWE	Hardy-Weinberg equilibrium
ICTV	International Committee on Taxonomy of Viruses
kbp	Kilobase pair
km	Kilometer
k-mer	Motif of length k in a sequence
LCA	Lowest common ancestor
LMA	Lima Department, Peru
LR	Loreto Department, Peru
LRT	Likelihood ratio test
MgCl$_2$	Magnesium chloride
mL	Milliliter
mM	Millimolar
mm	Millimeter
mRNA	Messenger RNA
$\boldsymbol{N_A}$	Number of alleles
NaCl	Sodium chloride
NaN$_3$	Sodium azide
$\boldsymbol{N_c}$	Census population size
NCBI	National Center for Biotechnology Information
ng	Nanogram
Nr	Non-redundant
NSP	Non-structural protein
Nt	Nucleotide

ORF	Open reading frame
PC	Principal component
PCA	Principal component analysis
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pg	Picogram
PFU	Plaque forming units
PicoV	Picornavirus
pmol	Picomole
Pol	*Retroviridae* DNA polymerase gene
qPCR	Quantitative PCR
RdRp	RNA-dependent RNA polymerase
RFFIT	Rapid fluorescent focus inhibition test
RLT	Qiagen Lysis Buffer
RNA	Ribonucleic acid
RNAse	Ribonuclease
RPE	Qiagen Wash Buffer
rRNA	Ribosomal ribonucleic acid
RV	Rotavirus
SBV	Schmallenberg virus
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
TD	Viral taxonomic diversity
U	Enzyme Unit
VBRV	Vampire bat rabies virus
VP	Viral protein
VTM	Viral transport medium
μL	Microliter
μM	Micromolar

1 Introduction

1.1 Microbial community diversity

Microbial communities play essential roles in ecosystem function (Suttle 2007; van der Heijden *et al.* 2008; Strom 2008; Strickland *et al.* 2009) and impact the health of organisms through their roles as either mutualists or pathogens (Costello *et al.* 2012; Virgin 2014; Manrique *et al.* 2016), therefore it is important to understand the factors shaping their diversity and distribution. Microbial communities exhibit extreme diversity (Breitbart & Rohwer 2005; Fierer *et al.* 2007) with spatial variation thought to be shaped by both contemporary environmental and historical processes (Martiny *et al.* 2006; Lindström & Langenheder 2011). The diversity of host-associated microbial communities must be considered on different hierarchical levels as communities exhibit variation between body habitats within a host (Costello *et al.* 2009), between individuals within a species (The Human Microbiome Project Consortium 2012; Wasimuddin *et al.* 2017), between species within an environment (McCord *et al.* 2013; Menke *et al.* 2014) and between environments (Sullam *et al.* 2012; Linnenbrink *et al.* 2013; Muletz Wolz *et al.* 2017). The factors involved in shaping this variation also depend on the scale being considered. For example, differences in host-associated bacterial communities between species are influenced by phylogeny and host diet (Ley *et al.* 2008; Muegge *et al.* 2011; Phillips *et al.* 2012), while differences in communities within a species may be influenced by factors such as geographic location (Linnenbrink *et al.* 2013; Kueneman *et al.* 2013) and host sex (Bolnick *et al.* 2014; Menke *et al.* 2017).

In addition to the importance of host dynamics in the structure of their associated microbial communities, important insights have come from applying ecological theories developed at the macro-organismal level to micro-organisms (Prosser *et al.* 2007; Christian *et al.* 2015; Dudaniec & Tesson 2016). For example, metacommunity theory has provided a useful framework for generating hypotheses about spatial structure in micro-organisms (Mihaljevic 2012); microbial community composition could depend on host dispersal, differences in host traits, and environmental variation between patches (Costello *et al.* 2012). Interactions among different microbes can also influence community composition (Lozupone *et al.* 2012; Koskella *et al.* 2017). A community-level perspective is critical for understanding both the factors that structure host-associated microbial communities and their roles in health and disease (Costello *et al.* 2012; Vayssier-Taussat *et al.* 2014; Virgin 2014; Christian *et al.* 2015). The ability to simultaneously characterize all members of

bacterial communities and test hypotheses about the factors that structure them has recently become possible due to advances in sequencing technology (Costello *et al.* 2009; Sullam *et al.* 2012; The Human Microbiome Project Consortium 2012). However, the viral component of host-associated microbial communities remains relatively poorly characterized (Stulberg 2016), despite the critical roles played by viruses in ecosystems and within hosts (Suttle 2007; Virgin 2014; Manrique *et al.* 2016) and their unique biology which might lead to very different patterns of diversity and distribution compared to other microbes.

1.2 Viral communities

1.2.1 Technical challenges to studying viral communities

Viruses remain understudied relative to bacteria and other micro-organisms because they lack a conserved genetic marker (Rohwer & Edwards 2002; Mokili *et al.* 2012), which has prohibited the use of a metabarcoding approach such as the 16S ribosomal RNA (rRNA) gene that has been used to examine bacterial communities across highly multiplexed samples (Creer *et al.* 2016). For this reason, viral communities have primarily been studied using untargeted shotgun metagenomic sequencing, which is the random sequencing of the genomic fragments of an entire sample, and in the case of viruses including additional steps to deplete non-viral material (Hall *et al.* 2014; Kleiner *et al.* 2015; Kohl *et al.* 2015; Conceição-Neto *et al.* 2015). As well as being relatively unbiased, shotgun sequencing has other advantages such as more precise classification of taxa and the ability to assess functional traits of the microbial community (Jovel *et al.* 2016). However, there are technical challenges specific to viral metagenomics that need to be addressed before broad comparative studies equivalent to those in other microbial systems are possible.

In contrast to the human-associated viral communities which have been studied extensively (Breitbart *et al.* 2003; Breitbart & Rohwer 2005; Pride *et al.* 2011; Reyes *et al.* 2012; Wylie *et al.* 2014; Hannigan *et al.* 2015), characterizing viral communities from a wide range of hosts likely requires sampling in resource-limited environments. The storage and transportation of field-collected samples is an important consideration in viral community studies, as RNA viruses are highly sensitive to degradation in sub-optimal temperature and storage conditions (Cardona *et al.* 2012). Diverse methods have been used to store field-collected samples destined for viral discovery in previous studies (e.g. Donaldson *et al.* 2010; Baker *et al.* 2013; Wu *et al.* 2016). Standardized collection procedures that can be

implemented in resource-limited environments will be essential in future studies that aim to compare viral communities across host species, space and time. Establishing these procedures will require testing different storage buffers and extraction methods in order to generate sufficient viral nucleic acid for shotgun sequencing.

Another technical consideration is the relatively small size of viral genomes, which can be swamped out by larger host and bacterial genomes in an unbiased sequencing approach (Nakamura *et al.* 2009; Yang *et al.* 2011), making enrichment an important component of viral community sequencing. Methods of viral enrichment relative to host and bacteria include filtration of host/bacterial particles, density gradient centrifugation, nuclease treatment, and removal of rRNA (Hall *et al.* 2014; Kleiner *et al.* 2015; Kohl *et al.* 2015). Filtration and centrifugation are known to bias the inferred taxonomic composition of samples (Thurber *et al.* 2009; Kleiner *et al.* 2015), and are not ideal for ecological studies in which non-viral sequences may also be of interest. However, nuclease treatment using DNAse (Allander *et al.* 2001) and depletion of host rRNA (He *et al.* 2010; Matranga *et al.* 2014) are two effective enrichment methods that are less likely to bias the viral community. Although DNAse could cause bias towards RNA viruses, previous studies including a DNAse treatment step have also detected DNA viruses (Baker *et al.* 2013; Hall *et al.* 2014). A challenge of comparative non-invasive viral metagenomics will be identifying a combination of field and laboratory methods that maximize the proportion of viral reads while minimizing bias and preserving information about other taxa of interest.

The bioinformatic challenges of analyzing viral metagenomic data are also significant, as typical studies contain many millions of reads per sample, the majority of which belong to host or bacterial genomes (Kunin *et al.* 2008; Fancello *et al.* 2012; Soueidan *et al.* 2015). Taxonomic assignment is typically performed using reference-based methods, in which reads are compared to a database of sequences, although reference-independent methods exist in which metagenomic samples are directly compared to one another (Dutilh *et al.* 2012). Reference-based methods may not be as useful for characterizing novel viruses, as existing databases are incomplete and viruses are characterized by high genetic diversity and fast evolutionary rates (Edwards & Rohwer 2005; Fancello *et al.* 2012), but are useful for identifying well-characterized viruses such as pathogens of concern to human health. There is therefore a need to develop standardized, yet flexible, bioinformatic approaches for comparing viral communities from non-human taxa.

1.2.2 Characterizing viral communities

Despite the challenges in generating and interpreting data, metagenomic studies have already revolutionized our understanding of viral communities in humans and the environment. Given the prominent roles played by viruses in health and disease, many viral metagenomic studies to date have focused on humans (Breitbart *et al.* 2003; Breitbart & Rohwer 2005; Pride *et al.* 2011; Reyes *et al.* 2012; Wylie *et al.* 2014; Hannigan *et al.* 2015). In addition to being compartmentalized by body habitat (Paez-Espino *et al.* 2016), viral communities in healthy humans are individually distinct and stable over time (Abeles *et al.* 2014; Wylie *et al.* 2014), with shared living environment increasing bacteriophage community similarity (Robles-Sikisaka *et al.* 2013). Disease has also been linked to changes in the viral community within humans (Ly *et al.* 2014; Norman *et al.* 2015). Looking outwards into the environment has revealed a stunning diversity of viral taxa (Edwards & Rohwer 2005; López-Bueno *et al.* 2009; Roux *et al.* 2012), including unique viral communities in extreme environments such as the Antarctic (Adriaenssens *et al.* 2017), perennial desert ponds (Fancello *et al.* 2013), and hypersaline ponds (Roux *et al.* 2016a). Studies of viral communities across a wide taxonomic range of invertebrates (Shi *et al.* 2016) and vertebrates (Shi *et al.* 2018) have shed light on viral diversity in an evolutionary context, but questions remain about the extent to which viral communities vary across populations and individuals within non-human hosts.

Viral metagenomic studies are beginning to realize comparisons on a shallower evolutionary scale. For example, Wu *et al.* (2016) examined pooled samples from 4,400 bats from China across 29 provinces and 40 species, describing broad patterns of diversity and characterizing novel viruses in a phylogenetic context. Other descriptive studies of viral communities within non-human hosts have discovered a variety of novel taxa (Li *et al.* 2011; Bodewes *et al.* 2013; Sasaki *et al.* 2015; Amimo *et al.* 2016; Conceição-Neto *et al.* 2017; Yinda *et al.* 2018), with some studies beginning to incorporate individual information about hosts, such as disease status in pandas (Zhang *et al.* 2017). The accumulation of in-depth studies of viral diversity will undoubtedly continue to reveal novel taxa and will eventually allow the establishment of general principles about the variation and organization of viral communities within and between diverse hosts.

1.2.3 Drivers of viral communities

Describing the variation in viral diversity across space and connecting differences in viral richness and community composition with environmental and host-specific traits are exciting opportunities offered by metagenomics. Previous studies testing the factors that structure viral communities have often focused on large ecological scales, for example, establishing that viral communities vary across biomes (Dinsdale *et al.* 2008; Fierer *et al.* 2012; Hurwitz *et al.* 2014). However, local environmental factors that vary over smaller scales, such as soil pH and elevation, can also play a role in viral community structure (Adriaenssens *et al.* 2017). In addition to furthering our knowledge of viral ecology in the environment, understanding the ecological drivers of viral communities has implications for understanding and predicting disease emergence from wildlife into humans and domestic animals.

Predicting viral emergence from wildlife is a complex field of study with relevance to global health (Karesh *et al.* 2012; Hassell *et al.* 2017; Kelly *et al.* 2017), as the majority of emerging infectious diseases are zoonotic and 25 - 44% are estimated to be viral (Taylor *et al.* 2001; Jones *et al.* 2008). All else being equal, a reservoir host with higher viral richness overall is likely to harbor more viruses that are able to infect other species (Morse 1993), but landscape and local factors are also important considerations in anticipating disease emergence (Hassell *et al.* 2017). Examining spatial patterns of viral diversity and ecological correlates of viral richness and community composition represent early steps in predicting viral emergence. For example, Anthony *et al.* (2015) tested whether viral communities in macaques were assembled through predictable or stochastic processes. Their finding of non-random patterns in viral communities across individuals and locations lends support to the idea that we might be able to predict how changing environmental conditions could lead to viral emergence from wildlife into humans (Anthony *et al.* 2015). Another study found that viral richness in mice differed across geographic sites and was positively correlated with mouse weight (Williams *et al.* 2018). However, few studies to date have robustly characterized viral communities within a single host species, and even fewer have identified population or landscape level features that explain intraspecific variation in viral diversity.

1.3 Viral diversity in bats

Though viruses are ubiquitous throughout the natural world, bats (Order Chiroptera) are a particularly interesting host group to focus on. Bats are highly diverse, with 1,242 described species (IUCN 2017) that vary dramatically in their dispersal, aggregation, dietary ecology and degree of association with humans and domestic animals. Moreover, bats are implicated as the likely reservoir host in a number of high-profile viral zoonoses such as Severe Acute Respiratory Syndrome (SARS-CoV), Ebola, and Nipah virus (Li *et al.* 2005; Leroy *et al.* 2005; Pulliam *et al.* 2011). Bats display both deep evolutionary relationships with individual viral families (Cui *et al.* 2007; Drexler *et al.* 2012a; Quan *et al.* 2013; Drexler *et al.* 2013) and associations with diverse communities of viruses (Li *et al.* 2010a; Donaldson *et al.* 2010; Wu *et al.* 2012; He *et al.* 2013a; Baker *et al.* 2013; Dacheux *et al.* 2014; Wu *et al.* 2016; Salmier *et al.* 2017; Hu *et al.* 2017), and as such have been described as "special" viral reservoirs (Calisher *et al.* 2006; Luis *et al.* 2013; Brook & Dobson 2015; Hayman 2016). The high viral diversity found in bats has been hypothesized to result from their unique biology and ecology, with a variety of mechanisms proposed.

One hypothesis links flight in bats with elevated body temperature and metabolic rate, such that bats are able to tolerate viruses without being adversely affected (O'Shea *et al.* 2014). Migration also influences disease dynamics (Altizer *et al.* 2011), and some migratory bat species may encounter different viral communities due to long-distance movements and use of different habitats (Hayman *et al.* 2012). Migratory and non-migratory populations of the same species may provide an opportunity for susceptible individuals to mix with those harboring viruses (Calisher *et al.* 2006). Population structure is another factor potentially allowing viruses to persist in bat populations with periodic outbreaks among spatially separated populations (Turmelle & Olival 2009). Many bats roost colonially, with some species maintaining local population densities of several million bats, providing the opportunity for sustained disease transmission (Calisher *et al.* 2006). Torpor in bats has been shown to influence the transmission dynamics of some bat viruses which are able to persist through overwintering in hosts until a birth pulse (George *et al.* 2011). For example, the evolutionary rate of rabies virus is slower in bat species that live in temperate zones as compared to those living in tropical zones, which may be explained by seasonal hibernation in temperate bats (Streicker *et al.* 2012a). It is thought that torpor can reduce pathogen replication rates, lengthen incubation time, and allow viruses to overwinter (Calisher *et al.* 2006; Hayman *et al.* 2012). Bats have a long life span for their body size (Munshi-South & Wilkinson 2010), with many species living at least 25 years (Calisher *et*

al. 2006). This has important epidemiological consequences as a persistent infection in a long-lived organism could lead to many opportunities for virus transmission (Calisher *et al.* 2006). Finally, unique aspects of bat immune systems have been noted for the potential to allow unusual tolerance of viral pathogens that cause disease in other species (Baker *et al.* 2012; Brook & Dobson 2015), emphasizing that virulence is not a trait inherent to viruses themselves, but rather a product of host-virus interactions (Mandl *et al.* 2015).

In support of the idea that variation in bat life history and ecology could influence viral diversity, a literature-based analysis of bats as viral reservoirs by Luis *et al.* (2013) revealed that bats hosted more zoonotic viruses per species than rodents, while Olival *et al.* (2017) found that this also holds true compared to all mammals. Sympatry with other bat species, longevity, litter size, and number of litters per year were correlated with increased viral richness in bats (Luis *et al.* 2013). In another comparative study, Turmelle and Olival (2009) found that near-threatened and vulnerable bats had higher viral richness than species of least-concern, which is opposite to what has been found in primates (Altizer *et al.* 2007). They also found that increasing levels of host population structure were associated with greater viral diversity (Turmelle & Olival 2009). Gay *et al.* (2014) used virus data from the literature to conclude that species distribution shape, and specifically increased levels of habitat fragmentation, resulted in lower viral diversity in bats. Recent comparative studies showed that viral richness increases with colony size before plateauing in colonies over a hundred thousand individuals (Webber *et al.* 2017) and that cave-roosting was associated with increased viral sharing in bats (Willoughby *et al.* 2017). However, such literature-based studies are inherently challenged by different diagnostic techniques used across records, biased study effort across host species and geographic areas, and low specificity in distinguishing related viruses. Previous studies have also focused exclusively on interspecific comparisons, while intraspecific drivers of viral diversity remain relatively unexplored. Empirical tests of how natural variability in ecological factors influences observed viral diversity would provide a powerful advance towards identifying how changing environmental conditions or contact with humans could affect viral emergence.

Understanding the role of human activities is a key research priority (Hassell *et al.* 2017; Kelly *et al.* 2017), as anthropogenic land-use change has been implicated in recent viral emergence events involving bats (e.g. Field *et al.* 2001). Bats have likely long co-existed with these viruses, but human modification of the landscape is increasingly bringing humans and livestock into contact with bats and their viruses. For example, land-use

change associated with agriculture acted as an important driver in a spillover of Nipah virus from bats into livestock and humans (Pulliam *et al.* 2011). Bats are globally distributed and often live in large numbers in close association with humans or domestic animals (Wood *et al.* 2012). It is important to understand the ecological factors leading to disease spillover as it is not possible or desirable to control zoonotic diseases by eliminating bat hosts. For example, decades of culling of vampire bats have failed to eliminate rabies virus in Latin American (WHO 2013), and in one study in Peru, culling did not reduce rabies seroprevalence in wild vampire bats (Streicker *et al.* 2012b). Ecological factors leading to pathogen spillover are a complex mixture of processes, but our ability to understand and predict future zoonotic disease outbreaks first requires a holistic understanding of viral diversity in connection with bat ecology (Wood *et al.* 2012; Plowright *et al.* 2014).

1.4 Vampire bats

Common vampire bats (*Desmodus rotundus*) are an ideal study system in which to examine viral diversity, as they display distinctive social behaviors and interact ecologically with many different species, providing opportunities for the acquisition or maintenance of diverse viruses (Figure 1.1). They belong to the Neotropical family *Phyllostomidae* and the subfamily *Desmodontinae*, which contains two other less common vampire bat genera (*Diphylla* and *Diaemus*) (Baker *et al.* 2003). *D. rotundus* are native to Central and South America and occupy diverse habitats from rainforest to semi-arid desert, ranging in elevation from sea level to 3600 meters (Quintana & Pacheco 2007; Martins *et al.* 2009).
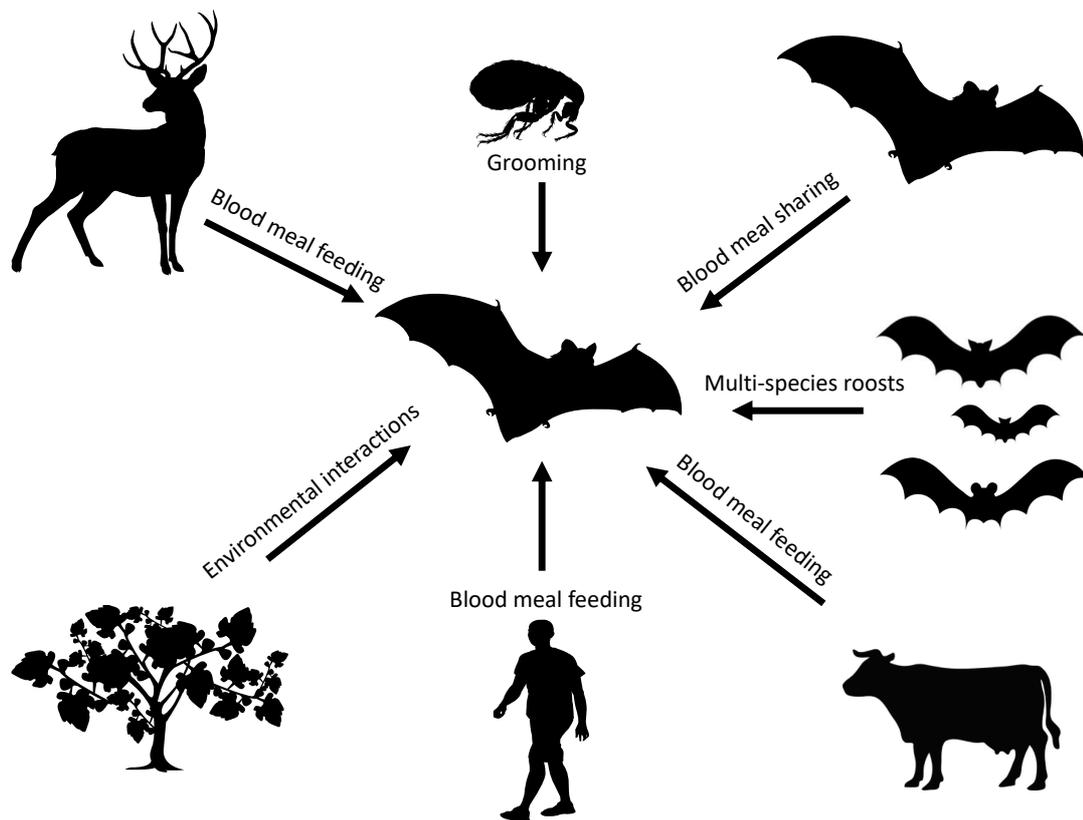
Figure 1 1 Summary of vampire bat ecological interactions.
Arrows and text represent specific behaviors or interactions of a vampire bat that could
result in virus acquisition. Adapted from Wray *et al.* (2016).

Vampire bats are a key wildlife reservoir in Latin America for rabies virus (family *Rhabdoviridae*) (Carini 1911; Pawan 1936) and have been associated with other viral families including *Paramyxoviridae* (Drexler *et al.* 2012a), *Coronaviridae* (Brandão *et al.* 2008), *Polyomaviridae* (Fagrouch *et al.* 2012), *Adenoviridae* (Lima *et al.* 2013; Wray *et al.* 2016), *Retroviridae* (Escalera-Zamudio *et al.* 2015) and *Herpesviridae* (Wray *et al.* 2016; Escalera-Zamudio *et al.* 2016; Salmier *et al.* 2017). Two recent metagenomic studies of vampire bats detected additional viral taxa, including some vertebrate-infecting families (Salmier *et al.* 2017; Escalera-Zamudio *et al.* 2017).

1.4.1 Behavioral features

Vampire bats, which are obligate blood-feeders, have evolved to specialize on native mammals (Greenhall *et al.* 1983), but now commonly feed on introduced livestock (Voigt & Kelm 2006; Bobrowiec *et al.* 2015), resulting in population growth and range expansions (Delpietro *et al.* 1992; Lee *et al.* 2012a). This shift in foraging behavior could impact viral richness in vampire bats if the livestock themselves harbor high viral diversity and prevalence due to close living conditions. However, bats continue to feed on diverse

wildlife in undisturbed areas (Voigt & Kelm 2006; Streicker & Allgeier 2016), which could expose the bats to the different communities of viruses unique to each prey species. Vampire bats also appear more likely to feed on humans in areas with less livestock (Schneider *et al.* 2009; Stoner-Duncan *et al.* 2014; Streicker & Allgeier 2016), providing another prey source from which bats might acquire viruses and creating the opportunity for disease transmission from bats to humans. A study of isotopic niche in vampire bats suggested that ecological connectivity via feeding is highest in regions of intermediate disturbance (Streicker & Allgeier 2016). Additionally, Becker *et al.* (2018) found that livestock density was associated with lower prevalence of two bacterial pathogens in vampire bats, potentially due to the positive effects of provisioning on host immunity. In summary, there is not a clear prediction as to how human modification of the landscape and its impact on foraging might affect viral diversity, but blood-feeding is an important behavioral feature of vampire bats that presents opportunities for viral sharing between bats and livestock, humans, and native wildlife.

Behavioral features such as allogrooming and food sharing through regurgitation of blood meals (Wilkinson 1984) might further promote virus transmission. Blood sharing occurs primarily between mother and offspring, but also may occur in small groups of females and males (Wilkinson 1984; DeNault & McFarlane 1995; Voigt *et al.* 2011; Carter & Wilkinson 2012). The sharing of blood meals within a colony may contribute to higher levels of viral richness within individuals than would be expected in other bat species due to sharing of fluids and the direct contact required.

1.4.2 Population-level features

Colony size and connectivity are two other factors that could affect viral richness and community composition in vampire bats. Host population size generally increases parasite persistence and diversity (Arneberg *et al.* 1998; Nunn *et al.* 2003; Vitone *et al.* 2004; Torres *et al.* 2006; Ezenwa *et al.* 2006; Lindenfors *et al.* 2007; Rifkin *et al.* 2012; Kamiya *et al.* 2013), which could be associated with increased probability of encountering novel viruses or with enhanced persistence within a colony, related to the epidemiological concept of critical community size (Bartlett 1957; Lloyd-Smith *et al.* 2005). Although Streicker *et al.* (2012b) did not find colony size to strongly impact seroprevalence of rabies in vampire bats, viruses with different transmission modes (i.e. density versus frequency dependent) might have varying probabilities of extinction depending on colony size (McCallum *et al.* 2001; Lloyd-Smith *et al.* 2005). A literature-based study found that viral

richness in bats increased up to a group size of 10,000 individuals (Webber *et al.* 2017); vampire bat colonies fall below that threshold, typically ranging from 10 - 200 bats (Greenhall *et al.* 1983), so colony size might therefore be important. Connectivity could also play a role, as many pathogens are best able to persist in a metapopulation (e.g. Grenfell & Harwood 1997; Swinton *et al.* 1998). Indeed, increasing levels of population structure have been associated with increased viral diversity in bats (Turmelle & Olival 2009) and vampire bat rabies can only persist in the context of a metapopulation (Blackwood *et al.* 2013). There is evidence of large-scale population structure in vampire bats (Martins *et al.* 2007; 2009) but haplotype sharing between neighboring colonies within regions and lack of local population structure in microsatellites suggests that there is local gene flow (Streicker *et al.* 2016). Levels of host movement between colonies might have implications for viral sharing, such that highly connected colonies could exhibit more similar viral community composition.

Host geographic range size is often positively correlated with parasite diversity (Torres *et al.* 2006; Lindenfors *et al.* 2007; Garrido-Olvera *et al.* 2012), which may arise because hosts with larger geographic ranges overlap with greater numbers of other host species, making them more likely to encounter and acquire novel parasites through cross-species transmission (Poulin 2014). Indeed, greater range overlap with other hosts (Davies & Pedersen 2008; Krasnov *et al.* 2010; Huang *et al.* 2013) and greater local host species richness (Harris & Dunn 2010; Kamiya *et al.* 2014) have both been associated with increased parasite diversity. This suggests that viral richness could be positively associated with bat species richness, which has been shown globally in the viral family *Coronaviridae* (Anthony *et al.* 2017b). Vampire bats live in colonies with other bat species including those in the genera *Micronycteris, Glossophaga, Carollia, Sturnia, Saccopteryx*, and *Artibeus* (Greenhall *et al.* 1983). Though each species typically has its own territory within the roost (Arellano-Sota 1988), aerosolized viruses can be transmitted within a bat roost (Winkler 1968). Other species, particularly smaller species, may try to avoid sharing a roost with vampire bats if possible (Wohlgenant 1994). However, in areas where trees are scarce, bats species may have to compete for roost space, which could lead to aggressive interactions between vampire bats and other species (Wohlgenant 1994). The comingling of species sharing a roost could result in virus transmission to vampire bats from other bat species, especially if species are phylogenetically closely related (Streicker *et al.* 2010), resulting in the accumulation of viral diversity as hypothesized by Luis *et al.* (2013).

1.4.3 Demographic features

Different demographic groups within a colony might exhibit differences in viral diversity. Host age can positively influence parasite diversity, as older hosts have had more time to accumulate different parasite species (Lo *et al.* 1998). However, the importance of juveniles in driving disease dynamics has been noted for both bacterial and viral pathogens in bats (Dietrich *et al.* 2015). In vampire bats, there is evidence that juveniles have higher infection prevalence for both viruses and bacteria (Streicker *et al.* 2012b; Volokhov *et al.* 2017). In addition, there might be sex-specific differences in viral diversity, as males and females exhibit different propensity for parasite infection based on behavior and physiology (Zuk & McKean 1996; Poulin 1996b; Reimchen & Nosil 2001; Negro *et al.* 2010). Studies have demonstrated differences in symbiotic microbial communities between healthy males and females (Mueller *et al.* 2006; Abeles *et al.* 2014), as well as sex-specific responses to illness in humans (Fish 2008). There is also evidence that the connection between parasitism and anthropogenic impact differs between male and female bats (Frank *et al.* 2016), suggesting the importance of considering sex-specific differences in studies of viral diversity.

1.4.4 Environmental features

Environmental conditions on both broad and small scales can influence viral diversity (Dinsdale *et al.* 2008; Hurwitz *et al.* 2014; Sunagawa *et al.* 2015). Studies have found that diversity generally tends to decline with both increasing distance from the equator (Guernier *et al.* 2004; Nunn *et al.* 2005; Lindenfors *et al.* 2007; Randhawa & Poulin 2010; Bordes *et al.* 2011; Guilhaumon *et al.* 2011) and with increasing elevation (Lomolino 2001), although this is not always true in bacterial communities (Fierer *et al.* 2011; Wang *et al.* 2011; Muletz Wolz *et al.* 2017). Local conditions such as temperature can play a role, with more diverse parasite communities occurring in warmer habitats (Poulin & Rohde 1997; Luque & Poulin 2008). However, these relationships may ultimately be driven by the relationship between host and parasite diversity, with higher host diversity in certain latitudes and habitats leading to higher parasite diversity in these environments (Poulin 2014). Habitat differences on a smaller scale can also influence the transmission and persistence of pathogens. For example, intact habitat has been associated with maintenance of rabies virus (de Thoisy *et al.* 2016), while habitat fragmentation decreased richness across multiple pathogens (Gay *et al.* 2014).

Due to their unique feeding habits and anthropogenic modification of the landscape which has resulted in increased levels of contact, vampire bats are ecologically closely connected to humans, livestock and wildlife. This may create the opportunity for viral emergence, but predicting the risk of cross-species transmission in this system first requires an understanding of vampire bat viral communities and factors that influence them.

1.5 Aims of the thesis

Viruses play important roles in health and disease of the organisms with which they interact, but our understanding of host-associated viral communities lies far behind that of other microbes such as bacteria (Stulberg 2016). This knowledge gap has potential implications for global health, as recent disease outbreaks in humans and domestic animals have been traced back to wildlife viruses (Karesh *et al.* 2012) and hosts with diverse viral communities may be more likely to contain viral taxa that can emerge in a new host species (Morse 1993; Wolfe *et al.* 2000). The first step in addressing this gap is to establish a baseline understanding of viral diversity by characterizing natural viral communities in wildlife and testing associated environmental and host demographic traits. If viral communities are structured in predictable ways, this might ultimately allow us to anticipate when changing conditions could lead to disease emergence from wildlife into humans and domestic animals (Anthony *et al.* 2015). Smaller scale investigations of viral communities within a single host species complement ongoing studies of single viral taxa at a global scale (e.g. Drexler *et al.* 2012a; Drexler *et al.* 2012b; Anthony *et al.* 2017b), allowing us to understand both the evolutionary and ecological forces contributing to viral diversity at different scales (Anthony *et al.* 2017b). This thesis presents a new field-laboratory-bioinformatic approach for comparative studies of wildlife-associated viral communities, then uses that method to characterize viral communities and examine potential drivers of variation on a country-wide scale in an ecologically important bat host.

First, I addressed field and laboratory challenges inherent to studying wildlife-associated viral communities by developing a minimally-biased shotgun metagenomic sequencing method for non-invasive field-collected samples (Chapter 2). This method was applied to pooled samples collected from *D. rotundus* in Peru, in which I statistically validated the ability to characterize the viral community using subsampling of reads. The results presented in this chapter are currently under review in the journal *Molecular Ecology Resources.*

Next I characterized novel viruses detected in vampire bat saliva and feces from across Peru (Chapter 3). Using multi-colony pools from eight localities, I described the overall diversity of viral taxa and the natural host groups of detected viruses, and examined differences in viral diversity between sample types. Viral metagenomic data were also generated for 48 single-colony pools of fecal and saliva samples. Combining multi-colony and single-colony datasets, large vertebrate-infecting viral contigs were phylogenetically analyzed compared to previously characterized bat viruses. The results presented in this chapter are in preparation for submission to the journal *mBio.*

Finally, I examined demographic and environmental correlates of viral diversity in vampire bats (Chapter 4). Saliva and fecal viral diversity at 23 colonies were tested for broad-scale differences between ecological regions, and correlations with geographic and host genetic distance. Colony-level demographic and local environmental variables were also tested for effects on viral richness and community composition. The results presented in this chapter are in preparation for submission to the journal *Ecology Letters.*

The results presented in the three data chapters were based on non-invasive swab samples collected from individual vampire bats captured at colonies across Peru. Individual samples were combined in different ways for metagenomic sequencing and analyses across different chapters. To clarify which samples were used in which analyses, Table A1 (Appendix A) presents the different metagenomic pools along with experiments or analyses in which they were included, as well as the individuals and colonies included in each pool.

Chapter 5 provides a general discussion of the results from previous chapters in a broader context, along with future directions. Together, the results paint a detailed picture of vampire bat viral communities on a country-wide scale and provide insight into factors that influence viral diversity in an ecologically important bat host.

1.6 Key collaborators

This project would not have been possible without the work and input of many collaborators in Glasgow and Peru. The flowchart below describes the sample collection and processing pipeline, and the people or institutions that were essential at each step (Figure 1.2).

Field sampling in Peru (2013-2016)
Carlos Tello
Dan Becker
Jorge Carrera
Anthony Almeyda
Jaime Pacheco
Jean Castro
Giuseppy Calizaya
Saori Grillo
Nicolas Tarmeño
Luiggi Carrasco
Yosselym Luzon
Marisel Flores

Storage in Lima, permits and transport to Scotland
Universidad Peruana Cayetano Heredia
Armando Hung
Nestor Falcon
Carlos Shiva
Carlos Tello
John Claxton

Viral metagenomic labwork
Centre for Virus Research, University of Glasgow
Ana da Silva Filipe
Chris Davis
Andrew Shaw
Alice Broos
Mariana Varela
Gavin Wilkie
Daniel Mair
Lily Tong

Bioinformatic pipeline development
Richard Orton

Figure 1 2 Key collaborators and institutions in the study of vampire bat viral communities.

2 Using non-invasive metagenomics to characterize viral communities from wildlife

2.1 Abstract

Microbial communities play an important role in organismal and ecosystem health. While high throughput metabarcoding has revolutionized the study of bacterial communities, generating comparable viral communities has proven elusive, particularly in wildlife samples where the diversity of viruses present and limited quantities of viral nucleic acid present distinctive challenges. Metagenomic sequencing is a promising solution for studying viral communities, but the lack of standardized methods currently precludes comparisons across host taxa or localities. Here I developed an untargeted shotgun metagenomic sequencing protocol to generate comparable viral communities from non-invasively collected fecal and oropharyngeal swabs. Using samples from common vampire bats (*Desmodus rotundus*), a key species for virus transmission to humans and domestic animals, I tested how different storage media, nucleic acid extraction procedures and enrichment steps affect viral community detection. Based on finding viral contamination in fetal bovine serum, it is recommended to store swabs in RNALater or another non-biological medium. Based on qPCR tests, nucleic acid should be extracted directly from swabs rather than from supernatant or pelleted material, which had undetectable levels of viral RNA. Using a low-input RNA library preparation protocol, I established that ribosomal RNA depletion and light DNAse treatment reduce host and bacterial nucleic acid and improve virus detection. Finally, applying the final protocol to twelve pooled samples from seven localities in Peru, I showed that detected viral communities saturated at the attained sequencing depth, allowing unbiased comparisons of viral community composition. Future studies using the methods outlined here will elucidate the determinants of viral communities across host species, environments and time.

2.2 Introduction

Microbial communities of bacteria and viruses play important roles in ecosystem function (Suttle 2007; van der Heijden *et al.* 2008; Strom 2008; Strickland *et al.* 2009) and in maintaining the health of organisms (Ley *et al.* 2006; Muegge *et al.* 2011; Manrique *et al.* 2016). Despite the importance of studying microbial communities in the environment and within hosts, classical methods of microbe discovery are not easily applied at the community level. For example, characterization by isolation and culturing are unsuitable

for members of the microbial community that are difficult to grow in culture (Fancello *et al.* 2012). Serological tests of antibody presence are targeted towards specific taxa, and can be difficult to interpret due to antibody cross-reactivity and inconsistent cut-off thresholds for positivity (Gilbert *et al.* 2013). Molecular detection of nucleic acids by targeted PCR remains an important technique for sequencing specific genomic regions, but these approaches cannot identify all taxa present and are inappropriate for discovering new, highly divergent taxa as designing primers or probes requires prior knowledge of nucleotide sequences (Fancello *et al.* 2012; Temmam *et al.* 2014). In contrast, unbiased deep sequencing has the potential to capture a snapshot of microbial communities in a large number of samples without prior expectations about what taxa will be detected.

Deep sequencing has illuminated the structure and function of microbial communities across time and space in ways that would not have been possible using traditional methods. In the field of ecology, theories developed at macro-organismal level have been tested in microbial communities, such as the cycling of predator and prey populations (Rodriguez-Brito *et al.* 2010) and the existence of elevational diversity gradients (Fierer *et al.* 2011). Deep sequencing has also demonstrated that both bacterial and viral communities differ across abiotic environments (Dinsdale *et al.* 2008) in such diverse systems as soil bacteria (Fierer *et al.* 2012) and marine viruses (Hurwitz *et al.* 2014). In the context of human and animal health, deep sequencing can identify candidate pathogens in unexplained disease (Cox-Foster *et al.* 2007; Palacios *et al.* 2008; Honkavuori *et al.* 2008; Briese *et al.* 2009) and potential hosts and vectors of emerging pathogens (Masembe *et al.* 2012; Veikkolainen *et al.* 2014). Studies of host-associated microbial communities have revealed that microbes vary across body habitats, space, and time (Costello *et al.* 2009; Blekhman *et al.* 2015), and that a community-level perspective of host-associated microbes is critical for understanding health and disease (Lecuit & Eloit 2013; Vayssier-Taussat *et al.* 2014; Virgin 2014). Sequencing host-associated bacterial communities in wildlife has revealed that communities vary over time (Bobbie *et al.* 2017), that social interactions are key determinants of community composition (Tung *et al.* 2015; Grieneisen *et al.* 2017) and that dietary changes due to habitat degradation can alter bacterial communities (Amato *et al.* 2013). While host-associated viral communities in wildlife remain relatively unexplored, the divergent responses of host-associated bacteria and viruses to experimental diet modification (Howe *et al.* 2015) and the biological differences between the two types of microbes suggest that viral communities in wildlife might exhibit different patterns to those observed in bacteria.

Deep sequencing studies of microbial communities typically employ either metagenomics, which is the random sequencing of genomic fragments of an entire sample, or metabarcoding, which is a sequence-specific PCR-based approach (Creer *et al.* 2016). Studies of bacterial communities frequently use 16S ribosomal rRNA metabarcoding to examine highly multiplexed samples. However, viral communities lack a similarly conserved marker across or even within viral families (Rohwer & Edwards 2002; Mokili *et al.* 2012) and therefore are more commonly characterized using metagenomics. Although this approach is currently less cost- and time-efficient than metabarcoding for large numbers of samples, it can assign taxa at higher resolution (depending on factors such as read length, genomic region, and reference database) and avoids PCR biases (Jovel *et al.* 2016). Shotgun metagenomics also allows the simultaneous characterization of different microbial communities (e.g. bacterial and viral) (Chandler *et al.* 2015; Schneeberger *et al.* 2016) as well as host population structure and diet (Srivathsan *et al.* 2016). Furthermore, metagenomics can detect viruses at or below the sensitivity of taxon-specific PCR and qPCR (Greninger *et al.* 2010; Yang *et al.* 2011; Li *et al.* 2015), implying that broader taxonomic coverage does not necessarily trade off with sensitivity. Targeted approaches also likely underestimate or bias measures of viral diversity, potentially impacting downstream comparative analyses. The ability of metagenomics to sensitively detect taxa that are not specifically targeted and/or were previously undescribed has the potential to overturn prior understandings of viral community diversity and distribution based on serology and PCR.

Despite the great promise of metagenomics for studying viral communities, challenges inherent to sequencing viral genomes and technical uncertainties need to be addressed to maximize comparability. Viral communities include single and double stranded viruses with both DNA and RNA genomes, ranging in size from 1,259,197bp (*Megavirus chilensis;* Arslan *et al.* 2011) to 1,700bp (*Hepatitis deltavirus;* Taylor 2006). Larger viral genomes that have a higher probability of being sequenced may be overrepresented in the inferred community (Fancello *et al.* 2012). The RNA virus component of viral communities is highly sensitive to degradation due to temperature and storage conditions, raising questions about how samples should be preserved and transported (Cardona *et al.* 2012). Indeed, different storage media alter viral detection in PCR-based studies (Forster *et al.* 2008; Osborne *et al.* 2011) and it is reasonable to assume the same in metagenomic studies. Two popular methods for preserving viruses from field or clinical samples are viral transport media (VTM), an aqueous solution that typically contains protective proteins, antibiotics, and buffers to control the pH (Johnson 1990) and RNALater, a commercial

reagent that penetrates tissues and stabilizes RNA (Ambion). VTM has historically been used to preserve samples when viruses are to be detected by PCR or cultured in vitro (Jensen & Johnson 1994; Druce *et al.* 2012). Given the large number of historically-collected samples in VTM, it would be ideal to include these in metagenomic studies. However, VTM may not be an appropriate medium because one commonly used component, fetal bovine serum (FBS), may be contaminated with bovine viruses. RNALater is another popular medium for storing microbial samples collected in the field (Drexler *et al.* 2011; Gomez *et al.* 2015; Frick *et al.* 2017; Bányai *et al.* 2017), as it preserves RNA without requiring immediate freezing. However, its high salt content, while not problematic for solid tissue samples, creates challenges for nucleic acid extraction from the kinds of non-invasive swab samples that are typical of ecological field studies (e.g. blood, urine, feces, saliva). While viruses are often extracted from an aliquot of supernatant (Tse *et al.* 2012; Wu *et al.* 2012; Baker *et al.* 2013), extraction from the swab itself may be desirable for samples stored in RNALater (Vo & Jedlicka 2014). These extraction procedures need to be tested and optimized for more widespread use in non-invasive viral metagenomics.

Another challenge for viral metagenomics is that since genomes are sequenced at random, larger host and bacterial genomes are preferentially detected relative to smaller viral genomes (Nakamura *et al.* 2009; Yang *et al.* 2011). For this reason, samples are often enriched for viruses using methods including nuclease treatment, filtration of host/bacterial particles, density gradient centrifugation, and removal of host rRNA (Hall *et al.* 2014; Kleiner *et al.* 2015; Kohl *et al.* 2015). DNAse treatment is a well-established and effective method of enrichment (Allander *et al.* 2001), while filtration and centrifugation are sometimes used but can bias the inferred viral community composition (Thurber *et al.* 2009; Kleiner *et al.* 2015) and are impractical for ecological studies given the large numbers of samples typically processed and interest in generating community data rather than focusing on a particular pathogen. Depletion of host rRNA is unlikely to bias the viral community (He *et al.* 2010; Matranga *et al.* 2014), but may affect the distribution of coverage across the viral genome (Li *et al.* 2016). There is therefore a need to identify a combination of laboratory methods that will maximize the proportion of viral reads while minimizing bias, allowing greater multiplexing and enabling metagenomic studies of viral communities on an ecological or evolutionary scale.

The aim of this chapter was to develop a method for generating comparable viral community data from non-invasively collected samples from wildlife. Specifically, a field-

laboratory-bioinformatic pipeline was developed to characterize viral communities in fecal and oropharyngeal swabs from common vampire bats (*D. rotundus*) in Peru. Two preliminary sequencing runs were performed, the results of which motivated four pilot studies to inform the optimized comparative metagenomic protocol. The following questions were addressed in pilot studies: (1) are samples stored in VTM containing FBS appropriate for viral metagenomics? (2) what is the most effective way to extract viral nucleic acid from swabs stored in RNALater? and do the enrichment methods of (3) rRNA depletion and (4) DNAse treatment increase the number of viral reads or viral taxa detected? The optimized protocol was then applied to field-collected samples to validate whether viral communities were reliably characterized at commonly attained depths of sequencing.

2.3 Materials and Methods

2.3.1 Authorizations

Bat capture and sampling methods were approved by the Research Ethics Committee of the University of Glasgow School of Medical Veterinary and Life Sciences (Ref081/15) and the University of Georgia Animal Care and Use Committee (A2014 04-016-Y3-A5). Bat capture and sampling was approved by the Peruvian Government under permits RD-009-2015-SERFOR-DGGSPFFS, RD-264-2015-SERFOR-DGGSPFFS, and RD-142-2015-SERFOR-DGGSPFFS. Access to genetic resources was granted under permit RD-054-2016-SERFOR-DGGSPFFS.

2.3.2 Field sampling of common vampire bats

Wild common vampire bats were captured and sampled at colonies across Peru. Roosts were either natural (caves, trees) or man-made structures (abandoned houses, tunnels, mines) inhabited by bats. Bats were captured within roosts using hand nets, or while exiting roosts using mist nets and harp traps. For nocturnal captures, nets were open from approximately 18:00 – 6:00 and checked every 30 minutes; a combination of 1-3 mist nets and 1 harp trap were used depending on the size and number of roost exits identified. When exact roost locations were unknown, bats were captured while foraging at nearby livestock pens. Upon capture, bats were placed into individual cloth holding bags before being processed and sampled. Bats were also given a uniquely numbered wing band (Porzana Inc) for identification of recaptures in ongoing longitudinal studies.

Oropharyngeal (saliva) samples were collected by allowing bats to chew on cotton-tipped wooden swabs (Fisherbrand) for 10 seconds. Fecal samples were collected by rectal swab, using a 3-mm diameter rayon-tipped aluminum swab (Technical Service Consultants Ltd) dipped in sterile Dulbecco's Phosphate Buffered Saline (DPBS) (Gibco). Swabs were stored in uniquely numbered cryovials containing 1 mL RNALater (Ambion) or VTM (10% fetal bovine serum, penicillin-streptomycin, fungizone antimycotic). Following the manufacturer's instructions, swabs in RNALater were stored overnight at 4°C before being transferred to dry ice (around -80°C), while those in VTM were immediately placed on dry ice until both were permanently stored in -70°C freezers.

2.3.3 Preliminary sequencing

The four pilot experiments described below (Section 2.3.6 - 2.3.9), which informed the final protocol, were developed based on the results of two preliminary sequencing runs. These runs aimed to characterize viral communities based on nucleic acid extracted directly from buffer (VTM and RNALater) in which swabs were stored. It was subsequently found that VTM is not a suitable medium for storing metagenomic samples and that extracting from buffer is not optimal for RNALater samples, thus comparisons of viral communities detected in those experiments are difficult to interpret. However, a description of preliminary sequencing runs and some results are presented to give context to the pilot experiments.

2.3.4 Viral communities from swab samples stored in VTM

The first preliminary run tested whether it was possible to characterize viral communities from swab samples stored in VTM, where nucleic acid was extracted directly from supernatant. Samples analyzed were collected from four vampire bat colonies in the southern Andes region of Peru (Figure B1) in July - August 2015. Field sampling was conducted as described in Section 2.3.2.

2.3.4.1 Extraction and library preparation

Total nucleic acid was extracted from the VTM buffer in which swab samples had been stored. Swabs were removed using sterile forceps and 100 μL of buffer was aliquoted to 96 well extraction plates along with 40 μL Proteinase K (Qiagen). Samples were inactivated at this stage by adding 600 μL of a mixture containing Buffer RLT, MagAttract Suspension G (both Qiagen) and isopropyl alcohol. All of these steps were performed according to

CL2+ guidelines. Plates for wash steps were prepared, including one plate containing 700 μL wash buffer AW1 and two plates containing 500 μL wash buffer RPE (both Qiagen). All plates were loaded onto a Kingfisher Flex 96 automated extraction machine (Thermo). The instrument settings, provided by Qiagen ('Protocol for purification of viral nucleic acid and bacterial DNA with Thermo Scientific KingFisher Flex'), consist of a lysis and binding step, followed by three wash steps, and a final elution in 80 μL Buffer AVE (RNase-free water with 0.04% NaN_3). Extracted nucleic acid was stored at -80°C. Eight sequencing pools were prepared by combining nucleic acid from ten vampire bat samples that were pooled by four colonies and three sample types (Table 2.1; Table A1).

Table 2 1 Description of samples used to test viral communities from swabs stored in VTM. Samples were analyzed in preliminary sequencing run 1. Pools were made up of nucleic acids extracted from 10 individual swabs of the same sample type from the same site.

Sample ID†	Sample Type	Site‡	Department	Raw reads	Viral reads
H1	Feces	AYA15	Ayacucho	3,098,838	38,097
H2	Feces	API17	Apurimac	2,870,572	137,293
H3	Feces	AYA14	Ayacucho	2,818,423	34,452
H4	Feces	CUS8	Cusco	4,367,722	20,216
SG1	Saliva	AYA15	Ayacucho	4,106,626	286
SV1	Saliva	API17	Apurimac	2,623,124	311
SV2	Saliva	AYA14	Ayacucho	2,426,277	229
SV3	Blood	AYA15	Ayacucho	2,811,141	318

†Sample ID reflects sample type where H is feces, SV is saliva, and SG is whole blood
‡Sites are depicted in Figure B1

All nucleic acid extracts were quantified using a Qubit 3.0 fluorometer and a Qubit RNA HS Assay (Life Technologies) to determine RNA concentration for pooling. As nucleic acid was undetectable following extraction, sample input into pools was normalized by volume rather than concentration; 50 μL was taken from each of 10 extracts for a total of 500 μL. Pools were then concentrated using 1.8X Agencourt RNAClean XP beads (Beckman Coulter) before reverse transcription and library preparation (Appendix B1.3). Libraries were pooled in equimolar ratios for sequencing on an Illumina MiSeq instrument with v3 2x201 bp chemistry.

2.3.4.2 Bioinformatic pipeline and viral community datasets

A bioinformatic pipeline was created for virus discovery and viral community analyses in vampire bat samples (Appendix B2; Figure B2). Briefly, the pipeline filtered out low-quality reads and duplicates, then filtered out non-viral reads including those matching the vampire bat genome (Zepeda-Mendoza *et al.* 2018; NCBI BioProject Accession PRJNA414273), the PhiX Illumina sequencing control, ribosomal RNA, and other reads

with high matches to prokaryote/eukaryote sequences. Remaining reads were assembled into contigs, and then both raw reads and assembled contigs were assigned to viral taxa by comparison to the NCBI Viral RefSeq database.

Viral reads and contigs were converted into lists of viral taxa at different taxonomic levels using MEGAN Community Edition (Huson *et al.* 2016). Viral taxa were filtered for vertebrate-infecting viruses using a list of vertebrate-infecting viral families and genera (Table B1) that was compiled from the 2017 ICTV Taxonomy (Adams *et al.* 2017), then viral family and genus richness were calculated using the R package *vegan* (R Core Team 2017; Oksanen *et al.* 2017). Further details on the bioinformatic pipeline and generating viral community datasets are found in Appendix B2.

2.3.4.3 Results

The preliminary sequencing run testing the ability to generate viral community data from samples stored in VTM and extracted directly from buffer yielded a total of 25,378,779 raw reads, which were relatively evenly divided among samples (Table 2.1). The proportion of sequences matching to viral taxa was low (<5%) in all samples, with the majority of reads mapping to the vampire bat genome (Figure B3A). In the blood sample SG1, there was a lower proportion of host reads, potentially due to mammalian red blood cells being non-nucleated, and this sample had a higher proportion of low complexity/PCR duplicate reads. Fecal and saliva samples had similar numbers of non-viral reads that were filtered out in the bioinformatic pipeline (Figure B3A).

Of the reads assigned to viral taxa, the vast majority were assigned to *Podoviridae,* a bacteriophage family (Table 2.2). Several other viral taxa were detected, including vertebrate-infecting families such as *Herpesviridae, Parvoviridae* and *Adenoviridae*. There were more viral reads in fecal samples due to the abundance of reads assigned to bacteriophage taxa, with an average of 57,515 viral reads per fecal sample (20,216 – 137,293 reads) compared to 286 reads on average for saliva and blood samples (229 – 318 reads).

Table 2 2 Read counts from different viral families detected from swabs stored in VTM. These samples were sequenced using a shotgun metagenomic approach of nucleic acid extracted directly from VTM buffer. Sample names correspond to Table 2.1.

	H1	H2	H3	H4	SG1	SV1	SV2	SV3
Adenoviridae	10	6	0	0	0	1	0	0
Baculoviridae	1	0	0	0	0	0	0	0
Myoviridae	11	23	17	12	6	6	14	20
Podoviridae	38,009	137,168	34,329	20,132	241	236	122	115
Siphoviridae	23	36	41	35	19	21	27	41
Herpesviridae	2	1	19	0	2	38	57	112
Iridoviridae	0	0	0	0	0	1	0	0
Papillomaviridae	0	3	0	2	0	0	0	0
Phycodnaviridae	0	2	1	0	2	0	1	2
Polydnaviridae	0	0	0	1	0	0	0	0
Poxviridae	0	0	0	0	0	0	1	0
Retroviridae	0	1	0	0	0	0	0	1
Anelloviridae	0	0	3	0	0	0	0	0
Circoviridae	4	2	0	0	0	0	0	0
Microviridae	0	0	0	0	1	0	0	0
Parvoviridae	11	2	1	0	0	1	0	0
Flaviviridae	0	2	0	1	6	2	0	14
Hepeviridae	2	1	0	0	0	0	0	0
Picornaviridae	0	0	0	6	0	0	0	0
Tombusviridae	0	3	1	5	0	0	2	0
Tymoviridae	0	4	0	1	0	0	0	0

The most notable finding was the detection in five samples of Bovine Viral Diarrhea Virus (BVDV; family *Flaviviridae*; genus *Pestivirus*), a known cell-culture contaminant often found in fetal bovine serum (FBS), one of the components of VTM. Reads assigned to the genus *Pestivirus* were selected from one representative of each sample type (H2, SG1, SV3) for further analysis by nucleotide blast (Table 2.3); all reads analyzed yielded matches with high percent identity to Bovine Viral Diarrhea Virus 3 (BVDV-3).

Table 2 3 Description of blast hits matching to BVDV-3 from samples stored in VTM. Blast results are shown from three samples in which BVDV was detected (SV3, SG1, H2) which represent different sites and sample types.

Read	Sample	Read Length (nt)	Blast Genbank ID	Blast Query cover	Blast % ID	Description of blast hit
15147/1	SV3	187	FR873802.1	100%	98%	Bovine viral diarrhea virus 3 strain Au/A55110-1162/09
13734/1	SV3	187	KY762287.1	99%	98%	Bovine viral diarrhea virus 3 strain PB22487
19330/1	SV3	185	KC297709.1	100%	99%	Bovine viral diarrhea virus 3 strain LVRI/cont-1
21648/1	SV3	185	KY762287.1	100%	97%	Bovine viral diarrhea virus 3 strain PB22487
16830/1	SV3	171	KY767958.1	100%	98%	Bovine viral diarrhea virus 3 strain SV478/07
5215/2	SV3	146	KU563155.1	100%	97%	Bovine viral diarrhea virus 3 isolate HN1507
15147/2	SV3	200	JX985409.1	100%	98%	Bovine viral diarrhea virus 3 isolate CH-KaHo/cont
13734/2	SV3	201	KY762287.1	100%	98%	Bovine viral diarrhea virus 3 strain PB22487
19330/2	SV3	199	KC297709.1	100%	98%	Bovine viral diarrhea virus 3 strain LVRI/cont-1
21648/2	SV3	200	KY683847.1	100%	94%	Bovine viral diarrhea virus 3 strain SV757/15
16830/2	SV3	185	KY767958.1	100%	98%	Bovine viral diarrhea virus 3 strain SV478/07
22574/1	SG1	187	KU563155.1	100%	95%	Bovine viral diarrhea virus 3 isolate HN1507
2794/1	SG1	187	AB871953.1	99%	96%	Bovine viral diarrhea virus 3 strain: D32/00_'HoBi'
13581/1	SG1	185	KY683847.1	100%	98%	Bovine viral diarrhea virus 3 strain SV757/15
22574/2	SG1	198	AB871953.1	98%	95%	Bovine viral diarrhea virus 3 strain: D32/00_'HoBi'
2794/2	SG1	201	KY767958.1	100%	98%	Bovine viral diarrhea virus 3 strain SV478/07
13581/2	SG1	199	KY767958.1	100%	98%	Bovine viral diarrhea virus 3 strain SV478/07
15012/1	H2	185	KY762287.1	100%	95%	Bovine viral diarrhea virus 3 strain PB22487
15012/2	H2	200	KY762287.1	100%	95%	Bovine viral diarrhea virus 3 strain PB22487

These samples were not taken from the same sites, making it even more unlikely that this virus originated from bat samples. The contamination of samples with viral nucleic acid originating from cows is particularly problematic here, given the possibility of vampire bats being truly infected with viruses of bovine origin that have been acquired through diet. This finding motivated the second preliminary sequencing run, which aimed to compare viral communities from theoretically identical samples stored in VTM and RNALater, as well as the utility of an enrichment protocol entailing DNAse treatment and the filtration of viral particles.

2.3.5 Testing the effect of storage medium and enrichment on viral community detection

A second preliminary sequencing run was performed to test the effect of storage medium and the utility of an enrichment protocol on viral detection when nucleic acid was extracted from the buffer in which swab samples were stored. To test the effect of storage medium, paired fecal and saliva swabs were collected from ten bats, five each at two different sites in the department of Lima (Figure B1) in April 2016. Field sampling was conducted as described in Section 2.3.2. Of these paired swabs, one was stored in 1 mL VTM and the other in 1 mL RNALater. To test the effect of enrichment, samples were split following pooling and before extraction; half of each sample was extracted using enrichment of DNAse treatment and filtration of viral particles, and half was extracted without enrichment.

2.3.5.1 Extraction and library preparation

Samples were processed as in the first preliminary run (Section 2.3.4.1), except that pooling was performed by sample type and storage medium prior to extraction. After thawing samples, swabs were removed using sterile forceps and samples were centrifuged at 10,000xg for 10 minutes to remove debris. Four pools were created by combining 20 μL supernatant of each sample type x storage medium combination from ten individuals. For each pool, half was placed on ice to be extracted unenriched, and the other half was enriched for viral particles, resulting in a total of eight pools (Table 2.4; Table A1).

Table 2 4 Description of samples analyzed to test enrichment and storage medium. Samples were sequenced using a metagenomic approach in preliminary sequencing run 2. Pools were made up of nucleic acids extracted from 10 individual swabs of the same sample type pooled across two sites.

Sample ID†	Sample Type	Enrichment	Storage medium	Raw reads	Viral reads
H_EN_VTM	Feces	Enriched	VTM	15,172,421	14,896
H_UN_VTM	Feces	Unenriched	VTM	15,489,950	951,213
SV_EN_VTM	Saliva	Enriched	VTM	16,286,719	141
SV_UN_VTM‡	Saliva	Unenriched	VTM	-	-
H_EN_RL	Feces	Enriched	RNALater	1,705,558	301
H_UN_RL‡	Feces	Unenriched	RNALater	-	-
SV_EN_RL	Saliva	Enriched	RNALater	4,400,423	1,164
SV_UN_RL	Saliva	Unenriched	RNALater	21,063,046	1,172

†Sample ID reflects sample type (H feces or SV saliva), enrichment protocol (EN enriched or UN unenriched) and storage medium (VTM or RNALater)
‡SV_UN_VTM and H_UN_RL were excluded from analyses so read counts are not shown

For enrichment, 100 μL of each pool was sequentially filtered through 0.45 and 0.22 uM filters (Millipore Ultrafree MC); samples were applied to spin column filters and centrifuged at 12,000xg for 4 minutes. After filtration, each pool was DNAse treated. The reactions consisted of 0.1 U/μL DNAse Turbo (Ambion) and 1x DNAse Turbo buffer incubated at 37°C for 30 minutes. Pool volumes were adjusted to 500 μL using sterile DPBS (Gibco), then concentrated and reverse filtered using Amicon spin column filters (Millipore Amicon Ultra Centrifugal Filters 10K MWCO), which were centrifuged at 14,000xg for 5 minutes. Filters were then removed, placed upside down in a clean tube, and centrifuged at 1,000xg for 2 minutes. Final volume for each enriched pool was adjusted to 100 μL using sterile DPBS. 100 μL of each of the ten samples was aliquoted to 96 well extraction plates along with 40 μL Proteinase K. From this point, samples were extracted, stored and quantified as described in Section 2.3.4.1.

Samples were converted to cDNA and libraries were prepared (Appendix B1.3). For most samples, 10 μL nucleic acid was used directly as input into cDNA synthesis but for two samples (H_EN_RL, SV_UN_RL), there was not sufficient nucleic acid in the 10 μL aliquots to prepare a library. For these samples, 40 μL aliquots were first concentrated to 10 μL using 1.8X Agencourt RNAclean XP beads. Samples were pooled in equimolar ratios along with other metagenomic samples for sequencing on an Illumina NextSeq500 with Mid Output v2 2x150 bp chemistry. Sequences were processed through the bioinformatic pipeline (Appendix B2).

2.3.5.2 Results

Sample SV_UN_VTM failed to sequence properly due to low clustering; this was because it was likely mixed up with H_UN_RL during library preparation, so both were excluded from analyses, yielding a total of 74,118,117 paired-end reads from the remaining six samples (Table 2.4). Compared with the previous run, a higher proportion of reads were lost to PCR duplicate filtering (Figure B3B). This was potentially due to the low diversity of nucleic acid present following extraction, as all samples were initially unmeasurable and remained so throughout the library prep. Proportions of reads assigned to different non-viral sources were highly variable between samples (Figure B3B).

As viral samples may have historically been collected in VTM, it would be useful to know whether they are usable material for metagenomics studies, and whether they can be compared with samples stored in other media. Unfortunately, the viral community

comparisons in this study would not be easy to interpret; several samples had to be excluded, VTM was subsequently found to be not a good storage buffer and extracting directly from RNALater buffer was found to be not very effective. Therefore, the results from this set of samples are not further discussed.

Based on results from the two preliminary sequencing runs, specifically the detection of bovine viruses in VTM-stored samples (Table 2.3) and the low number of viruses detected after sequencing nucleic acid extracted from swabs stored in RNALater (Table 2.4), I conducted four pilot studies which informed the final extraction and sequencing methods. These pilot studies aimed to determine whether samples stored in VTM containing FBS are appropriate for metagenomic analyses, what is the most effective way to extract viral nucleic acid from swab samples stored in RNALater, and whether rRNA depletion and DNAse treatment are useful as viral enrichment strategies.

2.3.6 Pilot study 1: Are samples stored in viral transport media appropriate for viral metagenomic analysis?

Based on the detection of BVDV in the first preliminary sequencing run, two different batches of FBS were analyzed using a shotgun metagenomic approach to evaluate the presence of bovine viruses and to determine whether another storage medium, such as RNALater, would be more appropriate. Total nucleic acid was extracted and quantified from sterile 100 μL aliquots of FBS (Gibco) as described in Section 2.3.4.1.

For one aliquot, 10 μL nucleic acid was used directly as input into cDNA synthesis but for the other there was not sufficient nucleic acid in the 10 μL aliquot to prepare a library, so 40 μL was first concentrated to 10 μL using 1.8X Agencourt RNAclean XP beads (Beckman Coulter). Nucleic acid was then converted to cDNA and libraries were prepared as described in Appendix B1.3. Samples were pooled in equimolar ratios for sequencing on an Illumina NextSeq500 with Mid Output v2 2x150 bp chemistry. The resulting reads were processed through the bioinformatic pipeline (Appendix B2).

2.3.7 Pilot study 2: What extraction method for swabs stored in RNALater maximizes nucleic acid?

This experiment used swabs that were inoculated with known concentrations of viral particles to identify the extraction method that maximized viral nucleic acid from swabs stored in RNALater and to assess the efficiency and repeatability of the extraction

protocol. Swabs were designed to mimic samples collected from the field, with the caveat that they did not include host material (e.g. feces, saliva), bacteria, parasites or the community of viruses expected to be present in field-collected samples. These other components of samples could impact extraction and PCR efficiency, for example, by acting as a carrier to enhance RNA extraction or through the presence of compounds that can act as extraction or PCR inhibitors. Also only one virus was tested in this experiment, which may limit the ability to extrapolate results to other types of viruses. However, rather than inoculate field-collected swabs, in which differences between sample types or pathogen communities could introduce uncontrolled variation, "clean" mock swabs were used that would allow the evaluation of differences in viral detection between extraction methods.

Extraction tests used Schmallenberg virus (SBV), a single-stranded RNA orthobunyavirus (Hoffmann *et al.* 2012). A 3.9×10^5 plaque forming units (PFU)/mL stock of SBV was serially diluted using sterile DPBS (Gibco), and 10 μL of cell-free virus at a range of dilutions from $10^6 - 10^3$ copies/mL was inoculated into the same swabs used in field studies (Fisherbrand, Technical Service Consultants Ltd). Swabs were stored in 1 mL of RNALater at -80°C overnight.

Two extraction methods were initially tested: a magnetic bead-based extraction using the reagents of the Biosprint One-for-all-vet kit (Qiagen) and extraction using TRIzol reagent (Invitrogen). The TRIzol method, a guanidinium thiocyanate-phenol-chloroform extraction, was ineffective as no virus was detected by qPCR in extractions from any of the components of the sample (results not shown). This could be due to salt components of RNALater, an aqueous sulfate salt solution, which appeared to adversely affect the phase separation during the TRIzol extraction. Therefore, only the magnetic bead-based extraction method is further discussed. Reagents and volumes are based upon the manufacturer's protocol, and steps are a manual approximation of the automatic extraction method performed on the Kingfisher Flex 96 machine (Appendix B1.1). Extractions were performed manually because the CVR, where the work was carried out, has a workflow which does not allow for lab propagated samples, such as the SBV used in this experiment, to be extracted on machines that are used for clinical or field collected samples, such as the Kingfisher.

Samples were thawed in a CL2 flow hood and the swab was removed into a tube containing 288 μL Buffer RLT and 40 μL Proteinase K. The swab/lysis buffer tubes were

vortexed for 15 seconds, incubated for 5 minutes, vortexed again, and the swab removed. 100 µL DPBS was added to make up the volume required in the protocol.

The original tube without the swab was centrifuged at 13,000xg for 5 minutes. 100 µL supernatant was then removed to another tube containing 288 µL Buffer RLT and 40 µL Proteinase K. Finally, the remaining supernatant was removed from the original tube and the pellet resuspended in 100 µL DPBS. The resuspended pellet was transferred to a third tube containing 288 µL Buffer RLT and 40 µL Proteinase K.

To each of the three tubes now containing sample (swab, supernatant, resuspended pellet), lysis buffer and Proteinase K, 288 µL isopropanol and 24 µL MagAttract beads were added. Samples were mixed on a rotating tube mixer for 5 minutes, briefly spun down, beads pelleted using a magnetic bead separation rack and supernatant removed. All mixing and pelleting steps were performed in this way. Three wash steps were performed – 700 µL Buffer AW1 and 1 minute of mixing, followed by two steps of 500 µL Buffer RPE and 1 minute of mixing. After removing supernatant, beads were air dried for 15 minutes, 100 µL Buffer AVE was added and mixed for 5 minutes. Beads were pelleted and the supernatant was removed and kept. The extracted RNA was temporarily stored at -20°C before proceeding to cDNA synthesis.

cDNA synthesis was performed with 5 µL RNA using random primers and SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. The cDNA was then used as input into a qPCR assay to determine viral copy number in each extraction. The qPCR assay was performed using the Brilliant III Ultra Fast qPCR kit (Agilent) and previously designed SBV primers and probe (Hoffmann *et al.* 2012) on an ABI7500 Fast Real Time PCR System (Applied Biosystems). qPCR reactions of samples and standards were run in triplicate, and quantity of viral copies was assessed against a standard curve of SBV concentrations ranging from 10^9-10^1 copies/µL. Controls containing no template were run in triplicate for each set of qPCR reactions. The approximate accuracy of the qPCR assay was confirmed by quantifying undiluted virus that had been extracted using the same method (concentration $3.9x10^5$ PFU/mL or approximately $4.02x10^8$ copies/mL) which was estimated as $2.2x10^8$ copies/mL by qPCR.

Two qPCR tests were performed on mock swab samples. The first test aimed to establish where in the sample the most extractable virus was located. RNA was extracted from three components of mock swabs (swab, supernatant, and pellet). Three extraction replicates

were performed for each component of swabs which had been inoculated at a concentration of 10^5 copies/mL. All extraction replicates were quantified by qPCR as described above in triplicate along with standards and no template controls.

The second test aimed to approximate the minimal detectable viral concentration by qPCR using this method, to assess repeatability, and to estimate extraction efficiency using the cotton-tipped wooden base swabs and rayon-tipped aluminum base swabs used to collect samples in the field. Three extraction replicates were performed for each concentration from 10^6 – 10^3 copies/mL for cotton-tipped wooden base swabs and three extraction replicates were performed for aluminum base swabs at 10^5 copies/mL. RNA was extracted from swabs, converted into cDNA, and quantified by qPCR as described above.

2.3.8 Pilot study 3: Is rRNA depletion a useful enrichment method for characterizing viral communities?

The effect of rRNA depletion on the number of viral reads and viral taxa detected was evaluated using swabs from 40 fecal and 10 saliva samples, which were taken from vampire bats captured at 16 sites in 7 departments (administrative regions) across Peru (Figure 2.1) between 2015-2016. Swab samples were extracted individually, quantified and pooled as described in Appendix B1.1. Five pools were created using nucleic acid extracts from the same sample type from 10 individuals across 1-2 sites in the same locality (between 0.14 -74.1 km apart) within each department of Peru (Table 2.5; Figure 2.1; Table A1).

Table 2 5 Multi colony pools sequenced for enrichment tests and subsampling. Pools were created by combining nucleic acid from 10 individual swabs of the same sample type from the same site or locality. All pools were sequenced using a shotgun metagenomic approach but different forms of enrichment were tested on different pools.

Pool ID†	Sample Type	Raw Reads	Viral Reads	Colony 1‡	Colony 2‡	Test§	Treatment
AAC_H_F*	Feces	12,166,001	10,870	AYA7	AYA14	-	-
AAC_H_SV*	Saliva	9,507,979	431	AYA7	AYA14	-	-
AAC_L_F*	Feces	12,000,988	2,417	API1	AYA11	-	-
AAC_L_SV*	Saliva	15,121,355	609	API1	AYA11	-	-
AMA_L_ F_NR	Feces	17,827,799	2,062	AMA2	AMA6	rRNA	Non-enriched
AMA_L_F_R	Feces	17,760,709	28,344	AMA2	AMA6	rRNA	Enriched
AMA_L_SV*	Saliva	9,363,273	305	AMA2	AMA4	-	-
CAJ_L_F_NR	Feces	15,940,753	1,179	CAJ4	-	rRNA	Non-enriched
CAJ_L_F_R	Feces	15,843,806	5,945	CAJ4	-	rRNA	Enriched
CAJ_L_SV*	Saliva	8,685,456	600	CAJ4	-	-	-
CAJ_H_F_1*	Feces	8,661,617	8,085	CAJ1	CAJ2	DNAse	Light
CAJ_H_F_2*	Feces	9,272,152	8,187	CAJ1	CAJ2	DNAse	Harsh
CAJ_H_SV*	Saliva	11,830,542	534	CAJ1	CAJ2	-	-
HUA_H_F*	Feces	10,814,816	11,285	HUA1	HUA2	-	-
HUA_H_SV*	Saliva	8,931,393	517	HUA1	HUA2	-	-
LMA_L_F_NR	Feces	19,605,605	1,425	LMA5	LMA6	rRNA	Non-enriched
LMA_L_F_R	Feces	17,365,381	8,206	LMA5	LMA6	rRNA	Enriched
LMA_L_SV_NR	Saliva	18,698,730	75	LMA5	LMA6	rRNA	Non-enriched
LMA_L_SV_R	Saliva	15,953,442	483	LMA5	LMA6	rRNA	Enriched
LR_L_F_NR	Feces	19,531,234	1,535	LR1	LR2	rRNA	Non-enriched
LR_L_F_R	Feces	13,843,629	4,544	LR1	LR2	rRNA	Enriched
LR_L_SV*	Saliva	9,023,821	478	LR1	LR2	-	-

†All Pool IDs reflect the locality (AAC, Ayacucho-Apurímac-Cusco; AMA, Amazonas; CAJ, Cajamarca; HUA, Huánuco; LMA, Lima; LR, Loreto) and sample type (F, feces; SV, saliva). Some IDs also reflect elevation (H, high; L, low) to differentiate localities with multiple pools. NR and R correspond to ribosomal treatment, either non-enriched or enriched, and one sample (CAJ_H_F) has associated numbers (1 and 2) referring to two batches that received different DNAse treatments during viral enrichment. Pools processed using the final protocol are shown in bold.

‡Colony codes correspond to department within Peru. Colony locations and pool midpoints are shown in Figure 2.1.

§Enrichment tests are abbreviated as rRNA (ribosomal RNA depletion) and DNAse (light or harsh DNAse treatment)

*Pool IDs with an asterisk are included in subsampling analyses
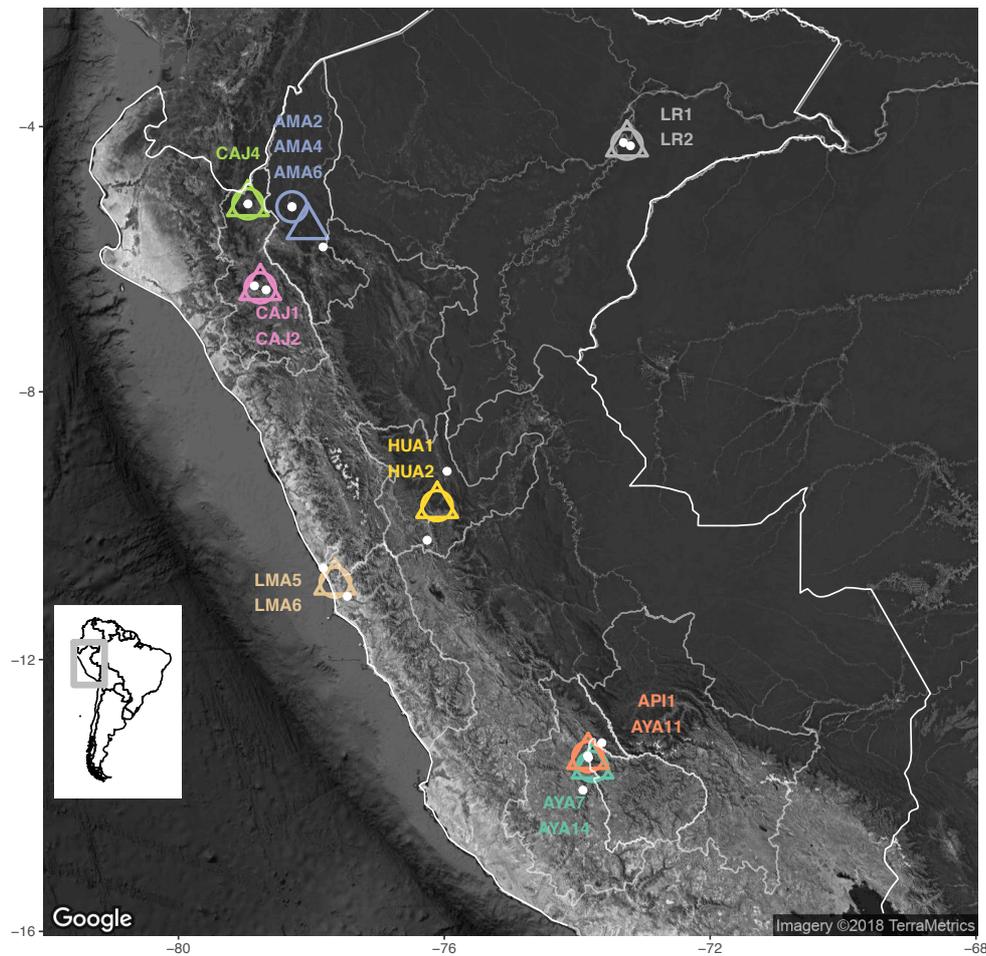
Figure 2 1 Vampire bat colonies and pools used in enrichment tests and subsampling. Individual colonies are represented as white points and midpoints for each pool, in which 1-2 colonies were combined, are represented as circles (feces) or triangles (saliva). Colony names are shown in the same color as the pools in which they are included. Peru country borders and departments within Peru where samples were collected are outlined in white. The inset map shows South America, with Peru highlighted in the gray box.

Pools were treated with DNAse I (Ambion), with buffer and enzyme scaled such that all reactions contained 1X DNAse buffer and 2U DNAse per 100 μL. Reactions were incubated at 37°C for 5 minutes, then cleaned up with 1.8X Agencourt RNAClean XP beads, eluted in RNAse-free water, and split in half. Half of each DNAse treated pool was enriched by rRNA depletion using the Ribo-Zero rRNA Removal Kit (Human/Mouse/Rat) (Illumina) according to the manufacturer's instructions, while the other half was library prepared directly, such that two libraries were prepared from each initial pool for a total of ten libraries.

cDNA synthesis and library preparation were performed as described in Appendix B1.3 with a variable number of PCR cycles: 12 cycles were used for non-enriched samples and 16 cycles were used for enriched samples. As rRNA depletion significantly decreased the

quantity of nucleic acid, increased PCR cycles were necessary to generate sufficient material for sequencing for enriched samples; however, this difference is not expected to influence the proportion or composition of viral reads. Final libraries were quantified, pooled, and sequenced (Appendix B1.3) and processed through the bioinformatic pipeline (Appendix B2).

2.3.9 Pilot study 4: Does intensive DNAse treatment further enrich viral communities?

A more intensive DNAse treatment was also tested for its effect on the number of viral reads and viral taxa detected. Fecal swabs from 10 individuals across 2 sites in the Cajamarca Department (Table 2.5; Figure 2.1; Table A1) were extracted, quantified and pooled (Appendix B1.1). The sample was split in half after pooling; one half was subjected to "light" treatment of 2U DNAse and incubated at 37°C for 5 minutes (as above), and the other half was subjected to "harsh" treatment of 10U DNAse and incubated at 37°C for 15 minutes. Both halves were then cleaned up using a 1.8X ratio of Agencourt RNACleanXP beads. Following this step, pools were library prepared and sequenced according to the final protocol (Appendix B1.3) and processed through the bioinformatic pipeline (Appendix B2).

2.3.10   Subsampling analysis of viral community saturation using the optimized sequencing protocol

A subsampling analysis was conducted to test whether observed variation in the number of raw sequencing reads (Table 2.5) would affect the viral community detected (i.e. the number of viral reads, viral taxa and vertebrate-infecting viral taxa). The datasets analyzed included 12 multi-colony pools (5 fecal, 7 saliva; Table 2.5) that had been sequenced according to the final protocol. Fecal and saliva pools contained swabs from individuals from the same colony or colonies, except in the Amazonas Department where saliva pools contained individuals from sites AMA2 and AMA4, but fecal pools contained individuals from sites AMA2 and AMA6. Subsampling comprised randomly selecting raw reads at every 10 percent between 10 to 100 percent of the total reads and was repeated five times per pool. Viruses from subsampled datasets were classified using the bioinformatic pipeline without the assembly step (Appendix B2).

A generalized linear mixed model (GLMM) with a Poisson distribution was used to assess the effect of the percentage of raw reads sampled on the number of viral taxa (families and

genera) detected using the lme4 package of R (Bates *et al.* 2015). Separate models were constructed for each combination of sample type (fecal and saliva), filtering condition (all viruses and vertebrate-infecting), and taxonomic level (family and genus). The percentage of the total raw reads sampled was standardized by subtracting the mean and dividing by the standard deviation of percentages, and pool ID was included as a random effect in the model. For each dataset, linear and second-degree polynomial models were tested and compared using a likelihood ratio test and the change in Akaike information criterion (ΔAIC), with a better fitting polynomial model indicating a plateau in the number of viral taxa detected at attained read depths.

2.4 Results

2.4.1 Metagenomic sequencing reveals diverse viral nucleic acid in FBS (Pilot study 1)

A total of 21,501,182 raw reads were generated from metagenomic sequencing of the two batches of FBS. The bioinformatic pipeline detected 1,373 and 516 viral reads in each batch respectively, which spanned 14 families of RNA and DNA viruses (Table 2.6). In both samples, the majority of viral reads were assigned to the family *Flaviviridae*, with 41% and 30% of viral reads for the two FBS batches respectively assigned to bovine viral diarrhea virus 3 (BVDV-3). Contigs matching to BVDV (the longest were 1,396 and 775bp respectively, out of a full genome of around 12,000bp) had 96-98% identity to strain SV757/15 of BVDV-3 (Table B2).

Table 2 6 Viral families detected from shotgun metagenomic sequencing of FBS.
For each viral family, the number of reads and contigs are reported for each of two batches of FBS which were analyzed.

	FBS1‡		FBS2‡	
	Reads	Contigs	Reads	Contigs
Adenoviridae	27	0	40	2
Asfarviridae	2	0	0	0
Myoviridae	52	2	10	0
Podoviridae	29	0	47	5
Siphoviridae	73	4	32	2
Herpesviridae	2	2	6	1
Iridoviridae	1	0	0	0
Polydnaviridae	4	0	0	0
Poxviridae	9	0	0	0
Retroviridae	180	20	104	15
Microviridae	2	0	2	0
Nyamiviridae	0	0	1	0
Flaviviridae	950	15	267	11
Alphaflexiviridae	8	0	0	0
Total viral reads†	1373		516	
Raw reads	13,565,793		7,935,389	

‡FBS1 and FBS2 were two different batches of FBS that were sequenced
†Number of reads assigned to families do not add up to the total number of viral reads as some were classified as viral but not assigned to a family

2.4.2 Viral sequences are maximized by extracting RNA from intact swabs (Pilot study 2)

For swabs stored in RNALater, extracting directly from the swab itself yielded viral nucleic acid that was measurable by qPCR, while supernatant and pellet did not (data not shown). The limit of detection occurred with swabs that were initially inoculated with 220 viral copies; at this level, virus was inconsistently detectable by qPCR (Table 2.7). Virus became consistently detectable at 2,200 copies inoculated into the swab. Of the three aluminum-base swabs that were inoculated with 2,200 copies, two of the extractions contained undetectable virus in all three qPCR replicate reactions; potentially because these swabs were smaller, and it was difficult to determine whether the virus had absorbed into the rayon. However, the one aluminum-base swab with measurable virus was comparable in final copy number to the wooden-base swabs (Table 2.7). The qPCR replicates were generally consistent aside from samples on the edge of detectability, but Ct and copy number varied between extraction replicates of swabs containing the same initial quantity of virus. For the swabs inoculated with 2,200 copies (aluminum and wooden-base), there were on average 1,230 copies present following RNA extraction, yielding an

extraction efficiency of about 56% (there were 1,578 copies and 72% efficiency when excluding an outlier wooden-base swab replicate that had 0.94 qPCR copies).

Table 2 7 Summary of mock swabs tested for different extraction methods using qPCR. Swabs were inoculated with Schmallenberg virus and final virus concentration following extraction was measured using qPCR for different swab types and initial quantities of virus.

Swab type	Virus concentration (copies/mL) ‡	Initial swab quantity (copies) ‡	Extraction Replicate	Average Ct (SD)§	Average qPCR copies (SD)§
Wooden-base	10^4	220	1†	37.44 (0.45)	0.67 (0.20)
			2	No Ct	No Ct
			3†	36.69 (0.72)	1.16 (0.55)
Wooden-base	10^5	2,200	1	33.86 (0.36)	7.84 (1.94)
			2	33.74 (0.6)	8.83 (3.79)
			3	36.98 (0.57)	0.94 (0.37)
Aluminum-base	10^5	2,200	1	34 (0.12)	7 (0.59)
Wooden-base	10^6	22,000	1	33.49 (0.35)	10.13 (2.4)
			2	31.72 (0.13)	33.70 (2.86)
			3	32.92 (0.32)	14.90 (3.06)

†Indicates only two of the three qPCR replicates were measurable (one replicate was below the limit of detection). When all three qPCR replicates were below the limit of detection, this is indicated with No Ct. All other average Ct and average qPCR quantities are calculated based on three qPCR replicates.
‡Virus concentration and initial swab quantities were calculated based on qPCR measurements of undiluted virus, which was then diluted to obtain the concentrations used in this experiment.
§SD – Standard deviation

2.4.3 Viral enrichment is improved through rRNA depletion (Pilot study 3)

The sequenced fecal and saliva samples that were split and trialed for rRNA depletion yielded a total of 172,371,088 reads which were evenly distributed across samples (Table 2.5). Samples that were enriched contained on average 8,213 more viral reads (Figure 2.2A), with this difference being close to statistically significant (paired Wilcoxon signed rank test, p= 0.06) despite the small sample size (N=10). On average, there were 9 more viral families (paired Wilcoxon signed rank test, p=0.058) and 3.8 more vertebrate-infecting viral families (paired Wilcoxon signed rank test, p=0.06) per sample in enriched samples (Figure 2.2B).
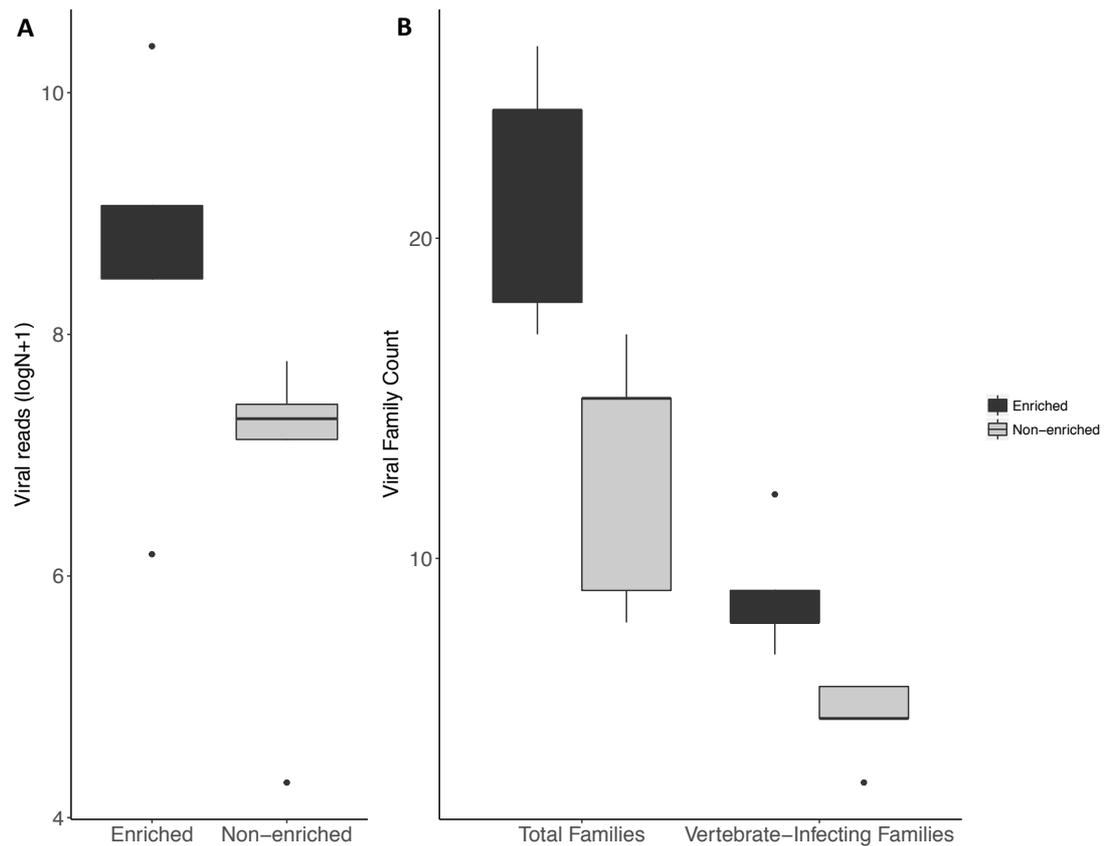
Figure 2 2 Comparison of viral reads and families in ribosomal depletion enrichment.
(A) Number of viral reads, shown as log (N +1), compared between enriched (N=5) and non-enriched (N=5) samples. (B) Total viral families and vertebrate-infecting viral families detected in samples enriched by rRNA depletion (N=5) compared to non-enriched samples (N=5).

Within vertebrate-infecting viral families, number of reads per family was higher in enriched samples with the exception of the family *Retroviridae* (Figure 2.3). Vertebrate-infecting viral families that were detected only after enrichment exhibited diverse genome structure including positive sense, single-stranded RNA (*Astroviridae, Nodaviridae*), negative sense, single-stranded RNA (*Rhabdoviridae, Paramyxoviridae*), double-stranded RNA (*Picobirnaviridae*) and double-stranded DNA (*Poxviridae*). Similar patterns were observed for all viruses, not just those infecting vertebrates, and results were consistent when analyses were repeated at the level of viral genera (data not shown). In summary, removal of host rRNA allowed detection of more viral taxa (Figure 2.2B) and improved the sequencing depth for detected viruses (Figure 2.3).
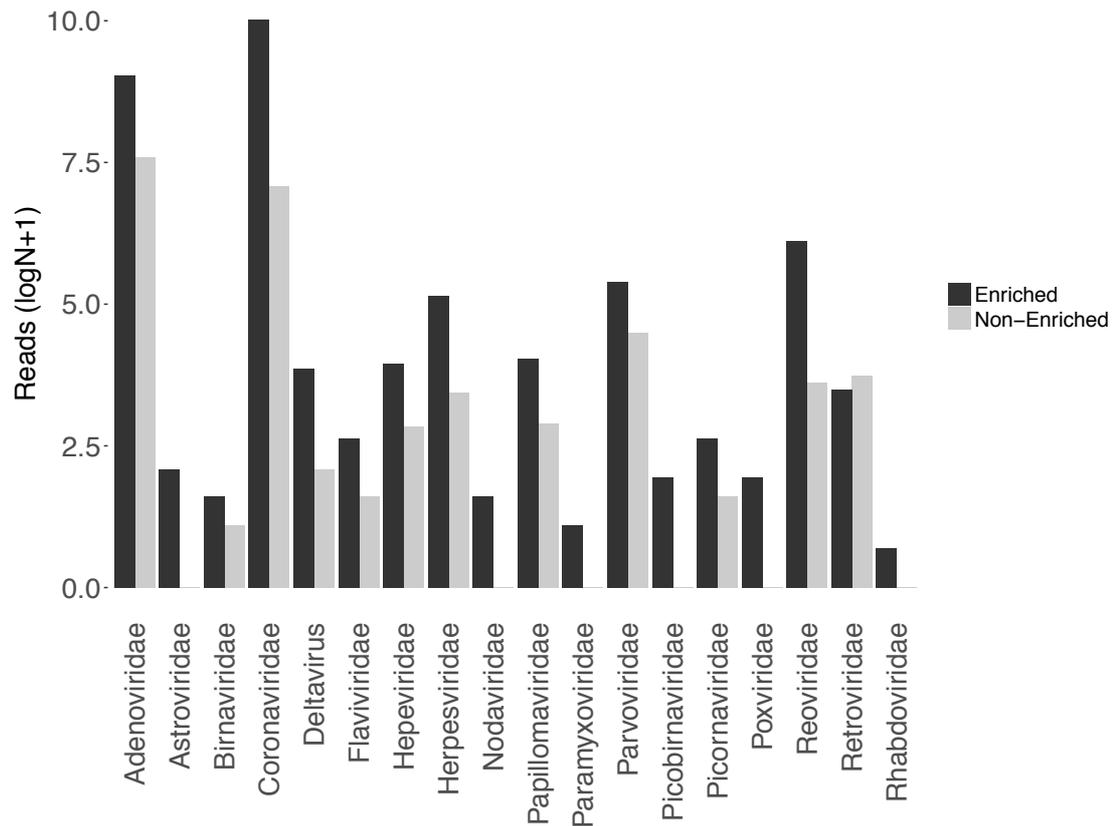
Figure 2 3 Comparison of reads per vertebrate-infecting viral family across samples. Comparisons are shown for reads per vertebrate-infecting viral family summed across samples enriched by ribosomal depletion (N=5) and non-enriched (N=5). Read number comparison is shown for summed reads (as opposed to the mean) to enable visualization on a log scale.

2.4.4 Viral enrichment is improved through light DNAse treatment (Pilot study 4)

The fecal sample that was split and trialed for light/harsh DNAse treatment yielded 17,933,769 reads that were evenly distributed across the two treatment pools (Table 2.5). Although the number of viral reads was comparable between the two pools, light DNAse treatment increased the taxonomic richness of viruses detected, both for all viruses and vertebrate-infecting viruses (Figure 2.4).
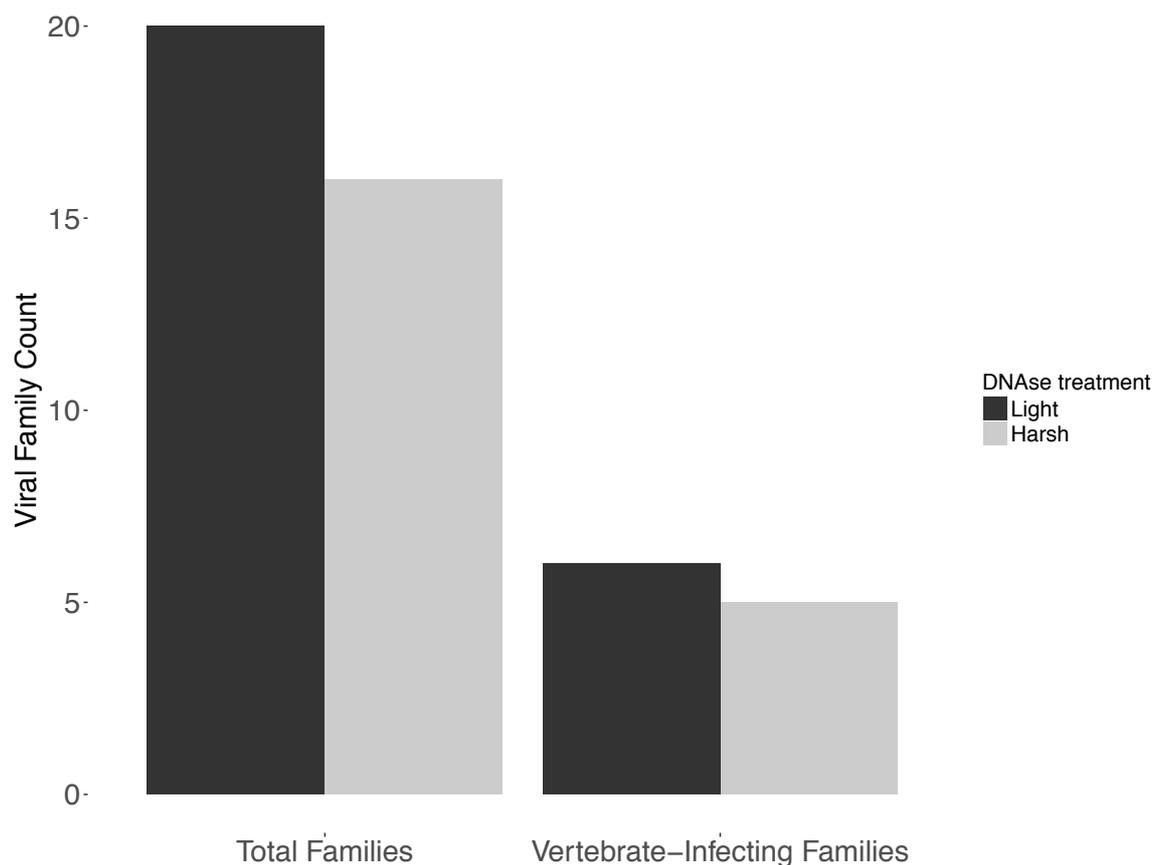
Figure 2 4 Comparison of number of viral families between DNAse treatments. Comparisons show a single split sample, with half receiving light DNAse treatment and half receiving harsh DNAse treatment.

The proportion of low complexity/PCR duplicate reads was also slightly higher in the harsh DNAse treatment (1,974,128 reads) compared to the light treatment (1,620,909 reads) (Figure B4). Low complexity reads in a metagenomic sample could originate from various sources and are difficult to interpret, but increased PCR duplicates suggest that the harsh DNAse treatment created a less diverse pool of nucleic acid prior to re-amplification, although it is difficult to draw conclusions based on only two pools. Viral families that were absent in the harsh treatment included those with single-stranded DNA genomes (*Circoviridae*), as well as single-stranded RNA genomes (*Flaviviridae*), suggesting that DNAse treatment may also degrade RNA viruses. However, RNA viruses were not always affected negatively by DNAse treatment, as the single-stranded RNA family *Paramyxoviridae* was present in the harsh treatment but not the light treatment. *Paramyxoviridae* was only represented by two reads in the harsh treatment so it could be a rare virus that was missing from the light treatment due to chance, but the effects of DNAse on different viral genome types appear complex and may require further study to resolve. Although only two pools were compared and they contained similar numbers of

viral reads, a greater diversity of viral families was detected following the light DNAse treatment.

2.4.5 Summary of samples sequenced using the optimized metagenomic protocol

Pooled samples processed according to the final protocol had similar numbers of raw reads, but the proportion of viral reads varied widely across samples (Table 2.5). Saliva samples consistently contained fewer viral reads than fecal samples. The proportion of reads filtered out during different stages of bioinformatic processing was fairly similar across samples (Figure B4), and detected sequences matched to vertebrates, arthropods, bacteria, and archaea in addition to the viral sequences that were the focus of the study (Figure B5).

2.4.6 Subsampling validates viral community saturation using the optimized protocol

The number of viral reads increased consistently with the number of raw reads, as would be expected with unbiased sequencing, though rate of increase differed among pools (Figure 2.5).
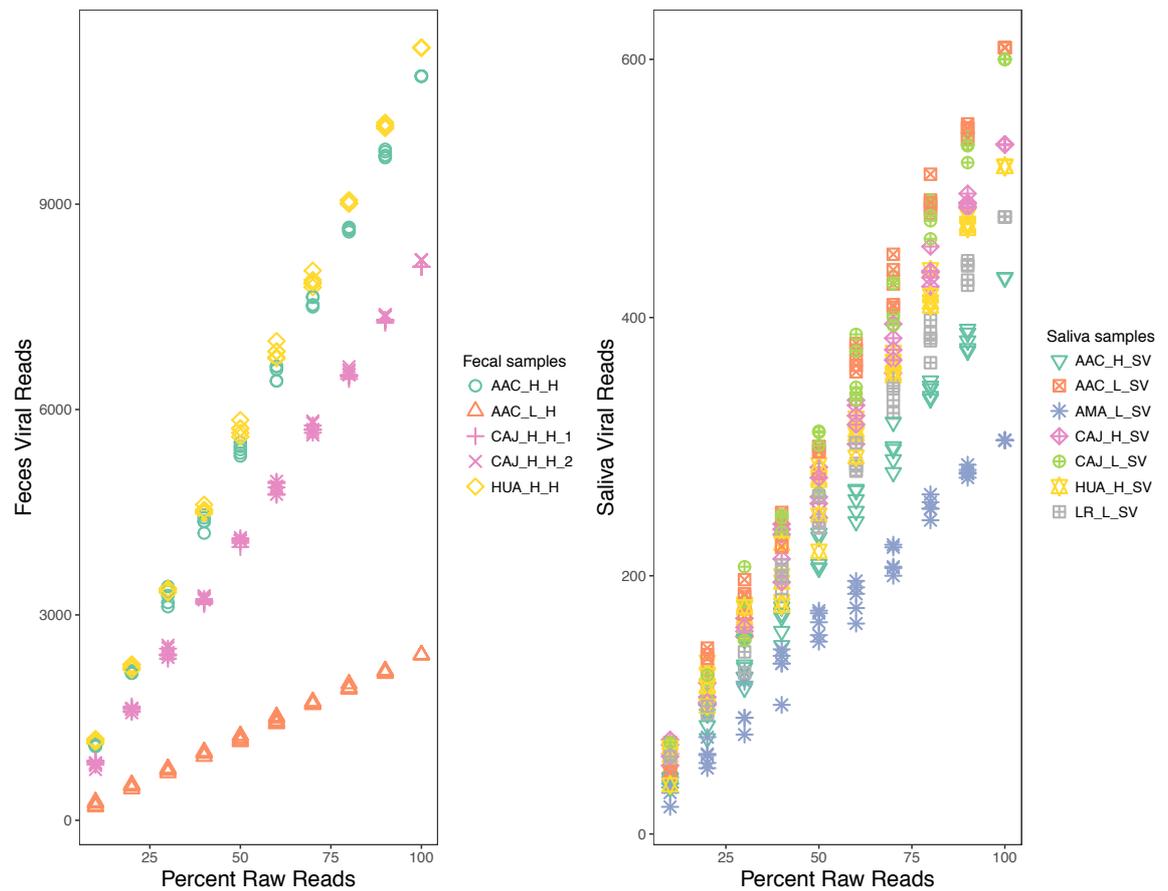
Figure 2 5 Viral reads increase proportionally to percentage of raw reads.
The number of reads assigned as viral for fecal (N=5) and saliva (N=7) samples are shown at increasing percentages of total raw reads. Five replicates of each sample are depicted using the same symbol and color; colors correspond to localities shown in Figure 2.1.

In contrast, the number of viral families and vertebrate-infecting viral families plateaued at higher percentages of the total number of raw reads sampled (Figure 2.6), and models explaining the number of viral families with a second-degree polynomial effect of percentage of raw reads generally fit the data better than linear models (Table 2.8). These results suggest that at the level of sequencing depth achieved, common viral families and genera are reliably identified.
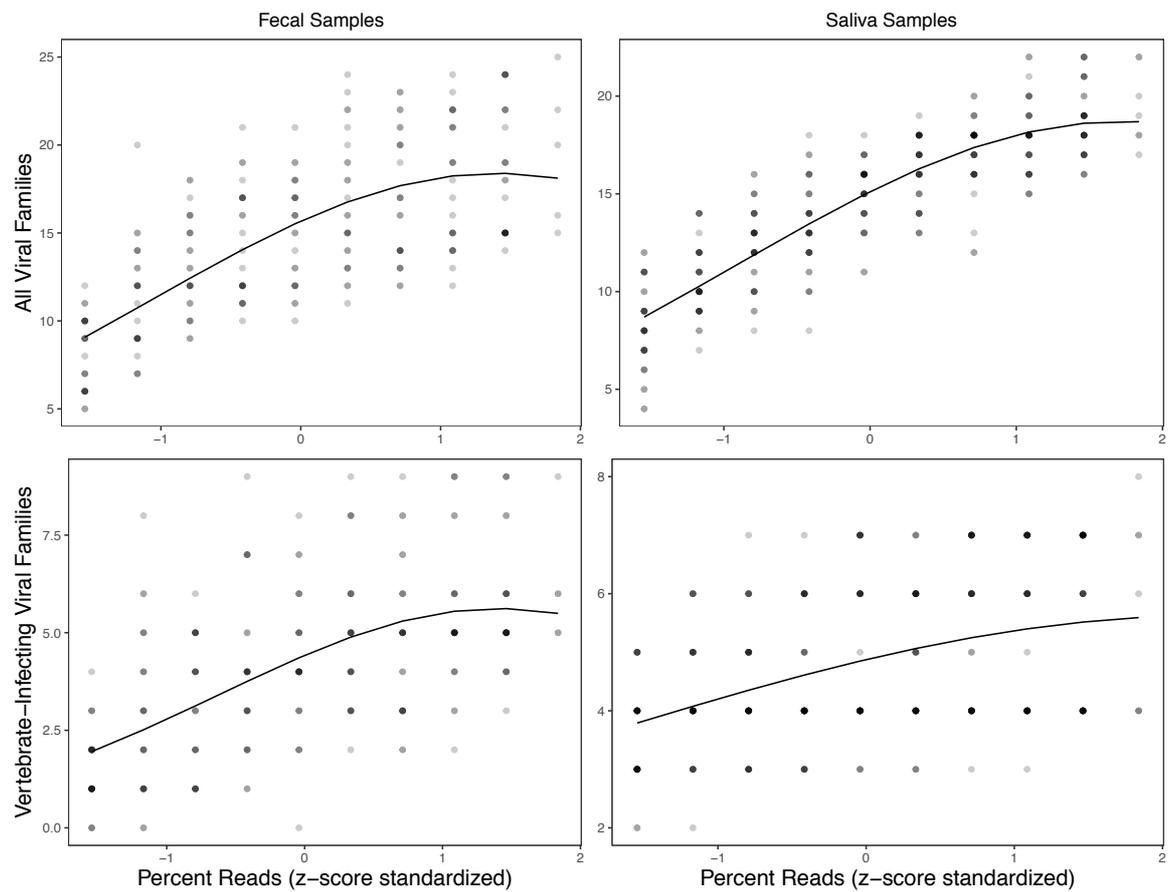
Figure 2 6 Viral family communities saturate at high read depths.
Panels show the number of viral families and vertebrate-infecting viral families detected in fecal (N=5) and saliva (N=7) samples at increasing percentages of the total raw reads. Percentage reads are z-score standardized by subtracting the mean and dividing by the standard deviation. Points are semi-transparent, with darker points indicating more subsamples with a given value, and show the rescaled original data. Lines show the model prediction.

Table 2 8 Model comparison for viral family detection in subsampling analysis.
Linear and polynomial models were compared for each sample type (feces and saliva) and filtering (all viral families and vertebrate-infecting only) combination at the family level. For each combination, two models were constructed and compared through likelihood ratio test (L, X^2, d.f., and P-value) and AIC (AIC and ΔAIC).

	Model	L	X^2	d.f.	P-value	AIC	ΔAIC
Viral Families Fecal	Linear	-556.1	17.271	1	3.24E-05	1118.2	15.271
	Polynomial	-547.47				1102.9	
Viral Families Saliva	Linear	-772.02	18.304	1	1.88E-05	1550	16.304
	Polynomial	-762.87				1533.7	
Vertebrate-infecting viral Families Fecal	Linear	-407.39	10.356	1	0.00129	820.79	8.3564
	Polynomial	-402.22				812.43	
Vertebrate-infecting viral Families Saliva	Linear	-573.15	0.8262	1	0.3634	1152.3	1.174
	Polynomial	-572.73				1153.5	

The exception was vertebrate-infecting viral families detected in saliva; however, detections did plateau at the viral genus level (Table B3, Figure B6). Aside from vertebrate-infecting viral families in saliva, viral richness plateaued at around 80% of the total reads (Figure 2.6; Figure B6). Converting the percentage of total reads at which the plateau typically occurred (80%) to the number of raw reads indicated that, on average, new detections began to level off at 8,358,626 reads (range: 6,929,294 - 12,097,084).

2.5 Discussion

A field-laboratory-bioinformatic protocol was developed for characterizing viral communities, incorporating the following findings from pilot studies to maximize viral detections:

1. Swab samples should be stored in RNALater rather than VTM containing FBS

2. Nucleic acids should be extracted directly from swabs, rather than from supernatant or pellet

3. Enrichment should use rRNA depletion and light DNAse treatment

The metagenomic pipeline yielded viral community data from swab samples taken from vampire bats across Peru, and detections in most cases plateaued within commonly attained levels of sequencing depth (Figure 2.6), suggesting that this is an effective non-invasive method for sampling viral communities from field samples collected from wildlife. The field protocol standardizes sample collection, storage and transportation among geographically widespread and remote study sites. The laboratory and bioinformatic protocols aim to capture and identify as many different types of viruses as possible, while processing large batches of samples and avoiding well-known sources of bias.

Metagenomic sequencing revealed diverse bovine viral nucleic acid in FBS. Importantly, these results are unlikely to indicate the presence of live viruses in FBS since commercial FBS is often heat inactivated and screened for live viruses. Instead these detections probably represent viral nucleic acids which persist after heat inactivation, but which could nevertheless impact metagenomic studies. Detecting BVDV is unsurprising, as it is a common cell culture contaminant that has previously been found in high quantities in FBS

(Allander *et al.* 2001; Gagnieur *et al.* 2014). Consistent with the South American origin of the FBS used in analyses, BVDV-3, or HoBi-like viruses, were initially reported in FBS from South America and are likely endemic to livestock in Brazil (Bauermann & Ridpath 2015). The consistent presence and proportion of BVDV as well as *Retroviridae* and several bacteriophage families (Table 2.6) across FBS batches suggests that this source of contamination could perhaps be accounted for in order to include VTM samples containing FBS in metagenomic analyses. However, reads in FBS also matched the family *Adenoviridae* (genus *Mastadenovirus*), which are common in bats (Li *et al.* 2010b; Drexler *et al.* 2011), including neotropical species (Wray *et al.* 2016). If bat samples stored in VTM were sequenced and filtered for viral genera detected in FBS, this would potentially exclude true bat viruses. The results therefore suggest that metagenomic results from historical samples stored in media containing FBS should be interpreted with caution and avoided where possible.

Using artificially inoculated swabs, the comparison of RNA extractions from different components of samples (swab, supernatant, pellet), showed that swab extraction, but not extraction from supernatant or pellet, typically yielded measurable nucleic acid. This could be due to the high salt concentrations in RNALater that are designed to inhibit RNAse activity, but which could also interfere with extraction from the supernatant/pellet. Typically, tissues stored in RNALater are blotted to remove excess solution, and other samples such as blood are centrifuged and the supernatant is removed prior to extraction. Unfortunately, it is not possible to completely remove the RNALater from swabs, but extracting from the swab itself might minimize salts relative to the other components of the sample. It is also possible that virus particles mostly remain within the swab itself when stored in RNALater.

Direct extraction from swabs was previously used to characterize bacterial communities from swabs stored in RNALater (Vo & Jedlicka 2014), and other studies have released particles bound to swabs through incubation in lysis buffer (Schweighardt *et al.* 2014) or lysis buffer and Proteinase K (Ghatak *et al.* 2013; Corthals *et al.* 2015). It may not be possible to extrapolate the estimated limit of detection or extraction efficiency to other viruses with different characteristics, or field-collected samples that include host cells and other material. In addition, the quantity of viral RNA extracted from swabs did not appear highly repeatable between extraction replicates. However, the results indicated that extracting directly from the swab improved viral detection relative to other components of the sample.

The study tested a variety of laboratory methods for enhancing unbiased detection of viruses. The rRNA depletion results suggested that removing host rRNA increased both the number of viral reads and number of viral taxa detected, without biasing the viral community, as has been observed in previous studies (He *et al.* 2010; Matranga *et al.* 2014). The only case in which there were more reads in the non-enriched samples was the family *Retroviridae*, however, retroviruses integrate into the host genome and are likely to behave differently than other viral taxa with respect to enrichment. Although the Ribo-Zero kit is described as being for Human/Mouse/Rat and should be tested before use on other sample types, it has been used effectively on samples from taxa as distantly related as mosquitos (Weedall *et al.* 2015), and it was found here to be effective for enriching samples taken from bats.

Although only one split sample was analyzed, the light DNAse treatment results suggested an increase in the number of viral taxa detected compared to the harsh treatment. DNAse is a well-established method to reduce the number of host and bacterial reads relative to virus (Allander *et al.* 2001). The light treatment was intended to knock down rather than remove all DNA, also potentially allowing for better detection of bacteria and parasites compared with an intensive enrichment, although this was not tested explicitly. Although this step could have caused bias towards RNA viruses, DNA virus reads occurred in all samples, as has been found in other viral metagenomic studies using an RNA-based approach (Hall *et al.* 2014; Kohl *et al.* 2015; Wu *et al.* 2016), including those with a DNAse treatment step (Baker *et al.* 2013; Hall *et al.* 2014). This could be explained by the presence of viral RNA transcripts, DNA viruses that replicate through an RNA intermediate (e.g. *Hepadnaviridae*), the ability of some DNA virus families to integrate into the genome of their host (e.g. *Herpesviridae*) or DNA being carried through the DNAse treatment into library prep due to the light treatment or less than perfect efficiency of the reaction. Although more intensive enrichment such as filtration or centrifugation could potentially have increased the number of viral reads, such methods are known to be biased against certain taxa (Kleiner *et al.* 2015; Wood-Charlson *et al.* 2015; Conceição-Neto *et al.* 2015). In addition, it would be impossible to include a filtration step since swabs were immediately treated with lysis buffer in the extraction, leading to lysis of the viral particles which would normally be selected for using filtration. In light of the results above, and despite the relatively small number of samples, light DNAse treatment and rRNA depletion are recommended as an effective combination for viral enrichment.

It is worth noting the caveats of analyzing non-invasively collected samples. First, although contamination has not been well characterized in viral metagenomic studies, it is a known problem in bacterial community studies. Samples with low microbial biomass are particularly sensitive to contamination with other microbes, for example from DNA extraction kits (Salter *et al.* 2014) or ultrapure water (Laurence *et al.* 2014). This protocol minimized this risk by pooling samples following extraction to increase the amount of target nucleic acid relative to potential reagent-derived contaminants in downstream steps. Second, non-invasive samples will only detect viruses that are actively shed in urine and feces, thus may miss latent viruses that are sporadically shed, but might be detectable by sequencing organs from sacrificed animals (Amman *et al.* 2012). Third, the protocol is not able to discriminate between viruses actively infecting hosts and transient viruses acquired from diet or the environment.

Although some sources of bias are unavoidable, and it is likely that not all viral taxa in a given sample will be identified, the same is true of all studies in community ecology where exhaustive sampling is not possible (Gotelli & Colwell 2001; Hughes *et al.* 2001), and viral communities were statistically shown to be adequately sampled (Figure 2.6). This approach yielded sufficient depth to confidently characterize viral communities at the viral family or genus level, while identification of species or strains might be achieved by further increasing read depths to generate longer contigs that could be more precisely assigned (Figure 2.5). Although a relatively small number of pools was examined here, the protocols described have been developed to enable scaling up to the larger numbers of samples typical of ecological and evolutionary studies. Performing extraction on an automated platform (e.g. the Kingfisher) allows high throughput nucleic acid extraction from swabs, the use of kits enables preparation of up to 96 libraries simultaneously, and the bioinformatic pipeline developed automates analyses of large amounts of sequence data. Future studies of viral communities could also consider the use of sequence-capture based approaches (e.g. Briese *et al.* 2015; Wylie *et al.* 2015), which have the potential to provide more on-target reads than shotgun metagenomics, although potentially at the expense of discovering novel or highly divergent taxa.

In summary, the pipeline described in this chapter simultaneously generated comparable viral communities from large numbers of non-invasively collected wildlife samples. A standardized approach to viral metagenomics opens many potential avenues of future research in disease and community ecology. For example, viral community data collected across multiple individuals, populations, and species allows the investigation of ecological

processes shaping host-associated viral community structure (Anthony *et al.* 2015; Olival *et al.* 2017). Taxonomic and functional patterns of bacterial diversity across host species are influenced by diet and phylogeny (Ley *et al.* 2008; Muegge *et al.* 2011; Zepeda-Mendoza *et al.* 2018), but drivers of host-associated viral communities may be different. In humans, viral communities are stable over time within individuals, but highly variable between individuals (Reyes *et al.* 2010; Minot *et al.* 2011). These observations suggest the potential to use viral communities as a host or environmental "fingerprint" to evaluate interactions between multiple hosts, or between hosts and environments, as has been proposed in humans and primates (Fierer *et al.* 2010; Franzosa *et al.* 2015; Stumpf *et al.* 2016). Finally, although it was not the focus of the study, reads were also detected from vertebrates, protozoa, and bacteria (Figure B5), suggesting that with appropriate bioinformatic modifications, shotgun metagenomic data generated using this protocol could simultaneously shed light on host genetics, diet, other non-viral pathogens, and commensal microbes. As metagenomics becomes an ever more popular and powerful tool for viral ecology, the use of standardized methods such as those developed here will be crucial for comparative insights from diverse host species and environments.

3 Metagenomics reveals the diverse RNA and DNA virus communities in common vampire bats across Peru

3.1 Abstract

Establishing a detailed understanding of viral diversity in key wildlife hosts represents an important first step in evaluating the risk of zoonotic disease emergence. As obligate blood feeders, common vampire bats (*Desmodus rotundus*) are an example of a wildlife host that poses a threat for disease transmission to humans and domestic animals, but we currently lack a holistic understanding of their viral communities. Therefore, this chapter aimed to thoroughly characterize viruses in vampire bat feces and saliva from individuals captured across Peru, evaluating differences in viral communities between body habitats and using phylogenetic analyses to assess whether novel viral taxa were most closely related to other bat-infecting taxa. A minimally biased shotgun metagenomic sequencing approach was used to describe vampire bat viral communities, which comprised both vertebrate-associated taxa potentially infecting vampire bats as well as viral taxa that typically infect bacteria, plants or insects. Viral communities in vampire bat feces and saliva were distinct, showing evidence of body habitat compartmentalization. Novel viruses from eight vertebrate-infecting families were phylogenetically analyzed in the context of previously characterized taxa, revealing that vampire bat viruses frequently fell into bat-specific clades, such as full genomes of novel viruses in the families *Coronaviridae* and *Hepeviridae* which were related to other bat-infecting taxa. However, full genomes of a divergent *Hepatitis delta*-like virus were also detected from vampire bat saliva at three sites, representing the first discovery of this virus outside of humans. These results suggest that most vampire bat viruses fall within bat-specific clades, without evidence of livestock or humans acting as a major source of viral diversity in vampire bats, adding to our understanding of viral diversity in an ecologically important bat species.

3.2 Introduction

Bats act as important hosts to a number of high-profile zoonotic viruses (Li *et al.* 2005; Leroy *et al.* 2005; Memish *et al.* 2013), exhibiting deep evolutionary relationships with a broad range of viral taxa (Cui *et al.* 2007; Taylor *et al.* 2010; Tong *et al.* 2012; Drexler *et al.* 2012a; Quan *et al.* 2013; Drexler *et al.* 2013; Sasaki *et al.* 2014; Escalera-Zamudio *et al.* 2016). All other factors being equal, a reservoir host with higher viral richness overall is

also likely to have higher richness of viral taxa that can infect other species, or a larger "zoonotic pool" (Morse 1993). Understanding the viral richness of key host species, particularly those exhibiting high contact rates with other species, is therefore important in evaluating the risk of zoonotic viral emergence (Turmelle & Olival 2009; Olival *et al.* 2017).

As an obligate blood feeding species, common vampire bats represent a particularly interesting system in which to investigate viral richness because of their high level of ecological connectivity. Vampire bats historically preyed upon wildlife such as tapirs and peccary, while more recently their diet has shifted towards mammalian livestock (Delpietro *et al.* 1992; Voigt & Kelm 2006; Streicker & Allgeier 2016), humans (Schneider *et al.* 2001; Gonçalves *et al.* 2002; Schneider *et al.* 2009), and birds (Bobrowiec *et al.* 2015). Feeding on a variety of other vertebrates creates the opportunity for exposure and infection of vampire bats by the viruses that infect their prey. Vampire bats also share roosts with bat species such as those in the genera *Micronycteris, Glossophaga, Carollia, Sturnira, Saccopteryx,* and *Artibeus* (Greenhall *et al.* 1983), with roost-sharing providing another potential route of pathogen transmission (Reckardt & Kerth 2007; Leu *et al.* 2010) from other bat species to vampire bats. Vampire bats exhibit traits common to bats generally which have been speculated to facilitate virus transmission, including long lifespan and colonial roosting (Calisher *et al.* 2006). The unique behavioral features of vampire bats, including allogrooming (Wilkinson 1986) and food sharing through regurgitation of blood meals (Wilkinson 1984), could also lead to opportunities for intraspecific virus transmission.

Vampire bats have been well-studied as a key wildlife reservoir in Latin America for rabies virus (family *Rhabdoviridae*) due to the high burden of disease for agriculture and public health (Schneider *et al.* 2009; Streicker *et al.* 2012b; Condori-Condori *et al.* 2013; de Thoisy *et al.* 2016; Benavides *et al.* 2016; Streicker *et al.* 2016). However, vampire bats are also an interesting species in which to examine viral diversity in general due to their unique behavioral traits. In addition to *Rhabdoviridae*, vampire bats are known to harbor other viral families, including *Paramyxoviridae* (Drexler *et al.* 2012a), *Coronaviridae* (Brandão *et al.* 2008), *Polyomaviridae* (Fagrouch *et al.* 2012), *Adenoviridae* (Lima *et al.* 2013; Wray *et al.* 2016), *Retroviridae* (Escalera-Zamudio *et al.* 2015), and *Herpesviridae* (Wray *et al.* 2016; Escalera-Zamudio *et al.* 2016; Salmier *et al.* 2017). Two recent viral metagenomic studies of vampire bats detected many of these families in addition to other vertebrate-infecting taxa (Salmier *et al.* 2017; Escalera-Zamudio *et al.* 2017). However,

these studies were primarily descriptive, examining relatively few samples across small geographic scales and employing sample collection and processing methods which have the potential to bias the viral communities detected.

In addition to vertebrate-infecting viral taxa, metagenomic studies in bats have simultaneously detected bacteriophages (Donaldson *et al.* 2010; Ge *et al.* 2012) and insect, plant and fungal viruses likely acquired from diet or the environment (Ge *et al.* 2012; Dacheux *et al.* 2014; Salmier *et al.* 2017). Considering viruses that are not infecting the bats themselves as part of the viral community could provide a signature of host diet or environment. Previous metagenomic studies in bats have discovered novel viruses in sample types including feces, saliva, urine and tissue (e.g. Donaldson *et al.* 2010; Baker *et al.* 2013; Dacheux *et al.* 2014; Salmier *et al.* 2017), and the body habitat in which a novel virus is detected can also yield clues to its tissue tropism and transmission route (Young & Olival 2016). Bacterial communities have been found to vary between body habitats in bats (Dietrich *et al.* 2017) as have viral communities in humans (Wylie *et al.* 2014; Hannigan *et al.* 2015). However, there has not yet been an explicit test of the compartmentalization of bat viral communities between different body habitats.

The aim of this chapter was to build upon previous viral metagenomic studies in vampire bats by expanding the geographic scale of sampling and employing minimally biased lab methods to thoroughly characterize viral communities. I generated viral metagenomic data from saliva and fecal samples from vampire bats captured across the country-wide scale of Peru. Specifically, I aimed to address (1) what are the natural host groups of viral taxa detected in vampire bats, (2) are there differences in viral richness (alpha diversity) and community composition (beta diversity) between body habitats in vampire bats and (3) are novel vampire bat viruses most closely related to other bat-infecting viral taxa? The results provide a detailed description of viral diversity in an ecologically important bat host.

3.3 Materials and methods

3.3.1 Dataset descriptions

Metagenomic sequence datasets were generated for 62 pools, of which 16 comprised pooled samples from multiple bat colonies (multi-colony pools) and 46 comprised pooled samples from one bat colony (single colony pools). The multi-colony dataset, described in Chapter 2, consisted of ten individual samples pooled across 1-2 colonies within the same

locality (Table 2.5; Figure 2.1; Table A1). Duplicated samples from Table 2.5 were consolidated to include only those that were processed according to the final laboratory protocol (Appendix B1) and are therefore comparable to one another (Table 3.1).

Table 3 1 Multi colony pools sequenced to characterize viral communities in vampire bats. Pools were created by combining nucleic acid from 10 individual swabs from the same sample type and the same site or locality. Pools were the same as in Table 2.5 but are consolidated to include only samples processed according to the final protocol.

Pool ID†	Sample Type	Raw Reads	Viral Reads	Colony 1‡	Colony 2‡
AAC_H_F	Feces	12,166,001	10,870	AYA7	AYA14
AAC_H_SV	Saliva	9,507,979	431	AYA7	AYA14
AAC_L_F	Feces	12,000,988	2,417	API1	AYA11
AAC_L_SV	Saliva	15,121,355	609	API1	AYA11
AMA_L_F	Feces	17,760,709	28,344	AMA2	AMA6
AMA_L_SV	Saliva	9,363,273	305	AMA2	AMA4
CAJ_L_F	Feces	15,843,806	5,945	CAJ4	-
CAJ_L_SV	Saliva	8,685,456	600	CAJ4	-
CAJ_H_F	Feces	8,661,617	8,085	CAJ1	CAJ2
CAJ_H_SV	Saliva	11,830,542	534	CAJ1	CAJ2
HUA_H_F	Feces	10,814,816	11,285	HUA1	HUA2
HUA_H_SV	Saliva	8,931,393	517	HUA1	HUA2
LMA_L_F	Feces	17,365,381	8,206	LMA5	LMA6
LMA_L_SV	Saliva	15,953,442	483	LMA5	LMA6
LR_L_F	Feces	13,843,629	4,544	LR1	LR2
LR_L_SV	Saliva	9,023,821	478	LR1	LR2

†All Pool IDs reflect the locality (AAC, Ayacucho- Apurímac -Cusco; AMA, Amazonas; CAJ, Cajamarca; HUA, Huánuco; LMA, Lima; LR, Loreto), sample type (F, feces; SV, saliva), and elevation (H, high; L, low) to differentiate localities with multiple pools.
‡Colony codes correspond to department within Peru. Locations and pool midpoints are shown in Figure 2.1.

Single colony viral community data were also generated by pooling individual fecal and saliva samples from 24 colonies, resulting in a total of 48 pools (Table 3.2; Table A1). Up to 10 individual swabs from a colony were pooled; different sample type pools within a colony often contained some or all of the same individuals and can therefore be considered as representing a colony-level viral community but not always the exact same individuals. Fecal and saliva swab samples were collected in the field (Section 2.3.2), individually extracted, pooled, and library prepared (Appendix B1). Forty-eight libraries were combined in equimolar ratios and sequenced in one High Output run (v2 300 cycles) on an Illumina NextSeq500 at the MRC-University of Glasgow CVR.

**Table 3 2 Single colony pools sequenced to characterize viral communities in vampire bats.
Pools were created by combining nucleic acid from up to 10 individual swab samples from
the same sample type and the same colony.**

Pool ID†	Sample Type	Colony‡	Raw Reads	Viral Reads	Proportion Viral
AMA7_F	Feces	AMA7	13,458,777	54	0.000004
AMA7_SV	Saliva	AMA7	10,908,218	160	0.000015
AMA2_F	Feces	AMA2	10,867,145	16,002	0.001473
AMA2_SV	Saliva	AMA2	8,583,589	360	0.000042
API1_F	Feces	API1	11,050,071	1,286	0.000116
API1_SV	Saliva	API1	11,605,602	210	0.000018
API17_F	Feces	API17	9,323,302	8,084	0.000867
API17_SV	Saliva	API17	9,312,286	128	0.000014
API140_F	Feces	API140	8,455,451	17,522	0.002072
API140_SV	Saliva	API140	11,407,649	722	0.000063
API141_F	Feces	API141	13,278,728	155,696	0.011725
API141_SV	Saliva	API141	10,955,775	558	0.000051
AYA1_F	Feces	AYA1	7,098,210	30,434	0.004288
AYA1_SV	Saliva	AYA1	9,922,871	94	0.000009
AYA7_F	Feces	AYA7	11,345,890	13,006	0.001146
AYA7_SV	Saliva	AYA7	10,363,555	490	0.000047
AYA11_F	Feces	AYA11	8,590,173	6,680	0.000778
AYA11_SV	Saliva	AYA11	11,207,240	324	0.000029
AYA12_F	Feces	AYA12	13,458,223	51,540	0.003830
AYA12_SV	Saliva	AYA12	11,608,698	110	0.000009
AYA14_F	Feces	AYA14	9,592,865	2,024	0.000211
AYA14_SV	Saliva	AYA14	11,075,284	1,178	0.000106
AYA15_F	Feces	AYA15	7,492,336	35,392	0.004724
AYA15_SV	Saliva	AYA15	9,448,671	254	0.000027
CAJ1_F	Feces	CAJ1	8,187,011	1,712	0.000209
CAJ1_SV	Saliva	CAJ1	9,047,393	262	0.000029
CAJ2_F	Feces	CAJ2	8,829,558	9,534	0.001080
CAJ2_SV	Saliva	CAJ2	14,671,986	1,122	0.000076
CAJ4_F	Feces	CAJ4	8,532,268	3,608	0.000423
CAJ4_SV	Saliva	CAJ4	9,468,189	316	0.000033
CUS8_F	Feces	CUS8	14,834,175	19,192	0.001294
CUS8_SV	Saliva	CUS8	13,942,320	990	0.000071
HUA1_F	Feces	HUA1	9,362,178	30,676	0.003277
HUA1_SV	Saliva	HUA1	17,852,828	606	0.000034
HUA2_F	Feces	HUA2	13,876,749	6,988	0.000504
HUA2_SV	Saliva	HUA2	12,764,201	416	0.000033
HUA3_F	Feces	HUA3	3,534,944	19,188	0.005428
HUA3_SV	Saliva	HUA3	13,357,838	210	0.000016
HUA4_F	Feces	HUA4	16,396,134	17,064	0.001041
HUA4_SV	Saliva	HUA4	12,705,637	448	0.000035
LMA5_F	Feces	LMA5	8,103,620	6,390	0.000789
LMA5_SV	Saliva	LMA5	11,437,853	234	0.000020
LMA6_F	Feces	LMA6	7,712,739	392	0.000051
LMA6_SV	Saliva	LMA6	12,372,323	168	0.000014
LR2_F	Feces	LR2	6,963,077	2,614	0.000375

LR2_SV	Saliva	LR2	12,988,290	210	0.000016
LR3_F	Feces	LR3	6,115,567	1,164	0.000190
LR3_SV	Saliva	LR3	12,976,159	96	0.000007

†Pool IDs reflect the colony and sample type (F, feces; SV, saliva)
‡Colony names reflect the department (AMA, Amazonas; API, Apurímac; AYA, Ayacucho; CAJ, Cajamarca; CUS, Cusco; HUA, Huánuco; LMA, Lima; LR, Loreto). Locations are shown in Figure 4.1.

3.3.2 Bioinformatic analyses

Sequence data were processed through the bioinformatic pipeline (Appendix B2). Description of overall diversity and sample type comparison were performed on the multi-colony read-level dataset. Phylogenetic analyses were performed on contigs from the multi-colony and single-colony datasets, which were further processed following assembly by taxonomic classification using Diamond (Buchfink *et al.* 2014) and KronaTools (Ondov *et al.* 2011) and a summary for each contig including length, coverage, BLAST e-value, and taxonomic assignment was generated using a custom script (R. Orton).

Sequencer-related carryover can be a problem in metagenomic studies, so it was important to ensure that results did not represent contamination with any other viruses sequenced at the CVR. The standard operating procedures for cleaning the sequencing machines reduce contamination to 1 read every 100,000 (0.001%; A. da Silva Filipe). For both read and contig datasets, assigned sequences were examined for, but did not contain, viruses sequenced on prior runs (specifically African swine fever virus, Equine influenza virus, and Parainfluenza virus).

Additionally, when viral contigs were assigned to genera known to have been sequenced previously, the nucleotide identity of these contigs was examined to confirm that they were not the result of contamination. Contigs matching to *Cyprinid herpesvirus*, *Ranavirus*, and *Cytomegalovirus* were checked by nucleotide blast; following this step none were deemed suspect contaminants due to poor match at the nucleotide level to the previously sequenced viral taxa (data not shown). Individual reads were not checked, but all read-based analyses were repeated with four different data subsets (all viruses, vertebrate-infecting viruses, taxa with >1 reads and taxa with >10 reads) to ensure that low-level contamination would not bias results. Although contamination can occur during extraction or library preparation, metagenomic samples were processed through a strict CVR laboratory pipeline in which samples were extracted in a room for only non-propagated clinical samples and libraries were prepared in a room for only non-amplified material. Based on these lines of evidence,

the effect of carryover contamination due to prior sequencing of other viruses was considered to be minimal.

3.3.3 Overall diversity

The overall presence and natural host groups of viral taxa at the read level was examined in 16 multi-colony pools. Many of the viral taxa detected infect vertebrates, but others could represent viruses associated with commensal bacteria or transient viruses acquired through diet or the environment. Therefore, the relative proportion of viral taxa infecting different natural host groups was examined using information from databases on which class of host a viral taxon typically infects (Hulo *et al.* 2011; Adams *et al.* 2017). Host groups included vertebrates, invertebrates, bacteria, plants, fungi, other (viral taxa infecting protozoans, amoebas, or algae), or any combinations of these groups (signifying either broad host range or low specificity). Multi- and single-colony contig datasets were combined to identify viral families that were frequently detected as large contigs (>1000bp). Comparisons of reads assigned to different host classes were performed using R version 3.4.2 (R Core Team 2017).

3.3.4 Sample type comparison

Differences in viral reads, richness (alpha diversity) and community composition (beta diversity) were compared between feces and saliva. Analyses were repeated at the family and genus level using four different data subsets (all viruses, vertebrate-infecting viruses, taxa with >1 reads and taxa with >10 reads) to ensure that rare or environmental viral reads were not biasing results. To examine whether sample types differed based on viral read number and distribution, the R package pheatmap (Kolde 2015) was used to construct a heatmap in which samples were hierarchically clustered using Ward's method (Ward 1963; Murtagh & Legendre 2014).

Viral richness at the family and genus level (alpha diversity) for each sample type was calculated using the package vegan (Oksanen *et al.* 2017). Differences in viral community composition between sample types (beta diversity) were visually assessed using a principal coordinate analysis (PCoA), in which Jaccard distance matrices were calculated from presence-absence data using vegan, and ordinations were then performed on the distance matrix using the package ape (Paradis *et al.* 2004).

A GLM-based approach for multivariate data in the mvabund package (Wang *et al.* 2012) was used to statistically test for differences in viral community composition between sample types. This method has greater power than the distance-based analyses often used to assess differences in community composition, and accounts for the positive mean-variance relationship which is common in such datasets (Warton *et al.* 2011; Wang *et al.* 2012). The function *manyglm* was used to test for differences in viral community composition (presence/absence of viral taxa) between sample types, while controlling for the potential effects of number of raw reads and sequencing run using a separate logistic regression for each taxon (generalized linear model with binomial error). Function *anova.manyglm* was then used to test for multivariate significance using the log-likelihood ratio test statistic and PIT-trap resampling (Warton *et al.* 2017) with 999 iterations. To identify viral taxa that significantly differed between sample types, univariate p-values were calculated using resampling with 999 iterations, and with adjustment for multiple comparisons using a step-down resampling method.

3.3.5 Phylogenetic analysis of select viral families

Viral sequences that assembled into large contigs (including potential full genomes), sequences from families that were of interest as potential zoonoses, and families containing sequences with widespread geographic distribution were selected for phylogenetic analysis. Nucleotide and protein blast analyses were performed against the Genbank nt and nr databases. Phylogenetic analyses varied depending on the nucleotide or amino acid sequence used to define relationships within each viral family. However, all analyses aligned newly generated sequences with previously published sequences using MAFFT (Katoh *et al.* 2002) within Geneious v.7.1.7 (Kearse *et al.* 2012). The best substitution model was then selected using jModelTest (Darriba *et al.* 2012) for nucleotide alignments and ProtTest3 (Darriba *et al.* 2011) for amino acid alignments. Phylogenetic analyses were then performed using maximum likelihood inference in RAxML (Stamatakis 2014) with 100 bootstraps using the rapid bootstrapping algorithm (Stamatakis *et al.* 2008). Trees were all visualized with a midpoint root.

RAxML can only accommodate GTR substitution models for nucleotide analyses, so these were performed using the GTR model including, if indicated by jModelTest, gamma distributed rate variation and a proportion of invariant sites in the model. When contigs matching a particular virus were detected in both multi-colony and single-colony datasets with individuals in common, potentially representing an infection in one or several of these

individuals, phylogenetic analyses were only performed on one dataset which was selected based on contig length, depth of coverage, or detection in a greater number of pools.

3.4 Results and Discussion

3.4.1 Overall diversity: read level

A total of 220 viral genera (65 families), including 81 genera (24 families) known to infect vertebrates, were detected in the read-level data across both multi-colony and single-colony pools. When reads from all multi-colony pools were combined, there were clear differences in host groups between sample types (Figure 3.1; Table C1, C2, C3).
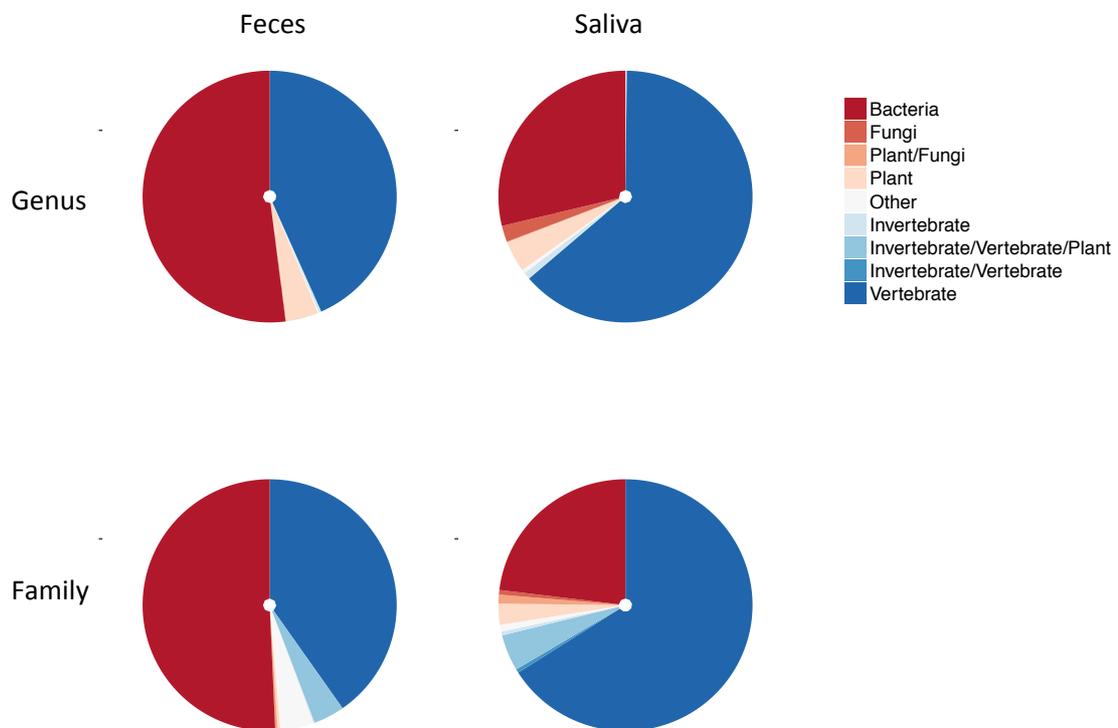
Figure 3 1 Natural host groups of viral reads in vampire bat feces and saliva combined across multi-colony pools. Natural hosts are the host class usually infected by a family or genus of virus, which were assigned based on Hulo *et al.* (2011) and Adams *et al.* (2017). Reads are combined from all multi-colony pools for each sample type (feces and saliva) and at different taxonomic levels (viral genus and family).

Fecal samples contained more viral reads than saliva samples (Table 3.1; Table 3.2), mainly due to sequences from bacteriophages, with almost half of all viral reads in fecal samples originating from viruses that infect bacteria (Figure 3.1). Saliva samples contained fewer reads but those detected were primarily from vertebrate-infecting viruses. In both sample types, host group patterns were similar at the family and genus level, but with higher specificity of host assignment at the viral genus level (i.e. fewer reads assigned to

families infecting both vertebrates/invertebrates or plants/fungi, but instead to a genus infecting one or the other).

When host group was examined in each multi-colony pool separately patterns remained similar; there were more bacteriophages in fecal samples, more vertebrate-infecting viruses in saliva samples, and higher specificity of assignments at the genus level (Figure 3.2).
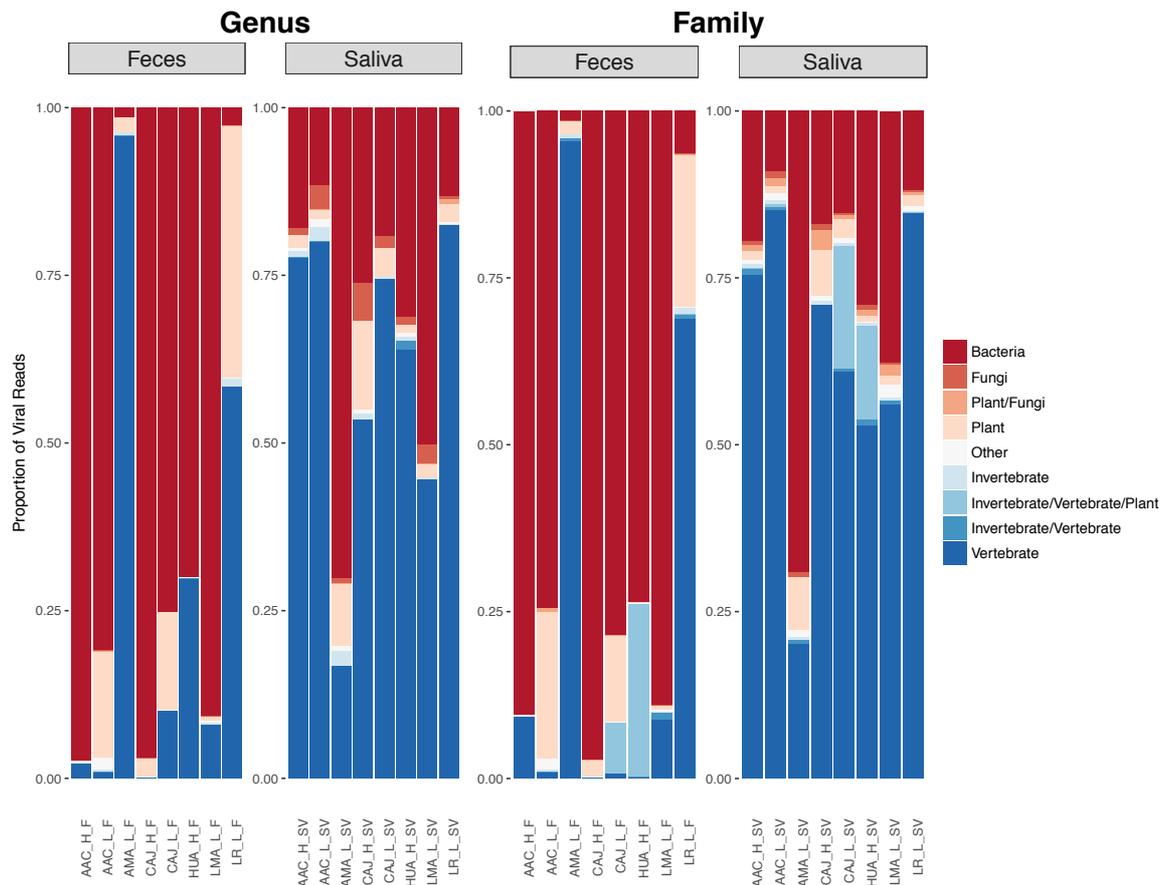
Figure 3 2 Natural host groups of viral reads in vampire bat feces and saliva split by locality. Natural hosts are the host class usually infected by a family or genus of virus, which were assigned based on Hulo *et al.* (2011) and Adams *et al.* (2017). Reads are separated by different sample types (feces and saliva) and at different taxonomic levels (genus and family). Each bar represents the total viral read assignments of one multi-colony pool.

However, some differences in host groups became apparent between pools. In fecal pools from two localities (AMA_L_F and LR_L_F), there was a higher proportion of reads assigned to vertebrate-infecting viruses than in other localities. There were also more bacteriophage-assigned reads in one saliva sample (AMA_L_SV). Samples with unusual host groupings (i.e. fecal samples with higher proportions of vertebrate-infecting viruses and the saliva sample with a higher proportion of bacteriophage-assigned reads) had a wide range in number of total viral reads (305 - 28,344) such that these observations are unlikely

to be an artefact of viral read number (i.e. the same number of bacteriophage reads across samples but fewer reads from other virus types).

The abundance of bacteriophage reads in fecal samples is likely due to the presence of more bacteria in the gut compared to saliva (Sender *et al.* 2016). The proportion of vertebrate-infecting viruses was higher in saliva samples, despite those samples containing fewer viral reads, which is interesting given that the majority of metagenomic viral discovery efforts in bats have focused on fecal samples (e.g. Li *et al.* 2010a; Ge *et al.* 2012; Yinda *et al.* 2017; Zheng *et al.* 2017), although some studies have examined both feces and saliva (e.g. Donaldson *et al.* 2010; Wu *et al.* 2012; 2016). Patterns of host grouping also varied between localities, suggesting that other location-specific factors influence viral communities in addition to sample type.

3.4.2 Overall diversity: contig level

Viral reads were assembled to form contigs up to 29,140bp (Table C4). Based on the contig dataset, there were a total of 139 viral genera (59 families) detected, of which 42 were vertebrate-infecting genera (23 families).

Retroviridae, *Herpesviridae*, and *Papillomaviridae* were the most common vertebrate-infecting viral families detected in saliva, while fecal samples more often contained families such as *Coronaviridae*, *Adenoviridae*, and *Hepeviridae* (Table C4). In fecal samples, bacteriophage families (*Podoviridae, Myoviridae, Siphoviridae*) were most widespread (i.e. contigs matching to *Podoviridae* were detected in all 8 multi-colony localities and all 24 single-colony sites). Bacteriophages are a unique sub-community of viruses in that they are often beneficial to the host, as they can be involved in regulating the gut microbiome (Manrique *et al.* 2016). Due to their unique blood feeding behavior, vampire bats could be an interesting system in which to study bacteriophages, as host diet influences the phage community within the gut (Minot *et al.* 2011). Phages can influence the immune system of hosts in return (Duerkop & Hooper 2013; Virgin 2014), and there is evidence that healthy humans share a core set of gut phages (Manrique *et al.* 2016). The widespread presence of the same bacteriophage families suggests that bats might also have a core "phageome" that contributes to their health, although this observation requires further investigation.

In addition to vertebrate-infecting viruses potentially associated with vampire bats, and bacteria-infecting viruses potentially associated with mutualistic gut bacteria, there were also large contigs from families that infect plants (*Tombusviridae* and *Tymoviridae*). Detecting large contigs from plant-infecting viral families was unexpected given that vampire bats are not thought to consume anything aside from blood (Voigt & Kelm 2006). Plant viruses could have been acquired through environmental interactions, such as from the skin of prey animals or conspecific grooming. Alternatively, herbivorous vampire bat prey (e.g. livestock) might consume plant viruses in the course of their own feeding, which could then be ingested by vampire bats through blood meals.

There were also detections of large contigs from insect-infecting viral families such as *Dicistroviridae*, which are commonly found in insectivorous bat families that acquire them through diet (Li *et al.* 2010a; Ge *et al.* 2012). Vampire bats are not believed to consume insects, although arthropods have been found in stomach contents (Aguirre *et al.* 2003) and a potentially insect-borne virus was detected in vampire bat fecal samples (Wray *et al.* 2016). Bats could accidentally ingest insects while flying or consume ectoparasites from the skin of prey while feeding. Vampire bats also groom one another socially (Wilkinson 1986; Carter & Leffer 2015) and are known to carry a variety of ectoparasites (Patterson *et al.* 2008; de Souza Aguiar & Antonini 2011), so a bat could consume an ectoparasite in the course of grooming. A recent metabarcoding diet study of vampire bats confirmed the presence of arthropod DNA in vampire bat stomach contents (Bohmann *et al.* 2018), so it is possible that insect-infecting viruses could be accidentally consumed by vampire bats along with arthropod hosts.

Some of the viral families detected have been found previously in vampire bats, while others which are novel for the species. For example, *Hepatitis deltavirus* is a viral taxa that has never been detected in bats, and indeed is only known previously from humans. In contrast, some detected viral families were well known from Neotropical bats such as *Coronaviridae* (Anthony *et al.* 2013; Corman *et al.* 2013) and some were known from vampire bats, such as *Adenoviridae* (Wray *et al.* 2016; Salmier *et al.* 2017) and *Herpesviridae* (Wray *et al.* 2016; Escalera-Zamudio *et al.* 2016; Salmier *et al.* 2017). Some of the viral families most commonly detected across different pools, including *Retroviridae* and *Papillomaviridae*, were found in vampire bats from French Guiana and Mexico (Salmier *et al.* 2017; Escalera-Zamudio *et al.* 2017).

However, some viral families were not detected that have been found in bats both globally and in the Neotropics, such as *Orthomyxoviridae* (Tong *et al.* 2012; 2013) and *Hepadnaviridae* (Drexler *et al.* 2013). There were also not any reads or contigs from several viral families that have previously been described in vampire bats, such as *Polyomaviridae* (Fagrouch *et al.* 2012; Salmier *et al.* 2017) and *Nairoviridae* (Salmier *et al.* 2017). It is possible that these families are present but were not detected using the metagenomic sequencing approach. However, prevalence rates of some viral families were low according to previous PCR-based and serological studies; for example, Hepadnaviruses were only detected at a rate of 0.3% by serology and a range of 6.3-9.3% by PCR (Drexler *et al.* 2013). It is possible that sampling more individuals would reveal these or other undetected viral families in vampire bats. Some zoonotic viral families known from bats were absent from samples, such as *Filoviridae*, which has only been previously described in Old World bats (Hayman 2016). However, trait-based modeling has shown that New World bat species possess similar traits to known *Filoviridae* bat hosts, suggesting that biogeographical processes could be responsible for their absence rather than host unsuitability (Han *et al.* 2016). Particularly at the read level where false positives and negatives are an issue, metagenomics is a method better suited to describing the presence of a virus than proving its absence, but can highlight families of interest (either present or absent) for more intensive future surveillance.

It is important to note that even for vertebrate-infecting viral families, a metagenomic study cannot establish that viruses are infecting or replicating in vampire bats; viral detections could equally represent transient viruses acquired from the vertebrates upon which vampire bats prey. It is also relevant to discuss whether viral families acquired from the environment, such as plant or insect families, should be considered members of the vampire bat viral community. The answer may depend on the goal of the analysis; for questions related to disease emergence and host/virus population structure, vertebrate-infecting viruses are likely to provide a more useful signal of host and virus movement, while other environmental viruses may only add noise. However, for spatial analyses in which the viral community is considered more of an environmental fingerprint, non bat-infecting viruses may be worth including.

3.4.3 Sample type comparison

A heatmap of viral reads at the genus level revealed that saliva and fecal samples clustered separately from one another (Figure 3.3A), with each sample type exhibiting different

genera with high read counts common across many localities (Figure 3.3B). Viral richness was higher on average in feces compared to saliva but this difference was non-significant (1.34 more viral genera; p=0.64), while saliva and fecal communities clearly separated along the first axis of a PCoA (Figure 3.3C,D). These results were largely consistent for different data subsets and when repeated at the family level (Figure C1 – Figure C5), with the exception of genus-level vertebrate-infecting viral richness (5.25 more viral genera in saliva samples; p=0.02).
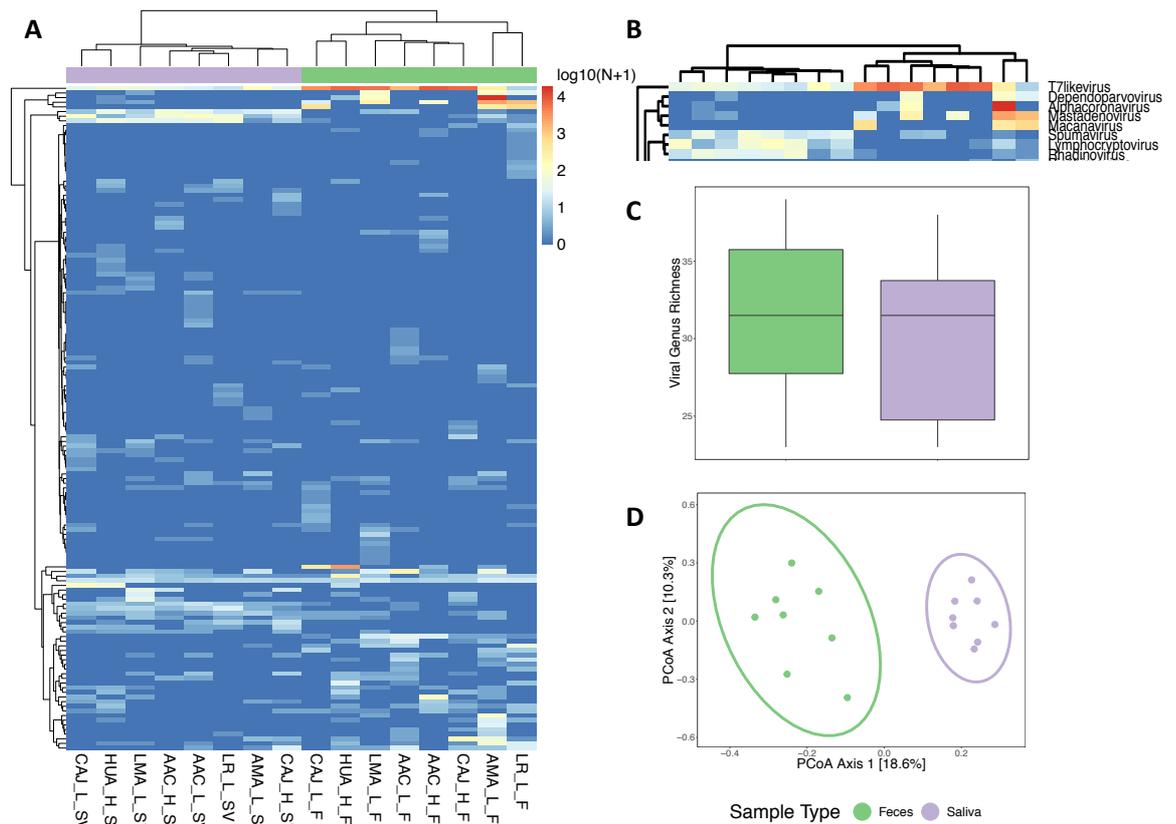
Figure 3 3 Comparison of genus-level fecal and saliva viral communities.
(A) Heatmap of read number in multi-colony pools where similar rows (viral genera detected together) and columns (pools containing viral genera in common) are clustered according to Ward's method (B) Inset of heatmap showing viral genera with high read abundance for fecal and saliva samples (C) Comparison of viral richness between sample types. Bold line shows the median, and upper and lower hinges show the first and third quartiles. Whiskers extend from the hinge to 1.5 * the inter-quartile range, (D) Comparison of viral community composition using principal coordinate analysis (PCoA). The first two axes are plotted, and axis labels show the percent of variation explained by each, with 95% confidence ellipses plotted for each group.

Sample type also had a significant multivariate effect on viral community composition, while sequencing run and number of raw reads did not (Table 3.3). Individual viral genera that differed significantly between sample types were *Percavirus* (LRT=15.902; p=0.016), *Gammapapillomavirus* (LRT=12.173; p=0.043), and *Coccolithovirus* (LRT=15.902; p=0.015). These results were consistent at the family level (Table C5).

Table 3 3 Genus-level viral community composition differs between sample types.
A GLM-based approach was used to test for a multivariate effect of sample type on viral community composition while controlling for the effects of sequencing run and number of raw reads.

Variable	Residual d.f.	d.f.	Dev†	*P*-value
Sample Type	**14**	**1**	**364.7**	**0.001**
Run	13	1	206.1	0.096
Raw Reads	12	1	293.1	0.062

†Deviance test statistic calculated using a log likelihood ratio test

Although fecal and saliva samples generally did not differ in viral richness, the exception was a significantly greater number of vertebrate-infecting viral genera in saliva, potentially due to the diversity of genera within the families *Papillomaviridae* and *Herpesviridae* (data not shown). These families were predominantly detected in saliva samples and have been found to exhibit high diversity at the genus level in human viral communities (Wylie *et al.* 2014). A previous study found that bacterial species richness in bats also did not differ between feces and saliva, although richness was significantly higher in urine (Dietrich *et al.* 2017) which was not examined in this study.

In contrast to richness, viral community composition clearly differed between sample types, as shown visually by ordination and in the multivariate GLM analysis (Figure 3.3D; Table 3.3). Differences in community composition have also been found between body sites in bacterial communities of bats (Dietrich *et al.* 2017), as well as for viral communities in humans (Wylie *et al.* 2014; Hannigan *et al.* 2015), and specifically for saliva and fecal viral communities in humans (Paez-Espino *et al.* 2016). These results emphasize the need to compare samples from the same body habitat when examining viral communities. In the case of novel viruses, the body habitat in which a virus is detected can yield clues about its biology, such as tissue tropism or mode of shedding (Young & Olival 2016).

3.4.4 A novel *Hepatitis deltavirus* in vampire bat saliva

Hepatitis deltavirus (HDV) is a subviral pathogen, which are parasitic viruses that depend on co-infection with a helper virus for replication and transmission. HDV causes severe viral hepatitis in humans, and an estimated 15-20 million people worldwide are chronically infected with the virus (Lempp & Urban 2017). HDV is found in combination with Hepatitis B virus (HBV) as it is dependent on three envelope proteins from HBV to form viral particles. Humans are the only known natural host for HDV, with eight human clades characterized (Le Gal *et al.* 2006). Experimental studies have shown that woodchuck cells

can be infected with the virus when a helper HBV-like virus is present (Ponzetto *et al.* 1984). An evolutionary relationship has been proposed between HDV and viroids, which are ssRNA satellites of plants (Elena *et al.* 1991), although competing hypotheses suggest that HDV arose from host RNA (Huang & Lo 2010) or from the RNA of a co-infecting virus such as HBV (Taylor 2014).

Saliva samples from three localities across Peru yielded apparent full genome contigs (1770 bp) matching most closely to HDV according to Diamond protein blast, which is hereafter referred to as *Desmodus rotundus Hepatitis deltavirus* (DrHDV). The amino acid sequence of the single coding region of HDV, which encodes the delta antigen protein, was extracted from sequences from three single-colony pools (LMA6_SV, CAJ1_SV, AYA14_SV) using getorf (Rice *et al.* 2000). The resulting phylogeny indicated that DrHDV sequences formed a unique, well-supported clade distinct from any characterized human virus (Figure 3.4).
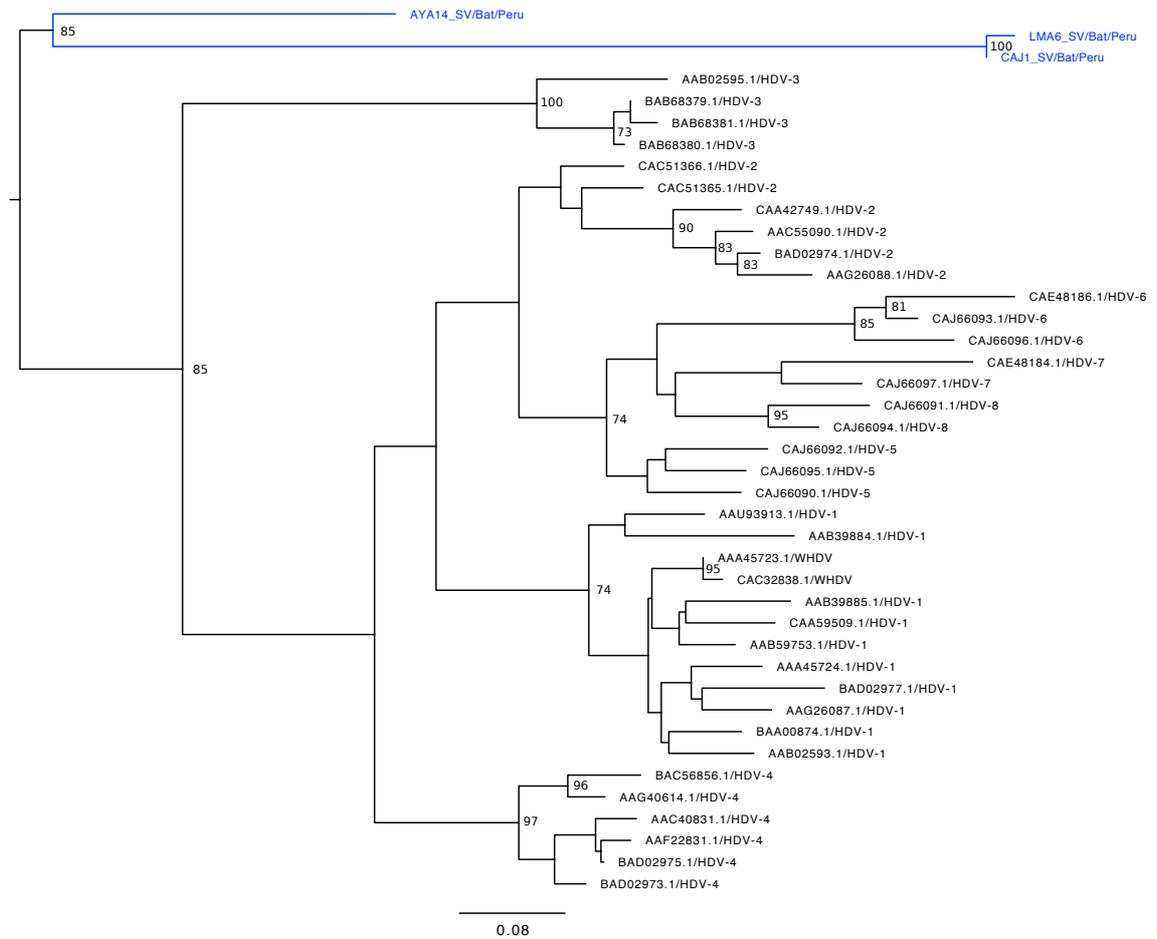
Figure 3 4 Hepatitis deltavirus phylogeny.
Maximum likelihood tree based on a 218 amino acid alignment of 41 sequences of the delta antigen protein from three Peru Hepatitis deltavirus (HDV) sequences (blue) and representative sequences from each of the characterized human HDV clades (black). Phylogenetic analysis was conducted using the JTT+G substitution model. Human HDV sequences include Genbank accession number and clade number in the tip labels. Peru sequence labels include pool, host species, and location of origin. Bootstrap values of >70 are displayed. Published viruses are detailed in Appendix C, Table C6. The scale bar represents the mean expected rate of substitutions per site.

DrHDV was distantly related to HDV at the nucleotide level, matching only over a small portion of the genome, and was also divergent at the protein level, although blast results matched over the entire protein sequence (Table 3.4). For known strains of human HDV, percent similarities at the nucleotide level within a clade are typically >80% (Le Gal *et al.* 2006), so the bat-derived sequences appear to represent a novel clade.

Table 3 4 Nucleotide and protein blast results for DrHDV sequences.
DrHDV genomes were detected in saliva of single and multi-colony pools from three localities. Full genomes were analyzed by nucleotide blast and open reading frames coding for the delta antigen were analyzed by protein blast to determine differences from previously characterized viruses from humans.

Sample ID	Length	Nucleotide†			Protein‡		
		Top hit	Query cover	Percent Identity	Top hit	Query cover	Percent Identity
LMA6_SV	1771	Hepatitis delta virus isolate C15	30	67	HDAg-large	98	55
CAJ1_SV	1770	Hepatitis delta virus, strain dFr1650	25	69	small delta antigen	100	54
AYA14_SV	1771	Hepatitis delta virus, strain dFr3006	52	71	delta antigen	100	64
CAJ_H_SV	1711	Hepatitis delta virus, strain dFr4824	13	74	small delta antigen	100	54
AAC_H_SV	1771	Hepatitis delta virus, strain dFr3006	52	71	delta antigen	100	64
LMA_L_SV	1771	Hepatitis delta virus from Somalia genotype IC	50	66	small delta antigen	100	54

†Nucleotide results show the top hit, query cover, and percent identity from nucleotide blast
‡Protein results show top hit, query cover, and percent identity from protein blast

This is the first detection of an HDV-like virus outside of humans, raising many questions about the evolutionary origins of the virus and whether its mechanisms of replication and transmission differ between bats and humans. There were three different DrHDV sequences in the three pools; the sequences from Lima and Cajamarca were more similar than the one from Ayacucho, which mirrors the pattern known to exist for rabies virus (Streicker *et al.* 2016). DrHDV was detected in saliva at all three sites, and while the virus typically exhibits liver tropism in humans (Lempp & Urban 2017), the presence of HDV RNA and antigen has recently been reported in human salivary glands without an accompanying HBV infection (Weller *et al.* 2016). There were no detections of HBV or any other Hepadnaviruses in the metagenomic results, although this viral family is well known in bats (Drexler *et al.* 2013; Rasche *et al.* 2016). HDV is able to replicate its genome and persist for at least a year in humans without the presence of HBV as a helper virus (Giersch *et al.* 2014), but assembly of infectious particles is only possible in the presence of a helper virus (Lempp & Urban 2017). Several possible explanations for the detection of HDV in vampire bat saliva without HBV detection include a) there is something particular about the salivary glands in which HDV can persist without the presence of a helper virus b) HBV is present but undetectable, either due to absence in

saliva/feces or the limitations of metagenomics as a method for detecting this virus c) HBV is not present and another virus is acting as a helper virus to allow HDV persistence or d) the HDV detected in vampire bat saliva represents a cross-species transmission of the virus from another animal and is not actively infecting the vampire bat. Further laboratory work is required to determine which virus, if any, is acting as a helper to the DrHDV virus detected in saliva, and whether HBV is present in any other tissues such as blood or liver.

3.4.5 Hepeviridae

Hepatitis E virus (HEV; family *Hepeviridae*) is an enterically transmitted pathogen that is one of the most common causes of acute hepatitis in the world (Perez-Gracia *et al.* 2014). HEV variants have been found in geographically widespread human populations, as well as in wildlife species such as pig, wild boar, and deer in which it poses a zoonotic threat to humans (Pavio *et al.* 2010; Smith *et al.* 2014). Bat HEVs appear to have a long-term association with their hosts, as previous studies have found that bat HEVs from different parts of the world (China, Germany, Panama) all group together in the *Orthohepevirus D* species (Drexler *et al.* 2012b; Wang *et al.* 2017). *Orthohepevirus D* has not been found in any non-bat host, suggesting fidelity between these viruses and bats.

HEV-like contigs were detected in fecal samples from six colonies, four of which (API17_F, AYA11_F, AYA14_F, LR3_F) were selected for phylogenetic analysis because they had complete (or nearly complete) genomes. Three of the colonies in which the virus was detected (API17, AYA11, AYA14) were located in close proximity to one another and had high nucleotide similarity at the genome level (89.1 - 94.9% nucleotide identity) while the fourth (LR3) was genetically divergent from the others (65.8 - 68.8% nucleotide identity). HEV is characterized by variable patterns of diversity across the genome (Smith *et al.* 2013) and phylogenetic analysis at the amino acid level using homologous subsections of the genome has been proposed (Smith *et al.* 2014); one of these subsections is the RdRp that has been previously sequenced in bat samples (Drexler *et al.* 2012b). Phylogenetic analysis of sequences from Peru were performed on both the whole genome nucleotide sequence and the RdRp amino acid sequence, in which the aim was to place the Peru vampire bat sequences within the known diversity of bat HEVs, including one other Neotropical bat sequence from Panama (Drexler *et al.* 2012b).

In the full genome nucleotide analysis, the Peru sequences fell within the *Orthohepevirus D* group along with other bat viruses (Figure 3.5). The RdRp amino acid phylogeny also

showed the Peru sequences within a monophyletic clade of bat HEVs that was distinct from all other mammalian HEVs, although placement of the bat clade differed relative to the full genome tree (Figure 3.6).

Figure 3 5 Hepatitis E virus full genome phylogeny.
Maximum likelihood tree based on a 7,735bp nucleotide alignment of 76 complete (or nearly complete) genome sequences including four Peru Hepatitis E virus (HEV) genomes (blue), other bat HEV genomes (green) and representative sequences from other HEV genotypes (black). Non-bat sequences are labeled with Genbank accession number, host species, and country of origin. Previous bat sequences are labeled with Genbank accession number, bat host species abbreviation, and country of origin. Peru sequences are labeled with pool, bat host species abbreviation, and country of origin. Contigs considered full genome sequences ranged from 6,646 - 6,683bp, with the exception of API17_F (4,108bp) which was nonetheless included as it comprised most of a genome, including the RdRp section of ORF1 and all of ORF2. Phylogenetic analyses were performed using the GTR+I+G substitution model. Published viruses are detailed in Appendix C, Table C6. Species within *Hepeviridae* are indicated by letters (A-D) on the right side, and major hosts within each species are shown as silhouettes. Bootstrap values of >70 are displayed. The scale bar represents the mean expected rate of substitutions per site.
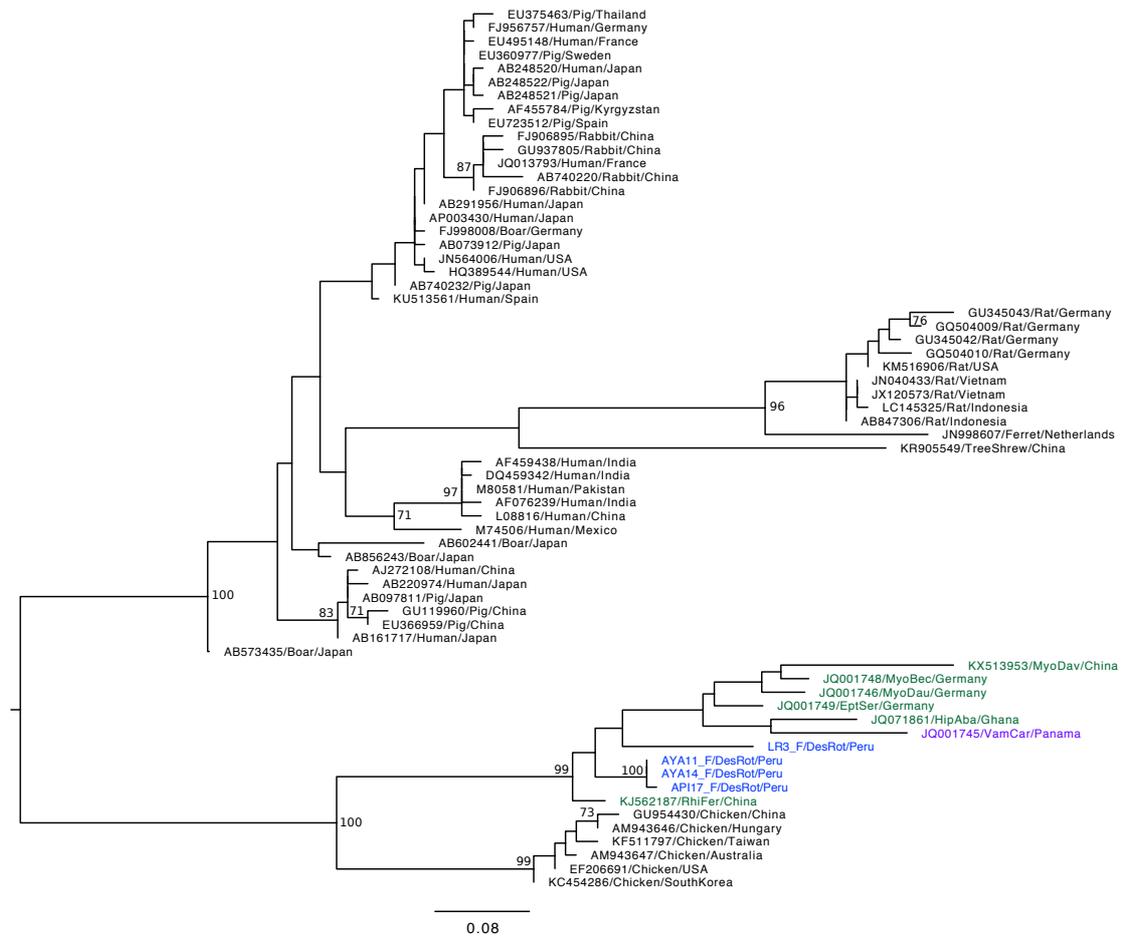
Figure 3 6 Hepatitis E virus RdRp phylogeny.
Maximum likelihood tree based on a 109 amino acid alignment of 65 RdRp sequences (ORF1-1419-ORF1-1527) from four Peru HEV sequences (blue), Neotropical bat sequences (purple), non-Neotropical bat sequences (green) and sequences from other HEV genotypes (black). Non-bat sequences are labeled with Genbank accession number, host species, and country of origin. Previous bat sequences are labeled with Genbank accession number, bat host species abbreviation, and country of origin. Peru sequences are labeled with pool, bat host species abbreviation, and country of origin. Phylogenetic analyses were performed using the LG+G substitution model. Published viruses are detailed in Appendix C, Table C6. Bootstrap values of >70 are displayed. The scale bar represents the mean expected rate of substitutions per site.

The placement of the bat clade in the full genome tree agreed with Drexler *et al.* (2012b), showing that all mammalian HEVs including bat HEVs share a common ancestor. In contrast, the RdRp tree was consistent with later studies (e.g. Smith *et al.* 2014) in which bat HEVs are most closely related to avian HEVs. The conflict between the two phylogenies might be explained by recombination, which has been previously demonstrated in HEV (van Cuyck *et al.* 2005; Wang *et al.* 2010) or by lower resolution in the RdRp tree that is based on a shorter sequence. The other Neotropical bat HEV sequence also fell within the bat clade, but there was no support for a closer relationship between the Panama and Peru sequences relative to bat HEVs from other parts of the world. The clade currently recognized as zoonotic (Species A) is apparently able to cross readily between mammalian hosts (e.g. pigs, boar, humans) while avian and bat HEVs

remain isolated within their host groups despite global distribution. Based on these results, and in agreement with previous studies, including those screening human samples for bat HEV (Drexler *et al.* 2012b) it appears that bat HEVs exhibit a stable association between virus and host.

3.4.6 Coronaviridae

Coronaviruses (CoVs; family *Coronaviridae*) are one of the viral families of highest concern as emerging zoonoses, with high profile species including Severe Acute Respiratory Syndrome (SARS-CoV) and Middle East Respiratory Syndrome (MERS-CoV) (Zhong *et al.* 2003; van Boheemen *et al.* 2012; Zaki *et al.* 2012). Evidence has accumulated for bats as a reservoir host for both SARS-CoV and MERS-CoV (Lau *et al.* 2005; Li *et al.* 2005; Memish *et al.* 2013; Anthony *et al.* 2017a). It is also well-established that CoVs occur frequently in bats and with high levels of genetic diversity, suggesting that bats have a long association with these viruses and act as a reservoir host, with occasional emergence into humans and other species (Poon *et al.* 2005; Woo *et al.* 2006; Tang *et al.* 2006; Drexler *et al.* 2010; Hu *et al.* 2015). The two high profile CoV spillovers both occurred from Old World bats, but there is also evidence for a diverse pool of CoVs existing in New World bats (Dominguez *et al.* 2007) including Neotropical bats (Carrington *et al.* 2008; Anthony *et al.* 2013; Corman *et al.* 2013) and one previous detection in a vampire bat from Brazil (Brandão *et al.* 2008).

A CoV-like contig of full genome length (~29,000bp) was detected in one multi-colony fecal pool (AMA_L_ F) and several contigs adding up to approximately a full genome (6,748; 9,926; 12,414bp) were detected in one single-colony fecal pool (HUA4_F). There were also smaller CoV-like contigs in multi-colony fecal pool LMA_L_F and single-colony fecal pool LMA5_F. The contigs were first examined using nucleotide blast, which indicated they were in the genus *Alphacoronavirus*, but distinct from previously described viruses. The genome organization of our vampire bat CoV was examined in the sample AMA_L_F, which contained the large open reading frames (ORFs) typical of CoVs (Figure 3.7).
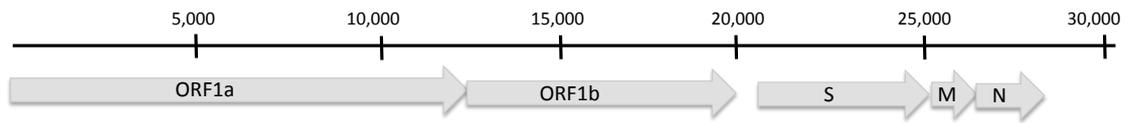
Figure 3 7 Genome organization of the novel *Alphacoronavirus* in vampire bats.
The genome is represented as the black line, with each 5,000bp interval marked. Open
reading frames are represented by grey arrows. The main open reading frames in CoVs
include the replicase genes ORF1a and ORF1b coding for the replicase polyproteins pp1a
and pp1ab, the latter of which is produced by ribosomal frameshifting. Other major open
reading frames depicted include the envelope proteins spike (S) and membrane (M), as well
as the nucleocapsid (N). Adapted from Tao *et al*. (2017).

The demarcation criteria for species within *Coronaviridae* used by the ICTV *Coronaviridae* study group requires a full-length genome, from which pairwise evolutionary distances are calculated from seven conserved domains from seven non-structural proteins encoded by the replicase gene. However, this method is not practical for the large number of samples in ecological studies (and information on the domains used is not publicly available) so an RdRp based grouping method was developed which is congruent with the official method (Drexler *et al.* 2010), and which was subsequently used to characterize novel CoVs in Neotropical bat species (Corman *et al.* 2013). Phylogenetic analysis was performed using this region of the RdRp in order to compare the Peruvian vampire bat sequences to other Neotropical bat sequences.

Getorf was used to extract the polyprotein amino acid sequence (ORF1a) from AMA_L_F and HUA4_F, and protein identity was confirmed by protein blast against the Genbank nr database. LMA pools were excluded from phylogenetic analysis because they did not contain sequence data in the RdRp region. The phylogeny based on the RdRp-based grouping method (Drexler *et al.* 2010; Corman *et al.* 2013) including *Alphacoronavirus* and *Betacoronavirus* sequences indicated that vampire bat sequences were related to other bat CoVs, and specifically fell within a clade of CoVs from other Neotropical bats in the family *Phyllostomidae* (Figure 3.8).
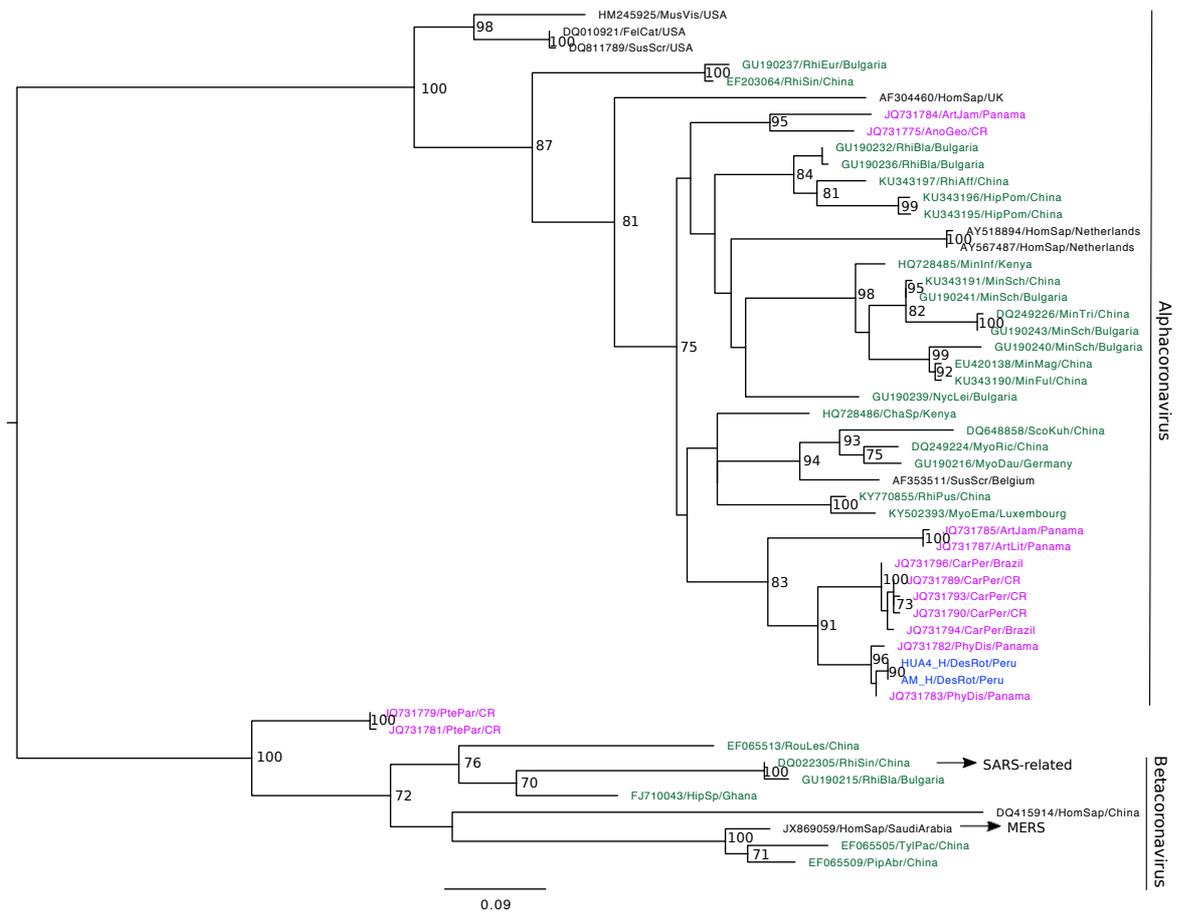
Figure 3 8 Coronavirus RdRp phylogeny.
Maximum likelihood tree based on a 272 amino acid alignment of 50 RdRp sequences including two Peru CoV sequences (blue), Neotropical bat RdRp sequences (pink), non-Neotropical bat RdRp sequences (green) and RdRp sequences from CoVs infecting other species (black). Phylogenetic analysis was conducted using the LG+I+G substitution model. Non-bat sequences are labeled with Genbank accession number, host species, and country of origin. Previous bat sequences are labeled with Genbank accession number, bat host species abbreviation, and country of origin. Peru sequences are labeled with pool, bat host species abbreviation, and country of origin. Published viruses are detailed in Appendix C, Table C6. Arrows indicate SARS-related and MERS sequences, and the genera *Alphacoronavirus* and *Betacoronavirus* are indicated on the right side. Bootstrap values of >70 are displayed. The scale bar represents the mean expected rate of substitutions per site.

In addition to phylogenetically placing the Peru sequences within the context of other Neotropical bat CoVs, the Peru vampire bat sequences were compared with a 52bp sequence generated from a vampire bat in Brazil (Brandão *et al.* 2008). This sequence was non-overlapping with the section of the RdRp used to construct the phylogeny, so a pairwise comparison was performed with the sequences from Peru, yielding 73.1% nucleotide identity. The Peruvian vampire bat CoV appears to differ from previously described sequences, although the other vampire bat sequence is so short that it is difficult to compare and it was not possible to classify sequences according to the official method.

Although CoVs are well known from bats, this is the first full CoV genome characterized in a Neotropical bat, adding to the knowledge of global diversity within the family. The novel vampire bat CoVs fall within a clade of other Neotropical bat CoVs (Figure 3.8), and specifically group with other bats in the family *Phyllostomidae* (*Phyllostomus discolor, Carollia perspicillata, Artibeus jamacensis, Artibeus lituratus*). It has previously been suggested that closely related CoVs are geographically widespread in different host species, and are potentially restricted at the level of host genus (Drexler *et al.* 2010; Corman *et al.* 2013). Given the feeding habits of vampire bats, there is the possibility of cross-species transmission to prey by biting; previous studies have identified CoVs in saliva samples (Anthony *et al.* 2013), although CoVs were only detected in vampire bat fecal samples here. There are not any known human CoV strains closely related to Neotropical bat CoVs, although it has been suggested that rapid deforestation in the region may lead to more opportunities for contact and the potential for CoV transmission from bats to humans in the future (Corman *et al.* 2013).

3.4.7 Rhabdoviridae

Rhabdoviridae is a diverse viral family infecting a wide range of host taxa from vertebrates to invertebrates to plants (Dietzgen *et al.* 2017). Within *Rhabdoviridae*, the species *Rabies lyssavirus* is well-known for the acute progressive encephalitis it causes in mammals. Vampire bats are the main wildlife reservoir for the virus in Latin America, playing an important role in transmission to livestock and humans through biting (Schneider *et al.* 2009). Vampire bat rabies virus (VBRV) is endemic to regions of Peru east of the Andes and in the Amazon (Streicker *et al.* 2016), but given the relatively low seroprevalence it was surprising to detect VBRV in the saliva and feces of apparently healthy bats at two colonies using metagenomic sequencing.

Diamond protein blast identified contigs closely matching VBRV in four saliva pools: two single-colony pools (CAJ4_SV and HUA1_SV) and two multi-colony pools (CAJ_L_SV and HUA_H_SV). Contigs closely matching VBRV were also detected in the corresponding single- (HUA1_F) and multi-colony (HUA_H_F) fecal pools from one locality. All pools in which VBRV was detected contained some of the same individuals, so these detections likely represent the presence of VBRV in two localities in one or several individuals. The contigs were analyzed by nucleotide blast against the Genbank nt database and all matched to VBRV with a high percent identity (Table 3.5), with top matches including previous sequences from Peru (Table C7). However, the shotgun nature

of the data means that reads are not necessarily located in the genes targeted by PCR in previous studies within Peru, and that some contigs had best matches to full genome sequences of other rabies variants (Troupin *et al.* 2016). Phylogenetic analysis was not performed for VBRV sequences generated in this study for the same reason. Many contigs matched best with a rabies variant found in a dog (Table C7; Genbank Accession KX148268), but this likely represents a cross-species transmission from vampire bat to dog, as this sequence falls within the bat clade of the rabies phylogeny (Troupin *et al.* 2016).

Table 3 5 Summary of blast analysis of VBRV contigs in vampire bat saliva.
VBRV contigs were detected in saliva of single and multi-colony pools from vampire bats in Peru. This table summarizes the individual contig nucleotide blast results presented in Table C7.

Sample ID	Mean contig length (range) (bp)†	Mean % ID to VBRV (range)‡
CAJ_L_SV	485.3 (254-1052)	98.3 (97-99)
HUA_H_F	258.5 (234-308)	98 (97-99)
HUA_H_SV	408.4 (243-549)	98 (97-99)
CAJ4_SV	395.8 (244-570)	98 (97-99)
HUA1_F	359 (274-444)	97 (96-98)
HUA1_SV	471.3 (248-842)	98.6 (98-100)

†Contig length mean and range describe the VBRV-matching contigs generated by SPAdes
‡Mean percent identity summarizes the percent identity at the nucleotide level of the blast results from the VBRV-matching contigs in each sample

The two localities where VBRV was detected are known to have endemic rabies virus circulation (Streicker *et al.* 2016), but the prevalence appears strikingly high. VBRV has been previously detected by serology in at least one of the sampled colonies (CAJ4) (Streicker *et al.* 2012b), so the colonies where bats were sampled could have been experiencing an outbreak of the virus, which is thought to be regionally maintained through meta-population dynamics (Streicker *et al.* 2012b; Blackwood *et al.* 2013). Yet a previous study found a global VBRV seroprevalence of 10.2%, with a range of 3.3 – 28.6% across years in colonies where at least four individuals were sampled (Streicker *et al.* 2012b). Given that these antibody detections could represent an infection that took place in the past, the number of current infections would be expected to be much lower. The metagenomic pools are made up of 10 individuals, so at least 1 in 10 bats at each of these sites is apparently shedding viral RNA. The individuals sampled for this study have not yet been tested by our group for VBRV exposure by serology, but will be examined using the method of rapid fluorescent focus inhibition test (RFFIT) to detect neutralizing antibodies. In other mammals, a rabies infection is often lethal (Rupprecht *et al.* 2011) but

high seroprevalence in vampire bats and other bat species suggests that bats may frequently acquire immunity after surviving an infection (Turmelle *et al.* 2010). In combination with recent findings suggesting that *Lyssavirus* RNA excretion and seropositivity are not perfectly correlated (Robardet *et al.* 2017), there is clearly more to understand about the dynamics of rabies infection, persistence, and shedding in vampire bats.

The close matches to sequences previously generated by our group raises the possibility of contamination, which is a real concern with a sensitive method such as metagenomics. The detection of VBRV in feces also appears somewhat suspect, but the results are believed to be real based on several lines of evidence. These reads are unlikely to be contamination from amplicon-based studies by our group (Streicker *et al.,* unpublished) as the metagenomic samples were processed through a strict CVR laboratory pipeline; samples were extracted in a room for only non-propagated clinical samples and library prepared in a room for only non-amplified material. The shotgun reads do not resemble amplicon data, as reads were scattered across the genome and not focused on genetic regions targeted by PCR in previous studies (e.g. Nucleoprotein gene; Streicker *et al.* 2016). VBRV amplicons generated by our group range from 358 - 398bp, while metagenomic contigs varied widely in length, with many being much longer or shorter than this (Table C7). VBRV was only detected in fecal pools where the virus was also detected in corresponding saliva pools, but if this was widespread contamination it would be expected in other samples that were not associated with the same individuals. Finally, fecal and saliva samples from these pools were not processed directly adjacent to one another during extraction or library preparation.

Detecting VBRV in saliva and fecal samples from apparently healthy individuals using metagenomics is methodologically novel. Molecular studies of rabies typically isolate RNA from brain tissue of known or suspected infected individuals (Streicker *et al.* 2016; Troupin *et al.* 2016)*,* although laboratory studies based on experimental injections of high doses of rabies have detected the virus in saliva (Aguilar-Setien *et al.* 1999; Almeida *et al.* 2005). A recent field-based study of European bat lyssavirus (EBLV), a *Lyssavirus* in the same genus as rabies, detected viral RNA in the saliva of apparently healthy bats (Robardet *et al.* 2017). Rabies virus RNA in feces has not to our knowledge been reported from a field surveillance study, although other *Lyssavirus* species have been detected in bat feces (Allendorf *et al.* 2012) and proposed to be excreted through the digestive system (Kading *et al.* 2013). The sequencing of VBRV in an untargeted metagenomic survey of field-

caught bats demonstrates a novel method of detecting rabies virus in wildlife, although a comparison of the detection limit of metagenomics compared to established PCR techniques would be a useful follow-up study.

3.4.8 Reoviridae

Rotaviruses (RV, family *Reoviridae*) cause acute diarrhea in humans as well as other mammals and birds, and are made up of 11 segments (lengths ranging from 200 bp – 3 kb) that encode six structural and five non-structural proteins. RVs are classified into antigenic species RVA-RVJ. Most bat RVs described thus far have been in the RVA group (Esona *et al.* 2010; He *et al.* 2013b; Xia *et al.* 2014; Asano *et al.* 2016; Yinda *et al.* 2016), although recent studies have reported RVH (Kim *et al.* 2016) and RVJ (Bányai *et al.* 2017) antigenic species in bats.

RV-like contigs were detected in four single-colony fecal pools (AYA14_F, AYA15_F, CAJ4_F, HUA1_F) and three multi-colony fecal pools (AAC_H_F, CAJ_L_F, HUA_H_F). The longest contig detected was 3,569bp (HUA_H_F), which is consistent with the segmented nature of RV genomes. An initial nucleotide blast analysis revealed that the closest hit for most contigs was the human RVH strain B219 (Genbank EF453355-60; DQ168032-36), which was first described in humans in Bangladesh (Alam *et al.* 2006; Nagashima *et al.* 2008). In addition to strain B219, RVH has been found elsewhere in humans (Jiang *et al.* 2008), pigs (Wakuda 2011), and recently in bats (Kim *et al.* 2016). The B219 genome was used to assess pairwise genetic distances across segments for the two pools containing the highest coverage (CAJ_L_F and HUA_H_F). The percent identities were variable across different segments and also suggested that the viruses from the two localities were distinct from one another (Table 3.6).

Table 3 6 Pairwise differences between vampire bat RVH sequences and human RVH. Pairwise percent IDs are shown for the two Peru sequences compared to one another as well as between each sequence and the most closely related published genome, which is a human-infecting RVH.

Segment	Pairwise % ID Peru§	CAJ_L_F			HUA_H_F		
		Contig lengths†	Pairwise % ID‡	% cov	Contig length†	Pairwise % ID‡	% cov
VP1	74.4	903 1058	70.4	55.4	3569	70.8	99.7
VP2	76.1	273 708 847	71.8	61.6	2953	69.2	99.2
VP3	70.5	411 515	64.6	41.9	2219	58.4	98
VP4	81.6	229 269	61.4	19.8	2518	54.2	99.2
VP6	94.6	1242	72.7	96.4	1252	72.1	97.3
NSP1	70.3	446 692 753	56.3	93	1189	56.9	90.2
NSP2	96.1	928	71.2	92.2	934	70.7	91.9
NSP3	86.7	639	64.4	68.6	872	64.4	93.2
NSP4	62.5	424	49.5	57.1	735	59	8
NSP5	-	-	-	-	648	51.9	98

†Contig lengths are shown for all those matching a particular segment in each sample
‡Pairwise percent ID is the pairwise identity between the sequence and B219 (Genbank accession Genbank EF453355-60; DQ168032-36)
§Pairwise percent IDs of the two Peru sequences compared to one another

Although there are not many full RVH genome sequences available, phylogenetic analyses were conducted to confirm the placement of the vampire bat virus within the RVH group, and compare it with two related bat RVs (Kim *et al.* 2016; Bányai *et al.* 2017). RV phylogenetic analyses were performed for the segments encoding three structural proteins (VP1, VP3, VP4) which were used to classify the previously detected bat RVH (Kim *et al.* 2016), although the previous sequences did not cover the full length of each segment. CAJ_L_F and HUA_H_F contigs from Peru were compared with two other human RVH sequences, eight porcine RVH sequences (11 for VP4), one bat RVH sequence, and representative sequences from related RV groups as available for each segment (RVB, RVI, RVJ, and RVG).

The VP1 and VP3 phylogenies indicated that the vampire bat RVH sequences formed their own clade which was most closely related to two human RVH species (Figure 3.9A-B), while the VP4 phylogeny placed the vampire bat clade outside of other RVH species, although this placement was not strongly supported (Figure 3.9C). Segmented genomes such as RVs can reassort, so it is not uncommon to observe discordant phylogenies

between different segments, and indeed this is an important mechanism for generating diversity and driving evolution (McDonald *et al.* 2016).
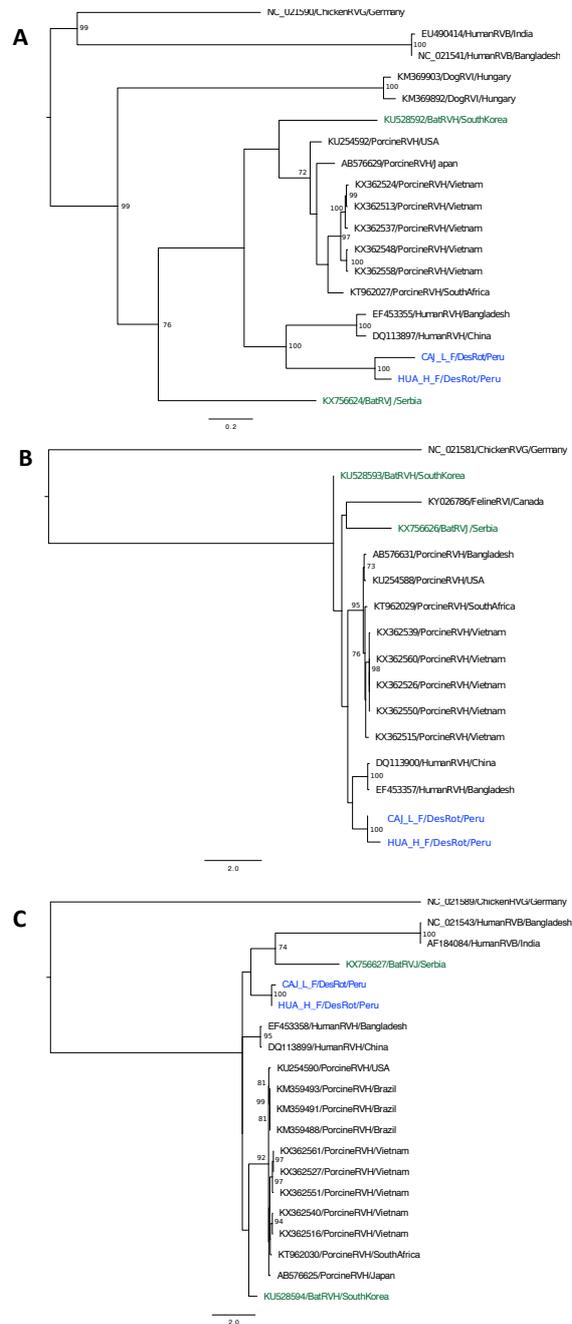
Figure 3 9 Rotavirus H phylogenies.
Maximum likelihood trees based on three RV segments: (A) 3,593 bp of VP1 (B) 2,484 bp of VP3 and (C) 2,648 bp of VP4. Analyses included two Peru RVH sequences (blue), other bat RV sequences (green) and RV sequences from other taxa and groups (black). Non-bat sequences are labeled with Genbank accession number, host species, and country of origin. Previous bat sequences are labeled with Genbank accession number, RV group, and country of origin. Peru sequences are labeled with pool, bat host species abbreviation, and country of origin. Phylogenetic analysis was performed using the models GTRGAMMAI (VP1 and VP3) and GTRGAMMA (VP4). Published viruses are detailed in Appendix C, Table C6. Bootstrap values of >70 are displayed. The scale bar represents the mean expected rate of substitutions per site.

Interestingly, the vampire bat RVH was not most closely related to the bat RVH detected in South Korea (Kim *et al.* 2016) based on any segments analyzed, and there was no evidence of bat RVH monophyly, which is consistent with previous studies of RVAs suggesting that there are not bat-specific clades of RVs (Yinda *et al.* 2016). The most recent RV species classification framework involves analyzing the VP6 gene (Matthijnssens *et al.* 2012), which was not undertaken here in the interest of comparing novel sequences to the existing bat RVH, for which VP6 was not sequenced. RVH remains a poorly characterized group, such that there could be unrecognized bat viruses closely related to the vampire bat RVH. However, the bat RVH sequences not grouping together and the vampire bat sequences grouping most closely with human RVH based on VP1 and VP3 suggest the possibility of historical cross-species transmission between bats and humans, which has also been noted in previous studies of bat RVs (Esona *et al.* 2010; Xia *et al.* 2014; Asano *et al.* 2016).

3.4.9 Adenoviridae

Adenoviruses (AdV, family *Adenoviridae*) have a broad host range including mammals, birds and reptiles, including diverse bat species globally (Maeda *et al.* 2008; Sonntag *et al.* 2009; Li *et al.* 2010b; Hackenbrack *et al.* 2017) and specifically vampire bats (Lima *et al.* 2013; Wray *et al.* 2016). Bat AdV genomes range in size from 29,162 – 38,073 bp and exhibit a wide range in G+C content, with high levels of genomic diversity suggesting that bats may have been an ancestral host for AdVs and played a key role in AdV evolution (Tan *et al.* 2017).

AdV-like contigs were detected in 21 multi-colony and single-colony pools; these included twelve single colony fecal pools, five multi colony fecal pools, three single colony saliva pools, and one multi colony saliva pool. AdV genomes are large and the contigs were scattered across the genome, so a 307 bp fragment of the DNA polymerase gene was examined which was also analyzed in a previous study of vampire bat AdVs (Wray *et al.* 2016). First, AdV-matching contigs were aligned to a representative vampire bat AdV DNA polymerase gene fragment (Genbank Accession number KX774307) and eight contigs were retained that overlapped mostly or fully with the published sequence (API140_F, API141_F, AYA7_F, AYA14_F, CAJ4_F, LMA5_F, LMA6_F, and LR2_F). The resulting phylogeny showed Peruvian vampire bat AdVs to be located within a clade of vampire bat AdVs, including previous sequences from Guatemala and Brazil (Lima *et al.* 2013; Wray *et al.* 2016) (Figure 3.10).
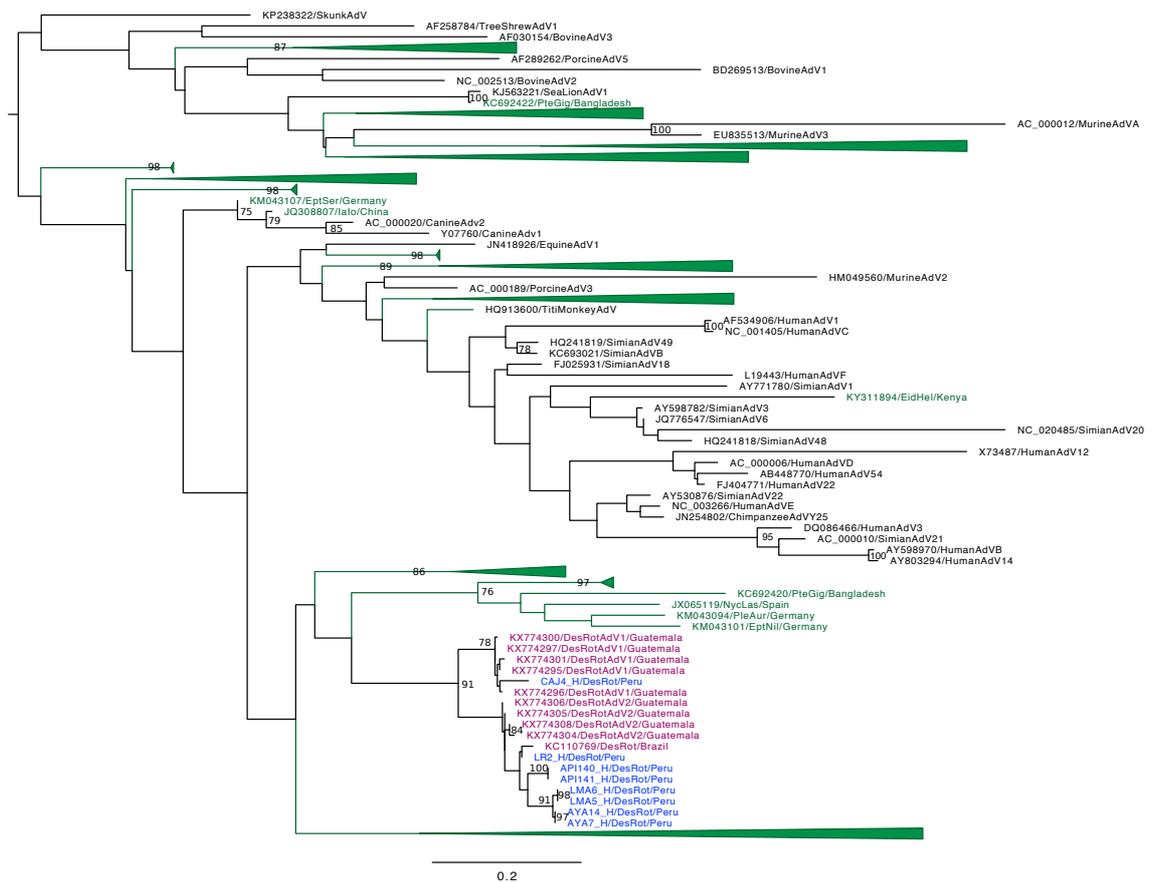
Figure 3 10 Adenovirus phylogeny
Maximum likelihood tree based on a 307 bp segment of 193 mammalian AdV sequences in the genus *Mastadenovirus* including eight Peru AdV sequences (blue), vampire bat AdV sequences (red), non-Neotropical bat AdV sequences (green) and sequences from AdVs infecting other species (black). Some non-Neotropical bat clades are collapsed and colored green. Non-bat sequences are labeled with Genbank accession number, host species, and country of origin. Previous bat sequences are labeled with Genbank accession number, bat host species abbreviation, and country of origin. Peru sequences are labeled with pool, bat host species abbreviation, and country of origin. Phylogenetic analyses were performed with the substitution model GTRGAMMAI. Published viruses are detailed in Appendix C, Table C6. Bootstrap values of >70 are displayed. The scale bar represents the mean expected rate of substitutions per site.

Most of the Peru contigs were within the previously described clade Dr-AdV2 (*Desmodus rotundus* Adenovirus 2), while one sequence (CAJ4_F) was in the clade Dr-AdV1 (*Desmodus rotundus* Adenovirus 1); however, this sequence was missing significant data (47%), so placement should be interpreted with caution. Within the Dr-AdV2 clade, there was some geographic structuring of AdVs, with samples from the same locality grouping together, suggesting that AdVs could be a useful viral family for examining host movement. These results corroborate previous findings suggesting that there is some degree of host specificity in vampire bat AdVs (Wray *et al.* 2016), even across a wide geographic scale in Latin America.

3.4.10   Picornaviridae

Picornaviruses (PicoV, family *Picornaviridae*) are a viral family that infects a wide range of host species, including several groups identified previously in bats (Li *et al.* 2010a; Lau *et al.* 2011; Kemenesi *et al.* 2015; Wu *et al.* 2016; Lukashev *et al.* 2017). As PicoVs are characterized by a fast mutation rate and high levels of genetic diversity, it has been suggested that PicoVs would be a useful viral marker for ecological studies of their hosts (Lukashev *et al.* 2017).

PicoV-like contigs (partial genome sequences ranging from 232 – 6774 bp) were found in twelve fecal pools, including four multi-colony fecal pools and eight single colony fecal pools. Contigs were first analyzed by nucleotide blast, from which it was evident that there were two different groups of sequences within the PicoV family. Contigs from pools HUA1_F and HUA4_F were similar to a group of previously characterized bat PicoVs (Lau *et al.* 2011; Lukashev *et al.* 2017) related to *Enterovirus*. Contigs from pools AYA12_F, LMA5_F, LMA6_F and CUS8_F matched most closely to a bat PicoV from China (Wu *et al.* 2016) that is similar to *Parechovirus*. As the PicoV family is highly genetically diverse, two separate phylogenetic analyses were performed, one for the *Enterovirus*-like contigs and one for the *Parechovirus*-like contigs.

For *Enterovirus*-like sequences, analyses were performed on a fragment of the 3D polymerase genome region that encodes the RdRp which had been previously characterized in bat species from Europe and Asia (Lukashev *et al.* 2017). In contrast, bat viruses are not as well-known from the *Parechovirus*-like part of the PicoV family and there is not a fragment of the genome that has been used in any amplicon-based studies. The other bat sequence from this part of the PicoV tree was also generated from a metagenomic study (Wu *et al.* 2016). For the *Parechovirus*-like analysis all the closely related genomes from the ICTV *Picornaviridae* family tree were included (https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/picornaviridae) as well as closely matching nucleotide blast hits. The *Enterovirus*-like contigs fell within the clade of previously characterized bat PicoVs that had been found across Europe and Asia (Figure 3.11A), while the *Parechovirus*-like contigs were most closely related to the bat PicoV from China (Figure 3.11B).
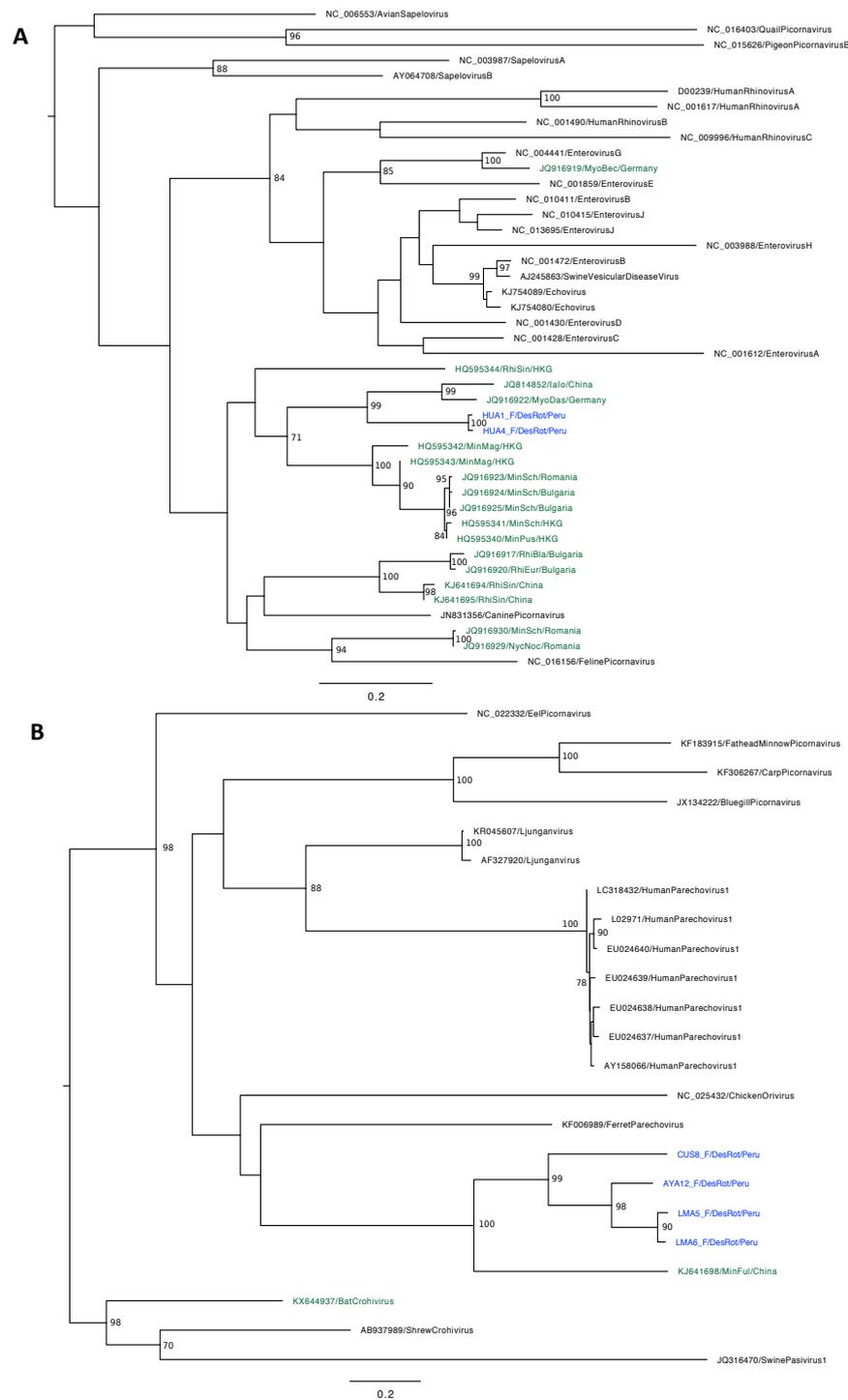
Figure 3 11 Picornavirus phylogenies.
Maximum likelihood trees of (A) *Enterovirus*-like sequences based on a 265 amino acid alignment of the 3D polyprotein of 43 sequences, including two Peru vampire bat fecal pool contigs (HUA1_F and HUA4_F) and 17 other bat sequences and (B) *Parechovirus*-like sequences based on a 503 amino acid alignment of the 3D polyprotein of 23 sequences including four Peru vampire bat contigs and two sequences from other bats. Peru sequences are in blue, other bat sequences are in green and sequences from viruses infecting other species are in black. Non-bat sequences are labeled with Genbank accession number and virus species name. Previous bat sequences are labeled with Genbank accession number, bat host species abbreviation, and country of origin. Peru sequences are labeled with pool, bat host species abbreviation, and country of origin. Phylogenetic analyses for both *Enterovirus*-like and *Parechovirus*-like contigs were performed with the substitution model LG+I+G. Published viruses are detailed in Appendix C, Table C6. Bootstrap values of >70 are displayed. The scale bar represents the mean expected rate of substitutions per site.

PicoVs have not been reported previously in Neotropical bats, although there would be no reason to suspect their absence given a wide distribution across Old World and North American bats (Li *et al.* 2010a; Wu *et al.* 2016; Lukashev *et al.* 2017). The vampire bat *Enterovirus*-like contigs fell within a clade of previously described bat viruses. The *Enterovirus*-like PicoVs could represent a new species, although taxonomic criteria for defining new species within this group of bat PicoVs are yet to be established (Lukashev *et al.* 2017). The vampire bat *Parechovirus*-like contigs are also most closely related to another bat virus, but no other bat viruses have been characterized in this part of the PicoV family. Nonetheless, the other *Parechovirus*-like bat virus was detected in China, so this groups of bat PicoVs appears to have a global distribution. The *Parechovirus*-like contigs appear to be widespread yet genetically distinct across Peru, supporting the idea that with their fast mutation rate PicoVs could be a useful viral family for examining host movement (Lukashev *et al.* 2017).

3.4.11   *Retroviridae*

Foamy viruses (FV, family *Retroviridae*) are a group of exogenous retroviruses that are common in various mammalian species (Pinto-Santini *et al.* 2017), and have been reported twice previously in bats including one detection in a Neotropical bat (Wu *et al.* 2012; Salmier *et al.* 2017). Although FVs generally co-speciate with their hosts in primates (Switzer *et al.* 2005), cross-species transmission has also been observed between simian FVs and humans (Betsem *et al.* 2011; Mouinga-Ondeme *et al.* 2011), suggesting the potential for transmission from bats to other species. FVs are transmitted primarily through saliva by biting, grooming and food-sharing (Pinto-Santini *et al.* 2017), behaviors in which vampire bats engage both with other vampire bats and across species.

FV-like contigs (partial genome sequences ranging from 233 – 2089 bp) were widespread in Peru, present in 20 single-colony saliva pools and two single-colony fecal pools. The DNA polymerase (*pol*) gene, which is typically used in FV phylogenetics, was analyzed by first aligning FV-matching contigs to the sequence available from a published bat FV genome (partial *pol* and *env* gene; Wu *et al.* 2012) and then extracting a fragment of the *pol* gene that was found in multiple vampire bat pools. Other mammalian FV sequences for which this region was available were included in the phylogenetic analysis to determine whether the bat sequences grouped together, or whether this could represent a potential cross-species transmission, for example, from non-human primates in Peru in which extensive diversity of FVs has been reported (Ghersi *et al.* 2015). The resulting phylogeny

showed the vampire bat FVs all grouped together, with the closest relative being the other bat FV (Figure 3.12).
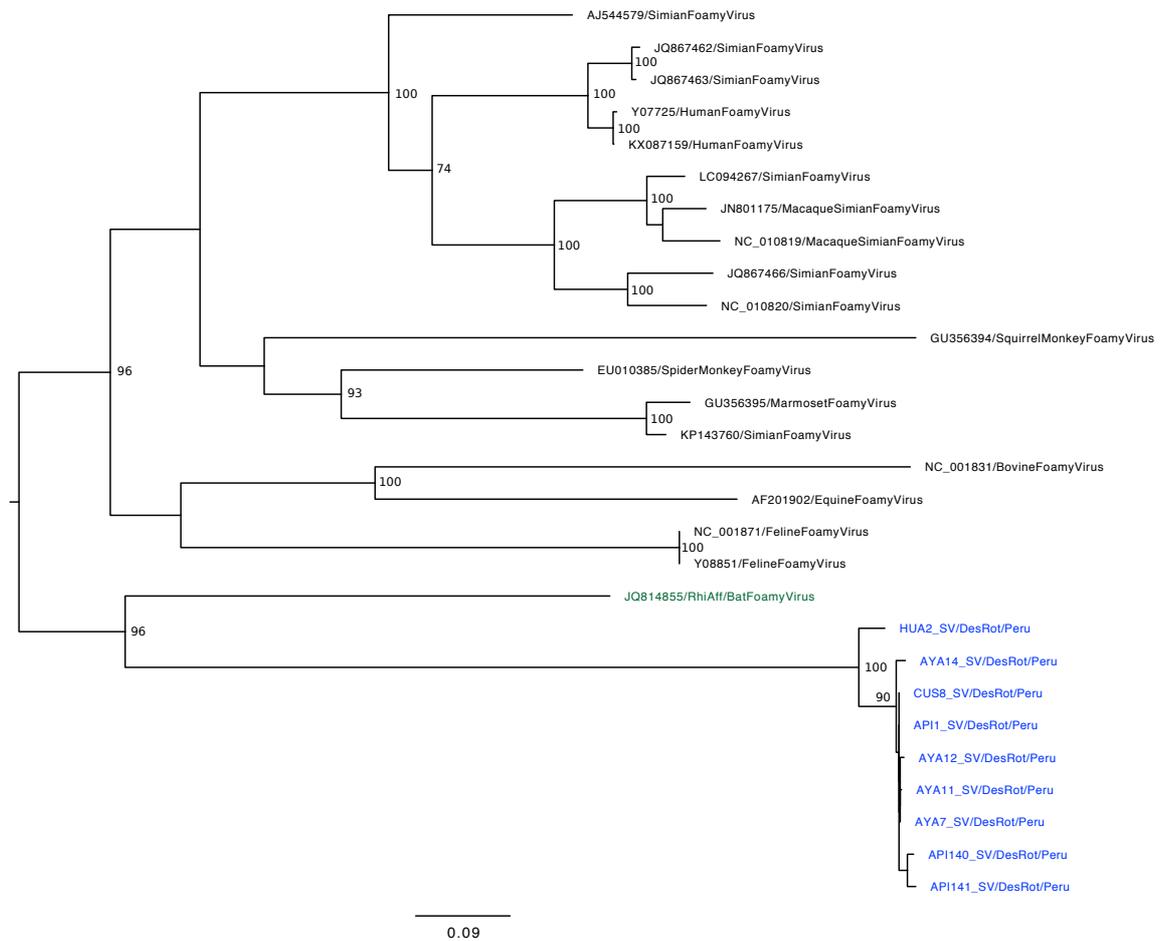
Figure 3 12 Foamy virus phylogeny.
Maximum likelihood tree based on a 1,322 bp of the *pol* gene from nine Peru FV sequences (blue), other bat FV sequence (green) and sequences from FVs infecting other species (black). Non-bat sequences are labeled with Genbank accession number and virus name. Previous bat sequences are labeled with Genbank accession number and virus name. Peru sequences are labeled with pool, bat host species abbreviation, and country of origin. The phylogenetic analysis was performed using the model GTRGAMMAI. Published viruses are detailed in Appendix C, Table C6. Bootstrap values of >70 are displayed. The scale bar represents the mean expected rate of substitutions per site.

FVs have only been described twice previously in bats, and the vampire bat FVs were most closely related to the previously characterized bat FV from China (Wu *et al.* 2012). Unfortunately it was not possible to compare vampire bat sequences to the other FV that had been characterized in a Neotropical bat (Salmier *et al.* 2017) because both studies used shotgun metagenomics and there was no overlap between sequences. FVs were almost ubiquitous among the vampire bat colonies sampled in Peru, and there appears to be some amount of geographical structuring within the vampire bat FVs. Other viral species in the family *Retroviridae* have been used as genetic markers in which the evolutionary relationships between viruses in different locations track the population structure of their

hosts (Biek *et al.* 2006; Antunes *et al.* 2008), suggesting a potential future avenue of research with vampire bat FVs.

An alternative explanation for the ubiquity of FVs is that they can occasionally endogenize into the host genome (Katzourakis *et al.* 2009; Han & Worobey 2012). It can be difficult to determine from short-read metagenomic sequence data whether a virus is endogenous or exogenous, but methods have been proposed based on patterns of read diversity (Mourier *et al.* 2015), so in the future it would be interesting to establish whether the vampire bat FV is widespread due to being highly prevalent or endogenous. Given also that FV are primarily transmitted by saliva, through activities such as biting and grooming, vampire bats might be able to transmit the FV to conspecifics through grooming or to livestock prey through biting.

3.5 Conclusions

Bats play important roles as viral hosts, with metagenomic studies revealing new and highly divergent taxa in bat species globally (e.g. Li *et al.* 2010a; Donaldson *et al.* 2010; Baker *et al.* 2013; Dacheux *et al.* 2014; Wu *et al.* 2016; Salmier *et al.* 2017). To evaluate the possible relationship between viral diversity and disease emergence from bats into humans or domestic animals, we first need a detailed understanding of bat viral communities. Ecologically well-connected hosts, such as vampire bats, represent an interesting system in which to conduct such in depth studies of baseline viral diversity. The work presented in this chapter builds upon previous descriptive metagenomic studies of novel viruses in vampire bats (e.g. Salmier *et al.* 2017; Escalera-Zamudio *et al.* 2017), sampling across a large geographic area and using an unbiased metagenomic protocol to thoroughly characterize vampire bat viral communities in Peru.

Numerous viral families were detected in feces and saliva, comprising both vertebrate-infecting viral taxa for which vampire bats serve as potential hosts and non vertebrate-infecting taxa associated with commensal bacteria or acquired from the environment. Viral community composition, but not richness, differed between fecal and saliva samples, demonstrating body habitat compartmentalization of viruses. Phylogenetic analyses of eight vertebrate-infecting viral families suggested that most vampire bat viruses fall within bat-specific clades, without evidence of livestock or humans acting as a major source of viral diversity in vampire bats. For example, CoVs and AdVs discovered in vampire bats from Peru fell within the previously described diversity in Neotropical bats. However,

DrHDV was highly distinct from human HDV and has never been reported before in bats. The vampire bat RVH did not group with other bat RVH species, although this group is poorly characterized and could include undescribed bat viruses. In contrast, vampire bat HEV clustered with other bat HEV sequences from around the globe, exhibiting an apparently stable relationship with bat hosts. Although PicoVs and FVs have not been well characterized previously in Neotropical bats, their relative ubiquity and evidence of genetic structuring between localities suggests that these groups could be useful for studies of host/virus relationships across space. In summary, the results presented in this chapter expand our understanding of vampire bat viral communities, creating a detailed picture of viral diversity in a host that is ecologically well-connected and serves as a key wildlife reservoir.

4 Ecological metagenomics reveals effects of host demography and environmental heterogeneity on viral diversity in vampire bats

4.1 Abstract

Bats host many viral taxa, exhibiting both associations with diverse communities of viruses and deep evolutionary relationships with individual viral families. Previous studies suggest that viral richness in bats is influenced by interspecific differences in ecological and life history traits, but the intraspecific factors structuring bat viral communities remain unexplored. Due to their unique feeding habits and anthropogenic modification of the landscape which has resulted in increased levels of contact with humans and domestic animals, common vampire bats (*Desmodus rotundus)* represent an interesting study species in which to examine the factors shaping intraspecific viral diversity. Here I tested the spatial, demographic and environmental correlates of viral diversity at the colony level in vampire bats. Generalized linear models were used to test for correlations between local ecological variables and three measures of diversity for each colony (richness, taxonomic diversity, and community composition). There was a longitudinal gradient of diversity in saliva viruses, with the northwest of Peru having the highest diversity. In contrast, sites in the Amazon generally had higher fecal viral richness and distinct community composition. Fecal viral diversity increased with the proportion of juveniles in a colony, and there appeared to be an effect of environmental context which encompassed elevation, climate, and sometimes local livestock density. Vertebrate-infecting viral communities tended to be highly distinct between colonies and were inconsistently correlated with geographic distance and host movement. Overall, these findings suggested that intraspecific drivers of viral communities are complex and that previous studies focusing on a single sample type, individual or environment are unlikely to have captured the full extent of viral diversity within a species.

4.2 Introduction

Viruses occur across all environments that support life, where they play crucial roles in ecosystem function, regulation of population dynamics and in the health of their hosts (Fuhrman 1999; Suttle 2005; Virgin 2014; Manrique *et al.* 2016). For human and animal health, the structure of viral communities (i.e. the number and identity of viral taxa present in a particular location, time or host species) is critically important because more diverse

pools of viruses may be more likely to contain taxa that can emerge in new host species (Morse 1993; Wolfe *et al.* 2000). While our ability to describe viral taxa within hosts has been transformed by technological developments in sequencing-based detection of viruses (Mokili *et al.* 2012; Lecuit & Eloit 2013), studies that compare viral communities across species, landscapes or time points remain a rarity (Suzán *et al.* 2015; Anthony *et al.* 2015; Brierley *et al.* 2016; Olival *et al.* 2017). The resulting gap in understanding the ecological determinants of viral richness and community composition (i.e. alpha and beta diversity) limits our ability to forecast viral diversity or predict its consequences on health or ecosystem processes.

The factors that structure viral communities likely operate across hierarchical scales, ranging from life history traits that vary between species (e.g. body size, range overlap) to population-specific factors that vary within a species (e.g. population size, connectivity, local climate) to variation among demographic groups within a population (e.g. age, sex, reproductive status). Most previous studies on the ecological determinants of viral communities have been comparative, focusing on interspecific traits correlated with viral richness across host taxa (Turmelle & Olival 2009; Luis *et al.* 2013; Gay *et al.* 2014; Olival *et al.* 2017). However, the intraspecific factors associated with differences in viral communities within and between populations remain poorly understood, aside from in the context of humans (Minot *et al.* 2011; Robles-Sikisaka *et al.* 2013; Manrique *et al.* 2016; Rampelli *et al.* 2017). There is therefore a need to explore the broad-scale and local factors that structure viral communities more widely.

A decrease in community similarity with increasing distance is a common ecological phenomenon (Nekola & White 1999), for which there is some evidence in environmental viral communities (Chow & Suttle 2015). For viruses, such a relationship might reflect host movement, which plays an important role in the spatial distribution of individual viral taxa (Biek & Real 2010; Côté *et al.* 2012; Brunker *et al.* 2012) and in structuring viral communities (Anthony *et al.* 2015). The extent to which host movement shapes viral communities could vary for individual viral taxa with different dispersal and transmission mechanisms, as well as those with differing degrees of obligate dependence on their host (Barrett *et al.* 2008). Correlations with geographic distance or host genetic distance might reflect different drivers of viral community structure; for example, if host movement plays a primary role, viral communities would likely correlate with host genetic distance, while geographic distance might be correlated if communities are strongly influenced by transient viruses from the environment. In addition to distance, host demographic and local

environmental factors may also be important in shaping viral richness and community composition.

Interspecific comparative studies of microparasites, as well as empirical studies of macroparasites, have identified demographic and environmental factors that differentiate parasite communities (Nunn *et al.* 2003; Guernier *et al.* 2004; Ezenwa *et al.* 2006; Lindenfors *et al.* 2007; Arriero & Moller 2008; Bordes *et al.* 2011; Kamiya *et al.* 2013; Gay *et al.* 2014; Nunn *et al.* 2014; Poulin 2014). For example, host population structure and levels of fragmentation can impact parasite richness and disease dynamics (Turmelle & Olival 2009; Gay *et al.* 2014; de Thoisy *et al.* 2016). The effect of host age on parasite diversity is complex, with some studies observing a positive relationship as hosts accumulate parasites over time (Lo *et al.* 1998; Nunn *et al.* 2003) and others a negative or non-linear relationship that has been related to age-specific differences in immunity or the probability of parasite encounter (Benavides *et al.* 2011; Poirotte *et al.* 2015). Males and females can exhibit different propensity for parasite infection due to behavioral and physiological differences (Poulin 1996a; Zuk & McKean 1996; Reimchen & Nosil 2001; Negro *et al.* 2010). Finally, host population size can affect the persistence of individual viral taxa (Bartlett 1957), which could have impacts at the viral community level, as has been shown for other parasites (Nunn *et al.* 2003; Lindenfors *et al.* 2007). These studies offer useful hypotheses about factors that might be important in structuring intraspecific viral communities, for which the drivers of diversity remain relatively unexplored.

Pathogen richness, including that of viruses, can also be influenced by local environmental conditions (Guernier *et al.* 2004; Nunn *et al.* 2005; Dunn *et al.* 2010; Schotthoefer *et al.* 2011). Latitude, and the climatic variables such as temperature and precipitation for which it often serves as a proxy, are significantly correlated with parasite richness in a variety of taxa (Guernier *et al.* 2004; Nunn *et al.* 2005; Lindenfors *et al.* 2007; Bordes *et al.* 2011). With increasing elevation, a decrease in macrofauna diversity is often observed (Lomolino 2001), although bacterial communities appear to deviate from this pattern (Fierer *et al.* 2011; Wang *et al.* 2011; Muletz Wolz *et al.* 2017). Greater range overlap with other hosts (Davies & Pedersen 2008; Krasnov *et al.* 2010; Huang *et al.* 2013) and greater local host species richness (Harris & Dunn 2010; Kamiya *et al.* 2014) have both been associated with increased parasite diversity. Finally, anthropogenic land-use conversion can have complex effects on parasite richness (Gillespie *et al.* 2005; McKenzie 2007; Gay *et al.* 2014; Bernardo *et al.* 2018) with potential implications for disease transmission (Murray & Daszak 2013; Gottdenker *et al.* 2014).

Bats are an ideal host group in which to investigate viral communities, as they are associated with many viruses (Calisher *et al.* 2006; Luis *et al.* 2013), including high-profile zoonotic pathogens (Li *et al.* 2005; Leroy *et al.* 2005; Memish *et al.* 2013). Literature reviews (Calisher *et al.* 2006; Hayman *et al.* 2012; Hayman 2016) and comparative analyses (Turmelle & Olival 2009; Luis *et al.* 2013; Webber *et al.* 2017) suggest that interspecific viral richness in bats is influenced by differences in ecological and life history traits such as host population structure, threatened status, sympatry, longevity, litter size, and number of litters per year (Turmelle & Olival 2009; Luis *et al.* 2013). However, the factors structuring viral communities within a bat species have yet to be examined.

Common vampire bats are an important reservoir host for rabies virus in Latin America (Carini 1911; Pawan 1936) and also harbor a variety of other viral taxa (Brandão *et al.* 2008; Drexler *et al.* 2012a; Fagrouch *et al.* 2012; Lima *et al.* 2013; Escalera-Zamudio *et al.* 2015; Salmier *et al.* 2017). Vampire bats feed preferentially on livestock and opportunistically on humans (Voigt & Kelm 2006; Schneider *et al.* 2009; Johnson *et al.* 2014), with anthropogenic modification of the landscape resulting in increased contact and opportunities for cross-species virus transmission (Delpietro *et al.* 1992; Lee *et al.* 2012a). The wide geographic range and diverse habitats occupied by vampire bats across Latin America (Quintana & Pacheco 2007; Martins *et al.* 2007; 2009) means that populations live in areas naturally varying in climate, richness of other host species, and anthropogenic impact. Colonies of vampire bats also vary in demographic traits that might impact viral community structure, such as population size, connectivity, age structure, and sex ratio (Greenhall *et al.* 1983; Delpietro *et al.* 1992; Streicker *et al.* 2012b; Blackwood *et al.* 2013; Streicker *et al.* 2016; Delpietro *et al.* 2017). Differing degrees of spatial and genetic separation between colonies of vampire bats, along with natural variation in environmental and demographic factors, make them an ideal system in which to test hypotheses about the factors that structure intraspecific viral richness and community composition (Table 4.1).

Table 4 1 Ecological factors that could influence viral richness and community composition in vampire bats. This table presents the general hypotheses addressed in this chapter, along with specifically tested variables, data source, and predicted effect on viral richness.

Hypothesized factor	Tested variable (colony-level)†	Predicted effect on richness
Host genetic connectivity	F_{IS}	↓ Isolation (reduced gene flow) decreases viral invasion of new colonies and increases extinction
Colony size	N_c	↑ Greater viral persistence within colonies and more viral encounters externally
Age structure	Proportion adults	↑↓ Adults obtain more chronic infections over a lifetime but juveniles play key roles in viral dynamics
Sex ratio	Proportion males	↑ Males are more susceptible to infections due to behavior and physiology
Local climate	PC1 of mean temperature, temperature range, and yearly rainfall PCA (Figure D1)	↓ Climates with higher productivity have higher viral diversity; sites with high temperature and rainfall tend to be negative for PC1
Elevation	Elevation	↓ Diversity tends to decrease with elevation
Location	Longitude (latitude excluded due to correlation with climate)	↑↓ Location effects encompass a number of factors so predictions are unclear
Other hosts	Presence/absence of other bat species	↑ Higher diversity of other species provides more opportunities for cross-species viral transmission
Anthropogenic impact	Livestock density (20km radius)	↓ Low livestock density increases diversity of prey fed upon by bats

†Color codes correspond to data source. Blue are lab-generated results, purple are field observations, green are data collected from public databases and orange is a combination of field and database sources.

This chapter presents an empirical investigation of the spatial, demographic and environmental correlates of viral diversity in common vampire bats. Specifically, I addressed the following questions (1) are colonies that are closely connected, either spatially or genetically, more similar in viral diversity compared to colonies that are less connected (2) are there broad scale differences in viral diversity between the ecological regions of Peru and (3) is variation in local demographic and environmental factors correlated with differences in colony-level viral diversity? The single-colony shotgun

metagenomic data described in Chapter 3 (Table 3.2) were used to address the above questions using three measures of viral diversity: richness (alpha diversity), a novel measure of taxonomically-informed diversity, and community composition (beta diversity). The results presented here shed new light on the relatively unexplored drivers of intraspecific viral diversity.

4.3 Materials and methods

4.3.1 Field sampling

Common vampire bats were captured at colonies at 24 sites in 8 departments across Peru (Figure 4.1) between 2013-2016 (Table D1). Distances between colonies ranged from 0.32 – 775.1 km; some colonies are likely within the distances reportedly traversed by vampire bats (Trajano 1996), although the relatively small home ranges and low dispersal of vampire bats could result in low connectivity between sites (Martins *et al.* 2009; Romero-Nava *et al.* 2014). Sampling occurred throughout the year, with sites in the Andes primarily sampled during the dry season (May – August) for reasons of accessibility. Due to the confounding of sampling location with time of year (Table D1), it was not possible to test for an effect of seasonality. Bats were captured within roosts using hand nets, or while exiting roosts using mist nets and harp traps. When bats were captured exiting roosts, nets were open during the night from approximately 18:00 – 6:00 and checked every 30 minutes; a combination of 1-3 mist nets and 1 harp trap were used depending on the size and number of roost exits identified. Roosts were either natural (caves, trees) or man-made structures (abandoned houses, tunnels, mines) inhabited by bats. When exact roost locations were unknown, bats were captured while foraging at nearby livestock pens.

Upon capture, bats were placed into individual cloth holding bags before being processed. Individual bat data were recorded including age, sex, weight, forearm length, and reproductive status. Bat age (adult, sub-adult, or juvenile) was determined by examining the level of fusion of the phalangeal epiphyses as described in Streicker *et al.* (2012b). Reproductive status was assessed by the presence of scrotal testes in males and pregnancy or lactation in females (Streicker *et al.* 2012b). Capture records from 2016 (the year most metagenomic samples were taken) were used to calculate the colony-level proportion of males (as opposed to females) and the proportion of adults (as opposed to sub-adults and juveniles). Capture records from 2011-2016 were also used to establish whether any other bat species were present at each colony (Table 4.2). As the effort in accurately identifying

other species was inconsistent across sites and years, only presence-absence of other bats was considered rather than species diversity. Vampire bats were also given a uniquely numbered wing band (Porzana Inc) for mark-recapture analyses.

Census population size (N_c) for each colony was estimated from mark-recapture data (D. Dekoski) using one of three methods. For sites sampled over multiple years, Cormack-Jolly-Seber models (Cormack 1989) were implemented in the package Rcapture (Baillargeon & Rivest 2007). For sites sampled in only one year, the Petersen estimator was calculated for sites with two capture occasions and the Schnabel estimator was calculated for sites with more than two capture occasions, both with Chapman correction (Chapman 1951) in the package FSA (Ogle 2017) (Table 4.2). Sampling intervals were not consistent between years and across sites, so the same estimator could not be used. Three sites (AMA7, LR2, and LR3) were excluded because roost locations were unknown, so sites were sampled around livestock and there were few recaptures. Two other sites (API17 and AYA15) were only sampled using hand nets during the day, making it difficult to account for recapture probability, so colony size was not estimated for those sites. Two datasets were created using the most recent estimates of N_c, as colonies can undergo major changes in N_c over a short number of years (Streicker *et al.* 2012b). One dataset included all sites where N_c was estimated and another more conservative dataset included only Petersen or Schnabel estimates from 2016-2017 (Table 4.2).

Table 4 2 Colony size (*Nc*) estimates and other species presence at vampire bat colonies. Colony size estimates were generated based on mark-recapture data using different methods depending on the data. The *Nc* estimate from the most recent year is presented along with the year of that estimate. The recorded presence of other bat species within each colony and the estimated count of other species is also presented.

Site†	N_c method	N_c estimate	N_c estimate year	Other species recorded (estimated species count)
AMA7	NA	NA	NA	N (0)
AMA2	Petersen	11	2012	Y (5)
API1	Petersen	74	2016	Y (2)
API17	NA	NA	NA	N (0)
API140	CJS	536	2015	N (0)
API141	Petersen	322	2016	N (0)
AYA1	Schnabel	276	2016	N (0)
AYA7	Petersen	25	2016	Y (2)
AYA11	Petersen	39	2016	Y (1)
AYA12	CJS	24	2015	Y (2)
AYA14	Petersen	94	2017	Y (1)
AYA15	NA	NA	NA	Y (1)
CAJ1	Petersen	22	2016	Y (2)
CAJ2	Petersen	77	2016	Y (3)
CAJ4	Petersen	312	2016	Y (6)
CUS8	Schnabel	168	2017	N (0)
HUA1	CJS	122	2015	Y (5)
HUA2	CJS	47	2013	N (0)
HUA3	Petersen	288	2014	N (0)
HUA4	CJS	31	2013	Y (4)
LMA5	CJS	510	2015	N (0)
LMA6	Schnabel	207	2016	Y (1)
LR2	NA	NA	NA	Y (6)
LR3	NA	NA	NA	Y (4)

†Sites in bold are in the more conservative dataset including only Petersen or Schnabel estimates from 2016-2017

4.3.2 Metagenomic characterization of viral communities

Both saliva and fecal metagenomic datasets were generated from pools of up to 10 individuals from 24 colonies as described in Section 3.3.1, resulting in a total of 48 pools sequenced (Table 3.2; Table A1). Different sample type pools within a colony often contained some or all of the same individuals; these pools can be considered as representing a colony-level viral community (Table D1). Proportions of males and adults within a sequencing pool were positively correlated with colony-level proportions (Pearson correlation; males r=0.85, p<0.001; adults r=0.84; p<0.001), so only colony-level proportions were considered in further analyses. The colony AMA7 was excluded from

analyses as it was sampled earlier using a different method for swabbing and exhibited low viral richness (data not shown).

Analyses in this chapter focused on viral communities including only contigs made up of at least two reads; after assembly with SPAdes and classification with Diamond, contigs were filtered by length, retaining only those >300bp. Length filtering was used to construct this dataset instead of e-value because novel viruses might not be very similar to those in existing databases. The filtered dataset included 79% of total contigs at the genus level (1,517 filtered contigs, 1,932 contigs total) and 61.5% of total contigs at the family level (1,811 filtered contigs, 2,944 contigs total). Saliva and fecal contig datasets were converted into lists of viral genera using MEGAN (Appendix B2), with all subsequent analyses performed separately for the two sample types in R version 3.4.2 (R Core Team 2017).

4.3.3 Microsatellite dataset

To examine vampire bat host population structure and test for correlations with viral diversity, individual bats from the same 24 colonies were genotyped at nine microsatellite loci. Two 2mm wing biopsy punches were collected from bats in the field and immediately stored in 95% ethanol. Cryovials containing biopsies were temporarily stored at 4°C in the field before long term storage at -20°C.

Nine microsatellite loci were amplified in two multiplex reactions (Table D2). For some individuals, loci were amplified according to previously optimized conditions (Piaggio *et al.* 2008), and fragment analysis was performed at the University of Georgia Genomics Facility on an Applied Biosystems 3730xl instrument (J. Winternitz). For other individuals, amplification was carried out using a Multiplex PCR Kit (Qiagen) in 15 μL reactions containing a final concentration of 3 mM MgCl2 and 0.2 μM each primer. PCR conditions were 15 minutes at 95°C, 40 cycles (Panel A) or 35 cycles (Panel B) consisting of 30 seconds at 94°C, 90 seconds at 52°C, and 60 seconds at 72°C, followed by 30 minutes at 60°C. Fragment analysis for these individuals was performed at the University of Dundee DNA Sequencing and Services on an Applied Biosystems 3730xl instrument. Microsatellite scoring was done using either Genemarker v.2.4.0 or the microsatellite plug-in for Geneious v. 7.17 (Kearse *et al.* 2012).

To account for potential scoring discrepancies between labs and microsatellite genotyping errors, 21 individuals were genotyped using both protocols, eight of which were included

in the final dataset (the other 13 were from colonies not included in the final dataset). Scores differed in a consistent manner between the two amplification and genotyping methods, as expected given differences between labs and protocols (Ellis *et al.* 2011) and scores from samples genotyped in Glasgow were converted to allow comparison across labs (Table D2). Nineteen individuals were genotyped twice in Glasgow to ensure consistent results across replicates within the same lab. The program PEDANT (Johnson & Haydon 2008) was used to calculate maximum likelihood estimates of genotyping error rate, and to estimate what proportion of errors were due to allelic dropout and false alleles.

To ensure that estimates of population structure would be accurate, the program MicroChecker (Van Oosterhout *et al.* 2004) was used to check for evidence of null alleles within loci or populations. The program FreeNA (Chapuis & Estoup 2007) was also used to calculate null allele frequencies. The inbreeding coefficient F_{IS} (Weir & Cockerham 1984) was estimated for each locus using FSTAT v.2.9.3.2 (Goudet 1995). FSTAT was used to test for significant departures from Hardy-Weinberg equilibrium for within population F_{IS} using 1000 permutations of a randomization test and Bonferroni correction.

Error rates per marker were relatively high for samples re-genotyped across different labs, particularly at loci DeroC12 and DeroD06 (Table D2). In contrast, replicates within Glasgow showed low error rates that were comparable to previous studies (Ellis *et al.* 2011). One locus (DeroH02) showed evidence of null alleles according to both methods, with 13 populations exhibiting evidence of null alleles in MicroChecker. F_{IS} values also deviated significantly from Hardy-Weinberg equilibrium (HWE) at DeroH02 and DeroD06 (Table D2). Overall there was evidence of genotyping error and null alleles at the three loci DeroC12, DeroD06 and DeroH02. Although the effects of genotyping errors may be less severe in population level analyses compared to analyses reliant on individual identification (Taberlet *et al.* 1999; Pompanon *et al.* 2005), analyses were repeated using a six loci subset, excluding potentially problematic loci to ensure that results were consistent using the two different datasets.

Per locus microsatellite diversity indices were calculated using the program FSTAT, including number of alleles (N_A) and allelic richness (A_R) (Table 4.3). Observed heterozygosity (H_O) and expected heterozygosity (H_E) were calculated using adegenet (Jombart 2008; Jombart & Ahmed 2011) in R and ENA-corrected F_{ST} was calculated using FreeNA.

Table 4 3 Microsatellite per locus diversity indices for vampire bats.
Diversity indices were calculated using the programs FSTAT and FreeNA, and the R
package adegenet. Per locus data are shown for all individuals combined across colonies.

Locus	Number of alleles (N_A)	Allelic richness (A_R)	Expected heterozygosity (H_E)	Observed heterozygosity (H_O)	F_{ST} ENA†
DeroB03	7	3.37	0.61	0.47	0.19
DeroB10	14	6.82	0.88	0.76	0.1
DeroB11	3	1.09	0.01	0.01	0.04
DeroC12	14	5.8	0.83	0.76	0.11
DeroD06	6	1.51	0.1	0.05	0.20
DeroC07	12	4.53	0.72	0.51	0.28
DeroD12	15	4.06	0.59	0.53	0.11
DeroG12	18	5.05	0.72	0.61	0.11
Dero H02	13	3.32	0.45	0.18	0.3

†F_{ST} corrected based on the ENA method to correct for the presence of null alleles (Chapuis & Estoup 2007)

Per population statistics (Table 4.4) were calculated using adegenet (N_A, A_R, percent missing data per site, H_E, and H_O) and FSTAT (F_{IS}). Two sites (AMA2 and CAJ4) were initially genotyped for larger numbers of individuals than other sites (61 and 88 respectively) but were randomly subsampled to 30 individuals to ensure that unequal sample sizes did not affect other analyses (Puechmaille 2016). Following subsampling, AMA2 deviated significantly from HWE in the 9 loci dataset while no populations deviated significantly in the 6 loci dataset (Table 4.4; Table D3).

Table 4 4 Microsatellite per population statistics for vampire bat colonies.
Diversity indices were calculated using the program FSTAT and the R package adegenet.
Data are shown for each colony separately.

Site	N	Number of alleles (N_A)	Allelic richness (A_R)	Percent missing data	Expected heterozygosity (H_E)	Observed heterozygosity (H_O)	F_{IS}†
AMA7	25	53	38.50	0	0.60	0.59	0.038
AMA2	30	56	40.82	6.3	0.63	0.56	**0.134**
API1	29	35	26.70	7.66	0.43	0.47	-0.067
API140	29	32	28.15	6.51	0.44	0.45	0.017
API141	29	41	27.82	0.38	0.44	0.43	0.015
API17	34	37	24.75	0	0.38	0.38	0.026
AYA1	29	30	23.31	0	0.33	0.35	-0.059
AYA11	29	26	23.55	0.38	0.34	0.38	0.096
AYA12	30	31	24.12	0	0.34	0.33	-0.083
AYA14	30	31	22.00	1.48	0.31	0.34	0.037
AYA15	29	29	24.05	0	0.34	0.35	-0.091
AYA7	24	29	20.86	0.46	0.26	0.24	0.006
CAJ1	31	48	34.32	7.89	0.53	0.48	0.117
CAJ2	13	39	31.49	6.84	0.48	0.43	0.156
CAJ4	30	48	35.40	3.33	0.54	0.49	0.103
CUS8	25	33	23.64	0	0.32	0.32	0.02
HUA1	24	50	37.42	4.17	0.58	0.53	0.124
HUA2	21	38	29.70	1.59	0.49	0.39	0.222
HUA3	22	36	30.24	8.59	0.49	0.44	0.125
HUA4	21	45	36.16	2.65	0.58	0.55	0.077
LMA5	28	38	27.99	1.19	0.48	0.41	0.164
LMA6	29	36	27.41	2.68	0.50	0.46	0.084
LR2	15	45	39.10	0	0.60	0.59	0.058
LR3	18	52	41.60	0	0.61	0.58	0.072

†Significant departure from Hardy-Weinberg equilibrium indicated in bold; p-values were adjusted using the Bonferroni correction (adjusted α= 0.00023)

4.3.4 Host population structure

Population structure between colonies was examined by calculating three differentiation measures: pairwise F_{ST} (Nei 1973) which was calculated using the hierfstat package (Goudet & Jombart 2015), and the pairwise differentiation estimators G''_{ST} (Hedrick 2005; Meirmans & Hedrick 2010) and D (Jost 2008), which were calculated using the mmod package (Winter 2012). In order to confirm population structure patterns based on differentiation measures, which may be difficult to estimate using microsatellite data (Whitlock 2011), population genetic clustering was further examined using a k-means discriminant analysis of principal components (DAPC), a Bayesian clustering method

implemented in Structure, and a maximum likelihood clustering method implemented in snapclust. DAPC was performed using the adegenet package, in which the optimal number of PCs was determined using *xvalDapc* and number of clusters determined using *find.clusters*. Bayesian clustering was performed in Structure v.2.3.3 (Pritchard *et al.* 2000; Falush *et al.* 2003a) using an admixture model with correlated allele frequencies and *K*=1-10. Larger values of K were not tested because a previous study of vampire bat population genetics in Peru detected only 2-3 genetic groups using a similar dataset (Streicker *et al.* 2016). Ten iterations (chain length 100,000 steps, burn-in 10,000 steps) were performed for each value of *K*. The number of distinct genetic clusters was inferred using the ΔK (Evanno *et al.* 2005) method implemented in Structure Harvester v. 0.6.94 (Earl & vonHoldt 2012). The program CLUMPAK (Kopelman *et al.* 2015) was used to summarize results for each value of *K*; analyses were performed through the web server using CLUMPP with Greedy algorithm (Jakobsson & Rosenberg 2007) and final plots were constructed using Distruct (Rosenberg 2003). Maximum likelihood clustering was performed using the *snapclust* function within adegenet (Beugin *et al.* 2018), with the number of clusters determined using the *snapclust.choose.k* function and Bayesian information criterion (BIC). All analyses were repeated using both 6 and 9 loci datasets.

4.3.5 Environmental variables

Latitude and longitude of each site were recorded in the field using GPS; for some sites, elevation was also recorded in the field and for others elevation was obtained from latitude and longitude coordinates using the *elevation* function in the R package rgbif (Chamberlain 2017). Climate variables for each site were gathered from the WorldClim database (Fick & Hijmans 2017) using the *getData* function from the package raster (Hijmans 2017) for each point location with resolution of 5 minutes of a degree. Three local climate variables of annual mean temperature (°C), annual precipitation (mm), and annual temperature range (°C) were then analyzed by principal component analysis (PCA) to classify sites into three ecological regions (ecoregions); the Coast (desert), Andes (mountains), and Amazon (rainforest) (Figure 4.1; Figure D1).

For each site the mean livestock density was calculated within a 20km buffer, using data from the FAO GLiPHA database (Food and Agriculture Organization of the United Nations 2012) including cows, pigs, sheep and goats. Density data was downloaded separately for each species, then extracted for each site and combined across species using the packages maptools (Bivand & Lewin-Koh 2017), rgdal (Bivand *et al.* 2017) and raster.

4.3.6 Description of viral richness and taxonomic diversity

Viral richness (number of viral genera) was calculated for each colony from the >300bp contig dataset using the package vegan (Oksanen *et al.* 2017). Vertebrate-infecting viral richness was similarly calculated after first filtering by a list of vertebrate-infecting viral genera (Table B1) based on the 2017 ICTV Taxonomy (Adams *et al.* 2017).

In addition to viral richness based on the presence-absence of viral genera at each colony, a new index was created that accounted for the pairwise relatedness of viral taxa present (Appendix D3). Because existing methods based on phylogenetic distance are inappropriate for viruses, which do not share conserved genes across families (Rohwer & Edwards 2002; Mokili *et al.* 2012), hierarchical taxonomic distances were calculated between viral taxa (Warwick & Clarke 1995). Specifically, an increasing score was assigned to each pair of viral genera present based on whether they are in the same family, have the same genome structure (dsDNA, ssDNA, dsRNA, ssRNA), are composed of the same type of molecule (DNA or RNA), or none of those. A pairwise distance matrix was generated for each possible combination of viruses detected in saliva and fecal samples separately. The distance matrices were hierarchically clustered using Ward's method (Ward 1963) using the package ape (Paradis *et al.* 2004). The package picante (Kembel *et al.* 2010) was then used to calculate Faith's phylogenetic diversity measure (Faith 1992), equal to the summed branch lengths for all viral genera found at a site. This taxonomically-informed measure is referred to here as viral "taxonomic diversity" (TD). As relatedness distances were assigned arbitrary units across the viral taxonomy, the scale cannot be related directly to viral species richness; both richness and TD are included in analyses as they provide different perspectives on viral diversity (Appendix D3).

4.3.7 Geographic, environmental and demographic correlates of viral richness

Differences in viral richness and TD were compared with geographic and host genetic distances between sites. Geographic distances between sites were calculated using the function *rdist* in the package fields (Nychka *et al.* 2015). Genetic distances were calculated as F_{ST} for both 6 and 9 loci datasets, and patterns were confirmed by repeating the analysis using the differentiation measures G''_{ST} and D. Differences in viral richness were calculated by taking the absolute value of the difference in viral richness between each pair of sites. Statistical correlations were assessed using a Mantel test with 10,000 permutations.

Broad-scale differences in viral diversity between ecoregions were evaluated using generalized linear models (GLMs) for viral richness (Poisson distribution) and TD (Gaussian distribution). ANOVA Type II tests were performed using the *Anova* function of the car package (Fox & Weisberg 2011) to calculate the likelihood ratio X^2 test statistic and assess model significance. All datasets met assumptions of homogeneity of variance, while TD measures for fecal viruses were transformed by squaring to normalize model residuals. However, saliva TD residuals could not be normalized by transformation, so differences were assessed by non-parametric Kruskal-Wallis tests. When significant differences were detected by ANOVA, the multcomp package (Hothorn *et al.* 2008) was used to perform post-hoc Tukey pairwise comparisons between ecoregions.

Local drivers of viral diversity were evaluated using GLMs to test for demographic and environmental factors correlated with viral richness and TD (Table 4.1). For each dataset, a global model was built including all possible explanatory variables, which was then used to generate the sub-models upon which automated model selection (Bartoń 2018) was performed. Sub-models were built for each combination of explanatory variables, restricting the number of explanatory variables per model to 2 due to the small number of observations (N=23 colonies). All continuous explanatory variables were examined for pairwise correlations using the package corrplot (Wei & Simko 2017) and any pair of variables with a Pearson correlation coefficient r>0.5 were excluded from the same model (Figure D2). In GLM analyses of local factors influencing viral diversity, PC1 of the PCA used to define ecoregions (Figure D1) was included as a variable representing local climate rather than the categorical variable of ecoregion, which was examined separately as described above. Models were compared with Akaike information criterion corrected for sample size (AIC_c) and R^2 values were calculated for each model. Model averaging (Burnham & Anderson 2002) was then performed to estimate effect sizes and 95% confidence intervals for each potential explanatory variable using the set of GLMs in which the cumulative Akaike weight summed to 0.95. Automated model selection and model averaging were performed using the package MuMIn (Bartoń 2018). Effect sizes were standardized using partial standard deviation to account for multi-collinearity between explanatory variables (Cade 2015).

Relationships found to be important in model averaging were confirmed by constructing univariate GLMs to examine the effect of each variable individually on viral richness and TD. Univariate p-values were corrected for multiple testing within each dataset using the false discovery rate method (Benjamini & Hochberg 1995). Finally, for each dataset and

sample type, a model was built including all variables in which the model averaged effect size significantly differed from 0. These model results were examined to verify that direction and relative magnitude of effect sizes were consistent with results from model averaging. For models with a Poisson distribution, final models were tested for overdispersion using the function *dispersiontest* in the package AER (Kleiber & Zeileis 2008); no models exhibited evidence of overdispersion. For all datasets, Moran's I was calculated using the package ape to test for evidence of spatial autocorrelation in raw data and final model residuals; no datasets or model residuals exhibited evidence of spatial autocorrelation.

The estimate of colony size (N_c) was missing from five sites, so the more and less conservative datasets (detailed above) were each analyzed for a univariate effect on richness and TD. In most cases, no relationship was found with either N_c dataset for saliva or feces and N_c was excluded from GLMs as an explanatory variable as its inclusion generated missing data. However, N_c was positively correlated with fecal vertebrate-infecting TD, using both the more and less conservative N_c datasets. The same model averaging and univariate testing above was performed using the less conservative dataset including N_c to determine whether colony size remained significant in the context of other variables, and whether patterns observed in other variables changed with its inclusion.

4.3.8 Geographic, environmental and demographic correlates of viral community composition

Differences in viral community composition between sites were assessed using Jaccard distances, which were calculated from presence-absence data using the package vegan. Differences between ecoregions for each sample type were statistically assessed using PERMANOVA (McArdle & Anderson 2001) with 10,000 permutations in the package vegan and visualized using a principal coordinate analysis (PCoA) in the package ape. Jaccard distances of viral communities were compared with geographic and host genetic distances between sites using a Mantel test with 10,000 permutations.

Demographic and environmental factors that were significantly correlated with viral richness and TD were also examined for a potential effect on viral community composition using PERMANOVA with 10,000 permutations and a GLM-based approach implemented in mvabund (Wang *et al.* 2012) as described in Section 3.3.4. When multiple variables

were tested for the same dataset, p-values were corrected using the false discovery rate method.

4.4 Results

4.4.1 Metagenomic sequencing summary

Using the length-filtered contig dataset, metagenomic sequencing revealed 108 viral genera (46 families) of which 44 were vertebrate-infecting (20 families) (Table 3.2). The mean number of total viral genera detected per colony was 7.9 for saliva samples (range 1-18) and 10.2 for fecal samples (range 5-16) (Figure 4.1). There was a mean of 5.7 vertebrate-infecting viral genera in saliva samples (range 0-12) and 2.3 vertebrate-infecting viral genera in fecal samples (range 0-5). There was a clear separation between viral communities in saliva and fecal samples, with the exception of the CUS8 saliva sample which grouped with fecal samples (Figure D3).
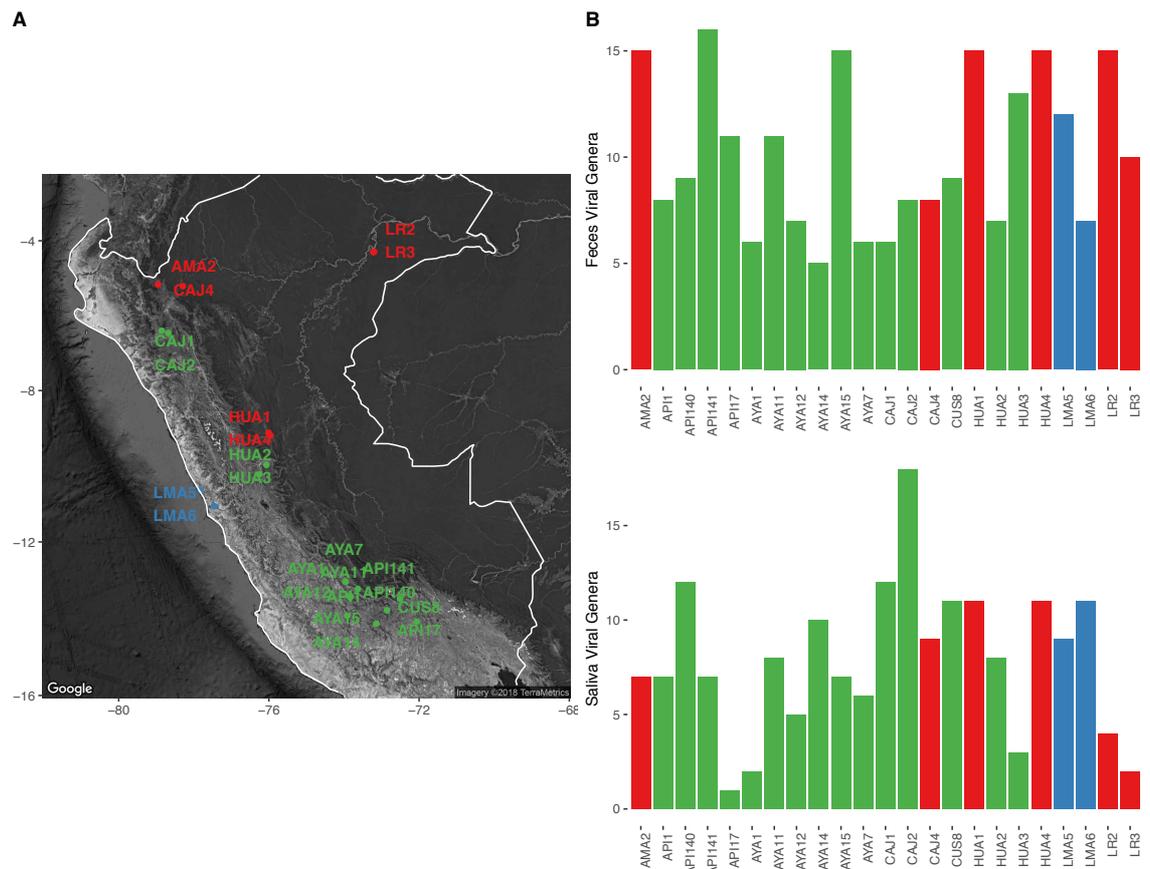
Figure 4 1 Vampire bat colony sampling and viral richness summary.
(A) Vampire bat colonies in Peru where host genetic and metagenomic samples were taken and (B) summary of viral genera richness in fecal and saliva samples at each colony. Individual colonies in (A) are represented as colored points and colony names located nearby, with colors corresponding to the different ecoregions within Peru (Figure D1). Peru country borders are outlined in white. Colors of bars in (B) also correspond to ecoregions (red, Amazon; green, Andes; blue, Coast).

4.4.2 Genetic structure summary

Microsatellite genotypes at 9 loci were generated for 624 vampire bat individuals from 24 colonies. F_{ST} comparisons between colonies indicated three main genetic clusters roughly corresponding to the Andes, Amazon, and Coast (Figure 4.2A), consistent with Streicker *et al.* (2016). Similar patterns were observed using a dataset including only 6 loci, although with less differentiation between the Amazon and Coast (Figure D4A). DAPC (Figure 4.2B; D4B), Structure (Figure 4.2C, D5A-B), and snapclust (Figure D4C, D5C) analyses performed using both 9 and 6 microsatellite loci confirmed that population structure at the country level consisted of three main groups, although in agreement with F_{ST} values, the Amazon and Coast were not as clearly differentiated. Colonies exhibited genetic isolation by distance, as genetic distance between sites increased with geographic distance (Figure D6; Mantel r=0.67; p=0.001).
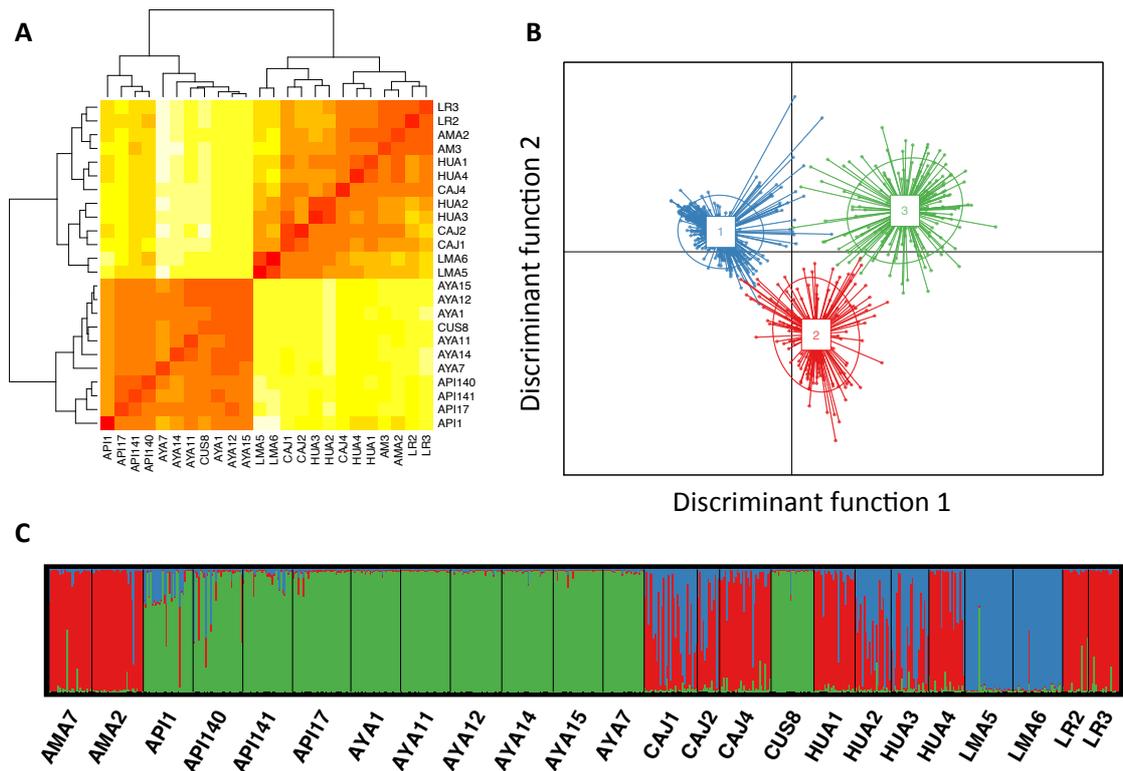
Figure 4 2 Population genetic structure of vampire bats across Peru.
Genetic structure plots are based on 9 microsatellite loci and show (A) heatmap of F_{ST} values between colonies (B) scatterplot based on k-means clustering in DAPC and (C) Bayesian clustering in Structure. In panel A darker shades of red correspond to lower F_{ST}, or higher connectivity, while lighter yellow and white correspond to higher F_{ST}, or lower connectivity. Colors in panels B and C roughly correspond different ecoregions within Peru as shown in Figure 4.1 (red, Amazon; green, Andes; blue, Coast).

4.4.3 Description of viral richness and community composition across ecoregions

Neither saliva viral richness (ANOVA; X^2=1.3; p= 0.52) nor TD (Kruskal-Wallis; X^2=2.69; p=0.26) varied across ecoregions (Figure 4.3), and the same was true for saliva vertebrate-infecting viral richness (ANOVA; X^2=1.57; p=0.46) and TD (Kruskal-Wallis; X^2=1.38; p=0.5) (Figure D7). In contrast, ecoregion had a significant effect on fecal viral richness (ANOVA; X^2=6.07; p=0.05), with a Tukey post-hoc test revealing this was due to significantly higher richness in the Amazon compared to the Andes (p=0.03). Fecal viral TD showed similar results, though only trending towards significance both overall (ANOVA; X^2=5.09; p=0.08) and in the Amazon-Andes comparison (p=0.06) (Figure 4.3). There were slightly different patterns for fecal vertebrate-infecting viruses, with a non-significant effect of ecoregion on richness (ANOVA; X^2=4.59; p=0.1) but a significant effect on TD (ANOVA; X^2=13.78; p=0.001) due to significant differences between both Amazon-Andes (p=0.005) and Coast-Andes (p=0.03) (Figure D7).
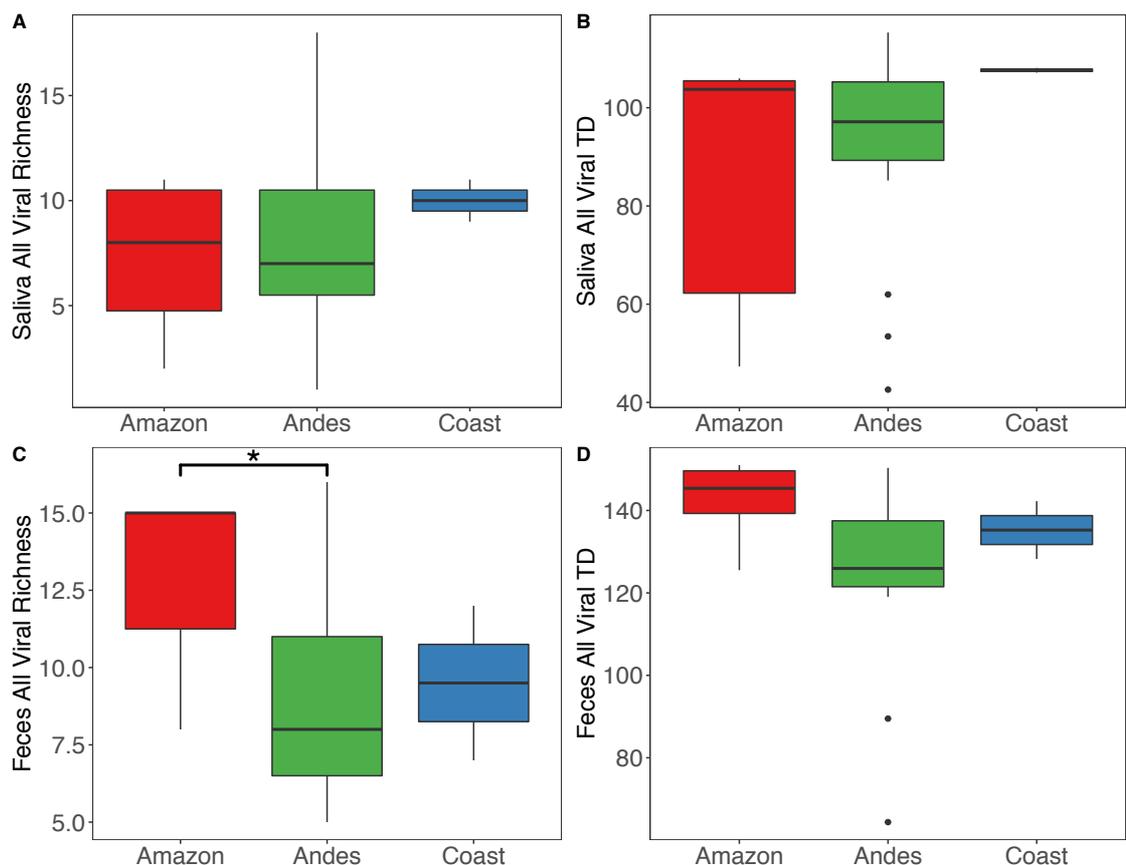
Figure 4 3 Viral richness and TD compared across ecoregions in vampire bats.
Plots show comparisons across ecoregions in saliva (A-B) and feces (C-D). In boxplots, bold line shows the median, and upper and lower hinges show the first and third quartiles. Whiskers extend from the hinge to 1.5 x the inter-quartile range. Stars indicate significance level of post-hoc Tukey pairwise comparisons, where * indicates p<0.05. Colors correspond to different ecoregions within Peru (red, Amazon; green, Andes; blue, Coast).

For both datasets (all and vertebrate-infecting viruses) and sample types (saliva and feces), most sites showed high normalized TD compared to normalized richness (Figure 4.4), suggesting that communities tend to be made up of diverse groups of viruses rather than closely related viruses, and highlighting the importance of taking relatedness into account in viral community analyses. However, there was variation within this pattern, in that sites with similar richness could exhibit either very high or very low TD.
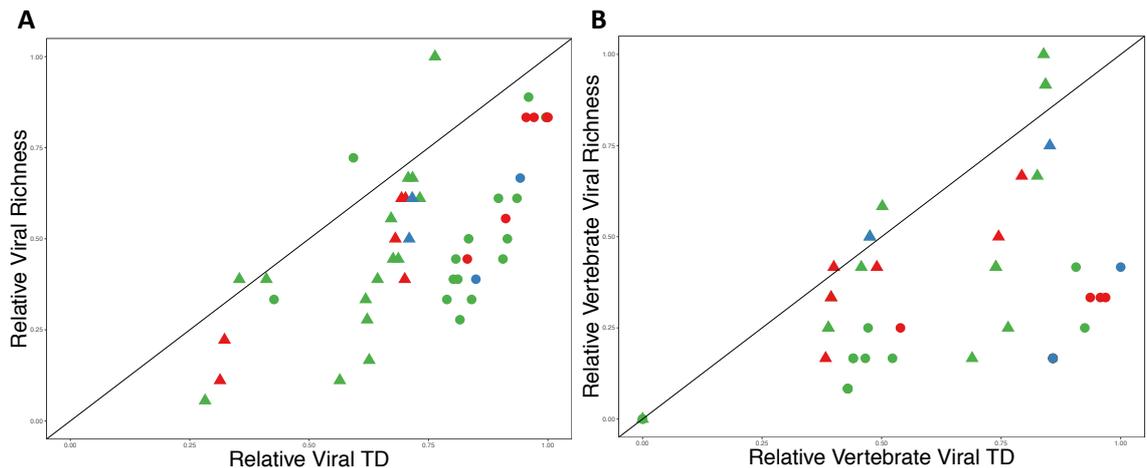
Figure 4 4 Viral communities in vampire bats show high diversity relative to richness. Comparisons of richness and TD are shown for datasets of all viruses (A) and vertebrate-infecting viruses (B). Viral richness and TD are normalized for comparison by dividing each value by the maximum value for each measure. Colors correspond different ecoregions within Peru (red, Amazon; green, Andes; blue, Coast); triangles represent saliva samples and circles represent fecal samples. The diagonal line represents equality between richness and TD; points below the diagonal line represent high diversity relative to richness and points above the diagonal line represent high richness relative to diversity.

PCoA of viral communities did not separate saliva samples by location in either the all virus or vertebrate-infecting virus dataset (Figure 4.5 A-B) and ecoregion explained little variation in either all saliva viruses (PERMANOVA; $F_{2,22}$ = 0.91; p= 0.58) and vertebrate-infecting viruses ($F_{2,21}$ = 1.18; p= 0.3) (Table D4). In contrast, fecal samples from the Amazon separated from those from the Coast and Andes along the second axis in the all virus dataset and along the first axis in the vertebrate-infecting virus dataset (Figure 4.5 C-D) with ecoregion explaining 16% of variation in all viruses (PERMANOVA; $F_{2,22}$ = 1.9; p= 0.005) and 14.6% of variation in vertebrate-infecting viruses ($F_{2,20}$ = 1.5; p= 0.09) (Table D4).
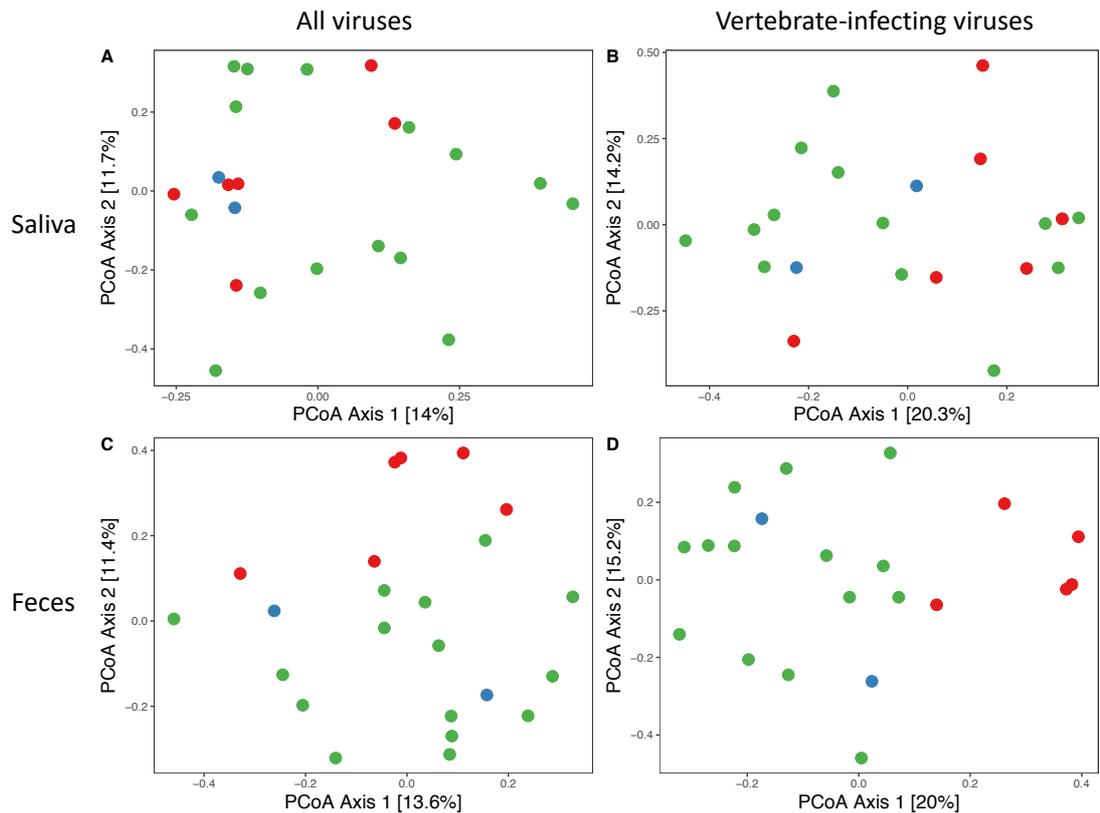
Figure 4 5 Principal coordinate analysis of viral communities in vampire bats.
Plots show ordinations of saliva all (A) and vertebrate-infecting (B) viruses and fecal all (C)
and vertebrate-infecting (D) viruses. Colors correspond to different ecoregions within Peru
(red, Amazon; green, Andes; blue, Coast).

4.4.4 Viral diversity compared with geographic and genetic distance

Differences in saliva viral richness and TD were positively correlated with geographic distance between sites (Figure 4.6A, 4.7A), but not with host genetic distance (Figure 4.6B, 4.7B) regardless of number of loci used, measure of genetic differentiation, or data subset (Figure D8, D9, D10, D11). Fecal viral richness and TD were not related to either geographic (Figure 4.6C, 4.7C) or genetic distance (Figure 4.6D, 4.7D, D8, D9, D10, D11), with the exception of a significantly positive correlation between TD and F_{ST} calculated using 6 loci (Figure D10). There was no relationship between saliva and fecal viral diversity within the same site, for either viral richness or TD (Figure D12).
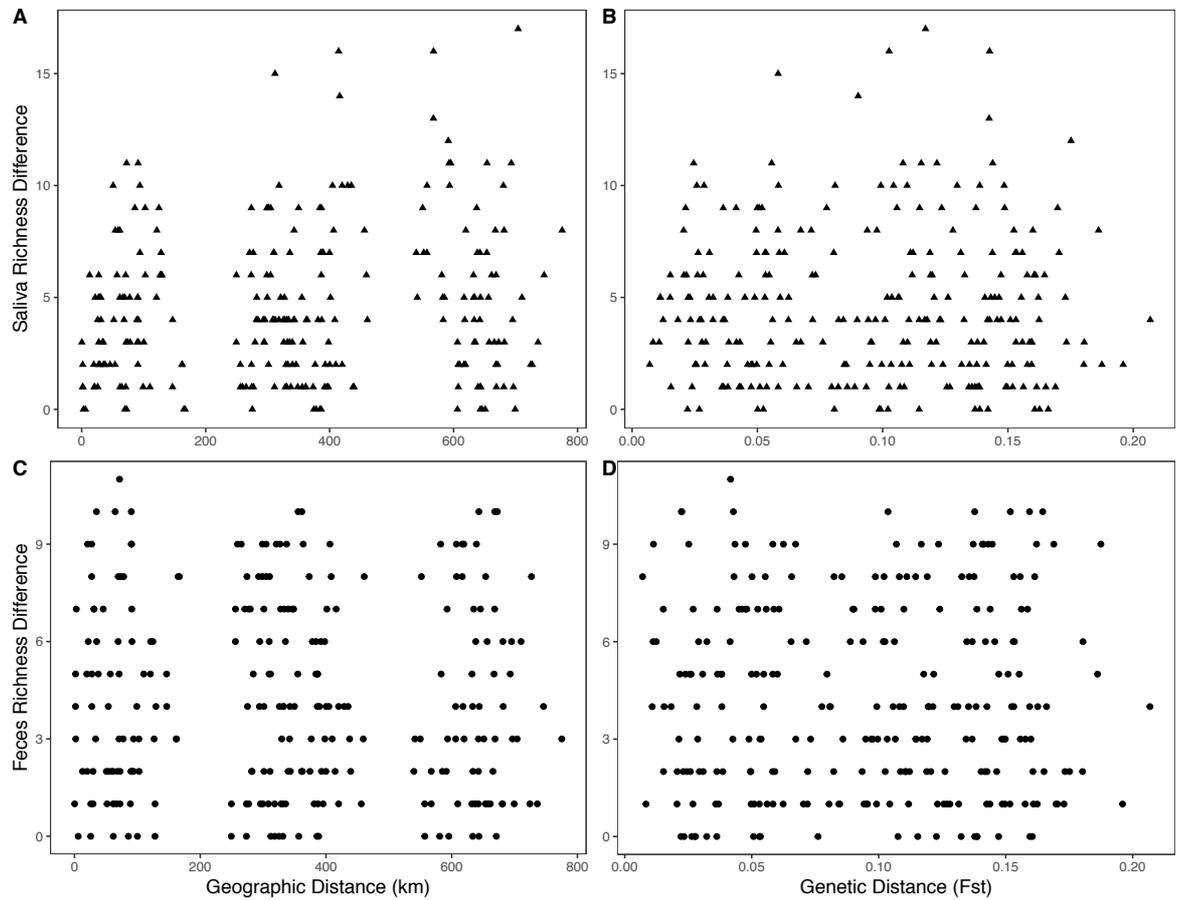
Figure 4 6 Correlations between viral richness and geographic or genetic distance.
Plots show saliva viral richness differences compared with colony geographic distance (km)
(Panel A; Mantel r=0.14; p=0.09) and genetic distance F_{ST} calculated using 9 microsatellite
loci (Panel B; Mantel r=-0.007; p=0.51), and correlations of fecal viral richness differences
with colony geographic distance (km) (Panel C; Mantel r=-0.04; p=0.67) and genetic distance
F_{ST} calculated using 9 microsatellite loci (Panel D; Mantel r=-0.004; p=0.44).
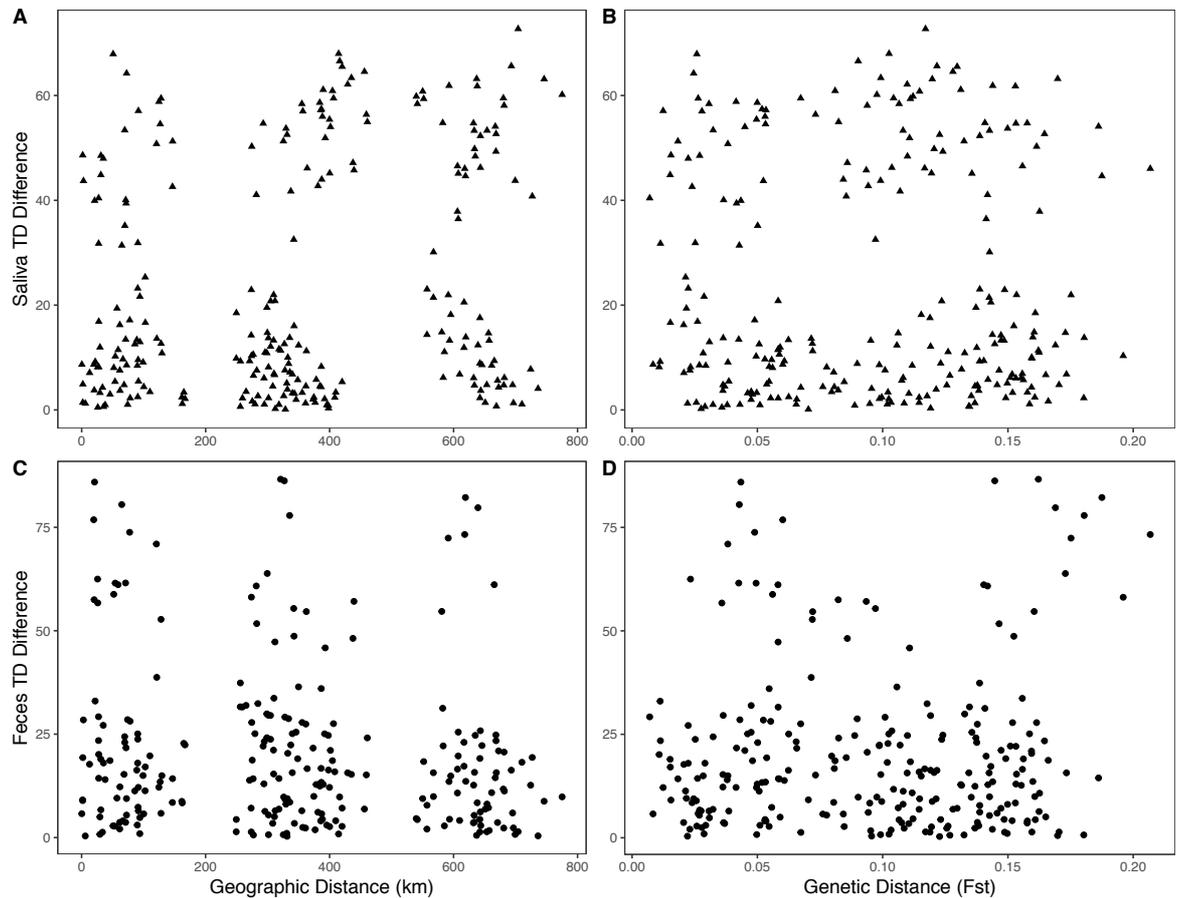
Figure 4 7 Correlations between viral TD and geographic or genetic distances.
Plots show relationships between saliva viral TD differences with colony geographic
distance (km) (Panel A; Mantel r= 0.22; p= 0.03) and genetic distance F_{ST} calculated using 9
microsatellite loci (Panel B; Mantel r= -0.02; p= 0.56). Fecal viral TD differences are also
compared with colony geographic distance (km) (Panel C; Mantel r= -0.15; p= 0.99) and
genetic distance F_{ST} calculated using 9 microsatellite loci (Panel D; Mantel r= 0.07; p=0.16).

Saliva virus community composition was not correlated with geographic or genetic
distance (Figure 4.8 A-B), while both geographic and genetic distance were significantly
correlated with fecal virus community composition (Figure 4.8 C-D). However, saliva
community composition often reached the maximum value of 1 (complete dissimilarity of
communities) making it difficult to detect patterns. Results were consistent when analyses
were repeated with F_{ST} based on 6 loci or with alternative differentiation measures (Figure
D13). There was not a significant correlation with distance in fecal and saliva vertebrate-
infecting viral community distances, although for both sample types Jaccard distances also
often reached the maximum value of 1 (Figure D14).
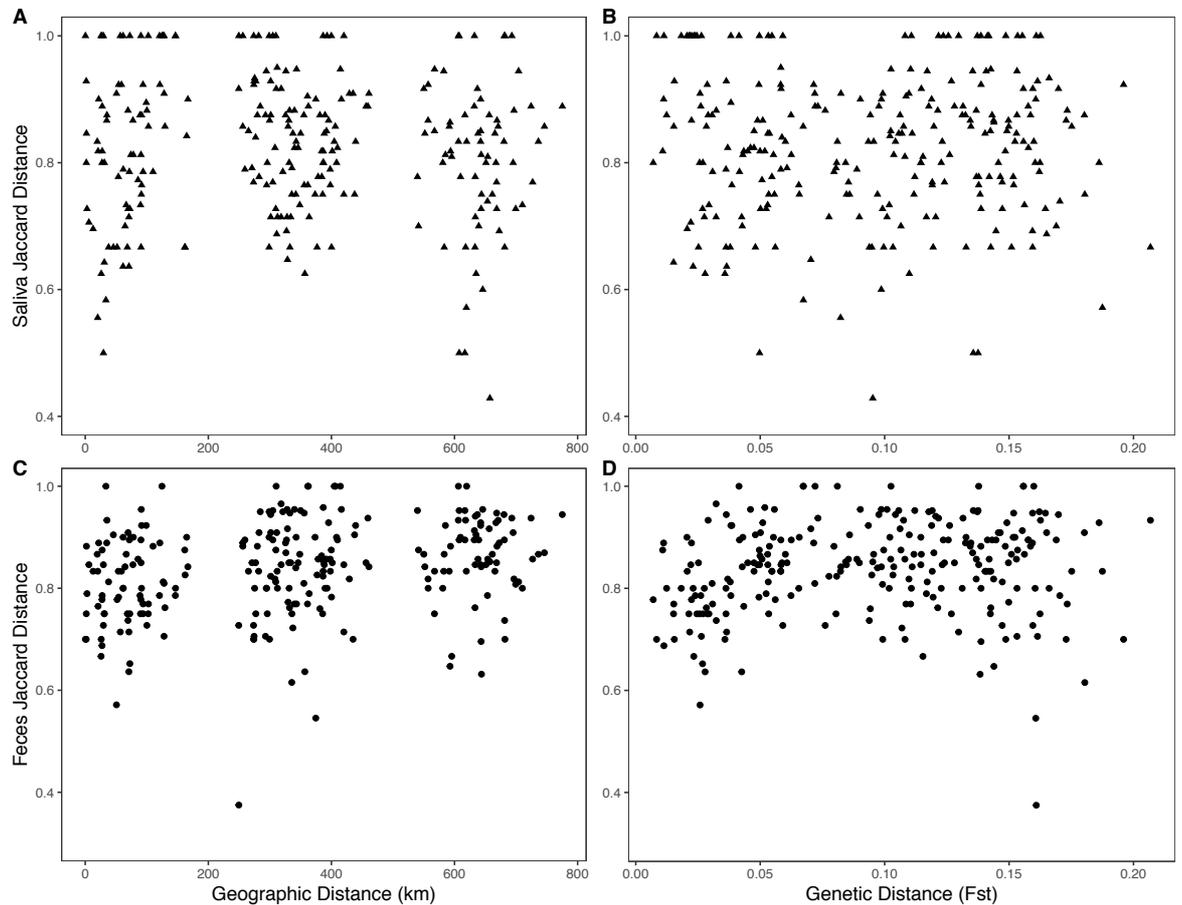
Figure 4 8 Correlations between viral community distance and geographic or genetic distance.
Plots show saliva virus community Jaccard distances compared with colony geographic distance (km) (Panel A; Mantel r=-0.05; p=0.67) and genetic distance F_{ST} calculated using 9 microsatellite loci (Panel B; Mantel r=-0.007; p=0.53), and fecal virus community Jaccard distances correlated with colony geographic distance (km) (Panel C; Mantel r=0.25; p=0.003) and genetic distance F_{ST} calculated using 9 microsatellite loci (Panel D; Mantel r=0.13; p=0.03).

4.4.5 Ecological drivers of viral richness and community composition

Colony-level demographic and local environmental variables were examined as potential drivers of viral richness and community composition. For saliva samples, overall viral diversity was negatively correlated with longitude for both richness and TD and positively correlated with raw sequencing reads for richness only (Figure 4.9; Table D6; Table D7). The model averaged results were consistent for richness and TD (Figure 4.9A-B), and final variables with effect sizes differing significantly from 0 all remained significant in the final models (Table D5). Longitude retained the significant relationship with viral richness and TD in univariate models following p-value correction, but raw sequencing reads became non-significant (Figure 4.9C). For vertebrate-infecting saliva viruses, longitude was the only variable significant following model averaging, and remained significant in the univariate model for richness, but not TD (Figure D15; Table D8; Table D9; Table D5).
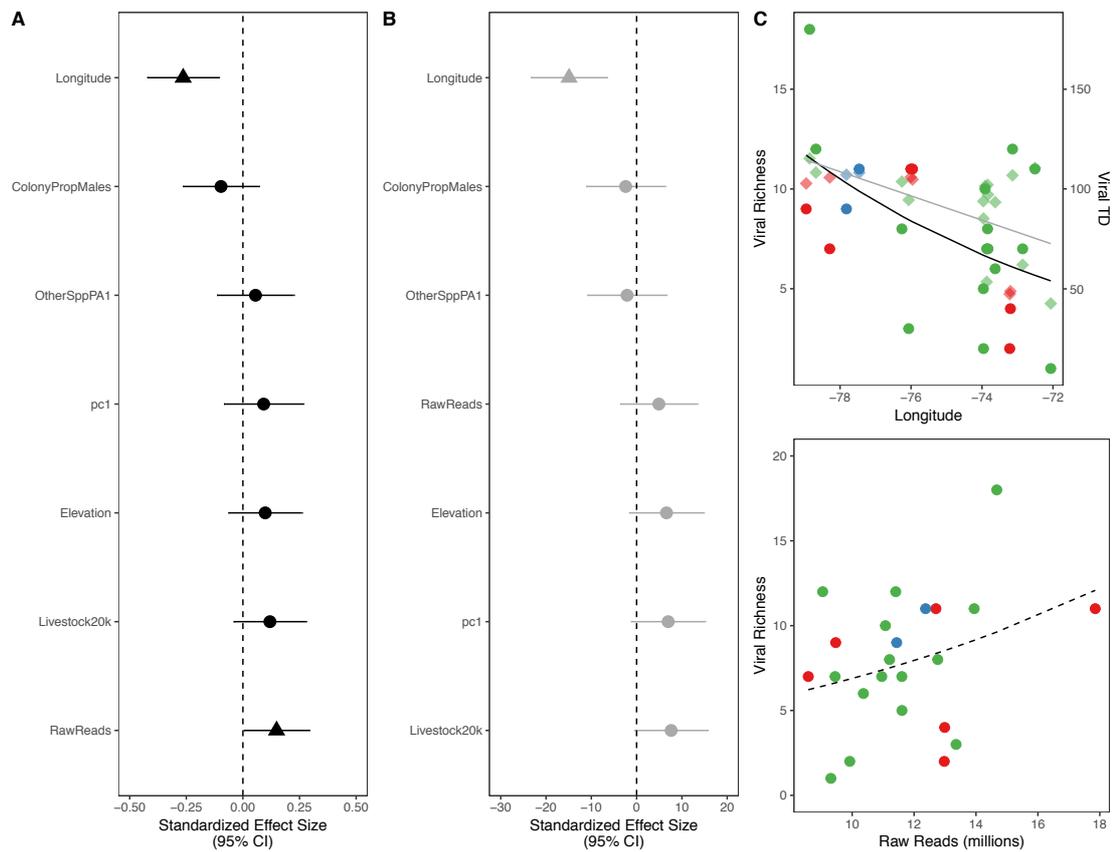
Figure 4 9 Ecological correlates of viral richness in vampire bat saliva samples.
Model averaged relationships of demographic and environmental factors correlated with (A)
richness and (B) TD and (C) univariate correlations of significant factors. Viral richness
model results shown in black and TD results are shown in gray. In panels (A) and (B) the
model averaged effect sizes are shown for each factor across the 95% confidence set of
GLMs with 95% confidence intervals. Factors that remained significant in the final model are
shown as triangles. The vertical dashed line shows an effect size of zero, such that any
confidence intervals overlapping the dashed line indicate a non-significant effect of the
factor in model averaged results. In panel (C) richness (left) and TD (right) are plotted
together for each variable that was significant according to model averaging. Solid lines
show GLM predictions for univariate relationships that remained significant following
correction for multiple testing, while dashed lines are univariate relationships that were no
longer significant after correction. Points are colored according to ecoregions; solid points
are values for richness and translucent diamonds are values for TD. Richness represents
the number of genera detected while the scale of TD cannot be directly related to number of
taxa.

For fecal samples, proportion of adults, elevation, and local climate variables were
negatively correlated with richness and TD, while livestock density was negatively
correlated with richness only (Figure 4.10; Table D10; Table D11). Model averaged results
were largely consistent between richness and TD (Figure 4.10A-B). In final models
including all significant variables, livestock density and the proportion of adults remained
significant for richness while all effects became non-significant for TD (Table D5). All
variables with significant effect sizes from model averaging remained significant in
univariate models following p-value correction, aside from the climate variable PC1 for
TD (Figure 4.10C). For vertebrate-infecting viruses, the proportion of adults and elevation
of the colony were negatively correlated with viral richness and TD in fecal samples

(Figure D16; Table D12; Table D13), but fewer effects remained significant in the final
models or in univariate relationships following p-value correction (Table D5).
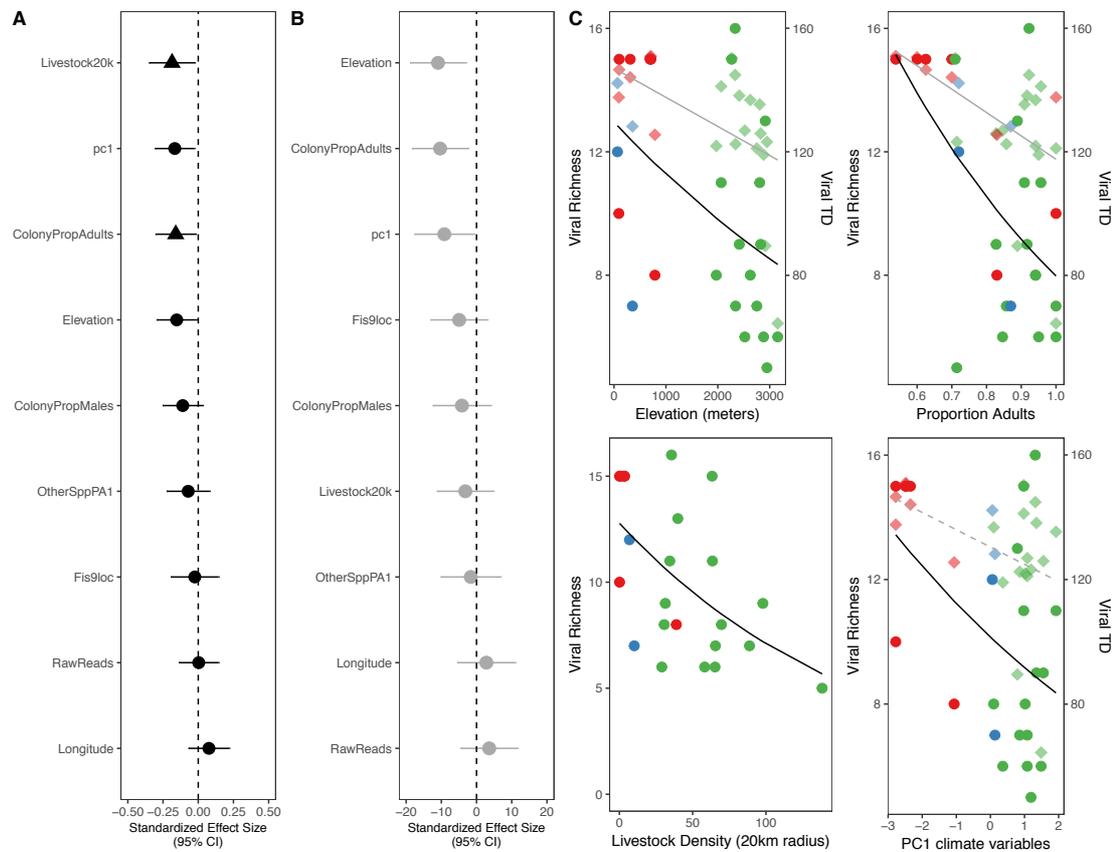
Figure 4 10 Ecological correlates of viral richness in vampire bat fecal samples.
Model averaged relationships of demographic and environmental factors correlated with (A)
richness and (B) TD and (C) univariate correlations of significant factors. Viral richness
model results shown in black and TD results are shown in gray. In panels (A) and (B) the
model averaged effect sizes are shown for each factor across the 95% confidence set of
GLMs with 95% confidence intervals. Factors that remained significant in the final model are
shown as triangles. The vertical dashed line shows an effect size of zero, such that any
confidence intervals overlapping the dashed line indicate a non-significant effect of the
factor in model averaged results. In panel (C) richness (left) and TD (right) are plotted
together for each variable that was significant according to model averaging. Solid lines
show GLM predictions for univariate relationships that remained significant following
correction for multiple testing, while dashed lines are univariate relationships that were no
longer significant after correction. Points are colored according to ecoregions; solid points
are values for richness and translucent diamonds are values for TD. Richness represents
the number of genera detected while the scale of TD cannot be directly related to number of
taxa.

As colony size exhibited a significant univariate effect on fecal vertebrate-infecting viral
TD (conservative N_c dataset t=2.02, p=0.04; less conservative N_c dataset t=1.73, p=0.08),
the same model averaging protocol as above was performed on the less conservative N_c
dataset, with the caveat that this dataset contains fewer observations (N=18 bat colonies)
and one additional variable compared to the dataset including all colonies. N_c had a
significant and positive effect both alone and in combination with other variables (Figure

D17), suggesting that larger colonies maintain a higher diversity of fecal vertebrate-infecting viruses.

Local demographic and environmental factors that were significantly correlated with viral richness for each dataset were also tested for an effect on viral community composition (Table 4.5). No variables were significantly associated with the community composition of all viruses in saliva, but longitude was significantly correlated with the community composition of vertebrate-infecting saliva viruses. For all fecal viruses, livestock and climate were consistently associated with differences in viral community composition while the proportion of adults and elevation were variably significant according to the PERMANOVA and GLM analyses; no variables were significant for vertebrate-infecting fecal viruses.

Table 4 5 Multivariate analyses of vampire bat viral community composition.
PERMANOVA and GLM analyses were used to test whether variables found to impact viral richness in vampire bats also affect community composition in the same dataset. Analyses were performed separately for different sample types (feces and saliva) and virus datasets (all viruses and vertebrate-infecting viruses).

		PERMANOVA				GLM	
		d.f.	F	R^2	P-value†	LR	P-value†
All Virus Community Saliva	Longitude	1,22	1.61	0.07	0.14	110.1	0.08
	Raw Reads	1,22	1.38	0.06	0.18	77.41	0.18
Vertebrate-infecting Virus Community Saliva	Longitude	1,21	2.19	0.1	**0.03**	90.81	**0.005**
All Virus Community Feces	Livestock	1,22	2.193	0.095	**0.01**	142.6	**0.03**
	Proportion Adults	1,22	1.45	0.06	0.13	135.4	**0.05**
	Climate	1,22	2.94	0.12	**0.002**	171.1	**0.03**
	Elevation	1,22	2.36	0.10	**0.007**	143.1	0.06
Vertebrate-infecting Virus Community Feces	Proportion Adults	1,20	1.25	0.06	0.28	39.01	0.07
	Elevation	1,20	1.67	0.08	0.24	49.1	0.07

†P-values for multiple analyses applied to the same dataset are adjusted using the Benjamini-Hochberg false-discovery rate method, with significant values shown in bold

4.5 Discussion

The determinants of viral diversity have largely been studied using multi-species comparative analyses of published records or snapshot surveillance focused on pre-selected viral groups in small numbers of host populations (Turmelle & Olival 2009; Luis *et al.* 2013; Maganga *et al.* 2014; Anthony *et al.* 2015; Olival *et al.* 2017). Here, by applying

unbiased shotgun metagenomic sequencing of saliva and fecal viruses across multiple populations of vampire bats in Peru, I found that ecological factors can influence viral communities within a single host species. I tested the hypotheses that highly connected vampire bat colonies share more similar viral communities, that there are broad differences in viral diversity between ecoregions in Peru, and that local demographic and environmental factors can influence colony-level viral diversity.

A new taxonomically-weighted measure of viral richness was developed which demonstrated that incorporating relatedness in some cases altered conclusions about the determinants of alpha and beta diversity, as has been shown in previous studies (Zhang *et al.* 2012; Huang *et al.* 2012). For example, raw reads were positively correlated with saliva viral richness (Figure 4.9) and livestock density was negatively correlated with fecal viral richness (Figure 4.10), but neither was correlated with TD. Yet in most cases the two measures showed broad agreement, for instance, the consistent negative correlation of longitude with saliva viral richness and TD (Figure 4.9, D15). Comparing normalized versions of the two measures showed that most samples had high diversity relative to richness and that the two diversity measures provided a different picture of the viral community (Figure 4.4). The communities with high richness relative to diversity were primarily vertebrate-infecting saliva viruses; these saliva communities often contained closely related viral genera (e.g. genera within the families *Herpesviridae* and *Papillomaviridae*) which could represent infections with multiple closely related genera. Alternatively, it is possible that a single genus was present, but was mis-classified because contigs aligned to poorly characterized genomic regions. The TD measure should counteract misassignments to some degree, as lower TD values were assigned to communities with many genera in the same family.

Saliva viral richness and community composition did not differ between ecoregions, and there was no effect of any genetic distance measure on richness and TD. The absence of effects of ecoregion and host genetic structure were not driven by homogeneity of taxa across saliva communities, as the Jaccard values for saliva communities were relatively high (Figure 4.8), but rather show that saliva communities are not predictable based on these traits. However, there was a positive correlation between geographic distance and both saliva viral richness and TD (Figure 4.6, 4.7), indicating that closer colonies have similar saliva viral richness, but that this effect might be driven more by location than by host movement, considering that genetic distance was not found to have an effect in this context. The potential importance of small-scale geographic differences, but not host

genetic structure, on microbial communities has also been observed in a gut microbiome study of wild mice (Linnenbrink *et al.* 2013).

Longitude was the only variable that was consistently correlated with saliva viral richness and community composition (Figure 4.9, D15; Table 4.5). Longitude corresponds roughly to a northwest-southeast gradient of viral diversity across Peru, with sites in the northwest (Cajamarca and Amazonas Departments) having the highest saliva viral diversity. This region does not neatly correspond to an ecoregion, which exhibited no detectable effect. The observation of high saliva richness and community composition in this region adds to previous observations suggesting its uniqueness as a potential bat corridor between the Coast and Andes/Amazon based on host gene flow (Streicker *et al.* 2016), the presence of unique mitochondrial COI haplotypes (Bohmann *et al.* 2018), and a high frequency of vampire bat rabies outbreaks in humans (Stoner-Duncan *et al.* 2014). Taken together, the effect of geographic distance on viral richness and the effect of longitude on richness and community composition indicate the importance of location in saliva communities, suggesting the northwest region of Peru as a hotspot of saliva viral diversity. However, none of the demographic or environmental variables included in models was able to provide a better explanation of this pattern.

Fecal viral communities differed between ecoregions, with the Amazon exhibiting higher richness and distinct community composition (Figure 4.3, 4.5). Broad-scale climate effects such as this are typically tested using latitude in studies of parasite richness (Guernier *et al.* 2004; Nunn *et al.* 2005; Bordes *et al.* 2011), but the scale of this study was such that ecoregion might be a better broad-scale climate proxy than latitude. Differences in fecal community composition were related to geographic and genetic distance between colonies, with closer colonies having more similar taxa, but distance correlations disappeared in vertebrate-infecting viral communities as Jaccard distance often reached its maximum value (Figure D14). This suggests that colony-level vertebrate-infecting fecal viral communities are highly distinct over the spatial scale of this study and that previous studies characterizing fecal viral diversity based on single site sampling may not capture the full picture of diversity within a host species. The importance of host dispersal in viral community structure has been observed in other systems (Anthony *et al.* 2015), and could play a role in shaping fecal viral communities given the link with host genetic structure. However, fecal viral richness and community composition were associated with environmental and demographic variables, particularly in the all-virus dataset, suggesting that local factors, as well as distance, might play an important role.

The proportion of adults in bat colonies had a consistent negative effect on the richness of viruses in bat feces such that colonies with a higher proportion of juveniles had a larger and more diverse viral pool (Figure 4.10, D16). The importance of juveniles in viral dynamics of bats has been related to seasonal birth pulses, which bring immunologically naïve individuals into the colony, facilitating virus transmission (George *et al.* 2011; Amman *et al.* 2012). Juvenile bats have been found to drive infection dynamics of both viruses and bacteria within a colony (Dietrich *et al.* 2015). The prevalence of coronavirus infection was found to be increased in juvenile bats (Anthony *et al.* 2017b), and a metagenomic study detected a novel coronavirus only in pools of samples from juvenile bats (Donaldson *et al.* 2010); indeed the single-colony metagenomic pool in which a full coronavirus genome was detected (HUA4; Chapter 3) contained a lower than average proportion of adults in the sequencing pool and within the colony (data not shown). In vampire bats, subadults and juveniles are more frequently exposed to rabies virus and have higher infection rates by Mycoplasma bacteria (Streicker *et al.* 2012b; Volokhov *et al.* 2017). In summary, this observation fits into a broad pattern indicating the importance of juvenile bats in driving viral dynamics, and suggests that future efforts targeting viral discovery or control should emphasize these individuals (Anthony *et al.* 2017b).

These results show the first evidence of reduced viral richness in higher elevation populations of a single host species (Figure 4.10, D16), while previous elevation effects of viral diversity in bats have been confounded with host species composition (Afelt *et al.* 2018). The negative elevation effect could be explained by the declining diversity of prey, alternative host species or vectors in high elevations (Lomolino 2001), or by environmental factors correlated with elevation that influence the survival of environmentally transmitted viruses. In humans, local climate variables such as temperature and precipitation range which are correlated with elevation have been associated with reduced viral richness on a global scale (Guernier *et al.* 2004). However, that elevation itself was more strongly correlated with viral richness than these environmental variables suggests that other factors that co-vary with elevation (i.e., host community composition) may be more important. Indeed, it has previously been hypothesized that cross-species transmission may be an important component of bat viral diversity (Luis *et al.* 2013; 2015). While this study did not find effects of other bat species on viral richness, it was only possible to measure the presence or absence of bats within the same roosting structures, which may not fully reflect the diversity of bats in the local environment. Finally, given that macrofauna are also expected to decline at higher elevations, it is possible that the observed reduction in viral diversity could be driven by the lower diversity of prey available to vampire bats.

In support of the importance of prey diversity, bat colonies located in areas of high livestock density had lower viral richness in fecal samples. If prey constitute an important source of viruses for vampire bats, it is possible that the lower diversity of native prey available in areas of high livestock density (Voigt & Kelm 2006; Bobrowiec *et al.* 2015) reduced the diversity of viruses that bats were exposed to. Alternatively, dietary resource provisioning from livestock could enhance the bat immune system leading to lower viral diversity, as was hypothesized for bacterial infections (Becker *et al.* 2018). However, there was only a correlation of livestock density with total viral richness, without any indication of a relationship with vertebrate-infecting viruses, so this effect might be driven by transient environmental viruses rather than viruses that are actively infecting bats. More broadly, all the environmental correlations suggest an association between vampire bat fecal viral diversity and the context from which viruses are acquired, whether that be diversity of prey, variables that were not measured in this study such as the diversity of other bat species (only presence/absence was reliably recorded) or arthropods, or environmental conditions such as temperature and humidity that facilitate viral persistence in the environment and therefore enhance transmission.

The analysis of colony sizes suggested that larger colonies (particularly those at low elevations with high proportions of juveniles) have higher richness of fecal viruses (Figure D17). The importance of N_c may be explained by a higher probability of encounter or enhanced persistence of novel viruses in large colonies. Larger colonies might encounter novel viruses more often through more individuals interacting with other bat species, prey and the environment. Viral persistence within a colony can be related to population size through the epidemiological concept of critical community size (Bartlett 1957; Lloyd-Smith *et al.* 2009), which will be specific to each pathogen species, but larger population sizes are likely to meet this minimum threshold across a variety of pathogen species. A follow-up to this observation would be the analysis of a larger N_c dataset and potentially other measures of population size such as genetic effective population size (N_e) (Luikart *et al.* 2010; Palstra & Fraser 2012). N_e reflects a more complex view of population size than N_c, as it is affected by census population size as well as sex ratio, variance in reproductive success, mating system, and mode of inheritance (Wang 2005). A relationship between viral richness and N_c could indicate a more contemporary population size effect on diversity, while an effect of N_e on viral richness might reflect the impact of population size changes over a longer time scale. However N_c and N_e are often correlated such that in practice it might be difficult to disentangle the effects of one or the other on viral diversity.

One important caveat to this study is the spatial and taxonomic resolution of the dataset; studies of viral communities over a smaller spatial scale with more precise taxonomic assignment (i.e. to species or strain level) might be more useful in answering epidemiological questions, such as predicting viral spread based on host movement. However, the results presented here provide novel insight into the ecological factors that structure genus-level viral communities. While this study aimed to assess country-level patterns in viral communities, the number of colonies sampled was small relative to the number of potentially important variables, limiting the statistical approaches that could be taken. Therefore model averaging, univariate and multivariate tests were combined to establish multiple lines of evidence to support significant variables. Finally, all viruses were considered as one community, but interesting patterns might be revealed in separating communities by viral traits, such as DNA/RNA viruses or bacteriophages/vertebrate-infecting viruses. Finally, viruses might be transient taxa acquired from the environment and not actively infecting bats, which is likely to be particularly true of the all virus dataset, as there were a variety of genera detected that are typically associated with plant and insect infecting viruses (Table C4). Even within the vertebrate-infecting virus dataset, detected viruses could originate from vertebrate prey and are not infecting bats. Thus, these results represent a colony-level viral fingerprint that might be a combination between bat-infecting viruses and transient taxa obtained from the local environment.

In summary, vampire bat viral communities are highly distinct over a country-wide scale and are inconsistently correlated with geographic distance and host movement, such that predicting viral communities may be difficult. Additionally, these findings confirm that previous studies of viral communities that have analyzed single individuals or a single environment are unlikely to have captured the full extent of viral diversity within a species. Saliva and fecal viral diversity were uncorrelated, and richness and community composition for the two sample types were associated with different demographic and environmental variables, implying that observations based on one sample type cannot be applied to another. Finally, both demographic and environmental factors influenced viral diversity, representing the first empirical test of patterns hypothesized from comparative analyses across bat species (Turmelle & Olival 2009; Luis *et al.* 2013; Webber *et al.* 2017). These results represent an important step in understanding natural viral communities, with the eventual goal of anticipating disease emergence (Anthony *et al.* 2015; Olival *et al.* 2017).

5 Discussion

5.1 Collective discussion of data chapters

The overarching aim of this PhD thesis was to advance our understanding of intraspecific variation in wildlife-associated viral communities and to develop an approach that could be applied widely in comparative studies. Specifically, I aimed to characterize viral communities in vampire bat colonies across Peru, examine spatial patterns in viral diversity, and test demographic and environmental correlates of diversity on a country-wide scale. These objectives were initially challenging due to the lack of standardized methods for comparative studies of wildlife-associated viral communities, so I developed a new metagenomic sequencing approach specifically for non-invasive samples. I characterized viral communities in fecal and saliva samples from vampire bats across Peru, identifying novel viral taxa and establishing that viral communities differ between body habitats. Finally I examined demographic and environmental correlates of viral richness and community composition, providing the first empirical insights into ecological factors associated with intraspecific viral diversity on a country-wide scale.

Despite advances in deep sequencing that have revolutionized our understanding of viral ecology, including the development of viral metagenomic laboratory protocols (Hall *et al.* 2014; Kleiner *et al.* 2015; Kohl *et al.* 2015; Conceição-Neto *et al.* 2015) and bioinformatic pipelines (Roux *et al.* 2011; Wommack *et al.* 2012; Rampelli *et al.* 2016), none have addressed the specific challenges associated with non-invasive samples from wildlife. To this end, I developed a field-laboratory-bioinformatic protocol for generating comparable viral community data, aiming to maximize viral reads while minimizing bias (Chapter 2). Pilot studies were performed to address several uncertainties relevant to ecological studies including which sample storage buffer to use in the field, how to maximize nucleic acid extracted from non-invasive samples, and whether depleting host and bacterial material improved detection of viruses. I aimed to make these protocols scalable for the larger numbers of samples associated with ecological and evolutionary studies. In the final protocol, samples were stored in RNALater and nucleic acid was extracted directly from non-invasive fecal and saliva swabs. rRNA depletion and DNAse treatment were included in the protocol to minimize host and bacterial nucleic acid while leaving a relatively unbiased representation of the viral community. A bioinformatic pipeline was developed specifically for analyzing viral communities of vampire bats, but which could be adapted for other host species or pathogen communities of interest.

Applying the final protocol to field-collected samples from vampire bats, I sequenced pooled samples from across Peru and statistically validated that viral communities were being thoroughly sampled using this approach. The depth of sequencing achieved did not allow the confident identification of viral taxa to species or strain level, while further increasing read depth would have generated longer contigs that could be more precisely assigned. However, characterizing viral taxa at the family or genus level allows insight into the structure and drivers of viral communities at higher taxonomic levels which are nonetheless informative. For example, host-associated bacterial communities exhibit differences between habitats within the host body when examined at levels as high as phylum (The Human Microbiome Project Consortium 2012; Linnenbrink *et al.* 2013) and between environmental habitats when examined at levels as high as order (Sullam *et al.* 2012). Thus it does not appear necessary to classify microbes to species or strain level in order to detect ecologically relevant differences in community composition. In summary, the newly developed shotgun metagenomic sequencing protocol generated comparable viral community data that could be used to test ecological hypotheses about drivers of community composition.

Using the metagenomic sequencing protocol, I characterized viral diversity in vampire bats across Peru (Chapter 3). Diverse taxa were detected from vertebrate and non-vertebrate infecting viral families, including some families previously described in vampire bats such as *Adenoviridae, Herpesviridae, Retroviridae* and *Papillomaviridae* (Wray *et al.* 2016; Salmier *et al.* 2017; Escalera-Zamudio *et al.* 2017). I also sequenced the full genome of a novel *Alphacoronavirus* and discovered viral families of potential zoonotic interest such as *Hepeviridae* and *Reoviridae,* which had not been previously associated with vampire bats. There were broad differences between fecal and saliva viral communities, showing the first evidence of body habitat compartmentalization in viral communities outside of humans (Wylie *et al.* 2014; Hannigan *et al.* 2015). There was also a surprising presence of plant- and insect-infecting viral taxa, particularly in fecal samples, which could be explained by vampire bats acquiring these viruses from the environment through diet or grooming (Bohmann *et al.* 2018).

Focusing on large contigs from vertebrate-infecting viral families, phylogenetic analyses were performed to assess novelty and relationships to previously characterized viral taxa. Full genomes of a novel HDV were detected from saliva samples at three localities, which was unexpected and exciting given that the virus had only previously been found in humans (Le Gal *et al.* 2006; Lempp & Urban 2017). Although follow-up work is required

to understand the mysterious presence of HDV in vampire bat saliva, based on the phylogenetic analysis it might well be another virus with which bats exhibit a deep evolutionary relationship (e.g. Drexler *et al.* 2012a; Quan *et al.* 2013). In contrast, all the other viral families investigated in depth had been previously described in bats, such as the CoVs and AdVs which fell within the known diversity of Neotropical bats (Drexler *et al.* 2010; Corman *et al.* 2013; Wray *et al.* 2016). Vampire bat RVH was closely related to human viruses, suggesting potential historical transmission between bats and humans, while vampire bat HEV clustered with other bat viruses from around the globe, exhibiting an apparently ancient relationship with bat hosts (Drexler *et al.* 2012b; Smith *et al.* 2014). The widespread distribution and geographic structure of AdVs, PicoVs and FVs suggested these taxa could be used to examine movement of bat hosts, as has been done previously in other host-virus systems (Biek *et al.* 2006; Antunes *et al.* 2008; Lee *et al.* 2012b). In summary, this chapter expanded our knowledge of viral diversity in an ecologically important bat host and suggested future avenues of research into individual viral taxa.

In Chapter 4, I investigated the ecological drivers of vampire bat viral diversity across Peru. Feces and saliva exhibited differences in viral richness and community composition, and were impacted by different broad and small-scale factors. Vertebrate-infecting viral communities were also highly dissimilar between sites, emphasizing that earlier studies characterizing viral communities based on only one individual or sampling location likely underestimate viral diversity within a species. Saliva viral diversity did not differ between ecoregions, and there was no effect of any genetic distance measure on richness and TD, indicating that saliva communities are not predictable based on these traits. However, longitude was significantly correlated with differences in saliva richness and community composition, and there was an effect of geographic distance on richness, with the highest diversity being found in the northwest of Peru. In contrast, ecoregion was associated with differences in fecal viral diversity, and both geographic and host genetic distance were positively correlated with differences in fecal viral community composition. Fecal viral diversity increased with the proportion of juveniles in a colony, suggesting that future efforts targeting viral discovery or control should emphasize these individuals (Anthony *et al.* 2017b). There was also an effect of environmental context on fecal viral diversity, including elevation, climate, and in some cases local livestock density. In summary, both demographic and environmental factors influenced vampire bat viral diversity in an empirical test of patterns hypothesized from comparative analyses across bat species (Turmelle & Olival 2009; Luis *et al.* 2013; Webber *et al.* 2017).

Taken together the results in the three data chapters represent a step forward in understanding intraspecific variation in wildlife-associated viral communities. The development of a method for standardized, comparable viral metagenomics will be applicable in diverse fields such as disease ecology, conservation, and other field-based studies of pathogens. The novel viruses characterized in vampire bats added to previous knowledge about viral taxa associated with this ecologically important bat species and suggested future avenues of research within specific viral taxa. Finally, both environmental and demographic factors appeared to influence viral richness and community composition within a single host species. The results complement ongoing studies of viral taxa at a global scale and contribute to our understanding of both the evolutionary and ecological forces that shape viral diversity.

5.2 Project limitations

Although these results advance our understanding of wildlife-associated viral communities, there are several important caveats to discuss. When collecting metagenomic samples from bats in the field, sampling site was often confounded with time of year (Table D1) such that it was not possible to assess seasonal differences in viral community composition, although seasonality is an important feature of other bat viruses (e.g. George *et al.* 2011). There were also differences between sites in the efficacy of the cold chain, for example, sites in Loreto were sampled earlier without ready access to dry ice. Although this could potentially lead to RNA degradation (Cardona *et al.* 2012), these sites exhibited relatively high viral richness. Different types of swabs were used for sampling feces (rayon-tipped) and saliva (cotton-tipped), such that variation in efficacy of these materials for binding viruses could potentially lead to differences in richness between the sample types. Although there were consistently fewer saliva reads compared to feces, this discrepancy disappeared for both family and genus level richness. Some sites were sampled in different years, which could mean that communities are not comparable, but we lack data on the stability of viral communities over time. Wildlife-associated bacterial communities display short-term temporal stability (e.g. Loudon *et al.* 2013; Bobbie *et al.* 2017) although they change predictably over stages of development within individuals (Costello *et al.* 2012; Christian *et al.* 2015; Prest *et al.* 2018). However, viral communities might be subject to different selective forces as compared to the primarily commensal bacteria that have been studied in this context. In summary, it is not clear whether variation due to field sampling logistics had a major impact on the viral communities detected but based on the available evidence it does not appear very problematic.

Although the metagenomic protocol aimed for minimal bias within the viral community, it was still necessary to deplete samples of DNA and rRNA to ensure that host and bacterial genomes did not swamp out viral nucleic acid. The DNAse treatment could have caused bias towards RNA viruses, but DNA virus reads were detected in all samples, as has been found in other viral metagenomic studies with a DNAse treatment step (Baker *et al.* 2013; Hall *et al.* 2014). A more intensive enrichment such as filtration or centrifugation could potentially have increased the number of viral reads, but these are known to be biased against certain taxa (Kleiner *et al.* 2015; Wood-Charlson *et al.* 2015; Conceição-Neto *et al.* 2015). While it is always preferable to avoid unnecessary PCR steps during library preparation (van Dijk *et al.* 2014), a re-amplification PCR was included due to the low RNA input from non-invasively collected samples. Any PCR bias introduced by this step would be consistent across samples such that viral communities should remain comparable. To avoid some of the pitfalls generally associated with shotgun viral metagenomics, the development of sequence capture-based approaches to viral discovery will provide an exciting option for future studies of viral communities, offering many more targeted viral reads with minimal bias (Briese *et al.* 2015; Wylie *et al.* 2015).

The inability to consider read counts as a measure of viral abundance limited the approaches to quantifying diversity that could be taken. The viral communities described here represent pools made up of multiple bats, with potential variation in viral load or shedding between individuals, such that there is not a straightforward association between read count and number of individuals infected or severity of infection. Ever for samples taken from single individuals, considering read counts as indicating relative abundance is generally problematic in shotgun metagenomic studies (Morgan *et al.* 2010). Methods for quantitative analyses have been developed for subsets of the viral community (Roux *et al.* 2016b) and bioinformatic methods exist to account for abundance in shotgun metagenomic data (Segata *et al.* 2012). However, the goal of this study was to include the entire viral community, while the quantitative lab methods are presently limited to dsDNA and ssDNA viruses, and bioinformatic methods that account for abundance in metagenomic data are similarly constrained by the reference sequences available within each program. Thus only presence-absence of viral taxa are considered in the analyses presented here. In bacterial community analyses, weighted diversity measures that take into account abundance provide a useful perspective that differs from that of richness alone (Lozupone *et al.* 2007), suggesting the value of eventually incorporating abundance into viral community studies.

It is essential to note that the viral communities detected in vampire bat samples are not necessarily viruses that are actively infecting the host, and that unsampled tissues within the same host may contain additional viral taxa. Although criteria for viral identification and taxonomy have been revised in the age of metagenomics (Simmonds *et al.* 2017), there remains a need to perform follow-up cellular studies to determine which viruses are actually associated closely with hosts and which are transient (Hayman 2016). Even in the vertebrate-infecting virus dataset, the viruses detected could represent prey viruses that are not actively infecting bats. A non-invasive metagenomic approach may also miss latent viruses that might be detectable by sequencing organs rather than fecal and saliva samples (Amman *et al.* 2012). I found evidence of compartmentalization of viral communities between body habitats (Chapter 3) and it was previously known that bat bacterial communities in urine are distinct from feces and saliva (Dietrich *et al.* 2017), suggesting that there might be still different viruses detected in vampire bat urine or other sample types such as blood and tissue. Indeed, literature-based analysis of viral discovery in bats found that there is a knowledge gap regarding which sample types should be targeted for detecting different viral families (Young & Olival 2016). In addition, metagenomic sequencing was performed on pooled samples, which limits the ability to evaluate viral communities at the individual bat level or to assess the prevalence of given viral taxa within the colony. The viral communities described here therefore represent a colony-level viral fingerprint, likely specific to feces and saliva, that could potentially represent viruses infecting vampire bats as well as transient taxa from prey and the environment.

Although the modeling approach in Chapter 4 aimed to incorporate many of the main factors that could affect viral diversity within a host species, the small number of colonies sampled meant that all potentially important variables could not be included. It was also not possible to generate empirical data for all predictors; for example, the effect of diet on richness could only be addressed indirectly by including local livestock density. However, in the future it might be possible to explicitly address the effect of diet by extracting prey sequences from shotgun metagenomic data (Srivathsan *et al.* 2016). Another potential diet-associated predictor would be stable isotopic niche, which was previously found to differ between ecoregions and reflect predominant prey type in vampire bats (Streicker & Allgeier 2016). Finally, it was necessary to combine local climate variables to reduce the number of predictors in models, so it would be interesting in future studies to tease apart the potentially differing effects of variables such as temperature and precipitation.

5.3 Future directions

There are many exciting future directions to consider in the field of comparative wildlife-associated viral metagenomics. The work presented in this thesis represents a detailed description of intraspecific viral diversity and ecological factors associated with variation in vampire bats across Peru. However, different factors may drive patterns of variation depending on the scale at which viral communities are examined. This work was conducted at a country-wide scale, while drivers of viral communities might differ at either a larger global scale or a finer landscape scale. The temporal stability of viral communities remains relatively unexplored at both the individual or population level. Although literature-based studies have identified factors associated with interspecific differences in viral diversity (Turmelle & Olival 2009; Luis *et al.* 2013), these hypotheses have not been empirically tested with viral community data. Future comparative studies will illuminate patterns and drivers of viral diversity across space, time, and host species.

Grouping all viruses as one community might obscure potentially important differences in the ways that subsets of viruses interact with hosts. Viral pathogens cause disease in bat hosts, while bacteriophages are presumably part of a healthy digestive microbiome in bats as in humans, and transient plant or insect viruses acquired through environmental interactions would likely be neutral with regard to the health of the bat host. In the future it would be interesting to separately analyze subgroups of viruses when evaluating the structure and drivers of viral community composition. For example, host movement is essential to the persistence of rabies virus, which exhibits wave-like movement across the landscape in vampire bats (Blackwood *et al.* 2013; Benavides *et al.* 2016), such that presence or absence of the virus at a given colony might be correlated with host genetic distance from an infected colony. In contrast, reads from bacteriophage families *Podoviridae, Myoviridae* and *Siphoviridae* were found at every colony sampled, suggesting the same bacteriophage taxa are ubiquitous over the scale of the study. Bacterial community composition is shaped by taxonomy and diet in bats (Phillips *et al.* 2012; Carrillo-Araujo *et al.* 2015), such that if microbiomes are conserved within a bat species, there might be a core gut "phageome" of their associated bacteriophages akin to that in humans (Manrique *et al.* 2016). Even excluding bacteriophages, some viral taxa might be so widespread among bats with minimally pathogenic effects such that they constitute a "core virome"; for example, the families *Herpesviridae* and *Papillomaviridae* were detected in most colonies, and can cause latent or chronic infections in humans (Virgin *et al.* 2009; Lecuit & Eloit 2013). Finally, transient plant or insect viruses occasionally

introduced to bats through diet might only be present in a localized area. For example, the proportion of plant-infecting viral reads was notably higher in samples pooled from the Loreto locality as compared to other areas (Figure 3.2). If this represents a virus infecting a plant species common in this locality, the virus might likely be detected in colonies that are geographically close to one another.

There are also biological traits that differ between viruses and are thought to affect the probability of emergence into other species. These include traits such as genome type, location of replication (nucleus or cytoplasm), and the presence or absence of an envelope (Pulliam 2008; Pulliam & Dushoff 2009). Although viral trait information might not be available for poorly characterized viral species from bats, it would be possible to extrapolate what is known about similar taxa in humans to examine patterns of distribution based on viral traits. Splitting the viral community into subsets based on host and viral traits that affect zoonotic potential might illuminate different distribution patterns and could have implications for predicting where viruses might be likelier to emerge from vampire bats into other species.

Given that the metagenomic laboratory approach preserves DNA and RNA from diverse origins, with appropriate bioinformatic modifications this approach could also be used to simultaneously characterize host diet, population structure, commensal bacteria and other host-associated parasites. Other metagenomic studies have described diverse pathogens (Chandler *et al.* 2015; Schneeberger *et al.* 2016) as well as diet and host population structure (Srivathsan *et al.* 2016). A targeted search approach has already been used to examine the metagenomic sequences described here for the bacterial pathogens Mycoplasma (Volokhov *et al.* 2017) and Bartonella (Becker *et al.*, under review), and it would also be possible to examine commensal bacteria, although the extent to which DNAse treatment and rRNA depletion bias the bacterial community would first have to be evaluated. Combining data from host-associated bacterial and viral communities provides a more holistic view of the microbial community (e.g. Hannigan *et al.* 2017), just as combining data on individual bacterial and viral taxa can illuminate important patterns in disease dynamics (Dietrich *et al.* 2015). There is also so-called "viral dark-matter" in the metagenomic data, which are viral sequences of unknown taxonomy and function (Brum *et al.* 2016). Reference-independent methods have been used to include dark matter in viral community analyses (Hannigan *et al.* 2015). As the bioinformatic approach taken here was reference-based these sequences were excluded, but there were contigs of significant length (>3,000bp) that were assigned as viral but did not match any characterized families.

What these sequences are remains a mystery for the moment, but given the ever-expanding understanding of the viral world it is likely that it will eventually be possible to incorporate them into analyses of viral diversity.

In addition to concern about viral emergence in humans and domestic animals, understanding more about viral communities in bats could lead to insights into their conservation. Emerging pathogens can threaten wildlife populations (Daszak 2000; Fisher *et al.* 2012), such as white-nose syndrome in North American bats (Blehert *et al.* 2009; Frick *et al.* 2010). Identifying typical members of the microbial community in apparently healthy populations of bats would make it easier to detect unusual or pathogenic microbes that might be introduced in the future. Establishing the extent of variation in viral communities within and between individuals will also allow us to better evaluate how communities are altered as a result of disease or changing environmental conditions. For example, understanding normal bacterial community variation within healthy humans has implications for designing studies to measure how disease alters normal communities (Zhou *et al.* 2013), and the same is true for examining alterations to the viral community due to environmental changes. Microbial communities can reflect the effects of habitat degradation on their hosts in various ways. Several studies have detected a decrease in microbial diversity following habitat fragmentation or captivity (Amato *et al.* 2013; Barelli *et al.* 2015) while another study found that disturbance broke apart an association between host genetic distance and microbial community diversity (Wegner *et al.* 2013). For viruses, individual taxa are known to respond to habitat fragmentation; intact habitat is associated with maintenance of rabies virus (de Thoisy *et al.* 2016), while fragmentation decreased richness across multiple pathogens (Gay *et al.* 2014). However, measuring the effects of changing environmental conditions on viral communities will first require establishing a baseline understanding of what taxa are present in each body habitat in healthy populations.

In addition to anticipating threats to bats by detecting emerging diseases in their viral communities, viral community data could be used to understand more about vampire bat biology. Host-associated microbes often track the movement patterns of their hosts (Falush *et al.* 2003b; Wirth *et al.* 2005) and primate microbiomes have been used to track individual movement between groups (Gomez *et al.* 2015), as microbial fingerprints are stable over time (Fierer *et al.* 2010). Although viral communities of individual bats were not generated here, the identification of widespread viral families (e.g. AdV, PicoV, FV) provides information about taxa that could be targeted by PCR in the future for

phylogenetic studies, in which multiple widespread viral taxa could provide independent perspectives on host movement patterns.

Understanding whether viral communities are structured according to consistent environmental or demographic factors might allow us to anticipate how disturbances such as land-use change or selective culling of certain demographic groups could affect patterns of viral diversity, ultimately leading to predictions of disease emergence (Anthony *et al.* 2015). The results described in this thesis represent an early step on the complex path leading from describing baseline viral diversity and its drivers to predicting where and how diseases might emerge from wildlife into humans and domestic animals (Plowright *et al.* 2014; Hassell *et al.* 2017; Carroll *et al.* 2018). Incorporating ecological and evolutionary concepts in microbial community studies has already led to many new insights and perspectives on host-associated bacterial communities (Costello *et al.* 2012; Christian *et al.* 2015; Davenport *et al.* 2017). In order to better understand the viral component of microbial communities, with its potential implications for global health, we need more standardized comparative studies of viral communities across species, landscapes and time points (Suzán *et al.* 2015; Anthony *et al.* 2015; Brierley *et al.* 2016; Olival *et al.* 2017). The work presented in this thesis provides both the laboratory methods and an analytical framework for conducting such future comparative studies of intraspecific viral diversity.

Appendices

Appendix A: Metagenomic sequencing pools

Table A 1 Detailed description of metagenomic sequencing pools.
This table summarizes the different sequenced pools discussed throughout the thesis including what experiments and analyses they were included in, which section these experiments can be found in, colony/colonies of origin of individual samples, and individual bat samples included in the pool. Sampling dates for individual bat samples included in single colony pools can be found in Table D1.

Pool ID	Experiment/ Analysis	Section	Colony 1	Colony 2	Individual bat samples in pool
H1	Preliminary Sequencing Run 1	2.3.4	AYA15	-	7871|7872|7873|7874|7875|7876| 7877|7878|7879|7880
H2	Preliminary Sequencing Run 1	2.3.4	API17	-	7954|7955|7956|7957|7958|7959| 7960|7961|7962|7963
H3	Preliminary Sequencing Run 1	2.3.4	AYA14	-	7767|7768|7769|7770|7771|7772| 7773|7775|7774|7776
H4	Preliminary Sequencing Run 1	2.3.4	CUS8	-	7989|7990|7991|7992|7993|7994| 7995|7996|7997|7998
SV1	Preliminary Sequencing Run 1	2.3.4	AYA15	-	7871|7872|7873|7874|7875|7876| 7877|7878|7879|7880
SV2	Preliminary Sequencing Run 1	2.3.4	API17	-	7954|7955|7956|7957|7958|7959| 7960|7961|7962|7963
SV3	Preliminary Sequencing Run 1	2.3.4	AYA14	-	7767|7768|7769|7770|7771|7772| 7773|7774|7775|7776
SG1	Preliminary Sequencing Run 1	2.3.4	AYA15	-	7871|7872|7873|7874|7875|7876| 7877|7878|7879|7880
H_EN_VTM	Preliminary Sequencing Run 2	2.3.5	LMA5	LMA6	9079|9080|9081|9082|3274|9105| 9106|9107|9108|9109
H_UN_VTM	Preliminary Sequencing Run 2	2.3.5	LMA5	LMA6	9079|9080|9081|9082|3274|9105| 9106|9107|9108|9109
SV_EN_VTM	Preliminary Sequencing Run 2	2.3.5	LMA5	LMA6	9079|9080|9081|9082|3274|9105| 9106|9107|9108|9109
SV_UN_VTM	Preliminary Sequencing Run 2	2.3.5	LMA5	LMA6	9079|9080|9081|9082|3274|9105| 9106|9107|9108|9109
H_EN_RL	Preliminary Sequencing Run 2	2.3.5	LMA5	LMA6	9079|9080|9081|9082|3274|9105| 9106|9107|9108|9109
H_UN_RL	Preliminary Sequencing Run 2	2.3.5	LMA5	LMA6	9079|9080|9081|9082|3274|9105| 9106|9107|9108|9109
SV_EN_RL	Preliminary Sequencing Run 2	2.3.5	LMA5	LMA6	9079|9080|9081|9082|3274|9105| 9106|9107|9108|9109
SV_UN_RL	Preliminary	2.3.5	LMA5	LMA6	9079|9080|9081|9082|3274|9105|

	Sequencing Run 2				9106|9107|9108|9109
AAC_H_F	Subsampling; Sample Type Comparison; Viral Phylogenetics	2.3.10; 3.3.4; 3.3.5	AYA7	AYA14	7064|8286|8287|8288|7047|8297| 8298|8299|8300|8301
AAC_H_SV	Subsampling; Sample Type Comparison; Viral Phylogenetics	2.3.10; 3.3.4; 3.3.5	AYA7	AYA14	7064|8286|8287|8288|7047|8297| 8298|8299|8300|8301
AAC_L_F	Subsampling; Sample Type Comparison; Viral Phylogenetics	2.3.10; 3.3.4; 3.3.5	API1	AYA11	5942|9260|9261|9262|9263|8201| 8202|8203|7082|8204
AAC_L_SV	Subsampling; Sample Type Comparison; Viral Phylogenetics	2.3.10; 3.3.4; 3.3.5	API1	AYA11	5942|9260|9261|9262|9263|8201| 8202|8203|7082|8204
AMA_L_ F_NR	rRNA Depletion	2.3.8	AMA2	AMA6	D203|D94|D95|D96|D97|D98|D99| D212|D213|SP3
AMA_L_F_R; AMA_L_F	rRNA Depletion; Sample Type Comparison; Viral Phylogenetics	2.3.8 3.3.4; 3.3.5	AMA2	AMA6	D203|D94|D95|D96|D97|D98|D99| D212|D213|SP3
AMA_L_SV	Subsampling; Sample Type Comparison; Viral Phylogenetics	2.3.10; 3.3.4; 3.3.5	AMA2	AMA4	D234|D235|D96|D236|D237|8410| 8411|8412|8413|8414
CAJ_L_F_NR	rRNA Depletion	2.3.8	CAJ4	-	D83|D84|D85|D86|D87|D88|D89| D90|D91|D92
CAJ_L_F_R; CAJ_L_F	rRNA Depletion; Sample Type Comparison; Viral Phylogenetics	2.3.8; 3.3.4; 3.3.5	CAJ4	-	D83|D84|D85|D86|D87|D88|D89| D90|D91|D92
CAJ_L_SV	Subsampling; Sample Type Comparison; Viral Phylogenetics	2.3.10; 3.3.4; 3.3.5	CAJ4	-	8361|8362|8363|8364|8365|8366| 8367|8368|8369|8370
CAJ_H_F_1; CAJ_H_F	DNAse Treatment; Subsampling; Sample Type Comparison; Viral Phylogenetics	2.3.9; 2.3.10; 3.3.4; 3.3.5	CAJ1	CAJ2	8080|8081|8082|8083|4156|8096| 8097|8098|8099|8100
CAJ_H_F_2	DNAse Treatment; Subsampling	2.3.9; 2.3.10	CAJ1	CAJ2	8080|8081|8082|8083|4156|8096| 8097|8098|8099|8100
CAJ_H_SV	Subsampling; Sample Type Comparison; Viral Phylogenetics	2.3.10; 3.3.4; 3.3.5	CAJ1	CAJ2	8080|8081|8082|8083|4156|8096| 8097|8098|8099|8100

HUA_H_F	Subsampling; Sample Type Comparison; Viral Phylogenetics	2.3.10; 3.3.4; 3.3.5	HUA1	HUA2	8332|8333|6750|8334|8335|8009| 8010|8011|8012|8013
HUA_H_SV	Subsampling; Sample Type Comparison; Viral Phylogenetics	2.3.10; 3.3.4; 3.3.5	HUA1	HUA2	8332|8333|6750|8334|8335|8009| 8010|8011|8012|8013
LMA_L_F_NR	rRNA Depletion	2.3.8	LMA5	LMA6	9083|2926|4557|9084|9085|9110| 7681|3859|9111|9112
LMA_L_F_R; LMA_L_F	rRNA Depletion; Sample Type Comparison; Viral Phylogenetics	2.3.8; 3.3.4; 3.3.5	LMA5	LMA6	9083|2926|4557|9084|9085|9110| 7681|3859|9111|9112
LMA_L_ SV_NR	rRNA Depletion	2.3.8	LMA5	LMA6	9083|2926|4557|9084|9085|9110| 7681|3859|9111|9112
LMA_L_SV_R LMA_L_SV	rRNA Depletion; Sample Type Comparison; Viral Phylogenetics	2.3.8; 3.3.4; 3.3.5	LMA5	LMA6	9083|2926|4557|9084|9085|9110| 7681|3859|9111|9112
LR_L_F_NR	rRNA Depletion	2.3.8	LR1	LR2	D79|D80|D81|D39|D57|D74|D75| D76|D77|D59
LR_L_F_R	rRNA Depletion; Sample Type Comparison; Viral Phylogenetics	2.3.8; 3.3.4; 3.3.5	LR1	LR2	D79|D80|D81|D39|D57|D74|D75| D76|D77|D59
LR_L_SV	Subsampling; Sample Type Comparison; Viral Phylogenetics	2.3.10; 3.3.4; 3.3.5	LR1	LR2	D251|D252|D253|D254|D255| D256|D257|D258|D259|D260
AMA7_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AMA7	-	D16|D17|D20|D23|D25|D5|D6| D14|D10|D12
AMA7_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AMA7	-	D16|D2|D3|D13|D18|D20|D24| D26|D5|D7
AMA2_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AMA2	-	D203|D94|D95|D96|D97|D98|D99| SP3|D237|D235
AMA2_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AMA2	-	D234|D235|D96|D236|D237|8400| 8401|8402|8403|8405
API1_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	API1	-	5942|9260|9261|9262|9263|9266| 9267|9268|7603|9269
API1_SV	Viral Phylogenetics; Ecological	3.3.5; 4.3.2	API1	-	5942|9260|9261|9262|9263|9274| 9275|9276|9277|9278

	Metagenomics				
API17_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	API17	-	8325|7959|8326|7986|8327|8328| 7980|8329|8330|8331
API17_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	API17	-	8325|7959|8326|7986|8327|8328| 7980|8329|8330|8331
API140_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	API140	-	9237|9238|7132|7141|9239|9240| 9241|9242|9245|9244
API140_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	API140	-	9237|9238|7132|7141|9239|9240| 9241|9242|9243|9244
API141_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	API141	-	9126|6849|9128|6877|9127|9130| 9129|9131|9132|9133
API141_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	API141	-	9126|6849|9128|6877|9127|9130| 9129|9131|9132|9133
AYA1_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AYA1	-	8228|7188|7181|7802|8229|7184| 8230|8231|6933|6934
AYA1_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AYA1	-	7184|8230|8231|6933|6934|8232| 8233|8234|8235|8237
AYA7_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AYA7	-	7064|8286|8287|8288|7047|8289| 8290|8291|8292|8293
AYA7_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AYA7	-	7064|8286|8287|8288|7047|8294| 8290|8291|8292|8293
AYA11_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AYA11	-	8201|8202|8203|7082|8204|7864| 8205|8206|8207|8208
AYA11_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AYA11	-	8201|8202|8203|7082|8204|8217| 8218|8206|8205|8221
AYA12_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AYA12	-	8222|7169|8223|8224|8225|8226| 8227
AYA12_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AYA12	-	8222|7169|8223|8224|8225|8226| 8227
AYA14_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AYA14	-	8297|8298|8299|8300|8301|8302| 8303|8304|8305|8306
AYA14_SV	Viral	3.3.5;	AYA14	-	8297|8298|8299|8300|8301|8320|

	Phylogenetics; Ecological Metagenomics	4.3.2			8321|8322|8323|7774
AYA15_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AYA15	-	9286|9284|9285|9287|9288|9289| 9290|9291|9292|9293
AYA15_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	AYA15	-	9286|9284|9285|9287|9288|9289| 9290|9291|9292|9293
CAJ1_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	CAJ1	-	8080|8081|8082|8083|4156|8084| 8085|8086|8087|8088
CAJ1_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	CAJ1	-	8080|8081|8082|8083|4156|8084| 8085|8086|8087|8088
CAJ2_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	CAJ2	-	8096|8097|8098|8099|8100|8101| 8102|8103|8104|6181
CAJ2_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	CAJ2	-	8096|8097|8098|8099|8100|8109| 8110|8111|8112|4693
CAJ4_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	CAJ4	-	D83|D84|D85|D86|D87|D88|D89| D90|D91|D92
CAJ4_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	CAJ4	-	8361|8362|8363|8364|8365|8366| 8367|8368|8369|8370
CUS8_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	CUS8	-	9216|9217|7503|9218|9219|7995| 9220|9221|9222|9223
CUS8_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	CUS8	-	9216|9217|7503|9218|9219|7995| 9220|9221|9222|9223
HUA1_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	HUA1	-	8009|8010|8011|8012|8013|8014| 8015|8016|8017|8018
HUA1_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	HUA1	-	8009|8010|8011|8012|8013|8014| 8015|8016|8017|8018
HUA2_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	HUA2	-	8332|8333|6750|8334|8335|7707| 8336|8337|8338|8339
HUA2_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	HUA2	-	8332|8333|6750|8334|8335|7707| 8336|8337|8338|8339
HUA3_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	HUA3	-	8003|8004|8005|8006|8007|7738| 9050|9022|8008

HUA3_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	HUA3	-	8003|8004|8005|8006|8007|7738|9050|9022|8008
HUA4_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	HUA4	-	8022|8023|8024|8025|8026|8027|8028|8029|8030|8031
HUA4_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	HUA4	-	8022|8023|8024|8025|8026|8027|8028|8029|8030|8031
LMA5_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	LMA5	-	9110|7681|3859|9111|9112
LMA5_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	LMA5	-	9110|7681|3859|9111|9112
LMA6_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	LMA6	-	9083|2926|4557|9084|9085|8060|8061|8062|8063|8064
LMA6_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	LMA6	-	9083|2926|4557|9084|9085|8072|5367|2941|8074|8075
LR2_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	LR2	-	D39|D57|D74|D75|D76|D77|D59|D256|D260|D261
LR2_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	LR2	-	D256|D257|D258|D259|D260|D261|D262|D0059
LR3_F	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	LR3	-	D242|D244|D243|D246|D248|D247|D249|D245|D250|D73
LR3_SV	Viral Phylogenetics; Ecological Metagenomics	3.3.5; 4.3.2	LR3	-	D249|D248|D243|D246|D245|D244|D247|D250|D242|D216

Appendix B: Supporting Material Chapter 2

B1: NGS final protocol

B1.1: Nucleic acid extraction

Total nucleic acid was extracted from samples using a Biosprint One-for-all Vet Kit (Qiagen) using a modified version of the manufacturer's protocol for purifying viral nucleic acids from swabs. Prior to inactivation, samples were processed in a MSC flow cabinet in a CL2+ laboratory. Extractions were performed by thawing samples and removing swabs using sterile forceps.

Each swab was first incubated in 144 μL Buffer RLT and 20 μL Proteinase K (both Qiagen) at 56°C for 15 minutes, vortexed for 15 seconds, then transferred to a second tube with 144 μL Buffer RLT and Proteinase K, incubated again for 15 minutes and vortexed again for 15 seconds. Samples were considered inactivated at this stage due to Buffer RLT, a lysis buffer containing guanidine isothiocyanate. The swab was then discarded and the two lysis buffer/Proteinase K solutions were briefly centrifuged, then combined and placed into a deep-well 96 sample extraction block with 25 μL MagAttract Suspension G (Qiagen) and 300 μL isopropyl alcohol. All of these steps were performed according to CL2+ guidelines.

Plates for wash steps were prepared; these included one plate containing 700 μL wash buffer AW1 and two plates containing 500 μL wash buffer RPE (both Qiagen). All plates were loaded onto a Kingfisher Flex 96 automated extraction machine (Thermo). The instrument settings, provided by Qiagen ('Protocol for purification of viral nucleic acid and bacterial DNA with Thermo Scientific KingFisher Flex'), consisted of a lysis and binding step, followed by three wash steps, and a final elution in 80 μL Buffer AVE (RNase-free water with 0.04% NaN_3). Extracted nucleic acid was stored at -80°C. All samples were quantified using a Qubit 3.0 fluorometer and a Qubit RNA HS Assay (Life Technologies) to determine RNA concentration for pooling. Samples with measurable RNA were pooled at approximately 120 ng RNA and unmeasurable samples were pooled up to a maximum volume of 30 μL as possible.

B1.2: Pooling and viral enrichment

Pools were treated with DNAse I (Ambion) to digest high molecular weight genomic DNA. Pool volume varied but buffer and enzyme were scaled such that all reactions contained 1X DNAse buffer and 2 Units (U) DNAse per 100 µL. Reactions were incubated at 37°C for 5 minutes, then immediately cleaned up with 1.8X Agencourt RNAClean XP beads (Beckman Coulter), washing the beads three times with 80% ethanol. Samples were then eluted in 20 µL nuclease-free water; 10 µL of eluate was used as input into the rRNA depletion step.

Pools were enriched by rRNA depletion using the Ribo-Zero rRNA Removal Kit (Human/Mouse/Rat) (Illumina) according to the manufacturer's instructions. Briefly, room temperature magnetic beads were prepared by washing twice in RNAse-free water, and then beads were resuspended in 65 µL Magnetic Bead Resuspension Solution. To each 10 µL RNA sample was added 18 µL RNAse-free water, 4 µL Ribo-Zero rRNA Reaction Buffer and 8 µL Ribo-Zero Removal Solution. Reactions were incubated at 68°C for 10 minutes, then at room temperature for 5 minutes to hybridize rRNA to probes. Pre-hybridized samples were added to the magnetic bead solution, then incubated at room temperature for 5 minutes followed by 50°C for 5 minutes. Samples were placed on a magnetic stand and 90 µL supernatant was removed, while beads containing hybridized rRNA were discarded. The enriched sample was cleaned up using 1.8X RNAClean XP beads and eluted in 10 µL RNAse-free water.

B1.3: Library preparation

First strand cDNA synthesis was performed using the Maxima H Minus First Strand cDNA Synthesis Kit (Thermo) by incubating 10 µL nucleic acid with 1 µL dNTPs and 1 µL random hexamers at 65°C for 5 minutes. Reactions were chilled on ice and then 4 µL 5X reverse transcriptase buffer, 1 µL reverse transcriptase, and 3 µL PCR-grade water were added for a total reaction volume of 20 µL. Reactions were incubated on a thermocycler at 25°C for 10 minutes, 60°C for 45 minutes, and 85°C for 5 minutes. Single strand cDNA was then immediately converted to double strand cDNA using the NEBNext mRNA Second Strand Synthesis Module (New England Biolabs) by adding 8 µL 10X second strand synthesis buffer, 4 µL second strand synthesis enzyme, and 48 µL PCR-grade water for a total reaction volume of 80 µL. Reactions were incubated on a thermocycler at 16°C for 2.5 hours, and dsDNA was stored at -20°C until library preparation.

Samples were library prepared using the KAPA DNA Library Preparation Kit for Illumina (KAPA Biosystems) modified for low input RNA samples. DNA was first cleaned up with 80 μL (1:1 bead:sample ratio) of Ampure XP beads (Beckman Coulter). Samples were eluted in 52 μL 10 mM Tris (pH 8.5) and beads left in solution. End repair was performed by adding 6 μL 10X end repair buffer and 2 μL end repair enzyme, and the 60 μL reaction was incubated at 20°C for 30 minutes. Samples were then cleaned up by adding 60 μL (1:1 ratio) of KAPA PEG/NaCl SPRI solution (KAPA Biosystems) to re-bind the DNA to the beads. Samples were eluted in 25 μL of Tris following clean-up, leaving the beads in solution.

A-tailing reactions were performed by adding 3 μL 10X A-tail buffer and 2 μL A-tail enzyme, then the 30 μL reaction volume was incubated at 30°C. 70 μL Tris was added to the reaction for a total volume of 100 μL, to which 100 μL SPRI solution (1:1 ratio) was added for cleanup. Samples were eluted in 15 μL Tris, leaving the beads in solution.

At this point, sample concentration was measured using a Qubit dsDNA HS Assay to calculate the volume of NEBNext Adaptor for Illumina (New England Biolabs) to add to adapter ligation reactions. Qubit readings were converted into picomoles (pmol) according to the formula (ng DNA * volume)/(average size (kbp) * 660). The amount of adapter added (pmol) was calculated as pmol DNA * 20, and for libraries with unmeasurable DNA, 0.5 μL of a 1:100 dilution of 15 μM adapter was added. Tris was added to calculated volume of adapters up to a total volume of 5 μL and this was combined with 14 μL A-tailed DNA, 5 μL 5X buffer and 1 μL T4 DNA ligase. The reaction was incubated at 20°C, then 1 μL USER enzyme was added to the reaction and incubated at 37°C to cleave a hairpin in the adapter. Cleanup was performed by adding 74 μL Tris for a total volume of 100 μL, followed by 100 μL SPRI solution. Samples were eluted in 11 μL Tris, this time removing DNA from beads.

PCR re-amplification was performed by first combining 12.5 μL PCR mastermix with 10 μL DNA. Each sample was barcoded using 1.25 μL universal primer and 1.25 μL individually barcoded primer (NEBNext Multiplex Oligos for Illumina Index Primers Set 1, New England Biolabs) or 1.25 μL of two different individually barcoded primers (Dual Index Primers Set 1, New England Biolabs). Number of PCR cycles varied depending on DNA concentration, from 12 cycles for higher concentrations up to a maximum of 16 cycles for undetectable DNA. Thermocycling parameters were: 3 minutes at 95°C, 12-16 cycles of 20s at 98°C, 15s at 65°C, and 30s at 72°C, followed by 2 minutes at 72°C.

Following PCR, 75 μL Tris was added for a total volume of 100 μL. Then 90 μL Ampure beads (0.9:1 bead:sample ratio) was added; the smaller ratio is intended to eliminate small fragments in the library such as primer dimer and adapter dimer. After bead cleanup, samples were eluted in 15 μL Tris. Libraries were validated using a Qubit dsDNA HS Assay and TapeStation 4200 D5000 ScreenTape (Agilent). Post-PCR libraries were often found to have high molecular weight peaks which can affect calculations for pooling and loading the sequencing instrument; in this case, Ampure beads were used in a size selection step. A 0.6X ratio of Ampure beads was added to the samples and the supernatant was removed, with larger fragments being retained by the beads. Ampure beads were then added to the supernatant for a final ratio of 1.4X beads/PEG-NaCl to sample, and samples were eluted in 15 μL Tris. Final libraries were pooled in equimolar ratios, and validation of the final pool was performed using a Qubit dsDNA HS Assay and a TapeStation D1000 ScreenTape. Sequencing was performed on an Illumina MiSeq or NextSeq500 at the MRC-University of Glasgow Centre for Virus Research.

B2: Bioinformatic pipeline and viral community datasets

A bioinformatic pipeline was created for analyzing vampire bat shotgun viral metagenomic data (Figure B2). Sequences were first demultiplexed according to barcode by the sequencing facility, then quality filtered using Trim Galore (Martin 2011; Andrews 2010). Low complexity reads and PCR duplicates were filtered out using prinseq-lite (Schmieder & Edwards 2011). Reads were mapped against a draft version of the vampire bat genome (Zepeda-Mendoza *et al.* 2018; NCBI BioProject Accession PRJNA414273) as well as the genome of the PhiX virus that is used as a positive control in Illumina sequencing, and is a widespread contaminant of previously published microbial genomes (Mukherjee *et al.* 2015). Mapping was performed using bowtie2 (Langmead *et al.* 2009), and only unmapped reads were retained for further analysis. The program RiboPicker (Schmieder *et al.* 2012) was then used to remove reads associated with rRNA, as some will be sequenced despite the rRNA depletion treatment.

The program Diamond (Buchfink *et al.* 2014) was further used to remove reads mapping to other eukaryotes and prokaryotes by comparing reads with the non-redundant NCBI database. Reads mapping to viruses by Diamond were retained for the next step of the analysis using a custom script ('Allmond') written for this purpose. Reads with no hits were also retained because they could be of viral origin, but not yet characterized in

databases, or subsequently form larger contigs that can be classified as viral. The remaining reads (viral reads and reads with no hits) were then characterized through comparison to the Viral Refseq Protein NCBI database using Diamond with a maximum e-value of 0.001 and retaining only the top hit as the final viral classification.

In addition to analyses at the read level, the remaining reads (viral reads and reads with no hits) were then *de novo* assembled into contigs using the program SPAdes (Bankevich *et al.* 2012). Contigs were first compared to the non-redundant NCBI database using Diamond to remove other prokaryotic and eukaryotic matches, and then compared to the Viral Refseq Protein database, and contigs matching viral sequences were retained as a final set of viral contigs. An additional step in the pipeline predicts open reading frames (ORFs) in all contigs using the program getORF (Rice *et al.* 2000), including those that have not been assigned to any known viral taxa, and which could represent new viral species or groups.

For reproducibility, the Diamond databases used for analyses (non-redundant, ViralRefSeq, and RefSeq protein) were standardized for samples that were compared to one another.

Viral reads and contigs from the bioinformatic pipeline were converted from tabular blast output into lists of viral taxa at different taxonomic levels using MEGAN Community Edition (Huson *et al.* 2016). The default parameters of the lowest common ancestor (LCA) assignment algorithm were used except that minimum score and minimum support percent were set to zero, such that all hits passing the filters of the bioinformatic pipeline (maximum e-value of 0.001 for each Diamond blast step) were included in initial analyses. Taxa lists were exported at family and genus levels so that downstream comparisons could be performed at various levels of taxonomic hierarchy. For read-level analysis, species-level assignments were not considered trustworthy as reads were only 150bp long and a read of this length could match equally well to numerous species within a genus. However, genera were included that are not yet assigned to a family and species that are not yet assigned to a genus. For example, the genus *Deltavirus* is not assigned to a family, but was considered as a taxonomic group in family-level analyses. Taxa lists were filtered for vertebrate-infecting viruses using R version 3.4.2 (R Core Team 2017) and a list of vertebrate-infecting viral families and genera (Table B1) that was compiled based on the 2017 ICTV Taxonomy (Adams *et al.* 2017).

B3: Tables

Table B 1 Families and genera of viruses that infect vertebrates.
This list was compiled based on the 2017 ICTV Taxonomy (Adams *et al*. 2017) and was used to filter viral datasets for analyses that included only vertebrate-infecting viruses. The list is comprehensive and includes vertebrate-infecting viral taxa that were not detected in any samples analyzed.

Family	Genus	Family	Genus
Hantaviridae	*Orthohantavirus*	*Caliciviridae*	*Norovirus*
Nairoviridae	*Orthonairovirus*		*Sapovirus*
Peribunyaviridae	*Herbevirus*		*Vesivirus*
	Orthobunyavirus	*Circoviridae*	*Circovirus*
Phenuiviridae	*Goukovirus*		*Cyclovirus*
	Phlebovirus	*Flaviviridae*	*Flavivirus*
Alloherpesviridae	*Batrachovirus*		*Hepacivirus*
	Cyprinivirus		*Pegivirus*
	Ictalurivirus		*Pestivirus*
	Salmonivirus	*Genomoviridae*	*Gemycircularvirus*
Herpesviridae	*Iltovirus*		*Gemygorvirus*
	Mardivirus		*Gemykibivirus*
	Scutavirus		*Gemykolovirus*
	Simplexvirus		*Gemykrogvirus*
	Varicellovirus		*Gemykroznavirus*
	Cytomegalovirus		*Gemytondvirus*
	Muromegalovirus		*Gemyvongvirus*
	Proboscivirus	*Hepadnaviridae*	*Avihepadnavirus*
	Roseolovirus		*Orthohepadnavirus*
	Lymphocryptovirus	*Hepeviridae*	*Orthohepevirus*
	Macavirus		*Piscihepevirus*
	Percavirus	*Iridoviridae*	*Lymphocystivirus*
	Rhadinovirus		*Megalocytivirus*
Bornaviridae	*Bornavirus*		*Ranavirus*
Filoviridae	*Cuevavirus*	*Nodaviridae*	*Alphanodavirus*
	Ebolavirus		*Betanodavirus*
	Marburgvirus	*Orthomyxoviridae*	*Influenzavirus A*
Nyamiviridae	*Nyavirus*		*Influenzavirus B*
Paramyxoviridae	*Aquaparamyxovirus*		*Influenzavirus C*
	Avulavirus		*Influenzavirus D*
	Ferlavirus		*Isavirus*
	Henipavirus		*Quaranjavirus*
	Morbillivirus		*Thogotovirus*
	Respirovirus	*Papillomaviridae*	*Alphapapillomavirus*
	Rubulavirus		*Betapapillomavirus*
Pneumoviridae	*Metapneumovirus*		*Chipapillomavirus*
	Orthopneumovirus		*Deltapapillomavirus*
Rhabdoviridae	*Curiovirus*		*Dyochipapillomavirus*
	Ephemerovirus		*Dyodeltapapillomavirus*
	Hapavirus		*Dyoepsilonpapillomavirus*

Family	Genus	Family	Genus
	Ledantevirus		*Dyoetapapillomavirus*
	Lyssavirus		*Dyoiotapapillomavirus*
	Novirhabdovirus		*Dyokappapapillomavirus*
	Perhabdovirus		*Dyolambdapapillomavirus*
	Sprivivirus		*Dyomupapillomavirus*
	Sripuvirus		*Dyonupapillomavirus*
	Tibrovirus		*Dyoomegapapillomavirus*
	Tupavirus		*Dyoomikronpapillomavirus*
	Vesiculovirus		*Dyophipapillomavirus*
Sunviridae	*Sunshinevirus*		*Dyopipapillomavirus*
Arteriviridae	*Dipartevirus*		*Dyopsipapillomavirus*
	Equartevirus		*Dyorhopapillomavirus*
	Nesartevirus		*Dyosigmapapillomavirus*
	Porartevirus		*Dyotaupapillomavirus*
	Simartevirus		*Dyothetapapillomavirus*
Coronaviridae	*Alphacoronavirus*		*Dyoupsilonpapillomavirus*
	Betacoronavirus		*Dyoxipapillomavirus*
	Deltacoronavirus		*Dyozetapapillomavirus*
	Gammacoronavirus		*Epsilonpapillomavirus*
	Bafinivirus		*Etapapillomavirus*
	Torovirus		*Gammapapillomavirus*
Picornaviridae	*Ampivirus*		*Iotapapillomavirus*
	Aphthovirus		*Kappapapillomavirus*
	Aquamavirus		*Lambdapapillomavirus*
	Avihepatovirus		*Mupapillomavirus*
	Avisivirus		*Nupapillomavirus*
	Cardiovirus		*Omegapapillomavirus*
	Cosavirus		*Omikronpapillomavirus*
	Dicipivirus		*Phipapillomavirus*
	Enterovirus		*Pipapillomavirus*
	Erbovirus		*Psipapillomavirus*
	Gallivirus		*Rhopapillomavirus*
	Harkavirus		*Sigmapapillomavirus*
	Hepatovirus		*Taupapillomavirus*
	Hunnivirus		*Thetapapillomavirus*
	Kobuvirus		*Treisdeltapapillomavirus*
	Kunsagivirus		*Treisepsilonpapillomavirus*
	Limnipivirus		*Treisetapapillomavirus*
	Megrivirus		*Treiszetapapillomavirus*
	Mischivirus		*Upsilonpapillomavirus*
	Mosavirus		*Xipapillomavirus*
	Oscivirus		*Zetapapillomavirus*
	Parechovirus	*Parvoviridae*	*Amdoparvovirus*
	Pasivirus		*Aveparvovirus*
	Passerivirus		*Bocaparvovirus*
	Potamipivirus		*Copiparvovirus*
	Rabovirus		*Dependoparvovirus*

	Rosavirus			Erythroparvovirus
	Sakobuvirus			Protoparvovirus
	Salivirus			Tetraparvovirus
	Sapelovirus	Picobirnaviridae	Picobirnavirus
	Senecavirus	Polyomaviridae	Alphapolyomavirus
	Sicinivirus			Betapolyomavirus
	Teschovirus			Deltapolyomavirus
	Torchivirus			Gammapolyomavirus
	Tremovirus	Poxviridae	Avipoxvirus
Adenoviridae	Atadenovirus			Capripoxvirus
	Aviadenovirus			Centapoxvirus
	Ichtadenovirus			Cervidpoxvirus
	Mastadenovirus			Crocodylidpoxvirus
	Siadenovirus			Leporipoxvirus
Anelloviridae	Alphatorquevirus			Molluscipoxvirus
	Betatorquevirus			Orthopoxvirus
	Deltatorquevirus			Parapoxvirus
	Epsilontorquevirus			Suipoxvirus
	Etatorquevirus			Yatapoxvirus
	Gammatorquevirus	Reoviridae	Orbivirus
	Gyrovirus			Rotavirus
	Iotatorquevirus			Seadornavirus
	Kappatorquevirus			Aquareovirus
	Lambdatorquevirus			Coltivirus
	Thetatorquevirus			Orthoreovirus
	Zetatorquevirus	Retroviridae	Alpharetrovirus
Arenaviridae	Mammarenavirus			Betaretrovirus
	Reptarenavirus			Deltaretrovirus
Asfarviridae	Asfivirus			Epsilonretrovirus
Astroviridae	Avastrovirus			Gammaretrovirus
	Mamastrovirus			Lentivirus
Birnaviridae	Aquabirnavirus			Spumavirus
	Avibirnavirus	Togaviridae	Alphavirus
	Blosnavirus			Rubivirus
Caliciviridae	Lagovirus	Deltavirus	Deltavirus
	Nebovirus	Tilapinevirus	Tilapinevirus

Table B 2 Bovine viral diarrhea virus nucleotide blast hits from selected FBS contigs.
Long contigs (>700bp) were selected for further analysis by nucleotide blast from the two
batches of FBS that were metagenomically sequenced.

Contig	Contig Length (bp)†	K-mer coverage†	Blast Genbank ID	Blast Query Cover	Blast Identity (%)	Description of blast hit
FBS1 NODE_57	1396	3.33	KY683847.1	100%	96%	Bovine viral diarrhea virus 3 strain SV757/15
FBS1 NODE_74	1280	2.49	KY683847.1	100%	97%	Bovine viral diarrhea virus 3 strain SV757/15
FBS1 NODE_99	1165	4.07	KY683847.1	100%	97%	Bovine viral diarrhea virus 3 strain SV757/15
FBS2 NODE_244	775	1.37	KY683847.1	100%	98%	Bovine viral diarrhea virus 3 strain SV757/15

†Contig length and k-mer coverage reported are based on the SPAdes assembly

Table B 3 Model comparison for viral genera detection in subsampling analysis.
Linear and polynomial models were compared for each sample type (feces and saliva) and
filtering (all viral genera and vertebrate-infecting only) combination at the genus level. For
each combination, two models were constructed and compared through likelihood ratio test
(L, X^2, d.f., and P-value) and AIC (AIC and ΔAIC).

	Model	L	X^2	d.f.	P-value	AIC	ΔAIC
All Genera Fecal	Linear	-590.54	31.59	1	1.90E-08	1187.1	29.591
	Polynomial	-574.75				1157.5	
All Genera Saliva	Linear	-847.37	42.641	1	6.58E-11	1700.7	40.641
	Polynomial	-826.04				1660.1	
Vertebrate-infecting Genera Fecal	Linear	-400.72	8.8299	1	0.002963	807.45	6.83
	Polynomial	-396.31				800.62	
Vertebrate-infecting Genera Saliva	Linear	-715.48	15.786	1	7.09E-05	1437	13.786
	Polynomial	-707.59				1423.2	

B4: Figures
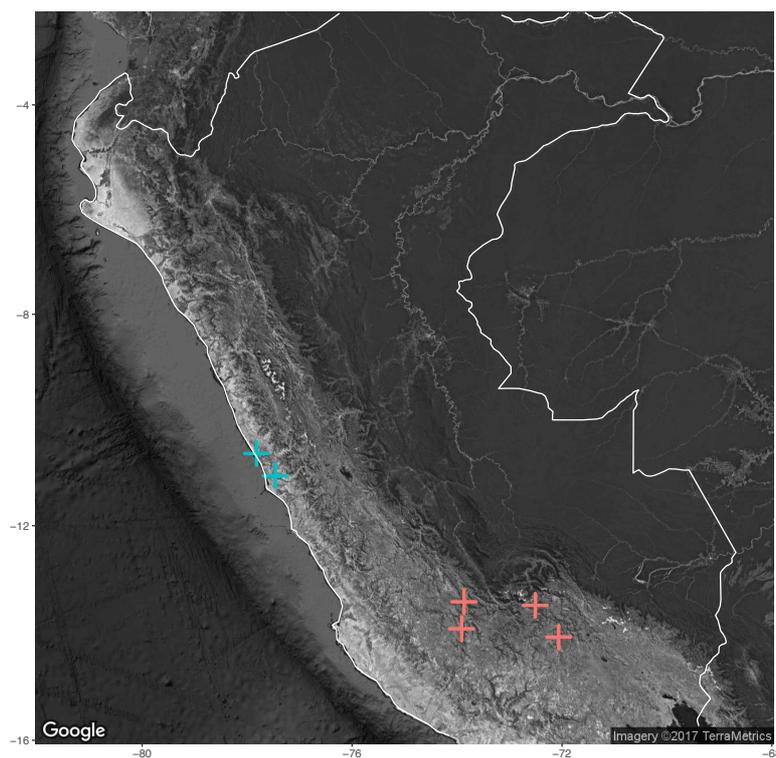
Figure B 1 Bat colonies in Peru analyzed in preliminary metagenomic sequencing.
Red crosses represent four sites in the departments of Ayacucho, Apurimac and Cusco
from which samples were stored in VTM and analyzed in preliminary sequencing run 1. Blue
crosses represent two sites in the department of Lima, from which samples were were
paired VTM/RNALater swabs that were tested for the effects of storage buffer and
enrichment in preliminary sequencing run 2. White lines show the outline of Peru.

Key Output Files	Step Name	Settings	Other Output Files	Script Name

| FBS1_S9_R1_001.fastq
FBS1_S9_R2_001.fastq | (1) Raw Reads | | FBS1_log.txt | |

| FBS1_S9_R1_001_val_1.fastq
FBS1_S9_R2_001_val_2.fastq | (2) Adapter/Quality
trim_galore | quality (25), length (75), right clip (14) | trim_galore_out.txt
FBS1_R1_001.fastq_trimming_report.txt
FBS1_R2_001.fastq_trimming_report.txt | trim.sh |

| FBS1_S9_R1_001_val_1.fastq
FBS1_S9_R2_001_val_2.fastq | (3) LowComp/Dups
prinseq | low complexity DUST (7)
derep duplicates (1,4) | prinseq_out.txt | trim.sh |

| FBS1_unmap_1.fastq
FBS1_unmap_2.fastq | (4) Host Mapping
bowtie2 | vampire bat genome
PhiX genome | FBS1_bt2_vampire.txt
FBS1_bt2_phix.txt | host.sh |

| FBS1_nonrrna_1.fastq
FBS1_nonrrna_2.fastq | (5) Ribosome Filter
ribopicker | ssr123 db (small subunit)
slr123 db (large subunit) | FBS1_ribo_ssr123.tsv
FBS1_ribo_slr123.tsv | ribo.sh |

| FBS1_clean_1.fastq
FBS1_clean_2.fastq | (6) Clean Euk/Pro
diamond/allmond | refseq protein DB | FBS1_diam.daa
FBS1_diam.html | clean.sh |

| FBS1_diam_ref.html
FBS1_diam_ref_sort_out.txt | (7) Classify Reads
diamond/allmond | Viral refseq protein DB
Top diamond hit only | FBS1_diam_ref.daa
allmond_output.txt | diam.sh |

| contigs.fasta
contigs_sort.fasta | (8) De Novo
spades | only-assembler | ./spades_fil/ directory | spades.sh |

| FBS1_diam_ref.html
FBS1_diam_ref_sort_out.txt | (9) Classify Contigs
diamond/allmond | Viral refseq protein DB
Top diamond hit only | FBS1_diam_ref.daa | spades.sh |

| novel_contigs.fasta
novel_orfs.fasta | (10) Novel ORFs
getORF/allmond | minsize 90
find 1 (start - stop) | | spades.sh |

Illustrated for FBS1 sample

Figure B 2 Schematic diagram of bioinformatic pipeline.
The bioinformatic pipeline was developed specifically to analyze vampire bat viral communities at both the read and contig levels. The diagram depicts each step, the script used to perform the analysis, settings or specifications used, and output files generated (using sample FBS1 as an example).
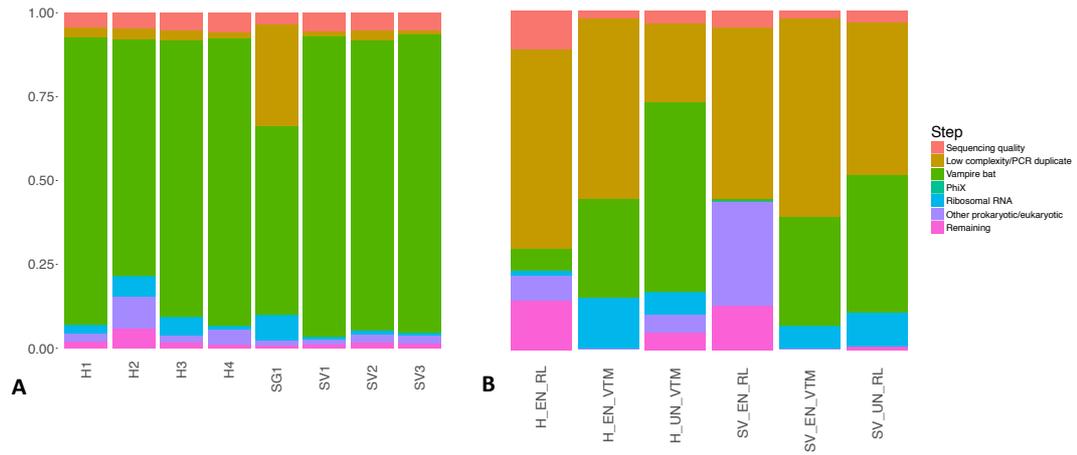
Figure B 3 Reads from preliminary sequencing runs filtered during bioinformatic processing
Sample names correspond to A) Table 2.1 and B) Table 2.4. Sequencing quality reads are those removed by Trim Galore, low complexity/duplicate reads are removed by prinseq, vampire bat are reads mapping the vampire bat genome, PhiX are reads mapping to the PhiX genome, ribosomal RNA are those removed by RiboPicker, and other prokaryote/eukaryote are reads assigned to those taxa when using Diamond to compare reads to the nr database. Remaining reads are assigned to viruses or unassigned.
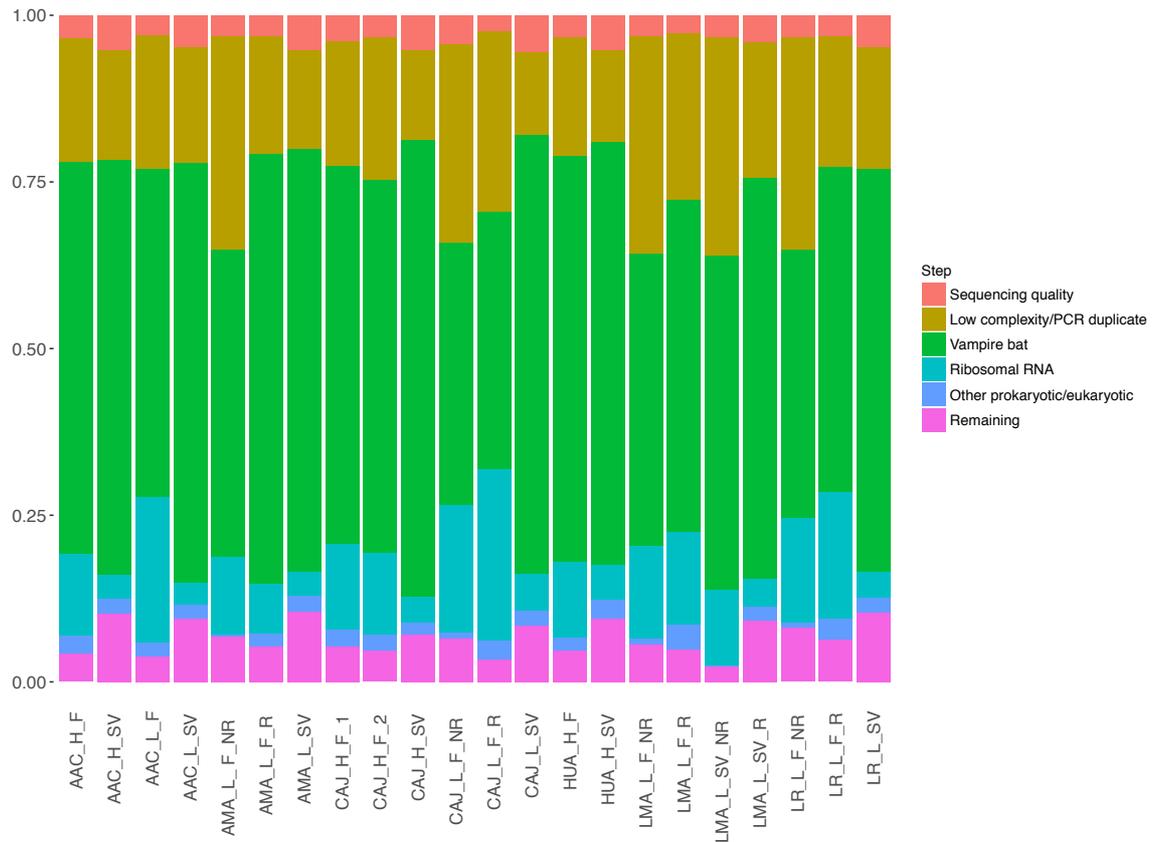
Figure B 4 Reads from multi-colony pools filtered during bioinformatic processing
Sample names correspond to Table 2.5. Sequencing quality reads are those removed by Trim Galore, low complexity/duplicate reads are removed by prinseq-lite, vampire bat are reads mapping the vampire bat genome, Ribosomal RNA are those removed by RiboPicker, and other prokaryote/eukaryote are reads assigned to those taxa when using Diamond to compare reads to the Genbank nr database. Remaining reads are those assigned to viruses or unassigned.
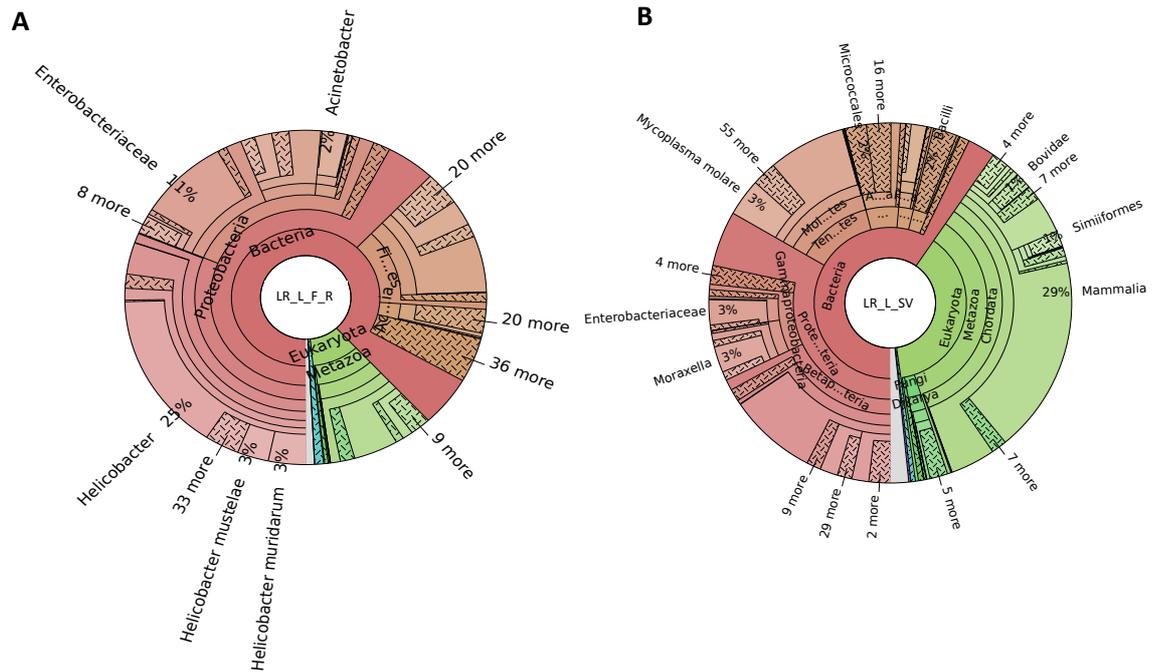
Figure B 5 Krona plots of read assignments in vampire bat feces and saliva.
Reads are from locality LR in (A) fecal and (B) saliva pools which were processed according to the optimized protocol. Reads are shown following quality filtering, rRNA depletion, and host subtraction but prior to subtraction of reads closely matching other prokaryotic/eukaryotic taxa based on Diamond blast comparison to the Genbank nr database. Red segments are bacterial taxa and green are eukaryotic taxa, while archaea and viruses are represented by smaller blue and purple segments respectively, and unassigned taxa are shown in gray. Taxa with a high percentage of reads out of the total have names shown, while names of taxa with lower percentages are not depicted to facilitate visualization. Sample names are shown inside plots and correspond to Table 2.5.
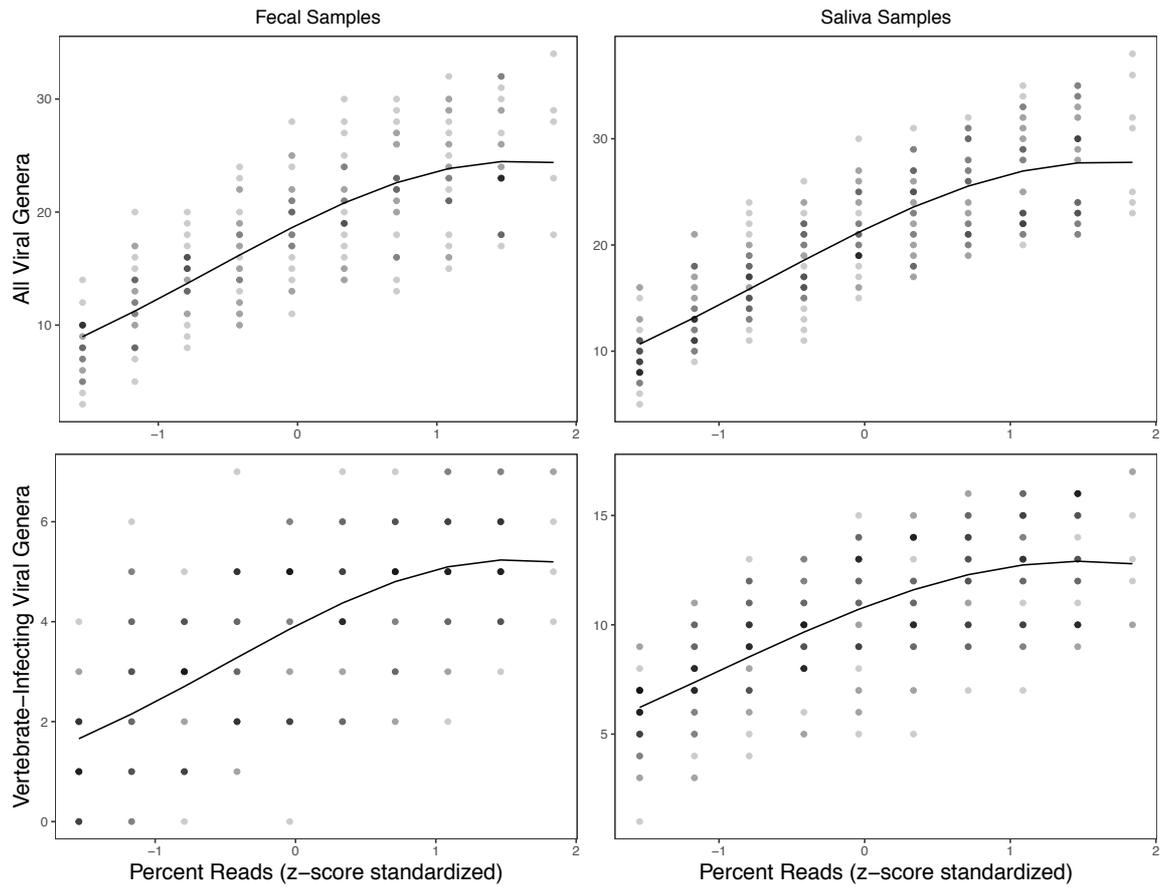
Figure B 6 Viral genera communities saturate at the genus level.
Number of viral genera and vertebrate viral genera detected in fecal (N=5) and saliva (N=7)
samples at increasing percentage of the original raw reads. Percent reads are z-score
standardized by subtracting the mean and dividing by the standard deviation. Points are
semi-transparent with darker points indicating more subsamples with a given value and
show the rescaled original data. Lines show the model prediction.

Appendix C: Supporting Material Chapter 3

C1: Tables

Table C 1 Natural host groups of viral reads summed across multi-colony pools.
This table shows the number of reads assigned to viruses that typically infect a given class of host. Reads were assigned at both the family and genus level. Host assignments are shown for each viral taxon at the family level in Table C2 and the genus level in Table C3.

	Fecal Reads		Saliva Reads	
	Genus	Family	Genus	Family
Bacteria	33,186	40,236	676	865
Fungi	23	2	49	22
Invertebrate	175	249	21	18
Invertebrate/Vertebrate†	16	3,637	4	193
Other	92	117	11	33
Plant	2,659	3,116	94	101
Plant/Fungi†	5	49	2	43
Vertebrate	27,661	31,945	1,494	2,474

†Invertebrate/Vertebrate and Plant/Fungi reflect typically families (or less often genera) with low specificity in host range

Table C 2 Host assignments at the family level for viral taxa in multi-colony pools.
Viral taxa are presented along with the class of host that is typically infected based on Hulo *et al.* (2011) and Adams *et al.* (2017).

Taxa	Infects		Taxa	Infects
Deltavirus	Vertebrate		*Microviridae*	Bacteria
Adenoviridae	Vertebrate		*Parvoviridae*	Invertebrate/ Vertebrate
Baculoviridae	Invertebrate		*Paramyxoviridae*	Vertebrate
Myoviridae	Bacteria		*Rhabdoviridae*	Invertebrate/ Vertebrate/Plant
Podoviridae	Bacteria		*Astroviridae*	Vertebrate
Siphoviridae	Bacteria		*Bromoviridae*	Plant
Herpesviridae	Vertebrate		*Carmotetraviridae*	Invertebrate
Iridoviridae	Invertebrate/ Vertebrate		*Flaviviridae*	Invertebrate/ Vertebrate
Marseilleviridae	Other (Amoeba)		*Hepeviridae*	Vertebrate
Mimiviridae	Other (Amoeba)		*Leviviridae*	Bacteria
Papillomaviridae	Vertebrate		*Luteoviridae*	Plant
Phycodnaviridae	Other (Alga)		*Narnaviridae*	Fungi
Polydnaviridae	Invertebrate		*Coronaviridae*	Vertebrate
Poxviridae	Invertebrate/ Vertebrate		*Nodaviridae*	Invertebrate/ Vertebrate
Birnaviridae	Invertebrate/ Vertebrate		*Ourmiavirus*	Plant
Chrysoviridae	Fungi		*Dicistroviridae*	Invertebrate
Cystoviridae	Bacteria		*Iflaviridae*	Invertebrate
Hypoviridae	Fungi		*Picornaviridae*	Vertebrate
Partitiviridae	Plant/Fungi		*Potyviridae*	Plant

Picobirnaviridae	Vertebrate		*Sobemovirus*	Plant
Reoviridae	Invertebrate/ Vertebrate/Plant		*Togaviridae*	Invertebrate/ Vertebrate
Totiviridae	Other (Fungi/Protozoa)		*Tombusviridae*	Plant
Caulimoviridae	Plant		*Alphaflexiviridae*	Plant/Fungi
Retroviridae	Vertebrate		*Betaflexiviridae*	Plant/Fungi
Anelloviridae	Vertebrate		*Tymoviridae*	Plant
Circoviridae	Vertebrate		*Virgaviridae*	Plant
Inoviridae	Bacteria		*Fusariviridae*	Fungi

Table C 3 Host assignments at the genus level for viral taxa in multi-colony pools.
Viral taxa are presented along with the class of host that is typically infected based on Hulo
et al. **(2011) and Adams** *et al.* **(2017).**

Taxa	Host group		Taxa	Host group
Deltavirus	Vertebrate		*Hypovirus*	Fungi
Mastadenovirus	Vertebrate		*Alphapartitivirus*	Plant/Fungi
Alphabaculovirus	Invertebrate		*Betapartitivirus*	Plant/Fungi
Bcep78likevirus	Bacteria		*Gammapartitivirus*	Fungi
Bcepmulikevirus	Bacteria		*Picobirnavirus*	Vertebrate
Cp220likevirus	Bacteria		*Orbivirus*	Vertebrate
Cp8unalikevirus	Bacteria		*Rotavirus*	Vertebrate
I3likevirus	Bacteria		*Totivirus*	Fungi
Mulikevirus	Bacteria		*Victorivirus*	Fungi
P2likevirus	Bacteria		*Caulimovirus*	Plant
PhiCD119likevirus	Bacteria		*Alpharetrovirus*	Vertebrate
Phikzlikevirus	Bacteria		*Betaretrovirus*	Vertebrate
Twortlikevirus	Bacteria		*Gammaretrovirus*	Vertebrate
Schizot4likevirus	Bacteria		*Spumavirus*	Vertebrate
T4likevirus	Bacteria		*Inovirus*	Bacteria
Phikmvlikevirus	Bacteria		*Chlamydiamicrovirus*	Bacteria
Sp6likevirus	Bacteria		*Microvirus*	Bacteria
T7likevirus	Bacteria		*Dependoparvovirus*	Vertebrate
Bcep22likevirus	Bacteria		*Morbillivirus*	Vertebrate
Bppunalikevirus	Bacteria		*Respirovirus*	Vertebrate
Epsilon15likevirus	Bacteria		*Rubulavirus*	Vertebrate
F116likevirus	Bacteria		*Lyssavirus*	Vertebrate
N4likevirus	Bacteria		*Vesiculovirus*	Vertebrate
P22likevirus	Bacteria		*Avastrovirus*	Vertebrate
Phi29likevirus	Bacteria		*Mamastrovirus*	Vertebrate
Andromedalikevirus	Bacteria		*Cucumovirus*	Plant
Barnyardlikevirus	Bacteria		*Alphacarmotetravirus*	Invertebrate
Bronlikevirus	Bacteria		*Flavivirus*	Vertebrate
C5likevirus	Bacteria		*Hepacivirus*	Vertebrate
L5likevirus	Bacteria		*Pegivirus*	Vertebrate
Lambdalikevirus	Bacteria		*Hepevirus*	Vertebrate
Phie125likevirus	Bacteria		*Allolevivirus*	Bacteria
Phietalikevirus	Bacteria		*Levivirus*	Bacteria

Phijlunalikevirus	Bacteria		*Luteovirus*	Plant
T5likevirus	Bacteria		*Mitovirus*	Fungi
Simplexvirus	Vertebrate		*Alphacoronavirus*	Vertebrate
Cytomegalovirus	Vertebrate		*Betacoronavirus*	Vertebrate
Muromegalovirus	Vertebrate		*Deltacoronavirus*	Vertebrate
Lymphocryptovirus	Vertebrate		*Alphanodavirus*	Invertebrate/ Vertebrate
Macavirus	Vertebrate		*Ourmiavirus*	Plant
Percavirus	Vertebrate		*Aparavirus*	Invertebrate
Rhadinovirus	Vertebrate		*Cripavirus*	Invertebrate
Lymphocystivirus	Vertebrate		*Iflavirus*	Invertebrate
Cafeteriavirus	Other (Amoeba)		*Avihepatovirus*	Vertebrate
Mimivirus	Other (Amoeba)		*Parechovirus*	Vertebrate
Alphapapillomavirus	Vertebrate		*Sakobuvirus*	Vertebrate
Betapapillomavirus	Vertebrate		*Salivirus*	Vertebrate
Chipapillomavirus	Vertebrate		*Potyvirus*	Plant
Dyodeltapapillomavirus	Vertebrate		*Sobemovirus*	Plant
Dyolambdapapillomavirus	Vertebrate		*Rubivirus*	Vertebrate
Dyomupapillomavirus	Vertebrate		*Alphanecrovirus*	Plant
Dyonupapillomavirus	Vertebrate		*Aureusvirus*	Plant
Gammapapillomavirus	Vertebrate		*Carmovirus*	Plant
Kappapapillomavirus	Vertebrate		*Macanavirus*	Plant
Lambdapapillomavirus	Vertebrate		*Machlomovirus*	Plant
Omegapapillomavirus	Vertebrate		*Panicovirus*	Plant
Omikronpapillomavirus	Vertebrate		*Tombusvirus*	Plant
Pipapillomavirus	Vertebrate		*Umbravirus*	Plant
Rhopapillomavirus	Vertebrate		*Allexivirus*	Plant
Upsilonpapillomavirus	Vertebrate		*Vitivirus*	Plant
Chlorovirus	Other (Alga)		*Maculavirus*	Plant
Coccolithovirus	Other (Alga)		*Marafivirus*	Plant
Prasinovirus	Other (Alga)		*Tymovirus*	Plant
Prymnesiovirus	Other (Alga)		*Acyrthosiphon pisum virus*	Invertebrate
Bracovirus	Invertebrate		*Chronic bee paralysis virus*	Invertebrate
Avipoxvirus	Vertebrate		*Diaporthe ambigua RNA virus 1*	Fungi
Orthopoxvirus	Vertebrate		*Halastavi arva RNA virus*	Vertebrate
Micromonas pusilla virus 12T	Other (Alga)		*Jingmen tick virus*	Invertebrate
Pandoravirus salinus	Other (Amoeba)		*Solenopsis invicta virus 2*	Invertebrate
Aquabirnavirus	Vertebrate		*Tobamovirus*	Plant
Entomobirnavirus	Invertebrate		*Tobravirus*	Plant
Chrysovirus	Fungi			

Table C 4 Summary of viral contigs >3000 bp found in multi and single colony pools. The pool in which each contig was found is presented along with contig length, assigned family and where possible assigned genus; some contigs are only assigned to a family when matching equally well to multiple genera.

Pool	Length	Family	Genus
AMA_F	29,140	*Coronaviridae*	*Alphacoronavirus*
LMA_F	27,407	*Podoviridae*	
LMA_F	18,009	*Podoviridae*	*T7virus*
CAJ_L_F	15,126	*Podoviridae*	*T7virus*
AAC_H_F	14,431	*Podoviridae*	*T7virus*
CAJ_H_F	13,312	*Podoviridae*	*T7virus*
AMA_F	13,154	*Adenoviridae*	*Mastadenovirus*
AMA_F	10,078	*Dicistroviridae*	*Cripavirus*
LMA_F	9,848	*Podoviridae*	
LMA_F	8,012	*Podoviridae*	*Kf1virus*
LR_F	7,339	*Adenoviridae*	*Mastadenovirus*
HUA_F	6,874	*Podoviridae*	*T7virus*
AAC_L_F	6,809	*Podoviridae*	*T7virus*
CAJ_H_F	6,749	*Podoviridae*	*T7virus*
AAC_L_F	6,492	*Tymoviridae*	*Marafivirus*
LR_F	6,481	*Tymoviridae*	*Marafivirus*
LMA_F	6,214	*Podoviridae*	*T7virus*
HUA_F	6,137	*Podoviridae*	*T7virus*
LMA_F	5,919	*Myoviridae*	
AAC_H_F	5,841	*Podoviridae*	*T7virus*
AAC_H_F	5,590	*Astroviridae*	
AAC_L_F	5,545	*Podoviridae*	*T7virus*
HUA_F	5,539	*Podoviridae*	*T7virus*
CAJ_L_F	5,366	*Podoviridae*	*Kf1virus*
AMA_F	5,330	*Adenoviridae*	*Mastadenovirus*
LMA_F	5,238	*Podoviridae*	*T7virus*
LMA_F	5,231	*Podoviridae*	*T7virus*
HUA_F	5,184	*Podoviridae*	*T7virus*
AAC_L_F	5,109	*Tombusviridae*	*Alphacarmovirus*
CAJ_L_F	5,096	*Tombusviridae*	*Pelarspovirus*
AAC_H_F	5,072	*Myoviridae*	*Cp8virus*
AAC_L_F	5,047	*Tombusviridae*	*Alphacarmovirus*
AAC_L_F	5,031	*Tombusviridae*	*Alphacarmovirus*
AMA_F	5,005	*Tombusviridae*	*Alphacarmovirus*
CAJ_L_F	4,932	*Podoviridae*	*Kf1virus*
AMA_F	4,833	*Tombusviridae*	*Betacarmovirus*
CAJ_H_F	4,827	*Podoviridae*	*T7virus*
CAJ_L_F	4,660	*Myoviridae*	
AAC_H_F	4,576	*Myoviridae*	*Cp8virus*
AMA_F	4,256	*Adenoviridae*	*Mastadenovirus*
CAJ_H_F	4,104	*Podoviridae*	*T7virus*
AAC_H_F	4,094	*Podoviridae*	*T7virus*

CAJ_L_F	4,025	*Tombusviridae*	*Macanavirus*
HUA_F	3,969	*Podoviridae*	*T7virus*
LR_F	3,949	*Tombusviridae*	*Macanavirus*
CAJ_L_F	3,945	*Podoviridae*	*T7virus*
AMA_F	3,924	*Adenoviridae*	*Mastadenovirus*
AMA_F	3,915	*Tombusviridae*	*Macanavirus*
LMA_SV	3,843	*Leviviridae*	
AAC_H_SV	3,833	*Leviviridae*	*Levivirus*
LR_F	3,820	*Adenoviridae*	*Mastadenovirus*
HUA_SV	3,582	*Baculoviridae*	*Alphabaculovirus*
LR_F	3,578	*Tombusviridae*	*Pelarspovirus*
HUA_F	3,569	*Reoviridae*	*Rotavirus*
LMA_SV	3,553	*Leviviridae*	
AAC_L_F	3,480	*Podoviridae*	*T7virus*
CAJ_L_F	3,445	*Siphoviridae*	
CAJ_L_F	3,296	*Siphoviridae*	
AAC_L_F	3,210	*Podoviridae*	*T7virus*
LR_F	3,186	*Dicistroviridae*	*Aparavirus*
CAJ_L_F	3,156	*Podoviridae*	
LR_F	3,116	*Adenoviridae*	*Mastadenovirus*
CAJ_L_SV	3,001	*Myoviridae*	
CAJ2_F	24,397	*Podoviridae*	*T7virus*
LMA5_F	22,769	*Podoviridae*	
API17_F	21,021	*Podoviridae*	*T7virus*
API141_F	19,896	*Podoviridae*	*T7virus*
AYA12_F	18,291	*Podoviridae*	*T7virus*
LMA5_F	16,982	*Podoviridae*	*T7virus*
HUA4_F	16,149	*Podoviridae*	*T7virus*
HUA1_F	14,917	*Podoviridae*	*T7virus*
LMA5_F	13,701	*Podoviridae*	
HUA4_F	12,414	*Coronaviridae*	*Alphacoronavirus*
HUA1_F	12,270	*Podoviridae*	*T7virus*
API17_F	12,060	*Podoviridae*	*T7virus*
AYA12_F	12,057	*Podoviridae*	*T7virus*
LR2_F	10,947	*Adenoviridae*	*Mastadenovirus*
AYA15_F	10,178	*Podoviridae*	*T7virus*
HUA4_F	10,085	*Dicistroviridae*	*Cripavirus*
HUA4_F	9,926	*Coronaviridae*	*Alphacoronavirus*
CAJ4_F	9,247	*Podoviridae*	*T7virus*
CUS8_F	8,784	*Podoviridae*	*T7virus*
AMA2_F	8,582	*Adenoviridae*	*Mastadenovirus*
AYA7_F	8,564	*Podoviridae*	*T7virus*
HUA4_F	8,509	*Podoviridae*	*T7virus*
AMA2_F	8,345	*Dicistroviridae*	*Cripavirus*
HUA3_F	8,268	*Podoviridae*	*Kf1virus*
HUA2_F	8,121	*Podoviridae*	*T7virus*

AYA15_F	7,955	*Podoviridae*	*T7virus*
CAJ4_F	7,828	*Podoviridae*	*T7virus*
CAJ1_F	7,669	*Podoviridae*	*T7virus*
LMA5_F	7,603	*Podoviridae*	*T7virus*
API140_F	7,513	*Podoviridae*	*T7virus*
AMA2_F	6,826	*Coronaviridae*	*Alphacoronavirus*
HUA3_F	6,805	*Podoviridae*	*T7virus*
CUS8_F	6,774	*Picornaviridae*	*Parechovirus*
HUA4_F	6,748	*Coronaviridae*	*Alphacoronavirus*
LR3_F	6,683	*Hepeviridae*	
AYA14_F	6,647	*Hepeviridae*	
AYA11_F	6,646	*Hepeviridae*	
CUS8_F	6,493	*Podoviridae*	*T7virus*
HUA4_F	6,482	*Tymoviridae*	*Marafivirus*
AYA11_F	6,394	*Tymoviridae*	*Marafivirus*
LMA5_F	6,198	*Podoviridae*	*Kf1virus*
API141_F	6,130	*Podoviridae*	*T7virus*
HUA2_F	6,065	*Podoviridae*	*T7virus*
HUA1_F	5,964	*Podoviridae*	*T7virus*
AYA15_F	5,609	*Podoviridae*	*T7virus*
AMA2_F	5,590	*Coronaviridae*	*Alphacoronavirus*
AYA14_F	5,310	*Astroviridae*	
HUA4_F	5,160	*Tombusviridae*	*Alphacarmovirus*
CAJ4_F	5,110	*Tombusviridae*	*Pelarspovirus*
AYA12_F	5,091	*Tombusviridae*	*Pelarspovirus*
HUA2_SV	5,081	*Papillomaviridae*	
HUA2_F	5,077	*Podoviridae*	*T7virus*
API141_F	5,069	*Tombusviridae*	*Alphacarmovirus*
API1_F	5,067	*Tombusviridae*	*Alphacarmovirus*
CUS8_F	5,057	*Tombusviridae*	*Alphacarmovirus*
CAJ1_F	5,046	*Tombusviridae*	*Pelarspovirus*
HUA4_F	5,022	*Tombusviridae*	*Alphacarmovirus*
AYA11_F	5,020	*Tombusviridae*	*Alphacarmovirus*
HUA2_F	5,015	*Tombusviridae*	*Alphacarmovirus*
AYA1_F	5,005	*Tombusviridae*	*Pelarspovirus*
API17_F	4,992	*Tombusviridae*	*Alphacarmovirus*
API1_F	4,976	*Tombusviridae*	*Alphacarmovirus*
API17_F	4,974	*Tombusviridae*	
AMA2_F	4,957	*Tombusviridae*	*Alphacarmovirus*
API140_F	4,892	*Podoviridae*	*T7virus*
CAJ4_F	4,796	*Podoviridae*	*Kf1virus*
API17_F	4,788	*Tombusviridae*	*Alphacarmovirus*
CUS8_F	4,753	*Podoviridae*	*T7virus*
HUA3_F	4,727	*Podoviridae*	*T7virus*
AMA2_F	4,692	*Tombusviridae*	*Betacarmovirus*
LMA5_F	4,668	*Podoviridae*	*T7virus*

HUA4_F	4,630	*Tombusviridae*	
AYA11_F	4,613	*Tombusviridae*	*Betacarmovirus*
HUA4_F	4,543	*Podoviridae*	*T7virus*
LMA6_F	4,465	*Picornaviridae*	*Parechovirus*
CAJ2_F	4,462	*Podoviridae*	*T7virus*
LR2_F	4,431	*Tymoviridae*	*Marafivirus*
AYA15_F	4,338	*Podoviridae*	*T7virus*
API140_F	4,292	*Tymoviridae*	*Marafivirus*
CUS8_F	4,279	*Podoviridae*	*T7virus*
AYA1_F	4,234	*Podoviridae*	*T7virus*
HUA3_F	4,177	*Podoviridae*	
API1_F	4,170	*Podoviridae*	*T7virus*
API17_F	4,108	*Hepeviridae*	
API140_F	4,063	*Podoviridae*	*T7virus*
CAJ1_F	4,050	*Podoviridae*	*T7virus*
HUA3_F	4,023	*Podoviridae*	*T7virus*
LR2_F	3,933	*Tombusviridae*	*Macanavirus*
CUS8_F	3,907	*Podoviridae*	*T7virus*
AMA2_F	3,894	*Coronaviridae*	*Alphacoronavirus*
LMA6_SV	3,844	*Leviviridae*	*Levivirus*
HUA4_F	3,754	*Podoviridae*	*T7virus*
HUA3_F	3,703	*Podoviridae*	*T7virus*
AYA15_F	3,693	*Siphoviridae*	*Spbetavirus*
API141_F	3,615	*Podoviridae*	*T7virus*
API140_F	3,577	*Podoviridae*	*T7virus*
CUS8_F	3,569	*Podoviridae*	*T7virus*
AMA2_F	3,507	*Adenoviridae*	*Mastadenovirus*
CUS8_F	3,505	*Podoviridae*	*T7virus*
API141_F	3,468	*Podoviridae*	*T7virus*
API140_F	3,460	*Podoviridae*	*T7virus*
API140_F	3,430	*Podoviridae*	*T7virus*
CUS8_F	3,426	*Podoviridae*	*T7virus*
AMA2_F	3,376	*Adenoviridae*	*Mastadenovirus*
HUA1_F	3,336	*Reoviridae*	*Rotavirus*
CAJ1_F	3,326	*Podoviridae*	*T7virus*
LMA5_F	3,297	*Podoviridae*	*T7virus*
CUS8_F	3,291	*Podoviridae*	*T7virus*
API141_F	3,282	*Podoviridae*	*T7virus*
HUA3_F	3,194	*Podoviridae*	*T7virus*
API141_F	3,102	*Podoviridae*	*T7virus*
HUA4_F	3,097	*Tombusviridae*	*Alphacarmovirus*
HUA3_F	3,088	*Podoviridae*	*T7virus*
HUA2_F	3,060	*Podoviridae*	*T7virus*
HUA1_F	3,050	*Podoviridae*	*Kf1virus*
CAJ1_F	3,015	*Podoviridae*	*T7virus*
API141_F	3,000	*Tymoviridae*	

Table C 5 Family-level viral community composition differs between sample types.
A GLM-based approach was used to test for a multivariate effect of sample type on viral
community composition, while controlling for sequencing run and the number of raw reads.

Variable	Residual d.f.	d.f.	Dev†	*P*-value
Sample Type	**14**	**1**	**112.31**	**0.014**
Run	13	1	62.40	0.320
Raw Reads	12	1	88.47	0.223

†Deviance test statistic was calculated using a log likelihood ratio test

Table C 6 Viral sequences included in individual viral family phylogenetic analyses.
Sequence data were obtained from Genbank and host details were either obtained from
Genbank or the relevant literature.

Virus ID†	Viral family‡	Host§	Accession
CAA42749.1/HDV-2	*Hepatitis deltavirus*	Human	CAA42749.1
AAC55090.1/HDV-2	*Hepatitis deltavirus*	Human	AAC55090.1
BAD02974.1/HDV-2	*Hepatitis deltavirus*	Human	BAD02974.1
AAG26088.1/HDV-2	*Hepatitis deltavirus*	Human	AAG26088.1
CAC51365.1/HDV-2	*Hepatitis deltavirus*	Human	CAC51365.1
CAC51366.1/HDV-2	*Hepatitis deltavirus*	Human	CAC51366.1
CAE48184.1/HDV-7	*Hepatitis deltavirus*	Human	CAE48184.1
CAJ66090.1/HDV-5	*Hepatitis deltavirus*	Human	CAJ66090.1
CAJ66095.1/HDV-5	*Hepatitis deltavirus*	Human	CAJ66095.1
CAJ66092.1/HDV-5	*Hepatitis deltavirus*	Human	CAJ66092.1
CAE48186.1/HDV-6	*Hepatitis deltavirus*	Human	CAE48186.1
CAJ66093.1/HDV-6	*Hepatitis deltavirus*	Human	CAJ66093.1
CAJ66096.1/HDV-6	*Hepatitis deltavirus*	Human	CAJ66096.1
CAJ66091.1/HDV-8	*Hepatitis deltavirus*	Human	CAJ66091.1
CAJ66094.1/HDV-8	*Hepatitis deltavirus*	Human	CAJ66094.1
CAJ66097.1/HDV-7	*Hepatitis deltavirus*	Human	CAJ66097.1
AAC40831.1/HDV-4	*Hepatitis deltavirus*	Human	AAC40831.1
AAF22831.1/HDV-4	*Hepatitis deltavirus*	Human†	AAF22831.1
BAD02975.1/HDV-4	*Hepatitis deltavirus*	Human	BAD02975.1
BAD02973.1/HDV-4	*Hepatitis deltavirus*	Human	BAD02973.1
AAG40614.1/HDV-4	*Hepatitis deltavirus*	Human	AAG40614.1
BAC56856.1/HDV-4	*Hepatitis deltavirus*	Human	BAC56856.1
BAA00874.1/HDV-1	*Hepatitis deltavirus*	Human	BAA00874.1
AAA45723.1/WHDV	*Hepatitis deltavirus*	Woodchuck	AAA45723.1
CAC32838.1/WHDV	*Hepatitis deltavirus*	Woodchuck	CAC32838.1
AAB59753.1/HDV-1	*Hepatitis deltavirus*	Human	AAB59753.1
AAA45724.1/HDV-1	*Hepatitis deltavirus*	Human	AAA45724.1
CAA59509.1/HDV-1	*Hepatitis deltavirus*	Human	CAA59509.1
AAG26087.1/HDV-1	*Hepatitis deltavirus*	Human	AAG26087.1
AAB39885.1/HDV-1	*Hepatitis deltavirus*	Human	AAB39885.1
AAB02593.1/HDV-1	*Hepatitis deltavirus*	Human	AAB02593.1
AAU93913.1/HDV-1	*Hepatitis deltavirus*	Human	AAU93913.1
BAD02977.1/HDV-1	*Hepatitis deltavirus*	Human	BAD02977.1

AAB39884.1/HDV-1	*Hepatitis deltavirus*	Human	AAB39884.1
AAB02595.1/HDV-3	*Hepatitis deltavirus*	Human	AAB02595.1
BAB68379.1/HDV-3	*Hepatitis deltavirus*	Human	BAB68379.1
BAB68380.1/HDV-3	*Hepatitis deltavirus*	Human	BAB68380.1
BAB68381.1/HDV-3	*Hepatitis deltavirus*	Human	BAB68381.1
KJ562187/RhiFer/China	*Hepeviridae*	*Rhinolophus ferrumequinum*	KJ562187
JQ001748/MyoBec/Germany	*Hepeviridae*	*Myotis bechsteinii*	JQ001748
JQ001746/MyoDau/Germany	*Hepeviridae*	*Myotis daubentonii*	JQ001746
JQ001749/EptSer/Germany	*Hepeviridae*	*Eptesicus serotinus*	JQ001749
KX513953/MyoDav/China	*Hepeviridae*	*Myotis davidii*	KX513953
JQ071861/HipAba/Ghana	*Hepeviridae*	*Hipposideros abae*	JQ071861
JQ001745/VamCar/Panama	*Hepeviridae*	*Vampyrodes caraccioli*	JQ001745
EF206691/Chicken/USA	*Hepeviridae*	Chicken	EF206691
KC454286/Chicken/ SouthKorea	*Hepeviridae*	Chicken	KC454286
AM943647/Chicken/Australia	*Hepeviridae*	Chicken	AM943647
KF511797/Chicken/Taiwan	*Hepeviridae*	Chicken	KF511797
AM943646/Chicken/Hungary	*Hepeviridae*	Chicken	AM943646
GU954430/Chicken/China	*Hepeviridae*	Chicken	GU954430
AP003430/Human/Japan	*Hepeviridae*	Human	AP003430
JN564006/Human/USA	*Hepeviridae*	Human	JN564006
AB740232/Pig/Japan	*Hepeviridae*	Pig	AB740232
HQ389544/Human/USA	*Hepeviridae*	Human	HQ389544
FJ998008/Boar/Germany	*Hepeviridae*	Boar	FJ998008
AB291956/Human/Japan	*Hepeviridae*	Human	AB291956
AB073912/Pig/Japan	*Hepeviridae*	Pig	AB073912
KU513561/Human/Spain	*Hepeviridae*	Human	KU513561
AB248520/Human/Japan	*Hepeviridae*	Human	AB248520
AB248521/Pig/Japan	*Hepeviridae*	Pig	AB248521
AB248522/Pig/Japan	*Hepeviridae*	Pig	AB248522
AF455784/Pig/Kyrgyzstan	*Hepeviridae*	Pig	AF455784
EU723512/Pig/Spain	*Hepeviridae*	Pig	EU723512
FJ956757/Human/Germany	*Hepeviridae*	Human	FJ956757
EU360977/Pig/Sweden	*Hepeviridae*	Pig	EU360977
EU495148/Human/France	*Hepeviridae*	Human	EU495148
EU375463/Pig/Thailand	*Hepeviridae*	Pig	EU375463
FJ906895/Rabbit/China	*Hepeviridae*	Rabbit	FJ906895
FJ906896/Rabbit/China	*Hepeviridae*	Rabbit	FJ906896
JQ013793/Human/France	*Hepeviridae*	Human	JQ013793
GU937805/Rabbit/China	*Hepeviridae*	Rabbit	GU937805
AB740220/Rabbit/China	*Hepeviridae*	Rabbit	AB740220
AB573435/Boar/Japan	*Hepeviridae*	Boar	AB573435

AB856243/Boar/Japan	*Hepeviridae*	Boar	AB856243
AB161717/Human/Japan	*Hepeviridae*	Human	AB161717
AB097811/Pig/Japan	*Hepeviridae*	Pig	AB097811
AJ272108/Human/China	*Hepeviridae*	Human	AJ272108
EU366959/Pig/China	*Hepeviridae*	Pig	EU366959
GU119960/Pig/China	*Hepeviridae*	Pig	GU119960
AB220974/Human/Japan	*Hepeviridae*	Human	AB220974
AB602441/Boar/Japan	*Hepeviridae*	Boar	AB602441
M80581/Human/Pakistan	*Hepeviridae*	Human	M80581
DQ459342/Human/India	*Hepeviridae*	Human	DQ459342
L08816/Human/China	*Hepeviridae*	Human	L08816
AF459438/Human/India	*Hepeviridae*	Human	AF459438
AF076239/Human/India	*Hepeviridae*	Human	AF076239
M74506/Human/Mexico	*Hepeviridae*	Human	M74506
GQ504009/Rat/Germany	*Hepeviridae*	Rat	GQ504009
JX120573/Rat/Vietnam	*Hepeviridae*	Rat	JX120573
JN040433/Rat/Vietnam	*Hepeviridae*	Rat	JN040433
LC145325/Rat/Indonesia	*Hepeviridae*	Rat	LC145325
AB847306/Rat/Indonesia	*Hepeviridae*	Rat	AB847306
KM516906/Rat/USA	*Hepeviridae*	Rat	KM516906
GU345042/Rat/Germany	*Hepeviridae*	Rat	GU345042
GQ504010/Rat/Germany	*Hepeviridae*	Rat	GQ504010
GU345043/Rat/Germany	*Hepeviridae*	Rat	GU345043
JN998607/Ferret/Netherlands	*Hepeviridae*	Ferret	JN998607
KR905549/TreeShrew/China	*Hepeviridae*	Tree Shrew	KR905549
NC_015521/Trout/USA	*Hepeviridae*	Trout	NC_015521
AB091394/Human/Japan	*Hepeviridae*	Human	AB091394
AB222183/Boar/Japan	*Hepeviridae*	Boar	AB222183
AB189070/Boar/Japan	*Hepeviridae*	Boar	AB189070
AB189071/Deer/Japan	*Hepeviridae*	Deer	AB189071
AB189075/Human/Japan	*Hepeviridae*	Human	AB189075
AB443624/Pig/Japan	*Hepeviridae*	Pig	AB443624
AB089824/Human/Japan	*Hepeviridae*	Human	AB089824
AF082843/Pig/USA	*Hepeviridae*	Pig	AF082843
AF060669/Human/USA	*Hepeviridae*	Human	AF060669
AB591734/Mongoose/Japan	*Hepeviridae*	Mongoose	AB591734
AY115488/Pig/Canada	*Hepeviridae*	Pig	AY115488
AB290312/Pig/Mongolia	*Hepeviridae*	Pig	AB290312
FJ705359/Boar/Germany	*Hepeviridae*	Boar	FJ705359
AB291958/Human/Japan	*Hepeviridae*	Human	AB291958
AB097812/Human/Japan	*Hepeviridae*	Human	AB097812
AB480825/Human/Japan	*Hepeviridae*	Human	AB480825
AB521805/Human/Japan	*Hepeviridae*	Human	AB521805
AB602440/Boar/Japan	*Hepeviridae*	Boar	AB602440

D11092/Human/China	*Hepeviridae*	Human	D11092
AY230202/Human/Morocco	*Hepeviridae*	Human	AY230202
AY535004/Chicken/USA	*Hepeviridae*	Chicken	AY535004
JQ731783/PhyDis/Panama	*Coronaviridae*	*Phyllostomus discolor*	JQ731783
JQ731782/PhyDis/Panama	*Coronaviridae*	*Phyllostomus discolor*	JQ731782
JQ731789/CarPer/CR	*Coronaviridae*	*Carollia perspicillata*	JQ731789
JQ731790/CarPer/CR	*Coronaviridae*	*Carollia perspicillata*	JQ731790
JQ731793/CarPer/CR	*Coronaviridae*	*Carollia perspicillata*	JQ731793
JQ731796/CarPer/Brazil	*Coronaviridae*	*Carollia perspicillata*	JQ731796
JQ731794/CarPer/Brazil	*Coronaviridae*	*Carollia perspicillata*	JQ731794
JQ731785/ArtJam/Panama	*Coronaviridae*	*Artibeus jamaicensis*	JQ731785
JQ731787/ArtLit/Panama	*Coronaviridae*	*Artibeus lituratus*	JQ731787
KY502393/MyoEma/Luxembourg	*Coronaviridae*	*Myotis emarginatus*	KY502393
KY770855/RhiPus/China	*Coronaviridae*	*Rhinolophus pusillus*	KY770855
GU190236/RhiBla/Bulgaria	*Coronaviridae*	*Rhinolophus blasii*	GU190236
GU190232/RhiBla/Bulgaria	*Coronaviridae*	*Rhinolophus blasii*	GU190232
KU343197/RhiAff/China	*Coronaviridae*	*Rhinolophus affinis*	KU343197
KU343196/HipPom/China	*Coronaviridae*	*Hipposideros pomona*	KU343196
KU343195/HipPom/China	*Coronaviridae*	*Hipposideros pomona*	KU343195
HQ728486/ChaSp/Kenya	*Coronaviridae*	*Chaerephon sp*	HQ728486
KU343190/MinFul/China	*Coronaviridae*	*Miniopterus fuliginosus*	KU343190
EU420138/MinMag/China	*Coronaviridae*	*Miniopterus magnater*	EU420138
GU190240/MinSch/Bulgaria	*Coronaviridae*	*Miniopterus schreibersii*	GU190240
GU190243/MinSch/Bulgaria	*Coronaviridae*	*Miniopterus schreibersii*	GU190243
DQ249226/MinTri/China	*Coronaviridae*	*Miniopterus magnater*	DQ249226
KU343191/MinSch/China	*Coronaviridae*	*Miniopterus schreibersii*	KU343191
GU190241/MinSch/Bulgaria	*Coronaviridae*	*Miniopterus schreibersii*	GU190241
HQ728485/MinInf/Kenya	*Coronaviridae*	*Miniopterus inflatus*	HQ728485
GU190239/NycLei/Bulgaria	*Coronaviridae*	*Nyctalus leisleri*	GU190239
DQ249224/MyoRic/China	*Coronaviridae*	*Myotis ricketti*	DQ249224
GU190216/MyoDau/Germany	*Coronaviridae*	*Myotis daubentonii*	GU190216
DQ648858/ScoKuh/China	*Coronaviridae*	*Scotophilus kuhlii*	DQ648858
AF353511/SusScr/Belgium	*Coronaviridae*	Pig	AF353511
JQ731775/AnoGeo/CR	*Coronaviridae*	*Anoura geoffroyi*	JQ731775

JQ731784/ArtJam/Panama	*Coronaviridae*	*Artibeus jamaicensis*	JQ731784
AY567487/HomSap/ Netherlands	*Coronaviridae*	Human	AY567487
AY518894/HomSap/ Netherlands	*Coronaviridae*	Human	AY518894
HM245925/MusVis/USA	*Coronaviridae*	American Mink	HM245925
DQ010921/FelCat/USA	*Coronaviridae*	Cat	DQ010921
DQ811789/SusScr/USA	*Coronaviridae*	Pig	DQ811789
GU190237/RhiEur/Bulgaria	*Coronaviridae*	*Rhinolophus euryale*	GU190237
EF203064/RhiSin/China	*Coronaviridae*	*Rhinolophus sinicus*	EF203064
AF304460/HomSap/UK	*Coronaviridae*	Human	AF304460
EF065505/TylPac/China	*Coronaviridae*	*Tylonycteris pachypus*	EF065505
EF065509/PipAbr/China	*Coronaviridae*	*Pipistrellus abramus*	EF065509
JX869059/HomSap/ SaudiArabia	*Coronaviridae*	Human	JX869059
GU190215/RhiBla/Bulgaria	*Coronaviridae*	*Rhinolophus blasii*	GU190215
DQ022305/RhiSin/China	*Coronaviridae*	*Rhinolophus sinicus*	DQ022305
FJ710043/HipSp/Ghana	*Coronaviridae*	*Hipposideros sp*	FJ710043
EF065513/RouLes/China	*Coronaviridae*	*Rousettus lechenaulti*	EF065513
JQ731781/PtePar/CR	*Coronaviridae*	*Pteronotus parnellii*	JQ731781
JQ731779/PtePar/CR	*Coronaviridae*	*Pteronotus parnellii*	JQ731779
DQ415914/HomSap/China	*Coronaviridae*	Human	DQ415914
DQ113897/HumanRVH/ China	*Reoviridae*	Human	DQ113897
EF453355/HumanRVH/ Bangladesh	*Reoviridae*	Human	EF453355
AB576629/PorcineRVH/ Japan	*Reoviridae*	Pig	AB576629
KU254592/PorcineRVH/USA	*Reoviridae*	Pig	KU254592
KX362513/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362513
KX362537/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362537
KX362524/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362524
KX362548/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362548
KX362558/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362558
KT962027/PorcineRVH/ SouthAfrica	*Reoviridae*	Pig	KT962027
KU528592/BatRVH/ SouthKorea	*Reoviridae*	Unknown	KU528592
KX756624/BatRVJ/Serbia	*Reoviridae*	*Miniopterus schreibersii*	KX756624
NC_021590/ChickenRVG/ Germany	*Reoviridae*	Chicken	NC_021590
KM369903/DogRVI/Hungary	*Reoviridae*	Dog	KM369903

KM369892/DogRVI/Hungary	*Reoviridae*	Dog	KM369892
NC_021541/HumanRVB/ Bangladesh	*Reoviridae*	Human	NC_021541
EU490414/HumanRVB/India	*Reoviridae*	Human	EU490414
DQ113900/HumanRVH/ China	*Reoviridae*	Human	DQ113900
EF453357/HumanRVH/ Bangladesh	*Reoviridae*	Human	EF453357
AB576631/PorcineRVH/ Bangladesh	*Reoviridae*	Pig	AB576631
KU254588/PorcineRVH/USA	*Reoviridae*	Pig	KU254588
KT962029/PorcineRVH/ SouthAfrica	*Reoviridae*	Pig	KT962029
KX362515/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362515
KX362560/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362560
KX362550/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362550
KX362526/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362526
KX362539/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362539
KU528593/BatRVH/ SouthKorea	*Reoviridae*	Unknown	KU528593
KX756626/BatRVJ/Serbia	*Reoviridae*	*Miniopterus schreibersii*	KX756626
KY026786/FelineRVI/ Canada	*Reoviridae*	Cat	KY026786
NC_021581/ChickenRVG/ Germany	*Reoviridae*	Chicken	NC_021581
EF453358/HumanRVH/ Bangladesh	*Reoviridae*	Human	EF453358
DQ113899/HumanRVH/ China	*Reoviridae*	Human	DQ113899
KU528594/BatRVH/ SouthKorea	*Reoviridae*	Unknown	KU528594
AB576625/PorcineRVH/ Japan	*Reoviridae*	Pig	AB576625
KU254590/PorcineRVH/USA	*Reoviridae*	Pig	KU254590
KM359493/PorcineRVH/ Brazil	*Reoviridae*	Pig	KM359493
KM359488/PorcineRVH/ Brazil	*Reoviridae*	Pig	KM359488
KM359491/PorcineRVH/ Brazil	*Reoviridae*	Pig	KM359491
KX362540/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362540
KX362516/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362516
KT962030/PorcineRVH/ SouthAfrica	*Reoviridae*	Pig	KT962030
KX362561/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362561
KX362527/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362527
KX362551/PorcineRVH/ Vietnam	*Reoviridae*	Pig	KX362551
KX756627/BatRVJ/Serbia	*Reoviridae*	*Miniopterus schreibersii*	KX756627

NC_021543/HumanRVB/Bangladesh	*Reoviridae*	Human	NC_021543
AF184084/HumanRVB/India	*Reoviridae*	Human	AF184084
NC_021589/ChickenRVG/Germany	*Reoviridae*	Chicken	NC_021589
HQ595344/RhiSin/HKG	*Picornaviridae*	*Rhinolophus sinicus*	HQ595344
JQ814852/IaIo/China	*Picornaviridae*	*Ia io*	JQ814852
JQ916922/MyoDas/Germany	*Picornaviridae*	*Myotis dasycneme*	JQ916922
KJ641694/RhiSin/China	*Picornaviridae*	*Rhinolophus sinicus*	KJ641694
KJ641695/RhiSin/China	*Picornaviridae*	*Rhinolophus sinicus*	KJ641695
JQ916917/RhiBla/Bulgaria	*Picornaviridae*	*Rhinolophus blasii*	JQ916917
JQ916920/RhiEur/Bulgaria	*Picornaviridae*	*Rhinolophus euryale*	JQ916920
JQ916924/MinSch/Bulgaria	*Picornaviridae*	*Miniopterus schreibersii*	JQ916924
JQ916925/MinSch/Bulgaria	*Picornaviridae*	*Miniopterus schreibersii*	JQ916925
JQ916923/MinSch/Romania	*Picornaviridae*	*Miniopterus schreibersii*	JQ916923
HQ595341/MinSch/HKG	*Picornaviridae*	*Miniopterus schreibersii*	HQ595341
HQ595340/MinPus/HKG	*Picornaviridae*	*Miniopterus pusillus*	HQ595340
HQ595343/MinMag/HKG	*Picornaviridae*	*Miniopterus magnater*	HQ595343
HQ595342/MinMag/HKG	*Picornaviridae*	*Miniopterus magnater*	HQ595342
JN831356/CaninePicornavirus	*Picornaviridae*	*Dog*	JN831356
JQ916930/MinSch/Romania	*Picornaviridae*	*Miniopterus schreibersii*	JQ916930
JQ916929/NycNoc/Romania	*Picornaviridae*	*Nyctalus noctula*	JQ916929
NC_016156/FelinePicornavirus	*Picornaviridae*	Cat	NC_016156
NC_001612/EnterovirusA	*Picornaviridae*	Human	NC_001612
NC_001490/HumanRhinovirusB	*Picornaviridae*	Human	NC_001490
NC_003988/EnterovirusH	*Picornaviridae*	Simian	NC_003988
KJ754089/Echovirus	*Picornaviridae*	Human	KJ754089
KJ754080/Echovirus	*Picornaviridae*	Human	KJ754080
NC_001472/EnterovirusB	*Picornaviridae*	Human	NC_001472
AJ245863/SwineVesicularDiseaseVirus	*Picornaviridae*	Porcine	AJ245863
NC_010415/EnterovirusJ	*Picornaviridae*	Simian	NC_010415
NC_013695/EnterovirusJ	*Picornaviridae*	Simian	NC_013695
NC_010411/EnterovirusB	*Picornaviridae*	Human	NC_010411
NC_001430/EnterovirusD	*Picornaviridae*	Human	NC_001430
NC_001428/EnterovirusC	*Picornaviridae*	Human	NC_001428
NC_004441/EnterovirusG	*Picornaviridae*	Porcine	NC_004441
JQ916919/MyoBec/Germany	*Picornaviridae*	*Myotis bechsteinii*	JQ916919

NC_001859/EnterovirusE	*Picornaviridae*	Bovine	NC_001859
NC_009996/ HumanRhinovirusC	*Picornaviridae*	Human	NC_009996
NC_003987/SapelovirusA	*Picornaviridae*	Porcine	NC_003987
AY064708/SapelovirusB	*Picornaviridae*	Simian	AY064708
NC_006553/ AvianSapelovirus	*Picornaviridae*	Duck	NC_006553
NC_001617/ HumanRhinovirusA	*Picornaviridae*	Human	NC_001617
D00239/HumanRhinovirusA	*Picornaviridae*	Human	D00239
NC_015626/ PigeonPicornavirusB	*Picornaviridae*	Pigeon	NC_015626
NC_016403/ QuailPicornavirus	*Picornaviridae*	Quail	NC_016403
KJ641698/MinFul	*Picornaviridae*	*Miniopterus fuliginosus*	KJ641698
NC_022332/EelPicornavirus	*Picornaviridae*	Eel	NC_022332
L02971/ HumanParechovirus1	*Picornaviridae*	Human	L02971
AY158066/ HumanParechovirus1	*Picornaviridae*	Human	AY158066
LC318432/ HumanParechovirus1	*Picornaviridae*	Human	LC318432
EU024639/ HumanParechovirus1	*Picornaviridae*	Human	EU024639
EU024640/ HumanParechovirus1	*Picornaviridae*	Human	EU024640
EU024637/ HumanParechovirus1	*Picornaviridae*	Human	EU024637
EU024638/ HumanParechovirus1	*Picornaviridae*	Human	EU024638
AF327920/Ljunganvirus	*Picornaviridae*	Ljunganvirus	AF327920
KR045607/Ljunganvirus	*Picornaviridae*	Ljunganvirus	KR045607
KF006989/ FerretParechovirus	*Picornaviridae*	Ferret	KF006989
AB937989/ShrewCrohivirus	*Picornaviridae*	Shrew	AB937989
KX644937/BatCrohivirus	*Picornaviridae*	Eidolon helvum	KX644937
JX134222/ BluegillPicornavirus	*Picornaviridae*	Bluegill Fish	JX134222
KF183915/ FatheadMinnowPicornavirus	*Picornaviridae*	Fathead Minnow	KF183915
KF306267/CarpPicornavirus	*Picornaviridae*	Carp	KF306267
NC_025432/ChickenOrivirus	*Picornaviridae*	Chicken	NC_025432
JQ316470/SwinePasivirus1	*Picornaviridae*	Pig	JQ316470
KX774306/DesRotAdV2/ Guatemala	*Adenoviridae*	*Desmodus rotundus*	KX774306
KX774305/DesRotAdV2/ Guatemala	*Adenoviridae*	*Desmodus rotundus*	KX774305
KX774308/DesRotAdV2/ Guatemala	*Adenoviridae*	*Desmodus rotundus*	KX774308
KX774304/DesRotAdV2/ Guatemala	*Adenoviridae*	*Desmodus rotundus*	KX774304
KC110769/DesRot/Brazil	*Adenoviridae*	*Desmodus rotundus*	KC110769
KX774297/DesRotAdV1/ Guatemala	*Adenoviridae*	*Desmodus rotundus*	KX774297
KX774300/DesRotAdV1/	*Adenoviridae*	*Desmodus*	KX774300

Guatemala		*rotundus*	
KX774296/DesRotAdV1/ Guatemala	*Adenoviridae*	*Desmodus rotundus*	KX774296
KX774295/DesRotAdV1/ Guatemala	*Adenoviridae*	*Desmodus rotundus*	KX774295
KX774301/DesRotAdV1/ Guatemala	*Adenoviridae*	*Desmodus rotundus*	KX774301
JX065119/NycLas/Spain	*Adenoviridae*	*Nyctalus lasiopterus*	JX065119
KM043101/EptNil/Germany	*Adenoviridae*	*Eptesicus nilssoni*	KM043101
KM043094/PleAur/Germany	*Adenoviridae*	*Plecotus auritus*	KM043094
KM043091/PipPyg/Germany	*Adenoviridae*	*Pipistrellus pygmaeus*	KM043091
KM043096/PipPip/Germany	*Adenoviridae*	*Pipistrellus pipistrellus*	KM043096
KM043093/PipPyg/Germany	*Adenoviridae*	*Pipistrellus pygmaeus*	KM043093
KY009645/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009645
KY009640/MyoRic/China	*Adenoviridae*	*Myotis ricketti*	KY009640
KY009649/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009649
KY009648/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009648
KY009642/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009642
KY009644/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009644
KY009660/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009660
KY009635/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009635
KY009663/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009663
KY009638/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009638
KY009647/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009647
KY009659/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009659
KY009654/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009654
KY009661/MyoRic/China	*Adenoviridae*	*Myotis ricketti*	KY009661
KY009636/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009636
KY009662/MyoRic/China	*Adenoviridae*	*Myotis ricketti*	KY009662
GU226966/MyoRic/China	*Adenoviridae*	*Myotis ricketti*	GU226966
GU226960/MyoRic/China	*Adenoviridae*	*Myotis ricketti*	GU226960
GU226963/MyoRic/China	*Adenoviridae*	*Myotis ricketti*	GU226963
KY783853/MyoRic/Macau	*Adenoviridae*	*Myotis ricketti*	KY783853
JQ308809/MyoRic/China	*Adenoviridae*	*Myotis ricketti*	JQ308809
KM043106/MyoMyo/Hungary	*Adenoviridae*	*Myotis myotis*	KM043106
KM043085/EptSer/Hungary	*Adenoviridae*	*Eptesicus serotinus*	KM043085
JN167523/PleAus/Hungary	*Adenoviridae*	*Plecotus austriacus*	JN167523

JX065124/NycLei/Spain	*Adenoviridae*	*Nyctalus leisleri*	JX065124
KM043102/NycLei/Germany	*Adenoviridae*	*Nyctalus leisleri*	KM043102
KM043103/NycLei/Germany	*Adenoviridae*	*Nyctalus leisleri*	KM043103
JX065129/NycLei/Spain	*Adenoviridae*	*Nyctalus leisleri*	JX065129
JX065127/NycLei/Spain	*Adenoviridae*	*Nyctalus leisleri*	JX065127
KM043104/NycLei/Germany	*Adenoviridae*	*Nyctalus leisleri*	KM043104
KM043105/NycLei/Germany	*Adenoviridae*	*Nyctalus leisleri*	KM043105
JX065118/NycLas/Spain	*Adenoviridae*	*Nyctalus lasiopterus*	JX065118
JX065128/NycLas/Spain	*Adenoviridae*	*Nyctalus lasiopterus*	JX065128
JX065125/NycLas/Spain	*Adenoviridae*	*Nyctalus lasiopterus*	JX065125
KM043112/NycNoc/Germany	*Adenoviridae*	*Nyctalus noctula*	KM043112
GU198877/NycNoc/Hungary	*Adenoviridae*	*Nyctalus noctula*	GU198877
KM043097/PipPip/Germany	*Adenoviridae*	*Pipistrellus pipistrellus*	KM043097
KM043098/NycNoc/Germany	*Adenoviridae*	*Nyctalus noctula*	KM043098
JX065126/NycLas/Spain	*Adenoviridae*	*Nyctalus lasiopterus*	JX065126
KM043111/NycNoc/Germany	*Adenoviridae*	*Nyctalus noctula*	KM043111
KM043110/NycNoc/Germany	*Adenoviridae*	*Nyctalus noctula*	KM043110
JX065117/NycLas/Spain	*Adenoviridae*	*Nyctalus lasiopterus*	JX065117
JX065121/HypSav/Spain	*Adenoviridae*	*Hypsugo savii*	JX065121
JX065122/HypSav/Spain	*Adenoviridae*	*Hypsugo savii*	JX065122
KY009641/EptSer/China	*Adenoviridae*	*Eptesicus serotinus*	KY009641
KY009637/EptSer/China	*Adenoviridae*	*Eptesicus serotinus*	KY009637
GU226970/MyoRic/China	*Adenoviridae*	*Myotis ricketti*	GU226970
GU226953/MyoRic/China	*Adenoviridae*	*Myotis ricketti*	GU226953
GU226954/MyoRic/China	*Adenoviridae*	*Myotis ricketti*	GU226954
GU226961/MyoSp/China	*Adenoviridae*	*Myotis sp*	GU226961
GU226962/MyoSp/China	*Adenoviridae*	*Myotis sp*	GU226962
KX871230/CorRaf/USA	*Adenoviridae*	*Corynorhinus rafinesquii*	KX871230
AC_000020/CanineAdv2	*Adenoviridae*	Dog	AC_000020
Y07760/CanineAdv1	*Adenoviridae*	Dog	Y07760
KM043107/EptSer/Germany	*Adenoviridae*	*Eptesicus serotinus*	KM043107
JQ308807/IaIo/China	*Adenoviridae*	*Ia io*	JQ308807
JQ308808/MyoRic/China	*Adenoviridae*	*Myotis ricketti*	JQ308808
KM043086/MyoBly/Hungary	*Adenoviridae*	*Myotis blythii*	KM043086
KM043084/MyoEma/Hungary	*Adenoviridae*	*Myotis emarginatus*	KM043084
KY311900/MinMin/Kenya	*Adenoviridae*	*Miniopterus minor*	KY311900
JN252129/PipPip/Germany	*Adenoviridae*	*Pipistrellus pipistrellus*	JN252129
KM043109/PipPip/Germany	*Adenoviridae*	*Pipistrellus pipistrellus*	KM043109

KM043092/PipPyg/Germany	*Adenoviridae*	*Pipistrellus pygmaeus*	KM043092
KM043108/PipPip/Germany	*Adenoviridae*	*Pipistrellus pipistrellus*	KM043108
KM043090/PipPyg/Hungary	*Adenoviridae*	*Pipistrellus pygmaeus*	KM043090
JX065123/NycLas/Spain	*Adenoviridae*	*Nyctalus lasiopterus*	JX065123
KT698855/RhiSin/China	*Adenoviridae*	*Rhinolophus sinicus*	KT698855
KT698853/RhiSin/China	*Adenoviridae*	*Rhinolophus sinicus*	KT698853
KT698854/RhiSin/China	*Adenoviridae*	*Rhinolophus sinicus*	KT698854
JN254802/ ChimpanzeeAdVY25	*Adenoviridae*	Chimpanzee	JN254802
NC_003266/HumanAdVE	*Adenoviridae*	Human	NC_003266
AY530876/SimianAdV22	*Adenoviridae*	Simian	AY530876
FJ404771/HumanAdV22	*Adenoviridae*	Human	FJ404771
AC_000006/HumanAdVD	*Adenoviridae*	Human	AC_000006
AB448770/HumanAdV54	*Adenoviridae*	Human	AB448770
HQ913600/TitiMonkeyAdV	*Adenoviridae*	Titi Monkey	HQ913600
KC693021/SimianAdVB	*Adenoviridae*	Simian	KC693021
HQ241819/SimianAdV49	*Adenoviridae*	Simian	HQ241819
FJ025931/SimianAdV18	*Adenoviridae*	Simian	FJ025931
AY598782/SimianAdV3	*Adenoviridae*	Simian	AY598782
JQ776547/SimianAdV6	*Adenoviridae*	Simian	JQ776547
HQ241818/SimianAdV48	*Adenoviridae*	Simian	HQ241818
AC_000010/SimianAdV21	*Adenoviridae*	Simian	AC_000010
DQ086466/HumanAdV3	*Adenoviridae*	Human	DQ086466
AY598970/HumanAdVB	*Adenoviridae*	Human	AY598970
AY803294/HumanAdV14	*Adenoviridae*	Human	AY803294
L19443/HumanAdVF	*Adenoviridae*	Human	L19443
AF534906/HumanAdV1	*Adenoviridae*	Human	AF534906
NC_001405/HumanAdVC	*Adenoviridae*	Human	NC_001405
AY771780/SimianAdV1	*Adenoviridae*	Simian	AY771780
AC_000189/PorcineAdV3	*Adenoviridae*	Porcine	AC_000189
JN418926/EquineAdV1	*Adenoviridae*	Equine	JN418926
X73487/HumanAdV12	*Adenoviridae*	Human	X73487
NC_020485/SimianAdV20	*Adenoviridae*	Simian	NC_020485
KP238322/SkunkAdV	*Adenoviridae*	Skunk	KP238322
NC_002513/BovineAdV2	*Adenoviridae*	Bovine	NC_002513
AF289262/PorcineAdV5	*Adenoviridae*	Porcine	AF289262
KY311889/RouAeg/Kenya	*Adenoviridae*	*Rousettus aegyptiacus*	KY311889
KY311890/RouAeg/Kenya	*Adenoviridae*	*Rousettus aegyptiacus*	KY311890
KY311902/RouAeg/Kenya	*Adenoviridae*	*Rousettus aegyptiacus*	KY311902
KY311903/RouAeg/Kenya	*Adenoviridae*	*Rousettus aegyptiacus*	KY311903

KY311908/RouAeg/Kenya	*Adenoviridae*	*Rousettus aegyptiacus*	KY311908
KY311886/RouAeg/Kenya	*Adenoviridae*	*Rousettus aegyptiacus*	KY311886
KY311909/RouAeg/Kenya	*Adenoviridae*	*Rousettus aegyptiacus*	KY311909
KY311901/OtoMar/Kenya	*Adenoviridae*	*Otomops martiensseni*	KY311901
KY311888/HipCaf/Kenya	*Adenoviridae*	*Hipposideros caffer*	KY311888
KY311893/RouAeg/Kenya	*Adenoviridae*	*Rousettus aegyptiacus*	KY311893
KY311885/RouAeg/Kenya	*Adenoviridae*	*Rousettus aegyptiacus*	KY311885
KC692417/PteGig/ Bangladesh	*Adenoviridae*	*Pteropus giganteus*	KC692417
KY311897/EidHel/Kenya	*Adenoviridae*	*Eidolon helvum*	KY311897
KY311899/ColAfr/Kenya	*Adenoviridae*	*Coleura afra*	KY311899
KY311896/EidHel/Kenya	*Adenoviridae*	*Eidolon helvum*	KY311896
KY311895/EidHel/Kenya	*Adenoviridae*	*Eidolon helvum*	KY311895
KY311887/RouAeg/Kenya	*Adenoviridae*	*Rousettus aegyptiacus*	KY311887
KY311891/RouAeg/Kenya	*Adenoviridae*	*Rousettus aegyptiacus*	KY311891
KY311905/RhiFum/Kenya	*Adenoviridae*	*Rhinolophus fumigatus*	KY311905
KY311910/CarCor/Kenya	*Adenoviridae*	*Cardioderma cor*	KY311910
KY311904/RhiFum/Kenya	*Adenoviridae*	*Rhinolophus fumigatus*	KY311904
KC692420/PteGig/ Bangladesh	*Adenoviridae*	*Pteropus giganteus*	KC692420
GU226957/ScoKuh/China	*Adenoviridae*	*Scotophilus kuhlii*	GU226957
GU226952/ScoKuh/China	*Adenoviridae*	*Scotophilus kuhlii*	GU226952
GU226969/ScoKuh/China	*Adenoviridae*	*Scotophilus kuhlii*	GU226969
GU226967/ScoKuh/China	*Adenoviridae*	*Scotophilus kuhlii*	GU226967
GU226968/ScoKuh/China	*Adenoviridae*	*Scotophilus kuhlii*	GU226968
JQ308810/IaIo/China	*Adenoviridae*	*Ia io*	JQ308810
GU226951/MyoHor/China	*Adenoviridae*	*Myotis horsfieldii*	GU226951
KM043083/MyoDas/Hungary	*Adenoviridae*	*Myotis dasycneme*	KM043083
GU226964/HipArm/China	*Adenoviridae*	*Hipposideros armiger*	GU226964
KY783852/HipArm/Macau	*Adenoviridae*	*Hipposideros armiger*	KY783852
BD269513/BovineAdV1	*Adenoviridae*	Bovine	BD269513
AF030154/BovineAdV3	*Adenoviridae*	Bovine	AF030154
AF258784/TreeShrewAdV1	*Adenoviridae*	Tree Shrew	AF258784
KM043099/VesMur/Germany	*Adenoviridae*	*Vespertilio murinus*	KM043099
KM043100/PipKuh/Germany	*Adenoviridae*	*Pipistrellus kuhlii*	KM043100
KM043095/PipNat/Germany	*Adenoviridae*	*Pipistrellus nathusii*	KM043095

KM043087/NycNoc/Germany	*Adenoviridae*	*Nyctalus noctula*	KM043087
KY009652/MyoPeq/China	*Adenoviridae*	*Myotis pequinius*	KY009652
KY009653/MyoPeq/China	*Adenoviridae*	*Myotis pequinius*	KY009653
KY009643/MyoPeq/China	*Adenoviridae*	*Myotis pequinius*	KY009643
KY009658/MyoRic/China	*Adenoviridae*	*Myotis ricketti*	KY009658
KM043079/RhiEur/Hungary	*Adenoviridae*	*Rhinolophus euryale*	KM043079
KM043080/RhiEur/Hungary	*Adenoviridae*	*Rhinolophus euryale*	KM043080
KM043081/RhiEur/Hungary	*Adenoviridae*	*Rhinolophus euryale*	KM043081
KT698852/MinSch/China	*Adenoviridae*	*Miniopterus schreibersi*	KT698852
AB303301/PteDas/Japan	*Adenoviridae*	*Pteropus dasumallus*	AB303301
KC692430/PteGig/ Bangladesh	*Adenoviridae*	*Pteropus giganteus*	KC692430
KC692429/PteGig/ Bangladesh	*Adenoviridae*	*Pteropus giganteus*	KC692429
HQ529709/RouLes/India	*Adenoviridae*	*Rousettus leschenaultii*	HQ529709
KX961095/RouLes/China	*Adenoviridae*	*Rousettus leschenaultii*	KX961095
KX961096/RouLes/China	*Adenoviridae*	*Rousettus leschenaultii*	KX961096
KC692426/PteGig/ Bangladesh	*Adenoviridae*	*Pteropus giganteus*	KC692426
KC692427/PteGig/ Bangladesh	*Adenoviridae*	*Pteropus giganteus*	KC692427
KC692425/PteGig/ Bangladesh	*Adenoviridae*	*Pteropus giganteus*	KC692425
KC692423/PteGig/ Bangladesh	*Adenoviridae*	*Pteropus giganteus*	KC692423
KC692424/PteGig/ Bangladesh	*Adenoviridae*	*Pteropus giganteus*	KC692424
KC692428/PteGig/ Bangladesh	*Adenoviridae*	*Pteropus giganteus*	KC692428
KJ563221/SeaLionAdV1	*Adenoviridae*	Sea Lion	KJ563221
KC692422/PteGig/ Bangladesh	*Adenoviridae*	*Pteropus giganteus*	KC692422
KC692421/PteGig/ Bangladesh	*Adenoviridae*	*Pteropus giganteus*	KC692421
KY783854/PipAbr/Macau	*Adenoviridae*	*Pipistrellus abramus*	KY783854
KY311907/RouAeg/Kenya	*Adenoviridae*	*Rousettus aegyptiacus*	KY311907
KC692419/PteGig/ Bangladesh	*Adenoviridae*	*Pteropus giganteus*	KC692419
JN167522/RhiFer/Hungary	*Adenoviridae*	*Rhinolophus ferrumequinum*	JN167522
HM049560/MurineAdV2	*Adenoviridae*	Mouse	HM049560
AC_000012/MurineAdVA	*Adenoviridae*	Mouse	AC_000012
EU835513/MurineAdV3	*Adenoviridae*	Mouse	EU835513
GU226956/MyoHor/China	*Adenoviridae*	*Myotis horsfieldii*	GU226956
KY311894/EidHel/Kenya	*Adenoviridae*	*Eidolon helvum*	KY311894
KY009639/MyoFim/China	*Adenoviridae*	*Myotis fimbriatus*	KY009639

Virus ID†	Viral family‡	Host species§	Accession
JQ814855/RhiAff/ BatFoamyVirus	*Retroviridae*	*Rhinolophus affinis*	JQ814855
JQ867462/ SimianFoamyVirus	*Retroviridae*	Simian	JQ867462
JQ867463/ SimianFoamyVirus	*Retroviridae*	Simian	JQ867463
KX087159/ HumanFoamyVirus	*Retroviridae*	Human	KX087159
Y07725/HumanFoamyVirus	*Retroviridae*	Human	Y07725
AJ544579/ SimianFoamyVirus	*Retroviridae*	Simian	AJ544579
NC_010819/ MacaqueSimianFoamyVirus	*Retroviridae*	Simian	NC_010819
JN801175/ MacaqueSimianFoamyVirus	*Retroviridae*	Simian	JN801175
LC094267/ SimianFoamyVirus	*Retroviridae*	Simian	LC094267
NC_010820/ SimianFoamyVirus	*Retroviridae*	Simian	NC_010820
JQ867466/ SimianFoamyVirus	*Retroviridae*	Simian	JQ867466
EU010385/ SpiderMonkeyFoamyVirus	*Retroviridae*	Spider Monkey	EU010385
GU356395/ MarmosetFoamyVirus	*Retroviridae*	Marmoset	GU356395
KP143760/ SimianFoamyVirus	*Retroviridae*	Simian	KP143760
GU356394/ SquirrelMonkeyFoamyVirus	*Retroviridae*	Squirrel Monkey	GU356394
NC_001871/ FelineFoamyVirus	*Retroviridae*	Feline	NC_001871
Y08851/FelineFoamyVirus	*Retroviridae*	Feline	Y08851
AF201902/ EquineFoamyVirus	*Retroviridae*	Equine	AF201902
NC_001831/ BovineFoamyVirus	*Retroviridae*	Bovine	NC_001831

†Virus ID is the name of the virus as depicted in the phylogeny
‡Viral family analysis in which sequence was included
§Host species according to Genbank or the relevant literature. Bat host species are represented by scientific name (as possible) while non-bat host species are represented by common name (e.g. Pig) or description (e.g. Porcine)

C2: Figures

Figure C 1 Heatmap of viral reads in multi colony pools.
Data are shown at the family level where similar rows (viral families detected together) and columns (pools containing viral families in common) are clustered using Ward's method.
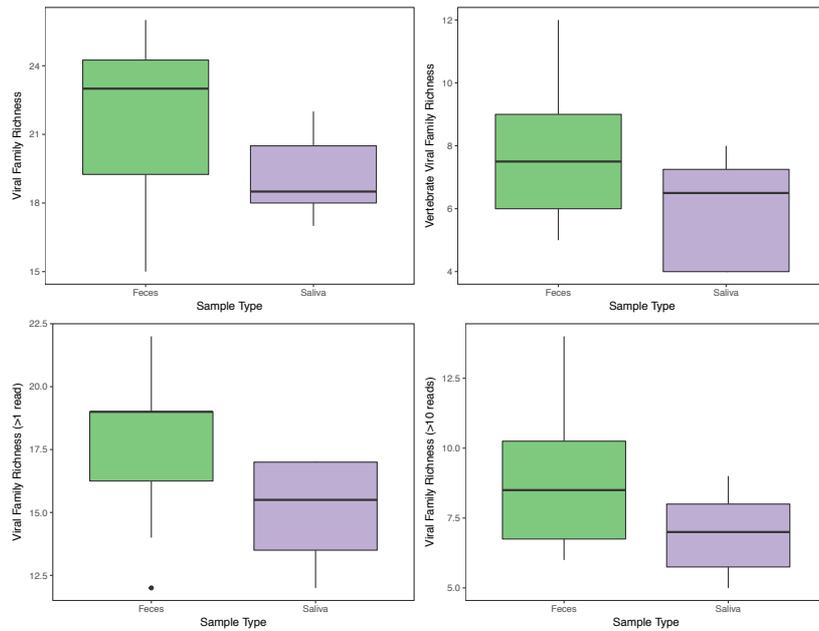
Figure C 2 Viral family richness comparisons between sample types.
Fecal (green) and saliva (purple) samples are compared using different data subsets. Data subsets were filtered by (clockwise from top left) all viral families, vertebrate-infecting families, >10 reads, and >1 read. Bold line shows the median, and upper and lower hinges show the first and third quartiles. Whiskers extend from the hinge to 1.5 * the inter-quartile range, and outlying points are shown individually.
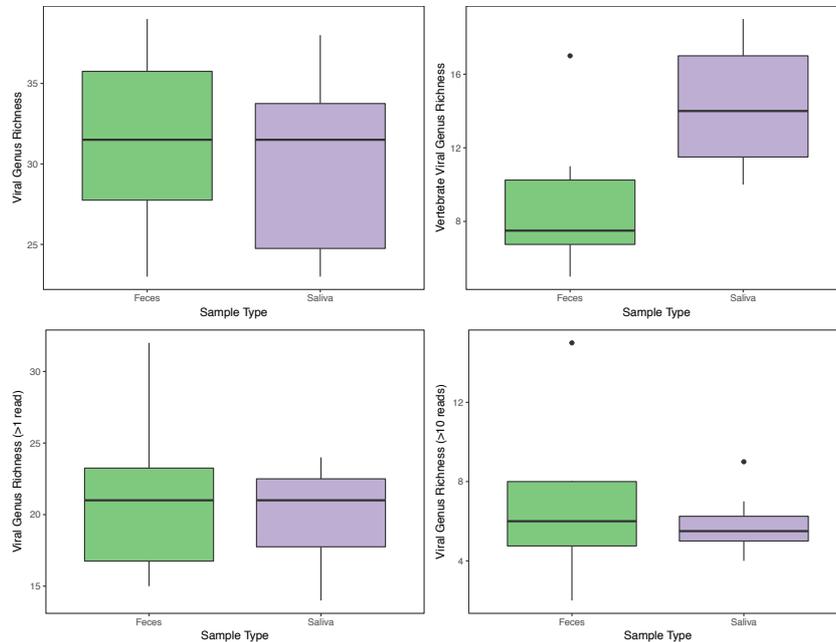
Figure C 3 Viral genus richness comparisons between sample types.
Fecal (green) and saliva (purple) samples are compared using different data subsets. Data subsets were filtered by (clockwise from top left) all viral genera, vertebrate-infecting genera, >10 reads, and >1 read. Bold line shows the median, and upper and lower hinges show the first and third quartiles. Whiskers extend from the hinge to 1.5 * the inter-quartile range, and outlying points are shown individually.
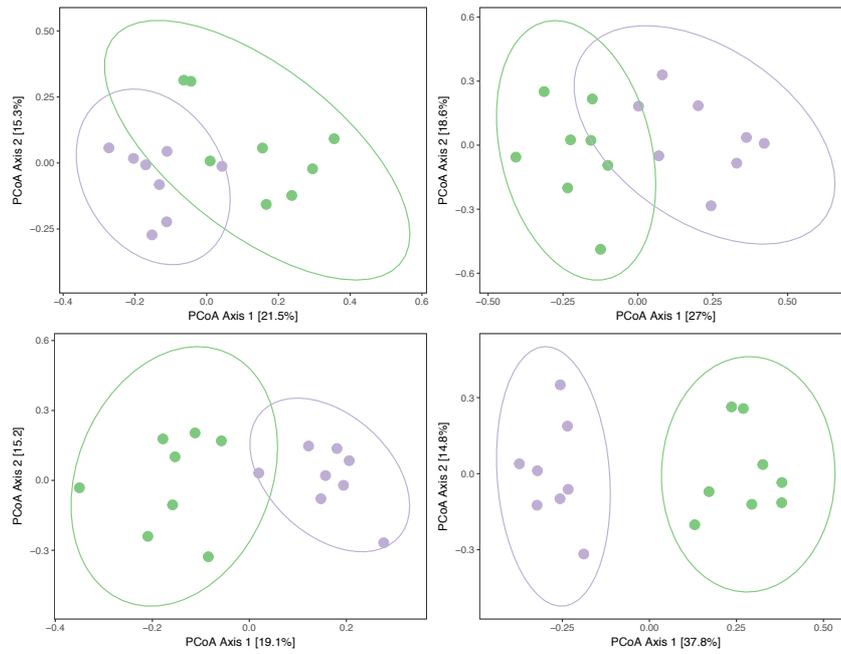
Figure C 4 Viral family community comparison.
Fecal (green) and saliva (purple) samples are plotted using principal coordinate analysis (PCoA) with different levels of filtering. Filtering included (clockwise from top left) all viral families, vertebrate-infecting families, >10 reads, and >1 read. The first two axes are plotted, and axis labels show the percent of variation explained by each, with 95% confidence ellipses plotted for each group
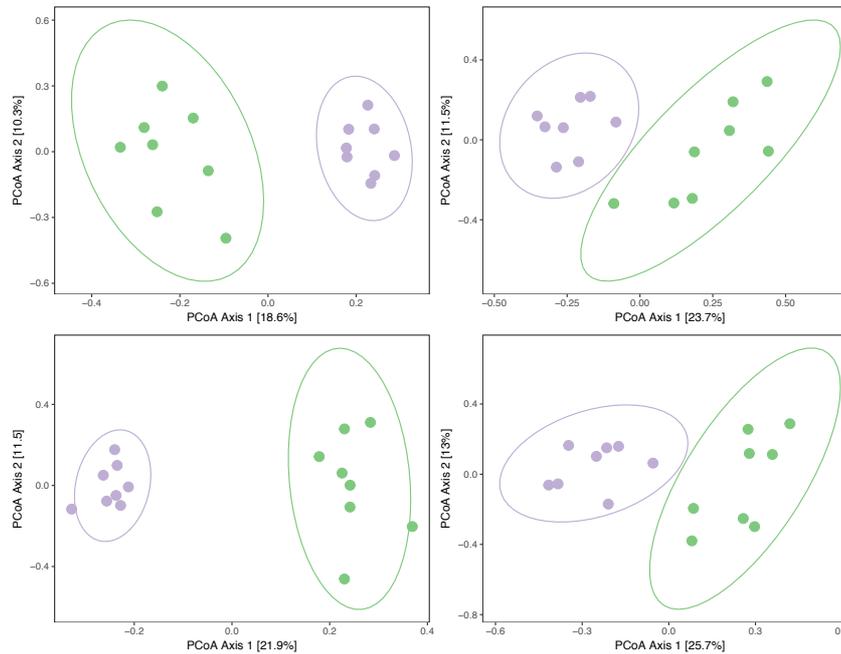
Figure C 5 Viral genus community composition comparison.
Fecal (green) and saliva (purple) samples are plotted using principal coordinate analysis (PCoA) with different levels of filtering. Filtering included (clockwise from top left) all viral families, vertebrate-infecting families, >10 reads, and >1 read. The first two axes are plotted, and axis labels show the percent of variation explained by each with 95% confidence ellipses plotted for each group

Appendix D: Supporting Material Chapter 4

D1: Tables

Table D 1 Collection details for individual bat samples in metagenomic pools. Bat IDs reflect individuals from which nucleic acid extracts were included in single colony saliva and fecal pools for each site. Month and year of sampling dates are shown for individual bat samples.

Saliva pool	Bat ID†	Month	Year	Fecal pool	Bat ID†	Month	Year
MSV5.1 AMA2	D234	8	2016	MH5.1 AMA2	D203	10	2015
	D235	8	2016		D94	10	2015
	D96	8	2015		D95	10	2015
	D236	8	2016		**D96**	8	2015
	D237	8	2016		D97	10	2015
	8400	12	2016		D98	10	2015
	8401	12	2016		D99	10	2015
	8402	12	2016		SP3	10	2015
	8403	12	2016		**D237**	8	2016
	8405	12	2016		**D235**	8	2016
MSV5.2 AMA7	**D16**	7	2013	MH5.2 AMA7	**D16**	7	2013
	D2	7	2013		D17	7	2013
	D3	7	2013		**D20**	8	2013
	D13	7	2013		D23	8	2013
	D18	7	2013		D25	8	2013
	D20	8	2013		**D5**	7	2013
	D24	8	2013		D6	7	2013
	D26	8	2013		D14	8	2013
	D5	7	2013		D10	7	2013
	D7	7	2013		D12	7	2013
MSV5.3 API1	**5942**	6	2016	MH5.3 API1	**5942**	6	2016
	9260	6	2016		**9260**	6	2016
	9261	6	2016		**9261**	6	2016
	9262	6	2016		**9262**	6	2016
	9263	6	2016		**9263**	6	2016
	9274	6	2016		9266	6	2016
	9275	6	2016		9267	6	2016
	9276	6	2016		9268	6	2016
	9277	6	2016		7603	6	2016
	9278	6	2016		9269	6	2016
MSV5.4 API17	**8325**	7	2016	MH5.4 API17	**8325**	7	2016
	7959	7	2016		**7959**	7	2016
	8326	7	2016		**8326**	7	2016
	7986	7	2016		**7986**	7	2016
	8327	7	2016		**8327**	7	2016
	8328	7	2016		**8328**	7	2016

	7980	7	2016		**7980**	7	2016
	8329	7	2016		**8329**	7	2016
	8330	7	2016		**8330**	7	2016
	8331	7	2016		**8331**	7	2016
MSV5.5 API140	**9237**	5	2016	MH5.5 API140	**9237**	5	2016
	9238	5	2016		**9238**	5	2016
	7132	5	2016		**7132**	5	2016
	7141	5	2016		**7141**	5	2016
	9239	5	2016		**9239**	5	2016
	9240	5	2016		**9240**	5	2016
	9241	5	2016		**9241**	5	2016
	9242	5	2016		**9242**	5	2016
	9243	5	2016		9245	5	2016
	9244	5	2016		**9244**	5	2016
MSV5.6 API141	**9126**	5	2016	MH5.6 API141	**9126**	5	2016
	6849	5	2016		**6849**	5	2016
	9128	5	2016		**9128**	5	2016
	6877	5	2016		**6877**	5	2016
	9127	5	2016		**9127**	5	2016
	9130	5	2016		**9130**	5	2016
	9129	5	2016		**9129**	5	2016
	9131	5	2016		**9131**	5	2016
	9132	5	2016		**9132**	5	2016
	9133	5	2016		**9133**	5	2016
MSV5.7 AYA1	**7184**	7	2016	MH5.7 AYA1	8228	7	2016
	8230	7	2016		7188	7	2016
	8231	7	2016		7181	7	2016
	6933	7	2016		7802	7	2016
	6934	7	2016		8229	7	2016
	8232	7	2016		**7184**	7	2016
	8233	7	2016		**8230**	7	2016
	8234	7	2016		**8231**	7	2016
	8235	7	2016		**6933**	7	2016
	8237	7	2016		**6934**	7	2016
MSV5.8 AYA7	**7064**	7	2016	MH5.8 AYA7	**7064**	7	2016
	8286	7	2016		**8286**	7	2016
	8287	7	2016		**8287**	7	2016
	8288	7	2016		**8288**	7	2016
	7047	7	2016		**7047**	7	2016
	8294	7	2016		8289	7	2016
	8290	7	2016		**8290**	7	2016
	8291	7	2016		**8291**	7	2016
	8292	7	2016		**8292**	7	2016
	8293	7	2016		**8293**	7	2016
MSV5.9 AYA11	**8201**	7	2016	MH5.9 AYA11	**8201**	7	2016
	8202	7	2016		**8202**	7	2016

	8203	7	2016		**8203**	7	2016
	7082	7	2016		**7082**	7	2016
	8204	7	2016		**8204**	7	2016
	8217	7	2016		7864	7	2016
	8218	7	2016		**8205**	7	2016
	8206	7	2016		**8206**	7	2016
	8205	7	2016		8207	7	2016
	8221	7	2016		8208	7	2016
MSV5.10 AYA12	**8222**	7	2016	MH5.10 AYA12	**8222**	7	2016
	7169	7	2016		**7169**	7	2016
	8223	7	2016		**8223**	7	2016
	8224	7	2016		**8224**	7	2016
	8225	7	2016		**8225**	7	2016
	8226	7	2016		**8226**	7	2016
	8227	7	2016		**8227**	7	2016
MSV5.11 AYA14	**8297**	7	2016	MH5.11 AYA14	**8297**	7	2016
	8298	7	2016		**8298**	7	2016
	8299	7	2016		**8299**	7	2016
	8300	7	2016		**8300**	7	2016
	8301	7	2016		**8301**	7	2016
	8320	7	2016		8302	7	2016
	8321	7	2016		8303	7	2016
	8322	7	2016		8304	7	2016
	8323	7	2016		8305	7	2016
	7774	7	2016		8306	7	2016
MSV5.12 AYA15	**9286**	6	2016	MH5.12 AYA15	**9286**	6	2016
	9284	6	2016		**9284**	6	2016
	9285	6	2016		**9285**	6	2016
	9287	6	2016		**9287**	6	2016
	9288	6	2016		**9288**	6	2016
	9289	6	2016		**9289**	6	2016
	9290	6	2016		**9290**	6	2016
	9291	6	2016		**9291**	6	2016
	9292	6	2016		**9292**	6	2016
	9293	6	2016		**9293**	6	2016
MSV5.13 CAJ1	**8080**	11	2016	MH5.13 CAJ1	**8080**	11	2016
	8081	11	2016		**8081**	11	2016
	8082	11	2016		**8082**	11	2016
	8083	11	2016		**8083**	11	2016
	4156	11	2016		**4156**	11	2016
	8084	11	2016		**8084**	11	2016
	8085	11	2016		**8085**	11	2016
	8086	11	2016		**8086**	11	2016
	8087	11	2016		**8087**	11	2016
	8088	11	2016		**8088**	11	2016
MSV5.14	**8096**	12	2016	MH5.14	**8096**	12	2016

CAJ2	**8097**	12	2016	CAJ2	**8097**	12	2016
	8098	12	2016		**8098**	12	2016
	8099	12	2016		**8099**	12	2016
	8100	12	2016		**8100**	12	2016
	8109	12	2016		8101	12	2016
	8110	12	2016		8102	12	2016
	8111	12	2016		8103	12	2016
	8112	12	2016		8104	12	2016
	4693	12	2016		6181	12	2016
MSV5.15 CAJ4	8361	12	2016	MH5.15 CAJ4	D83	10	2015
	8362	12	2016		D84	10	2015
	8363	12	2016		D85	10	2015
	8364	12	2016		D86	10	2015
	8365	12	2016		D87	10	2015
	8366	12	2016		D88	10	2015
	8367	12	2016		D89	10	2015
	8368	12	2016		D90	10	2015
	8369	12	2016		D91	10	2015
	8370	12	2016		D92	10	2015
MSV5.16 CUS8	**9216**	5	2016	MH5.16 CUS8	**9216**	5	2016
	9217	5	2016		**9217**	5	2016
	7503	5	2016		**7503**	5	2016
	9218	5	2016		**9218**	5	2016
	9219	5	2016		**9219**	5	2016
	7995	5	2016		**7995**	5	2016
	9220	5	2016		**9220**	5	2016
	9221	5	2016		**9221**	5	2016
	9222	5	2016		**9222**	5	2016
	9223	5	2016		**9223**	5	2016
MSV5.17 HUA1	**8009**	10	2016	MH5.17 HUA1	**8009**	10	2016
	8010	10	2016		**8010**	10	2016
	8011	10	2016		**8011**	10	2016
	8012	10	2016		**8012**	10	2016
	8013	10	2016		**8013**	10	2016
	8014	10	2016		**8014**	10	2016
	8015	10	2016		**8015**	10	2016
	8016	10	2016		**8016**	10	2016
	8017	10	2016		**8017**	10	2016
	8018	10	2016		**8018**	10	2016
MSV5.18 HUA2	**8332**	10	2016	MH5.18 HUA2	**8332**	10	2016
	8333	10	2016		**8333**	10	2016
	6750	10	2016		**6750**	10	2016
	8334	10	2016		**8334**	10	2016
	8335	10	2016		**8335**	10	2016
	7707	10	2016		**7707**	10	2016
	8336	10	2016		**8336**	10	2016

	8337	10	2016		**8337**	10	2016
	8338	10	2016		**8338**	10	2016
	8339	10	2016		**8339**	10	2016
MSV5.19 HUA3	**8003**	10	2016	MH5.19 HUA3	**8003**	10	2016
	8004	10	2016		**8004**	10	2016
	8005	10	2016		**8005**	10	2016
	8006	10	2016		**8006**	10	2016
	8007	10	2016		**8007**	10	2016
	7738	10	2016		**7738**	10	2016
	9050	10	2016		**9050**	10	2016
	9022	10	2016		**9022**	10	2016
	8008	10	2016		**8008**	10	2016
MSV5.20 HUA4	**8022**	10	2016	MH5.20 HUA4	**8022**	10	2016
	8023	10	2016		**8023**	10	2016
	8024	10	2016		**8024**	10	2016
	8025	10	2016		**8025**	10	2016
	8026	10	2016		**8026**	10	2016
	8027	10	2016		**8027**	10	2016
	8028	10	2016		**8028**	10	2016
	8029	10	2016		**8029**	10	2016
	8030	10	2016		**8030**	10	2016
	8031	10	2016		**8031**	10	2016
MSV5.21 LMA5	**9110**	4	2016	MH5.21 LMA5	**9110**	4	2016
	7681	4	2016		**7681**	4	2016
	3859	4	2016		**3859**	4	2016
	9111	4	2016		**9111**	4	2016
	9112	4	2016		**9112**	4	2016
MSV5.22 LMA6	**9083**	4	2016	MH5.22 LMA6	**9083**	4	2016
	2926	4	2016		**2926**	4	2016
	4557	4	2016		**4557**	4	2016
	9084	4	2016		**9084**	4	2016
	9085	4	2016		**9085**	4	2016
	8072	11	2016		8060	11	2016
	5367	11	2016		8061	11	2016
	2941	11	2016		8062	11	2016
	8074	11	2016		8063	11	2016
	8075	11	2016		8064	11	2016
MSV5.23 LR2	**D256**	9	2016	MH5.23 LR2	D39	9	2015
	D257	9	2016		D57	9	2015
	D258	9	2016		D74	9	2015
	D259	9	2016		D75	9	2015
	D260	9	2016		D76	9	2015
	D261	9	2016		D77	9	2015
	D262	9	2016		**D59**	5	2016
	D0059	5	2016		**D256**	9	2016
					D260	9	2016

					D261	9	2016
MSV5.24 LR3	**D249**	9	2016	MH5.24 LR3	**D242**	9	2016
	D248	9	2016		**D244**	9	2016
	D243	9	2016		**D243**	9	2016
	D246	9	2016		**D246**	9	2016
	D245	9	2016		**D248**	9	2016
	D244	9	2016		**D247**	9	2016
	D247	9	2016		**D249**	9	2016
	D250	9	2016		**D245**	9	2016
	D242	9	2016		**D250**	9	2016
	D216	5	2015		D73	9	2015

†Bat IDs in bold show individual IDs for those that were included in pools for both fecal and saliva swabs.

Table D 2 Microsatellite error rate estimates, null alleles and F_{IS} estimates per locus. Conversion shows the number of base pairs different between microsatellite loci scored in different labs. Error rate estimates across and between labs were calculated using PEDANT. Number of populations showing evidence of null alleles were based on MicroChecker. Average null allele frequency across populations were based on FreeNA.

Locus	Convert	Across labs		Within lab		Null pops	Null freq	F_{IS}†
		Allelic dropout	False alleles	Allelic dropout	False alleles			
DeroB03	-1	0	0	0.000001	0	1	0.028	0.038
DeroB10	--	0	0.05	0	0	2	0.024	0.046
DeroB11	--	--	--	--	--	0	0.001	-0.048
DeroC12	2	0.26	0.24	0.000002	0	0	0.014	-0.026
DeroD06	2	0.7	0.05	--	--	4	0.034	**0.405**
DeroC07	2	0	0	0.000001	0	1	0.022	0.023
DeroD12	1	0	0	0	0	1	0.013	-0.01
DeroG12	1	0	0	0.000001	0.000001	3	0.036	0.056
Dero H02	2	0	0	0	0	13	0.105	**0.509**

†Significant departure from Hardy-Weinberg equilibrium indicated in bold; p-values were adjusted using the Bonferroni correction (adjusted α=0.006)

Table D 3 Microsatellite per population statistics calculated from the 6 loci dataset. Diversity indices were calculated using adegenet in R and FSTAT.

Site	Allelic richness (A_R)	Expected heterozygosity (H_E)	Observed heterozygosity (H_O)	F_{IS}
AMA7	25.35	0.63	0.63	0.022
AMA2	26.94	0.67	0.63	0.076
API1	19.15	0.47	0.52	-0.103
API140	18.24	0.47	0.47	0.023
API141	18.97	0.47	0.51	-0.058
API17	16.52	0.43	0.41	0.073
AYA1	15.77	0.37	0.41	-0.1
AYA11	17.42	0.42	0.45	-0.039
AYA12	16.93	0.40	0.38	0.063
AYA14	15.06	0.34	0.38	-0.083
AYA15	16.41	0.40	0.40	0.005
AYA7	13.55	0.26	0.22	0.167
CAJ1	21.82	0.51	0.47	0.101
CAJ2	19.99	0.48	0.44	0.129
CAJ4	25.38	0.59	0.56	0.059
CUS8	16.77	0.38	0.37	0.04
HUA1	23.49	0.59	0.56	0.075
HUA2	19.09	0.52	0.41	0.229
HUA3	18.38	0.49	0.51	0.017
HUA4	23.68	0.61	0.65	-0.041
LMA5	19.86	0.55	0.52	0.067
LMA6	19.49	0.55	0.54	0.035
LR2	25.51	0.64	0.66	0.005
LR3	28.30	0.65	0.64	0.04

Table D 4 Multivariate PERMANOVA testing the effect of ecoregion on viral community composition. Separate tests were performed for each sample type (saliva and feces) and virus data set (all viruses and vertebrate-infecting).

	d.f.	F	R^2	P-value†
All Virus Community Saliva	2,22	0.91	0.08	0.58
Vertebrate-infecting Virus Community Saliva	2,21	1.18	0.11	0.3
All Virus Community Feces	2,22	1.93	0.16	**0.005**
Vertebrate-infecting Virus Community Feces	2,20	0.15	0.15	0.09

†Significant p-values shown in bold

Table D 5 Viral richness final models with variables significant in model averaging. All variables included in final models had a model-averaged effect size that significantly differed from 0. Final model results were examined to verify that direction and relative magnitude of effect sizes were consistent with those from model averaging and univariate models. Final models were constructed for each sample type (saliva and feces), diversity measure (richness and TD) and virus dataset (all viruses and vertebrate-infecting).

	Best model	R^2	Adj R^2	Variable	Estimate (Std error)	Test stat	P-value	Partial R^2
Saliva All Richness	Richness~ Longitude + RawReads	0.5	0.51	Longitude	-0.11 (0.03)	-3.45	**0.001**	0.4
				Raw Reads	0.00000007 (0.00000003)	2.14	**0.03**	0.17
Saliva All TD	TD~ Longitude	0.35	0.35	Longitude	6.08 (1.79)	3.4	**0.003**	-
Feces All Richness	Richness~ Livestock + PropAdults + pc1 + Elevation	0.47	0.47	Livestock	-0.007 (0.003)	-2.15	**0.03**	0.19
				Proportion Adults	-1.4 (0.63)	-2.22	**0.03**	0.2
				Climate	0.04 (0.08)	0.45	0.7	0.01
				Elevation	0.00006 (0.0001)	0.51	0.6	0.01
Feces All TD	TD~ PropAdults + Elevation + pc1	0.37	0.37	Proportion Adults	-53.18 (32.6)	-1.63	0.12	0.11
				Elevation	-0.009 (0.006)	-1.45	0.16	0.09
				Climate	2.07 (4.36)	0.48	0.64	0
Saliva Vertebrate-infecting Richness	Richness~ Longitude	0.34	0.34	Longitude	-0.120 (0.038)	-3.13	**0.002**	-
Saliva Vertebrate-infecting TD	TD~ Longitude	0.21	0.21	Longitude	-5.61 (2.39)	-2.34	**0.03**	-
Feces Vertebrate-infecting Richness	Richness~ Elevation	0.21	0.22	Elevation	0.00026 (0.0001)	-2.16	**0.03**	-
Feces Vertebrate-infecting TD	TD~ PropAdults + Elevation	0.32	0.32	Proportion Adults	-68.13 (58.41)	-1.17	0.26	0.06
				Elevation	-0.013 (0.007)	-1.85	0.08	0.15

Table D 6 95% confidence set of GLMs for total saliva viral richness.
Model averaging of the models presented was used to estimate effect sizes and
confidence intervals for each explanatory variable. Models are ranked by ΔAICc; for
each model AICc, ΔAICc, Akaike weights (*wi*) and *R*^{2} are shown.

Model	AICc	ΔAICc	*wi*	R^2
Longitude + RawReads + 1	126.59	0	0.383	0.503
Longitude + 1	128.32	1.73	0.161	0.399
Livestock20k + Longitude + 1	128.62	2.028	0.139	0.458
Elevation + Longitude + 1	129.37	2.774	0.096	0.44
ColonyPropMales + Longitude + 1	129.54	2.951	0.088	0.436
Longitude + pc1 + 1	129.75	3.154	0.079	0.431
Longitude + OtherSppPA + 1	130.51	3.919	0.054	0.411

Table D 7 95% confidence set of GLMs for total saliva viral TD.
Model averaging of the models presented was used to estimate effect sizes and
confidence intervals for each explanatory variable. Models are ranked by ΔAICc; for
each model AICc, ΔAICc, Akaike weights (*wi*) and *R*^{2} are shown.

Model	AICc	delta AICc	weight	R^2
Livestock20k + Longitude + 1	204.5	0	0.291	0.46
Longitude + pc1 + 1	205.23	0.727	0.202	0.442
Elevation + Longitude + 1	205.6	1.095	0.168	0.433
Longitude + 1	205.63	1.129	0.165	0.355
Longitude + RawReads + 1	206.98	2.478	0.084	0.398
ColonyPropMales + Longitude + 1	208.21	3.71	0.046	0.365
Longitude + OtherSppPA + 1	208.3	3.801	0.043	0.363

Table D 8 95% confidence set of GLMs for vertebrate-infecting saliva viral richness.
Model averaging of the models presented was used to estimate effect sizes and
confidence intervals for each explanatory variable. Models are ranked by ΔAICc; for
each model AICc, ΔAICc, Akaike weights (*wi*) and *R*^{2} are shown.

Model	AICc	delta AICc	weight	R^2
Longitude + 1	113.27	0	0.243	0.342
Longitude + pc1 + 1	113.82	0.545	0.185	0.4
Livestock20k + Longitude + 1	114.02	0.745	0.167	0.395
Elevation + Longitude + 1	114.27	0.993	0.148	0.388
ColonyPropAdults + Longitude + 1	115.02	1.747	0.101	0.368
Longitude + RawReads + 1	115.28	2.007	0.089	0.361
ColonyPropMales + Longitude + 1	115.87	2.591	0.066	0.344

Table D 9 95% confidence set of GLMs for vertebrate-infecting saliva viral TD. Model averaging of the models presented was used to estimate effect sizes and confidence intervals for each explanatory variable. Models are ranked by ΔAICc; for each model AICc, ΔAICc, Akaike weights (*wi*) and R^2 are shown.

Model	AICc	delta AICc	weight	R^2
Livestock20k + Longitude + 1	218.5	0	0.217	0.319
Elevation + Longitude + 1	218.95	0.449	0.173	0.306
Longitude + 1	219.04	0.541	0.165	0.207
Longitude + pc1 + 1	220.29	1.797	0.088	0.264
ColonyPropAdults + Longitude + 1	221.27	2.777	0.054	0.232
Longitude + RawReads + 1	221.34	2.845	0.052	0.23
ColonyPropMales + Longitude + 1	221.82	3.318	0.041	0.214
Longitude + OtherSppPA + 1	221.85	3.35	0.041	0.213
Livestock20k + 1	223.32	4.827	0.019	0.045
OtherSppPA + 1	223.52	5.025	0.018	0.037
Elevation + 1	223.86	5.362	0.015	0.023
RawReads + 1	223.86	5.363	0.015	0.023
ColonyPropMales + 1	224.16	5.661	0.013	0.01
ColonyPropAdults + 1	224.27	5.772	0.012	0.005
Fis9loc + 1	224.28	5.783	0.012	0.005
pc1 + 1	224.28	5.786	0.012	0.004
Livestock20k + OtherSppPA + 1	224.99	6.495	0.008	0.097
Livestock20k + RawReads + 1	225.13	6.635	0.008	0.092
Elevation + OtherSppPA + 1	225.31	6.811	0.007	0.085
Fis9loc + Livestock20k + 1	225.37	6.874	0.007	0.082
OtherSppPA + pc1 + 1	225.68	7.184	0.006	0.07
ColonyPropMales + Livestock20k + 1	225.91	7.414	0.005	0.06
OtherSppPA + RawReads + 1	225.97	7.469	0.005	0.058
Elevation + RawReads + 1	226	7.505	0.005	0.057

Table D 10 95% confidence set of GLMs for total fecal viral richness.
Model averaging of the models presented was used to estimate effect sizes and
confidence intervals for each explanatory variable. Models are ranked by ΔAICc; for
each model AICc, ΔAICc, Akaike weights (*wi*) and *R*² are shown.

Model	AICc	delta AICc	weight	R^2
ColonyPropAdults + Livestock20k + 1	116.7	0	0.277	0.445
Livestock20k + 1	118.61	1.911	0.106	0.323
ColonyPropMales + Livestock20k + 1	118.62	1.916	0.106	0.397
ColonyPropAdults + 1	118.98	2.283	0.088	0.312
Livestock20k + Longitude + 1	119.97	3.27	0.054	0.361
Livestock20k + OtherSppPA + 1	120.64	3.939	0.039	0.342
Fis9loc + Livestock20k + 1	120.68	3.974	0.038	0.341
ColonyPropAdults + Longitude + 1	120.75	4.049	0.037	0.338
ColonyPropAdults + OtherSppPA + 1	121.08	4.383	0.031	0.329
OtherSppPA + pc1 + 1	121.22	4.524	0.029	0.325
Livestock20k + RawReads + 1	121.25	4.548	0.028	0.324
pc1 + 1	121.45	4.746	0.026	0.234
ColonyPropAdults + Fis9loc + 1	121.52	4.818	0.025	0.316
ColonyPropAdults + RawReads + 1	121.64	4.94	0.023	0.312
Elevation + 1	121.89	5.192	0.021	0.219
ColonyPropMales + pc1 + 1	122.13	5.432	0.018	0.297
ColonyPropMales + Elevation + 1	122.18	5.482	0.018	0.296
Longitude + pc1 + 1	122.61	5.905	0.014	0.283
Elevation + Longitude + 1	122.99	6.288	0.012	0.271
ColonyPropMales + 1	123.38	6.675	0.01	0.167

Table D 11 95% confidence set of GLMs for total fecal viral TD.
Model averaging of the models presented was used to estimate effect sizes and
confidence intervals for each explanatory variable. Models are ranked by ΔAICc; for
each model AICc, ΔAICc, Akaike weights (*wi*) and R^2 are shown.

Model	AICc	delta AICc	weight	R^2
Elevation + 1	201.88	0	0.165	0.278
ColonyPropAdults + 1	202.32	0.448	0.132	0.263
Elevation + Fis9loc + 1	202.51	0.639	0.12	0.347
Elevation + RawReads + 1	203.54	1.667	0.072	0.317
ColonyPropMales + Elevation + 1	203.73	1.851	0.065	0.312
Elevation + Longitude + 1	203.95	2.07	0.059	0.305
ColonyPropAdults + Fis9loc + 1	204.49	2.611	0.045	0.288
pc1 + 1	204.5	2.626	0.044	0.19
Elevation + OtherSppPA + 1	204.51	2.63	0.044	0.288
ColonyPropAdults + Livestock20k + 1	204.52	2.643	0.044	0.287
ColonyPropAdults + RawReads + 1	204.73	2.852	0.04	0.281
ColonyPropAdults + Longitude + 1	205.07	3.197	0.033	0.27
Fis9loc + pc1 + 1	205.24	3.364	0.031	0.265
ColonyPropAdults + OtherSppPA + 1	205.28	3.407	0.03	0.263
ColonyPropMales + pc1 + 1	206.33	4.454	0.018	0.229
pc1 + RawReads + 1	206.54	4.664	0.016	0.222
OtherSppPA + pc1 + 1	206.64	4.768	0.015	0.219
Longitude + pc1 + 1	206.85	4.97	0.014	0.212
ColonyPropMales + 1	207.02	5.146	0.013	0.096

Table D 12 95% confidence set of GLMs for vertebrate-infecting fecal viral richness. Model averaging of the models presented was used to estimate effect sizes and confidence intervals for each explanatory variable. Models are ranked by ΔAICc; for each model AICc, ΔAICc, Akaike weights (*wi*) and R^2 are shown.

Model	AICc	delta AICc	weight	R^2
Elevation + 1	78.96	0	0.146	0.181
ColonyPropAdults + 1	80.04	1.072	0.086	0.142
Elevation + OtherSppPA + 1	80.22	1.26	0.078	0.229
Elevation + Longitude + 1	80.59	1.621	0.065	0.217
pc1 + 1	80.91	1.947	0.055	0.109
Livestock20k + 1	80.99	2.027	0.053	0.106
OtherSppPA + pc1 + 1	81.29	2.324	0.046	0.193
ColonyPropAdults + Livestock20k + 1	81.55	2.589	0.04	0.184
Elevation + Fis9loc + 1	81.61	2.65	0.039	0.181
Elevation + RawReads + 1	81.62	2.653	0.039	0.181
ColonyPropMales + Elevation + 1	81.62	2.659	0.039	0.181
ColonyPropAdults + OtherSppPA + 1	82.01	3.041	0.032	0.167
ColonyPropAdults + Longitude + 1	82.32	3.359	0.027	0.156
ColonyPropAdults + Fis9loc + 1	82.58	3.617	0.024	0.146
ColonyPropAdults + RawReads + 1	82.63	3.661	0.023	0.145
Longitude + pc1 + 1	82.97	4.001	0.02	0.132
Livestock20k + OtherSppPA + 1	83.06	4.092	0.019	0.128
Fis9loc + 1	83.19	4.226	0.018	0.016
Livestock20k + Longitude + 1	83.22	4.254	0.017	0.122
OtherSppPA + 1	83.36	4.391	0.016	0.009
ColonyPropMales + 1	83.37	4.409	0.016	0.008
Longitude + 1	83.48	4.52	0.015	0.003
RawReads + 1	83.52	4.554	0.015	0.002
pc1 + RawReads + 1	83.57	4.605	0.015	0.109
ColonyPropMales + pc1 + 1	83.57	4.609	0.015	0.109
Fis9loc + pc1 + 1	83.57	4.61	0.015	0.109
ColonyPropMales + Livestock20k + 1	83.62	4.651	0.014	0.107
Fis9loc + Livestock20k + 1	83.63	4.661	0.014	0.107

Table D 13 95% confidence set of GLMs for vertebrate-infecting fecal viral TD. Model averaging of the models presented was used to estimate effect sizes and confidence intervals for each explanatory variable. Models are ranked by ΔAICc; for each model AICc, ΔAICc, Akaike weights (*wi*) and *R*^{2} are shown.

Model	AICc	delta AICc	weight	R^2
Elevation + OtherSppPA + 1	230.27	0	0.208	0.362
Elevation + 1	230.28	0.015	0.207	0.274
Elevation + RawReads + 1	232.17	1.907	0.08	0.307
ColonyPropAdults + 1	232.4	2.139	0.071	0.204
Elevation + Longitude + 1	232.56	2.29	0.066	0.296
ColonyPropMales + Elevation + 1	232.58	2.316	0.065	0.295
Elevation + Fis9loc + 1	233.16	2.891	0.049	0.277
OtherSppPA + pc1 + 1	233.19	2.928	0.048	0.276
pc1 + 1	233.92	3.654	0.033	0.15
ColonyPropAdults + OtherSppPA + 1	234.37	4.102	0.027	0.238
ColonyPropAdults + Livestock20k + 1	234.66	4.396	0.023	0.228
ColonyPropAdults + RawReads + 1	234.97	4.7	0.02	0.218
ColonyPropAdults + Longitude + 1	235.28	5.018	0.017	0.207
ColonyPropAdults + Fis9loc + 1	235.28	5.018	0.017	0.207
ColonyPropMales + 1	235.91	5.648	0.012	0.073
ColonyPropMales + pc1 + 1	236.09	5.827	0.011	0.178
Livestock20k + 1	236.18	5.916	0.011	0.062
pc1 + RawReads + 1	236.2	5.933	0.011	0.175
Longitude + pc1 + 1	236.55	6.286	0.009	0.162
Fis9loc + pc1 + 1	236.84	6.577	0.008	0.151
RawReads + 1	237.29	7.027	0.006	0.016

D2: Figures
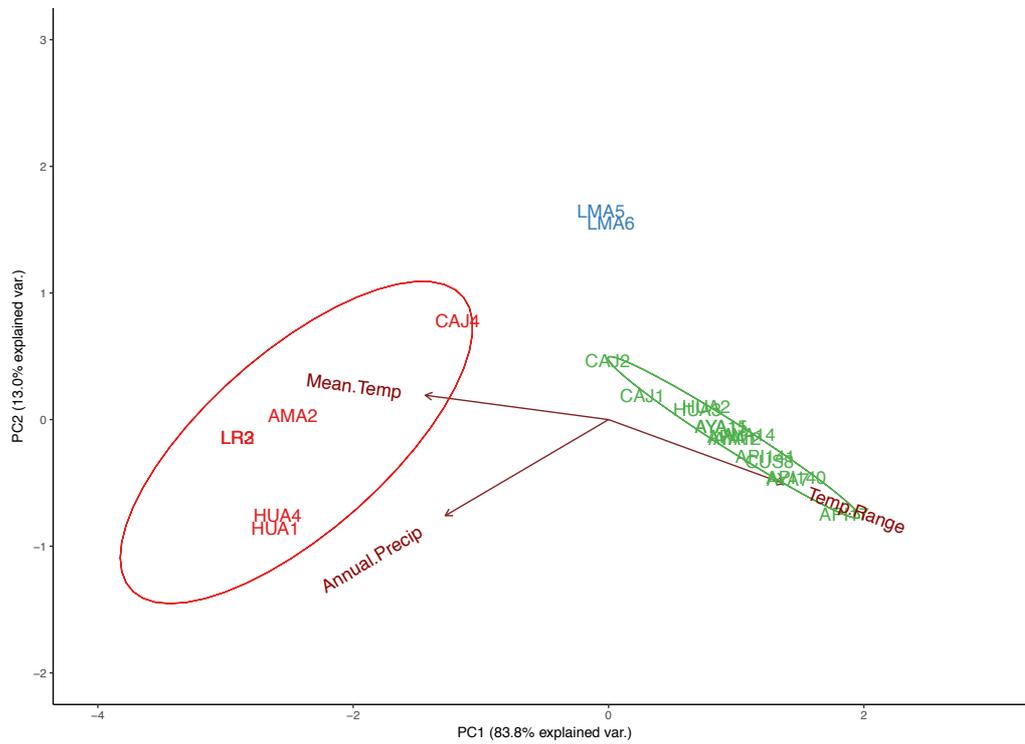
Figure D 1 Principal component analysis (PCA) of sites by environmental variables. The PCA was performed with centering and scaling on the variables annual mean temperature, annual precipitation, and annual temperature range. Sites are colored by ecoregion and circles show the 95% normal probability ellipse for each group.
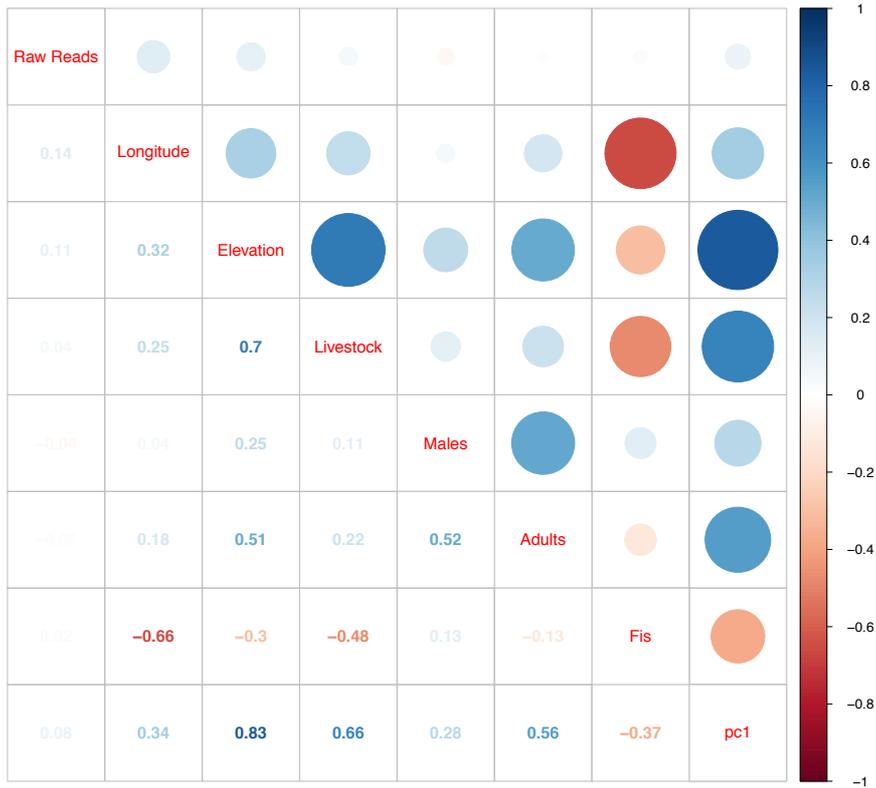
**Figure D 2 Pearson correlations between variables in viral richness modeling.
Pairwise correlations between all continuous explanatory variables were examined for
potential multi-collinearity. Variables with a correlation coefficient of r>0.5 were
excluded from the same model in model averaging.**
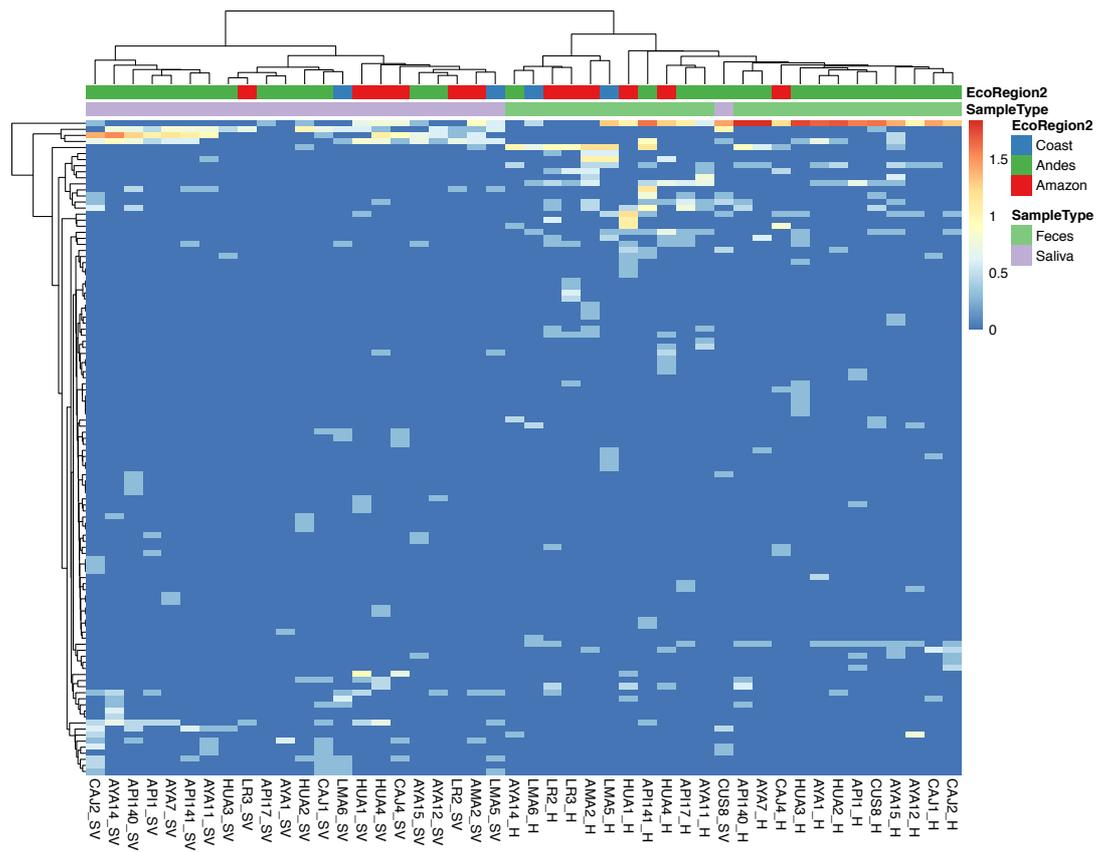
Figure D 3 Heatmap of viral read number in single colony pools.
Read data is depicted at the genus level where similar rows (viral genera detected together) and columns (pools containing viral families in common) are clustered according to Ward's method. Colored bars above indicate ecoregion and sample type.
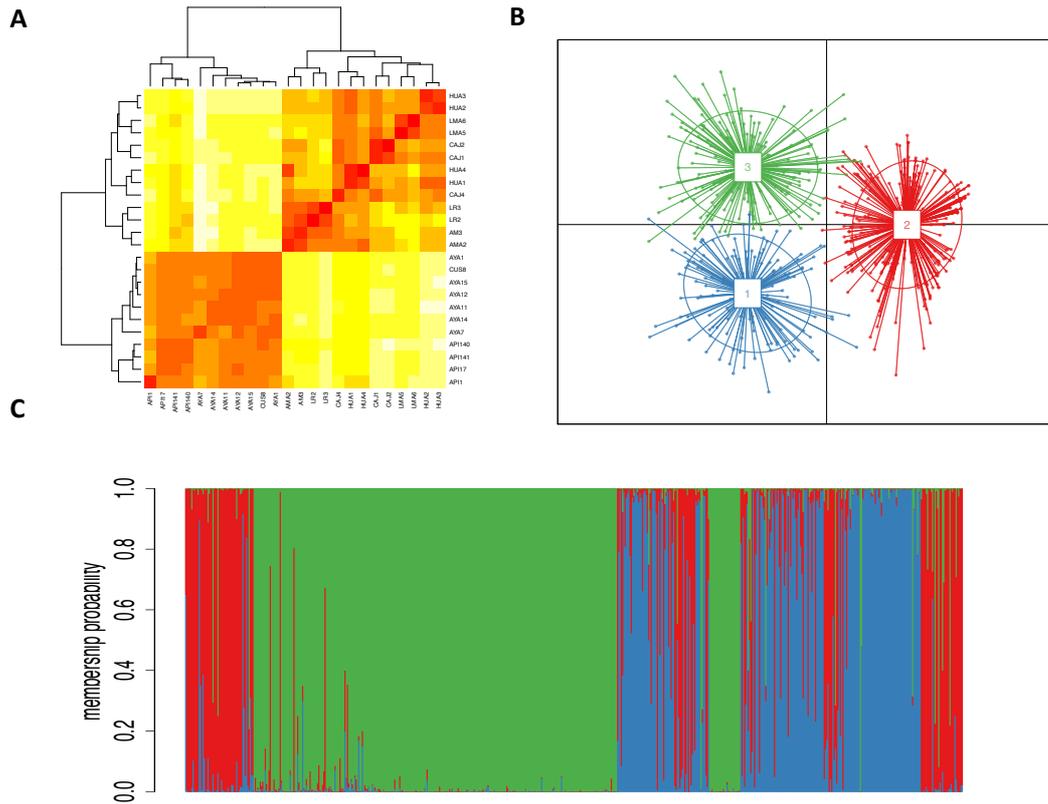
Figure D 4 Population genetic structure of vampire bats based on 6 microsatellite loci.
Genetic structure across Peru is shown as (A) heatmap of F_{ST} values between colonies
(B) scatterplot based on k-means clustering in DAPC (C) maximum likelihood
clustering in snapclust. In panel A darker shades of red correspond to lower F_{ST}, or
higher connectivity, while lighter yellow and white correspond to higher F_{ST}, or lower
connectivity. Colors in panels B and C correspond to different ecoregions within Peru
as shown in Figure 4.1.
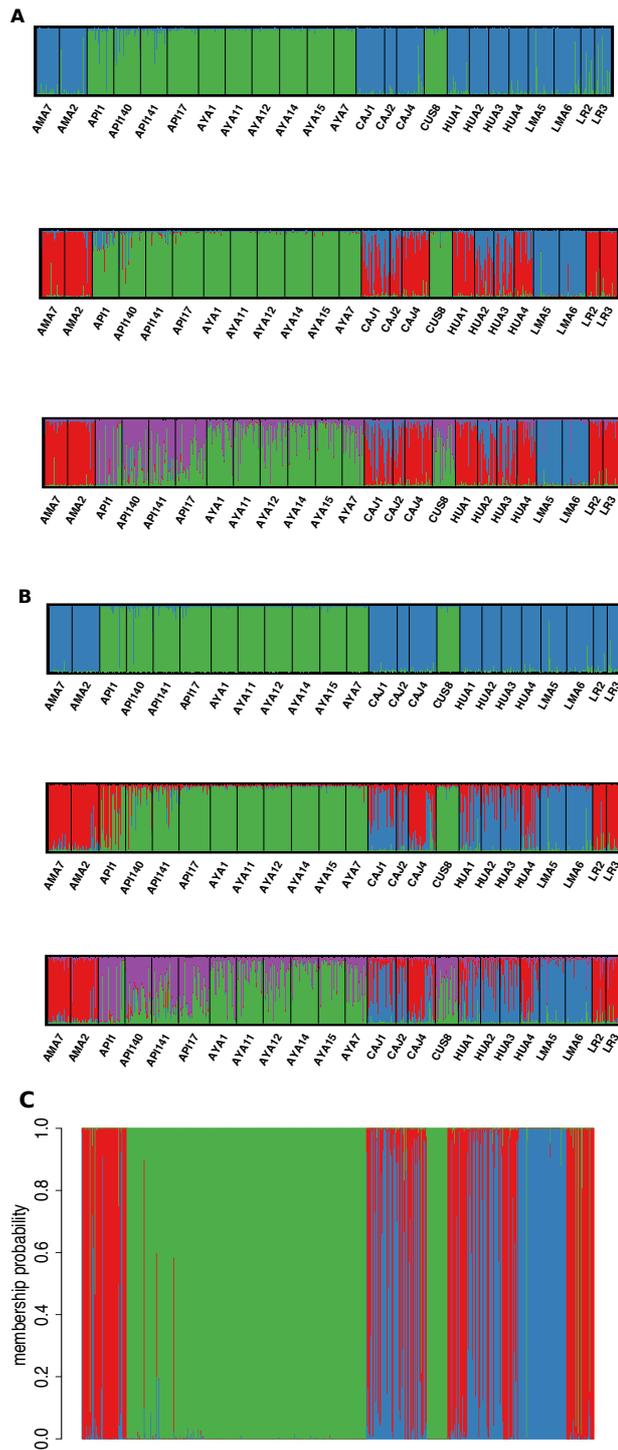
Figure D 5 Supplementary membership plots for microsatellite analyses.
Plots show (A-B) comparison of different values of K in microsatellite datasets with different numbers of loci and (C) the snapclust analysis results using 9 loci. Plots show values of K=2-4 for the (A) 6 and (B) 9 microsatellite loci datasets. Colony names and colors correspond to sites and ecoregions.
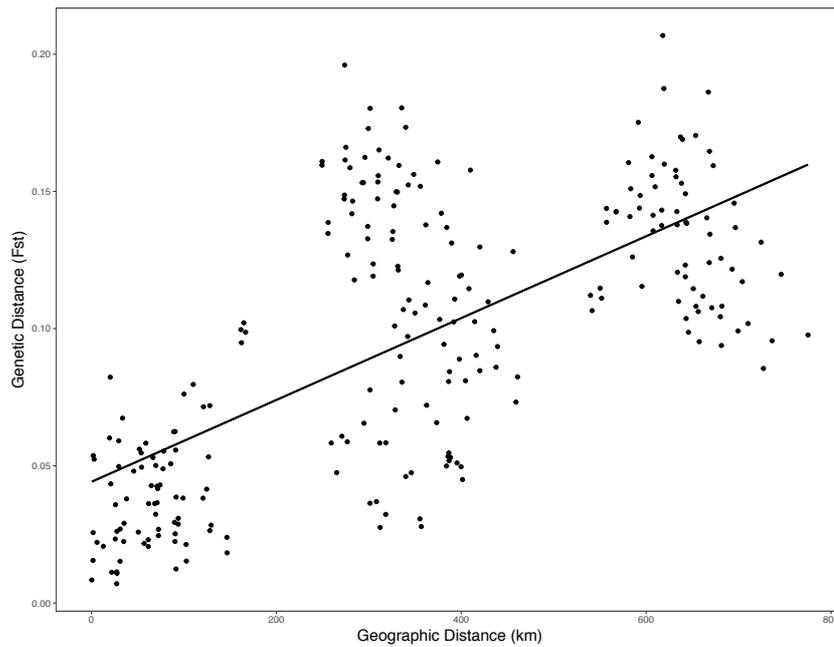
Figure D 6 Genetic isolation by distance between vampire bat colonies.
Isolation by distance plot showing the correlation between geographic distances and genetic distances between sites. Genetic distance is measured as F_{ST} from the dataset of 9 microsatellite loci, but results were consistent in an analysis with 6 loci.
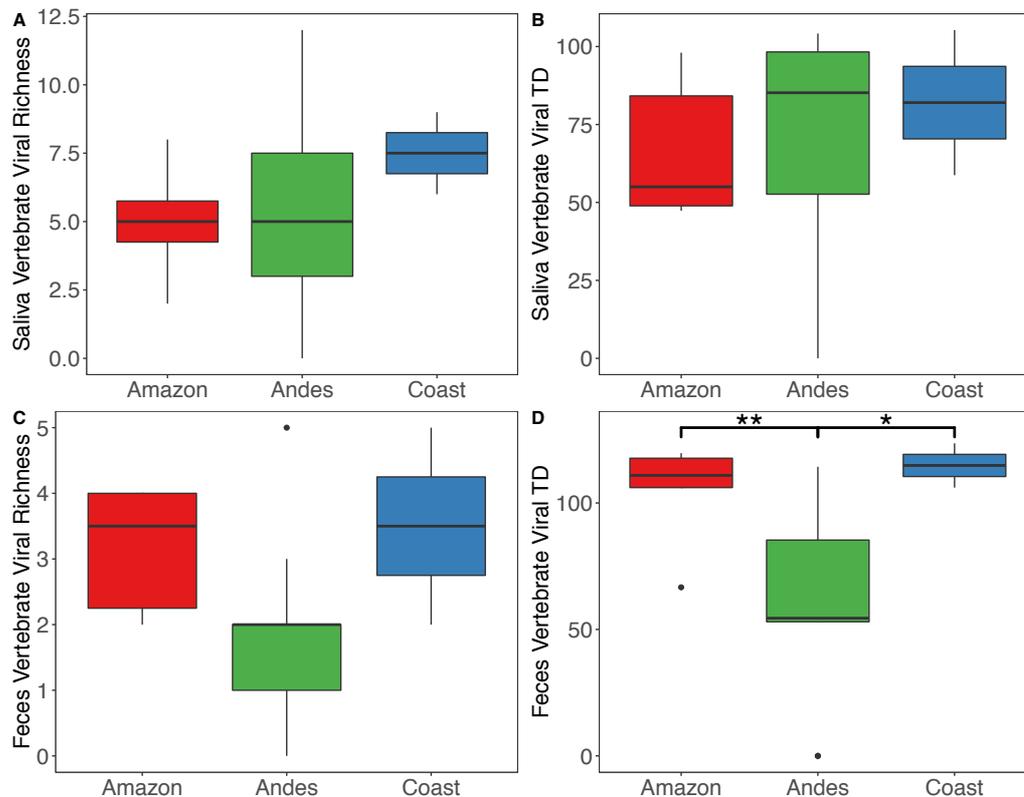
Figure D 7 Vertebrate-infecting viral richness and TD compared across ecoregions.
Comparisons between ecoregions are shown for saliva (A-B) and feces (C-D). In boxplots, bold line shows the median, and upper and lower hinges show the first and third quartiles. Whiskers extend from the hinge to 1.5 x the inter-quartile range. Stars indicate significance level of post-hoc Tukey pairwise comparisons, where * indicates p<0.05 and ** indicates p<0.01. Colors correspond to different ecoregions within Peru.
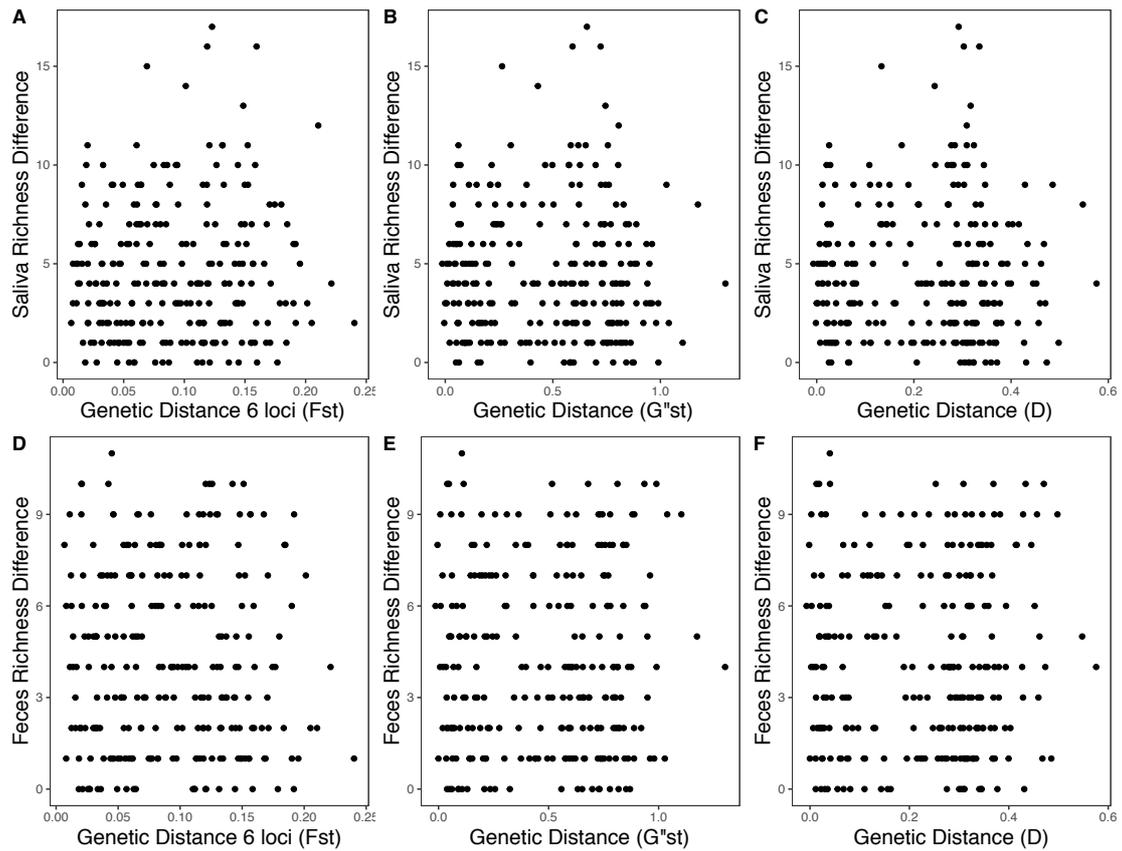
Figure D 8 Correlations between viral richness and host genetic distance.
Plots depict relationships between saliva viral richness differences and F_{ST} calculated using 6 microsatellite loci (Panel A; Mantel r=0.06; p=0.23), G''_{ST} (Panel B; Mantel r=-0.002; p=0.48) and D (Panel C; Mantel r=0.006; p=0.43). Fecal viral richness differences are also compared with F_{ST} calculated using 6 microsatellite loci (Panel D; Mantel r=-0.035; p=0.69), G''_{ST} (Panel E; Mantel r=0.023; p=0.28), and D (Panel F; Mantel r=0.043; p=0.21)
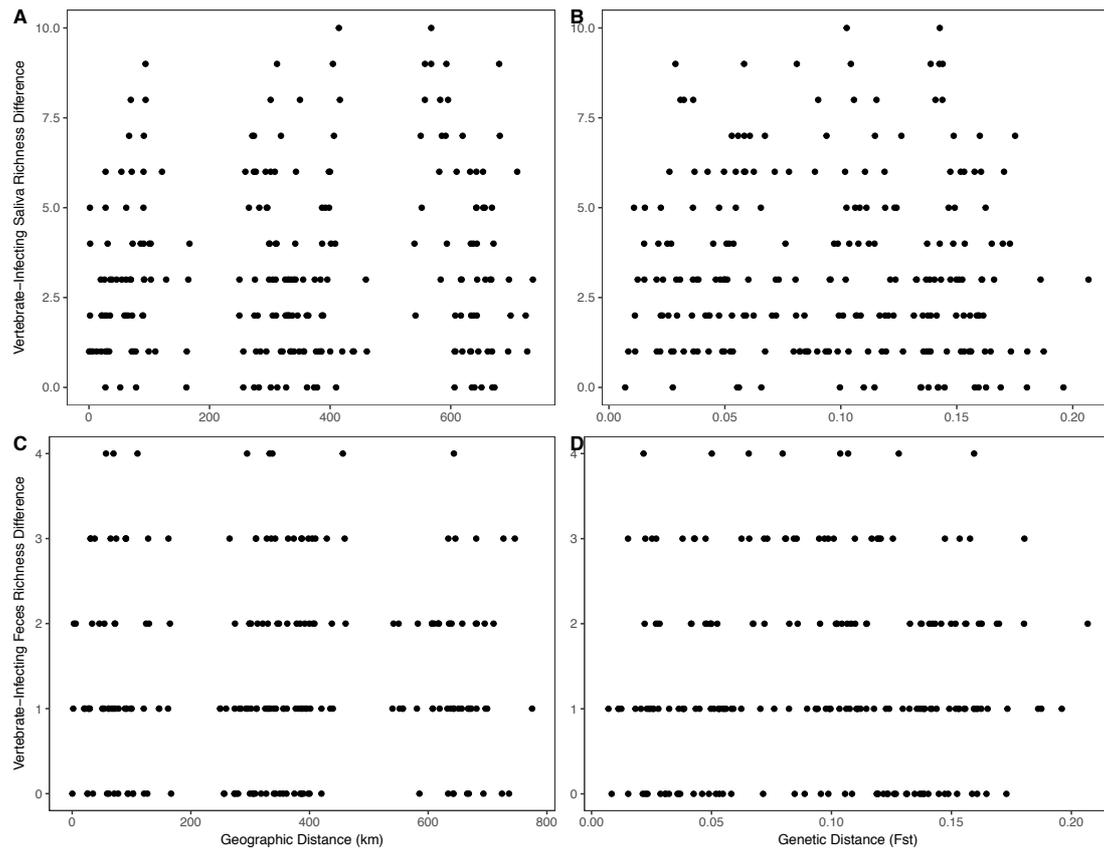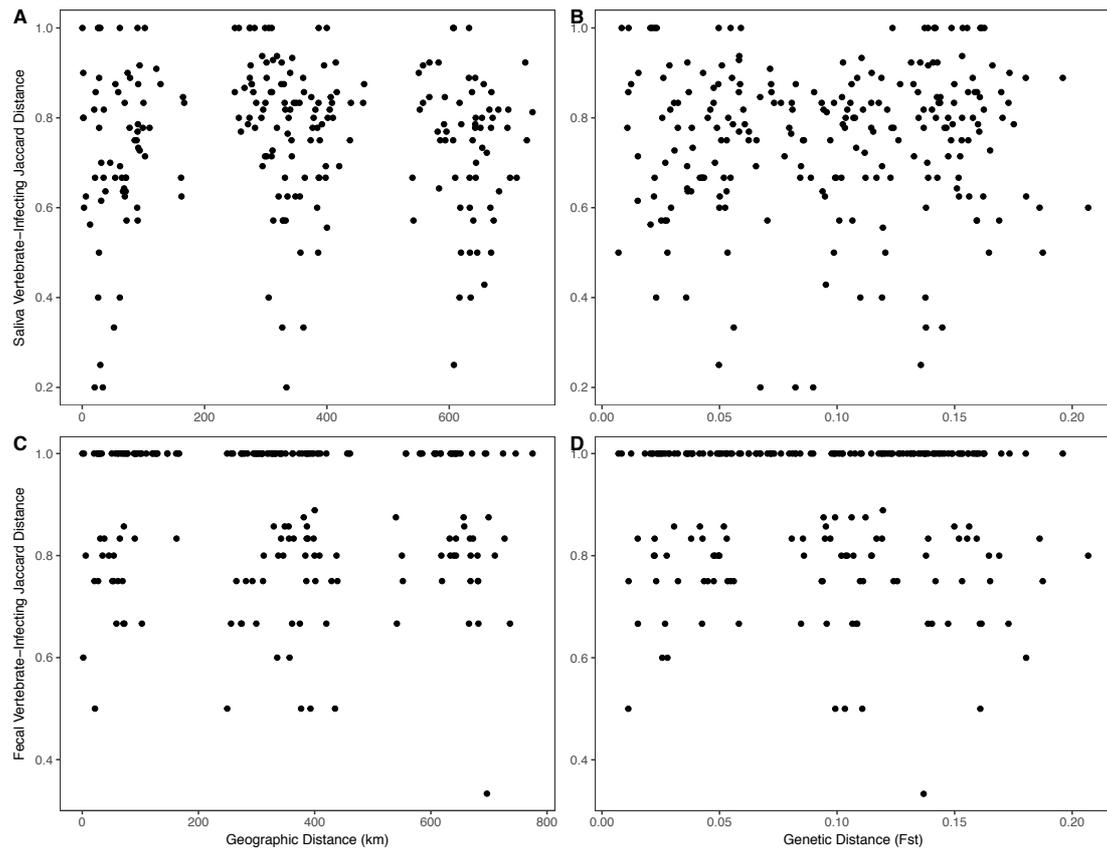
Figure D 9 Correlations between vertebrate-infecting viral richness and geographic or genetic distance. Plots show saliva vertebrate-infecting viral richness differences compared with colony geographic distance (km) (Panel A; r= 0.11; p= 0.12) and genetic distance F_{ST} calculated using 9 microsatellite loci (Panel B; Mantel r= -0.047; p= 0.75). Fecal vertebrate-infecting viral richness differences are also compared with colony geographic distance (km) (Panel C; Mantel r= 0.029; 0.34) and genetic distance F_{ST} calculated using 9 microsatellite loci (Panel D; Mantel r= -0.0009; p= 0.48).
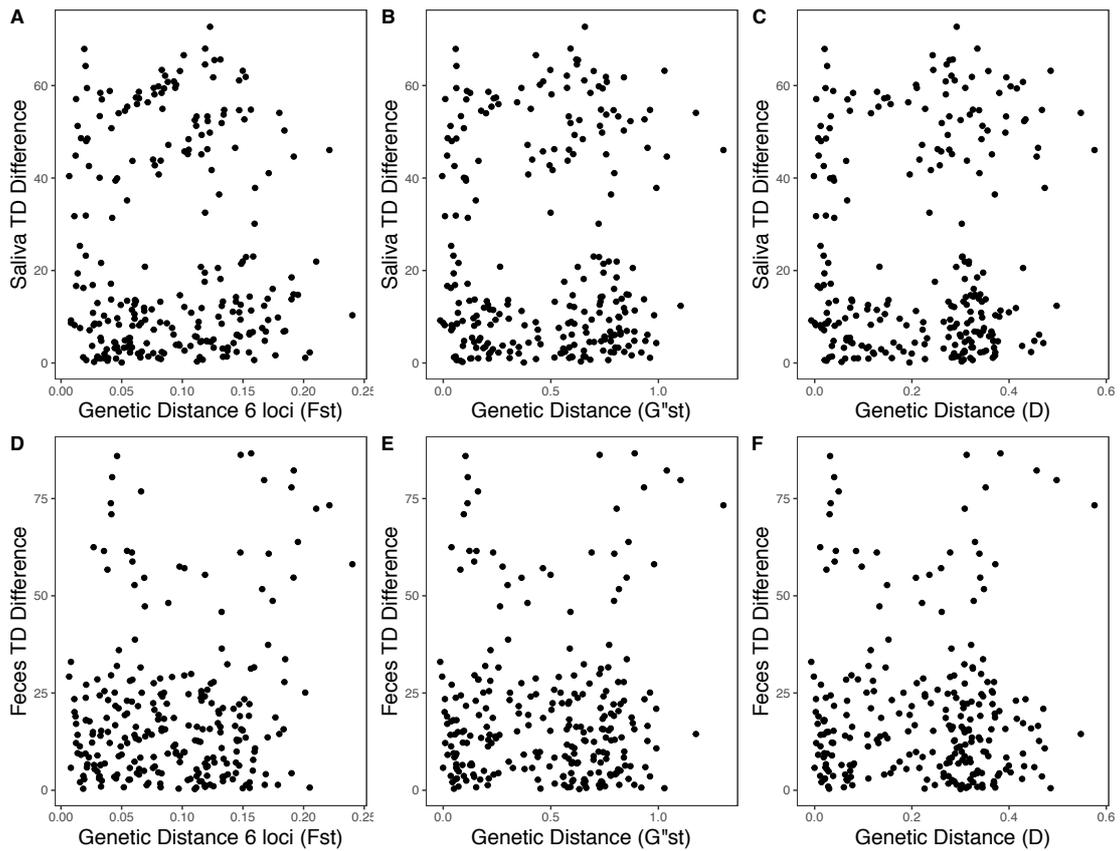
Figure D 10 Correlations between viral TD and host genetic distances.
Plots show saliva viral TD differences and F_{ST} calculated using 6 microsatellite loci
(Panel A; Mantel r= -0.004; p= 0.49), G''_{ST} (Panel B; Mantel r= 0.017; p= 0.37) and D
(Panel C; Mantel r= 0.05; p= 0.21). Feces viral TD differences are also compared with
F_{ST} calculated using 6 microsatellite loci (Panel D; Mantel r= 0.15; p= 0.04), G''_{ST} (Panel
E; Mantel r= 0.04; p= 0.3), and D (Panel F; Mantel r= -0.01; p= 0.5)
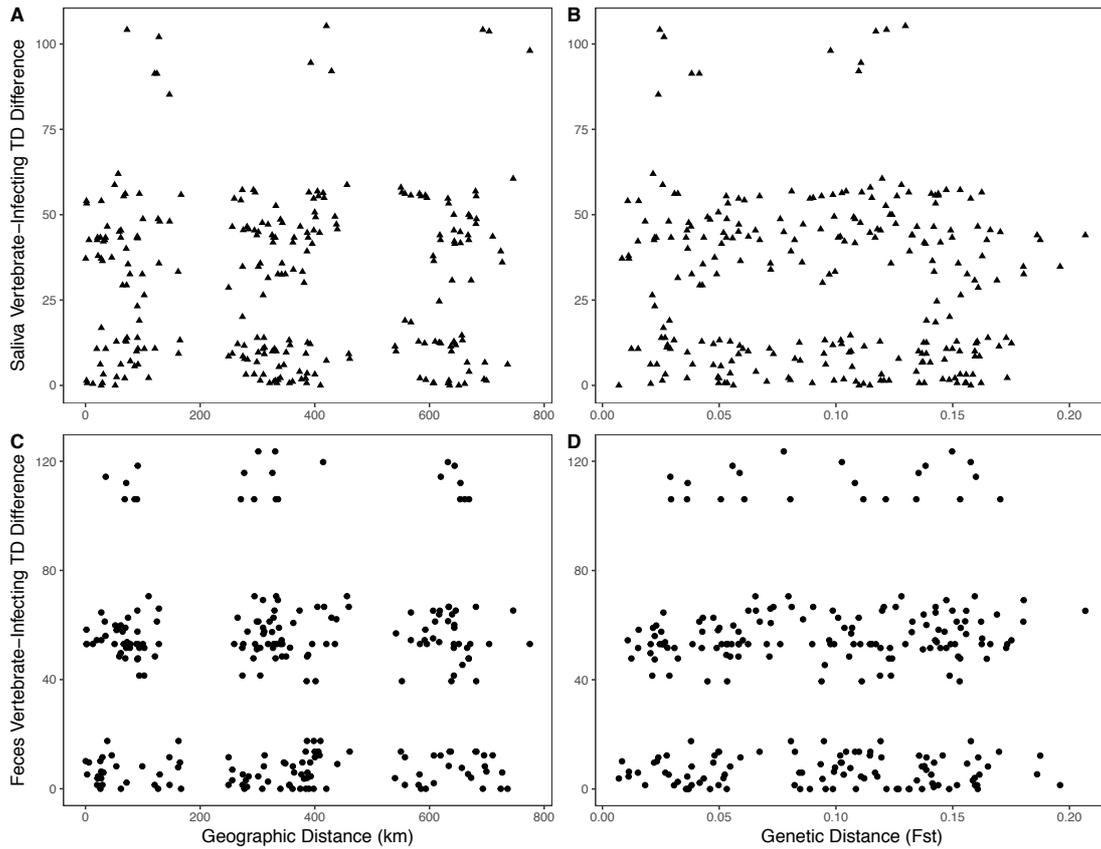
Figure D 11 Correlations between vertebrate-infecting viral TD and geographic or genetic distance. Correlations are shown between saliva vertebrate-infecting viral TD differences with colony geographic distance (km) (Panel A; r= 0.02; p= 0.33) and genetic distance F_{ST} calculated using 9 microsatellite loci (Panel B; Mantel r= -0.07; p= 0.88). Fecal vertebrate-infecting viral TD differences are also compared with colony geographic distance (km) (Panel C; Mantel r= 0.007; 0.43) and genetic distance F_{ST} calculated using 9 microsatellite loci (Panel D; Mantel r= 0.03; p= 0.29).
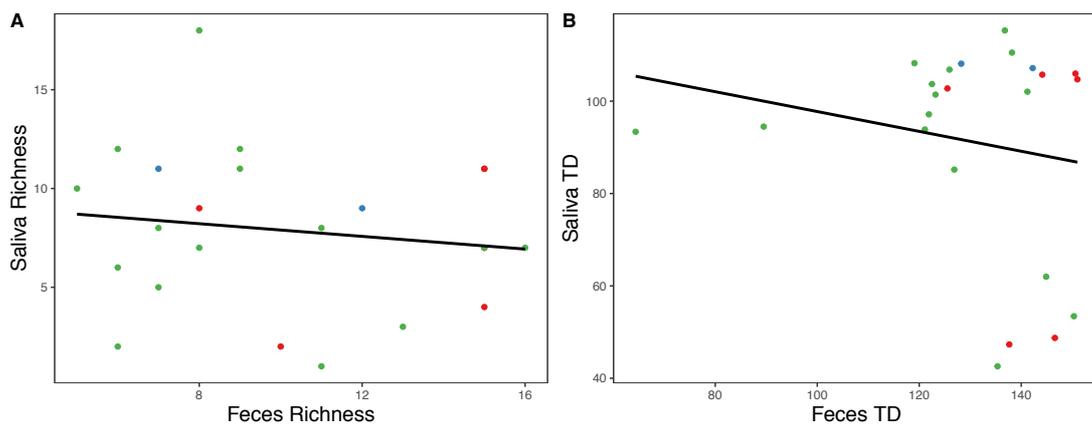
Figure D 12 Comparison between saliva and fecal viral richness and TD. Comparisons are shown between sample types for both viral richness (Panel A; R^2= 0.02; p=0.51) and TD (Panel B; R^2= 0.04; p= 0.39). Points each represent a site and are colored according to ecoregion.
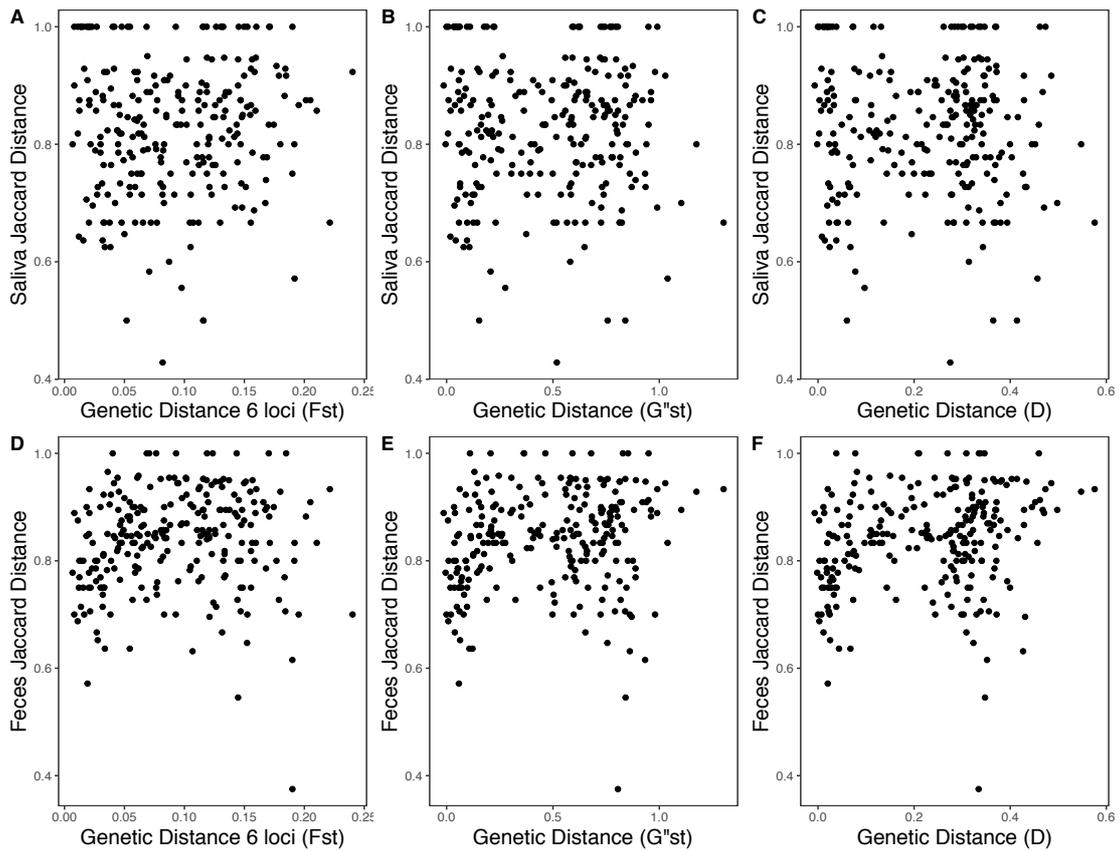
Figure D 13 Correlations between viral community distance and host genetic distance.
Plots depict saliva viral Jaccard community distances and F_{ST} calculated using 6
microsatellite loci (Panel A; Mantel r=0.06; p=0.22), G''_{ST} (Panel B; Mantel r=-0.02;
p=0.61) and D (Panel C; Mantel r=-0.05; p=0.75). Fecal viral Jaccard community
distances are also compared with F_{ST} calculated using 6 microsatellite loci (Panel D;
Mantel r=0.1; p=0.09), G''_{ST} (Panel E; Mantel r=0.17; p=0.01), and D (Panel F; Mantel
r=0.23; p=0.003).

Figure D 14 Correlations between vertebrate-infecting viral community distance and geographic or genetic distance. Plots show correlations of vertebrate-infecting saliva virus community Jaccard distances with colony geographic distance (km) (Panel A; Mantel r= 0.006; p= 0.44) and genetic distance F_{ST} calculated using 9 microsatellite loci (Panel B; Mantel r= 0.06; p= 0.19). Fecal vertebrate-infecting virus community Jaccard distances are also compared with colony geographic distance (km) (Panel C; Mantel r= -0.076; p= 0.82) and genetic distance F_{ST} calculated using 9 microsatellite loci (Panel D; Mantel r= -0.015; p= 0.56).
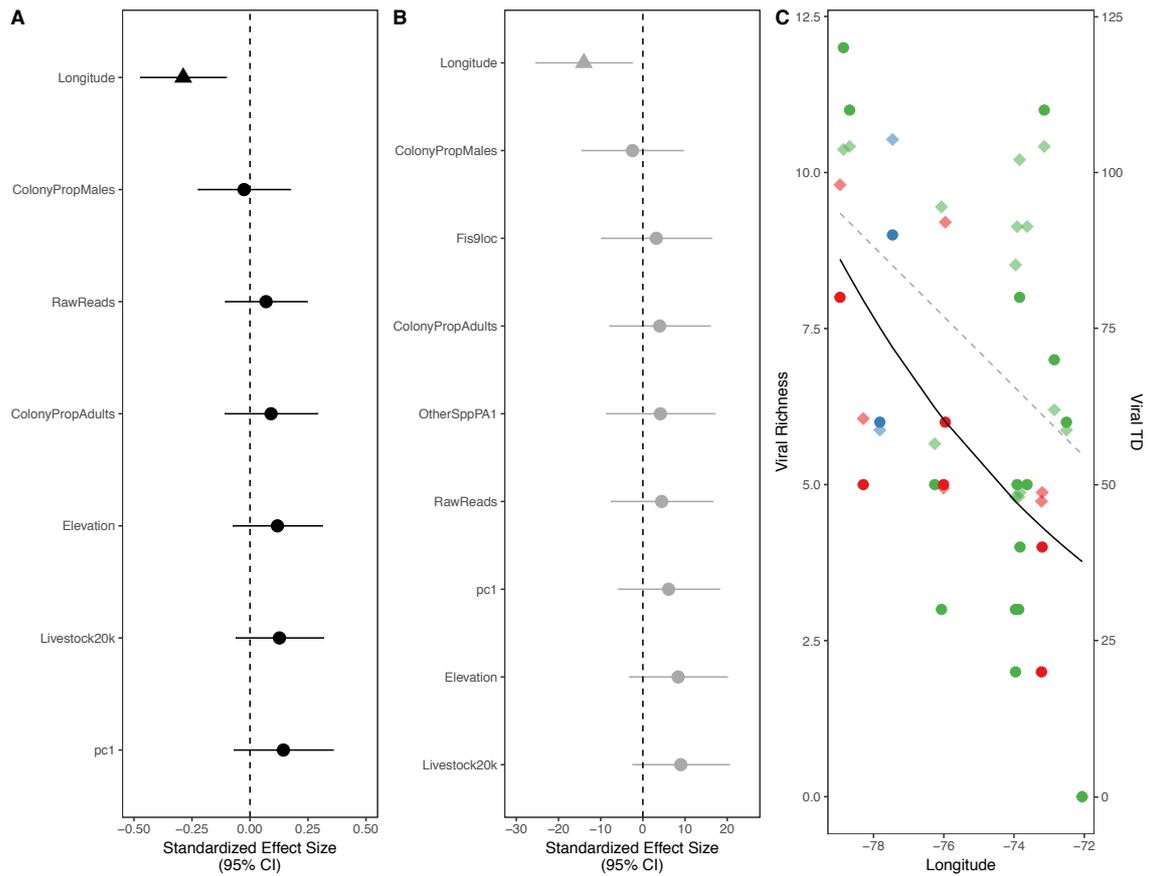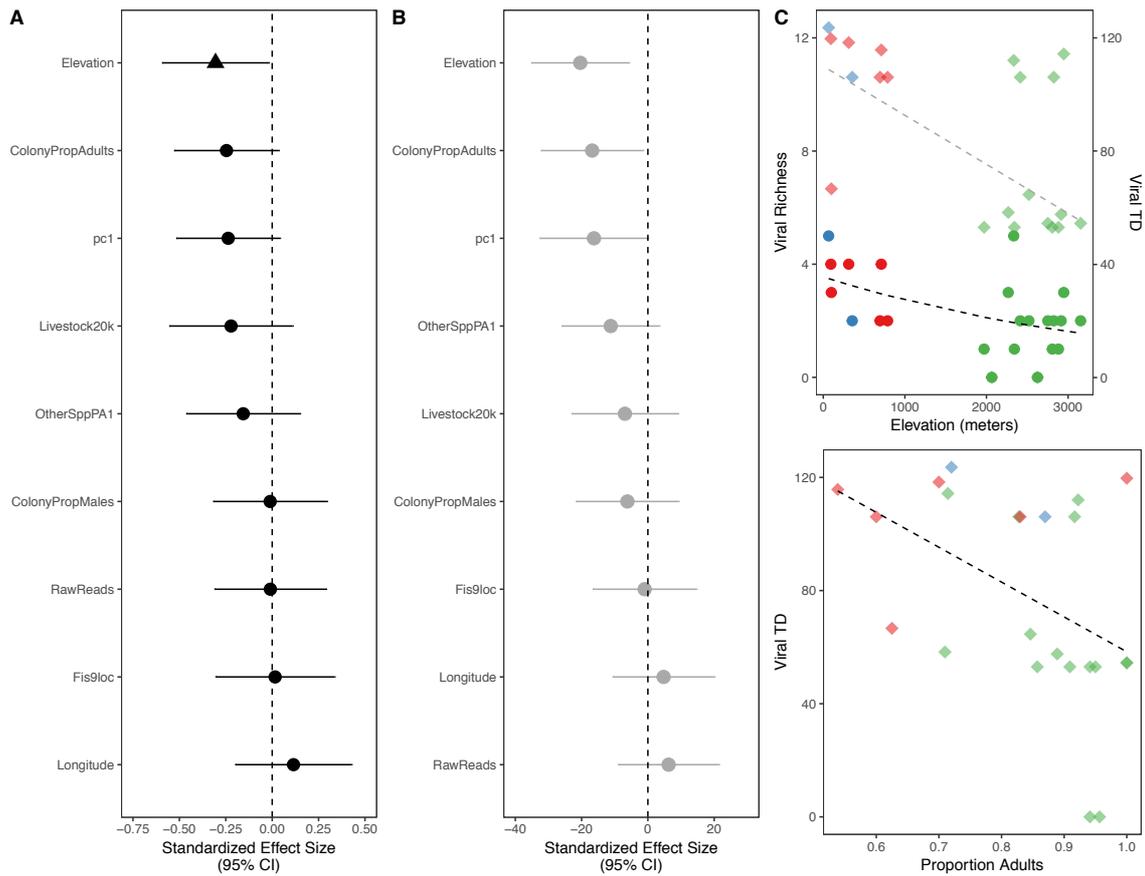
Figure D 15 Ecological correlates of vertebrate-infecting viral richness in bat saliva samples. Model averaged relationships of demographic and environmental factors correlated with (A) richness and (B) TD and (C) univariate correlations of significant factors. Viral richness model results shown in black and TD results are shown in gray. In panels (A) and (B) the model averaged effect sizes are shown for each factor across the 95% confidence set of GLMs with 95% confidence intervals. Factors that remained significant in the final model are shown as triangles. The vertical dashed line shows an effect size of zero, such that any confidence intervals overlapping the dashed line indicate a non-significant effect of the factor in model averaged results. In panel (C) richness (left) and TD (right) are plotted together for each variable that was significant according to model averaging. Solid lines show GLM predictions for univariate relationships that remained significant following correction for multiple testing, while dashed lines are univariate relationships that were no longer significant after correction. Points are colored according to ecoregions; solid points are values for richness and translucent diamonds are values for TD. Richness represents the number of genera detected while the scale of TD cannot be directly related to number of taxa.
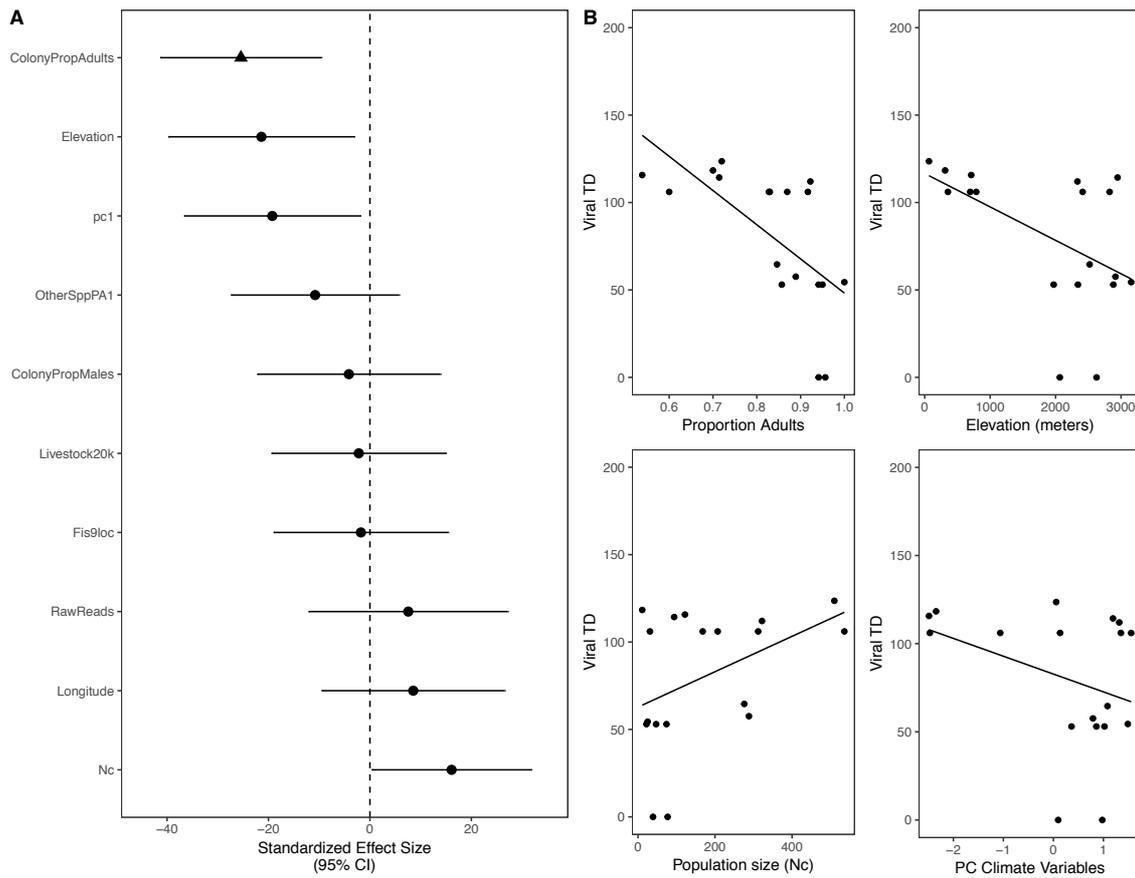
Figure D 16 Ecological correlates of vertebrate-infecting viral richness in bat fecal samples. Model averaged relationships of demographic and environmental factors correlated with (A) richness and (B) TD and (C) univariate correlations of significant factors. Viral richness model results shown in black and TD results are shown in gray. In panels (A) and (B) the model averaged effect sizes are shown for each factor across the 95% confidence set of GLMs with 95% confidence intervals. The vertical dashed line shows an effect size of zero, such that any confidence intervals overlapping the dashed line indicate a non-significant effect of the factor in model averaged results. In panel (C) richness (left) and TD (right) are plotted together for each variable that was significant according to model averaging. Solid lines show GLM predictions for univariate relationships that remained significant following correction for multiple testing, while dashed lines are univariate relationships that were no longer significant after correction. Points are colored according to ecoregions; solid points are values for richness and translucent diamonds are values for TD. Richness represents the number of genera detected while the scale of TD cannot be directly related to number of taxa.

Figure D 17 Ecological correlates of fecal vertebrate-infecting viral TD including colony size. Model averaged relationships are shown for host and environmental factors for the smaller dataset including colony size (N_c). Panel A shows the model averaged effect sizes for each factor across the 95% confidence set of GLMs with 95% confidence intervals. Factors that remained significant in the final model are shown as triangles. The vertical dashed line shows an effect size of zero, such that any confidence intervals overlapping the dashed line indicate a non-significant effect of the factor in model averaged results. In panel (B) univariate relationships are plotted for each variable that was significant according to model averaging

D3: Viral TD Calculation

This section provides further detail and an example for the calculation of hierarchical viral "taxonomic diversity" (TD). Each pair of viral genera found in fecal or saliva communities was assigned a pairwise score (1-4) depending on hierarchical relatedness (Table D3.1). Higher scores represent more distance in relatedness between taxa, akin to a longer phylogenetic branch length. Viral TD was calculated for each site using Faith's phylogenetic diversity measure by summing branch lengths for all viral genera found at a site. An example of two saliva communities with the same viral richness but very different viral TD are shown in Figure D3.1 to emphasize the potential for differences between the two measures. Values for viral TD range from $0 - 151$, where 0 represents no viral genera detected and 151 many distantly related genera. However, as relatedness distances were assigned arbitrary units across the viral taxonomy, the scale cannot be related directly to viral species richness.

Table D3 1 Details for the calculation of pairwise viral TD scores.
Pairwise relationship describes potential relationships of two viruses along with examples and the score given to them in the calculation of TD. These scores are then used to calculate TD at the colony level.

Pairwise relationship	Example	Score
Same family	*Herpesviridae, Papillomaviridae, Coronaviridae*	1
Same genome structure	dsDNA, ssDNA, dsRNA, ssRNA	2
Same molecule type	DNA, RNA	3
None of those	-	4
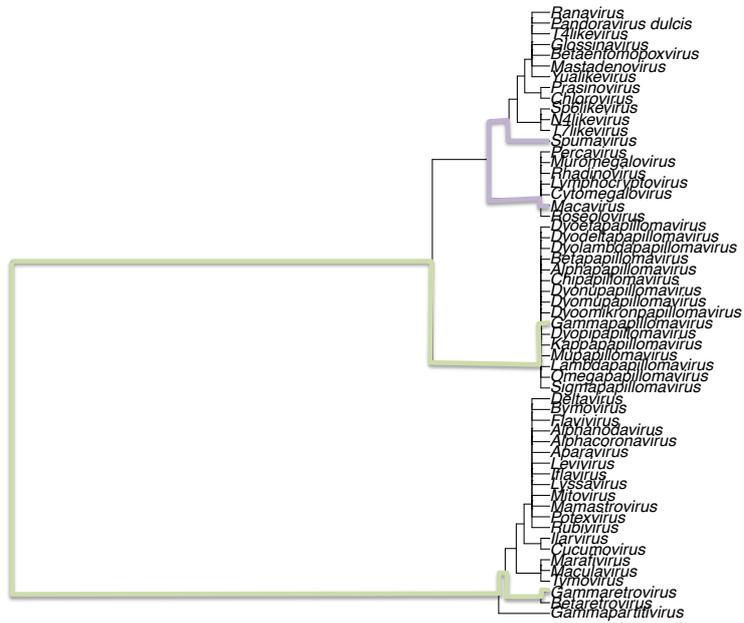
Figure D3 1 Example of TD score calculation for two saliva viral communities
TD score calculations are shown for two saliva viral communities with the same richness. The tree shows all viral genera that were detected in any saliva community, and branches highlighted in green and purple show distances between genera found at two sites. AYA1 is shown in green (viral richness = 2; viral TD = 85.19) and LR3 is shown in purple (viral richness = 2; viral TD = 47.32)

Appendix E: Additional Publications

Volokhov, D., Becker, D., **Bergner, L.**, Camus, M., Orton, R., Chizhikov, V., Altizer, S., Streicker, D. (2017). Novel hemotropic mycoplasmas are widespread and genetically diverse in vampire bats. *Epidemiology and Infection, 145*(15), 3154-3167. doi:10.1017/S095026881700231X

Novel hemotropic mycoplasmas are widespread and genetically diverse in vampire bats

D. V. VOLOKHOV[1]†*, D. J. BECKER[2,3]†, L. M. BERGNER[4], M. S. CAMUS[5], R. J. ORTON[4,6], V. E. CHIZHIKOV[1], S. M. ALTIZER[2,3] AND D. G. STREICKER[2,4,6]

[1] *Center for Biologics Evaluation & Research, Food and Drug Administration, Silver Spring, MD, USA*
[2] *Odum School of Ecology, University of Georgia, Athens, GA, USA*
[3] *Center for the Ecology of Infectious Disease, University of Georgia, Athens, GA, USA*
[4] *Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, UK*
[5] *Department of Pathology, College of Veterinary Medicine, University of Georgia, Athens, GA, USA*
[6] *MRC-University of Glasgow Centre for Virus Research, Glasgow, UK*

SUMMARY

Bats (Order: Chiroptera) have been widely studied as reservoir hosts for viruses of concern for human and animal health. However, whether bats are equally competent hosts of non-viral pathogens such as bacteria remains an important open question. Here, we surveyed blood and saliva samples of vampire bats from Peru and Belize for hemotropic *Mycoplasma* spp. (hemoplasmas), bacteria that can cause inapparent infection or anemia in hosts. 16S rRNA gene amplification of blood showed 67% (150/223) of common vampire bats (*Desmodus rotundus*) were infected by hemoplasmas. Sequencing of the 16S rRNA gene amplicons revealed three novel genotypes that were phylogenetically related but not identical to hemoplasmas described from other (non-vampire) bat species, rodents, humans, and non-human primates. Hemoplasma prevalence in vampire bats was highest in non-reproductive and young individuals, did not differ by country, and was relatively stable over time (i.e., endemic). Metagenomics from pooled *D. rotundus* saliva from Peru detected non-hemotropic *Mycoplasma* species and hemoplasma genotypes phylogenetically similar to those identified in blood, providing indirect evidence for potential direct transmission of hemoplasmas through biting or social contacts. This study demonstrates vampire bats host several novel hemoplasmas and sheds light on risk factors for infection and basic transmission routes. Given the high frequency of direct contacts that arise when vampire bats feed on humans, domestic animals, and wildlife, the potential of these bacteria to be transmitted between species should be investigated in future work.

Key words: Chiroptera, *Desmodus rotundus*, hemoplasmas, metagenomics, phylogenetic analysis, 16S rRNA gene, wildlife.

* Author for correspondence: D. V. Volokhov, The US Food and Drug Administration, Center for Biologics Evaluation and Research, Laboratory of Method Development, 10903 New Hampshire Avenue, Building 52, Room 1120, Silver Spring, MD 20993-0002, USA.
(Email: dmitriy.volokhov@fda.hhs.gov)
† Both authors contributed equally to this work.

INTRODUCTION

Bats (Order: Chiroptera) have been widely studied as reservoir hosts for pathogens of concern for human and animal health [1,2], with particular attention paid to RNA viruses in the *Coronaviridae*, *Filoviridae*, *Rhabdoviridae*, and *Paramyxoviridae*

List of References

Abeles SR, Robles-Sikisaka R, Ly M *et al.* (2014) Human oral viruses are personal, persistent and gender-consistent. *The ISME Journal*, **8**, 1753–1767.

Adams MJ, Lefkowitz EJ, King AMQ *et al.* (2017) Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). *Archives of Virology*, **162**, 2505–2538.

Adriaenssens EM, Kramer R, Van Goethem MW *et al.* (2017) Environmental drivers of viral community composition in Antarctic soils identified by viromics. *Microbiome*, **5**, 1–14.

Afelt A, Lacroix A, Zawadzka-Pawlewska U *et al.* (2018) Distribution of bat-borne viruses and environment patterns. *Infection, Genetics and Evolution*, **58**, 181–191.

Aguilar-Setien A, Loza-Rubio E, Salas-Rojas M *et al.* (1999) Salivary excretion of rabies virus by healthy vampire bats. *Epidemiology and Infection*, **133**, 517–522.

Aguirre LF, Herrel A, Van Damme R, Matthysen E (2003) The implications of food hardness for diet in bats. *Functional Ecology*, **17**, 201–212.

Alam MM, Kobayashi N, Ishino M *et al.* (2006) Genetic analysis of an ADRV-N-like novel rotavirus strain B219 detected in a sporadic case of adult diarrhea in Bangladesh. *Archives of Virology*, **152**, 199–208.

Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J (2001) A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 11609–11614.

Allendorf SD, Cortez A, Heinemann MB *et al.* (2012) Rabies virus distribution in tissues and molecular characterization of strains from naturally infected non-hematophagous bats. *Virus Research*, **165**, 119–125.

Almeida MF, Martorelli LFA, Aires CC *et al.* (2005) Experimental rabies infection in haematophagous bats Desmodus rotundus. *Epidemiology and Infection*, **133**, 523–527.

Altizer S, Bartel R, Han BA (2011) Animal Migration and Infectious Disease Risk. *Science*, **331**, 296–302.

Altizer S, Nunn CL, Lindenfors P (2007) Do threatened hosts have fewer parasites? A comparative study in primates. *Journal of Animal Ecology*, **76**, 304–314.

Amato KR, Yeoman CJ, Kent A *et al.* (2013) Habitat degradation impacts black howler monkey (Alouatta pigra) gastrointestinal microbiomes. *The ISME Journal*, **7**, 1344–1353.

Amimo JO, Zowalaty El ME, Githae D *et al.* (2016) Metagenomic analysis demonstrates the diversity of the fecal virome in asymptomatic pigs in East Africa. *Archives of Virology*, **161**, 887-897.

Amman BR, Carroll SA, Reed ZD *et al.* (2012) Seasonal Pulses of Marburg Virus Circulation in Juvenile Rousettus aegyptiacus Bats Coincide with Periods of Increased Risk of Human Infection. *PLoS Pathogens*, **8**, e1002877–11.

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.

Anthony SJ, Gilardi K, Menachery VD *et al.* (2017a) Further Evidence for Bats as the Evolutionary Source of Middle East Respiratory Syndrome Coronavirus. *mBio*, **8**, e00373–17–13.

Anthony SJ, Islam A, Johnson C *et al.* (2015) Non-random patterns in viral diversity.

Nature Communications, **6**, 1–7.

Anthony SJ, Johnson CK, Greig DJ *et al.* (2017b) Global patterns in coronavirus diversity. *Virus Evolution*, **3**, 1–15.

Anthony SJ, Ojeda-Flores R, Rico-Chavez O *et al.* (2013) Coronaviruses in bats from Mexico. *Journal of General Virology*, **94**, 1028–1038.

Antunes A, Troyer JL, Roelke ME *et al.* (2008) The Evolutionary Dynamics of the Lion Panthera leo Revealed by Host and Viral Population Genomics. *PLoS Genetics*, **4**, e1000251.

Arellano-Sota C (1988) Biology, Ecology, and Control of the Vampire Bat. *Reviews of Infectious Diseases*, **10**, S615–S619.

Arneberg P, Skorping A, Grenfell B, Read AF (1998) Host densities as determinants of abundance in parasite communities. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **265**, 1283–1289.

Arriero E, Moller AP (2008) Host ecology and life-history traits associated with blood parasite species richness in birds. *Journal of Evolutionary Biology*, **21**, 1504–1513.

Arslan D, Legendre M, Seltzer V, Abergel C, Claverie J-M (2011) Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 17486–17491.

Asano KM, Gregori F, Hora AS *et al.* (2016) Group A rotavirus in Brazilian bats: description of novel T15 and H15 genotypes. *Archives of Virology*, **161**, 3225–3230.

Baillargeon S, Rivest L-P (2007) The Rcapture Package: Loglinear Models for Capture-Recapture in R. *Journal of Statistical Software*, **19**.

Baker KS, Leggett RM, Bexfield NH *et al.* (2013) Metagenomic study of the viruses of African straw-coloured fruit bats: Detection of a chiropteran poxvirus and isolation of a novel adenovirus. *Virology*, **441**, 95–106.

Baker ML, Schountz T, Wang LF (2012) Antiviral Immune Responses of Bats: A Review. *Zoonoses and Public Health*, **60**, 104–116.

Baker RJ, Hoofer SR, Porter CA, Van Den Bussche RA (2003) Diversification among New World leaf-nosed bats: an evolutionary hypothesis and classification inferred from digenomic congruence of DNA sequence. *Occasional Papers Museum of Texas Tech University*, **230**.

Bankevich A, Nurk S, Antipov D *et al.* (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, **19**, 455–477.

Barelli C, Albanese D, Donati C *et al.* (2015) Habitat fragmentation is associated to gut microbiota diversity of an endangered primate: implications for conservation. *Scientific Reports*, **5**, 1–12.

Barrett LG, Thrall PH, Burdon JJ, Linde CC (2008) Life history determines genetic structure and evolutionary potential of host–parasite interactions. *Trends in Ecology and Evolution*, **23**, 678–685.

Bartlett MS (1957) Measles periodicity and community size. *Journal of the Royal Statistical Society. Series A (General)*, **120**, 48–70.

Bartoń K (2018) MuMIn: Multi-Model Inference. R package version 1.40.4. https://CRAN.R-project.org/package=MuMIn

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**.

Bauermann FV, Ridpath JF (2015) HoBi-like viruses--the typical 'atypical bovine

pestivirus'. *Animal Health Research Reviews*, **16**, 64–69.

Bányai K, Kemenesi G, Budinski I *et al.* (2017) Candidate new rotavirus species in Schreiber's bats, Serbia. *Infection, Genetics and Evolution*, **48**, 19–26.

Becker DJ, Czirják GÁ, Volokhov DV *et al.* (2018) Livestock abundance predicts vampire bat demography, immune profiles and bacterial infection risk. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **373**, 20170089–12.

Benavides JA, Huchard E, Pettorelli N *et al.* (2011) From parasite encounter to infection: Multiple-scale drivers of parasite richness in a wild social primate population. *American Journal of Physical Anthropology*, **147**, 52–63.

Benavides JA, Valderrama W, Streicker DG (2016) Spatial expansions and travelling waves of rabies in vampire bats. *Proceedings of the Royal Society B: Biological Sciences*, **283**, 20160328.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B*, **57**, 289–300.

Bernardo P, Charles-Dominique T, Barakat M *et al.* (2018) Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *The ISME Journal*, **12**, 173–184.

Betsem E, Rua R, Tortevoye P, Froment A, Gessain A (2011) Frequent and Recent Human Acquisition of Simian Foamy Viruses Through Apes' Bites in Central Africa. *PLoS Pathogens*, **7**, e1002306–15.

Beugin M-P, Gayet T, Pontier D, Devillard S, Jombart T (2018) A fast likelihood solution to the genetic clustering problem. *Methods in Ecology and Evolution*, **21**, 199–11.

Biek R, Real LA (2010) The landscape genetics of infectious disease emergence and spread. *Molecular Ecology*, **19**, 3515–3531.

Biek R, Drummond AJ, Poss M (2006) A Virus Reveals Population Structure and Recent Demographic History of Its Carnivore Host. *Science*, **311**, 535–538.

Bivand R, Keitt T, Rowlingson B (2017) rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.2-16. https://CRAN.R-project.org/package=rgdal

Bivand R, Lewin-Koh N (2017) maptools: Tools for Reading and Handling Spatial Objects. R package version 0.9-2. https://CRAN.R-project.org/package=maptools

Blackwood JC, Streicker DG, Altizer S, Rohani P (2013) Resolving the roles of immunity, pathogenesis, and immigration for rabies persistence in vampire bats. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 20837–20842.

Blehert DS, Hicks AC, Behr M *et al.* (2009) Bat White-Nose Syndrome: An Emerging Fungal Pathogen? *Science*, **323**, 227.

Blekhman R, Goodrich JK, Huang K *et al.* (2015) Host genetic variation impacts microbiome composition across human body sites. *Genome Biology*, **16**, 191.

Bobbie CB, Mykytczuk NCS, Schulte-Hostedde AI (2017) Temporal variation of the microbiome is dependent on body region in a wild mammal (Tamiasciurus hudsonicus). *FEMS Microbiology Ecology*, **93**.

Bobrowiec PED, Lemes MR, Gribel R (2015) Prey preference of the common vampire bat (Desmodus rotundus, Chiroptera) using molecular analysis. *Journal of Mammalogy*, **96**, 54–63.

Bodewes R, van der Giessen J, Haagmans BL, Osterhaus ADME, Smits SL (2013) Identification of Multiple Novel Viruses, Including a Parvovirus and a Hepevirus,

in Feces of Red Foxes. *Journal of Virology*, **87**, 7758–7764.

Bohmann K, Gopalakrishnan S, Nielsen M *et al.* (2018) Using DNA metabarcoding for simultaneous inference of common vampire bat diet and population structure. *Molecular Ecology Resources*, 1–34.

Bolnick DI, Snowberg LK, Hirsch PE *et al.* (2014) Individual diet has sex-dependent effects on vertebrate gut microbiota. *Nature Communications*, **5**, 4500.

Bordes F, Guégan J-F, Morand S (2011) Microparasite species richness in rodents is higher at lower latitudes and is associated with reduced litter size. *Oikos*, **120**, 1889–1896.

Brandão PE, Scheffer K, Villarreal LY *et al.* (2008) A Coronavirus Detected in the Vampire Bat Desmodus rotundus. *The Brazilian Journal of Infectious Diseases*, **12**, 466–468.

Breitbart M, Rohwer F (2005) Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *BioTechniques*, **39**, 729–736.

Breitbart M, Hewson I, Felts B *et al.* (2003) Metagenomic Analyses of an Uncultured Viral Community from Human Feces. *Journal of Bacteriology*, **185**, 6220–6223.

Brierley L, Vonhof MJ, Olival KJ, Daszak P, Jones KE (2016) Quantifying Global Drivers of Zoonotic Bat Viruses: A Process-Based Perspective. *American Naturalist*, **187**, E53–E64.

Briese T, Kapoor A, Mishra N *et al.* (2015) Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. *mBio*, **6**, e01491–15.

Briese T, Paweska JT, McMullan LK *et al.* (2009) Genetic Detection and Characterization of Lujo Virus, a New Hemorrhagic Fever–Associated Arenavirus from Southern Africa. *PLoS Pathogens*, **5**, e1000455–8.

Brook CE, Dobson AP (2015) Bats as "special" reservoirs for emerging zoonotic pathogens. *Trends in Microbiology*, **23**, 172–180.

Brum JR, Ignacio-Espinoza JC, Kim E-H *et al.* (2016) Illuminating structural proteins in viral "dark matter" with metaproteomics. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 2436–2441.

Brunker K, Hampson K, Horton DL, Biek R (2012) Integrating the landscape epidemiology and genetics of RNA viruses: rabies in domestic dogs as a model. *Parasitology*, **139**, 1899–1913.

Buchfink B, Xie C, Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, **12**, 59–60.

Burnham KP, Anderson DR (2002) *Model selection and multimodel inference: a practical information-theoretic approach*. Springer-Verlag, New York, New York, USA.

Cade BS (2015) Model averaging and muddled multimodel inferences. *Ecology*, **96**, 2370–2382.

Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T (2006) Bats: Important Reservoir Hosts of Emerging Viruses. *Clinical Microbiology Reviews*, **19**, 531–545.

Cardona S, Eck A, Cassellas M *et al.* (2012) Storage conditions of intestinal microbiota matter in metagenomic analysis. *BMC Microbiology*, **12**, 158.

Carini A (1911) Sur une grande épizootie de rage. *Annals of the Institute Pasteur,* **25**, 843–846.

Carrillo-Araujo M, Tas N, Alcantara-Hernandez RJ *et al.* (2015) Phyllostomid bat microbiome composition is associated to host phylogeny and feeding strategies.

Frontiers in Microbiology, **6**, 447.

Carrington CVF, Foster JE, Zhu HC *et al.* (2008) Detection and Phylogenetic Analysis of Group 1 Coronaviruses in South American Bats. *Emerging Infectious Diseases*, **14**, 1890–1893.

Carroll D, Daszak P, Wolfe ND *et al.* (2018) The Global Virome Project. *Science*, **359**, 872–874.

Carter G, Leffer L (2015) Social Grooming in Bats: Are Vampire Bats Exceptional? *PLoS ONE*, **10**, e0138430–11.

Carter GG, Wilkinson GS (2012) Food sharing in vampire bats: reciprocal help predicts donations more than relatedness or harassment. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20122573–20122573.

Chamberlain S (2017) Interface to the Global "Biodiversity" Information Facility API. R package version 0.9.9. https://CRAN.R-project.org/package=rgbif

Chandler JA, Liu RM, Bennett SN (2015) RNA shotgun metagenomic sequencing of northern California (USA) mosquitoes uncovers viruses, bacteria, and fungi. *Frontiers in Microbiology*, **06**, 403–16.

Chapman DH (1951) Some properties of the hypergeometric distribution with applications to zoological censuses. *Univ Calif Public Stat*, **1**, 131–160.

Chapuis MP, Estoup A (2007) Microsatellite Null Alleles and Estimation of Population Differentiation. *Molecular Biology and Evolution*, **24**, 621–631.

Chow C-ET, Suttle CA (2015) Biogeography of Viruses in the Sea. *Annual Review of Virology*, **2**, 41–66.

Christian N, Whitaker BK, Clay K (2015) Microbiomes: unifying animal and plant systems through the lens of community ecology theory. *Frontiers in Microbiology*, **6**, 100–15.

Conceição-Neto N, Godinho R, Álvares F *et al.* (2017) Viral gut metagenomics of sympatric wild and domestic canids, and monitoring of viruses: Insights from an endangered wolf population. *Ecology and Evolution*, **367**, 2840.

Conceição-Neto N, Zeller M, Lefrère H *et al.* (2015) Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Scientific Reports*, **5**, 1–14.

Condori-Condori RE, Streicker DG, Cabezas-Sanchez C, Velasco-Villa A (2013) Enzootic and Epizootic Rabies Associated with Vampire Bats, Peru. *Emerging Infectious Diseases*, **19**, 1463–1469.

Cormack RM (1989) Log-Linear Models for Capture-Recapture. *Biometrics*, **45**, 395.

Corman VM, Rasche A, Diallo TD *et al.* (2013) Highly diversified coronaviruses in neotropical bats. *Journal of General Virology*, **94**, 1984–1994.

Corthals A, Martin A, Warsi OM *et al.* (2015) From the Field to the Lab: Best Practices for Field Preservation of Bat Specimens for Molecular Analyses. *PLoS ONE*, **10**, e0118994.

Costello EK, Lauber CL, Hamady M *et al.* (2009) Bacterial Community Variation in Human Body Habitats Across Space and Time. *Science*, **326**, 1694–1697.

Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA (2012) The application of ecological theory toward an understanding of the human microbiome. *Science*, **336**, 1255–1262.

Cox-Foster DL, Conlan S, Holmes EC *et al.* (2007) A Metagenomic Survey of Microbes in Honey Bee Colony Collapse Disorder. *Science*, **318**, 283–287.

Côté H, Garant D, Robert K, Mainguy J, Pelletier F (2012) Genetic structure and rabies spread potential in raccoons: the role of landscape barriers and sex-biased dispersal. *Evolutionary Applications*, **5**, 393–404.

Creer S, Deiner K, Frey S *et al.* (2016) The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, **7**, 1008–1018.

Cui J, Han N, Streicker D *et al.* (2007) Evolutionary relationships between bat coronaviruses and their hosts. *Emerging Infectious Diseases*, **13**, 1526–1532.

Dacheux L, Cervantes-Gonzalez M, Guigon G *et al.* (2014) A Preliminary Study of Viral Metagenomics of French Bat Species in Contact with Humans: Identification of New Mammalian Viruses. *PLoS ONE*, **9**, e87194.

Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics,* **27**, 1164-1165.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772–772.

Daszak P (2000) Emerging Infectious Diseases of Wildlife-- Threats to Biodiversity and Human Health. *Science*, **287**, 443–449.

Davenport ER, Sanders JG, Song SJ *et al.* (2017) The human microbiome in evolution. *BMC Biology*, **15**, 1–12.

Davies TJ, Pedersen AB (2008) Phylogeny and geography predict pathogen community similarity in wild primates and humans. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 1695–1701.

de Souza Aguiar LM, Antonini Y (2011) Descriptive ecology of bat flies (Diptera: Hippoboscoidea) associated with vampire bats (Chiroptera: Phyllostomidae) in the cerrado of Central Brazil. *Memórias do Instituto Oswaldo Cruz*, **106**, 170–176.

de Thoisy B, Bourhy H, Delaval M *et al.* (2016) Bioecological Drivers of Rabies Virus Circulation in a Neotropical Bat Community. *PLoS Neglected Tropical Diseases*, **10**, e0004378.

Delpietro HA, Marchevsky N, Simonetti E (1992) Relative population densities and predation of the common vampire bat (*Desmodus rotundus*) in natural and cattle-raising areas in north-east Argentina. *Preventive Veterinary Medicine*, **14**, 13–20.

Delpietro HA, Russo RG, Carter GG, Lord RD, Delpietro GL (2017) Reproductive seasonality, sex ratio and philopatry in Argentina's common vampire bats. *Royal Society Open Science*, **4**, 160959.

DeNault LK, McFarlane DA (1995) Reciprocal altruism between male vampire bats, *Desmodus rotundus*. *Animal Behaviour*, **49**, 855–856.

Dietrich M, Kearney T, Seamark ECJ, Markotter W (2017) The excreted microbiota of bats: evidence of niche specialisation based on multiple body habitats. *FEMS Microbiology Letters*, **364**, fnw284.

Dietrich M, Wilkinson DA, Benlali A *et al.* (2015) Leptospira and paramyxovirus infection dynamics in a bat maternity enlightens pathogen maintenance in wildlife. *Environmental Microbiology*, **17**, 4280–4289.

Dietzgen RG, Kondo H, Goodin MM, Kurath G, Vasilakis N (2017) The family Rhabdoviridae: mono- and bipartite negative-sense RNA viruses with diverse genome organization and common evolutionary origins. *Virus Research*, **227**, 158–170.

Dinsdale EA, Edwards RA, Hall D *et al.* (2008) Functional metagenomic profiling of nine biomes. *Nature*, **452**, 629–632.

Dominguez SR, O'Shea TJ, Oko LM, Holmes KV (2007) Detection of group 1 coronaviruses in bats in North America. *Emerging Infectious Diseases*, **13**, 1295–1300.

Donaldson EF, Haskew AN, Gates JE *et al.* (2010) Metagenomic Analysis of the Viromes of Three North American Bat Species: Viral Diversity among Different

Bat Species That Share a Common Habitat. *Journal of Virology*, **84**, 13004–13018.

Drexler JF, Corman VM, Müller MA *et al.* (2012a) Bats host major mammalian paramyxoviruses. *Nature Communications*, **3**, 796–12.

Drexler JF, Corman VM, Wegner T *et al.* (2011) Amplification of Emerging Viruses in a Bat Colony. *Emerging Infectious Diseases*, **17**, 449–456.

Drexler JF, Geipel A, König A *et al.* (2013) Bats carry pathogenic hepadnaviruses antigenically related to hepatitis B virus and capable of infecting human hepatocytes. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 16151–16156.

Drexler JF, Gloza-Rausch F, Glende J *et al.* (2010) Genomic Characterization of Severe Acute Respiratory Syndrome-Related Coronavirus in European Bats and Classification of Coronaviruses Based on Partial RNA-Dependent RNA Polymerase Gene Sequences. *Journal of Virology*, **84**, 11336–11349.

Drexler JF, Seelen A, Corman VM *et al.* (2012b) Bats Worldwide Carry Hepatitis E Virus-Related Viruses That Form a Putative Novel Genus within the Family Hepeviridae. *Journal of Virology*, **86**, 9134–9147.

Druce J, Garcia K, Tran T, Papadakis G, Birch C (2012) Evaluation of Swabs, Transport Media, and Specimen Transport Conditions for Optimal Detection of Viruses by PCR. *Journal of Clinical Microbiology*, **50**, 1064–1065.

Dudaniec RY, Tesson SVM (2016) Applying landscape genetics to the microbial world. *Molecular Ecology*, **25**, 3266–3275.

Duerkop BA, Hooper LV (2013) Resident viruses and their interactions with the immune system. *Nature Immunology*, **14**, 654–659.

Dunn RR, Davies TJ, Harris NC, Gavin MC (2010) Global drivers of human pathogen richness and prevalence. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 2587–2595.

Dutilh BE, Schmieder R, Nulton J *et al.* (2012) Reference-independent comparative metagenomics using cross-assembly: crAss. *Bioinformatics*, **28**, 3225–3231.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Edwards RA, Rohwer F (2005) Viral metagenomics. *Nature Reviews Microbiology*, **3**, 504–510.

Elena SF, Dopazo J, Flores R, Diener TO, Moya A (1991) Phylogeny of viroids, viroidlike satellite RNAs, and the viroidlike domain of hepatitis delta virus RNA. *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 5631–5634.

Ellis JS, Gilbey J, Armstrong A *et al.* (2011) Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (Salmo salar L.). *Genetica*, **139**, 353–367.

Escalera-Zamudio M, Rojas-Anaya E, Kolokotronis S-O *et al.* (2016) Bats, Primates, and the Evolutionary Origins and Diversification of Mammalian Gammaherpesviruses. *mBio*, **7**, e01425–16–9.

Escalera-Zamudio M, Taboada B, Rojas-Anaya E *et al.* (2017) Viral Communities Among Sympatric Vampire Bats and Cattle. *EcoHealth*, **111**, 1–11.

Escalera-Zamudio M, Zepeda Mendoza ML, Heeger F *et al.* (2015) A novel endogenous betaretrovirus in the common vampire bat (Desmodus rotundus) suggests multiple independent infection and cross-species transmission events. *Journal of Virology*, **89**, 5180-5184.

Esona MD, Mijatovic-Rustempasic S, Conrardy C *et al.* (2010) Reassortant Group A Rotavirus from Straw-colored Fruit Bat (Eidolon helvum). *Emerging Infectious Diseases*, **16**, 1844–1852.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Ezenwa VO, Price SA, Altizer S, Vitone ND, Cook KC (2006) Host traits and parasite species richness in even and odd-toed hoofed mammals, Artiodactyla and Perissodactyla. *Oikos*, **115**, 526–536.

Fagrouch Z, Sarwari R, Lavergne A *et al.* (2012) Novel polyomaviruses in South American bats and their relationship to other members of the family Polyomaviridae. *Journal of General Virology*, **93**, 2652–2657.

Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation*, **61**, 1–10.

Falush D, Stephens M, Pritchard JK (2003a) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Falush D, Wirth T, Linz B *et al.* (2003b) Traces of human migrations in Helicobacter pylori populations. *Science*, **299**, 1582–1585.

Fancello L, Raoult D, Desnues C (2012) Computational tools for viral metagenomics and their application in clinical research. *Virology*, **434**, 162–174.

Fancello L, Trape S, Robert C *et al.* (2013) Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. *The ISME Journal*, **7**, 359–369.

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302–4315.

Field H, Young P, Yob JM *et al.* (2001) The natural history of Hendra and Nipah viruses. *Microbes and Infection*, **3**, 307–314.

Fierer N, Breitbart M, Nulton J *et al.* (2007) Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Applied and Environmental Microbiology*, **73**, 7059–7066.

Fierer N, Lauber CL, Zhou N *et al.* (2010) Forensic identification using skin bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 6477–6481.

Fierer N, Leff JW, Adams BJ *et al.* (2012) Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 21390–21395.

Fierer N, McCain CM, Meir P *et al.* (2011) Microbes do not follow the elevational diversity patterns of plants and animals. *Ecology*, **92**, 797–804.

Fish EN (2008) The X-files in immunity: sex-based differences predispose immune responses. *Nature Reviews Immunology*, **8**, 737–744.

Fisher MC, Henk DA, Briggs CJ *et al.* (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature*, **484**, 186–194.

Food and Agriculture Organization of the United Nations (2012) GLIPHA. Rome, Italy: FAO.

Forster JL, Harkin VB, Graham DA, McCullough SJ (2008) The effect of sample type, temperature and RNAlater™ on the stability of avian influenza virus RNA. *Journal of Virological Methods*, **149**, 190–194.

Fox J, Weisberg S (2011) *An R Companion to Applied Regression*. Sage, Thousand Oaks, CA.

Frank HK, Mendenhall CD, Judson SD, Daily GC, Hadly EA (2016) Anthropogenic impacts on Costa Rican bat parasitism are sex specific. *Ecology and Evolution*, **6**, 4898–4909.

Franzosa EA, Huang K, Meadow JF *et al.* (2015) Identifying personal microbiomes using metagenomic codes. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, E2930–E2938.

Frick WF, Cheng TL, Langwig KE *et al.* (2017) Pathogen dynamics during invasion and establishment of white-nose syndrome explain mechanisms of host persistence. *Ecology*, **98**, 624–631.

Frick WF, Pollock JF, Hicks AC *et al.* (2010) An Emerging Disease Causes Regional Population Collapse of a Common North American Bat Species. *Science*, **329**, 679–682.

Fuhrman JA (1999) Marine viruses and their biogeochemical and ecological effects. *Nature*, **399**, 541–548.

Gagnieur L, Cheval J, Gratigny M *et al.* (2014) Unbiased analysis by high throughput sequencing of the viral diversity in fetal bovine serum and trypsin used in cell culture. *Biologicals*, **42**, 145–152.

Garrido-Olvera L, Arita HT, Pérez-Ponce De León G (2012) The influence of host ecology and biogeography on the helminth species richness of freshwater fishes in Mexico. *Parasitology*, **139**, 1652–1665.

Gay N, Olival KJ, Bumrungsri S *et al.* (2014) Parasite and viral species richness of Southeast Asian bats: Fragmentation of area distribution matters. *International Journal for Parasitology*, **3**, 161–170.

Ge X, Li Y, Yang X *et al.* (2012) Metagenomic Analysis of Viruses from Bat Fecal Samples Reveals Many Novel Viruses in Insectivorous Bats in China. *Journal of Virology*, **86**, 4620–4630.

George DB, Webb CT, Farnsworth ML *et al.* (2011) Host and viral ecology determine bat rabies seasonality and maintenance. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 10208–10213.

Ghatak S, Muthukumaran RB, Nachimuthu SK (2013) A Simple Method of Genomic DNA Extraction from Human Samples for PCR-RFLP Analysis. *Journal of Biomolecular Techniques*, **24**, 224–231.

Ghersi BM, Jia H, Aiewsakun P *et al.* (2015) Wide distribution and ancient evolutionary history of simian foamy viruses in New World primates. *Retrovirology*, **12**, 1–19.

Giersch K, Helbig M, Volz T *et al.* (2014) Persistent hepatitis D virus mono-infection in humanized mice is efficiently converted by hepatitis B virus to a productive co-infection. *Journal of Hepatology*, **60**, 538–544.

Gilbert AT, Fooks AR, Hayman DTS *et al.* (2013) Deciphering Serology to Understand the Ecology of Infectious Diseases in Wildlife. *EcoHealth*, **10**, 298–313.

Gillespie TR, Chapman CA, Greiner EC (2005) Effects of logging on gastrointestinal parasite infections and infection risk in African primates. *Journal of Applied Ecology*, **42**, 699–707.

Gomez A, Petrzelkova K, Yeoman CJ *et al.* (2015) Gut microbiome composition and metabolomic profiles of wild western lowland gorillas (Gorilla gorilla gorilla) reflect host ecology. *Molecular Ecology*, **24**, 2551–2565.

Gonçalves MA, Sa-Neto RJ, Brazil TK (2002) Outbreak of aggressions and

transmission of rabies in human beings by vampire bats in northeastern Brazil. *Revista da Sociedade Brasileira de Medicina Tropical*, **35**, 453–460.

Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

Gottdenker NL, Streicker DG, Faust CL, Carroll CR (2014) Anthropogenic Land Use Change and Infectious Diseases: A Review of the Evidence. *EcoHealth*, **11**, 619-632.

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Goudet J, Jombart T (2015) hierfstat: Estimation and Tests of Hierarchical F-Statistics. *Molecular Ecology Notes*, **5**, 184–186.

Greenhall AM, Joermann G, Schmidt U (1983) *Desmodus rotundus*. In: *Mammalian Species*, pp. 1–6. American Society of Mammalogists, New York.

Grenfell B, Harwood J (1997) (Meta)population dynamics of infectious diseases. *Trends in Ecology and Evolution*, **12**, 395–399.

Greninger AL, Chen EC, Sittler T *et al.* (2010) A Metagenomic Analysis of Pandemic Influenza A (2009 H1N1) Infection in Patients from North America. *PLoS ONE*, **5**, e13381–16.

Grieneisen LE, Livermore J, Alberts S, Tung J, Archie EA (2017) Group Living and Male Dispersal Predict the Core Gut Microbiome in Wild Baboons. *Integrative and Comparative Biology*, **57**, 770–785.

Guernier V, Hochberg ME, Guégan J-F (2004) Ecology Drives the Worldwide Distribution of Human Diseases. *PLoS Biology*, **2**, e141.

Guilhaumon F, Krasnov BR, Poulin R, Shenbrot GI, Mouillot D (2011) Latitudinal mismatches between the components of mammal-flea interaction networks. *Global Ecology and Biogeography*, **21**, 725–731.

Hackenbrack N, Rogers MB, Ashley RE *et al.* (2017) Evolution and Cryo-electron Microscopy Capsid Structure of a North American Bat Adenovirus and Its Relationship to Other Mastadenoviruses. *Journal of Virology*, **91**, e01504–16–18.

Hall RJ, Wang J, Todd AK *et al.* (2014) Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *Journal of Virological Methods*, **195**, 194–204.

Han BA, Schmidt JP, Alexander LW *et al.* (2016) Undiscovered Bat Hosts of Filoviruses. *PLoS Neglected Tropical Diseases*, **10**, e0004815–10.

Han GZ, Worobey M (2012) An Endogenous Foamy Virus in the Aye-Aye (Daubentonia madagascariensis). *Journal of Virology*, **86**, 7696–7698.

Hannigan GD, Duhaime MB, Ruffin MT IV, Koumpouras CC, Schloss PD (2017) Viral and Bacterial Communities of Colorectal Cancer. *BioRxiv* 152868.

Hannigan GD, Meisel JS, Tyldsley AS *et al.* (2015) The Human Skin Double-Stranded DNA Virome: Topographical and Temporal Diversity, Genetic Enrichment, and Dynamic Associations with the Host Microbiome. *mBio*, **6**, e01578–15–13.

Harris NC, Dunn RR (2010) Using host associations to predict spatial patterns in the species richness of the parasites of North American carnivores. *Ecology Letters*, **13**, 1411–1418.

Hassell JM, Begon M, Ward MJ, Fèvre EM (2017) Urbanization and Disease Emergence: Dynamics at the Wildlife–Livestock–Human Interface. *Trends in Ecology and Evolution*, **32**, 55–67.

Hayman DTS (2016) Bats as Viral Reservoirs. *Annual Review of Virology*, **3**, 77–99.

Hayman DTS, Bowen RA, Cryan PM *et al.* (2012) Ecology of Zoonotic Infectious Diseases in Bats: Current Knowledge and Future Directions. *Zoonoses and Public Health*, **60**, 2–21.

He B, Li Z, Yang F *et al.* (2013a) Virome Profiling of Bats from Myanmar by Metagenomic Analysis of Tissue Samples Reveals More Novel Mammalian Viruses. *PLoS ONE*, **8**, e61950.

He B, Yang F, Yang W *et al.* (2013b) Characterization of a Novel G3P[3] Rotavirus Isolated from a Lesser Horseshoe Bat: a Distant Relative of Feline/Canine Rotaviruses. *Journal of Virology*, **87**, 12357–12366.

He S, Wurtzel O, Singh K *et al.* (2010) Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nature Methods*, **7**, 807–812.

Hedrick PW (2005) A Standardized Genetic Differentiation Measure. *Evolution*, **59**, 1633.

Hijmans RJ (2017) raster: Geographic Data Analysis and Modeling. R package version 2.6-7. https://CRAN.R-project.org/package=raster.

Hoffmann B, Scheuch M, Hoeper D *et al.* (2012) Novel Orthobunyavirus in Cattle, Europe, 2011. *Emerging Infectious Diseases*, **18**, 469–472.

Honkavuori KS, Shivaprasad HL, Williams BL *et al.* (2008) Novel Borna Virus in Psittacine Birds with Proventricular Dilatation Disease. *Emerging Infectious Diseases*, **14**, 1883–1886.

Hothorn T, Bretz F, Westfall P (2008) Simultaneous Inference in General Parametric Models. *Biometrical Journal*, **50**, 346–363.

Howe A, Ringus DL, Williams RJ *et al.* (2015) Divergent responses of viral and bacterial communities in the gut microbiome to dietary disturbances in mice. *The ISME Journal*, **10**, 1217–1227.

Hu B, Ge X, Wang L-F, Shi Z (2015) Bat origin of human coronaviruses. *Virology Journal*, **12**, 1–10.

Hu B, Zeng L-P, Yang X-L *et al.* (2017) Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathogens*, **13**, e1006698–27.

Huang C-R, Lo SJ (2010) Evolution and Diversity of the Human Hepatitis D Virus Genome. *Advances in Bioinformatics*, **2010**, 1–9.

Huang S, Bininda-Emonds ORP, Stephens PR, Gittleman JL, Altizer S (2013) Phylogenetically related and ecologically similar carnivores harbour similar parasite assemblages. *Journal of Animal Ecology*, **83**, 671–680.

Huang S, Stephens PR, Gittleman JL (2012) Traits, trees and taxa: global dimensions of biodiversity in mammals. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **279**, 4997–5003.

Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJ (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. *Applied and Environmental Microbiology*, **67**, 4399–4406.

Hulo C, de Castro E, Masson P, *et al.* (2011) ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Research*, **39**, D576–D582.

The Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature*, **486**, 207–214.

Hurwitz BL, Westveld AH, Brum JR, Sullivan MB (2014) Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 10714–10719.

Huson DH, Beier S, Flade I *et al.* (2016) MEGAN Community Edition - Interactive

Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Computational Biology*, **12**, e1004957–12.

IUCN (2017) *The IUCN Red List of Threatened Species.* Version 2017-3. http://www.iucnredlist.org

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jensen C, Johnson FB (1994) Comparison of various transport media for viability maintenance of herpes simplex virus, respiratory syncytial virus, and adenovirus. *Diagnostic Microbiology and Infectious Disease*, **19**, 137–142.

Jiang S, Ji S, Tang Q *et al.* (2008) Molecular characterization of a novel adult diarrhoea rotavirus strain J19 isolated in China and its significance for the evolution and origin of group B rotaviruses. *The Journal of General Virology*, **89**, 2622–2629.

Johnson FB (1990) Transport of Viral Specimens. *Clinical Microbiology Reviews*, **3**, 120–131.

Johnson N, Aréchiga-Ceballos N, Aguilar-Setien A (2014) Vampire Bat Rabies: Ecology, Epidemiology and Control. *Viruses*, **6**, 1911–1928.

Johnson PCD, Haydon DT (2008) Software for Quantifying and Simulating Microsatellite Genotyping Error. *Bioinformatics and biology insights*, **1**, BBI.S373.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

Jones KE, Patel NG, Levy MA *et al.* (2008) Global trends in emerging infectious diseases. *Nature*, **451**, 990–993.

Jost L (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Jovel J, Patterson J, Wang W *et al.* (2016) Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. *Frontiers in Microbiology*, **7**, 253–17.

Kading RC, Gilbert AT, Mossel EC *et al.* (2013) Isolation and molecular characterization of Fikirini rhabdovirus, a novel virus from a Kenyan bat. *Journal of General Virology*, **94**, 2393–2398.

Kamiya T, O'Dwyer K, Nakagawa S, Poulin R (2013) What determines species richness of parasitic organisms? A meta-analysis across animal, plant and fungal hosts. *Biological Reviews*, **89**, 123–134.

Kamiya T, O'Dwyer K, Nakagawa S, Poulin R (2014) Host diversity drives parasite diversity: meta-analytical insights into patterns and causal mechanisms. *Ecography*, **37**, 689–697.

Karesh WB, Dobson A, Lloyd-Smith JO *et al.* (2012) Ecology of zoonoses: natural and unnatural histories. *The Lancet*, **380**, 1936–1945.

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**, 3059-3066.

Katzourakis A, Gifford RJ, Tristem M, Gilbert MTP, Pybus OG (2009) Macroevolution of Complex Retroviruses. *Science*, **325**, 1512–1512.

Kearse M, Moir R, Wilson A *et al.* (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Kelly TR, Karesh WB, Johnson CK *et al.* (2017) One Health proof of concept: Bringing a transdisciplinary approach to surveillance for zoonotic viruses at the human-wild animal interface. *Preventive Veterinary Medicine*, **137**, 112–118.

Kembel SW, Cowan PD, Helmus MR *et al.* (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, **26**, 1463–1464.

Kemenesi G, Zhang D, Marton S *et al.* (2015) Genetic characterization of a novel picornavirus detected in Miniopterus schreibersii bats. *Journal of General Virology*, **96**, 815–821.

Kim HK, Yoon SW, Kim DJ *et al.* (2016) Detection of Severe Acute Respiratory Syndrome-Like, Middle East Respiratory Syndrome-Like Bat Coronaviruses and Group H Rotavirus in Faeces of Korean Bats. *Transboundary and Emerging Diseases*, **63**, 365–372.

Kleiber C, Zeileis A (2008). *Applied Econometrics with R*. New York: Springer-Verlag. https://CRAN.R-project.org/package=AER

Kleiner M, Hooper LV, Duerkop BA (2015) Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genomics*, **16**, 7.

Kohl C, Brinkmann A, Dabrowski PW *et al.* (2015) Protocol for Metagenomic Virus Detection in Clinical Specimens. *Emerging Infectious Diseases*, **21**, 48–57.

Kolde R (2015) pheatmap: Pretty Heatmaps. R package version 1.0.8.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, **15**, 1179–1191.

Koskella B, Hall LJ, Metcalf CJE (2017) The microbiome beyond the horizon of ecological and evolutionary theory. *Nature Ecology & Evolution*, **1**, 1–10.

Krasnov BR, Mouillot D, Shenbrot GI *et al.* (2010) Similarity in ectoparasite faunas of Palaearctic rodents as a function of host phylogenetic, geographic or environmental distances: Which matters the most? *International Journal for Parasitology*, **40**, 807–817.

Kueneman JG, Parfrey LW, Woodhams DC *et al.* (2013) The amphibian skin-associated microbiome across species, space and life history stages. *Molecular Ecology*, **23**, 1238–1250.

Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P (2008) A Bioinformatician's Guide to Metagenomics. *Microbiology and Molecular Biology Reviews*, **72**, 557–578.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, **10**, R25.

Lau SKP, Woo PCY, Lai KKY *et al.* (2011) Complete Genome Analysis of Three Novel Picornaviruses from Diverse Bat Species. *Journal of Virology*, **85**, 8819–8828.

Lau S, Woo P, Li K *et al.* (2005) Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 14040–14045.

Laurence M, Hatzis C, Brash DE (2014) Common Contaminants in Next-Generation Sequencing That Hinder Discovery of Low-Abundance Microbes. *PLoS ONE*, **9**, e97876–8.

Le Gal F, Gault E, Ripault M-P *et al.* (2006) Eighth major clade for hepatitis delta virus. *Emerging Infectious Diseases*, **12**, 1447–1450.

Lecuit M, Eloit M (2013) The human virome: new tools and concepts. *Trends in*

Microbiology, **21**, 510–515.

Lee DN, Papeş M, Van Den Bussche RA (2012a) Present and Potential Future Distribution of Common Vampire Bats in the Americas and the Associated Risk to Cattle. *PLoS ONE*, **7**, e42466.

Lee JS, Ruell EW, Boydston EE *et al.* (2012b) Gene flow and pathogen transmission among bobcats (Lynx rufus) in a fragmented urban landscape. *Molecular Ecology*, **21**, 1617–1631.

Lempp F, Urban S (2017) Hepatitis Delta Virus: Replication Strategy and Upcoming Therapeutic Options for a Neglected Human Pathogen. *Viruses*, **9**, 172–19.

Leroy EM, Kumulungui B, Pourrut X *et al.* (2005) Fruit bats as reservoirs of Ebola virus. *Nature*, **438**, 575–576.

Leu ST, Kappeler PM, Bull CM (2010) Refuge sharing network predicts ectoparasite load in a lizard. *Behavioral Ecology and Sociobiology*, **64**, 1495–1503.

Ley RE, Hamady M, Lozupone C *et al.* (2008) Evolution of mammals and their gut microbes. *Science*, **320**, 1647–1651.

Ley RE, Turnbaugh PJ, Klein S, Gordon JI (2006) Microbial ecology: human gut microbes associated with obesity. *Nature*, **444**, 1022–1023.

Li D, Li Z, Zhou Z *et al.* (2016) Direct next-generation sequencing of virus-human mixed samples without pretreatment is favorable to recover virus genome. *Biology Direct*, **11**, 1–10.

Li L, Deng X, Mee ET *et al.* (2015) Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. *Journal of Virological Methods*, **213**, 139–146.

Li L, Shan T, Wang C *et al.* (2011) The Fecal Viral Flora of California Sea Lions. *Journal of Virology*, **85**, 9909–9917.

Li L, Victoria JG, Wang C *et al.* (2010a) Bat Guano Virome: Predominance of Dietary Viruses from Insects and Plants plus Novel Mammalian Viruses. *Journal of Virology*, **84**, 6955–6965.

Li W, Shi Z, Yu M *et al.* (2005) Bats are natural reservoirs of SARS-like coronaviruses. *Science*, **310**, 676–679.

Li Y, Ge X, Zhang H *et al.* (2010b) Host Range, Prevalence, and Genetic Diversity of Adenoviruses in Bats. *Journal of Virology*, **84**, 3889–3897.

Lima FE de S, Cibulski SP, Elesbao F *et al.* (2013) First detection of adenovirus in the vampire bat (Desmodus rotundus) in Brazil. *Virus Genes*, **47**, 378–381.

Lindenfors P, Nunn CL, Jones KE *et al.* (2007) Parasite species richness in carnivores: effects of host body mass, latitude, geographical range and population density. *Global Ecology and Biogeography*, **16**, 496–509.

Lindström ES, Langenheder S (2011) Local and regional factors influencing bacterial community assembly. *Environmental Microbiology Reports*, **4**, 1–9.

Linnenbrink M, Wang J, Hardouin EA *et al.* (2013) The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Molecular Ecology*, **22**, 1904–1916.

Lloyd-Smith JO, Cross PC, Briggs CJ *et al.* (2005) Should we expect population thresholds for wildlife disease? *Trends in Ecology and Evolution*, **20**, 511–519.

Lloyd-Smith JO, George D, Pepin KM *et al.* (2009) Epidemic Dynamics at the Human-Animal Interface. *Science*, **326**, 1362–1367.

Lo C, Morand S, Galzin R (1998) Parasite diversity\ host age and size relationship in three coral-reef fishes from French Polynesia. *International Journal for Parasitology*, **28**, 1695–1708.

Lomolino MV (2001) Elevation gradients of species-density: historical and

prospective views. *Global Ecology and Biogeography*, **10**, 3–13.

Loudon AH, Woodhams DC, Parfrey LW *et al.* (2013) Microbial community dynamics and effect of environmental microbial reservoirs on red-backed salamanders (Plethodon cinereus). *The ISME Journal*, **8**, 830–840.

Lozupone CA, Hamady M, Kelley ST, Knight R (2007) Quantitative and Qualitative Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. *Applied and Environmental Microbiology*, **73**, 1576–1585.

Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R (2012) Diversity, stability and resilience of the human gut microbiota. *Nature*, **489**, 220–230.

López-Bueno A, Tamames J, Velázquez D *et al.* (2009) High diversity of the viral community from an Antarctic lake. *Science*, **326**, 858–861.

Luikart G, Ryman N, Tallmon DA, Schwartz MK, Allendorf FW (2010) Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics*, **11**, 355–373.

Luis AD, Hayman DTS, O'Shea TJ *et al.* (2013) A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20122753–20122753.

Luis AD, O'Shea TJ, Hayman DTS *et al.* (2015) Network analysis of host-virus communities in bats and rodents reveals determinants of cross-species transmission. *Ecology Letters*, **18**, 1153–62.

Lukashev AN, Corman VM, Schacht D *et al.* (2017) Close genetic relatedness of picornaviruses from European and Asian bats. *Journal of General Virology*, **98**, 955–961.

Luque JL, Poulin R (2008) Linking ecology with parasite diversity in Neotropical fishes. *Journal of Fish Biology*, **72**, 189–204.

Ly M, Abeles SR, Boehm TK *et al.* (2014) Altered Oral Viral Ecology in Association with Periodontal Disease. *mBio*, **5**, e01133–14–e01133–14.

Maeda K, Hondo E, Terakawa J *et al.* (2008) Isolation of novel adenovirus from fruit bat (Pteropus dasymallus yayeyamae). *Emerging Infectious Diseases*, **14**, 347–349.

Maganga GD, Bourgarel M, Vallo P *et al.* (2014) Bat Distribution Size or Shape as Determinant of Viral Richness in African Bats. *PLoS ONE*, **9**, e100172.

Mandl JN, Ahmed R, Barreiro LB *et al.* (2015) Reservoir Host Immune Responses to Emerging Zoonotic Viruses. *Cell*, **160**, 20–35.

Manrique P, Bolduc B, Walk ST *et al.* (2016) Healthy human gut phageome. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 10400–10405.

Martin M (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal,* **17**, 10-12.

Martins FM, Ditchfield AD, Meyer D, Morgante JS (2007) Mitochondrial DNA phylogeography reveals marked population structure in the common vampire bat, Desmodus rotundus (Phyllostomidae). *Journal of Zoological Systematics and Evolutionary Research*, **45**, 372–378.

Martins FM, Templeton AR, Pavan AC, Kohlbach BC, Morgante JS (2009) Phylogeography of the common vampire bat (Desmodus rotundus): Marked population structure, Neotropical Pleistocene vicariance and incongruence between nuclear and mtDNA markers. *BMC Evolutionary Biology*, **9**, 294.

Martiny JBH, Bohannan BJM, Brown JH *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, **4**, 102–112.

Masembe C, Michuki G, Onyango M *et al.* (2012) Viral metagenomics demonstrates that domestic pigs are a potential reservoir for Ndumu virus. *Virology Journal*, **9**, 218.

Matranga CB, Andersen KG, Winnicki S *et al.* (2014) Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. *Genome Biology*, **15**, 519.

Matthijnssens J, Otto PH, Ciarlet M *et al.* (2012) VP6-sequence-based cutoff values as a criterion for rotavirus species demarcation. *Archives of Virology*, **157**, 1177–1182.

McArdle BH, Anderson MJ (2001) Fitting Multivariate Models to Community Data: A Comment On Distance-Based Redundancy Analysis. *Ecology*, **82**, 290–297.

McCallum H, Barlow N, Hone J (2001) How should pathogen transmission be modelled? *Trends in Ecology and Evolution*, **16**, 295–300.

McCord AI, Chapman CA, Weny G *et al.* (2013) Fecal microbiomes of non-human primates in Western Uganda reveal species-specific communities largely resistant to habitat perturbation. *American Journal of Primatology*, **76**, 347–354.

McDonald SM, Nelson MI, Turner PE, Patton JT (2016) Reassortment in segmented RNA viruses: mechanisms and outcomes. *Nature Reviews Microbiology*, **14**, 448–460.

McKenzie VJ (2007) Human land use and patterns of parasitism in tropical amphibian hosts. *Biological Conservation*, **137**, 102–116.

Meirmans PG, Hedrick PW (2010) Assessing population structure: FST and related measures. *Molecular Ecology Resources*, **11**, 5–18.

Memish ZA, Mishra N, Olival KJ *et al.* (2013) Middle East Respiratory Syndrome Coronavirus in Bats, Saudi Arabia. *Emerging Infectious Diseases*, **19**, 1–5.

Menke S, Meier M, Mfune JKE *et al.* (2017) Effects of host traits and land-use changes on the gut microbiota of the Namibian black-backed jackal (Canis mesomelas). *FEMS Microbiology Ecology*, **93**, 10.

Menke S, Wasimuddin, Meier M *et al.* (2014) Oligotyping reveals differences between gut microbiomes of free-ranging sympatric Namibian carnivores (Acinonyx jubatus, Canis mesomelas) on a bacterial species-like level. *Frontiers in Microbiology*, **5**, 526.

Mihaljevic JR (2012) Linking metacommunity theory and symbiont evolutionary ecology. *Trends in Ecology and Evolution*, **27**, 323–329.

Minot S, Sinha R, Chen J *et al.* (2011) The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Research*, **21**, 1616–1625.

Mokili JL, Rohwer F, Dutilh BE (2012) Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology*, **2**, 63–77.

Moreno JA, Baer GM (1980) Experimental rabies in the vampire bat. *American Journal of Tropical Medicine and Hygiene*, **29**, 254–259.

Morgan JL, Darling AE, Eisen JA (2010) Metagenomic Sequencing of an In Vitro-Simulated Microbial Community. *PLoS ONE*, **5**.

Morse SS (1993) Examining the origins of emerging viruses. In: *Emerging Viruses* (ed Morse SS), pp. 10–28. Oxford University Press, New York.

Mouinga-Ondeme A, Caron M, Nkoghe D *et al.* (2011) Cross-Species Transmission of Simian Foamy Virus to Humans in Rural Gabon, Central Africa. *Journal of Virology*, **86**, 1255–1260.

Mourier T, Mollerup S, Vinner L *et al.* (2015) Characterizing novel endogenous retroviruses from genetic variation inferred from short sequence reads. *Scientific Reports*, **5**, 1–11.

Muegge BD, Kuczynski J, Knights D *et al.* (2011) Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans. *Science*, **332**, 970–974.

Mueller S, Saunier K, Hanisch C *et al.* (2006) Differences in Fecal Microbiota in Different European Study Populations in Relation to Age, Gender, and Country: a Cross-Sectional Study. *Applied and Environmental Microbiology*, **72**, 1027–1033.

Mukherjee S, Huntemann M, Ivanova N, Kyrpides NC, Pati A (2015) Large-scale contamination of microbial isolate genomes by Illumina PhiX control. *Standards in Genomic Sciences*, **10**, 18.

Muletz Wolz CR, Yarwood SA, Campbell Grant EH, Fleischer RC, Lips KR (2017) Effects of host species and environment on the skin microbiome of Plethodontid salamanders. *Journal of Animal Ecology*, **87**, 341–353.

Munshi-South J, Wilkinson GS (2010) Bats and birds: Exceptional longevity despite high metabolic rates. *Ageing Research Reviews*, **9**, 12–19.

Murray KA, Daszak P (2013) Human ecology in pathogenic landscapes: two hypotheses on how land use change drives viral emergence. *Current Opinion in Virology*, **3**, 79–83.

Murtagh F, Legendre P (2014) Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification*, **31**, 274–295.

Nagashima S, Kobayashi N, Ishino M *et al.* (2008) Whole genomic characterization of a human rotavirus strain B219 belonging to a novel group of the genus rotavirus. *Journal of Medical Virology*, **80**, 2023–2033.

Nakamura S, Yang C-S, Sakon N *et al.* (2009) Direct Metagenomic Detection of Viral Pathogens in Nasal and Fecal Specimens Using an Unbiased High-Throughput Sequencing Approach. *PLoS ONE*, **4**, e4219.

Negro SS, Caudron AK, Dubois M, Delahaut P, Gemmell NJ (2010) Correlation between Male Social Status, Testosterone Levels, and Parasitism in a Dimorphic Polygynous Mammal. *PLoS ONE*, **5**, e12507.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321–3323.

Nekola JC, White PS (1999) The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, **26**, 867–878.

Norman JM, Handley SA, Baldridge MT *et al.* (2015) Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell*, **160**, 447–460.

Nunn CL, Altizer SM, Sechrest W, Cunningham AA (2005) Latitudinal gradients of parasite species richness in primates. *Diversity and Distributions*, **11**, 249–256.

Nunn CL, Altizer S, Jones KE, Sechrest W (2003) Comparative Tests of Parasite Species Richness in Primates. *The American Naturalist*, **162**, 597–614.

Nunn CL, Brezine C, Jolles AE, Ezenwa VO (2014) Interactions between Micro- and Macroparasites Predict Microparasite Species Richness across Primates. *The American Naturalist*, **183**, 494–505.

Nychka D, Furrer R, Paige J, Sain S (2015) fields: Tools for spatial data. R package version 9.0. http://doi.org/10.5065/D6W957CT

O'Shea TJ, Cryan PM, Cunningham AA *et al.* (2014) Bat Flight and Zoonotic Viruses. *Emerging Infectious Diseases*, **20**, 741–745.

Ogle DH (2017) FSA: Fisheries Stock Analysis. R package version 0.8.17.

Oksanen J, Blanchet FG, Friendly M *et al.* (2017) vegan: Community Ecology Package. R package version 2.4-4. https://CRAN.R-project.org/package=vegan

Olival KJ, Hosseini PR, Zambrana-Torrelio C *et al.* (2017) Host and viral traits predict zoonotic spillover from mammals. *Nature*, **546**, 646–650.

Ondov BD, Bergman NH, Phillippy AM (2011) Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, **12**, 385.

Osborne C, Cryan PM, O'Shea TJ *et al.* (2011) Alphacoronaviruses in New World Bats: Prevalence, Persistence, Phylogeny, and Potential for Interaction with Humans. *PLoS ONE*, **6**, e19156.

Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA *et al.* (2016) Uncovering Earth's virome. *Nature*, **536**, 425–430.

Palacios G, Druce J, Du L *et al.* (2008) A new arenavirus in a cluster of fatal transplant-associated diseases. *New England Journal of Medicine*, **358**, 991–998.

Palstra FP, Fraser DJ (2012) Effective/census population size ratio estimation: a compendium and appraisal. *Ecology and Evolution*, **2**, 2357–2365.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289–290.

Patterson BD, Dick CW, Dittmar K (2008) Parasitism by bat flies (Diptera: Streblidae) on neotropical bats: effects of host body size, distribution, and abundance. *Parasitology Research*, **103**, 1091–1100.

Pavio N, Meng X-J, Renou C (2010) Zoonotic hepatitis E: animal reservoirs and emerging risks. *Veterinary Research*, **41**, 46–20.

Pawan JL (1936) The Transmission of Paralytic Rabies in Trinidad by the Vampire Bat (Desmodus Rotundus Murinus Wagner, 1840). *Annals of Tropical Medicine & Parasitology*, **30**, 101–130.

Perez-Gracia M, Suay B, Luisa Mateos-Lindemann M (2014) Hepatitis E: An emerging disease. *Infection, Genetics and Evolution*, **22**, 40–59.

Phillips CD, Phelan G, Dowd SE *et al.* (2012) Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography. *Molecular Ecology*, **21**, 2617–2627.

Piaggio AJ, Johnston JJ, Perkins SL (2008) Development of polymorphic microsatellite loci for the common vampire bat, Desmodus rotundus (Chiroptera: Phylostomidae). *Molecular Ecology Resources*, **8**, 440–442.

Pinto-Santini DM, Stenbak CR, Linial ML (2017) Foamy virus zoonotic infections. *Retrovirology*, **14**, 1–14.

Plowright RK, Eby P, Hudson PJ *et al.* (2014) Ecological dynamics of emerging bat virus spillover. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20142124–20142124.

Poirotte C, Basset D, Willaume E *et al.* (2015) Environmental and individual determinants of parasite richness across seasons in a free-ranging population of Mandrills (Mandrillus sphinx). *American Journal of Physical Anthropology*, **159**, 442–456.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–846.

Ponzetto A, Cote PJ, Popper H *et al.* (1984) Transmission of the hepatitis B virus-associated delta agent to the eastern woodchuck. *Proceedings of the National Academy of Sciences of the United States of America*, **81**, 2208–2212.

Poon LLM, Chu DKW, Chan KH *et al.* (2005) Identification of a Novel Coronavirus in Bats. *Journal of Virology*, **79**, 2001–2009.

Poulin R (1996a) Sexual inequalities in helminth infections: a cost of being a male? *American Naturalist*, 287–295.

Poulin R (1996b) Helminth growth in vertebrate hosts: does host sex matter?

International Journal for Parasitology, **26**, 1311–1315.

Poulin R (2014) Parasite biodiversity revisited: frontiers and constraints. *International Journal for Parasitology*, **44**, 581–589.

Poulin R, Rohde K (1997) Comparing the richness of metazoan ectoparasite communities of marine fishes: controlling for host phylogeny. *Oecologia*, **110**, 278–283.

Prest TL, Kimball AK, Kueneman JG, McKenzie VJ (2018) Host-associated bacterial community succession during amphibian development. *Molecular Ecology*, **105**, 11512–15.

Pride DT, Salzman J, Haynes M *et al.* (2011) Evidence of a robust resident bacteriophage population revealed through analysis ofthe human salivary virome. *The ISME Journal*, **6**, 915–926.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Prosser JI, Bohannan BJM, Curtis TP *et al.* (2007) The role of ecological theory in microbial ecology. *Nature Reviews Microbiology*, **5**, 384–392.

Puechmaille SJ (2016) The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, **16**, 608–627.

Pulliam JRC (2008) Viral Host Jumps: Moving toward a Predictive Framework. *EcoHealth*, **5**, 80–91.

Pulliam JRC, Dushoff J (2009) Ability to Replicate in the Cytoplasm Predicts Zoonotic Transmission of Livestock Viruses. *The Journal of Infectious Diseases*, **199**, 565–568.

Pulliam JRC, Epstein JH, Dushoff J *et al.* (2011) Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis. *Journal of The Royal Society Interface*, **9**, 89–101.

Quan P-L, Firth C, Conte JM *et al.* (2013) Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 8194–8199.

Quintana H, Pacheco V (2007) Identificación y distribución de los murciélagos vampiros del Perú. *Revista Peruana de Medicina Experimental y Salud Publica*, **24**, 81–88.

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rampelli S, Soverini M, Turroni S *et al.* (2016) ViromeScan: a new tool for metagenomic viral community profiling. *BMC Genomics*, **17**, 165.

Rampelli S, Turroni S, Schnorr SL *et al.* (2017) Characterization of the human DNA gut virome across populations with different subsistence strategies and geographical origin. *Environmental Microbiology*, **19**, 4728–4735.

Randhawa HS, Poulin R (2010) Determinants of tapeworm species richness in elasmobranch fishes: untangling environmental and phylogenetic influences. *Ecography*, **33**, 866–877.

Rasche A, Souza, Drexler JF (2016) Bat hepadnaviruses and the origins of primate hepatitis B viruses. *Current Opinion in Virology*, **16**, 86–94.

Reckardt K, Kerth G (2007) Roost selection and roost switching of female Bechstein's bats (Myotis bechsteinii) as a strategy of parasite avoidance. *Oecologia*, **154**, 581–588.

Reimchen TE, Nosil P (2001) Ecological causes of sex-biased parasitism in threespine

stickleback. *Biological Journal of the Linnean Society*, **73**, 51–63.

Reyes A, Haynes M, Hanson N *et al.* (2010) Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature*, **466**, 334–338.

Reyes A, Semenkovich NP, Whiteson K, Rohwer F, Gordon JI (2012) Going viral: next-generation sequencing applied to phage populations in the human gut. *Nature Reviews Microbiology*, **10**, 607–617.

Rice P, Longden I, Bleasby A (2000) EMBOSS: The European molecular biology open software suite. *Trends in Genetics*, **16**, 276–277.

Rifkin JL, Nunn CL, Garamszegi LZ (2012) Do Animals Living in Larger Groups Experience Greater Parasitism? A Meta-Analysis. *The American Naturalist*, **180**, 70–82.

Robardet E, Borel C, Moinet M *et al.* (2017) Longitudinal survey of two serotine bat (Eptesicus serotinus) maternity colonies exposed to EBLV-1 (European Bat Lyssavirus type 1): Assessment of survival and serological status variations using capture-recapture models. *PLoS Neglected Tropical Diseases*, **11**, e0006048–20.

Robles-Sikisaka R, Ly M, Boehm T *et al.* (2013) Association between living environment and human oral viral ecology. *The ISME Journal*, **7**, 1710–1724.

Rodriguez-Brito B, Li L, Wegley L *et al.* (2010) Viral and microbial community dynamics in four aquatic environments. *The ISME Journal*, **4**, 739–751.

Rohwer F, Edwards R (2002) The Phage Proteomic Tree: a Genome-Based Taxonomy for Phage. *Journal of Bacteriology*, **184**, 4529–4535.

Romero-Nava C, León-Paniagua L, Ortega J (2014) Microsatellites loci reveal heterozygosis and population structure in vampire bats (Desmodus rotundus) (Chiroptera: Phyllostomidae) of Mexico. *Revista de biología tropical*, **62**, 659–669.

Rosenberg NA (2003) distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Roux S, Enault F, Ravet V *et al.* (2016a) Analysis of metagenomic data reveals common features of halophilic viral communities across continents. *Environmental Microbiology*, **18**, 889–903.

Roux S, Enault F, Robin A *et al.* (2012) Assessing the Diversity and Specificity of Two Freshwater Viral Communities through Metagenomics. *PLoS ONE*, **7**, e33641.

Roux S, Faubladier M, Mahul A *et al.* (2011) Metavir: a web server dedicated to virome analysis. *Bioinformatics*, **27**, 3074–3075.

Roux S, Solonenko NE, Dang VT *et al.* (2016b) Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ*, **4**, e2777.

Rupprecht CE, Turmelle A, Kuzmin IV (2011) A perspective on lyssavirus emergence and perpetuation. *Current Opinion in Virology*, **1**, 662–670.

Salmier A, Tirera S, de Thoisy B *et al.* (2017) Virome analysis of two sympatric bat species (Desmodus rotundus and Molossus molossus) in French Guiana. *PLoS ONE*, **12**, e0186943–25.

Salter SJ, Cox MJ, Turek EM *et al.* (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, **12**, 118.

Sasaki M, Orba Y, Ueno K *et al.* (2015) Metagenomic analysis of the shrew enteric virome reveals novel viruses related to human stool-associated viruses. *Journal of General Virology*, **96**, 440–452.

Sasaki M, Setiyono A, Handharyani E *et al.* (2014) Isolation and Characterization of a Novel Alphaherpesvirus in Fruit Bats. *Journal of Virology*, **88**, 9819–9829.

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic

datasets. *Bioinformatics*, **27**, 863–864.

Schmieder R, Lim YW, Edwards R (2012) Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics*, **28**, 433–435.

Schneeberger PHH, Becker SL, Pothier JF *et al.* (2016) Metagenomic diagnostics for the simultaneous detection of multiple pathogens in human stool specimens from Côte d'Ivoire: a proof-of-concept study. *Infection, Genetics and Evolution*, **40**, 389–397.

Schneider MC, Aron J, Santos-Burgoa C, Uieda W, Ruiz-Velazco S (2001) Common vampire bat attacks on humans in a village of the Amazon region of Brazil. *Cadernos de saúde pública*, **17**, 1531–1536.

Schneider MC, Romijn PC, Uieda W *et al.* (2009) Rabies transmitted by vampire bats to humans: an emerging zoonotic disease in Latin America? *Revista panamericana de salud pública*, **25**, 260–269.

Schotthoefer AM, Rohr JR, Cole RA *et al.* (2011) Effects of wetland vs. landscape variables on parasite communities of Rana pipiens: links to anthropogenic factors. *Ecological applications*, **21**, 1257–1271.

Schweighardt AJ, Tate CM, Scott KA, Harper KA, Robertson JM (2014) Evaluation of Commercial Kits for Dual Extraction of DNA and RNA from Human Body Fluids. *Journal of Forensic Sciences*, **60**, 157–165.

Segata N, Waldron L, Ballarini A *et al.* (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, **9**, 811–814.

Sender R, Fuchs S, Milo R (2016) Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology*, **14**, e1002533–14.

Shi M, Lin X-D, Chen X *et al.* (2018) The evolutionary history of vertebrate RNA viruses. *Nature*, **556**, 1–16.

Shi M, Lin X-D, Tian J-H *et al.* (2016) Redefining the invertebrate RNA virosphere. *Nature*, **540**, 539–543.

Simmonds P, Adams MJ, Benkő M *et al.* (2017) Consensus statement: Virus taxonomy in the age of metagenomics. *Nature*, **15**, 161–168.

Smith DB, Purdy MA, Simmonds P (2013) Genetic Variability and the Classification of Hepatitis E Virus. *Journal of Virology*, **87**, 4161–4169.

Smith DB, Simmonds P, members of the International Committee on the Taxonomy of Viruses Hepeviridae Study Group *et al.* (2014) Consensus proposals for classification of the family Hepeviridae. *Journal of General Virology*, **95**, 2223–2232.

Sonntag M, Mühldorfer K, Speck S, Wibbelt G, Kurth A (2009) New Adenovirus in Bats, Germany. *Emerging Infectious Diseases*, **15**, 2052–2055.

Soueidan H, Schmitt L-A, Candresse T, Nikolski M (2015) Finding and identifying the viral needle in the metagenomic haystack: trends and challenges. *Frontiers in Microbiology*, **5**, 739.

Srivathsan A, Ang A, Vogler AP, Meier R (2016) Fecal metagenomics for the simultaneous assessment of diet, parasites, and population genetics of an understudied primate. *Frontiers in Zoology*, **13**, 17.

Stamatakis A, Hoover P, Rougemont J (2008) A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology*, **57**, 758–771.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Stoner-Duncan B, Streicker DG, Tedeschi CM (2014) Vampire Bats and Rabies: Toward an Ecological Solution to a Public Health Problem. *PLoS Neglected Tropical Diseases*, **8**, e2867.

Streicker DG, Allgeier JE (2016) Foraging choices of vampire bats in diverse landscapes: potential implications for land-use change and disease transmission. *Journal of Applied Ecology*, **53**, 1280–1288.

Streicker DG, Lemey P, Velasco-Villa A, Rupprecht CE (2012a) Rates of Viral Evolution Are Linked to Host Geography in Bat Rabies. *PLoS Pathogens*, **8**, e1002720.

Streicker DG, Recuenco S, Valderrama W *et al.* (2012b) Ecological and anthropogenic drivers of rabies exposure in vampire bats: implications for transmission and control. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 3384–3392.

Streicker DG, Turmelle AS, Vonhof MJ *et al.* (2010) Host Phylogeny Constrains Cross-Species Emergence and Establishment of Rabies Virus in Bats. *Science*, **329**, 676–679.

Streicker DG, Winternitz JC, Satterfield DA *et al.* (2016) Host-pathogen evolutionary signatures reveal dynamics and future invasions of vampire bat rabies. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 10926–10931.

Strickland MS, Lauber C, Fierer N, Bradford MA (2009) Testing the functional significance of microbial community composition. *Ecology*, **90**, 441–451.

Strom SL (2008) Microbial ecology of ocean biogeochemistry: a community perspective. *Science*, **320**, 1043–1045.

Stulberg E (2016) An assessment of US microbiome research. *Nature Microbiology*, **1**, 15015.

Stumpf RM, Gomez A, Amato KR *et al.* (2016) Microbiomes, metagenomics, and primate conservation: New strategies, tools, and applications. *Biological Conservation*, **199**, 56–66.

Sullam KE, Essinger SD, Lozupone CA *et al.* (2012) Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Molecular Ecology*, **21**, 3363–3378.

Sunagawa S, Coelho LP, Chaffron S *et al.* (2015) Ocean plankton. Structure and function of the global ocean microbiome. *Science*, **348**, 1261359–1261359.

Suttle CA (2005) Viruses in the sea. *Nature*, **437**, 356–361.

Suttle CA (2007) Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology*, **5**, 801–812.

Suzán G, García-Peña GE, Castro-Arellano I *et al.* (2015) Metacommunity and phylogenetic structure determine wildlife and zoonotic infectious disease patterns in time and space. *Ecology and Evolution*, **5**, 865–873.

Swinton J, Harwood J, Grenfell BT, Gilligan CA (1998) Persistence thresholds for phocine distemper virus infection in harbour seal Phoca vitulina metapopulations. *Journal of Animal Ecology*, **67**, 54–68.

Switzer WM, Salemi M, Shanmugam V *et al.* (2005) Ancient co-speciation of simian foamy viruses and primates. *Nature*, **434**, 376–380.

Taberlet P, Waits L, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution*, **14**, 323–327.

Tan B, Yang X-L, Ge X-Y *et al.* (2017) Novel bat adenoviruses with low G+C content shed new light on the evolution of adenoviruses. *Journal of General Virology*, **98**, 739–748.

Tang XC, Zhang JX, Zhang SY *et al.* (2006) Prevalence and Genetic Diversity of Coronaviruses in Bats from China. *Journal of Virology*, **80**, 7481–7490.

Tao Y, Shi M, Chommanard C *et al.* (2017) Surveillance of Bat Coronaviruses in

Kenya Identifies Relatives of Human Coronaviruses NL63 and 229E and Their Recombination History. *Journal of Virology*, **91**, e01953–16.

Taylor DJ, Leach RW, Bruenn J (2010) Filoviruses are ancient and integrated into mammalian genomes. *BMC Evolutionary Biology*, **10**, 193.

Taylor JM (2006) Hepatitis delta virus. *Virology*, **344**, 71–76.

Taylor JM (2014) Host RNA circles and the origin of hepatitis delta virus. *World Journal of Gastroenterology*, **20**, 2971–9.

Taylor LH, Latham SM, Woolhouse MEJ (2001) Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **356**, 983–989.

Temmam S, Davoust B, Berenger J-M, Raoult D, Desnues C (2014) Viral Metagenomics on Animals as a Tool for the Detection of Zoonoses Prior to Human Infection? *International Journal of Molecular Sciences*, **15**, 10377–10397.

Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F (2009) Laboratory procedures to generate viral metagenomes. *Nature Protocols*, **4**, 470–483.

Tong S, Li Y, Rivailler P *et al.* (2012) A distinct lineage of influenza A virus from bats. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 4269–4274.

Tong S, Zhu X, Li Y *et al.* (2013) New world bats harbor diverse influenza A viruses. *PLoS Pathogens*, **9**, e1003657.

Torres J, Miquel J, Casanova J-C *et al.* (2006) Endoparasite Species Richness of Iberian Carnivores: Influences of Host Density and Range Distribution. *Biodiversity and Conservation*, **15**, 4619–4632.

Trajano E (1996) Movements of cave bats in southeastern Brazil, with emphasis on the population ecology of the common vampire bat, Desmodus rotundus (Chiroptera). *Biotropica*, **28**, 121–129.

Troupin C, Dacheux L, Tanguy M *et al.* (2016) Large-Scale Phylogenomic Analysis Reveals the Complex Evolutionary History of Rabies Virus in Multiple Carnivore Hosts. *PLoS Pathogens*, **12**, e1006041–20.

Tse H, Tsang AKL, Tsoi H-W *et al.* (2012) Identification of a Novel Bat Papillomavirus by Metagenomics. *PLoS ONE*, **7**, e43986.

Tung J, Barreiro LB, Burns MB, Grenier JC, Lynch J (2015) Social networks predict gut microbiome composition in wild baboons. *eLife, 4, e05224*

Turmelle AS, Olival KJ (2009) Correlates of Viral Richness in Bats (Order Chiroptera). *EcoHealth*, **6**, 522–539.

Turmelle AS, Jackson FR, Green D, McCracken GF, Rupprecht CE (2010) Host immunity to repeated rabies virus infection in big brown bats. *Journal of General Virology*, **91**, 2360–2366.

van Boheemen S, de Graaf M, Lauber C *et al.* (2012) Genomic Characterization of a Newly Discovered Coronavirus Associated with Acute Respiratory Distress Syndrome in Humans. *mBio*, **3**, e00473–12.

van Cuyck H, Fan J, Robertson DL, Roques P (2005) Evidence of recombination between divergent hepatitis E viruses. *Journal of Virology*, **79**, 9306–9314.

van der Heijden MGA, Bardgett RD, van Straalen NM (2008) The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters*, **11**, 296–310.

van Dijk EL, Jaszczyszyn Y, Thermes C (2014) Library preparation methods for next-generation sequencing: Tone down the bias. *Experimental Cell Research*, **322**, 12–20.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Vayssier-Taussat M, Albina E, Citti C *et al.* (2014) Shifting the paradigm from pathogens to pathobiome: new concepts in the light of meta-omics. *Frontiers in Cellular and Infection Microbiology*, **4**, 29.

Veikkolainen V, Vesterinen EJ, Lilley TM, Pulliainen AT (2014) Bats as Reservoir Hosts of Human Bacterial Pathogen, Bartonella mayotimonensis. *Emerging Infectious Diseases*, **20**, 960–967.

Virgin HW (2014) The Virome in Mammalian Physiology and Disease. *Cell*, **157**, 142–150.

Virgin HW, Wherry EJ, Ahmed R (2009) Redefining Chronic Viral Infection. *Cell*, **138**, 30–50.

Vitone ND, Altizer S, Nunn CL (2004) Body size, diet and sociality influence the species richness of parasitic worms in anthropoid primates. *Evolutionary Ecology Research*, **6**, 183–199.

Vo ATE, Jedlicka JA (2014) Protocols for metagenomic DNA extraction and Illumina amplicon library preparation for faecal and swab samples. *Molecular Ecology Resources*, **14**, 1183–1197.

Voigt CC, Kelm DH (2006) Host Preference Of The Common Vampire Bat (Desmodus Rotundus; Chiroptera) Assessed By Stable Isotopes. *Journal of Mammalogy*, **87**, 1–6.

Voigt CC, Voigt-Heucke SL, Schneeberger K (2011) Isotopic Data Do Not Support Food Sharing Within Large Networks of Female Vampire Bats (Desmodus rotundus). *Ethology*, **118**, 260–268.

Volokhov DV, Becker DJ, Bergner LM *et al.* (2017) Novel hemotropic mycoplasmas are widespread and genetically diverse in vampire bats. *Epidemiology and Infection*, **145**, 3154–3167.

Wakuda M (2011) Porcine rotavirus closely related to novel group of human rotaviruses. *Emerging Infectious Diseases*, **17**, 1491–1493.

Wang B, Yang X-L, Li W *et al.* (2017) Detection and genome characterization of four novel bat hepadnaviruses and a hepevirus in China. *Virology Journal*, **14**, 1–10.

Wang H, Zhang W, Ni B, Shen H, Song Y, Wang X *et al.* (2010) Recombination analysis reveals a double recombination event in hepatitis E virus. *Virology Journal*, **7**, 129.

Wang J (2005) Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1395–1409.

Wang J, Soininen J, Zhang Y *et al.* (2011) Contrasting patterns in elevational diversity between microorganisms and macroorganisms. *Journal of Biogeography*, **38**, 595–603.

Wang Y, Naumann U, Wright ST, Warton DI (2012) mvabund- an Rpackage for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution*, **3**, 471–474.

Ward JH (1963) Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, **58**, 236.

Warton DI, Wright ST, Wang Y (2012) Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution*, **3**, 89–101.

Warton DI, Thibaut L, Wang YA (2017) The PIT-trap—A "model-free" bootstrap

procedure for inference about regression models with discrete, multivariate responses. *PLoS ONE,* **12**, e0181790.

Warwick RM, Clarke KR (1995) New "biodiversity" measures reveal a decrease in taxonomic distinctness with increasing stress. *Marine Ecology Progress Series*, **129**, 301–305.

Wasimuddin, Menke S, Melzheimer J *et al.* (2017) Gut microbiomes of free-ranging and captive Namibian cheetahs: Diversity, putative functions and occurrence of potential pathogens. *Molecular Ecology*, **26**, 5515–5527.

Webber QMR, Fletcher QE, Willis CKR (2017) Viral Richness is Positively Related to Group Size, but Not Mating System, in Bats. *EcoHealth*, **34**, 1–10.

Weedall GD, Irving H, Hughes MA, Wondji CS (2015) Molecular tools for studying the major malaria vector Anopheles funestus: improving the utility of the genome using a comparative poly(A) and Ribo-Zero RNAseq analysis. *BMC Genomics*, **16**, 1–15.

Wegner KM, Volkenborn N, Peter H, Eiler A (2013) Disturbance induced decoupling between host genetics and composition of the associated microbiome. *BMC Microbiology*, **13**, 252.

Wei T, Simko V R R package (2017) "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution,* **38**, 1358–1370.

Weller ML, Gardener MR, Bogus ZC *et al.* (2016) Hepatitis Delta Virus Detected in Salivary Glands of Sjögren"s Syndrome Patients and Recapitulates a Sjögren"s Syndrome-Like Phenotype in Vivo. *Pathogens & Immunity*, **1**, 12–40.

Whitlock MC (2011) G'ST and D do not replace FST. *Molecular Ecology*, **20**, 1083–1091.

WHO (2013) WHO Expert Consultation on Rabies. *World Health Organization Technical Report Series*, **982**.

Wilkinson GS (1984) Reciprocal food sharing in the vampire bat. *Nature*, **308**, 181–184.

Wilkinson GS (1986) Social Grooming in the Common Vampire Bat, Desmodus-Rotundus. *Animal Behaviour*, **34**, 1880–1889.

Williams SH, Che X, Garcia JA *et al.* (2018) Viral Diversity of House Mice in New York City. *mBio*, **9**, e01354–17–17.

Willoughby A, Phelps K, PREDICT Consortium, Olival K (2017) A Comparative Analysis of Viral Richness and Viral Sharing in Cave-Roosting Bats. *Diversity*, **9**, 35–16.

Winkler WG (1968) Airborne rabies virus isolation. *Journal of Wildlife Diseases*, **4**, 37–40.

Winter DJ (2012) mmod: an R library for the calculation of population differentiation statistics. *Molecular Ecology Resources*, **12**, 1158–1160.

Wirth T, Meyer A, Achtman M (2005) Deciphering host migrations and origins by means of their microbes. *Molecular Ecology*, **14**, 3289–3306.

Wohlgenant TJ (1994) Roost interactions between the common vampire bat (Desmodus rotundus) and two frugivorous bats (Phyllostomus discolor and Sturnira lilium) in Guanacaste, Costa Rica. *Biotropica*, 344–348.

Wolfe ND, Ngole EM, Gockowski J *et al.* (2000) Deforestation, hunting and the ecology of microbial emergence. *Global Change and Human Health*, **1**, 10–25.

Wommack KE, Bhavsar J, Polson SW *et al.* (2012) VIROME: a standard operating procedure for analysis of viral metagenome sequences. *Standards in Genomic*

Sciences, **6**, 427–439.

Woo PCY, Lau SKP, Li KSM *et al.* (2006) Molecular diversity of coronaviruses in bats. *Virology*, **351**, 180–187.

Wood JLN, Leach M, Waldman L *et al.* (2012) A framework for the study of zoonotic disease emergence and its drivers: spillover of bat pathogens as a case study. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 2881–2892.

Wood-Charlson EM, Weynberg KD, Suttle CA, Roux S, van Oppen MJH (2015) Metagenomic characterization of viral communities in corals: mining biological signal from methodological noise. *Environmental Microbiology*, **17**, 3440–3449.

Wray AK, Olival KJ, Moran D *et al.* (2016) Viral Diversity, Prey Preference, and Bartonella Prevalence in Desmodus rotundus in Guatemala. *EcoHealth*, 716–774.

Wu Z, Ren X, Yang L *et al.* (2012) Virome analysis for identification of novel mammalian viruses in bat species from Chinese provinces. *Journal of Virology*, **86**, 10999–11012.

Wu Z, Yang L, Ren X *et al.* (2016) Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *The ISME Journal*, **10**, 609–620.

Wylie KM, Mihindukulasuriya KA, Zhou Y *et al.* (2014) Metagenomic analysis of double-stranded DNA viruses in healthy adults. *BMC Biology*, **12**, 2317–10.

Wylie TN, Wylie KM, Herter BN, Storch GA (2015) Enhanced virome sequencing using targeted sequence capture. *Genome Research*, **25**, 1910–1920.

Xia L, Fan Q, He B *et al.* (2014) The complete genome sequence of a G3P[10] Chinese bat rotavirus suggests multiple bat rotavirus inter-host species transmission events. *Infection, Genetics and Evolution*, **28**, 1–4.

Yang J, Yang F, Ren L *et al.* (2011) Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. *Journal of Clinical Microbiology*, **49**, 3463–3469.

Yinda CK, Conceição-Neto N, Zeller M *et al.* (2017) Novel highly divergent sapoviruses detected by metagenomics analysis in straw-colored fruit bats in Cameroon. *Emerging Microbes & Infections*, **6**, e38.

Yinda CK, Ghogomu SM, Conceição-Neto N *et al.* (2018) Cameroonian fruit bats harbor divergent viruses, including rotavirus H, bastroviruses, and picobirnaviruses using an alternative genetic code. *Virus Evolution*, **4**, 1411–15.

Yinda CK, Zeller M, Conceição-Neto N *et al.* (2016) Novel highly divergent reassortant bat rotaviruses in Cameroon, without evidence of zoonosis. *Scientific Reports*, **6**, 1–10.

Young CCW, Olival KJ (2016) Optimizing Viral Discovery in Bats. *PLoS ONE*, **11**, e0149237.

Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM (2012) Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *New England Journal of Medicine*, **367**, 1814–1820.

Zepeda-Mendoza ML, Xiong Z, Escalera-Zamudio M *et al.* (2018) Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. *Nature Ecology & Evolution*, **63**, 1–12.

Zhang J-L, Swenson NG, Chen S-B *et al.* (2012) Phylogenetic beta diversity in tropical forests: Implications for the roles of geographical and environmental distance. *Journal of Systematics and Evolution*, **51**, 71–85.

Zhang W, Yang S, Shan T *et al.* (2017) Virome comparisons in wild-diseased and healthy captive giant pandas. *Microbiome*, **5**, 1–19.

Zheng X-Y, Qiu M, Guan W-J *et al.* (2017) Viral metagenomics of six bat species in close contact with humans in southern China. *Archives of Virology*, **85**, 1–16.

Zhong NS, Zheng BJ, Li YM *et al.* (2003) Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *The Lancet*, **362**, 1353–1358.

Zhou Y, Gao H, Mihindukulasuriya KA *et al.* (2013) Biogeography of the ecosystems of the healthy human body. *Genome Biology*, **14**, R1.

Zuk M, McKean KA (1996) Sex differences in parasite infections: patterns and processes. *International Journal for Parasitology*, **26**, 1009–1023.