Primate phageomes are structured by superhost phylogeny and environment

Jan F. Gogarten[a,b,1], Malte Rühlemann[c], Elizabeth Archie[d], Jenny Tung[e,f,g], Chantal Akoua-Koffi[h], Corinna Bang[c], Tobias Deschner[i], Jean-Jacques Muyembe-Tamfun[j], Martha M. Robbins[i], Grit Schubert[a], Martin Surbeck[i,k], Roman M. Wittig[i,l], Klaus Zuberbühler[m], John F. Baines[n,o], Andre Franke[c], Fabian H. Leendertz[a], and Sébastien Calvignac-Spencer[a,b,1]

[a]Epidemiology of Highly Pathogenic Organisms, Robert Koch Institute, 13353 Berlin, Germany; [b]Viral Evolution, Robert Koch Institute, 13353 Berlin, Germany; [c]Institute of Clinical Molecular Biology, Christian-Albrecht-University of Kiel, 24105 Kiel, Germany; [d]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556; [e]Department of Biology, Duke University, Durham, NC 27708; [f]Duke University Population Research Institute, Duke University, Durham, NC 27708; [g]Department of Evolutionary Anthropology, Duke University, Durham, NC 27708; [h]Unité de Formation et Recherche des Sciences Médicales, Université Alassane Ouattara de Bouake, BP V1801 Bouaké, Côte d'Ivoire; [i]Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany; [j]National Institute for Biomedical Research, National Laboratory of Public Health, BP 1197 Kinshasa, Democratic Republic of the Congo; [k]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138; [l]Tai Chimpanzee Project, Centre Suisse de Recherches Scientifiques, BP 1301, Abidjan 01, Cote d'Ivoire; [m]Institute of Biology, University of Neuchatel, CH-2000 Neuchatel, Switzerland; [n]Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany; and [o]Institute for Experimental Medicine, Christian-Albrechts-University of Kiel, 24118 Kiel, Germany

Edited by James J. Bull, University of Idaho, Moscow, ID, and approved February 24, 2021 (received for review July 1, 2020)

Humans harbor diverse communities of microorganisms, the majority of which are bacteria in the gastrointestinal tract. These gut bacterial communities in turn host diverse bacteriophage (hereafter phage) communities that have a major impact on their structure, function, and, ultimately, human health. However, the evolutionary and ecological origins of these human-associated phage communities are poorly understood. To address this question, we examined fecal phageomes of 23 wild nonhuman primate taxa, including multiple representatives of all the major primate radiations. We find relatives of the majority of human-associated phages in wild primates. Primate taxa have distinct phageome compositions that exhibit a clear phylosymbiotic signal, and phage–superhost codivergence is often detected for individual phages. Within species, neighboring social groups harbor compositionally and evolutionarily distinct phageomes, which are structured by superhost social behavior. Captive nonhuman primate phageome composition is intermediate between that of their wild counterparts and humans. Phage phylogenies reveal replacement of wild great ape–associated phages with human-associated ones in captivity and, surprisingly, show no signal for the persistence of wild-associated phages in captivity. Together, our results suggest that potentially labile primate-phage associations have persisted across millions of years of evolution. Across primates, these phylosymbiotic and sometimes codiverging phage communities are shaped by transmission between groupmates through grooming and are dramatically modified when primates are moved into captivity.

bacteriophages | codivergence | phylosymbiosis | zoonotic transmission | fecal virome

Mammals harbor diverse communities of microorganisms, the majority of which are bacteria in the gastrointestinal tract. Gut bacterial communities in turn host diverse phage communities that influence their structure, function, colonization patterns, and ultimately superhost health [the superhost is the host for bacteria that in turn host the phages (1)]. For example, enriched phage communities in human intestinal mucus can act as an acquired immune system by limiting mucosal bacterial populations (2), while dysbiotic gut phageomes are associated with health conditions such as type II diabetes (3), colitis (4), and stunting (5). Transplantation of healthy viral filtrates restored health in *Clostridioides difficile* patients (6), while in vitro studies suggest phages from stunted children shape bacterial populations differently from those of healthy children (5), supporting a direct link between phageome composition and disease. However, despite their importance in gut microbial ecosystems, the ecological and evolutionary processes that gave rise to these communities remain poorly resolved. Recent work on the widespread crAssphage suggests it might demonstrate long-term associations with its superhosts (7), similar to patterns described for many bacteria (8, 9).

Primates host distinct bacterial communities, such that more phylogenetically related host taxa have more similar gut microbial composition (8, 10). The structure of these communities thus recapitulates the host phylogeny [i.e., phylosymbiosis (8, 10)], potentially reflecting widespread cospeciation of bacteria and hosts or phylogenetic conservation of the environments that shape bacterial communities (8, 9). Such long-term host–bacterial associations would imply restricted transmission of bacterial lineages within—rather than between—host lineages (8). This pattern of transmission may be facilitated by the tendency for primates to live in organized societies (11), creating opportunities for bacterial transmission to conspecifics (12, 13). When removed from their natural social and ecological environments and placed in captivity,

Significance

Mammals harbor diverse communities of gut microbes. The assembly and evolution of the bacterial components of these communities are influenced by host evolutionary histories and social behavior. Little is known about the ecological and evolutionary origins of the phages infecting these bacteria. We explore drivers of phage community assembly and phage lineage evolution in primates. Many phages codiverged with their superhosts. Furthermore, neighboring social groups harbor compositionally and evolutionary distinct phageomes, structured by superhost social behavior. Captive primate phageome composition is intermediate to humans and their wild primate counterparts, with phage phylogenies revealing replacement of wild-associated phages by human-associated lineages. This plasticity makes the long-term associations of phages with their superhosts observed across ecosystems and continents all the more striking.

Author contributions: J.F.G., J.F.B., A.F., F.H.L., and S.C.-S. designed research; J.F.G., M.R., E.A., J.T., C.A.-K., C.B., T.D., J.-J.M.-T., M.M.R., G.S., M.S., R.M.W., K.Z., F.H.L., and S.C.-S. performed research; J.F.G. and M.R. analyzed data; and J.F.G., M.R., E.A., J.T., C.A.-K., C.B., T.D., J.-J.M.-T., M.M.R., G.S., M.S., R.M.W., K.Z., J.F.B., A.F., F.H.L., and S.C.-S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: jan.gogarten@gmail.com or calvignacs@rki.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2013535118/-/DCSupplemental.

Published April 5, 2021.

primates quickly develop humanized bacterial microbiomes (14, 15). This apparent plasticity makes the long-term associations of primates with particular bacterial lineages all the more striking (8, 9).

Here, we investigate whether these key findings about primate-associated gut bacterial communities can be generalized to phages. We explore drivers of phage community composition and individual phage lineage evolution in primate superhosts across multiple scales and environments, with a particular emphasis on the potential role of social transmission. We then examine the phageomes of captive primates to understand the flexibility of phage communities in response to the environment and the potential of phage transmission between superhosts. Lastly, we explore whether temperate versus virulent phage lifestyles influence the observed patterns in phage community composition and/or individual phage lineage evolution.

Results

Detection of Nonhuman Primate Phages. We used a database of healthy human-associated phages (HHAPs; n = 4,301 double stranded DNA phages; 16) to identify related phages in fecal shotgun metagenomes from 23 wild nonhuman primate taxa ($N_{\text{individuals}}$ = 243; Figs. 1 and 2*A*). This database of HHAPs was assembled via ultradeep sequencing of purified virus-like particles isolated from the stool of healthy humans, and many of these phages have been found in metagenomic datasets of healthy individuals in different populations, supporting their broad association with humans (16). The nonhuman primate superhost taxa spanned all major radiations of primates, including wild catarrhines (Hominidae, hereafter great apes: N_{species} = 4; $N_{\text{subspecies}}$ = 6; Cercopithecidae, hereafter cercopithecids: N_{species} = 7), platyrrhines (Atelidae, hereafter atelids; N_{species} = 7), and lemurs (Lemuridae: N_{species} = 2; Indriidae: N_{species} = 1). In addition, we analyzed fecal shotgun metagenomes from humans living in Africa ($N_{\textit{Democratic Republic of Congo}}$ = 12, $N_{\textit{Côte d'Ivoire}}$ = 12) and Europe (N_{Germany} = 24; Fig. 1, Dataset S1, and *SI Appendix*, *Supplementary Information*). We performed de novo assembly of metagenomic contigs and blasted contigs ≥500 bp against the HHAP database (min. E-value = 1e-3; https://zenodo.org/record/4641870; Zenodo Dataset 1; ref. 17), allowing us to find both close matches to known human-associated phages and those that are related to human-associated phages but substantially divergent (min. identity = 68.6%; Dataset S2). We populated a phage community matrix by considering a phage present when we detected a ≥500 bp contig covering ≥10% of an HHAP's genome (minimum E-value = 1e-3). Of the 4,301 HHAPs, 2,639 (61.4%) have relatives in at least one wild nonhuman primate superhost taxon and 1,243 (28.9%) in five or more, while 32 (0.74%) had relatives in all the superhost taxa we considered (*SI Appendix*, Fig. S1).

Phage Community Composition across Primates. Fecal phage community composition differs by superhost taxon (analysis of variance using distance matrices: R^2 = 0.418, $F_{23,271}$ = 8.478, P = 0.001; Fig. 2*B*), indicative of superhost specificity. Phage community composition also differs by superhost family (analysis of variance using distance matrices: R^2 = 0.159, $F_{4,290}$ = 13.651, P = 0.001; Fig. 2*C*), though great apes and cercopithecids partially overlap in this ordination. To further test for potential phylosymbiosis,

Fig. 1. Map indicating locations where samples from superhost taxa included in the current study originated. Black circles indicate samples that were collected and sequenced as part of the current study, while white circles indicate superhost samples that were sequenced by others previously. Map was created with the packages ggreple, rnaturalearth, and sf (53–55).

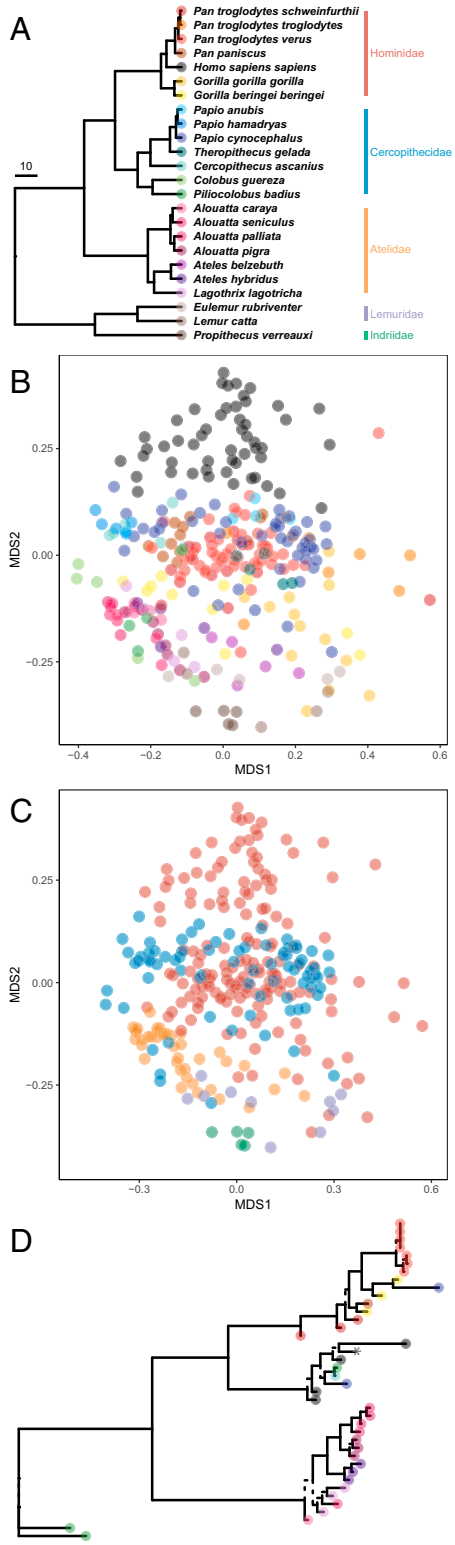

Fig. 2. Wild nonhuman primate and human phageomes. (*A*) A phylogeny of the wild primate taxa examined in this study. Scale in millions of years. (*B*) An ordination of phage community composition for these species (non-metric multidimensional scaling: NMDS, Sørensen's dissimilarity, stress = 0.182), with each point representing the phage community detected in an individual; colors correspond to the primate superhost species in *A*. (*C*) The same NMDS plot of phage community composition, now colored by the superhost's family, as indicated in *A*. (*D*) A phage phylogeny demonstrating evidence for host-specificity (i.e., within-superhost species distances were significantly lower than between-superhost species distances; categorical Mantel: $P = 0.001$). This phage phylogeny also shows evidence consistent with codivergence between superhosts and the phage (i.e., all of the 1,000 ParaFit tests after downsampling to one representative sequence per superhost taxa were significant). The * indicates the reference HHAP sequence generated in ref. 16. Branches supported by Shimodaira–Hasegawa-like approximate likelihood ratio test values <0.95 are dashed.

we downsampled to a single sample per superhost taxon ($N_{replicates}$ = 1,000), performed hierarchical clustering of phage community structure with the unweighted pair group method with arithmetic mean (UPGMA), and tested for congruence of the UPGMA dendrogram and the superhost phylogeny with a ParaFit test (18). ParaFit tests whether the similarity between two trees is higher than expected by chance (18). We find broad support for phylosymbiosis across primates (95.1% of downsampling replicates were significant; $P \leq 0.05$; Datasets S3 and S4), revealing that phage communities were less similar with greater divergence times between their superhosts (*SI Appendix*, Fig. S2*A*).

To formally assess whether this signal of phylosymbiosis is driven by deep branches separating primates living in Madagascar (lemurs), the Americas (atelids), and mainland Africa (great apes and cercopithecids), we repeated the ParaFit analysis described above, focusing separately on atelids ($n = 7$ taxa), cercopithecids ($n = 7$ taxa), and great apes ($n = 7$ taxa; the small number of lemur taxa sampled precluded an analysis within this group). Despite small sample sizes, phylosymbiotic signal is detected in 31.0% of great ape, 35.1% of cercopithecid, and 66.9% of atelid replicates (Dataset S4 and *SI Appendix*, Fig. S2*B*). The reduced signal identified in apes and cercopithecids might reflect the fact that many of these species have overlapping geographic distributions in mainland Africa (39.7% of species pairs, compared with 23.8% of atelid taxa; *SI Appendix*, Fig. S3).

To test whether observed patterns of phylosymbiosis are driven by isolation by distance, we focused on catarrhines in mainland Africa, which represented the largest dataset on one continent. We then tested whether phageome community composition was predicted by geographic distance, after controlling for the phylogenetic relatedness of superhosts. This dataset contained both wild species that are sympatric and those separated by more than 5,200 km (*SI Appendix*, Fig. S4). We found little support for a pattern of isolation by distance, with only 14 of 1,000 (1.4%) downsampling replicates showing a significant relationship between phageome community dissimilarity and geographic distance after controlling for phylogeny (Dataset S5). In contrast, when we tested whether phageome community composition was predicted by phylogenetic relatedness of superhosts, after controlling for their geographic distance, 453 of 1,000 (45.3%) of downsampling replicates showed a significant relationship (Dataset S5). Patterns of phage–primate phylosymbiosis are thus widespread within and between continents and across a range of phylogenetic scales, but at least in regards to the scales investigated here, these patterns do not appear to be driven by isolation or by distance.

Phage Evolution across Primates. Observed patterns of phylosymbiosis could reflect a process of codivergence between phages and their superhosts. To investigate this possibility, we generated maximum likelihood estimates of the phylogenies of 208 phages detected in at least 10 primate taxa and from which relatively long sequence alignments could be generated ($\bar{x}_{length}$ = 1,411 bp; min = 355; max = 16,851). For each phage phylogeny, we first tested whether phages from the same superhost taxon were more closely related to each other than to phages from other superhost taxa (i.e., superhost specificity), by testing whether within-superhost taxa distances were lower than between-superhost taxa distances with categorical Mantel tests. Superhost specificity is detected in the majority (87.8%) of phage phylogenies tested ($n = 189$ phylogenies including at least two sequences in each of three superhost species; Fig. 2*D*

and Dataset S6; https://zenodo.org/record/4641870; Zenodo Dataset 2). As a further test of superhost–phage codivergence, we then ran ParaFit tests for each phage phylogeny (18), accounting for the observed superhost specificity by downsampling to one representative per superhost taxa ($N_{replicates}$ = 1,000). These ParaFit analyses test the null hypothesis that phages associate randomly with superhosts (against the alternative hypothesis of codivergence between phages and superhosts; 18).

We detect patterns consistent with codivergence in 44 phages (22.1%; ≥95% of replicates significant, $P \leq 0.05$; n = 199 phylogenies with ≥5 superhost taxa represented; Datasets S7 and S8). Phage phylogenies with smaller numbers of superhost taxa represented (n = 96 of phylogenies with <10 superhost taxa) were less likely to exhibit evidence of codivergence (11.5%) compared to those with more superhost taxa represented (32.0%; n = 103 phylogenies with ≥10 taxa). Because power increases with sample size, our estimate of codivergence frequency is therefore likely to be conservative (goodness of fit test, G = 10.0, P = 0.0015).

To control for the effects of deep tree structure, we conducted parallel analyses focusing on cercopithecid and great ape phages. For these smaller phylogenies, 10 phages (5.2%) display a pattern consistent with codivergence (n = 192 distinct phage phylogenies with ≥5 taxa; Datasets S9 and S10). Results show that codivergence of host-specific phages with their superhosts occurs within families/continents, but the predominant signal for codivergence observed across primates is driven by deep, interfamily/intercontinental structure.

Within-Species Phage Community Composition and Evolution. To investigate how long-term host specificity and patterns of codivergence might arise, we examined phage communities in 48 baboon (*Papio cynocephalus*) fecal shotgun metagenomes from two social groups in the same population ("Mica's group" and "Viola's group:" Fig. 3*A*; 12). In a pattern that strongly parallels gut bacterial microbiome structure, phage communities from baboons in the same social group are more similar to each other than to members of different groups (analysis of variance using distance matrices: R^2 = 0.0355, $F_{1,46}$ = 1.692, P = 0.049; Fig. 3*B*). Further, for both groups, stronger within-group grooming relationships predict more similar phage communities, even after controlling for kinship (using pedigree-based pairwise relatedness estimates from: 12; Mica's group partial Mantel test: r = -0.258, P = 0.002; Viola's group partial Mantel test: r = -0.112, P = 0.015; Fig. 3*C*). The social network of Mica's group exhibits more substructure than that of Viola's group ($\bar{x}_{\text{weighted clustering coefficient Mica's group}}$ = 0.647; $\bar{x}_{\text{weighted clustering coefficient Viola's group}}$ = 0.527; t test: $t_{30.63}$ = 2.27, P = 0.030), which may explain the stronger relationship between social behavior and phage community composition in this group.

We next examined whether social group membership predicts the genetic structure of common baboon gut phages (n = 70 phages present in at least two baboons in each group). To do so, we compared the pairwise distance between sequences from group members and nongroup members using categorical Mantel tests. For 7 phages (10.0%), pairwise distances are significantly lower for sequences sampled from the same social group than between social groups (Fig. 3*D*; https://zenodo.org/record/4641870; Zenodo Dataset 3). To place these results in context, we also examined whether more related individuals harbor more closely related phages by substituting the group membership matrix for relatedness estimates. We observed positive correlations between pairwise relatedness and phage sequence distance for 4 baboon phages (Dataset S11), suggesting that some phage lineages are maintained within groups of related individuals (e.g., via vertical or horizontal transmission within matrilines).

These findings suggest that phylogeographic patterns observed for human-associated crAssphages between cities and across continents (7) hold for at least a subset of baboon-associated phages in groups that range within a few kilometers of each other. Microorganism transmission through social interactions within groups thus not only shapes baboon bacteria community composition (12) but phage community composition and evolution as well.

Cross-Species Phage Transmission. If primate social interactions reinforce superhost-specific phageome composition, phageomes might be affected by environments that substantially depart from natural conditions. To quantify this response, we analyzed the phage communities in fecal samples ($N_{individuals}$ = 55) of captive great apes from four taxa, as well as four zookeepers. Phage community composition was predicted by the location of the superhost (i.e., captive or wild for nonhuman primates; humans living in Africa or Europe, or those working as zookeepers; analysis of variance using distance matrices: R^2 = 0.263, $F_{4,200}$ = 17.832, P = 0.001; Fig. 4*A* and *SI Appendix*, Fig. S5). Zookeeper phage community composition falls within the diversity observed for humans (Fig. 4*A*), while phage community composition of captive great apes is intermediate between zookeepers and wild great apes. Interestingly, captive great ape phage community composition is more similar to zookeepers than that of their wild great ape counterparts (Fig. 4 *A* and *B*). For species sampled in the wild and captivity, captive great ape phageomes are also more similar to those of humans than their wild counterparts are to humans ($N_{randomizations}$ = 1,000, $P_{Pan\ troglodytes\ verus}$ = 0.001, $P_{Pan\ paniscus}$ = 0.001, *SI Appendix*, Fig. S6). Captive great ape phageome composition is therefore clearly distinct from that of their wild great ape counterparts and appears to be humanized (Fig. 4*A*).

To understand the processes driving the humanization of the captive great ape phageome, we examined the sequences of their phages. Overall, phage sequence identity to HHAP reference sequences was higher in humans and captive great ape superhost taxa than in wild great ape superhosts (Fig. 4*C*). We examined phage phylogenies containing at least one captive primate, one wild primate, and one human sequence to formally evaluate potential phage transmission in captivity. We downsampled to one phage from each superhost taxon/location combination and examined pairwise distances between the phages from captive and wild primates, as well as pairwise distances between captive primates and humans using categorical Mantel tests. In 31 of the 176 (17.6%) phages examined, pairwise distances are greater between sequences from captive and wild great apes than between captive great apes and humans and are also greater between wild great apes and humans than between captive great apes and humans (Fig. 4*B* and Dataset S12; https://zenodo.org/record/4641870; Zenodo Dataset 4). The inverse scenario, which would suggest a wild great ape–associated phage was retained in a captive great ape (i.e., pairwise distances are greater between sequences from captive apes and humans than between captive great apes and wild great apes and are also greater between sequences from wild great apes and humans than between captive great apes and wild great apes) did not occur in our dataset (0 of 176 phages; 0%; Datasets S12 and S13). These findings not only suggest phage transmission occurs in captivity, but also indicate little or no retention of wild great ape–associated phages in their captive conspecifics. Remaining differences in phageome composition between humans and captive great apes are likely a product of superhost filtering of human-derived bacterial communities and their phages.

Phage Lifestyle. Phages can lead either temperate or virulent lifestyles, which could affect their long-term association with superhosts. Virulent phages infect and obligately kill bacterial cells to release their phage progeny (a lytic cycle). While temperate phages can also kill their bacterial hosts to release progeny, they also have the ability to integrate into their bacterial host's genome and

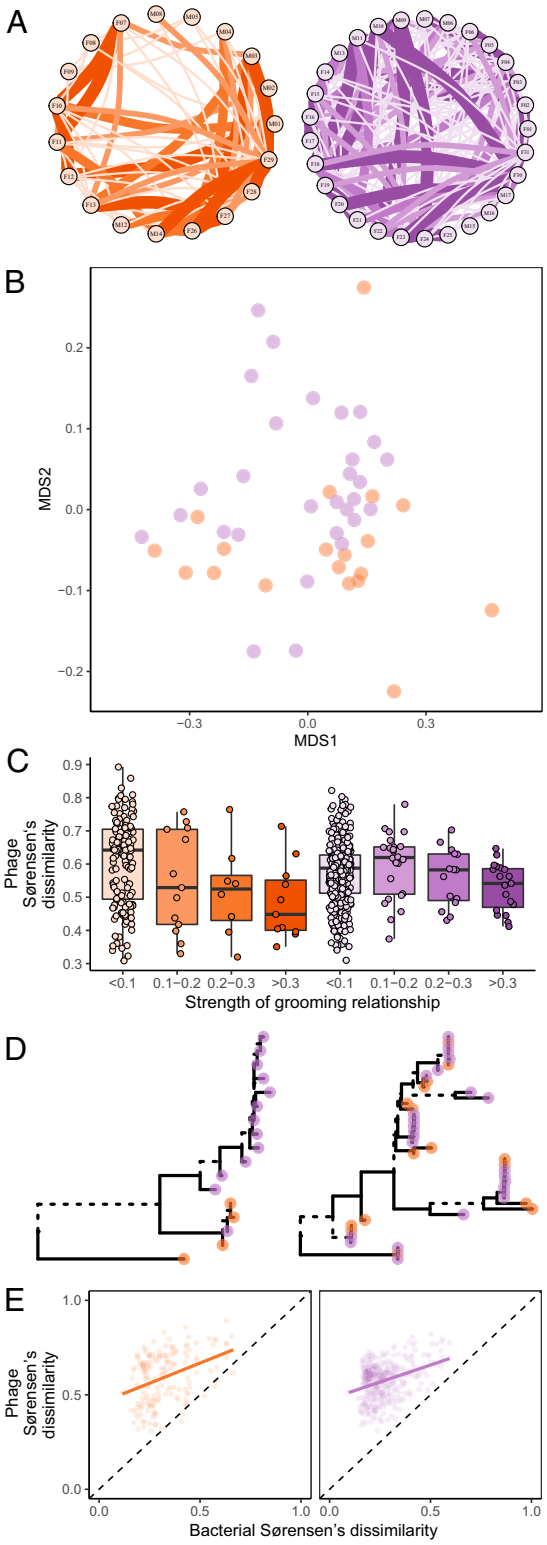

Fig. 3. Within-species phage ecology and evolution. (*A*) The social networks of two neighboring social groups of baboons (orange = Mica, purple = Viola). Circles represent individuals (individual ID shown within circle) and thickness and color of lines indicate the strength of the grooming relationships between individuals. (*B*) An ordination of phage community composition (NMDS, Sørensen's dissimilarity, stress = 0.153), with each point representing the phage community from a fecal sample (colors correspond to groups as shown in *A*). (*C*) Box plots showing the relationship between pairwise grooming bond strength and pairwise Sørensen's dissimilarity in phage community composition in the social groups. Raw data are plotted to aid in interpretation. (*D*) Baboon phage phylogenies; the left is an example where pairwise distance between phages from group members is lower than between nongroup members (categorical Mantel: $P = 0.001$). On the right, an example where there is no difference between the pairwise distances of phages from groupmates and nongroupmates (categorical Mantel: $P = 0.456$). Branches supported by Shimodaira–Hasegawa-like approximate likelihood ratio test values <0.95 are dashed. (*E*) The relationship between phageome community composition and bacterial community composition in these social groups. The dashed lines show the line of equality and the solid-colored line represents the fit of a linear model to aid in interpretation of the relationship; significance was assessed with Mantel tests, not these linear models (Mica's group: $Z = 33.15$, $P = 0.003$; Viola's group, $Z = 62.86$, $P = 0.001$).

replicate without lysing the cell (a lysogenic lifestyle). Temperate phage–host relationships may therefore be more mutualistic than virulent phage–host interactions, which are more likely to be antagonistic (19). As a result, coevolutionary processes with bacteria have been hypothesized to occur more frequently in temperate phages than in virulent phages (19). To test this hypothesis, we examined how phage lifestyles influence the patterns of phylosymbiosis and codivergence of primates and their phage communities. Traditionally, phage lifestyle has been determined experimentally, but such approaches are not easily scaled up to diverse phage communities. As an approximation, we estimated the likelihood that HHAPs were temperate or virulent using the BACterioPHage LIfestyle Predictor (BACPHLIP). BACPHILP identifies the presence/absence of conserved protein domains associated with temperate lifestyles, assigning a probability of being temperate to each phage (20).

We identified 182 phages that had a high probability (≥90%) of being temperate and selected the same number of phages with the highest probability of being virulent (Dataset S14). Mirroring the full dataset, the community composition of temperate phages and virulent phages covaried with superhost taxon (analysis of variance using distance matrices: temperate, $R^2 = 0.362$, $F_{23,255} = 6.276$, $P = 0.001$; virulent, $R^2 = 0.314$, $F_{23,269} = 6.065$, $P = 0.001$; *SI Appendix*, Fig. S7) and by superhost family (analysis of variance using distance matrices: temperate, $R^2 = 0.150$, $F_{4,274} = 12.114$, $P = 0.001$; virulent, $R^2 = 0.118$, $F_{4,288} = 9.670$, $P = 0.001$). We also found similar support for phylosymbiosis for temperate versus virulent phages across all sampled taxa (32.6% of downsampling replicates were significant for temperate phages; $P \leq 0.05$; compared with 26.7% of replicates for virulent phages; Datasets S15 and S16) and within major primate lineages (temperate: 11.9% of great ape, 0.18% of cercopithecids, and 63.6% of atelid replicates; virulent: 8.5% of great ape, 0.29% of cercopithecid, and 51.4% of atelid replicates; Datasets S15 and S16). Of the 44 phages in which we detected patterns consistent with codivergence, 3 were predicted to have a temperate lifestyle and 2 were predicted to have a virulent lifestyle. Overall, these results suggest that temperate phages are not more likely to codiverge with their superhosts than phages in general (6.8% of codiverging phages were temperate; temperate phages in dataset = 4.2%; goodness of fit test, $G = 0.617$, $P = 0.432$).

Discussion

Our results reveal striking parallels between the ecological and evolutionary patterns of primate-associated phageomes and those previously reported for primate-associated bacterial communities (8–10, 12–15, 21). These findings likely reflect the tight ecological links between phages and their bacterial hosts. Indeed, phage community composition and bacterial community composition are strongly positively correlated across species and locations (Figs. 3*E* and 4*C*). However, in one baboon group, closer grooming relationships predict increased phage community similarity even after controlling for similarity in bacterial community composition (partial Mantel test; $r = -0.276$, $P = 0.007$; Viola's group; $r = -0.056$,

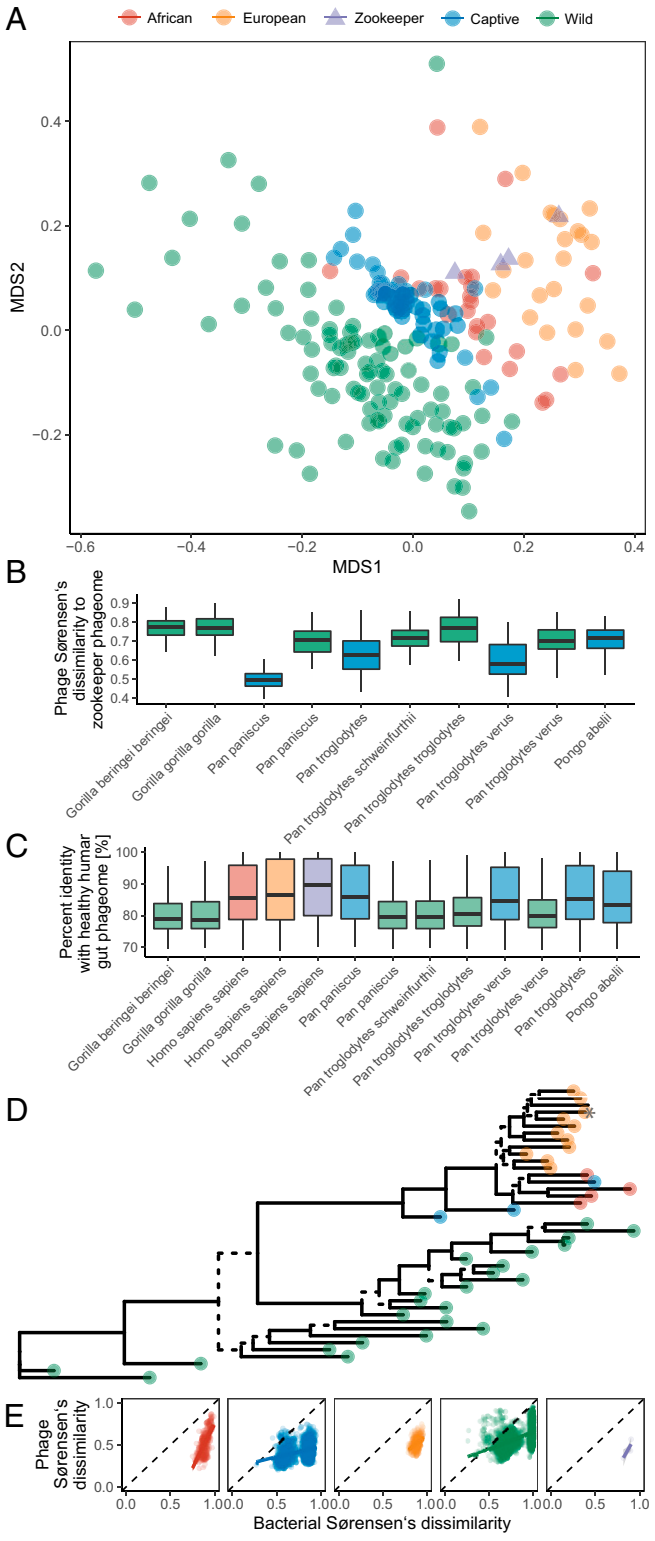

Fig. 4. Captive primate phageomes. (*A*) An ordination of great ape phage community composition (NMDS, Sørensen's dissimilarity, stress = 0.195) colored by their origin. (*B*) Box plots showing the dissimilarity of great ape phageome community composition to that of zookeepers. The phageomes of wild and captive samples from a species are compared separately, following the coloring scheme in *A*. (*C*) The percent identity of phages from wild great apes, captive great ape, and humans, compared to the HHAP, following the coloring scheme in *A*. (*D*) An example phage phylogeny suggestive of human to captive great ape phage transmission (captive great apes are nested within the branch containing all humans instead of the branch containing all wild great apes): colors are indicative of the categories indicated in *A*, the * indicates the reference HHAP, and branches supported by SH-like aLRT values <0.95 are dashed. (*E*) The relationship between bacterial community composition and phage community composition for the location categories indicated in *A*. The dashed lines show the lines of equality. The solid-colored lines represent the fit of linear models to aid in interpretation of the relationship; significance was assessed with Mantel tests—not these linear models (African humans, $Z = 132.06$, $P = 0.001$; captive great apes, $Z = 488.74$, $P = 0.001$; European humans, $Z = 105.66$, $P = 0.001$; wild great apes, $Z = 2774.68$, $P = 0.001$). For zookeepers, we did not detect a significant relationship ($Z = 2.29$, $P = 0.135$), though the relationship was in the expected direction and sample size was small ($n = 4$).

$P = 0.142$). Thus, phage community structure may not be shaped by bacterial community structure alone, but also by differential transmissibility of some phages between hosts or direct selection of specific phages by their superhosts (2, 22).

Our finding that captive primates have humanized phageomes mirrors findings for the bacterial communities of captive primates (14, 15). To date, no phylogenetic analyses of bacterial lineages demonstrate bacterial transmission in captivity. By constructing phage phylogenies, we were able to show widespread replacement of wild-associated lineages in captivity, with no evidence that phages found in wild apes are retained in captive animals. This result suggests that the humanization of great ape bacterial communities in captivity may also involve the replacement of lineages normally found in wild populations. It also indicates that primate phageomes are extremely labile, with captive great ape phageomes most similar to a modified human phageome. The fact that captive phageomes are not an exact replicate of the human phageome could be due to superhost factors (e.g., genetic factors) or differences in the environment of humans and captive primates (e.g., dietary differences). However, the convergence of great ape phageomes from different species in captivity suggests environmental differences are particularly important in modifying the phageomes of captive great apes. Determining the functional consequences of humanized phageomes for primate superhosts, as well as the scale and scope of phage transmission in environments that have shifted less dramatically (e.g., at human-wildlife interfaces, such as habituated populations visited by tourists), or between sympatric species in less-modified environments, represents exciting areas of future research.

Our analysis focused on relatives of the healthy human phageome in wild and captive primates. As catalogs of human phage diversity improve, they promise to reveal additional phage relatives in nonhuman primates that will expand the scope of the types of analyses reported here (23). At the same time, more complete descriptions of wild and captive primate phageomes through de novo prediction of phages in metagenomes should allow researchers to discover phages without relatives in human populations (24–26). Whether phage enrichment prior to sequencing simplifies such efforts when working with fecal samples represents a critical technical question that will help guide these efforts (23, 27). Notably, humans have species-poor bacterial communities compared to nonhuman primates (28) and a systematic exploration of primate phage diversity will allow an assessment of what phages humans may have lost during their recent evolutionary history.

Overall, our results suggest that primate phageomes (regardless of phage lifestyle) have been shaped by complex interactions with their bacterial hosts and primate superhosts, which resulted in evolutionary stable, but potentially labile, phage–host–superhost associations. These phylosymbiotic and sometimes codiverging communities are partly shaped by transmission between groupmates through grooming. Phage communities are substantially modified when superhosts are moved into captivity through replacement of wild-associated lineages by human-associated phages. Intriguingly, phage community structure is not shaped

by bacterial community structure alone and understanding the factors responsible for this uncoupling represents an important avenue of research (29). Together, these findings provide an essential backdrop for further investigations into the recent evolutionary trajectory of human phageomes.

Materials and Methods

Fecal Sampling. Fecal samples were collected from habituated, individually identifiable, wild great apes immediately after defecation ($N_{individuals}$ = 100; N_{taxa} = 6; N_{sites} = 5; $N_{countries}$ = 4; Fig. 1 and Dataset S1). Depending on infrastructure available at these field sites, fecal samples were either 1) stored in an equal volume of RNAlater and subsequently frozen at −20 °C or 2) put into a cryotube, kept cool in a thermos in the field, and snap frozen in liquid nitrogen upon returning to the field laboratory and subsequently maintained at ≤−80 °C. Appropriate government permits and permission to conduct research on these wild primates were granted by the relevant authorities (see *Acknowledgments* for site-specific details).

Fecal samples were collected from individually identifiable captive great apes in German zoos ($N_{individuals}$ = 55; N_{taxa} = 4; N_{zoos} = 4). Stool samples from zoo animals were collected in four different zoos within Germany (Gettorf in Schleswig-Holstein, Hagenbeck in Hamburg, Leipzig in Saxonia, and Schwaigern in Baden-Württemberg) by the responsible animal keepers and snap frozen at −80 °C. Detailed information on the captive animals' origins can be found at https://www.zootierliste.de/.

To create a comparable dataset from humans, focusing specifically on countries where nonhuman primates were sampled for this project, human fecal samples were collected in the Democratic Republic of Congo ($N_{individuals}$ = 12), the Cote d'Ivoire ($N_{individuals}$ = 12), and Germany ($N_{individuals}$ = 24; Fig. 1). In the Democratic Republic of Congo, fecal samples were stored in an equal volume of RNAlater and subsequently frozen at −20 °C. In the Cote d'Ivoire, feces were put into a cryotube, kept cool in a thermos until they could be snap frozen in liquid nitrogen upon returning to the field laboratory, and subsequently maintained at ≤ −80 °C. In addition, fecal samples were collected from zookeepers working with the captive primates at one German zoo and snap frozen at −80 °C ($N_{individuals}$ = 4). Human stool samples from Northern Germany were collected by the participants themselves at their respective homes in standard fecal collection tubes and mailed to the study center, where they were then flash frozen at −80 °C. The study on human samples was performed in accordance with the declaration of Helsinki. Written informed consent was obtained from all human study participants. Ethical approval was obtained from the Local Ethics Committee Germany, Kiel (reference number A156/03).

Published Data. In addition to the samples analyzed here, we included data generated in prior analyses using comparable methods. Specifically, we included data generated from wild nonhuman primate fecal samples ($N_{individuals}$ = 95; N_{taxa} = 18) from field sites across the globe (N_{sites} = 13; $N_{countries}$ = 9) that were preserved in RNAlater or 95% ethanol [Fig. 1 and Dataset S1 (30)]. We also included shotgun metagenomes generated from wild baboon feces preserved in 95% ethanol ($N_{individuals}$ = 48), collected by the Amboseli Baboon Research Project (Fig. 1). The baboon samples represent a nearly complete sampling of the adult members (92%) of two neighboring social groups over a single month (referred to as "Mica's group" and "Viola's group"; 12). In addition to these shotgun metagenomes, Tung et al. (12) collected representative grooming data from the year before fecal samples were collected, from both social groups (including the month they were collected). These data were used to calculate the number of observed grooming interactions between adult dyads in each of the two social groups. This allowed Tung et al. to generate a matrix of grooming relationship strength, by scoring the strongest dyadic grooming relationship in each group as a 1 and weighting all other dyadic relationships relative to this strongest bond (12). Social networks were constructed using grooming interactions of baboons in the year prior to and including the month of fecal sampling (data presented in ref. 9) and visualized using the igraph R package with a circular layout (31). To estimate the substructure of these networks, we estimated the vertex level transitivity by calculating the weighted clustering coefficient (32) as implemented in the igraph "transitivity" function. This metric can range from 0 to 1, with higher transitivity indicative of networks that are more subdivided into different modules or cliques. We compared the vertex level weighted clustering coefficients of the two groups using a *t* test. Tung et al. also estimated superhost pairwise genetic relatedness values from the extensive pedigree data available for the Amboseli population (12) and bacterial community composition using the program MetaPhlAn 2.0 (12).

Host Phylogeny. As an estimate of the superhost's evolutionary relationships, we used the consensus phylogeny from the 10kTrees project (Fig. 2*A*; V3; 33). The 10kTrees phylogeny includes over 300 species and subspecies and is based on 17 genes; from this phylogeny, we selected those superhost taxa for which shotgun metagenomes were available. *Ateles hybridus* was missing from this phylogeny; we added this taxon by using an estimated 4.5 Mydivergence time from *Ateles belzebuth* (estimated in ref. 34).

DNA Extraction. For DNA extraction out of stool samples, 200 mg were transferred to 0.70 mm Garnet Bead tubes (Qiagen) filled with 1.1 mL ASL buffer. Subsequently, bead beating was performed using the SpeedMill PLUS (Analytik Jena AG) for 45 s at 50 Hz. Samples were then heated to 95 °C for 5 min and centrifuged afterward. 200 μl of the resulting supernatant were transferred and processed with the QIAamp DNA Stool Mini Kit (Qiagen) automated on a QIAcube system (Qiagen) according to the manufacturer's protocol.

Shotgun Library Preparation and Sequencing. Quality as well as quantity of stool DNA samples were determined by Qubit measurements and by using the Genomic DNA ScreenTape (Agilent). Subsequently, metagenomic library preparation was performed using the Illumina Nextera DNA Library Preparation Kit (as described in detail in ref. 35). Sequencing was performed either with 2 × 125 bp reads on a HiSEq. 2500 platform or with 2 × 150 bp reads on a HiSEq. 4000 machine.

16S Amplicon Preparation and Sequencing. From all great ape and human fecal samples analyzed as part of the present study, we used 16S rRNA gene amplicon sequencing to characterize the bacterial communities. Extracted fecal DNA was subjected to PCR amplification of the V1-V2 fragment (primer pair 27F-338R) using barcoded fusion primers for Illumina sequencing. PCR products were pooled into sequencing libraries in equimolar amounts using the SequalPrep Normalization kit and sequenced on the Illumina MiSeq using v3 chemistry for 2 × 300 bp reads. This 16S amplicon sequencing was successful for all but two samples from German humans (Dataset S1).

Contig Assembly from Shotgun Metagenomes. Shotgun metagenomic data were quality controlled and preprocessed using the BBTools software suite (36). Briefly, Nextera and TruSeq sequencing adapters and low quality sequences were trimmed, followed by removal of sequencing artifacts and PhiX reads using bbduk.sh. Host reads were removed using a human reference database and a lenient threshold of 95% identity to account for a broader host range using bbmap.sh. Overlapping forward and reverse reads were merged with the bbmerge.sh module (37). Metagenome assembly was performed by metaSPAdes (38), using k-mers of size 21, 33, and 55 (https://github.com/mruehlemann/metagenome_preproc/blob/master/qc_and_assemble.slurm). We then selected contigs ≥500 bp for subsequent analyses (https://zenodo.org/record/4641870; Zenodo Dataset 1).

Populating a Phage Community Matrix. For each contig ≥500 bp, we used BLAST to compare it against the HHAP (17), using a minimum E-value of 1e-3. For each contig, we then removed hits that were less than 500 bp in length and that spanned less than 10% of a HHAP, and then kept only the maximum bitscore for a particular contig (i.e., each contig was only counted once; Dataset S2). From these results, we populated a presence–absence community matrix of all HHAP phages using the *data.table* R package (39).

Analysis of Phage Community Composition. To assess the dissimilarity of phage communities, we calculated the Sørensen's dissimilarity metric with the vegdist function in the vegan R package (40). For a pair of samples, this is calculated as follows:

$$D_{12}^{Sor} = 1 - \frac{2a}{2 + b + c},$$

where *a* is the number of phages that are found in both samples 1 and 2; *b* the number of phages found in sample 1 but not in sample 2; and *c* the number in sample 2 but not sample 1. This metric can range from 0 to 1, with 0 indicating a pair of samples with identical phage communities, and 1 indicating two samples that contain no shared phages. We elected to use Sørensen's dissimilarity metric, as it emphasizes community membership (presence–absence), deals well with outliers, and does not consider "shared absence" as indicative of similar communities.

Ordination of phage community composition was performed using the vegan R package (40) and visualized using the ggplot2 (41) package; only a single sample per individual was included in these analyses (Dataset S1). Analysis of variance using distance matrices was performed on the Sørensen's

dissimilarity matrix using the adonis function in the vegan R package (40). Visualizations of ordinations were created with the R packages ggplot2 (41) and gridExtra (42). For the reduced datasets of temperate or virulent phages, samples in which 1 or 0 temperate or virulent phages identified were detected were not included in the community composition analyses.

To test for phylosymbiosis, we downsampled to a single sample per superhost taxon ($N_{replicates}$ = 1,000) and performed hierarchical clustering of the Sørensen's dissimilarity matrix using the hclust function with the UPGMA agglomeration method in R. We tested for congruence of the UPGMA dendrogram and the superhost phylogeny with a ParaFit test (18). Simulations by Gottschling et al. demonstrated that a minimum of five associations are necessary for topological comparisons with ParaFit, so we focused on datasets that included at least five superhost taxa (43). Statistical analyses were performed in R (version 3.6.1; 44).

To test for isolation by distance in the catarrhine samples, we downsampled to a single sample per superhost taxon ($N_{replicates}$ = 1,000) and tested for a relationship between the Sørensen's dissimilarity matrix and geographic distance. We used a partial Mantel test, controlling for phylogenetic relatedness of the superhosts with the distance matrix of the superhost phylogeny.

To test whether the difference between the similarity of human and captive superhosts and those of human and wild superhosts was larger than expected by chance, we used randomization tests. Specifically, we randomized the status (captive or wild) of samples within each species and calculated the differences between the mean for these categories (human and captive superhosts versus human and wild superhosts). To estimate a *P* value, we calculated the proportion of randomizations for which the observed difference of the mean for these categories was greater than the observed value.

Constructing Phage Phylogenies. We generated phage phylogenies by mapping contigs (≥500 bp) from wild nonhuman primates and humans to the HHAP using BWA mem (min. seed length = 40). We sorted mapping files with the SortSam tool in the Picard suite (45) and used SAMtools to retain contigs mapping to ≥500 bp with a MAPQ score >30 (https://zenodo.org/record/4641870; Zenodo Dataset 5; ref. 46). For those phages that were present in at least 10 taxa, we manually selected regions of these mapped contigs present in many superhost taxa in Geneious (V11; https://zenodo.org/record/4641870; Zenodo Dataset 6). For 14 phages we selected multiple potentially informative regions and analyzed these separately and the additional trimmed files are annotated with "trim" (https://zenodo.org/record/4641870; Zenodo Dataset 6). In the main text, we report statistical results from the first phylogeny for each of those phages that survived the thresholds necessary for it to be included in a particular analysis. From these alignments, we removed sites with gaps or Ns in Geneious, and generated maximum likelihood phylogenies using PhyML with Smart Model Selection (v1.8.1), a full optimization approach, tree search using subtree pruning and regrafting, and the Bayesian information criterion for model selection (https://zenodo.org/record/4641870; Zenodo Dataset 7 ref. 47). We estimated branch robustness using a Shimodaira–Hasegawa-like approximate likelihood ratio test (48).

We repeated the process of constructing phage phylogenies for captive primates, by mapping contigs ≥500 bp and selecting the same region as was selected for the wild nonhuman primates and human dataset from a mapped file including all data (https://zenodo.org/record/4641870; Zenodo Dataset 8). We then generated maximum likelihood phylogenies with PhyML as described above (https://zenodo.org/record/4641870; Zenodo Dataset 9).

Statistical Analysis of Phage Phylogenies. As a test for superhost taxon specificity, we first tested whether phages from the same superhost taxon were more closely related to each other than to phages from other superhost taxa. To do so, we asked whether within-superhost taxa distances were lower than between-superhost taxa distances using categorical Mantel tests. We ran these tests separately for each of the phage phylogenies generated from the wild primate and human dataset, keeping only a single representative from each individual.

Based on evidence for widespread host specificity, we performed tests for codivergence by first downsampling to one representative per superhost taxa ($N_{replicates}$ = 1,000) for each phage phylogeny. As a test of superhost–phage codivergence, we then ran a ParaFit test (18) on each of these downsampled replicates, recording the proportion of significant tests ($P \leq 0.05$). The ParaFit test was implemented in the *ape* R package (49) with 1,000 permutations and the *cailliez* correction (18). We first ran this analysis on all wild primates and humans. We then repeated this analysis dropping all representatives from the phylogeny that were not from Catarrhini superhosts; small sample sizes precluded an examination at a finer taxonomic resolution.

To analyze the phages of the Amboseli baboons, we dropped all representatives from the wild primate and human phylogenies that were not from the baboons. We then compared the pairwise distance between sequences from group members and nongroup members using categorical Mantel tests. In addition, we compared the pairwise distance between sequences with the pairwise relatedness estimates of the baboons using Mantel tests.

Visualization of phylogenies were created using the ggtree package (50); host and phage phylogenies were handled in R with functions in the packages ape (49) and phytools (51).

Generation of Bacterial Community Matrix. Amplicon sequencing data were processed using the dada2 library for R following the recommendations of the developer (52), adjusting trimming parameters to 230 bp and 180 bp for the forward and reverse read, respectively, to fit the V1-V2 (27F-338R) 16S rRNA gene amplicon. Taxonomic annotations were performed using the Bayesian classifier and the Ribosomal Database Project database release version 16. To control for potential differences in sampling effort, we rarified to the smallest number of reads in a sample (9,247 reads) and then considered the presence–absence of particular bacterial oligotypes to create a dataset with a comparable resolution to the phage community composition data that we were able to generate. From the baboon dataset, we considered the presence or absence of bacterial species as estimated with MetaPhlAn 2.0 (elife-05224-supp2-v2; ref. 12); this represents a coarser taxonomic resolution than was possible for the oligotyping approach performed on great ape and human samples.

Statistical Analysis Examining the Relationship between Bacterial Communities and Phage Communities. Only a single sample per individual was included in these analyses. For specific taxonomic groups of samples, we compared the bacterial community similarity based on this presence–absence data (Sørensen's dissimilarity calculated with the vegdist function in the *vegan* R package) with the phage community similarity based on presence–absence data (also Sørensen's dissimilarity calculated with the vegdist function), using Mantel tests.

Assessing Comparability of Data. To assess the comparability of publicly available shotgun metagenomes and the reads generated as part of this study, which were generated in different laboratories and with samples stored in different ways, we first extracted a subset of reads (n = 5,000,000 read pairs) from each shotgun metagenome with seqtk (https://github.com/lh3/seqtk). We mapped these subsampled reads to the superhost's mitochondrial genome using BWA mem (min. seed length = 40) and sorted the mapping files with the SortSam tool in the Picard suite (18) retained reads with a MAPQ score >30 (46). We then calculated insert sizes with the function CollectInsertSizeMetrics in the Picard suite (18). We find no major differences in insert sizes based on the laboratory generating the reads or storage method ($\bar{x}_{Amato\ et\ al.,\ 2019\ ethanol}$ = 204.8 bp; $\bar{x}_{Amato\ et\ al.,\ 2019\ RNAlater}$ = 230.8 bp; $\bar{x}_{Kiel\ RNAlater}$ = 213.8 bp; $\bar{x}_{Kiel\ snap\ frozen}$ = 248.2 bp; $\bar{x}_{Tung\ et\ al.,\ 2015\ ethanol}$ = 237.7 bp; $\tilde{x}_{Amato\ et\ al.,\ 2019\ ethanol}$ = 186 bp; $\tilde{x}_{Amato\ et\ al.,\ 2019\ RNAlater}$ = 204 bp; $\tilde{x}_{Kiel\ RNAlater}$ = 172bp; $\tilde{x}_{Kiel\ snap\ frozen}$ = 218 bp; $\tilde{x}_{Tung\ et\ al.,\ 2015\ ethanol}$ = 239 bp; *SI Appendix*, Fig. S8).

Data Availability. The data supporting the conclusions of this article are available in the supporting information, with the exception of the larger files that are available through the Zenodo open-access repository (https://zenodo.org/record/4641870). Code for quality control and assembly of contigs is available here: https://github.com/mruehlemann/metagenome_preproc/blob/master/qc_and_assemble.slurm. All sequences generated as part of this study have been uploaded to the project accession number PRJNA692042. Data from previously published work are available through project accession number PRJNA271618 and ERP104379.

ACKNOWLEDGMENTS. We thank the IKMB Microbiome and NGS laboratories for excellent technical support. For fruitful discussion, we thank Thomas Briese, Adrian Caciula, Jonathan Davies, J. Peter Gogarten, Komal Jain, W. Ian Lipkin, Dorotta Nagy-Szakal, Luiz Thiberio Rangel, Brent Williams, and Simon Williams. We thank Maryke Gray, Freddy Makaya, Ulrich Bora Moussouami, Maude Pauly, Annika Starke, Erwan Theleste, Rodolphe Violleau, and the Zoo Hagenbeck (Hamburg), Zoo Leipzig, Zoo Gettorf, and Zoo Schwaigern for assistance with sample collection. This study was supported by the Deutsche Forschungsgemeinschaft (DFG) Research Group "Sociality and Health in Primates" (FOR2136; CA 1108/3-1) and the DFG Research Training Group 1743. It received infrastructure support from the Collaborative Research Center 1182, Origin and Function of Metaorganisms (https://www.metaorganism-research.com, no. SFB1182). J.F.G. was additionally supported by the DAAD with funds from the German Federal Ministry of Education and Research and the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA Grant Agreement No. 605728 (Postdoctoral Researchers International Mobility Experience). Research at Kokolopori was supported by the Max Planck Society and Harvard University. Research at Ozouga and Taï National Park was supported by the Max Planck Society. Research at Loango was supported by US Fish and Wildlife Service, Tusk Trust, Berggorilla Regenwald Direkthilfe, and the Max Planck Society. Research at

Bwindi was supported by the Max Planck Society. The survey of the Bwindi mountain gorillas was also funded from the International Gorilla Conservation Programme coalition members with supplemental funding from Wildlife Conservation Society. The microbiome and behavioral data from the Amboseli baboon population were generated by the Amboseli Baboon Research Project, with support from the NSF (IOS 1053461 to EAA and DEB 0919200 to S. Alberts) and the NIH (R01 AG-034513 and P01-AG 031719 to S. Alberts) and assistance in Kenya from the Kenya Wildlife Service, Institute of Primate Research, National Museums of Kenya, University of Nairobi, National Commission for Science, Technology, and Innovation, and National Environmental Management Authority (for full acknowledgements, please see: https://amboselibaboons.nd.edu/acknowledgements/). We are extremely grateful to the organizations that facilitated work at these field sites and granted permission for sample collection: Bwindi—The mountain gorilla survey was conducted by the Uganda Wildlife Authority, l'Institut Congolais pour la Conservation de la Nature, the Rwanda Development Board, the International Gorilla Conservation Programme, the Max Planck Institute for Evolutionary Anthropology, Conservation Through Public Health, the Mountain Gorilla Veterinary Project, the Institute for Tropical Forest Conservation, and The Dian Fossey Gorilla Fund and was conducted in compliance with the regulations of and permission of the Uganda National Council for Science and Technology and the Uganda Wildlife Authority; Kokolopori—permission was granted through the Ministere de Recherche Scientifique et Technologie, Democratic Republic of the Congo, and work was supported by the Vie Sauvage, the Bonobo Conservation Initiative; Loango—permission was granted by the Agence Nationale des Parcs Nationaux, the Centre National de la Recherche Scientifique et Technique of Gabon; Taï National Park—permission was granted by the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, the Ministère des Eaux et Fôrets in Côte d'Ivoire, and the Office Ivoirien des Parcs et Réserves, and work was supported by the Centre Suisse de Recherches Scientifiques en Côte d'Ivoire and the staff members of the Taï Chimpanzee Project. Ethical approval for work on human samples was obtained from the Local Ethics Committee Germany, Kiel (reference number A156/03), and we thank the national and local health authorities in Côte d'Ivoire and Democratic Republic of Congo as well as the according ethics commission for granting permission for this work.

1. M. K. Mirzaei, C. F. Maurice, Ménage à trois in the human gut: Interactions between host, bacteria and phages. *Nat. Rev. Microbiol.* **15**, 397–408 (2017).
2. J. J. Barr *et al.*, Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10771–10776 (2013).
3. Y. Ma, X. You, G. Mai, T. Tokuyasu, C. Liu, A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome* **6**, 24 (2018).
4. L. Gogokhia *et al.*, Expansion of bacteriophages is linked to aggravated intestinal inflammation and colitis. *Cell Host Microbe* **25**, 285–299.e8 (2019).
5. M. K. Mirzaei *et al.*, Bacteriophages isolated from stunted children can regulate gut bacterial communities in an age-specific manner. *Cell Host Microbe* **27**, 199–212.e5 (2020).
6. S. J. Ott *et al.*, Efficacy of sterile fecal filtrate transfer for treating patients with *Clostridium difficile* infection. *Gastroenterology* **152**, 799–811.e7 (2017).
7. R. A. Edwards *et al.*, Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. *Nat. Microbiol.* **4**, 1727–1736 (2019).
8. M. Groussin *et al.*, Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nat. Commun.* **8**, 14319 (2017).
9. A. H. Moeller *et al.*, Cospeciation of gut microbiota with hominids. *Science* **353**, 380–382 (2016).
10. H. Ochman *et al.*, Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biol.* **8**, e1000546 (2010).
11. C. P. van Schaik, P. M. Kappeler, Infanticide risk and the evolution of male-female association in primates. *Proc. Biol. Sci.* **264**, 1687–1694 (1997).
12. J. Tung *et al.*, Social networks predict gut microbiome composition in wild baboons. *elife* **4**, e05224 (2015).
13. J. F. Gogarten *et al.*, Factors influencing bacterial microbiome composition in a wild non-human primate community in Taï National Park, Côte d'Ivoire. *ISME J.* **12**, 2559–2574 (2018).
14. J. B. Clayton *et al.*, Captivity humanizes the primate microbiome. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10376–10381 (2016).
15. J. S. Frankel, E. K. Mallott, L. M. Hopper, S. R. Ross, K. R. Amato, The effect of captivity on the primate gut microbiome varies with host dietary niche. *Am. J. Primatol.* **81**, e23061 (2019).
16. P. Manrique *et al.*, Healthy human gut phageome. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10400–10405 (2016).
17. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
18. P. Legendre, Y. Desdevises, E. Bazin, A statistical test for host-parasite coevolution. *Syst. Biol.* **51**, 217–234 (2002).
19. E. Harrison, M. A. Brockhurst, Ecological and evolutionary benefits of temperate phage: What does or doesn't kill you makes you stronger. *BioEssays* **39**, 1700112 (2017).
20. A. J. Hockenberry, C. O. Wilke, BACPHLIP: Predicting bacteriophage lifestyle from conserved protein domains. *bioRxiv* [Preprint] (2020). 10.1101/2020.05.13.094805. Accessed 21 November 2020.
21. L. J. Funkhouser, S. R. Bordenstein, Mom knows best: The universality of maternal microbial transmission. *PLoS Biol.* **11**, e1001631 (2013).
22. E. K. Costello, K. Stagaman, L. Dethlefsen, B. J. Bohannan, D. A. Relman, The application of ecological theory toward an understanding of the human microbiome. *Science* **336**, 1255–1262 (2012).
23. A. C. Gregory *et al.*, The gut virome database reveals age-dependent patterns of virome diversity in the human gut. *Cell Host Microbe* **28**, 724–740.e8 (2020).
24. S. Roux, F. Enault, B. L. Hurwitz, M. B. Sullivan, VirSorter: Mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).
25. D. Amgarten, L. P. P. Braga, A. M. da Silva, J. C. Setubal, MARVEL, a tool for prediction of bacteriophage sequences in metagenomic bins. *Front. Genet.* **9**, 304 (2018).
26. J. Ren *et al.*, Identifying viruses from metagenomic data using deep learning. *Quant. Biol.* **8**, 64–77, https://doi.org/10.1007/s40484-019-0187-4 (2020).
27. A. N. Shkoporov *et al.*, The human gut virome is highly diverse, stable, and individual specific. *Cell Host Microbe* **26**, 527–541.e5 (2019).
28. A. H. Moeller, The shrinking human gut microbiome. *Curr. Opin. Microbiol.* **38**, 30–35 (2017).
29. E. C. Keen, G. Dantas, Close encounters of three kinds: Bacteriophages, commensal bacteria, and host immunity. *Trends Microbiol.* **26**, 943–954 (2018).
30. K. R. Amato *et al.*, Evolutionary trends in host physiology outweigh dietary niche in structuring primate gut microbiomes. *ISME J.* **13**, 576–587 (2019).
31. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal. Complex Syst.* **1695**, 1–9 (2006).
32. A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3747–3752 (2004).
33. C. Arnold, L. J. Matthews, C. L. Nunn, The 10kTrees website: A new online resource for primate phylogeny. *Evol. Anthropol.* **19**, 114–118 (2010).
34. A. L. Morales-Jimenez, T. Disotell, A. Di Fiore, Revisiting the phylogenetic relationships, biogeography, and taxonomy of spider monkeys (genus *Ateles*) in light of new molecular data. *Mol. Phylogenet. Evol.* **82**, 467–483 (2015).
35. J. Wang *et al.*, Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat. Genet.* **48**, 1396–1406 (2016).
36. B. Bushnell, BBMap. https://sourceforge.net/projects/bbmap/. Accessed 24 May 2020.
37. B. Bushnell, J. Rood, E. Singer, BBMerge–Accurate paired shotgun read merging via overlap. *PLoS One* **12**, e0185056 (2017).
38. S. Nurk, D. Meleshko, A. Korobeynikov, P. A. Pevzner, metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834, 10.1101/gr.213959.116 (2017).
39. M. Dowle, A. Srinivasan, Package 'Data. Table': Extension of 'Data. Frame (Version 1.12.6, R package, 2019). https://CRAN.R-project.org/package=data.table. Accessed 21 November 2020.
40. J. Oksanen *et al.*, vegan: Community Ecology Package (Version 2.5-6, R package, 2019). https://CRAN.R-project.org/package=vegan. Accessed 21 November 2020.
41. H. Wickham, *Ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
42. B. Auguie, GridExtra: Miscellaneous Functions for "Grid" Graphics (Version 2.3, R package, 2016). https://cran.r-project.org/web/packages/gridExtra. Accessed 21 November 2020.
43. M. Gottschling *et al.*, Quantifying the phylodynamic forces driving papillomavirus evolution. *Mol. Biol. Evol.* **28**, 2101–2113 (2011).
44. R Core Team, R: A Language and Environment for Statistical Computing, version 3.6.1. R Foundation for Statistical Computing, Vienna, Austria, 2019. https://www.R-project.org/. Accessed 21 November 2020.
45. Broad Institute, Picard. http://broadinstitute.github.io/picard. Accessed 24 March 2020.
46. H. Li *et al.*; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
47. S. Guindon *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
48. B. Q. Minh, M. A. T. Nguyen, A. von Haeseler, Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
49. E. Paradis, K. Schliep, Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
50. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Y. Lam, ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
51. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
52. B. J. Callahan *et al.*, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
53. K. Slowikowski, Ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2' (Version 0.8.2, R package, 2020). https://CRAN.R-project.org/package=ggrepel. Accessed 21 November 2020.
54. A. South, rnaturalearth: World Map Data from Natural Earth (Version 0.1.0, R package, 2017). https://CRAN.R-project.org/package=rnaturalearth. Accessed 21 November 2020.
55. E. Pebesma, Simple features for R: Standardized support for spatial vector data. *R J.* **10**, 439–446 (2018).