Metagenomic Analysis Reveals Previously Undescribed Bat Coronavirus Strains in Eswatini

Julie Teresa Shapiro, Sarah Mollerup, Randi Holm Jensen, Jill Katharina Olofsson, Nam-phuong D. Nguyen, Thomas Arn Hansen, Lasse Vinner, Ara Monadjem, Robert A. Mc Cleery, and Anders J. Hansen

1School of Natural Resources and Environment, University of Florida, Gainesville, FL
2Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL
3University of Lyon, CIRI INSERM U1111 - CNRS UMR5308 - ENS Lyon, 46 Allée d’Italie, 69364 Lyon, France
4Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark
5Centre for GeoGenetics, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
6Computer Science and Engineering, University of California, San Diego, La Jolla, CA
7Department of Zoology and Entomology, Mammal Research Institute, University of Pretoria, Pretoria, South Africa
8Department of Life Sciences, Ben-Gurion University of the Negev, Be’er Sheva, Israel

Abstract: We investigated the prevalence of coronaviruses in 44 bats from four families in northeastern Eswatini using high-throughput sequencing of fecal samples. We found evidence of coronaviruses in 18% of the bats. We recovered full or near-full-length genomes from two bat species: Chaerephon pumilus and Afronycteris nana, as well as additional coronavirus genome fragments from C. pumilus, Epomophorus wahlbergi, Mops condylurus, and Scotophilus dinganii. All bats from which we detected coronaviruses were captured leaving buildings or near human settlements, demonstrating the importance of continued surveillance of coronaviruses in bats to better understand the prevalence, diversity, and potential risks for spillover.

Keywords: Chiroptera, alphacoronavirus, betacoronavirus, emerging infectious diseases, zoonotic disease, human–wildlife interface

INTRODUCTION

Coronaviruses are a family of zoonotic viruses comprised of four genera, two of which, alpha- and betacoronaviruses, have an evolutionary origin in bats, while gamma- and deltacoronaviruses, originate in birds (Graham et al. 2013). Coronaviruses have since radiated to a variety of hosts (Drexler et al. 2014). Notably, in humans, coronaviruses have caused COVID-19 (Zhou et al. 2020; Gorbalenya et al. 2020), Severe Acute Respiratory Syndrome (SARS) (Marra et al. 2003; Li et al. 2005), and Middle East Respiratory Syndrome (MERS) (Memish et al. 2013). While recent studies have increased our knowledge of coronavirus diversity and ecology, large gaps in sampling mean there are probably still many undiscovered species and strains in bats (Anthony et al. 2013, 2017).
Southern Africa has a diverse bat community (Monadjem et al. 2020b) that appears to host many coronaviruses, including strains phylogenetically close to MERS-CoV (Geldenhuys et al. 2013, 2018; Ithete et al. 2013), although studies are still limited (Markotter et al. 2020). Globally, the diversity and distribution of coronaviruses in bats makes it likely that future transmission of these pathogens to humans or other animal species will occur (Woo et al. 2009; Anthony et al. 2017). Although there are no known cases of coronavirus spillover in Africa thus far (Markotter et al. 2020), this could occur where bat species come into frequent, close contact with humans or domestic animals (Monadjem 1998; Fenton et al. 2004; Jacobs and Barclay 2009; Noer et al. 2012; Monadjem et al. 2020b).

Therefore, we investigated the prevalence of coronaviruses in bats belonging to eight species from four families (Pteropodidae: Epomophorus wahlbergi; Emballonuridae: Taphozous mauritianus; Molossidae: Chaerephon pumilus, Mops condylurus, and Mops midas; and Vespertilionidae: Afronycteris nana, Scotophilus danganii, and Scotophilus viridis). These species are all widely distributed and abundant across southeastern Africa and are commonly found in or near human settlements in northeast Eswatini (Monadjem et al. 2020b, 2021; Shapiro et al. 2020). We subjected fecal samples to virion enrichment followed by RNA sequencing to noninvasively investigate the prevalence and types of coronavirus in the bats of this region. We used this approach to recover whole coronavirus genomes and thus more reliably characterize them (Drexler et al. 2014; De Sabato et al. 2019). This method also allowed us to detect both known and unknown coronaviruses regardless of the specific sequences or genomic region present in samples.

We captured bats at eight sites in northeast Eswatini (Fig. 1) from December 2013–May 2014 using mist-nets and/or a harp trap. Taxonomy follows Monadjem et al. (2010, 2020b, 2020a). To aid in the identification of species, we measured forearm length of each captured bat with calipers to the nearest 0.1 mm and mass to the nearest 0.5 g with a spring balance. Captured bats were placed individually in cloth holding bags for the deposition of feces. We trapped, handled, and released bats in accordance with a permit from the Eswatini National Trust Commission and University of Florida Institutional Animal Care and Use Committee approval (Protocol #201,508,751).

Fecal samples from insectivorous species were desiccated and preserved with silica gel (Sigma-Aldrich), while samples from the frugivorous Epomophorus wahlbergi were placed in RNALater (Thermo Fisher Scientific) because due to their wet condition they could not be properly dried. Samples were stored at — 10 °C until the end of the field season (May 2014), then transferred to — 80 °C.

Frozen fecal samples were vortexed vigorously in 600 µl of PBS with beads from the PowerFecal kit (MoBio) for 1 min and incubated at room temperature for 10 min. Following incubation, samples were vortexed for 1 min, then centrifuged at 2500 × g for 3 min. The supernatant was then filtered and the flow-through nuclease-treated following Jensen et al. (2015). Viral nucleic acids were subsequently extracted using Roche High Pure Viral RNA kit (Roche) according to the manufacturer’s guidelines after which 1 µl RNase Out (Invitrogen) was added to the final RNA extract (Jensen et al. 2015; Hansen et al. 2015).

Forty-four RNA libraries were produced, each one from an individual fecal sample, using ScriptSeq v2 RNA-seq library preparation kit (Epicentre, Illumina), according to the manufacturer’s guidelines. Samples were DNase-treated with Promega DNase for 30 min at 37 °C and purified on RNaseA columns (Qiagen). Seven or eight individually and uniquely single-indexed sequencing libraries were pooled together in equimolar ratios for sequencing with paired-end reads of 100 bp (PE100) on an Illumina HiSeq 2000 platform. The library from one sample

Figure 1. Map of study region. Site numbers indicate where bats were captured and are referenced in Table 2. Sites from which coronaviruses were detected in bats are marked in red, while coronaviruses were not detected in bats captured from sites marked in black. The area shaded in gray is Hlane National Park. Solid lines indicate national borders and dotted lines indicate roads.
(Bat50) was resequenced individually on one lane of PE100 on an Illumina Hiseq 2000 platform.

Reads with overlapping sections of sequences were assembled into longer contiguous sequences (contigs) using Ray Meta v2.2.0 with default settings (Boisvert et al. 2012). The contigs were searched for coronaviruses using megablast and BLASTn on the NCBI Nucleotide collection (nt) database (Altschul et al. 1990, 1997) and by mapping against NCBI’s nr database using DIAMOND (Buchfink et al. 2014).

De novo assembly with an alternative assembler was attempted on the eight coronavirus-positive samples using MEGAHIT v1.1.1 (Li et al. 2015) with the following parameters: minimum contig length = 100, minimum kmer size = 15, maximum kmer size = 101, increment of kmer size of each iteration = 2. To search for potential coronavirus genomes, the 20 longest contigs from each assembly were selected and analyzed using BLASTn on the nt/nr databases, which resulted in the identification of longer coronavirus contigs spanning and extending shorter contigs already identified. Further assembly was attempted on the combined set of contigs using Geneious v.11 software (https://www.geneious.com/), resulting in full or near-full genomes for four bats. Reads were mapped back to the genomes using bowtie2 (Langmead and Salzberg 2012) to correct ambiguous bases.

We also mapped all the sequenced reads from individual samples back to the coronavirus contigs using bowtie2 (v2.2.9) (Langmead and Salzberg 2012). We did this in order to confirm which samples the sequences came from and identify any potential cases of bleed over (the misidentification of the sample from which each sequence read originated) following Kircher et al. (2012) and Jensen et al. (2015).

We identified three full-length and one partial alphacoronavirus genomes from four individual bats (accession numbers OL807608, OL807609, OL807610, OL807611; Supplementary File 1). Three of these were isolated from the species *Chaerephon pumilus*: two of the full genomes (from Bat143 and Bat151; 27,956 nt and 28,061 nt respectively) and one partial genome (Bat180; 20,826 nt). The best hit using BLASTn for all three of these coronavirus genomes was *Chaerephon* bat coronavirus/Kenya/KY22/2006 from Kenya (Tong et al. 2009). When aligned in Geneious, all three were 86–87% identical to this species. When aligned in different tribes (Vespertilionini and Pipistrellini, respectively) (Monadjem et al. 2020a); thus it is not unexpected that the coronaviruses from these genera would be relatively similar.

In addition to these full- and near-full-length genomes, we identified 75 shorter coronavirus genome fragments ranging from 103 to 5241 nt from five bats belonging to the species: *Chaerephon pumilus* (Bat180), *Mops condylurus* (Bat166), *Scotophilus dinganii* (Bat167), and *Eptomops wahlbergi* (Bat50, Bat76) (accession numbers OM000306–OM000380; Supplementary File 3). Only alphacoronavirus...
sequences were isolated from molossids (Chaerephon pumilus and Mops condylurus) and vespertilionids (Scotophilus dinganii). The sequences detected in the fruit bat Epomophorus wahlbergi were either betacoronaviruses (the same genus as SARS-CoV-2 (Zhou et al. 2020; Gorbaleiya et al. 2020), SARS-CoV-1 (Li et al. 2005), and MERS-CoV (Memish et al. 2014)), or unclassified coronaviruses (Tables 1, and 2). These sequences were short, ranging from

| Family       | Species                        | Number of captured individuals | Detected CoV (no. of samples) | Proportion of CoV-positive individuals (%) |
|--------------|--------------------------------|-------------------------------|------------------------------|------------------------------------------|
| Pteropodidae | Epomophorus wahlbergi           | 9                             | 2                            | 15                                       |
| Emballonurida| Taphozous mauritianus           | 2                             | 0                            | 0                                        |
| Molossidae   | Chaerephon pumilus              | 18                            | 3                            | 17                                       |
|              | Mops condylurus                 | 7                             | 1                            | 14                                       |
|              | Mops midas                      | 1                             | 0                            | 0                                        |
| Vespertilionida| Afronycteris nana              | 1                             | 1                            | 100                                      |
|              | Scotophilus dinganii            | 3                             | 1                            | 33                                       |
|              | Scotophilus viridis             | 3                             | 0                            | 0                                        |
| Total        |                                | 44                            | 8                            | 18                                       |

Figure 2. Maximum likelihood phylogeny of coronaviruses (CoVs) based on full genomes, including reference genomes and the four full-length and partial genomes from this study, which are labelled in red and can be retrieved under accession numbers OL807608, OL807609, OL807610, OL807611. Stars indicate branches with 100% bootstrap support.

Table 1. Table summarizing coronavirus detection in bats in northeast Eswatini.
106–517 nt (mostly < 200 nt), with 71–93% percent identity to previously described coronaviruses. They appear most closely related to coronaviruses sequenced from other pteropodid fruit bats, particularly *Eidolon helvum* in Kenya (Tong et al. 2009), Cameroon (Yinda et al. 2018), and Nigeria (Leopardi et al. 2016), and to a lesser extent *Rousettus leschenaultii* in southern China (Woo et al. 2007). None appeared particularly closely related to any human betacoronavirus pathogens. Targeted PCR of specific genes, such as the RNA-dependent RNA polymerase (RdRp), which is widely used in studies of animal coronaviruses (Drexler et al. 2014), could provide further information about specific lineages of coronaviruses, including the betacoronaviruses. However, lack of material prevents such endeavors at this time.

In conclusion, from a sample of 44 bats in Eswatini, we detected both alpha- and betacoronaviruses. All eight bats from which coronaviruses were detected were captured leaving roosts in houses, churches, or within human settlements. More research is necessary to determine whether any of these detected coronaviruses could be a concern for the health of humans or livestock. Limiting direct contact with these bats or their feces might possibly aid in preventing future emerging infectious diseases, while continued monitoring may shed light on the diversity and ecology of coronaviruses.

**Acknowledgements**

We would like to thank Hervé Echecolona, Phumlile Simelane, Mduduzi Ngwenya, and Zanele Dlamini for assistance in the field and Mandla Motsa, Smart Shabangu, Tal Fineberg the staff at Mbuluzi Game Reserve, Thea Litschka-Koen, Clifton Koen, Nick Jackson, and Kim Roques and All-Out Africa, for help with logistics. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1315138 (J.T.S), a National Science Foundation Graduate Research Opportunities Worldwide grant (J.T.S.), Innovation Fund Denmark (The Genome Denmark platform, grant no. 019-2011-2), a Student Research Grant from Bat Conservation International (J.T.S), a University of Florida Biodiversity Institute Fellowship (J.T.S), the Zuckerman STEM Leadership Program (J.T.S), and an NIH Grant (1R01GM114362) (N.D.N). Initial phylogenetic analyses used the Extreme Science and Engineering Discovery Environment (XSEDE) resources, which is supported by National Science Foundation grant number ACI-1053575. XSEDE resources were provided by project allocation TG-ASC160034.

**Open Access**

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Altschul S, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410. https://doi.org/10.1016/0022-2836(90)90570-k

Anthony SJ, Epstein JH, Murray KA, Navarrete-Macias I, Zambrana-Torrello CM, Solovoyov A, Ojeda-Flores R, Arrigo NC, Islam A, Khan SA, Hosseini P, Bogich TL, Olival KJ, Sanchez-Leon MD, Karesh WB, Goldstein T, Luby SP, Morse SS, Mazet JAK, Daszak W, Lipkin WI (2013) A strategy to estimate unknown viral diversity in mammals. *mBio* 4:215:403–410. https://doi.org/10.1128/mBio.00598-13

Anthony SJ, Johnson CK, Greig DJ, Kramer S, Che X, Wells H, Hicks AL, Joly DO, Wolfe ND, Daszak W, Karabas W, Lipkin WI, Morse SS, Mazet JAK, Goldstein T (2017) Global patterns in coronavirus diversity. *Virus Evolution* 3:1814–1820. https://doi.org/10.1093/virusev/vey012

Boisvert S, Raymond F, Godzaridis É, Laviollette F, Corbeil J, Wold B, Myers R, Brenner S, McPherson J, Mardis E, Compeau P, Pezzer T, Tesler G, Flicek P, Birney E, Isqal Z, Caccamo M, Turner I, Flicek P, McVean G, Miller J, Koren S, Sutton G, Salzberg S, Teangten T, Salzberg S, Lorenz P, Eck J, Scholz M, Lo C, Chain P, Schoenfeldt T, Patterson M, Richardson P, Womack K, Young M, Mead D, Varin T, Lovejoy C, Jungblut A, Vincent W, Corbeil J, Varin T, Lovejoy C, Jungblut A, Vincent W, Corbeil J, Narasingarao P, Podell S, Ugalde J, Brochier-Armanet C, Emerson J, Brooks J, Heidelberg K, Bandfield J, Allen E, Tringe S, Mering C von, Kobayashi A, Salamov A, Chen K, Chang H, Podar M, Short J, Mathur E, Detter J, Bork P, Hugenholtz P, Rubin E, Tyson G, Chapman J, Hugenholtz P, Allen E, Ram R, Richardson P, Solovoy V, Rubin E, Rohkars D, Bandfield J, Cho I, Blaser M, Gill S, Pop M, Deboy R, Eckburg P, Turnbaugh P, Samuel B, Gordon J, Relman D, Fraser-Liggett C, Nelson K, Qin J, Raes J, Arumugam M, Burgdorf K, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende D, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto J, Hansen T, Paslier D Le, Linneberg A, Nielsen H, Pelletier H, Lepage P, Le, Yamada T, Mende D, Fernandes G, Tap J, Bruls T, Batto J, Bertalan M, Borzul N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattemori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen H, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tima S, Consortium T, Schloss P, Handelsman J, Liu B, Gibbons T, Ghodsi M, Pop M, Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Hutterbener C, McDonald D, Price M, Goodrich J, Njao JH, DeSantis T, Probst A, Andersen G, Knight R, Hugenholtz P, Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry J, Davis A, Dolinski K, Dwight S, Eppig J, Harris M, Hill D, Issel-Tarver L, Kasarinsks A, Lewis S, Matese J, Richardson J, Ringwald M, Rubin G, Sherlock G, Simpson J, Wong K, Jackman S, Schein J, Jones S, Birol I, Boisvert S, Laviollette F, Corbeil J, Schatz M, Langmead B, Salzberg S, Huson D, Mitra S, Ruchowey H, Weber N, Schuster S, Meyer F, Paarmann D, D M, Dixon P, Caporaso J, Kuczynski J, Stombaugh J, Bittinger K, Bushman F, Costello E, Fierer N, Pena A, Goodrich J, Gordon J, Hulttelly G, Kelley S, Knights D, Koenig J, Ley R, Lozupone C, McDonald D, Muegge B, Pirrung M, Reeder J, Sevinsky J, Turnbaugh P, Walters W, Widmann J, Yatsunenko T, Zaneveld J, Knight R, Krause L, Diz N, Goessmann A, Kelley S, Nat- kemper T, Rohwer F, Edwards R, Stove J, Brady A, Salzberg S, Namiki T, Hachiya T, Tanaka H, Sakakibara Y, Peng Y, Leung H, Yu S, Ching F, Larsenson J, Jojic V, Koller D, Wu G, Chen J, Hoffmann C, Bittinger K, Chen Y, Keilbaugh S, Betwra M, Knights D, Walters W, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman F, Lewis J, Pezzer N, Tan H, Waterman M, Kurtz S, Phillippy A, Delcher A, Smoot M, Shumway M, Antonescu C, Salzberg S, Schadt E, Linderman M, Sorenson J, Lee L, Nolan G, Benson D, Boguski M, Lipman D, Ostell J, Kulikova T, Aldebert P, Athorpe N, Baker W, Bates K, Browne P, Broek A van den, Cognacane G, Duggan K, Eberhardt R, Faruque N, Garcia-Pastor M, Harte N, Kanz C, Leinenon R, Lin Q, Lombard V, Lopez R, Mancuso R, Mc Hale M, Nardone F, Silviento N, Stoehr P, Stoesser G, Ann M, Tzouvara K, Vaughan R, Wu D, Zhu W, Apweiler R, Camon E, Magrane M, Barrett D, Lee V, Dimmer E, Maslen J, Birks D, Harte N, Lopez R, Apweiler R, Gabriel E, Fagg G, Bosilica G, Angskun T, Dongarra J, Squirees J, Sahay V, Kambadur P, Barrett B, Lumsdaine A, Castain R, Daniel D, Graham R, Woodall T, Gabriel E, Fagg G, Bosilica G, Angskun T, Dongarra J, Squirees J, Sahay V, Kambadur P, Barrett B, Lumsdaine A, Castain R, Daniel D, Graham R, Woodall T, Gropp W (2012) Ray Meta: Scalable de novo metagenome assembly and profiling. *Genome Biology* https://doi.org/10.1186/gb-2012-13-12-r22

Buchfink B, Xie C, Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12:59–60. https://doi.org/10.1038/nmeth.3176

De Sabato L, Lelli D, Faccin F, Canziani S, Di Bartolo I, Vaccari G, Moreno A (2019) Full genome characterization of two novel Alpha-coronavirus species from Italian bats. *Virus Research* 260:60–66. https://doi.org/10.1016/j.virusres.2018.11.007

Drexler JE, Corman VM, Drosten C (2014) Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. *Antiviral Research* 101:45–56. https://doi.org/10.1016/j.antivi.2013.10.013

Fenton MB, Jacobs DS, Richardson EJ, Taylor PJ, White W (2004) Individual signatures in the frequency-modulated sweep calls of African large-eared, free-tailed bats *Otomops martiensseni* (Chiroptera: Molossidae). *J Zool* 262:11–19. https://doi.org/10.1017/S095283690300431X

Geldenhuys M, Mortlock M, Weyer J, Bezuidt O, Seamark ECJ, Kearney T, Gleasner C, Erkhhila TH, Cui H, Markottker W (2018) A metagenomic viral discovery approach identifies potential zoonotic and novel mammalian viruses in *Neoromicia* bats within South Africa. *PLOS ONE* 13:0194527. https://doi.org/10.1371/journal.pone.0194527

Geldenhuys M, Weyer J, Nel LH, Markottker W (2013) Coronavirus in South African bats. *Vector-Borne and Zoonotic Diseases* 13:516–519. https://doi.org/10.1089/vbz.2012.1101

Gorbalenya EA, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, Haagmans BL, Lauber C, Leontovich AM, Neu- man BW, Ponzar D, Perlman S, Poon LLM, Samborskiy DV, Sidorov IA, Solis Z, Ehlberries J (2020) The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* 5:536–544
Woo PCY, Wang M, Lau SKP, Xu H, Poon RWS, Guo R, Wong BHL, Gao K, Tsoi H-w, Huang Y, Li KSM, Lam CSF, Chan K-h, Zheng B-j, Yuen K-y (2007) Comparative Analysis of Twelve Genomes of Three Novel Group 2c and Group 2d Coronaviruses Reveals Unique Group and Subgroup Features. *Journal of Virology* 81:1574–1585. https://doi.org/10.1128/JVI.02182-06

Yinda CK, Ghogomu SM, Conceição-Neto N, Beller L, Deboutte W, Vanhulle E, Maes P, Van Ranst M, Matthijnssens J (2018) Cameroonian fruit bats harbor divergent viruses, including rotavirus H, bastroviruses, and picobirnaviruses using an alternative genetic code. *Virus Evolution*. https://doi.org/10.1093/ve/vey008

Zhou P, Lou Yang X, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Di Jiang R, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270–273. https://doi.org/10.1038/s41586-020-2012-7