**Supplementary Information**

**SARS-related beta-coronavirus infection linked to gut microbial dysbiosis in bats**

Dominik W. Melville, Magdalena Meyer, Alice Risely, Kerstin Wilhelm, Heather J. Baldwin, Ebenezer K. Badu, Evans Ewald Nkrumah, Samuel Kingsley Oppong, Nina Schwensow, Marco Tschapka, Peter Vallo, Victor M. Corman, Christian Drosten, Simone Sommer

**Methods**

*Sampling design*

Sampling protocols and methods were described in detail elsewhere1. Briefly, for this study bats were captured at five different cave sites in central Ghana, West Africa: Buoyem 1 (N7°72’35.833” W1°98’79.167), Buoyem 2 (N7°72’38.056” W1°99’26.389), Forikrom (N7°58’97.5” W1°87’30.299), Kwamang 1 (N6°58’0.001” W1°16’0.001) and Kwamang 2 (N7°43’24.899” W1°59’16.501) over 12 sampling events spread evenly over the span of two years (September 2010-August 2012). Each sampling event consisted of mist net trapping at the cave entrance one hour after dusk until dawn for two, non-consecutive nights. *Hipposideros caffer* *D* was identified molecularly by sequencing the *cytb* from 2mm wing punches collected at sampling1–3. The species nomenclature of the *Hipposideros caffer* complex remains unresolved throughout the Afrotropics, but here we use *H. caffer* *D*, as described by Vallo et al.3 as an interim species name. Additionally, fecal samples were collected and stored in RNAlater (Life Technologies, USA) at -80 °C for subsequent virus and microbiome screening. The Wildlife Division of the Forestry Commission of the Ministry of Lands, Forestry and Mines granted Research (A04957) and ethics permit (CHRPE49/09/CITES).

*Fecal coronavirus screening and bacterial 16S rRNA gene sequencing*

Approximately 20mg of the fecal material was used to extract and purify viral RNA. Subsequently, real-time reverse transcription-PCR assay designed to detect several alpha- and beta-CoVs1,4–6. A total of four coronaviruses were described: A MERS-related Betacoronavirus termed 2C exclusively found in samples originating from *Nycteris* *macrotis*1,6, the Alphacoronavirus 229E-like as the closest known ancestral form to the HCoV-229E 4, and two SARS-related Betacoronaviruses, named 2B and 2Bbasal1,5. Our focal species *H. caffer* *D* hosts both SARS-related CoVs and the 229E-like CoV1. Notably, *H. caffer* *D* is thought to be particularly susceptible to infections with the SARS-related CoV-2B1, which could hint at the species’ role as main reservoir to this virus. After viral screening, we recorded the infection status (as category, i.e., positive for CoV-229E-like, CoV-2B, or CoV-2Bbasal) and infection intensity estimated based on cycle threshold (CT) value for each sample.

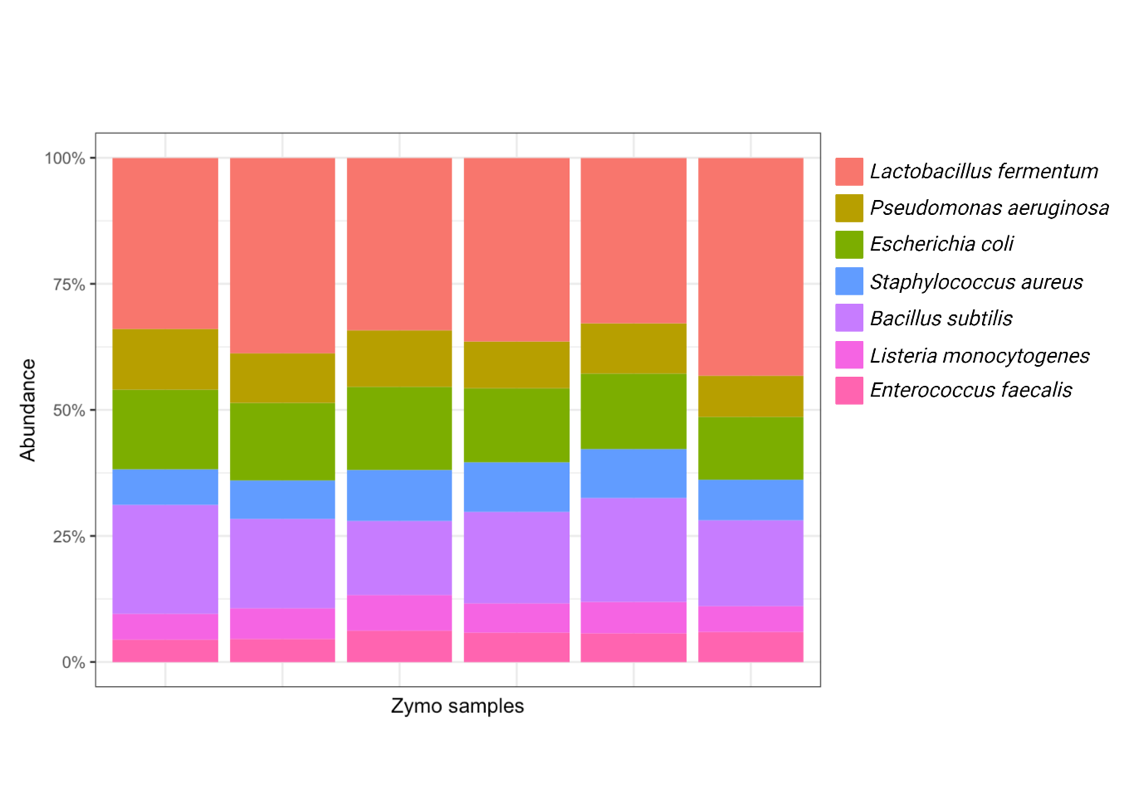
Bacterial DNA was extracted from the remainder of the fecal material for 221 samples as per instructions of the NucleoSpin Soil Kit (Macherey-Nagel, Germany). This includes a bead-beating step to mechanically lyse bacterial cells during two 3-minute pulses with ceramic beads using the SpeedMill PLUS (Analytik Jena, Germany). After centrifugation, the supernatant was transferred to new collection tubes just prior to precipitation. We followed the protocol for the remaining steps. We included ten extraction blanks and six standardised communities (ZymoBIOMICS Microbial Community DNA Standard, Zymo Research, Germany).

We amplified the V4 region of the 16S rRNA gene using the 515F-806R primer pair (Fwd: 5’-GTGCCAGCMGCCGCGGTAA-3’; Rvs: 5’- GGACTACHVGGGTWTCTAAT-3’) 7 and added Illumina adaptor sequences using the Fluidigm Access Array for Illumina Sequencing (Access Array System for Illumina Sequencing Systems, © Standard Bio Tools, USA). Post purification (NucleoMag NGS clean-up and size select, Macherey-Nagel, Germany) and quantification (QuantiFluor dsDNA System, Promega, USA) the normalized pooled sample library was sequenced as paired-end run on the Illumina MiSeq platform at the Institute for Human Genetics, University Hospital of Bonn. A total of nine PCR controls were included.

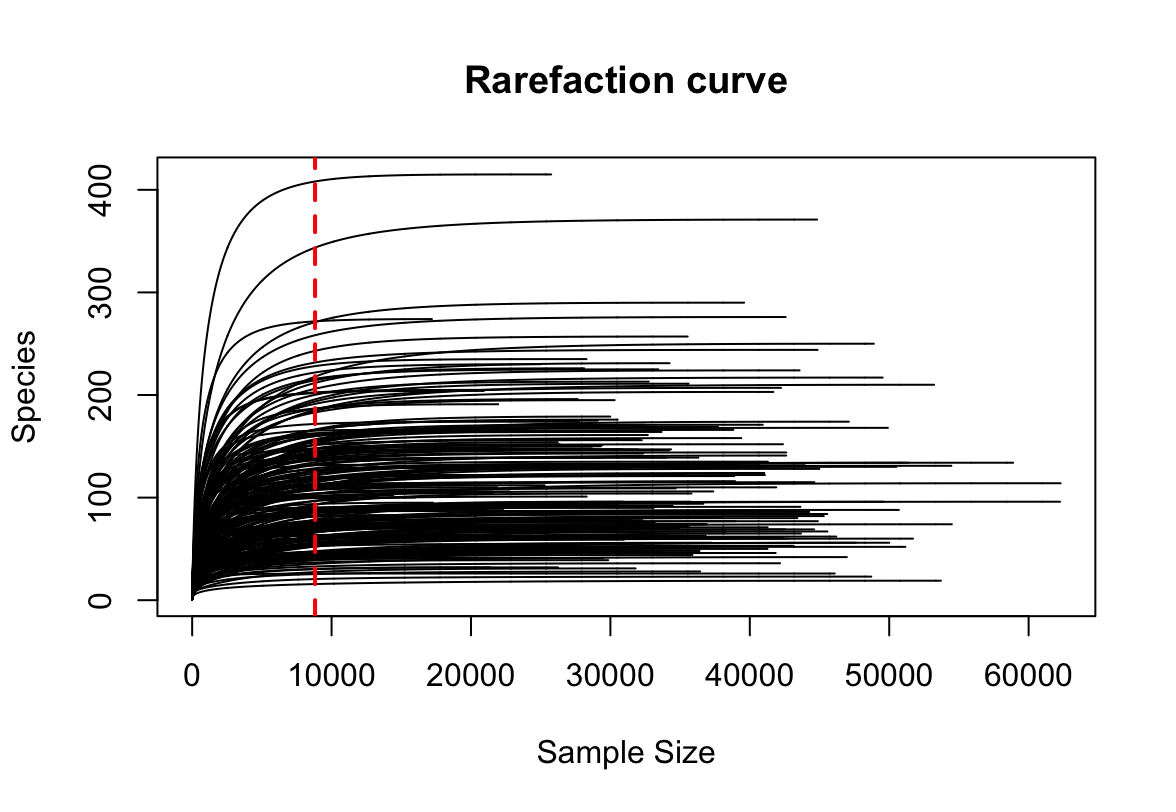
*Bioinformatic pipeline*

The reads were processed with the DADA2 plug-in in QIIME 2 (v2021.8.0)8, removing primers, denoising reads, detecting and removing chimeras, merging paired-end reads, and differentiating between single amino acid sequence variants (ASVs)9. ASVs were then assigned their taxonomy using SILVA (v138) as reference database10. ASVs unassigned at the phylum level or identified to originate from chloroplast or mitochondrial sequences were excluded from subsequent analyses. An unrooted, phylogenetic tree was build using Mafft11 and FastTree12. The tree was rooted in Dendroscope13 using an archaeon sequence (accession number: KU656649) as the outgroup, which was later removed. Metadata including each sample’s sex, age, location, infection status, the CT-value of the respective infection and sampling period, the taxonomy and ASV counts, and the phylogenetic tree were imported into the Rstudio interface of R (v4.3.2)14 using the ‘phyloseq’ package (v1.46.0)15. All sample processing and statistical analyses were henceforth performed in Rstudio.

We first confirmed the community composition of the six standardised microbial community references, which showed no large deviation in amplification across the extraction runs from the expected community composition (Supplementary Figure 1). Next, we pruned unassigned ASVs at the phylum level, ASVs with fewer than 10 reads and phyla with less than 0.3% read coverage across all samples leaving us with 4,852 taxa from an original of 10,975. Employing the prevalence-based contamination identification functions from the ‘decontam’ R package (v1.16) with the default P\*-threshold set to 0.116, we identified a possible 163 taxa (3.3%) from the 10 extraction blanks and 165 taxa (3.4%) from the 9 PCR blanks. These taxa were removed. Finally, the decontaminated phyloseq object contained 4,531 taxa. We plotted alpha diversity rarefaction curves for each sample with the rarecurve function from the ‘vegan’ R package (v2.6-4)17. Based on the curves (Supplementary Figure 2), two fecal samples below a sequencing depth of 5000 reads were eliminated from downstream analyses (including all controls) and one samples missing age information. Hence, a total of 218 high-quality fecal samples with on average 32,849 reads (range: 8,815-62,295) remained.



**Supplementary Figure 1. A total of six ZymoBIOMICS microbial community standards were included to assess consistency in extraction and sequencing protocols.**



**Supplementary Figure 2 Rarefaction curve for all samples.** Red dashed line indicates read count of the sample with the lowest number of reads (i.e., 8815) and, hence, forms the rarefying threshold.

*Statistical analysis*

Alpha diversity

Alpha diversity metrics (i.e., Observed ASVs, Chao1, and Faith’s Phylogenetic diversity) were calculated employing the ‘phyloseq’ (v1.46.0)15 and ‘picante’ package (v1.8.2)18 for samples rarefied to the sequencing depth of the sample with the lowest reads (i.e., 8815 reads, Supplementary Figure 2)19,20. While observed ASVs measures actual bacterial ASV richness, Chao1 accounts for possible undetected ASVs, and Faith’s Phylogenetic diversity considers abundance and phylogenetic proximity between ASVs. The effect of sampling locations, host sex, host age and CoV infection status on each metric was then estimated using a linear mixed effect model with capture period as random effect. Faith’s Phylogenetic Diversity was square-root, and Chao1 and Observed ASVs log-transformed to meet normality assumption. Subsequently, we assessed whether each alpha-diversity index correlated with infection intensity (measured as CT-value).

Beta diversity

To assess inter-sample differences, we first calculated unweighted and weighted Unifrac-distances as microbial beta-diversity indices on the rarefied data agglomerated to genus level. Both distances take phylogenetic distance between ASVs into account, but whereas weighted Unifrac considers reads as proxy for ASV abundance, and, thus, represents the structure of a microbial community, unweighted Unifrac treats ASVs as either absent or present and, hence, epitomises the composition of a microbial community. We used each index separately as response variable in a PERMANOVA to determine changes in centroid position associated with sampling location, host sex, host age and CoV infection status and PERMUtest to assess differences in distance to centroid. Following from this, we extracted the unweighted and weighted distances as matrix from all infected samples and calculated the average distance to other samples in the same infection category. The average distance, therefore, denotes how dissimilar a particular sample is from others. Finally, we correlated the average distance between samples with infection intensity (measured as CT-value).

Joint species distribution modelling to detect differentially abundant bacterial genera

In order to understand which bacterial genera differed between infection status, we used joint-species distribution modelling. Compared with traditional abundance-based analyses (e.g., ANCOM), joint species distribution models incorporate correlations between bacterial taxa when predicting their abundance with respect to the explanatory variable21. Since taxonomic databases remain biased towards identifying microbial taxa from common model organisms22, the taxonomic resolution for wildlife gut microbiota at the species level is lacking. As to avoid spurious results, we agglomerated reads at the genus level and restricted our analysis to genera with a prevalence >50% (i.e., common core)23. The abundance was centred log transformed (clr)24. Finally, we constructed a generalised linear latent variable model (GLLVM) using the ‘gllvm’ package21. Another advantages of a joint model approach over differential abundance analysis is that we can use the same model structure as for testing alpha- and beta-diversity to evaluate multivariate microbial abundance data, meaning that infection status, sampling location, host age and sex were kept as main explanatory variables and capture period was used as random effect. Additionally, we accounted for sequencing depth and specified a negative binomial distribution for our response variable.

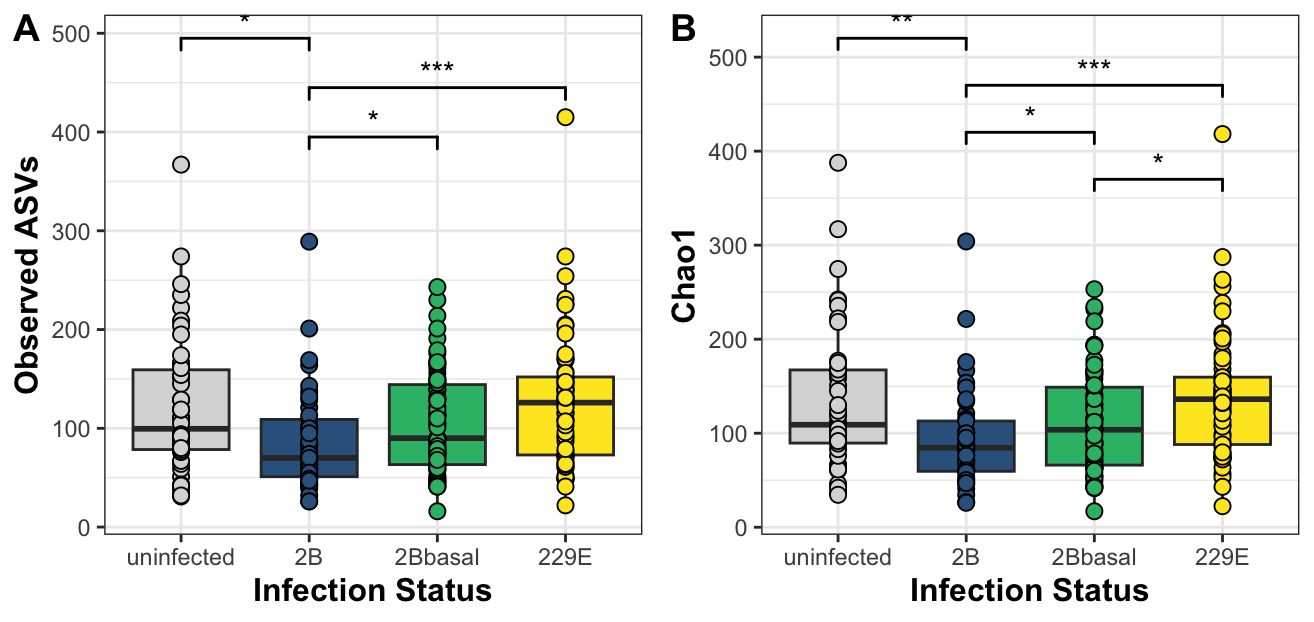
Linking bacterial abundances and infection intensity

For those genera that were identified to occur more or less frequent in samples from bats infected with CoV-2B, we wanted to ascertain whether their abundance was linked to infection intensity when compared to uninfected bats. In other words, whether there is a linear or non-linear relationship with CT-value over the course of an infection. We constructed generalised additive models with the clr-transformed abundances (i.e., reads) as response and sampling location, sex and age as categorical explanatory variables. The CoV-2B CT-value was included as smooth term. Sequencing depth was included as smooth term as way to control for differences in sequencing performance between samples. Model fit was assessed using gam.check() function of the ‘mgcv’ package25 and we visualized the model results using plot\_smooths() from the ‘tidymv’ package26.

**Extended Results & Discussion**

Alpha diversity

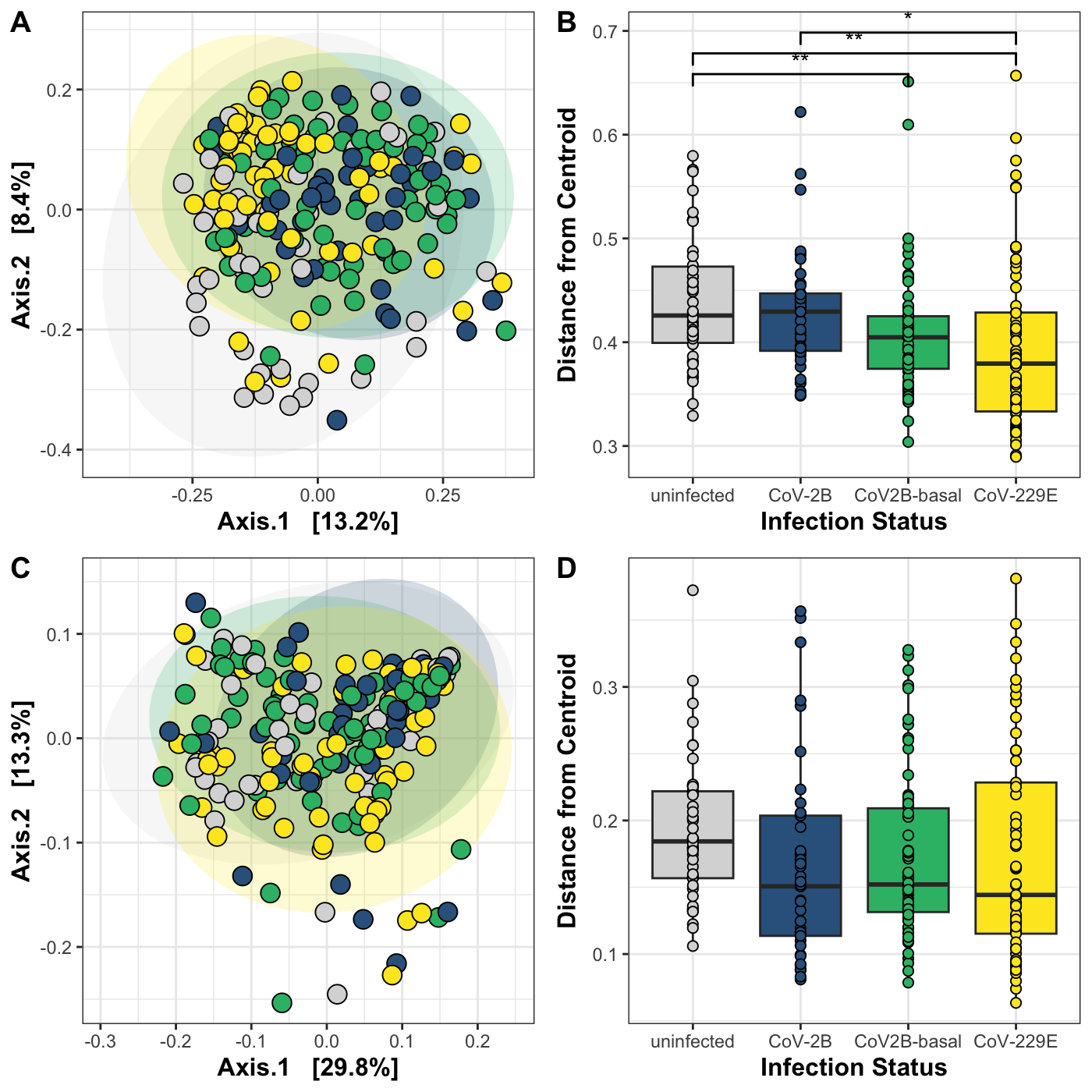
Mirroring the results from Faith’s phylogenetic diversity, infection status was as the single best explanatory variable for Chao1 and Observed ASVs (Supplementary table 1). The post-hoc test reiterated that only infections with the SARS-related CoV-2B lowered both Chao1 and Observed ASVs compared with the microbial diversity found in uninfected, CoV-2Bbasal or CoV-229E infected bats (Supplementary Fig 3AB, Supplementary Table 1). By comparison, in Jamaican fruit bats (*Atribeus jamaicensis*) the Observed ASVs and Chao1 was higher in subadults infected with Astroviridae27. Although we are not aware of another comparison of the effect of a viral infection on the chiropteran gut microbiome, viral infection frequently alter gut microbial alpha-diversity in other host species (lower: influenza A virus infected mallards (*Anas platyrhyncos*)28; higher: e.g., *Puumala orthohantavirus* in bank voles (*Myodes glareolus*)29), but, in some cases, viral infection had no effect (e.g., Adenovirus infection in mouse lemurs (*Microcebus griseorufus*)30). Importantly, the gut microbial response to infections is not unilateral and might depend on host genetics. A follow-up study on the Astrovirus-infected Jamaican fruit bats discovered that alpha-diversity in uninfected and infected bats was shaped by the identity and diversity of alleles from the major histocompatibility complex class II DRB region31. Collectively, these findings suggest that a three-way interaction between host, pathogen and microbiota determines disease outcomes.



**Supplementary Figure 3. Differences in gut microbial alpha-diversity measured as A) Observed ASVs and B) Chao1 between infection status.** Asterix indicate level of significance: \* < 0.05, \*\* < 0.01, \*\*\* <0.001.

Beta diversity

Inter-sample dissimilarity measured as unweighted (i.e., microbial composition) and weighted (i.e., microbial structure) Unifrac distances was also chiefly affected by infection status (Supplementary Fig 4AC; Supplementary Table 2). In Jamaican fruit bats, only the gut microbial structure in bats with Astrovirus infections was shifted27. When comparing unweighted Unifrac distances it was also noted that samples from bats infected with CoV-229E were on average less dispersed than uninfected and CoV-2B infected bats, and CoV-2Bbasal infected bats were also less dispersed than uninfected bats (Supplementary Fig 4B). Dissimilarity measured as weighted Unifrac distances did not differ in dispersion (Supplementary Fig 4D). Changes in centroids and dispersion of the gut microbial community is not an uncommon observation in studies investigating the impact of infections in wildlife more generally. For example, Adenovirus-infected mouse lemurs (*M. griseorufus*) had a less dispersed (i.e., more similar) gut microbial composition32, whereas it was more widely dispersed (i.e., more dissimilar) in chimpanzees (*Pan troglodytes*) infected with SIV33.



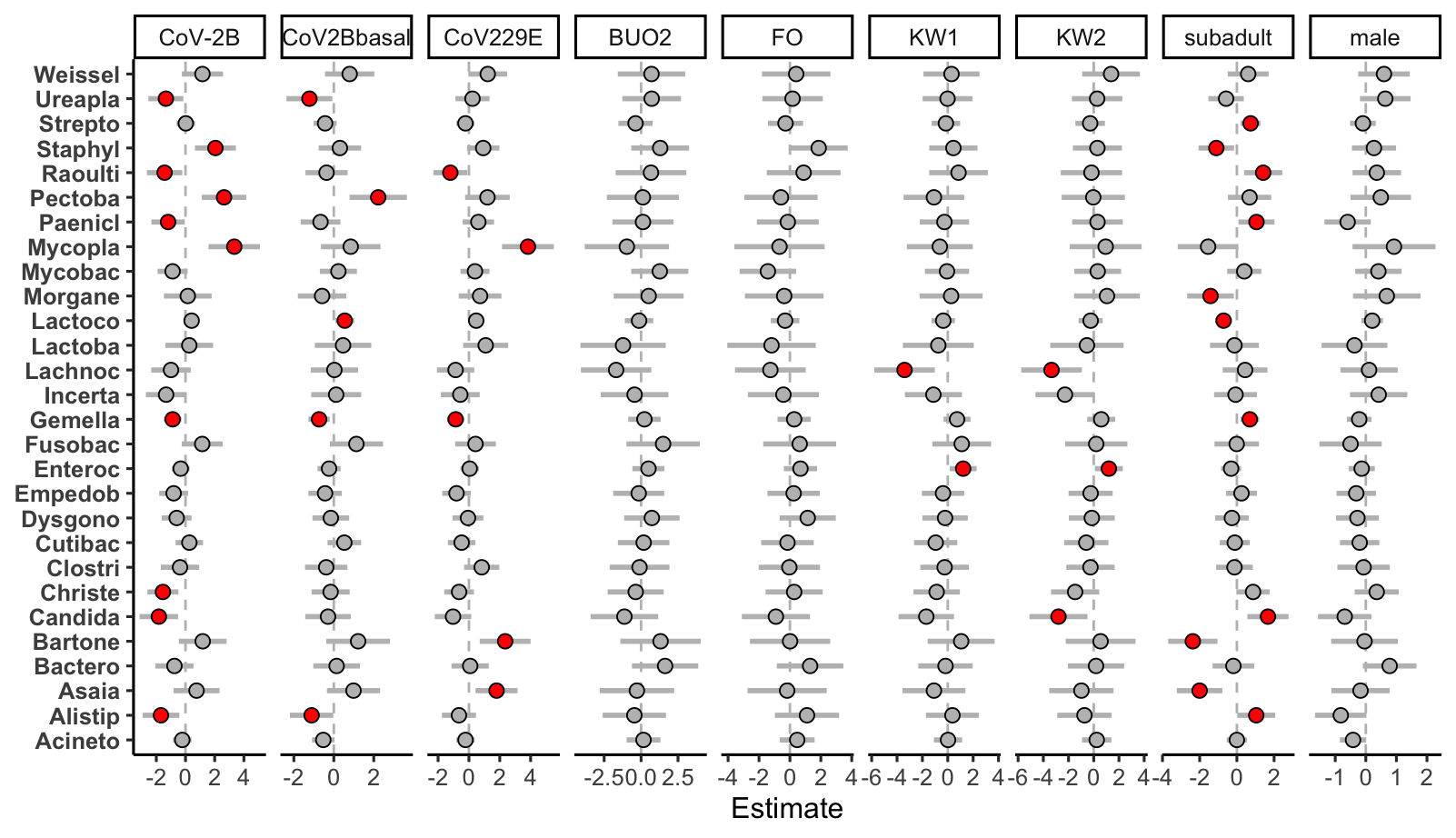
**Supplementary Figure 4 A-B) Unweighted and C-D) weighted Unifrac distances depicted in a PCoA and as distances from centroid comparing infection statuses and cave site.** Ellipses are drawn using 95% confidence intervals. Asterix indicate level of significance: \* < 0.05, \*\* < 0.01, \*\*\* <0.001.

Although host sex and age had no detectable impact on the gut microbial structure and composition (Supplementary Table 2), both were associated with shifts in weighted and unweighted Unifrac in Jamaican fruit bats27. Host age effects were also recovered in mouse lemurs30, hyenas (*Crutta crutta*)34, meerkats (*Suricata suricatta*)35 and Assamese macaques (*Macaca assamensis*)36 and an almost universal feature believed to be connection to shifts in diet and maturation of immunity during mammal ontogeny. The lack of an age effect might be explained by the fact that our samples were gathered from volant animals, which likely have already undergone a switch to an adult and, hence, insect-dominated diet, while still having an unfused epiphysis. In addition, we found some differences between habitats and thereby an effect of environmental variability on the gut microbial community composition in *H. caffer* *D* (Supplementary Fig 4E-F; Supplementary Table 2). Colonialization of the enteric tract by environmentally acquired bacteria and the expansion of bacteria profiting from local food resources are reasons for differences between natural habitats (e.g., Lesser long-nosed bat (*Leptonycteris yerbabuenae*)37; Rocky Mountain Elk (*Cervus canadensis*)38 but also in comparison with human-dominated environments (e.g., Tiger (*Panthera tigris jacksoni*)39; rodents and marsupials40; black-and-white ruffed lemur (*Varecia variegata editorum*)41). Yet, this, again, is by no means a universal pattern (e.g., elephants42, primates43).

Taken together, we find that host characteristics, such as sex and age, play a small part in shaping the gut microbial beta-diversity of *H. caffer D* when compared to the environment and pathogens. The seemingly largest impact on gut microbial community composition had infection status. Our work is in line with current thoughts on how host characteristics, pathogen infections and the environment concurrently impact the gut microbial composition (i.e., disease pyramid, 44). It’s important to emphasise though that we cannot fully disentangle whether the connection between virus infection and gut microbial composition is a consequence of the CoV-2B infection or lowered microbial resistance to viral invaders in an already depauperated gut microbial community (e.g., due to nutritional stress of the host or prior infections). Whereas other systems could explore the role of nutritional stress in relation to microbiota-mediated viral resistance45, an interesting follow-up question for our system could be to investigate the role of multiple CoVs co-infected *H. caffer* *D*32, and gather repeated samples from the same bats as they progress through a viral infection.

Joint species distribution modelling to compare differentially abundant genera

Several common bacterial genera seemed to differ between uninfected and CoV-2B infected bats, but also the two other CoVs caused slight changes in abundant genera (Supplementary Table 3, Supplementary Figure 5). *Mycoplasma*, *Staphylococcus*, and *Pectobacterium* were more common in CoV-2B infected bats. *Raoultibacter*, *Paeniclostridium*, *Gemella*, *Christensenellaceae R-7 group*, *Candidatus Soleaferrea*, *Ureaplasma* and *Alistipes*, on the other hand, were found at lower abundances in CoV-2B infected bats. We even uncovered some commonalities among different pathogens: *Mycoplasma* was, for instance, more common in both CoV-2B and CoV-229E infected bats, while *Gemella* was reduced in any infected bat. Moreover, Jamaican fruit bats infected with Astroviridae were also enriched in Mycoplasma and impoverished in *Bacteroidiales* (Family to *Alistipes*)27,31. Other genera, such as *Lactobacillus* or *Streptococcus*, showed differences in Astrovirus-infected A*. jamaicensis* but not in CoV-infected *H. caffer* *D*. Lastly, gut bacterial genera varied in decreasing strength with age and sampling location but not significantly between sexes (Supplementary Table 3, Supplementary Figure 5).



**Supplementary Figure 5. Generalized linear latent variable model indicating which common gut bacterial genera (prevalence > 0.5) are found significantly (dots colored red) more or less common in *H. caffer* *D* with an infection (CoV-2B, CoV-2Bbasal or CoV-229E) or between cave sites, age classes and sexes compared with the reference.** Abbreviations: Weissel=*Weissella*; Ureapla=*Ureaplasma*; Strepto=*Streptococcus*; Staphyl=*Staphylococcus*; Raoulti=*Raoultibacter*; Pectoba= *Pectobacterium*; Paenicl=*Peniclostridium*; Mycopla=*Mycoplasma*; Mycobac=*Mycobacterium*; Morgane=*Morganella*; Lactoco=*Lactococcus*; Lactoba=*Lactobacillus*; Lachnoc=*Lachnoclostridium*; Incerta=*Incertae Sedis*; Fusobac=*Fusobacterium*; Enteroc=*Enterococcus*; Dysgono=*Dysgonomonas*;; Clostri=*Clostridium sensu stricto 1*; Christe=*Christensenellaceae R-7 group*; Candida = *Candidatus Soleaferrea*; Bartone=*Bartonella*; Bactero=*Bacteroides*; Alistip=*Alistipes*; Acineto=*Acinetobacter*.

Generalised additive model results assessing the impact of infection intensity on bacterial abundances

Aside from the positive relationship with *Alistipes* and *Christensenellaceae R-7 group* Candidatus Soleaferrea increased with infection intensity. Pectobacterium decreased non-linearly, which was the exception in these genera. The majority changed a linear manner (Supplementary Table 4; Figure 2). Noteworthy is the negative relationship between the Firmicute *Christensenellaceae R-7* group and infection intensity, which was also unveiled in non-human primates experimentally infected with SARS-CoV-246. Several members of *Christensenellaceae* are known to benefit host health in humans. Our finding raises the possibility of conserved gut microbial responses to similar pathogens across evolutionarily distant host taxa.

**References**

1. Meyer, M. et al. *Nat Commun* **15**, 2887 (2024).

2. Schmid, D.W. et al. *Molecular Ecology* **32**, 3989–4002 (2023).

3. Vallo, P., Guillén-Servent, A., Benda, P., Pires, D.B. & Koubek, P. *acta* **10**, 193–206 (2008).

4. Corman, V.M. et al. *J Virol* **89**, 11858–11870 (2015).

5. Pfefferle, S. et al. *Emerg Infect Dis* **15**, 1377–1384 (2009).

6. Annan, A. et al. *Emerg Infect Dis* **19**, 456–459 (2013).

7. Caporaso, J.G. et al. *Proceedings of the National Academy of Sciences* **108**, 4516–4522 (2011).

8. Bolyen, E. et al. *Nat Biotechnol* **37**, 852–857 (2019).

9. Callahan, B.J. et al. *Nat Methods* **13**, 581–583 (2016).

10. Quast, C. et al. *Nucleic Acids Research* **41**, D590–D596 (2012).

11. Katoh, K. & Standley, D.M. *Molecular Biology and Evolution* **30**, 772–780 (2013).

12. Price, M.N., Dehal, P.S. & Arkin, A.P. *PLoS ONE* **5**, e9490 (2010).

13. Huson, D.H. & Scornavacca, C. *Syst Biol* **61**, 1061–1067 (2012).

14. R Core Team (2022).

15. McMurdie, P.J. & Holmes, S. *PLoS ONE* **8**, e61217 (2013).

16. Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A. & Callahan, B.J. *Microbiome* **6**, 226 (2018).

17. Oksanen, J. et al. (2022).

18. Kembel, S.W. et al. *Bioinformatics* **26**, 1463–1464 (2010).

19. Weiss, S. et al. *Microbiome* **5**, 27 (2017).

20. Schloss, P.D. *mSphere* **9**, e00355-23 (2023).

21. Niku, J., Hui, F.K.C., Taskinen, S. & Warton, D.I. *Methods in Ecology and Evolution* **10**, 2173–2182 (2019).

22. Schloss, P.D., Jenior, M.L., Koumpouras, C.C., Westcott, S.L. & Highlander, S.K. *PeerJ* **4**, e1869 (2016).

23. Risely, A. *Journal of Animal Ecology* **89**, 1549–1558 (2020).

24. Quinn, T.P. et al. *GigaScience* **8**, giz107 (2019).

25. Wood, S. (Chapman and Hall/CRC: 2017).at <https://doi.org/10.1201/9781315370279>

26. Coretta, S., van Rij, J. & Wieling, M.

27. Wasimuddin et al. *ISME J* **12**, 2883–2893 (2018).

28. Ganz, H.H. et al. *mSystems* **2**, 10.1128/msystems.00188-16 (2017).

29. Brila, I. et al. *Journal of Animal Ecology* **92**, 826–837 (2023).

30. Wasimuddin et al. *Sci Rep* **9**, 13410 (2019).

31. Fleischer, R. et al. *Molecular Ecology* **31**, 3342–3359 (2022).

32. Schmid, D.W. et al. *anim microbiome* **4**, 48 (2022).

33. Moeller, A.H. et al. *Cell Host & Microbe* **14**, 340–345 (2013).

34. Rojas, C.A. et al. *mSystems* **8**, e00965-22 (2023).

35. Risely, A. et al. *Proceedings of the Royal Society B: Biological Sciences* **289**, 20220609 (2022).

36. Sadoughi, B., Schneider, D., Daniel, R., Schülke, O. & Ostner, J. *Microbiome* **10**, 95 (2022).

37. Víquez-R, L. et al. *Microbiol Spectr* **9**, e01525-21 (2021).

38. Pannoni, S.B., Proffitt, K.M. & Holben, W.E. *Ecology and Evolution* **12**, e8564 (2022).

39. Gani, M. et al. *World J Microbiol Biotechnol* **40**, 111 (2024).

40. Heni, A.C. et al. *Animal Microbiome* **5**, 22 (2023).

41. McManus, N. et al. *BMC Ecol Evo* **21**, 222 (2021).

42. Budd, K. et al. *Ecology and Evolution* **10**, 5637–5650 (2020).

43. McCord, A.I. et al. *American J Primatol* **76**, 347–354 (2014).

44. Bernardo-Cravo, A.P., Schmeller, D.S., Chatzinotas, A., Vredenburg, V.T. & Loyau, A. *Trends in Parasitology* **36**, 616–633 (2020).

45. Eby, P. et al. *Nature* **613**, 340–344 (2023).

46. Sokol, H. et al. *Gut Microbes* **13**, 1893113