PNAS
www.pnas.org

Supplementary Information for
Primate phageomes are structured by superhost phylogeny and environment

Jan F. Gogarten[1,2*], Malte Rühlemann[3], Elizabeth Archie[4], Jenny Tung[5,6,7], Chantal Akoua-Koffi[8], Corinna Bang[3], Tobias Deschner[9], Jean-Jacques Muyembe-Tamfun[10], Martha M. Robbins[9], Grit Schubert[1], Martin Surbeck[9,11], Roman M. Wittig[9,12], Klaus Zuberbühler[13], John F. Baines[14,15], Andre Franke[3], Fabian H. Leendertz[1], Sébastien Calvignac-Spencer[1,2*]

[1]Epidemiology of Highly Pathogenic Organisms, Robert Koch Institute, Berlin, Germany.

[2]Viral Evolution, Robert Koch Institute Berlin, Berlin, Germany.

[3]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany.

[4]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, USA.

[5]Department of Biology, Duke University, Durham, NC, USA.

[6]Duke University Population Research Institute, Duke University, Durham, NC, USA.

[7]Department of Evolutionary Anthropology, Duke University, Durham, NC, USA.

[8]Université Alassane Ouattara de Bouake, Bouaké, Côte d'Ivoire.

[9]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

[10]National Institute for Biomedical Research, National Laboratory of Public Health, Kinshasa, Democratic Republic of the Congo.

[11]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA.

[12]Tai Chimpanzee Project, CSRS, BP 1301, Abidjan 01, Cote d'Ivoire.

[13]Institute of Biology, University of Neuchatel, Rue Emile Argand 11, CH-2000 Neuchatel, Switzerland

[14]Max Planck Institute for Evolutionary Biology, Plön, Germany.

[15]Institute for Experimental Medicine, Christian-Albrechts-University of Kiel, Kiel, Germany.

*Correspondence: Jan F. Gogarten and Sébastien Calvignac-Spencer.

Email: jan.gogarten@gmail.com, CalvignacS@rki.de

This PDF file includes:

Figures S1 to S8
Legends for Datasets S1 to S16
Legends for Zenodo hosted Datasets 1 to 9
Supplementary references

Other supplementary materials for this manuscript include the following:

Datasets S1 to S16
Zenodo hosted Datasets 1 to 9

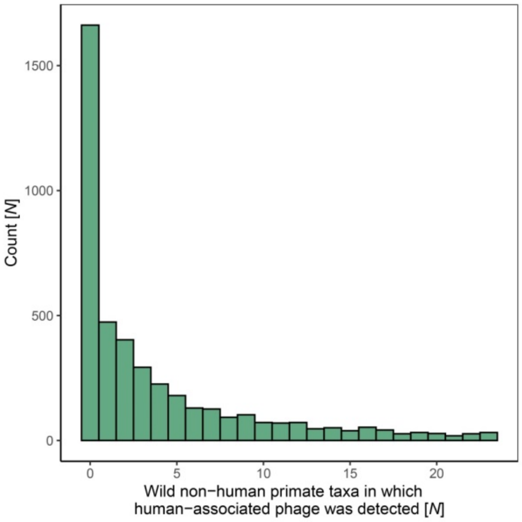

Fig. S1. Histogram of the number of wild non-human primate taxa in which each of the 4,301 HHAP are detected.

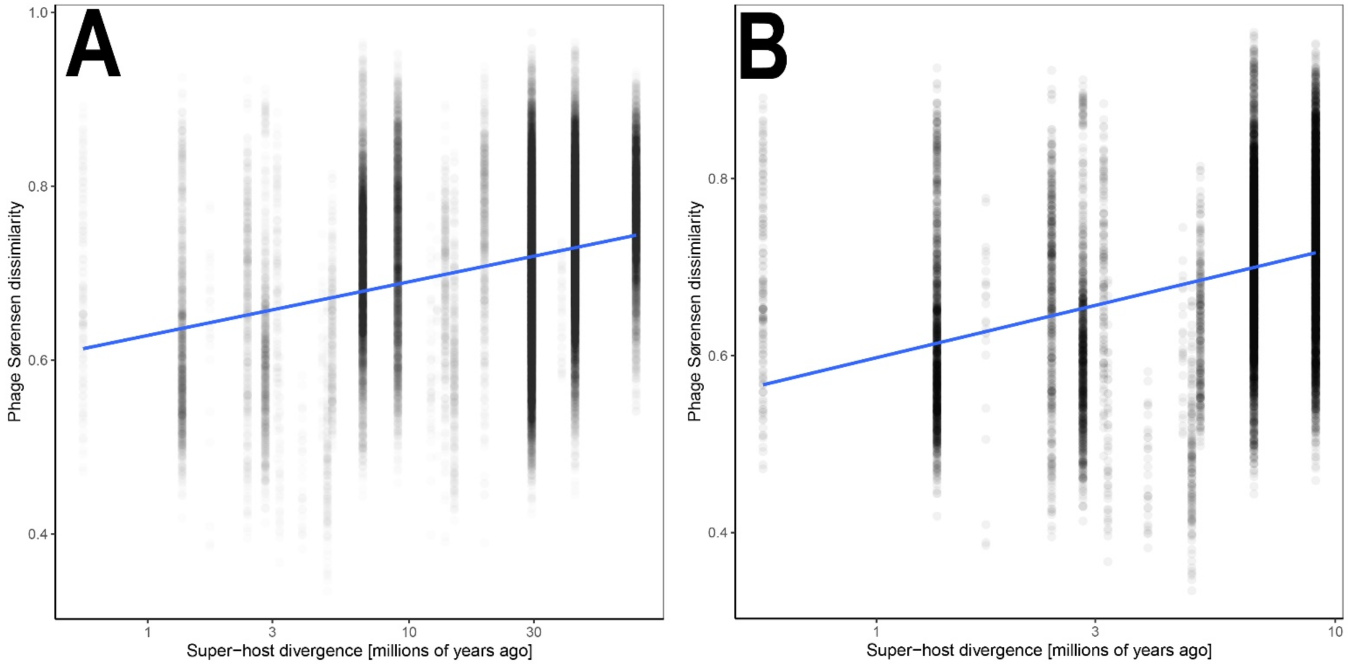

Fig. S2. Phage community dissimilarity for samples from different superhosts plotted against the divergence time between the superhosts. In A, comparison of all pairs of superhosts are shown. In B, only comparisons between superhosts that diverged less than 10 million years ago are shown. The solid blue line represents the fit of a linear model to aid in interpretation of the relationship.

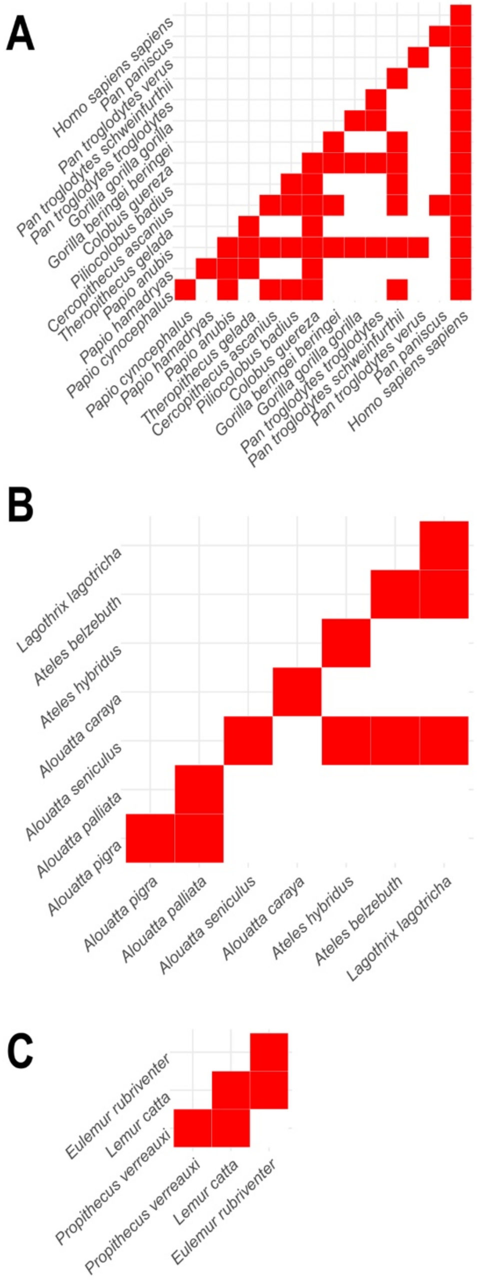

Fig. S3. Matrices of wild and human A) catarrhine, B) atelid, and C) lemur taxa, with red squares in the lower triangle indicating pairs of taxa that overlap geographically somewhere in their home ranges (data from IUCN red list; when taxon was not available, we used species distributions compiled through the All the Worlds Primates database (55).

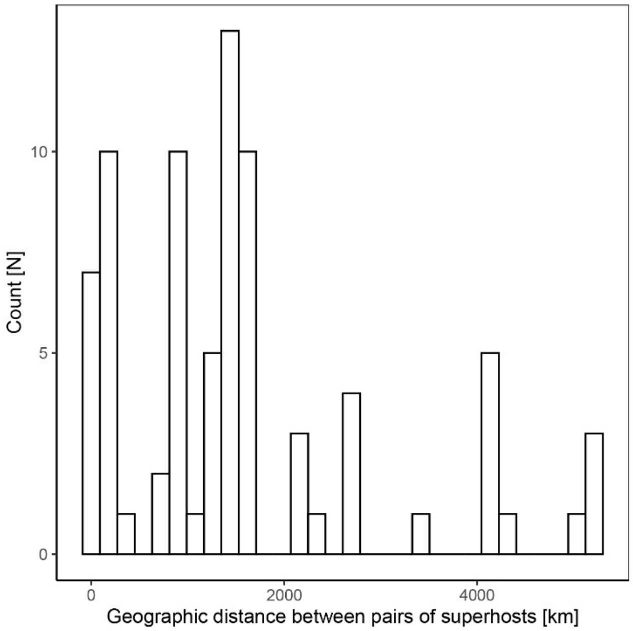

Fig. S4. Histogram of the geographic distance between wild non-human catarrhine primate taxa in which each of the 4,301 HHAP are detected.

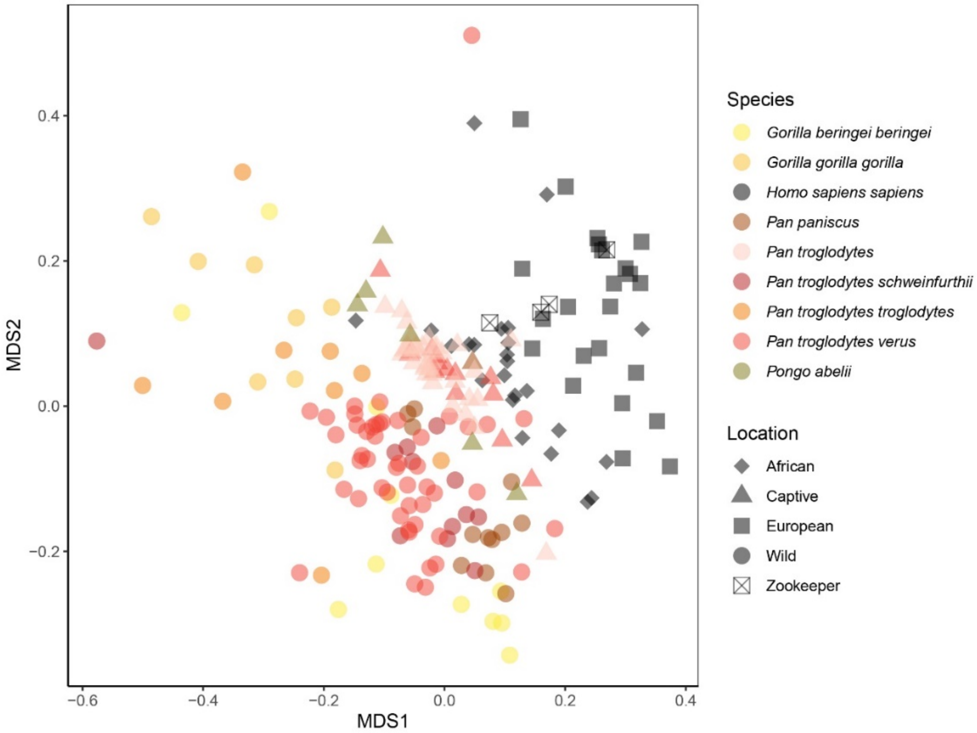

Fig. S5. An ordination of great ape phage community composition (NMDS, Sørensen's dissimilarity, stress=0.195) colored by the superhost species, with the superhost location indicated by the shape.

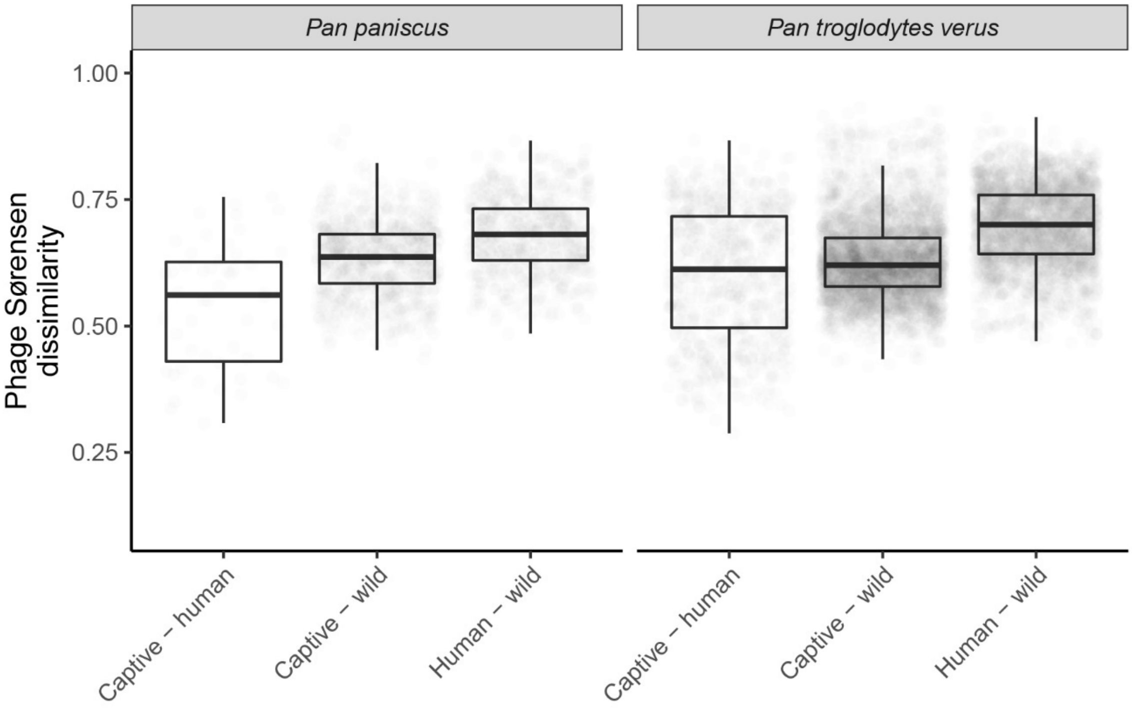

Fig. S6. Box plots showing the pairwise Sørensen's dissimilarity of the phage community composition of samples from two great ape species sampled in the wild and captivity. Raw data are plotted to aid in interpretation.

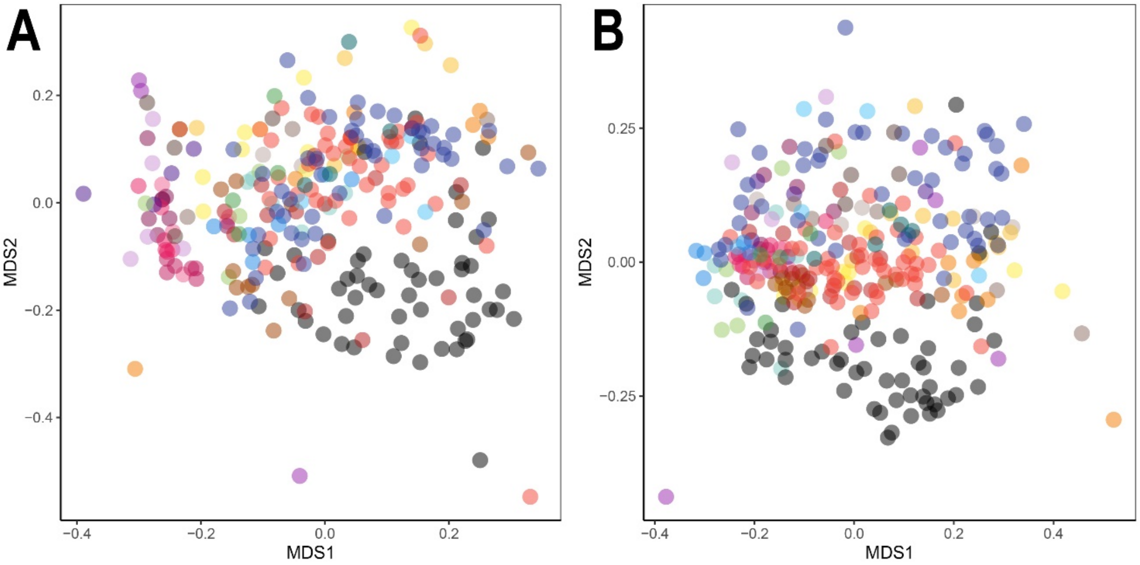

Fig. S7. Ordination of temperate (A) and virulent (B) phage community composition of wild and human superhosts (non-metric multidimensional scaling: NMDS, Sørensen's dissimilarity, $\text{stress}_{\text{temperate}}$=0.206, $\text{stress}_{\text{virulent}}$=0.239), with each point representing the phage community detected in an individual. Colors correspond to the primate superhost species in Fig. 1A.

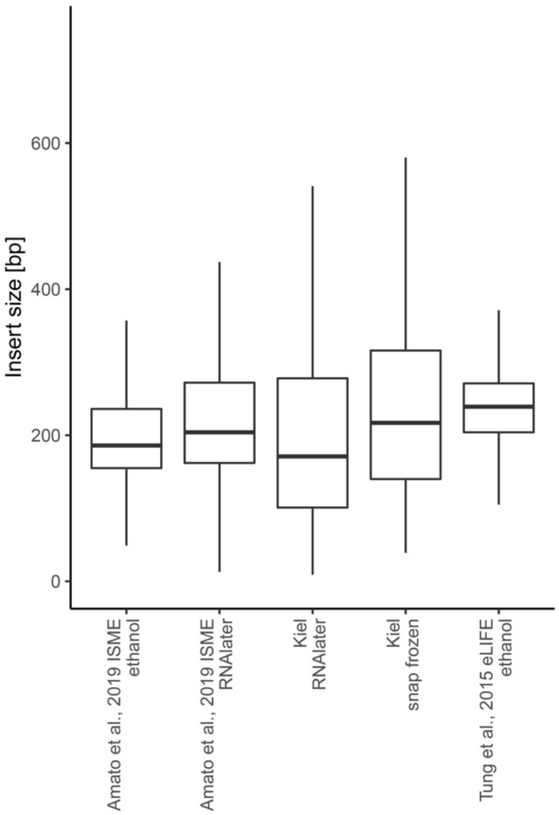

Fig. S8. Insert sizes from the different datasets include in the current study, separated by the storage method for the sample and the group that generated the data. Insert sizes were estimated by mapping a subset of reads to the mitochondrial genome of the superhost. Kiel indicates shotgun metagenomes generated as part of the current study.

Supplementary Dataset Legends (available through PNAS)

Dataset S1 (separate file). Metadata. Table including metadata for the samples included in the current study and information about where raw data for each sample are available.

Dataset S2 (separate file). Blast results. A table of BLAST results for contigs from each sample compared to the HHAP, along with an estimate of the % of the phage genome covered.

Dataset S3 (separate file). UPGMA clustering of the phage community composition for all samples from wild non-human primates and humans. Colors correspond to the primate superhost species in Fig. 1A.

Dataset S4 (separate file). Results of tests for phylosymbiosis (wild primates and humans). Results for the tests for phylosymbiosis (*Z* score and *P*-value from correlation between UPGMA distances generated from the community composition matrix and the distance matrix of the superhost phylogeny) for different taxonomic groups of wild primates and sometimes humans (Primates, Hominidae, Cercopithecidae, Atelidae, Homindae without humans, catarrhines without humans, and primates without humans). Each row represents one downsampling replicate, retaining just one sample per superhost taxon belonging to the taxonomic group in question.

Dataset S5 (separate file). Results of tests for an effect of isolation by distance on phage community similarity and for an effect of phylogenetic relatedness of superhosts on phage community similarity. Each row represents downsampling replicate, showing the results of a partial mantel test between phageome community dissimilarity and geographic distance after controlling for phylogeny and the results of a partial mantel test between phageome community dissimilarity and phylogenetic relatedness of superhosts after controlling for geographic distance.

Dataset S6 (separate file). Results of tests for superhost-specificity (wild primates and humans). Summary of the results of the test for whether within-superhost taxa distances were lower than between-superhost taxa distances for phage phylogenies containing at least two sequences representing three superhost species [categorical Mantel test: for superhost-specificity across phage phylogenies]. To summarize results across phylogenies of phages in the manuscript and not repeatedly count phages that were trimmed multiple times (i.e., there were multiple phylogenies), we selected only the first trim (indicated in table) for each phage that fulfilled the criteria (at least two sequences representing three superhost species).

Dataset S7 (separate file). Summary of results of tests of superhost-phage co-divergence (wild primates and humans). Table summarizing evidence for superhost-phage co-divergence for each phage phylogeny that contained representatives from ≥5 superhost taxa (considering wild primates and humans). To summarize the results across phylogenies of phages in the manuscript and not repeatedly count phages that were trimmed multiple times (i.e., there were multiple phylogenies), we selected only one (the first) trim for each phage that fulfilled the criteria for a ParaFit analysis, when summarizing the results and these are indicated in the table.

Dataset S8 (separate file). Raw data for the tests of superhost-phage co-divergence (wild primates and humans). The raw data summarized in Data S7; results from ParaFit tests run across the phage phylogenies (for wild primate and human superhosts). Each row represents one downsampling replicate, retaining just one representative per superhost taxon.

Dataset S9 (separate file). Summary of results of tests of superhost-phage co-divergence (Catarrhini). Table summarizing evidence for superhost-phage co-divergence for each phage phylogeny that contained representatives from ≥5 superhost taxa (considering only Catarrhini superhosts). To summarize the results across phylogenies of phages in the manuscript and not repeatedly count phages that were trimmed multiple times (i.e., there were multiple phylogenies), we selected only one (the first) trim for each phage that fulfilled the criteria for a ParaFit analysis, when summarizing the results and these are indicated in the table.

Dataset S10 (separate file). Raw data for tests of superhost-phage co-divergence (Catarrhini). The raw data summarized in Data S9; results from ParaFit tests run across the phage phylogenies (for Catarrhini superhosts). Each row represents one downsampling replicate, retaining just one representative per superhost taxon.

Dataset S11 (separate file). Results of Mantel tests of a correlation between pairwise baboon phage distances and pairwise baboon relatedness estimates.

Dataset S12 (separate file). Evidence for wild phage replacement in captivity. Results of categorical Mantel tests for each phage, comparing pairwise distances between phages from captive primates and wild primates with the pairwise distances between captive primates and humans. In addition, this table shows the results of categorical Mantel tests for each phage comparing the pairwise distances between phages from wild primates and humans with the pairwise distances between captive primates and humans. To summarize the results across phylogenies of phages in the manuscript and not repeatedly count phages that were trimmed multiple times (i.e., there were multiple phylogenies), we selected only one (the first) trim for each phage that fulfilled the criteria for this categorical mantel test comparison when summarizing the results.

Dataset S13 (separate file). Evidence for wild phage retainment in captivity. Results of categorical Mantel tests for each phage, comparing pairwise distances between phages from captive primates and wild primates with the pairwise distances between captive primates and humans. In addition, this table shows the results of categorical Mantel tests for each phage comparing the pairwise distances between phages from wild primates and humans with the pairwise distances between captive primates and wild primates. To summarize the results across phylogenies of phages in the manuscript and not repeatedly count phages that were trimmed multiple times (i.e., there were multiple phylogenies), we selected only one (the first) trim for each phage that fulfilled the criteria for this categorical mantel test comparison when summarizing the results.

Dataset S14 (separate file). BACPHILP assigned probabilities for each HHAP being temperate or virulent.

Dataset S15 (separate file). Results of tests for phylosymbiosis (wild primates and humans) for temperate phages. Results for the tests for phylosymbiosis (*Z* score and *P*-value from correlation between UPGMA distances generated from the community composition matrix and the distance matrix of the superhost phylogeny) for different taxonomic groups of wild primates and sometimes humans (Primates, Hominidae, Cercopithecidae, Atelidae, Homindae without humans, catarrhines without humans, and primates without humans). Each row represents one downsampling replicate, retaining just one sample per superhost taxon belonging to the taxonomic group in question.

Dataset S16 (separate file). Results of tests for phylosymbiosis (wild primates and humans) for virulent phages. Results for the tests for phylosymbiosis (*Z* score and *P*-value from correlation between UPGMA distances generated from the community composition matrix and the distance matrix of the superhost phylogeny) for different taxonomic groups of wild primates and sometimes humans (Primates, Hominidae, Cercopithecidae, Atelidae, Homindae without humans, catarrhines without humans, and primates without humans). Each row represents one downsampling replicate, retaining just one sample per superhost taxon belonging to the taxonomic group in question.

Supplementary Dataset Legends (available through: https://zenodo.org/record/4641870)

Zenodo Dataset 1 (separate file). Contigs. Contigs generated for each sample that were ≥500bp in length.

Zenodo Dataset 2 (separate file). Visualizations of phage phylogenies with tests for superhost-specificity (wild primates and humans). Phage phylogenies, with the superhost indicated by the color of the circles at the tip, as in Fig. 2A. Branches supported by SH-like aLRT values <0.95 are dashed. The *P*-values in the lower left corner of each phylogeny indicate the results of the test for whether within-superhost taxa distances were lower than between-superhost taxa distances [categorical Mantel test].

Zenodo Dataset 3 (separate file). Baboon phage phylogenies. The superhost's social group indicated by the color of the circles at the tip, as in Fig. 3A. Branches supported by Shimodaira-Hasegawa-like approximate likelihood ratio test values <0.95 are dashed. The *P*-values in the lower left corner of each phylogeny indicate the results of the comparison of the pairwise distance between sequences from group members and non-group members using a categorical Mantel test.

Zenodo Dataset 4 (separate file). Phage phylogenies including sequences from captive and wild non-human primates, as well as those from humans. Tips are colored as in Fig. 3. Next to each phylogeny is a box plot of pairwise phage distances, after downsampling to one phage from each superhost taxon/location. The comparisons of phages from captive primates and wild primates, captive primates and humans, as well as between wild primates and humans are shown in these boxplots.

Zenodo Dataset 5 (separate file). Mapping files of contigs ≥500bp for each sample mapped to the HHAP.

Zenodo Dataset 6 (separate file). Manually trimmed mapped contigs (wild non-human primates and humans). Manually selected region conserved across many samples and taxa.

Zenodo Dataset 7 (separate file). Maximum likelihood estimates of the phylogenies of 208 phages generated using alignment of conserved region (wild non-human primates and humans).

Zenodo Dataset 8 (separate file). Manually trimmed mapped contigs of the complete dataset (captive and wild non-human primates and humans). Considered the same region as was selected for the wild primate and human dataset.

Zenodo Dataset 9 (separate file). Maximum likelihood estimates of the phylogenies of 208 phages generated using alignment of conserved region (captive and wild non-human primates and humans).

SI Reference

55. Rowe N & Myers M (2017) All the Worlds Primates database. www.alltheworldsprimates.org