Title: Bat coronavirus phylogeography in the western Indian Ocean

Running title: Bat coronavirus in the western Indian Ocean

Authors: Léa Joffrin*, Steven M. Goodman, David A. Wilkinson, Beza Ramasindrazana, Erwan Lagadec, Yann Gomard, Gildas Le Minter, Andréa Dos Santos, M. Corrie Schoeman, Rajendraprasad Sookhareea, Pablo Tortosa, Simon Julienne, Eduardo S. Gudo, Patrick Mavingui and Camille Lebarbenchon

Author affiliations :

Université de La Réunion, UMR Processus Infectieux en Milieu Insulaire Tropical (PIMIT) INSERM 1187, CNRS 9192, IRD 249, Sainte-Clotilde, La Réunion, France (L. Joffrin, B. Ramasindrazana, E. Lagadec, Y. Gomard, G. Le Minter, D.A. Wilkinson, P. Tortosa, P. Mavingui, C. Lebarbenchon)

Association Vahatra, Antananarivo, Madagascar (S.M. Goodman, B. Ramasindrazana)

Field Museum of Natural History, Chicago, USA (S.M. Goodman)

Veterinary Faculty, Eduardo Mondlane University, Maputo, Mozambique (A. Dos Santos)

School of Life Sciences, University of Kwa-Zulu Natal, Kwa-Zulu Natal, South Africa (M.C. Schoeman)

National Parks and Conservation Service, Réduit, Mauritius (R. Sookhareea)

Seychelles Ministry of Health, Victoria, Mahe, Seychelles (S. Julienne)

Instituto Nacional de Saúde, Maputo, Mozambique (E.S. Gudo).

*Corresponding author: lea.joffrin@univ-reunion.fr

UMR PIMIT, 2 rue Maxime Rivière, 97490 Sainte-Clotilde, Reunion Island, France.

Tel: +262 262 93 88 00

Abstract – 147 words

Bats are important reservoirs of zoonotic pathogens, including coronaviruses (CoVs). The Western Indian Ocean (WIO) islands are a biodiversity hotspot with more than 50 bat species. Here we tested 1,099 bats belonging to 39 species from Mozambique, Madagascar, Mauritius, Mayotte, Reunion Island and Seychelles. Based on molecular screening and partial sequencing of the RNA-dependent RNA polymerase gene, a total of 88 bats (8.0% ± 1.6%) tested positive for bat-borne coronaviruses (CoVs), with higher prevalence in Mozambican bats (19.6% ± 4.7%) as compared to those sampled on islands (4.2% ± 1.2%). Phylogenetic analyses revealed that a large diversity of α- and β-CoVs are maintained in bat populations of the WIO, some being genetically related to human CoVs (e.g. NL63, MERS). Finally, we found a strong signal of co-evolution between CoVs and their bat host species with limited evidence for host-switching, except for bat species sharing day roost sites.

Keywords: bat, coronavirus, islands, tropical, evolution, ecology

Text – 3356 words

Introduction

The burden of emerging infectious diseases has significantly increased over the last decades and is recognized as a major global health concern. In 2018, the World Health Organization (WHO) established the "Blueprint priority disease list", identifying viruses such as Ebola, Lassa fever, Middle East Respiratory Syndrome (MERS), and Nipah fever as significant threats to international biosecurity (1). This list also highlights the potential pandemic risk from the emergence of currently unknown zoonotic pathogens, collectively referring to these unknown threats as "disease X" (1). Investigation of the potential zoonotic pathogens in wild animals, particularly vertebrates, is thus critical for emerging infectious diseases preparedness and responses.

Bats represent nearly 1,400 species and live on all continents except Antarctica. They provide key ecosystem services such as crop pest regulation, pollination, seed dispersal, and soil fertilization. Bats are also recognized as reservoirs of many zoonotic pathogens. Several bat-borne coronaviruses (CoVs) have recently emerged in humans and livestock with sometimes major impacts to public health. For instance, in 2003, the Severe Acute Respiratory Syndrome (SARS) outbreak in China spread to 30 countries, infecting 8,096 people and leading to 774 deaths in less than a year (2). The Middle East Respiratory Syndrome (MERS) is caused by a camel-associated CoV that likely originated from bats, and in 2003 afflicted humans in Saudi Arabia and elsewhere, infecting 2,442 people with 842 associated deaths worldwide (3).

Our study area spans geographic locations across the islands of the western Indian Ocean (WIO) and southeastern continental Africa (SECA) (Figure 1). These land areas have diverse geological origins that have influenced the process of bat colonization and species distributions. The ecological settings and species diversity on these islands for bats are notably different. On Madagascar, more than 45 bat species are known to occur, of which more than 80

% are endemic to the island. The smaller studied islands of the WIO, Mauritius, Mayotte, Reunion Island, and Mahé (Seychelles), host reduced bat species diversity (e.g. three species on Reunion Island), whereas SECA supports a wide range of bat species. To date, several studies have identified bat-infecting CoVs in countries of continental Africa, including Zimbabwe (4), South Africa (5), and Kenya (6). CoVs have also been reported in fruit bat (Pteropodidae) populations of Madagascar, where beta-coronaviruses belonging to the D-subgroup were identified in the two bat species *Eidolon dupreanum*, and *Pteropus rufus* (7).

In this study, we investigated the presence of CoVs in over 1,000 individual bats belonging to 39 species and sampled on five islands (Madagascar, Mauritius, Mayotte, Reunion Island, and Mahé) and one continental area (Mozambique). Based on molecular screening and partial sequencing of the RNA-dependent RNA polymerase gene, we (i) estimated CoV prevalence in the regional bat populations, (ii) assessed CoVs genetic diversity, and (iii) identified potential association between bat families and CoVs and evolutionary drivers leading to these associations.

Materials and methods

Origin of the tested samples

Samples obtained from vouchered bat specimens during previous studies in Mozambique (February to May 2015), Mayotte (November to December 2014), Reunion Island (February 2015), Seychelles (February to March 2014), Mauritius (November 2012) and Madagascar (October to November 2014) were tested (8–11) (Technical Appendix). We also collected additional swab samples from several synanthropic bat species on Madagascar, in January 2018 (Technical Appendix). A total of 1,099 bats were tested (Figure 1). Details on sample types, bat families, species, and locations are provided in Appendix Table S1.

87

88 **Molecular detection**

89 RNA was extracted from 140 µL of each sample using the QIAamp Viral RNA mini kit
90 (QIAGEN, Valencia, California, USA), and eluted in 60 µL of Qiagen AVE elution buffer. For
91 bat organs, approximately 1 mm^3 of tissue (either lungs or intestines) was placed in 750 µL of
92 DMEM medium and homogenized in a TissueLyser II (Qiagen, Hilden, Germany) for 2 min at
93 25 Hz using 3 mm tungsten beads, prior to the RNA extraction. Reverse transcription was per-
94 formed on 10 µL of RNA using the ProtoScript II Reverse Transcriptase and Random Primer 6
95 (New England BioLabs, Ipswich, MA, USA) under the following thermal conditions: 70 °C for
96 5 min, 25 °C for 10 min, 42 °C for 50 min, and 65 °C for 20 min (12). cDNAs were tested for
97 the presence of the RNA-dependent RNA-polymerase (RdRp) gene using a multi-probe Real-
98 Time (RT) PCR (13). The primer set with Locked Nucleic Acids (LNA; underlined position in
99 probe sequences) was purchased from Eurogentec (Seraing, Belgium): 11-FW: 5'-TGA-TGA-
100 TGS-NGT-TGT-NTG-YTA-YAA-3' and 13-RV: 5'-GCA-TWG-TRT-GYT-GNG-ARC-
101 ARA-ATT-C-3'. Three probes were used: probe I (ROX): 5'-TTG-TAT-TAT-CAG-AAT-
102 GGY-GTS-TTY-AT-3', probe II (FAM): 5'-TGT-GTT-CAT-GTC-WGA-RGC-WAA-ATG-
103 TT-3', and probe III (HEX): 5'-TCT-AAR-TGT-TGG-GTD-GA-3'. RT-PCR was performed
104 with ABsolute Blue QPCR Mix low ROX 1X (Thermo Fisher Scientific, Waltham, MA, USA)
105 and 2.5 µL of cDNA under the following thermal conditions: 95 °C for 15 min, 95 °C for 30 s,
106 touchdowns from 56 °C to 50°C for 1 min and 50 cycles with 95 °C for 30 s and 50 °C for 60
107 s in a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA).

108 Because of the limited size of the sequence generated from the RT-PCR, a second PCR
109 targeting 440 bp of the RdRp gene was performed with 5 µL of cDNA of each positive sample,
110 with the following primer set: IN-6: 5'-GGT-TGG-GAC-TAT-CCT-AAG-TGT-GA-3' and IN-

7: 5'-CCA-TCA-TCA-GAT-AGA-ATC-ATC-ATA-3' (14). PCRs were performed with the GoTaq G2 Hot Start Green Master Mix (Promega, Madison, WI, USA) in an Applied Biosystems 2720 Thermal Cycler (Thermo Fisher Scientific, Waltham, MA, USA). After electrophoresis in a 1.5% agarose gel stained with 2% GelRed (Biotium, Hayward, CA, USA), the remaining amplicons of the expected size were directly sequenced on both strands by Genoscreen (Lille, France). All generated sequences were deposited in GenBank under the accession numbers MN183146 to MN183273.

Statistical analysis

We tested the effect of bat family, species, sex, as well as sampling location and roosting habitat (caves, outdoor or buildings) on the probability for detecting CoV RNA using Pearson χ^2 tests. The effect of bat age (adult *vs* juvenile) and female reproductive status (pregnant *vs* not pregnant) was investigated for species in which we detected at least 15 CoV positive individuals. Finally, the effect of sampling during the wet (February) or dry (May) season was investigated in Mozambique. Analyses were conducted with R v3.5.1 software (15).

Phylogenetic analyses

Sequences obtained with the second PCR system (14) were edited with the Chromas Lite Software package version 2.6.4 (16). We explored CoV diversity of the sequences with pairwise identity values obtained from *seqidentity* function in R *bio3d* package v2.3-4 (17) and identified the most similar CoV RdRp sequences referenced in GenBank using BLASTN 2.2.29+. An alignment was then generated using the 51 nucleotide sequences obtained in this study and 151 reference CoV sequences from a large diversity of host family and geographic origins (Europe, Asia, Oceania, America and Africa), with CLC Sequence viewer 8.0 Software

(CLC Bio, Aarhus, Denmark). A phylogenetic tree was obtained by maximum likelihood using MEGA Software v10.0.4 (18), with 1,000 bootstrap iterations, and the best evolutionary model for our dataset as selected by modelgenerator v0.85 (19).

Host-virus associations were investigated using the phylogeny of WIO bats and their associated CoVs. Bat phylogeny was generated from an alignment of 1,030 bp of mitochondrial Cytochrome *b* (Cyt *b*) gene sequences downloaded from GenBank and sequenced for this study (Appendix Table S2), for each CoV positive bat species with available sequence data. Finally, bat and pruned CoV phylogenies based on each 393 bp RdRp unique sequence fragment were generated by Neighbor-Joining with 1,000 bootstrap iterations, using CLC Sequence viewer 8.0 Software (CLC Bio, Aarhus, Denmark)(20). Phylogenetic congruence was tested (21) to assess the significance of the coevolutionary signal between bat host species and CoVs sequences, using *ParaFit* with 999 permutations in the '*ape'* package v5.0 in R 3.5.1 (22). Tanglegram representations of the co-phylogeny were visualized using the Jane software v4.01 (23).

Results

Prevalence of CoV

In total, 88 of the 1,099 bat samples tested positive for CoV by RT-PCR (mean detection rate ± 95% confidence interval: 8.0% ± 1.6%). The prevalence of positive bats was different according to the sampling locations ($\chi^2 = 70.2$; $p<0.001$), with a higher prevalence in Mozambique (19.6% ± 4.7%) than on all WIO islands (4.2% ± 1.2%) (Figure 2). A significant difference in the prevalence of positive bats was also detected between families ($\chi^2 = 44.2$; $p<0.001$; Appendix Figure S1). The highest prevalence was observed in the families Nycteridae (28.6 % ± 23.6%), Rhinolophidae (24.6% ± 10.5%), Hipposideridae (11.9% ± 6.9%), and Rhinonycteridae (10.7% ± 5.5%). Bat species ($\chi^2 = 156.27$; $p<0.001$; Appendix Figure S2) and roosting

habitat (χ^2 = 23.76; p<0.001; Figure 3) had a significant effect on the probability of CoVs detection, but not host sex (χ^2 = 2.32; p = 0.13). For Mozambican *Hipposideros caffer* and *Mops condylurus*, bat age (adult *vs* juvenile) had no effect on the probability of CoVs detection (χ^2 = 0.1; p = 0.8 and χ^2 = 0.1; p = 0.7, respectively). For *Chaerephon pusillus*, female reproductive status (pregnant or not pregnant) had no effect on the probability of CoVs detection (χ^2 = 2.50; p = 0.3). Finally, the prevalence of CoV positive bats in Mozambique was significantly different (χ^2= 21.5; p<0.001; Appendix Figure S3) between February (35.1% ± 4.9%) and May (11.0% ± 2.3).

RdRp sequence diversity

Of the 88 positive samples, we obtained 77 partial RdRp sequences using the RT-PCR detection system (179 bp) and 51 longer partial RdRp sequences using the standard PCR system (440 bp): this latter was subsequently used for phylogenetic analyses. Details of the sequenced CoV-positive samples are given in Appendix Table S3. Pairwise comparison of these 51 sequences revealed 28 unique sequences, and sequences similarities ranging from 60.2% to 99.8%. The lowest sequence similarity was found in Mozambique (60.2% to 99.8%), then in Madagascar (64.0% to 99.8%). No genetic variation was observed for samples from Mayotte and Reunion Island.

Phylogenetic structure of CoVs

Sequence comparison indicated that WIO bats harbor a high diversity of both α and β-CoVs, with conserved clade groups clustering mostly by bat family (Figure 4). Specifically, 25 sequences were identified as α-CoVs, and three sequences were genetically related to the β-CoVs. For α-CoVs, all sequences detected in our study of members of the family Molossidae

formed a highly supported monophyletic group, including CoV sequences from Molossidae bats previously detected in continental Africa (Figure 5). CoVs detected in *Mops condylurus* (Mozambique), *Mormopterus francoismoutoui* (Reunion Island), *Chaerephon pusillus* and *Chaerephon* sp. (Mayotte), and *Mormopterus jugularis* (Madagascar) shared 90% - 98% nucleotide similarity with a CoV detected in *Chaerephon* sp. in Kenya (Appendix Table S4). All CoVs found in Miniopteridae clustered in a monophyletic group, including Miniopteridae CoVs sequences from Africa, Asia, and Oceania (Appendix Table S4). The great majority of α-CoVs detected in Rhinolophidae bats clustered in two monophyletic groups (Figure 4); one with African Rhinolophidae CoVs and one with Asian Rhinolophidae CoVs. We additionally detected one CoV from *Rhinolophus rhodesiae*, which was 100% similar to a Miniopteridae CoV from this study. Rhinonycteridae CoVs formed a single monophyletic group with NL63 Human CoVs. The Rhinonycteridae CoVs detected clustered with NL63-related bat sequences found in *Triaenops afer* in Kenya (Figure 6) and showed 85% similarity to NL63 Human CoVs (Appendix Table S4). Hipposideridae α-CoVs mainly clustered into a single monophyletic group. Hipposideridae CoV sequences from this study clustered with a 229E Human CoV-related bat sequence found in *Hipposideros vittatus* from Kenya (Figure 7) and demonstrated 93% similarity to 229E Human CoV (Appendix Table S4).

Regarding the β-CoVs, two sequences obtained from *Nycteris thebaica* clustered in the C-subgroup of β-CoVs together with other CoVs previously reported in African *Nycteris sp.* bats (Figure 8). The sequences showed 88% nucleotide identity to a β-C CoV found in *Nycteris gambiensis* in Ghana (Appendix Table S4). *Rousettus madagascariensis* CoV clustered with Pteropodidae CoVs belonging to the D-subgroup of β-CoVs (Figure 9). BLAST queries against the NCBI database showed 98% nucleotide identity between CoV sequences from *Rousettus*

madagascariensis and a β-D CoV sequence detected in *Eidolon helvum* from Kenya (Appendix Table S4).

Co-phylogeny between bats and CoVs

Co-phylogeny tests were conducted using 11 Cyt *b* sequences obtained from the 11 CoVs positive bat species and 27 partial CoV RdRp sequences (440 bp). Results supported co-evolution between the WIO bats and their CoVs (ParaFitGlobal = 0.04; p = 0.001) and a high level of phylogenetic congruence (Figure 10).

Discussion

We provide evidence for a high diversity of CoVs in bat populations on WIO islands. The overall prevalence of CoV positive bats was consistent with other studies from continental Africa (5) and island systems in the Australasian region (24). However, in the case of WIO islands, prevalence rates showed considerable variations based on bat family, species, landmass, and season.

We found a significant effect of the roosting habitat on the probability of CoV detection, with higher detection in bats occupying caves and buildings than outdoor habitats. Bats use different types of day roosts sites, including caves, rock crevices, tree cavities, forest vegetation, or synanthropic structures (25). Critical for this study, roosting site choice could be a risk factor for infection of bats (26). This was indeed the case with higher detection rates for bats using caves and buildings, as compared to other roost sites. Shelters like caves or buildings (bridges, houses, etc.) may protect excreted viral particles from rainfall, temperature, humidity variation, and ultraviolet radiation (27,28). Thus, these roost types may facilitate the maintenance and transmission of viruses between syntopic species. Moreover, the accumulation of guano in these

231 confined environments compared to outdoor habitats could favor virus transmission in bat pop-
232 ulations. However, our investigation of infectious CoV particles in bat guano only detected
233 RNA, without demonstrating the infectious potential of viral particles from this environment
234 (29).

235 Host specificity is well known for some bat CoVs subgenera (30,31). For example, β-C
236 CoVs are largely associated with Vespertilionidae bats, whereas β-D CoVs are found mostly in
237 Pteropodidae (31,32). In our study, we showed that WIO bats harbor genetic structured CoVs,
238 of both α-CoV and β-CoV subclades, clustering mostly by bat family. In the new CoV taxon-
239 omy based on full genomes proposed by the International Committee of Taxonomy of Viruses
240 (ICTV), α-CoVs and β-CoVs are split in subgenera mostly based on host families (33), reflected
241 in the subgenera names (e.g. Rhinacovirus for a Rhinolophidae α-CoV cluster, Minuacovirus
242 for a Miniopteridae α-CoV cluster, Hibecovirus for an Hipposideridae β-CoV cluster). Alt-
243 hough our classification was based on a partial sequence of the RdRp region, we identified
244 sequences from samples belonging to four of these subgenera (Minuacovirus, Duvinacovirus,
245 Rhinacovirus, and Nobecovirus) and three that could not be classified according to this taxo-
246 nomic scheme hence representing unclassified subgenera (we propose "Molacovirus", "Nycbe-
247 covirus", and "Rhinacovirus2").

248 In the context of WIO islands, we expect a strong geographical influence on CoVs diversity
249 rather than on host specificity, with independent evolution of CoVs on each island because of
250 spatial isolation and endemism. Anthony et al. (32) suggested that the dominant evolutionary
251 mechanism for African CoVs was host switching. In our study, congruence between host and
252 viral phylogenies suggests a strong signal for co-evolution between WIO bats and their associ-
253 ated CoVs. This strongly suggests that individual bat species harbor specific CoV lineages, but
254 CoV transmission between different bat species may nevertheless occur infrequently. However,

the geographical influence seems to occur within family specific groups, as in the WIO Molossidae CoV clade. Endemism resulting from geographic isolation may then have favored viral diversification within family specific viral lineages.

Although co-evolution could be the dominant mechanism, host-switching may take place in certain situations. For example, in Mozambique we found a potential Miniopteridae CoV in a Rhinolophidae bat co-roosting with Miniopteridae in the same cave. These host-switching events could be favored when several bat species roost in the same portion of a cave *in sympatria* (34). A similar scenario was described on Australia where Miniopteridae CoV was detected in Rhinolophidae bats (24). These infrequent host-switching events show that spillovers can happen but suggest that viral transmission is not maintained independently within the spillover host. Thus, the co-evolution of virus and host might have resulted in strong adaptation of the CoVs to the specific bat species. In addition, viral factors (mutation rate, recombination propensity, replication ability in the cytoplasm, changes in the ability to bind host cells), environmental factors (climate variation, habitat degradation, decrease of bat preys), and phylogenetic relatedness of host species are also critical for the viral establishment in a novel host (35). Nevertheless, apparent evidence of host switching as a dominant mechanism of CoV evolution could be an artifact of a lack of data for some potential bat hosts, leading to incomplete phylogenetic reconstructions (32).

Several bat CoVs we identified in Rhinonycteridae and Hipposideridae from Mozambique had between 85% and 93% nucleotide sequence similarity with NL63 Human CoVs and 229E Human CoVs, respectively. These two human viruses are widely distributed in the world and associated with mild to moderate respiratory infection in humans (36). Tao et al. established that the NL63 Human CoVs and 229E Human CoVs have a zoonotic recombinant origin from their most recent common ancestor, estimated to be about 1,000 years ago (37). During the past

279 decade, they were both detected in bats in Kenya, and in Ghana, Gabon, Kenya, and Zimbabwe,
280 respectively (4,6,38,39). Moreover, CoVs notably similar to both NL63 and 229E have been
281 described in Kenya and Mozambique bats, suggesting an East African bats origin of these hu-
282 man viruses. Intermediate hosts are important in the spillover of CoVs, despite gaps in direct
283 and indirect transmission routes of bat infectious agents to secondary hosts (40). This hypothe-
284 sis has been formulated for the 229E Human CoV, with an evolutionary origin in Hippo-
285 sideridae bats and with camelids as intermediate hosts (39). Similarly, the spillover of NL63
286 from Rhinonycteridae bats to humans might have occurred through a currently unidentified
287 intermediate host (6,41,42). Because receptor recognition by viruses is the first essential cellular
288 step to infect host cells, CoVs may have spilt over into humans from bats through an interme-
289 diate host possibly due to mutations on spike genes (6,43). Further investigations of CoVs in
290 Kenyan and Mozambican livestock and hunted animals could potentially provide information
291 on the complete evolutionary and emergence history of these viruses before their establishment
292 in humans.

293 MERS-like CoV, with high sequence similarity (>85%) to human and camel strains of
294 MERS-CoV, have been detected in *Neoromicia capensis* in South Africa and *Pipistrellus* cf.
295 *hesperidus* in Uganda, suggesting a possible origin of camel MERS-CoV in Vespertilionidae
296 bats (5,32,44). This family has been widely studied, with 30% of all reported bat CoVs se-
297 quences from the past 20 years coming from vespertilionids (45), including MERS-like CoVs.
298 No members of this family were positive for CoV in our study, which may be associated with
299 the low number of individuals sampled; additional material is needed to explore potential
300 MERS-like CoV on Madagascar.

301 CoV transmission risk also depends on ecological factors and human actions such as en-
302 croachment, landscape uses, and cultural traditions. On Madagascar, for example, bats are

hunted for commercial or personal consumption (46,47). Further, certain Malagasy ethnic groups have cultural rituals associated with caves, including those with bat roosts (47). Indirect contact with contaminated environments such as bat guano may also increase infection risk. Given the strong co-evolution of CoV and bats in the WIO, there is the potential of human populations being exposed to CoV spillovers, which in part depends on contact with day roost sites and different ecological contexts, as well as the bat families present. Moreover, there is evidence of a relationship between biodiversity loss and the risk of infectious disease emergence (48).

At a worldwide scale, epidemiological studies with serological surveys on human populations are necessary to investigate potential CoV transmission from bats to humans (49,50) and should be considered on WIO islands. Even if CoV prevalence in bats is higher on Madagascar than other oceanic islands around the world, the high CoV genetic diversity in Malagasy bats could complicate the development of detection protocols in human populations. On oceanic islands, however, the development of serological tests for particular coronaviruses found in endemic bat species is achievable, and underlines the need to develop virus surveillance in local populations based on results from bat studies.

Acknowledgments

We are very grateful to S. Muradrasoli for recommendations on the optimization of the multi-probe PCR protocol as well as for providing PCR controls and to C. Cordonin for providing bat Cyt *b* sequences. L. Biscornet, C. Dionisio, L. Domergue, M. Dietrich, T. Mbohoahy, T. Nekena, J. Rakotoarivelo, M. Rakotomanga, C. F. Rakotondramanana, and A. Randrenjarison are thanked for their assistance in the field. We also thank S. Bos and A. Hoarau for their help in the laboratory, and K. Dellagi and H. Pascalis for the development and the management of the 'partenariat Mozambique-Réunion dans la recherche en santé: pour une approche intégrée d'étude des maladies infectieuses à risque épidémique (MoZaR)' research program.

Funding

Field research was funded by the 'Pathogènes associés à la Faune Sauvage Océan Indien (FS-OI)', the 'Leptospirose Ocean Indien (LeptOI)', the 'Paramyxovirus Océan Indien (Para-myxOI)', and the 'Partenariat Mozambique-Réunion dans la recherche en santé : pour une ap-proche intégrée d'étude des maladies infectieuses à risque épidémique (MoZaR)' programs (Fond Européen de Développement Régional, Programme Opérationnel de Coopération Terri-toriale). Fieldwork on Mayotte was funded by the 'Centre National de la Recherche Scien-tifique' (Projets Exploratoires Premier Soutien BATMAN). Molecular analyses were finan-cially supported by tutorship institutions of the UMR PIMIT. LJ is a recipient of a 'Région Réunion, European Regional Development Funds (FEDER 2014-2020)' PhD fellowship. BR's post-doctoral fellowship was supported by the 'Run Emerge' European Union's Seventh Framework Program (FP7/2007–2013; Grant agreement NO 263958), the 'Fonds de Coopé-

ration Régionale, Prefecture de La Réunion', and The Field Museum of Natural History, Chicago, through the Ralph and Marian Falk Medical Research Trust. CL is supported by a 'Chaire mixte : Université de La Réunion – INSERM'. The funding agencies were not involved in the study design, implementation or publishing of this study, and the research presented herein represents the opinions of the authors but not necessarily the opinions of the funding agencies.

References

1. Reusken C, Obasanya J, Messori S, Vasconcelos P, Wilkinson A, Khellef N, et al. WHO Research and Development Blueprint 2018 Annual review of diseases prioritized under the Research and Development Blueprint Informal consultation. Geneva, Switzerland; 2018.
2. World Health Organization. Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003 Based on data as of the 31 December 2003. [Internet]. World Health Organization. 2003. Available from: https://www.who.int/csr/sars/country/table2004_04_21/en/
3. World Health Organization. MERS situation update [Internet]. 2018. Available from: http://www.emro.who.int/health-topics/mers-cov/mers-outbreaks.html
4. Bourgarel M, Pfukenyi DM, Boué V, Talignani L, Chiweshe N, Diop F, et al. Circulation of alphacoronavirus, betacoronavirus and paramyxovirus in *Hipposideros* bat species in Zimbabwe. Infect Genet Evol. 2018 Mar 1;58:253–7.
5. Ithete NL, Stoffberg S, Corman VM, Cottontail VM, Richards LR, Schoeman MC, et al. Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. Emerg Infect Dis. 2013;19(10):1697–9.
6. Tao Y, Shi M, Chommanard C, Queen K, Zhang J, Markotter W, et al. Surveillance of bat coronaviruses in Kenya identifies relatives of human coronaviruses NL63 and 229E and their recombination history. J Virol. 2017;91(5):e01953-16.
7. Razanajatovo NH, Nomenjanahary LA, Wilkinson DA, Razafimanahaka JH, Goodman SM, Jenkins RK, et al. Detection of new genetic variants of betacoronaviruses in endemic frugivorous bats of Madagascar. Virol J. 2015 Dec 12;12(1):42.
8. Wilkinson DA, Temmam S, Lebarbenchon C, Lagadec E, Chotte J, Guillebaud J, et al. Identification of novel paramyxoviruses in insectivorous bats of the Southwest Indian Ocean. Virus Res. 2012;170(1–2):159–63.
9. Wilkinson DA, Mélade J, Dietrich M, Ramasindrazana B, Soarimalala V, Lagadec E, et al. Highly diverse morbillivirus-related paramyxoviruses in wild fauna of the Southwestern Indian Ocean islands: evidence of exchange between introduced and endemic small mammals. J Virol. 2014;88(15):8268–77.
10. Gomard Y, Dietrich M, Wieseke N, Ramasindrazana B, Lagadec E, Goodman SM, et al. Malagasy bats shelter a considerable genetic diversity of pathogenic *Leptospira* suggesting notable host-specificity patterns. FEMS Microbiol Ecol. 2016;92(4):fiw037.
11. Mélade J, Wieseke N, Ramasindrazana B, Flores O, Lagadec E, Gomard Y, et al. An

eco-epidemiological study of morbilli-related paramyxovirus infection in Madagascar bats reveals host-switching as the dominant macro-evolutionary mechanism. Sci Rep. 2016 Apr 12;6:23752.

12. Lebarbenchon C, Ramasindrazana B, Joffrin L, Bos S, Lagadec E, Le Minter G, et al. Astroviruses in bats, Madagascar. Emerg Microbes Infect. 2017;6:e58.
13. Muradrasoli S, Mohamed N, Hornyák Á, Fohlman J, Olsen B, Belák S, et al. Broadly targeted multiprobe QPCR for detection of coronaviruses: coronavirus is common among mallard ducks (*Anas platyrhynchos*). J Virol Methods. 2009;159(2):277–87.
14. Poon LLM, Chu DKW, Chan KH, Wong OK, Ellis TM, Leung YHC, et al. Identification of a novel coronavirus in bats. J Virol. 2005 Feb;79(4):2001–9.
15. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018.
16. Technelysium Pty. Ltd. Chromas Lite Software. 2018.
17. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 2006 Nov 1;22(21):2695–6.
18. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35(6):1547–9.
19. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol. 2006 Mar 24;6:29.
20. QIAGEN. CLC Sequence Viewer 8. 2018.
21. Legendre P, Desdevises Y, Bazin E. A Statistical test for host-parasite coevolution. Syst Biol. 2002;51(2):217–34.
22. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2018;1–3.
23. Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R. Jane: a new tool for the cophylogeny reconstruction problem. Algorithms Mol Biol. 2010 Feb 3;5:16.
24. Smith CS, de Jong CE, Meers J, Henning J, Wang L-F, Field HE. Coronavirus infection and diversity in bats in the Australasian region. Ecohealth. 2016 Mar 5;13(1):72–82.
25. Kunz T, Lumsden L. Ecology of cavity and foliage roosting bats. In: Kunz TH, Fenton M, editors. Bat Ecology. Chicago, USA: University of Chicago Press, 2005; 2006. p. 24.
26. Mühldorfer K, Speck S, Kurth A, Lesnik R, Freuling C, Mü Ller T, et al. Diseases and causes of death in European bats: dynamics in disease susceptibility and infection rates. PLoS One. 2011;6(12):e29773.
27. Chan KH, Peiris JSM, Lam SY, Poon LLM, Yuen KY, Seto WH. The effects of temperature and relative humidity on the viability of the SARS coronavirus. Adv Virol. 2011;2011:734690.
28. Fogarty R, Halpin K, Hyatt AD, Daszak P, Mungall BA. Henipavirus susceptibility to environmental variables. Virus Res. 2008;132(1–2):140–4.
29. Wacharapluesadee S, Sintunawa C, Kaewpom T, Khongnomnan K, Olival KJ, Epstein JH, et al. Group C betacoronavirus in bat guano fertilizer, Thailand. Emerg Infect Dis. 2019;13(8):1349–51.
30. Leopardi S, Holmes EC, Gastaldelli M, Tassoni L, Priori P, Scaravelli D, et al. Interplay between co-divergence and cross-species transmission in the evolutionary history

of bat coronaviruses. Infect Genet Evol. 2018;58:279–89.
31. Cui J, Han N, Streicker D, Li G, Tang X, Shi Z, et al. Evolutionary relationships between bat coronaviruses and their hosts. Emerg Infect Dis. 2007;13(10):1526–32.
32. Anthony SJ, Johnson CK, Greig DJ, Kramer S, Che X, Wells H, et al. Global patterns in coronavirus diversity. Virus Evol. 2017 Jan 1;3(1):vex012.
33. Ziebuhr J, Baric RS, Baker S, de Groot RJ, Drosten C, Gulyaeva A, et al. ICTV Report 2017.013S. 2017.
34. Luis AD, Hayman DTS, O'Shea TJ, Cryan PM, Gilbert AT, Pulliam JRC, et al. A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? Proc R Soc B Biol Sci. 2013;280(1756):20122753.
35. Parrish CR, Holmes EC, Morens DM, Park E-C, Burke DS, Calisher CH, et al. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol Rev. 2008 Sep;72(3):457–70.
36. Dijkman R, van der Hoek L. Human coronaviruses 229E and NL63: close yet still so far. J Formos Med Assoc. 2009;108:270–9.
37. Pyrc K, Dijkman R, Deng L, Jebbink MF, Ross HA, Berkhout B, et al. Mosaic structure of human coronavirus NL63, one thousand years of evolution. J Mol Biol. 2006 Dec 15;364(5):964–73.
38. Pfefferle S, Oppong S, Drexler JF, Gloza-Rausch F, Ipsen A, Seebens A, et al. Distant relatives of Severe Acute Respiratory Syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. Emerg Infect Dis. 2009;15:1377–84.
39. Corman VM, Baldwin HJ, Fumie Tateno A, Zerbinati RM, Annan A, Owusu M, et al. Evidence for an ancestral association of human coronavirus 229E with bats. J Virol. 2015;89(23):11858–70.
40. Joffrin L, Dietrich M, Mavingui P, Lebarbenchon C. Bat pathogens hit the road: But which one? PLoS One. 2018;14(8):e1007134.
41. El-Duah P, Meyer B, Sylverken A, Owusu M, Gottula LT, Yeboah R, et al. Development of a whole-virus ELISA for serological evaluation of domestic livestock as possible hosts of human coronavirus NL63. Viruses. 2019;11(43):v11010043.
42. Huynh J, Li S, Yount B, Smith A, Sturges L, Olsen JC, et al. Evidence supporting a zoonotic origin of human coronavirus strain NL63. J Virol. 2012 Dec 1;86(23):12816–25.
43. Cui J, Li F, Shi Z-L. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019;17(March 2019):181–92.
44. Geldenhuys M, Mortlock M, Weyer J, Bezuidt O, Seamark ECJ, Kearney T, et al. A metagenomic viral discovery approach identifies potential zoonotic and novel mammalian viruses in *Neoromicia* bats within South Africa. PLoS One. 2018;13(3):e0194527.
45. Chen L, Liu B, Yang J, Jin Q. DBatVir: the database of bat-associated viruses. Database (Oxford). 2014;2014:bau021.
46. Hobbs JJ. People and caves in Madagascar. Focus Geogr. 2010;46(3):1–7.
47. Fernández-Llamazares Á, López-Baucells A, Rocha R, Andriamitandrina SM, Andriatafika ZE, Burgas D, et al. Are sacred caves still safe havens for the endemic bats of Madagascar? Oryx. 2019;25(2):271–5.
48. Wilkinson DA, Marshall JC, French NP, Hayman DTS. Habitat fragmentation, biodiversity loss and the risk of novel infectious disease emergence. J R Soc Interface. 2018;15(149):20180403.
49. Meyer B, Drosten C, Müller MA. Serological assays for emerging coronaviruses: Challenges and pitfalls. Virus Res. 2014 Dec 19;194:175–83.

50. Wang L-F, Anderson DE. Viruses in bats and potential spillover to animals and humans. Curr Opin Virol. 2019;34:79–89.

Data Accessibility:

- DNA sequences: Genbank accessions MN183146 to MN183273

Author Contributions: CL and LJ conceived and designed the study. BR, CL, DAW, EL, LJ, PT, SMG and YG collected biological material on Madagascar. SMG, RS and YG collected biological material on Mauritius. BR, EL and GLM, collected biological material on Mayotte. SMG, GLM, ADS and MCS, collected biological material in Mozambique. BR, DAW, EL and GLM collected biological material on Reunion Island. EL, GLM, SJ and YG collected biological material in the Seychelles. LJ performed the molecular analyses. CL and DAW analyzed the data. CL and LJ wrote the paper. SMG, ESG, MP, PM and SJ contributed to the project management in Malagasy, Mauritian, Seychelles, Mozambican, and French institutions. All authors edited, read, and approved the final manuscript.

Figures

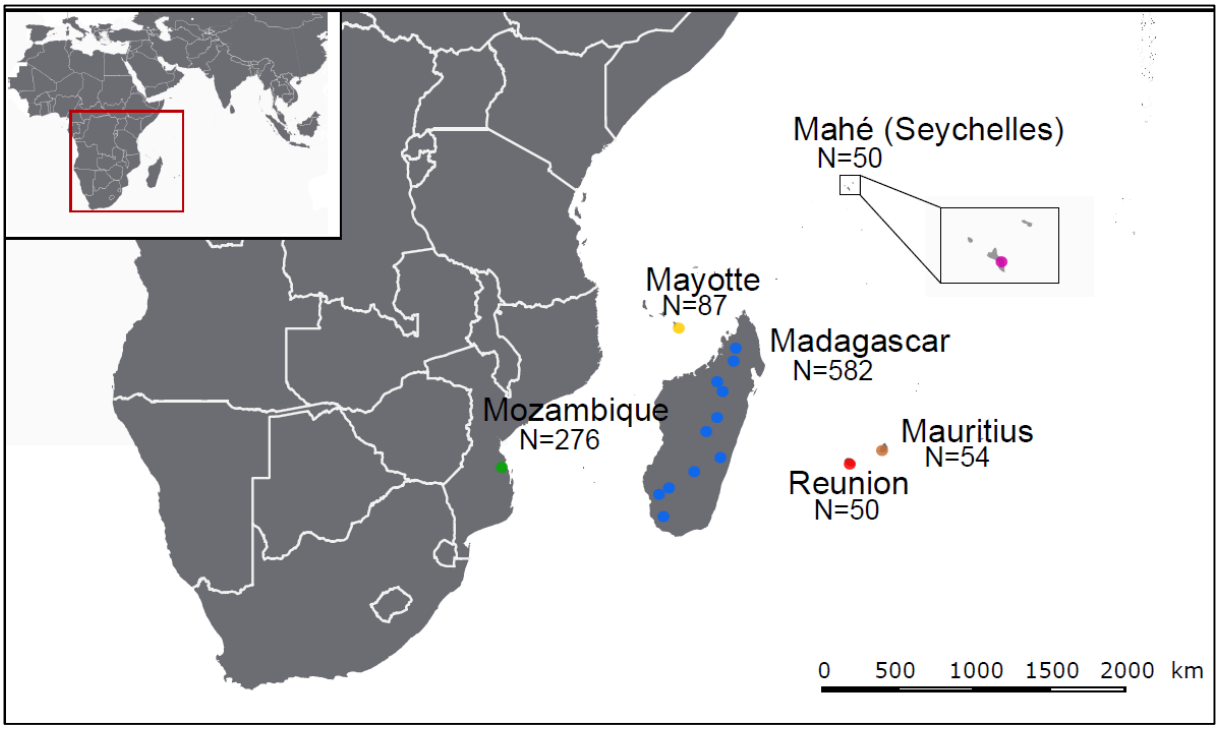


Figure 1. Geographic distribution of the tested samples. N: number of bats sampled for each locality.

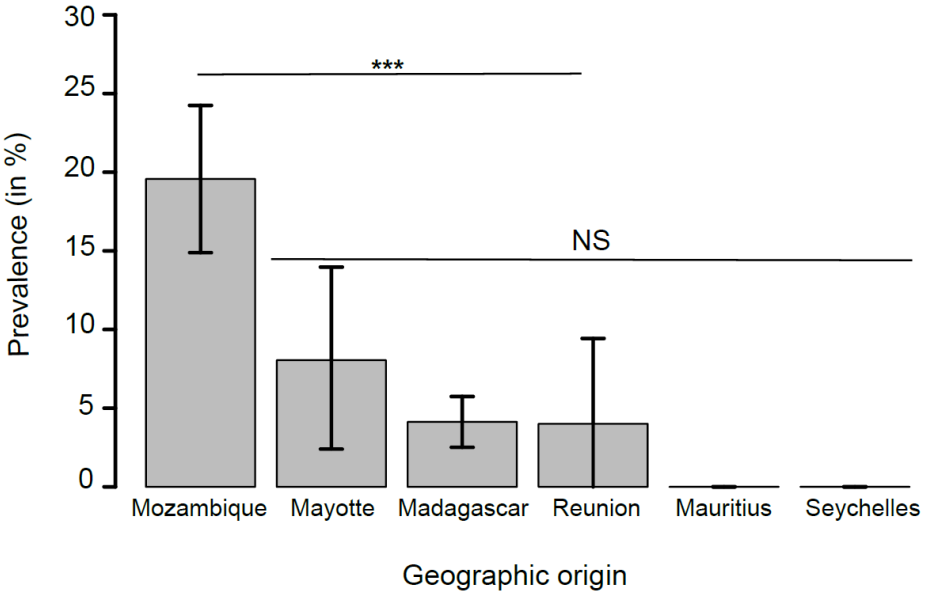


Figure 2. Mean CoV prevalence (± 95% confidence interval) in bats in the western Indian Ocean. Pairwise test; ***: $p<0.001$; NS: $p>0.05$, not significant.

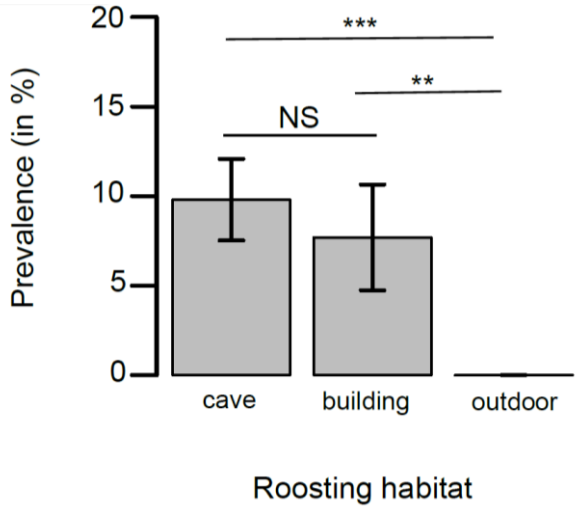
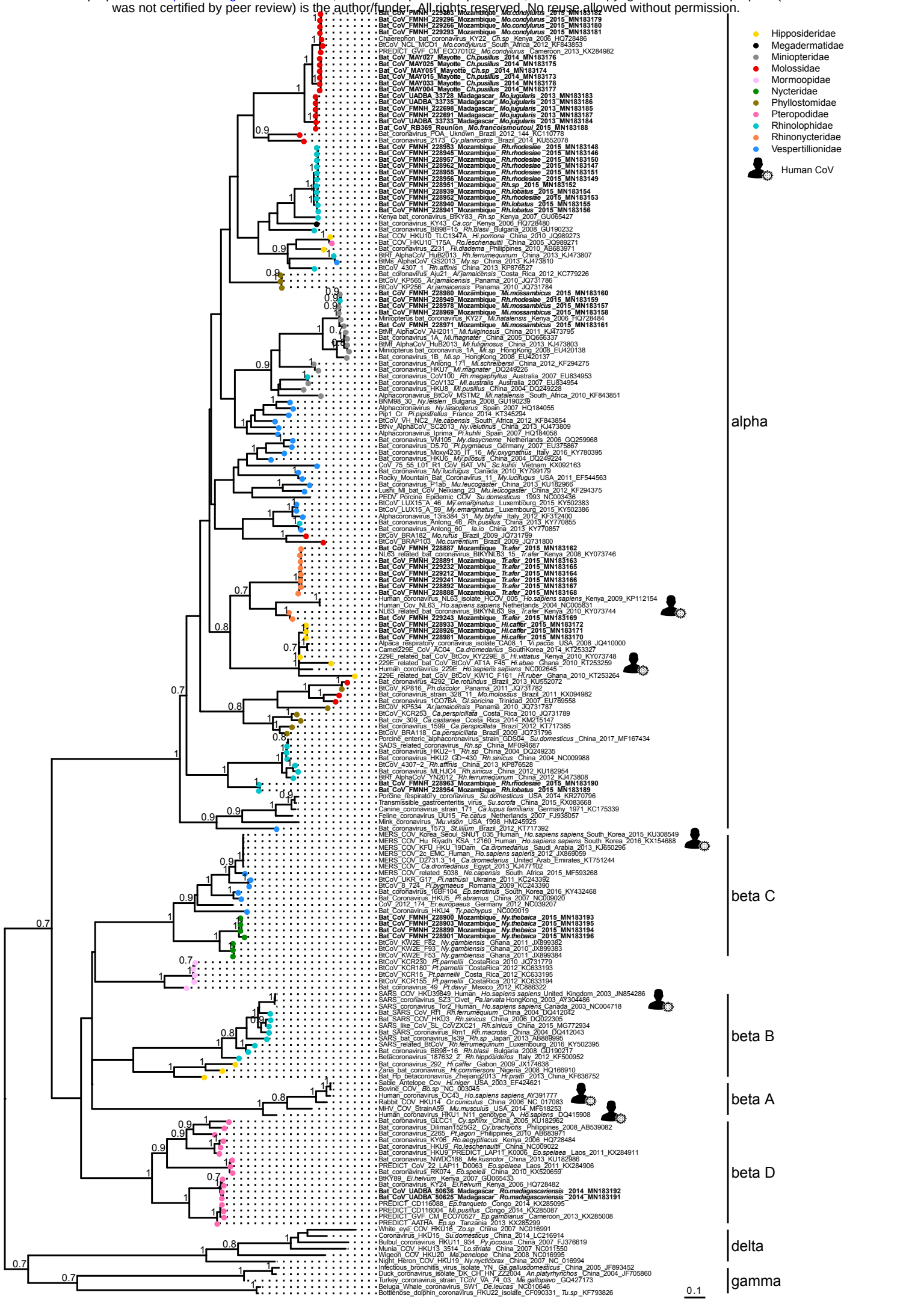
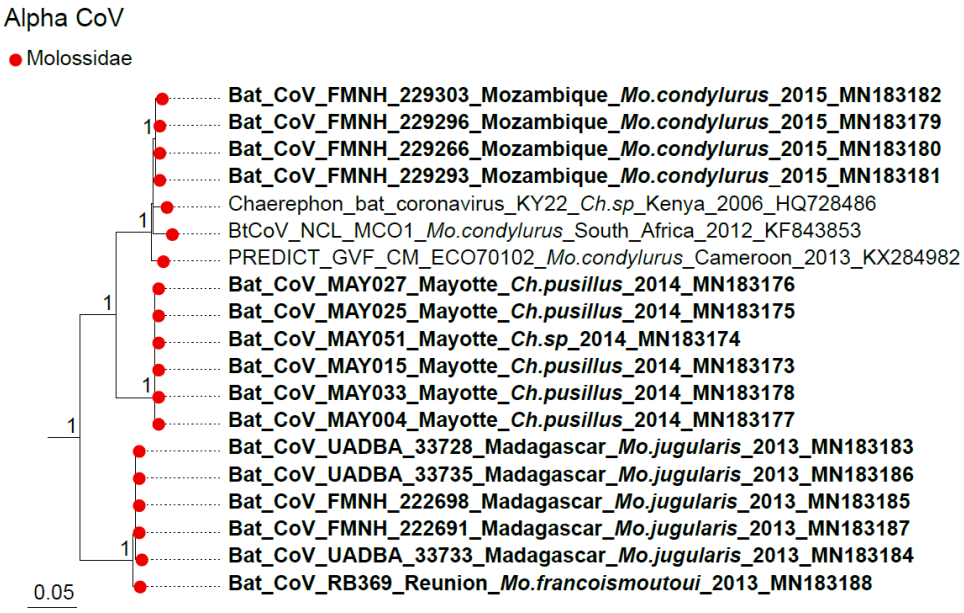


Figure 3. Mean CoV prevalence (± 95% confidence interval) as function of the type of roosting habitat. Pairwise test; ***: p<0.001; **: p<0.01; NS: p>0.05, not significant.

Figure 4. Maximum Likelihood (ML) consensus tree derived from 202 coronavirus (CoV) RNA-dependent RNA-polymerase partial nucleotide sequences (393 bp). Colored circles at the end of branches indicate bat family origin. Sequences in bold refer to bat CoVs detected in this study. Only bootstrap values >0.7 are indicated on the tree. Scale bar indicates mean number of nucleotide substitutions per site. The tree was generated with the General Time Reversible evolutionary model (GTR+I+Γ, I = 0.18, α = 0.64) and 1000 bootstrap replicates.





525 **Figure 5.** Detail of the alpha CoV clade. Molossidae CoVs generated in the study are indi-
526 cated in bold. This sub-tree is a zoom on Molossidae CoV clade from the tree depicted in Fig-
527 ure 4. Only bootstrap values >0.7 are indicated on the tree. Scale bar indicates mean number
528 of nucleotide substitutions per site.

529

530

531 **Figure 6.** Detail of the alpha CoV clade. NL63-like CoVs generated in the study are indicated
532 in bold. This sub-tree is a zoom on NL63 CoV clade from the tree depicted in Figure 4. Only
533 bootstrap values >0.7 are indicated on the tree. Scale bar indicates mean number of nucleotide
534 substitutions per site.

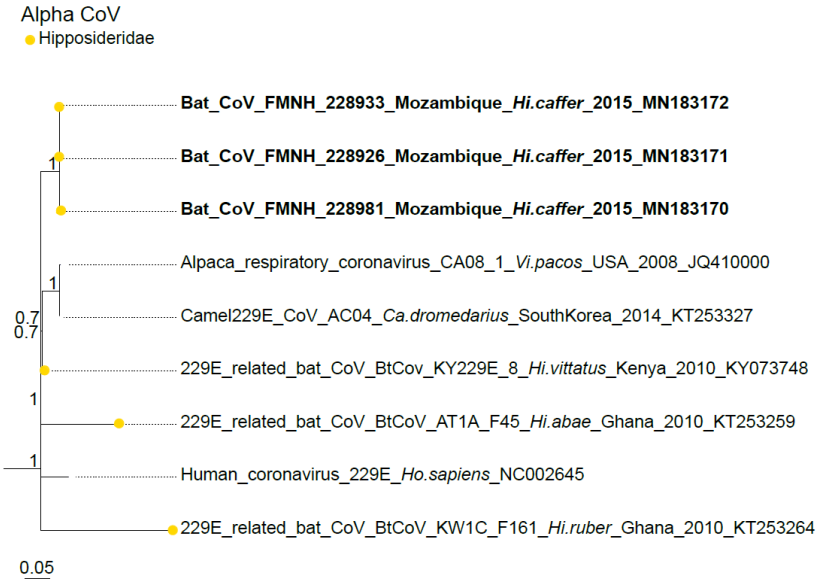
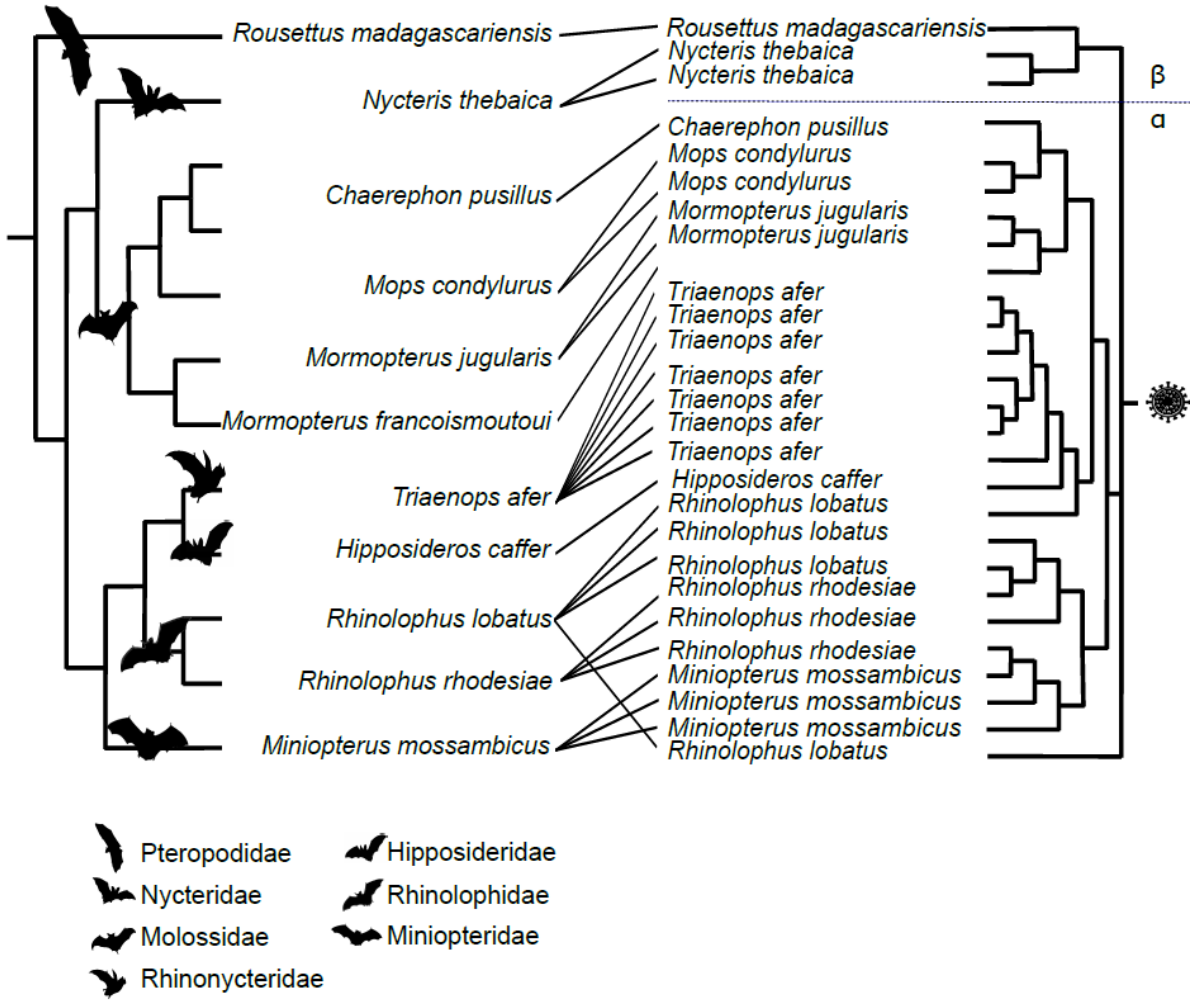


Figure 7. Detail of the alpha CoV clade. 229E-like CoVs generated in the study are indicated in bold. This sub-tree is a zoom on NL63 CoV clade from the tree depicted in Figure 4. Only bootstrap values >0.7 are indicated on the tree. Scale bar indicates mean number of nucleotide substitutions per site.

Figure 8. Detail of the beta-C CoV clade. CoVs generated in the study are indicated in bold. This sub-tree is a zoom on beta-C CoV clade from the tree depicted in Figure 4. Only bootstrap values >0.7 are indicated on the tree. Scale bar indicates mean number of nucleotide substitutions per site.

Figure 9. Detail of the beta-D CoV. CoVs generated in the study are indicated in bold. This sub-tree is a zoom on beta-D CoV clade from the tree depicted in Figure 4. Only bootstrap values >0.7 are indicated on the tree. Scale bar indicates mean number of nucleotide substitutions per site.



551

552 **Figure 10.** Tanglegram representing host-virus co-evolution between bats of the western In-
553 dian Ocean and their associated CoVs. Phylogeny of bats (left) was constructed with an align-
554 ment of 11 Cyt *b* sequences of 1,030 bp by Neighbor-Joining with 1,000 bootstrap iterations.
555 Pruned phylogeny of SWIO bats CoVs (right) was constructed with an alignment of 27
556 unique sequences of 393 bp from SWIO bats CoVs, by Neighbor-Joining with 1,000 boot-
557 strap iterations.