Predictable and host-species specific humanization of the gut microbiota in captive primates

Jennifer L. Houtz[1]†, Jon G. Sanders[1], Anthony Denice[2,3], Andrew H. Moeller[1]†

[1]Cornell University, Department of Ecology and Evolutionary Biology, Ithaca, New York 14850, USA

[2]Project Chimps, Blue Ridge, Georgia 30513, USA

[3]Chimpanzee Sanctuary Northwest, Cle Elum, Washington 98922, USA

†corresponding author: jlh498@cornell.edu, ahm226@cornell.edu

Abstract: Humans and non-human primates (NHPs) harbor complex gut microbial communities that affect phenotypes and fitness. The gut microbiotas of wild NHPs reflect their hosts' phylogenetic histories and are compositionally distinct from those of humans, but in captivity the endogenous gut microbial lineages of NHPs can be lost or replaced by lineages found in humans. Despite its potential contributions to gastrointestinal dysfunction, this humanization of the gut microbiota has not been investigated systematically across captive NHP species. Here we show through comparisons of well-sampled wild and captive populations of apes and monkeys that the fraction of the gut microbiota humanized by captivity varies significantly between NHP species but is remarkably reproducible between captive populations of the same NHP species. Conspecific captive populations displayed significantly greater than expected overlap in the sets of bacterial 16S rRNA gene variants that were differentially abundant between captivity and the wild. This overlap was evident even between captive populations residing on different continents but was never observed between heterospecific captive populations. In addition, we developed an approach incorporating human gut microbiota data to rank NHPs' gut microbial clades based on the propensity of their lineages to be lost or replaced by lineages found in humans in captivity. Relatively few microbial genera displayed reproducible degrees of humanization in different captive host species, but most microbial genera were reproducibly humanized or retained from the wild in conspecific pairs of captive populations. These results demonstrate that the gut microbiotas of captive NHPs display predictable, host-species specific responses to captivity.

Introduction

The gut microbial communities of humans and wild living primates reflect the evolutionary histories of their hosts (Ochman et al. 2010; Moeller et al. 2014; Amato et al. 2019). Each host species' gut microbiota contains a distinct set of microbial lineages, some of which have co-diversified with primate species (Moeller et al. 2016). However, gut microbial lineages that are confined to geographically separated wild populations of host species can be transmitted between host species when individuals come into direct contact. Similarly, shared environments can lead to the parallel gain or loss of lineages from the gut microbiota in co-occurring host populations, reducing differentiation of the gut microbiota. For example, studies of captive non-human primates (NHPs) have demonstrated that captivity leads to partial convergence of NHP gut microbiota with human gut microbiota (Clayton et al. 2016). In both apes (Uenishi et al. 2007; McKenzie et al. 2017; Frankel et al. 2019; Campbell et al. 2020; Narat et al. 2020) and monkeys (Nakamura et al. 2011; Clayton et al. 2016; McKenzie et al. 2017; Tsukayama et al. 2018; Frankel et al. 2019; Hale et al. 2019; Lee et al. 2019), the gut microbiotas of captive NHP populations are distinct from those of wild populations. In addition, captive the gut microbiotas of captive NHPs are often more compositionally similar to human gut microbiotas than are the gut microbiotas of wild living conspecific NHP populations (Clayton et al. 2016). The disruption and humanization of the endogenous gut microbiota in captive primates has been implicated in the gastrointestinal diseases often experienced by these populations (McKenna et al. 2008; Amato et al. 2016; Shigeno et al. 2018).

Despite the potential importance of humanization of the gut microbiota for the health of captive NHPs, this process, which can proceed through the gain of microbial lineages found in humans or the loss of gut microbiota constituents private to wild NHPs, has not been systematically evaluated across captive NHP populations. It is currently unclear whether humanization of the primate gut microbiota tends to be underlain by specific sets of microbial lineages or whether all lineages in the primate gut microbiota are equally prone to humanization. Similarly, the degree to which the probability of humanization of specific microbial lineages in captivity varies across primate populations and species is unknown.

One limitation of previous studies that has hindered the identification of specific microbial lineages that respond to captivity in NHP species is the lack of population-level host sampling. Several previous studies have examined the effects of captivity in the gut microbiota in multiple NHP species (Nakamura et al. 2011; McKenzie et al. 2017; Tsukayama et al. 2018; Frankel et al. 2019; Hale et al. 2019; Lee et al. 2019), but these have rarely examined more than ten individuals in captivity per NHP species. Sample size is often an insurmountable constraint given the limited sizes of captive NHP populations. Previous studies have had power to detect broad differences in gut microbiota composition between captive and wild populations (e.g., differences in microbiota alpha and beta diversity), but have typically not had power to test for effects of captivity on each individual microbial lineage in the gut microbiota. Another limitation of previous studies has been the lack of replicate captive NHP populations from the sample species, which are required in order to test the reproducibility and predictability of the effects of captivity on the gut microbiota of NHP species. One exception to these limitations is a previous study of wild and captive red-shanked doucs that sampled >30 captive individuals from two replicate captive populations, observing some evidence of reproducible effects of captivity (Clayton et al. 2016). Repeating such sampling regimes, in which relatively large numbers of individuals are sampled from multiple independent conspecific captive populations, in other NHP species promises to reveal whether NHP species display reproducible, NHP-species specific responses to captivity.

Here, we sequenced the gut microbiota of captive chimpanzees retired from the New Iberia Research Center and combined our data with gut microbiota datasets from humans, wild chimpanzees, and additional captive and wild populations of chimpanzees, gorillas, and red-shanked doucs. Each captive and wild population was represented by >15 individuals, enabling tests for differentially abundant microbial lineages between wild and captive individuals as well as microbial signatures of humanization within individual gut microbial clades. Results showed that gut microbial lineages were remarkably consistent in the degree to which they were humanized in replicate captive primate populations from the same species. However, the fraction of the gut microbiota that was humanized in captivity varied

significantly among primate species. These results indicate that the sets of microbial taxa that are humanized in the captive primate gut microbiota are predictable but dependent on host species identity.

Materials and Methods

Sample Collection and Processing

Fresh fecal samples were collected from 17 captive-born captive chimpanzees during their transport from the New Iberia Research Center, Louisiana, USA to Project Chimps Sanctuary, Georgia, USA in March and September of 2018. Samples were immediately stored at -20°C then transferred to -80°C within one month. DNA was extracted using DNeasy PowerSoil PowerLyzer DNA Extractions kits according to the manufacturer's protocol. Amplicons of the V4-V5 regions of the 16S rRNA gene were generated using 515F/926R primers based on the Earth Microbiome Project (EMP) 16S Illumina Amplicon protocol (Thompson et al. 2017). Amplicons were sequenced on an Illumina MiSeq platform using V3 chemistry and 2×250 bp paired-end reads at the Weill Cornell Medicine Microbiome Core Lab following methods of the EMP.

Combining NHP and human data for meta-analyses

Raw FASTQ sequence files and metadata from the captive chimpanzee samples were uploaded to the public database Qiita (Gonzalez et al. 2018) as study ID 12140 and processed using the database's implementation of the Quantitative Insights Into Microbial Ecology (QIIME2) software. We combined our data from captive chimpanzees (*Pan troglodytes subspp.*) retired from the New Iberia Research Center (USA2) with gut microbiota 16S rDNA datasets from wild chimpanzees (*Pan troglodytes schweinfurthii*) (n = 96) from Tanzania (Moeller et al. 2016b) (TZA), wild chimpanzees (*P. t. troglodytes*) (n = 18) and gorillas (*Gorilla gorilla gorilla*) (n = 28) from the Democratic Republic of the Congo (DRC), captive chimpanzees (*P. t. subspp.*) (n = 18) and gorillas (*G. g. subspp.*) (n = 15) from the United States (USA1) (Campbell et al. 2020), wild red-shanked doucs (*Pygathrix nemaeus*) from Vietnam (VNM) (n = 66), and captive red-shanked doucs (*P. n.*) from Singapore (SGP) (n = 15) and the United States (USA) (n = 12) (Clayton et al., 2016). This combined wild and captive NHP gut microbiota dataset was

then merged with a dataset from humans (n = 528) living in the USA (Yatsunenko et al. 2012). Forward and reverse sequences were demultiplexed and quality filtered using default settings. High-quality reads were trimmed to equal length and amplicon sequence variants (ASVs) were identified using Deblur. Data were rarefied to 10,000 sequences per sample and exported from Qiita for analyses in QIIME2 v 2018.11 (Boylen et al. 2019).

Taxonomy and diversity analyses

Taxonomy was assigned to ASVs by fitting a naive-Bayes classifier trained on the Greengenes v.13_8 99% OTUs reference using the sk-learn classifier. A phylogenetic tree was constructed by inserting the sequences into the Greengenes 13_8 reference tree using the QIIME2 plugin q2-fragment-insertion, which uses the SEPP insertion method. Shannon entropy and Chao1 alpha diversity indices were calculated for each sample. The Shannon index measures richness and evenness, whereas Chao1 measures the abundance weighted ASV richness within a microbiota. To examine the sharing of ASVs between captive and wild NHPs and humans, the binary Sorensen-Dice dissimilarly was calculated for each pair of samples (Dice 1945). Binary Sorensen-Dice dissimilarities were calculated as one minus the quotient of the total number of ASVs shared between two samples by the sum of the total number of ASVs in each sample.

Statistical analyses of alpha and beta diversity

Alpha diversity metrics including Shannon index and Chao1 were compared using Kruskal-Wallis tests. For Sorensen-Dice beta diversity, permutational multivariate analysis of variation (PERMANOVA) and Monte-Carlo nonparametric permutation tests were conducted in QIIME to statistically partition the sources of variation in community structure using 999 permutations. Sorensen-Dice dissimilarities among all samples were visualized through ordination using principal coordinates analysis (PCoA) and plots in EMPeror. Pearson correlations and Spearman ranked correlations among log-transformed HSSs and MCSs were conducted with the “stats” package in R v 4.0.4 (R Core Team 2021).

Detection of differentially abundant microbial ASVs and genera between captive and wild populations

ANCOM2 (Mandal et al. 2015) was employed in R (github.com/FrederickHuangLin/ANCOM) to test whether any individual microbial ASVs or genera were differentially abundant between the captive and wild individuals for every captive NHP population. Each analysis focused on ASVs or genera present in >10% of samples from the pair of captive and wild populations. Volcano plots displaying differentially abundant ASVs and genera were created in R with ggplot2. For these analyses, captive chimpanzees retired from the New Iberia Research Center (USA2) were paired with wild chimpanzees from Tanzania (TZA) and zoo captive chimpanzees (USA1) were paired with wild chimpanzees from the DRC (DRC). Pairings were chosen such that samples from wild and captive conspecific populations were processed and sequenced by the same laboratories, minimizing any potential study effects that could confound downstream analyses.

In addition, we tested whether microbial ASVs and genera differed repeatedly in relative abundance between the wild and captivity in multiple captive NHP populations. Specifically, we employed hypergeometric tests to assess the significance of the overlap between the sets of microbial taxa (i.e., ASVs and genera) whose abundances differ from the wild in pairs of captive NHP populations. These analyses tested whether the number of microbial taxa that shifted in relative abundance in the two captive NHP populations relative to wild conspecific NHP populations was significantly greater than the number expected by chance, given the observed number of microbial taxa shared between the multiple replicate captive NHP populations. In addition, we employed regression analyses to test for significant associations of W statistics between pairs of captive NHP populations.

Host Specificity Scores and Microbiota Convergence Scores

We developed a statistic, the Host Specificity Score (HSS), to assess microbial genus-specific patterns of humanization in captive NHP gut microbiotas. This statistic was defined for a microbial genus in a captive NHP population as the mean binary Sorensen-Dice dissimilarity of the ASV composition of the genus between the captive NHP population and humans divided by the mean binary Sorensen-Dice dissimilarity of the ASV composition of the genus between the captive NHP population and a conspecific wild-living population. HSSs of >1 indicate that the ASV composition of the microbial clade is more

similar between captive NHPs and wild-living conspecifics than between captive NHPs and humans. In contrast, a score of <1 indicates that the ASV composition of the microbial clade is more similar between captive NHPs and humans than between captive NHPs and wild-living conspecifics.

In order to calculate HSSs, we split the ASV table containing all samples by microbial genus and calculated genus-level mean binary Sorensen-Dice (BSD) dissimilarities between captive NHPs and wild conspecifics as well as between captive NHPs and humans. ASVs without taxonomic assignments and microbial genera detected by fewer than 5 reads were excluded from downstream analyses. As in the ANCOM analyses described above, captive chimpanzees retired from the New Iberia Research Center (USA2) were paired with wild chimpanzees from Tanzania (TZA) and zoo captive chimpanzees (USA1) were paired with wild chimpanzees from the DRC (DRC). The HSS for each genus was calculated and log-transformed for regression analyses.

In addition to HSSs, we also calculated for each microbial genus in each captive NHP population a Microbiota Convergence Score (MCS). This index measures the degree to which the ASV composition of the genus in a captive NHP has converged with the ASV composition of the genus in humans relative to the ASV composition of the genus in the wild conspecific NHP population. Descriptions of these analyses are presented in the Supplemental Materials and Methods. For both HSSs and MCSs, heatmaps were calculated in R using heatmap.2 with the parameter scale="col" and default settings. In addition to analyses based on the Yatsunenko et al. (2012) dataset, we also calculated per genus HSSs and MCSs for each captive NHP population using an American Gut dataset (McDonald et al., 2018) (Supplementary Materials and Methods). All scripts used to calculate HSSs and MCSs are publicly available at github.com/CUMoellerLab/Houtz_etal_2021-NHP_microbiome.

Results

Sequencing and comparisons of NHP and human gut microbiomes

We sequenced 16S rDNA extracted from 18 fecal samples collected from captive chimpanzees upon their arrival at Project Chimps sanctuary, generating 1,406,091 million high-quality sequences with an average

of 80,704 reads per sample. These data were combined with published V4-V5 16S rDNA amplicon datasets containing matched populations of captive and wild primates from the same species, including doucs, chimpanzees, and gorillas as well as 16S rDNA datasets collected from human populations in the United States, the country in which most captive NHP individuals represented in the combined dataset resided. One fecal sample from a captive chimpanzee (chimp.17) dominated by a single ASV (>50% of reads) was removed from downstream analyses. The total dataset contained 796 samples rarefied to an even depth of 10,000 reads per sample. A list of samples and their corresponding metadata are presented in Table S1. Sample metadata corresponding to published datasets are available in the supplementary materials of their respective studies (Clayton et al., 2016; Moeller et al., 2016; Campbell et al., 2020). A map of sampling locations for wild and captive NHPs is presented in Figure S1.

Altered microbiota diversity in captive NHPs relative to wild conspecific NHPs

We tested for effects of host species identity and captivity state on microbiota alpha and beta diversity using 16S rDNA sequences from captive chimpanzee samples generated by the present study (Captive chimpanzees USA2) and published data from humans (Humans), wild chimpanzees (Wild chimpanzees TZA; Wild chimpanzees DRC), and captive and wild red-shanked doucs (Captive douc SGP, Captive douc USA, Wild douc VNM). Captive populations tended to display reduced Shannon entropy relative to wild conspecifics (Figure S2; Table S2). However, results from Chao1 analyses were mixed, with captive doucs displaying lower alpha diversity than wild conspecifics but captive chimpanzees and gorillas displaying higher alpha diversity than wild conspecifics (Figure S2; Table S2).

Analyses of beta diversity indicated that the gut microbiotas of captive primate populations were compositionally distinct from the gut microbiotas of wild conspecifics (PERMANOVA p-value = 0.001 in each comparison). This distinctiveness of captive and wild NHP gut microbiotas was evident in principal coordinates analyses based on Sorensen-Dice distances (Figure 1A). Moreover, for each NHP species the gut microbiotas of captive populations were significantly more similar to those of humans than were the gut microbiotas of wild-living conspecific populations (Figure 1B) (Monte-Carlo nonparametric permutation tests p-value = 0.001). The relative abundances of ASVs for each sample are

220 presented in Table S3. The relative abundances of microbial genera for each sample are presented in
221 Table S4.
222 *No ASVs or genera were differentially abundant between captive and wild in all NHP species*
223 We next employed ANCOM to test for ASVs that were differentially abundant between captive and wild
224 individuals for each captive NHP population. These analyses indicated that hundreds of ASVs were
225 differentially abundant between captive and wild individuals in at least one captive NHP population
226 (Figure S3, Table S5). However, of the 262 ASVs detected in >2 of the 6 captive NHP populations, none
227 were differentially abundant between captive and wild individuals in >2 of the 6 captive NHP populations
228 (Table S5). Similarly, dozens of microbial genera were differentially abundant between captive and wild
229 individuals in at least on captive NHP population (Figure 2, Table S6), but none were differentially
230 abundant between captive and wild individuals in >3 of the 6 captive NHP populations (Table S6). This
231 observation indicates that the individual ASVs and genera that were overrepresented or underrepresented
232 in captive NHP population relative to wild conspecific populations varied across NHP species.
233 *Reproducible shifts of microbial ASVs and genera to captivity in replicate conspecific NHP populations*
234 We observed significant overlap in the microbial ASVs and genera that displayed significant shifts in
235 relative abundance from the wild, based on ANCOM analyses, between replicate captive NHP
236 populations of the same species (i.e., conspecific populations) (Table S5). Of the 627 ASVs detected in
237 both replicate captive populations of chimpanzees (USA1 and USA2), 102 and 153 were significantly
238 overrepresented or underrepresented in USA1 and USA2 chimpanzees, respectively, relative to their
239 matched wild populations (each captive chimpanzee population was matched with a different wild
240 chimpanzee population, Materials and Methods). Of these ASVs displaying significantly different relative
241 abundances between captive and wild chimpanzees, 53 displayed significant differences in both captive
242 populations of chimpanzees (USA1 and USA2) (2.1-fold more than random expectation, hypergeometric
243 *p*-value = 1.429e-11). Similarly, of the 573 ASVs detected in both replicate captive populations of doucs
244 (USA and SGP), 61 and 66 were significantly overrepresented or underrepresented in USA and SGP
245 doucs, respectively, relative to their matched wild population. Of these ASVs displaying significantly

different relative abundances between captive and wild doucs, 53 displayed significant differences in both captive populations of doucs (USA and SGP) (6.8-fold more than random expectation, hypergeometric *p*-value = 5.43e-43). In contrast, this degree of overlap between ASVs displaying significantly different abundances between captivity and the wild was not observed between heterospecific captive NHP populations. For example, of the 64 ASVs shared by USA1 chimpanzees and SGP doucs, 10 and 15 were significantly overrepresented or underrepresented in chimpanzees and doucs, respectively, relative to their matched wild populations. Of these ASVs, only 3 displayed significant differences in both USA1 chimpanzees and SGP doucs (1.3-fold more than expected, hypergeometric *p*-value = 0.215).

We also observed significant overlap between the sets of microbial genera that were differentially abundant between captivity and the wild in replicate captive conspecific NHP populations (Figure 2; Table S6). Of the 126 microbial genera detected in both replicate captive populations of chimpanzees (USA1 and USA2), 23 and 22 were significantly overrepresented or underrepresented in USA1 and USA2 chimpanzees, respectively, relative to their matched wild populations. Of these genera, 7 displayed significant differences in both captive populations of chimpanzees (USA1 and USA2) (4.2-fold more than random expectation, hypergeometric *p*-value = 0.0355) (Figure 2F). Similarly, of the 82 genera detected in both replicate captive populations of doucs (USA and SGP), 16 and 12 were significantly overrepresented or underrepresented in USA and SGP doucs, respectively, relative to their matched wild population. Of these genera, 5 displayed significant differences in both captive populations of doucs (USA and SGP) (6.8-fold more than random expectation, hypergeometric *p*-value = 5.95e-5) (Figure 2F). In contrast, this degree of overlap between genera displaying significantly different abundances between captivity and the wild was not observed between heterospecific captive NHP populations. For example, of the 67 genera shared by USA1 chimpanzees and SGP doucs, 15 and 11 were significantly overrepresented or underrepresented in chimpanzees and doucs, respectively, relative to their matched wild populations. Of these genera, only 3 displayed significant differences in both USA1 chimpanzees and SGP doucs (1.2-fold more than expected, hypergeometric *p*-value = 0.234) (Figure 2G).

In addition to hypergeometric tests for significant overlap of ANCOM results between pairs of captive NHP populations, we also conducted regression analyses to test whether W statistics of ASVs and genera we associated between pairs of captive NHP populations. These analyses revealed qualitatively similar results to the hypergeometric tests. Specifically, ASV and genus W statistics tended to display significant positive relationships between pairs of conspecific NHP populations, but less significant and more variable relationships between pairs of heterospecific NHP populations (Table S5, Table S6).

HSSs reveal reproducible and host-species specific humanization of microbial genera in captive NHPs

Identifying differentially abundant ASVs and genera between matched wild and captive populations can reveal the effects of captivity on specific constituents of the NHP gut microbiota. However, these analyses alone do not provide information about the extent to which specific microbial clades within the captive NHP gut microbiota are humanized. Here, we define humanization of a microbial clade in a captive NHP population as the replacement of endogenous ASVs in the clade (i.e., ASVs found in wild NHP populations) by ASVs found in humans or the parallel absence from captive NHP populations and humans of ASVs in the clade found in wild conspecific NHP populations.

To assess the degree of humanization for each gut microbial clade in each captive NHP population, we tested whether the ASV-compositions of individual microbial genera in captive NHPs were more similar to those of the genera in humans (i.e., humanized) or to those of the genera in wild living conspecific individuals. These analyses allowed us to identify microbial genera in each captive NHP population whose endogenous host-species specific ASVs (i.e., ASVs private to wild conspecific hosts) displayed evidence of loss or of replacement by ASVs found in humans. We developed a statistic, Host Specificity Score (HSSs) (Materials and Methods), which can be calculated for any microbial clade or taxonomic rank of arbitrary phylogenetic depth. Here, we calculated these scores at the level of microbial genus, which was the finest-scale taxonomic resolution allowed by the 16S rDNA amplicon sequences in our dataset. Log-transformed HSSs are normally distributed around 0 and provide a quantitative ranking of the degree of humanization for all microbial genera in the gut microbiotas of a captive population, with lower scores indicating a greater degree of humanization. Log-transformed HSSs

297 >0 indicate the ASV composition of the genus is more similar between captive primates and wild-living
298 conspecifics than between captive primates and humans. Conversely, log-transformed HSSs <0 indicate
299 that the ASV composition of the genus is more similar between captive primates and humans than
300 between captive primates and wild-living conspecifics. Humanization of the ASV composition of a
301 microbial genus in a captive NHP population, as indicated by a negative HSS, could be underlain by the
302 loss of ASVs found in wild conspecific NHPs but not found in humans or the gain of ASVs found in
303 humans but not found in wild conspecific NHPs.

304 Calculating HSSs for each microbial genus for each captive NHP population indicated that some
305 genera in the captive NHP gut microbiota were humanized, whereas others were retained from the wild.
306 An example of a bacterial genus that was humanized in captive chimpanzees in the United States (i.e., the
307 genus *Collinsella*) is shown in Figure 3A. An example of an archaeal genus that was retained from the
308 wild in captive chimpanzees in the United States (i.e., the genus *vadinCA11*) is shown in Figure 3B. A list
309 of all microbial genus-level HSSs for each captive NHP population are shown in Table S7.

310 Comparing HSSs between pairs of captive primate populations indicated that no microbial genus
311 was humanized in every NHP population (Table S7). However, replicate captive populations of the same
312 species displayed highly consistent genus-level patterns of gut microbiota humanization. The HSSs
313 between two independent captive populations of chimpanzees were significant positively associated (R^2 =
314 0.82; *p*-value = 5.51e-24), as were the HSSs between two independent captive populations of doucs (R^2 =
315 0.8; *p*-value = 8.46e-18). In addition, the HSS of chimpanzees were significantly associated with those of
316 gorillas (R^2 = 0.16; *p*-value = 0.000871 and R^2 = 0.25; *p*-value = 1.05e-05), but not with those of doucs
317 (Figure 4). Similar results, in which no genus displayed evidence of humanization in all captive NHP
318 populations but replicate captive populations of the same NHP species displayed similar patterns of
319 humanization of microbial genera, were observed in analyses of a related statistic, Microbiota
320 Convergence Scores (MCSs) (Supplementary Materials and Methods, Figure S4, Table S8). Heatmaps of
321 HSSs and MCSs of taxa found in all captive NHP populations are presented in Figure S5. Qualitatively
322 similar results, in which HSSs and MCSs were significantly positively associated between conspecific

captive NHP populations but tended to be not associated or negatively associated between heterospecific captive NHP populations, were observed in analyses based on data from the American Gut Project (McDonald et al., 2018) (Supplementary Materials and Methods; Table S9 and S10).

Discussion

Comparing multiple well-sampled wild and captive populations of chimpanzees, gorillas, and doucs with humans allowed us to test for reproducible effects of captivity on specific microbial lineages in the gut microbiotas of non-human primate (NHP) species. Our results indicate that captivity humanizes the primate gut microbiota, but that the microbial taxa underlying this process of humanization vary substantially and consistently among NHP species. No microbial ASV or genus was significantly overrepresented or underrepresented in all captive NHP populations relative to wild conspecific populations (Tables S5, Table S6). Similarly, no microbial genus displayed evidence of humanization, defined by Host-Specificity Scores (HSSs), in every captive NHP population (Table S7). In contrast, we observed striking consistency of the effects of captivity on gut microbiota constituents between replicate captive NHP populations of the sample species. The same ASVs and genera reproducibly shifted in relative abundance in replicate captive conspecific NHP populations relative to matched wild NHP populations (Table S5, Table S6, Figure 2). Similarly, most microbial genera displayed reproducible signatures of humanization in replicate captive conspecific NHP populations: the HSSs of one captive population predicted >80% of the variation in HSSs in the other captive population in both NHP species for which replicate captive populations have been sampled (i.e., chimpanzees and doucs) (Figure 4, Figure S4). The consistency of per-genus measures of humanization in replicate conspecific captive NHP populations was also observed in analyses based on human gut microbiota data from the American Gut Project (Supplementary Materials and Methods; Table S9-S10). Previous results have shown that the gut microbiotas of monkeys are humanized in captivity (Clayton et al., 2016); our results demonstrate that this humanization also occurs in captive ape populations. In addition, our results indicate that the humanization of the gut microbiota is underlain by distinct sets of microbial lineages in captive NHP

349 populations of different species, but that the sets of microbial lineages that are humanized by captivity can
350 be predicted with high accuracy by the species identity of the NHP population.
351 Interestingly, certain microbial taxa showed consistent patterns of humanization between
352 replicate captive NHP populations of the same species but opposite patterns between replicate captive
353 NHP populations of different species. For example, *Oribacterium* and *Roseburia*, genera within the
354 Lachnospiraceae, showed evidence of humanization, based on HSSs, in both captive douc populations,
355 but these genera were robust to humanization in both captive chimpanzee populations (Table S8, Figure
356 S5). The observations that different NHP species display discordant but reproducible patterns of
357 humanization of gut microbial lineages in captivity raises questions about the mechanisms underlying this
358 pattern. One explanation is that each NHP species contains a distinct set of microbial lineages that are
359 better adapted to the gut environment of their respective host species relative to the congeneric microbial
360 lineages found in humans. For example, microbial genera whose ASVs were retained from the wild in
361 captive NHP populations (i.e., genera displaying a log-transformed HSS > 0) represent microbial lineages
362 that may, in captive NHPs, outcompete microbial lineages found in humans. Discordance of per-genus
363 HSSs between NHP species could arise if the microbial lineages that outcompete congeneric lineages
364 derived from humans belong to different genera in different NHP species. In addition, differences in diet
365 may affect the process of gut microbiota humanization in captivity, leading to discordance between
366 species but reproducibility within species. For example, previous work has shown that global patterns of
367 gut microbiota composition (i.e., alpha and beta diversity) are more affected by captivity in folivorous
368 NHP species than in less dietary specialized NHP species (Frankel et al., 2019). Under this scenario,
369 differences in the degree of dietary shifts experienced in captivity could lead to different effects on
370 individual microbial lineages and genus-level patterns of humanization between NHP species.
371 Our results highlight the importance of population-level sampling of captive NHPs in order to
372 identify specific gut microbial lineages most affected by captivity. Studies of fewer numbers of captive
373 individuals per NHP species have statistical power to detect differences in alpha and beta diversity
374 between captive and wild gut microbiotas and have yielded important insights into the effects of captivity

on the gut microbiota (Uenishi et al. 2007; Nakamura et al. 2011; Clayton et al. 2016; McKenzie et al. 2017; McKenzie et al. 2017; Tsukayama et al. 2018; Frankel et al. 2019; Hale et al. 2019; Lee et al. 2019; Frankel et al. 2019; Campbell et al. 2020; Narat et al. 2020). However, identifying individual microbial lineages or clades (e.g., ASVs or genera) that display significantly different relative abundances between populations requires sampling sufficient numbers of individuals to overcome high false-discovery rates inherent in multiple testing. Our results suggest that future studies focused on identifying specific microbial lineages in the gut microbiota that are affected by captivity should when possible prioritize replication at the level of host individual.

The statistics that we developed (HSSs and MCSs) provide quantitative means to identify the specific clades of microbial lineages that display the strongest evidence of humanization in captive hosts. HSSs and MCSs can be applied to any microbial clade or taxonomic rank as well as any captive population for which microbiota data from wild conspecifics are available. In addition to captive populations, these statistics could also be applied to identify humanized gut microbial taxa in animals associated with humans in other contexts, such as urban settings or other habitat disruptions that may affect the gut microbiota (e.g., Amato et al. 2013). In the case of captive NHPs, the identification of individual microbial taxa that are repeatably humanized has important implications for managing the health of captive hosts. Humanized microbial taxa represent candidate contributors to gastrointestinal dysbiosis in captive NHP populations, in particular if they are reproducibly humanized in individuals that develop gastrointestinal dysfunction (Amato et al. 2016). For example, the genus *Collinsella* was repeatably humanized in replicate captive chimpanzee populations, and overrepresentation of this genus in the gut microbiota has been associated with reduced fiber intake, insulin resistance, and rheumatoid arthritis in human populations (Chen et al. 2016; Gomez-Arango et al., 2018; Mena-Vázquez et al. 2020). Similarly, *Clostridium* lineages were repeatably humanized in replicate captive chimpanzee populations, and lineages from this genus are the most common causes of gastrointestinal infections in humans (Smits et al. 2016). The gut microbial lineages that are reproducibly humanized in captive populations represent high priority for targeted culturing from wild NHP populations. Culturing and biobanking representatives

401 of these microbial lineages from wild NHPs could provide opportunities to restore the gut microbiotas of
402 captive NHP populations.
403

Acknowledgements: We gratefully acknowledge Dr. Weiwei Yan for assistance in the laboratory. This work was funded by a grant from the National Institutes of Health (1R35GM138284) to AHM. JLH was supported by a Cornell Presidential Life Sciences Fellowship and National Science Foundation Graduate Research Fellowship. All procedures conducted in this study were approved by the Institutional Animal Care and Use Committee at Cornell University. We declare we have no competing interests.

Data accessibility: All sequence data are available in the European Nucleotide Archive under accession PRJEB44013.

Authors' contributions: JLH carried out study design, conducted data analysis, and drafted the manuscript; JGS conducted data analysis and critically revised the manuscript; AD collected field data and critically revised the manuscript; AHM carried out study design, conducted data analysis, and drafted the manuscript. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

References

Amato, K. R., Metcalf, J. L., Song, S. J., Hale, V. L., Clayton, J., Ackermann, G., Humphrey, G., Niu, K., Cui, D., Zhao, H., and Schrenzel, M. D. (2016). Using the gut microbiota as a novel tool for examining colobine primate GI health. *Global Ecology and Conservation*, *7*, 225-237.

Amato, K. R., Sanders, J. G., Song, S. J., Nute, M., Metcalf, J. L., Thompson, L. R., Morton, J. T., Amir, A., McKenzie, V. J., Humphrey, G., and Gogul, G. (2019). Evolutionary trends in host physiology outweigh dietary niche in structuring primate gut microbiomes. *The ISME Journal*, *13*(3), 576-587.

Amato, K. R., Yeoman, C. J., Kent, A., Righini, N., Carbonero, F., Estrada, A., Gaskins, H. R., Stumpf, R. M., Yildirim, S., Torralba, M., and Gillis, M. (2013). Habitat degradation impacts black howler monkey (*Alouatta pigra*) gastrointestinal microbiomes. *The ISME journal*, *7*(7), 1344-1353.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., and Bai, Y. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852-857.

Campbell, T. P., Sun, X., Patel, V. H., Sanz, C., Morgan, D., and Dantas, G. (2020). The microbiome and resistome of chimpanzees, gorillas, and humans across host lifestyle and geography. *The ISME Journal*, *14*(6), 1584-1599.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, 265-270.

Chen, J., Wright, K., Davis, J. M., Jeraldo, P., Marietta, E. V., Murray, J., Nelson, H., Matteson, E. L., and Taneja, V. (2016). An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. *Genome Medicine*, *8*(1), 1-14.

Clayton, J. B., Vangay, P., Huang, H. U., Ward, T., Hillmann, B. M., Al-Ghalith, G. A., Travis, D. A., Long, H. T., Van Tuan, B., Van Minh, V., and Cabana, F. (2016). Captivity humanizes the primate microbiome. *Proceedings of the National Academy of Sciences*, *113*(37), 10376-10381.

Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, *26*(3), 297-302.

Frankel, J. S., Mallott, E. K., Hopper, L. M., Ross, S. R., and Amato, K. R. (2019). The effect of captivity on the primate gut microbiome varies with host dietary niche. *American Journal of Primatology*, *81*(12), e23061.

Gonzalez, A., Navas-Molina, J. A., Kosciolek, T., McDonald, D., Vázquez-Baeza, Y., Ackermann, G., DeReus, J., Janssen, S., Swafford, A. D., Orchanian, S. B., and Sanders, J. G. (2018). Qiita: rapid, web-enabled microbiome meta-analysis. *Nature Methods*, *15*(10), 796-798.

Gomez-Arango, L. F., Barrett, H. L., Wilkinson, S. A., Callaway, L. K., McIntyre, H. D., Morrison, M., and Nietert, M. D. (2018). Low dietary fiber intake increases *Collinsella* abundance in the gut microbiota of overweight and obese pregnant women. *Gut Microbes*, *9*(3), 189-201.

Hale, V. L., Tan, C. L., Niu, K., Yang, Y., Zhang, Q., Knight, R., and Amato, K. R. (2019). Gut microbiota in wild and captive Guizhou snub−nosed monkeys, *Rhinopithecus brelichi*. *American Journal of Primatology*, *81*(10-11), e22989.

Lee, W., Hayakawa, T., Kiyono, M., Yamabata, N., and Hanya, G. (2019). Gut microbiota composition of Japanese macaques associates with extent of human encroachment. *American Journal of Primatology*, *81*(12), e23072.

Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., and Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health and Disease*, *26*(1), 27663.

McDonald D., Hyde E., Debelius J. W., Morton J. T., Gonzalez A., Ackermann G., Aksenov A. A., Behsaz B., Brennan C., Chen Y., Goldasich L. D. (2018). American gut: an open platform for citizen science microbiome research. mSystems. 3(3), e00031-18.

McKenna, P., Hoffmann, C., Minkah, N., Aye, P. P., Lackner, A., Liu, Z., Lozupone, C. A., Hamady, M., Knight, R., and Bushman, F. D. (2008). The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. *PLoS Pathog*, *4*(2), e20.

McKenzie, V. J., Song, S. J., Delsuc, F., Prest, T. L., Oliverio, A. M., Korpita, T. M., Alexiev, A., Amato, K. R., Metcalf, J. L., Kowalewski, M., and Avenant, N. L. (2017). The effects of captivity on the mammalian gut microbiome. *Integrative and Comparative Biology*, *57*(4), 690-704.

Mena-Vázquez, N., Ruiz-Limón, P., Moreno-Indias, I., Manrique-Arija, S., Tinahones, F. J., and Fernández-Nebro, A. (2020). Expansion of Rare and Harmful Lineages is Associated with Established Rheumatoid Arthritis. *Journal of Clinical Medicine*, *9*(4), 1044.

Moeller, A. H., Caro-Quintero, A., Mjungu, D., Georgiev, A. V., Lonsdorf, E. V., Muller, M. N., Pusey, A. E., Peeters, M., Hahn, B. H., and Ochman, H. (2016a). Cospeciation of gut microbiota with hominids. *Science*, *353*(6297), 380-382.

Moeller, A. H., Foerster, S., Wilson, M. L., Pusey, A. E., Hahn, B. H., and Ochman, H. (2016b). Social behavior shapes the chimpanzee pan-microbiome. *Science Advances*, *2*(1), e1500997.

Moeller, A. H., Li, Y., Ngole, E. M., Ahuka-Mundeke, S., Lonsdorf, E. V., Pusey, A. E., Peeters, M., Hahn, B. H., and Ochman, H. (2014). Rapid changes in the gut microbiome during human evolution. *Proceedings of the National Academy of Sciences*, *111*(46), 16431-16435.

Nakamura, N., Amato, K. R., Garber, P., Estrada, A., Mackie, R. I., and Gaskins, H. R. (2011). Analysis of the hydrogenotrophic microbiota of wild and captive black howler monkeys (*Alouatta pigra*) in palenque national park, Mexico. *American Journal of Primatology*, *73*(9), 909-919.

Narat, V., Amato, K. R., Ranger, N., Salmona, M., Mercier-Delarue, S., Rupp, S., Ambata, P., Njouom, R., Simon, F., Giles-Vernick, T., and LeGoff, J. (2020). A multi-disciplinary comparison of great ape gut microbiota in a central African forest and European zoo. *Scientific Reports*, *10*(1), 1-15.

Ochman, H., Worobey, M., Kuo, C. H., Ndjango, J. B. N., Peeters, M., Hahn, B. H., and Hugenholtz, P. (2010). Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biology*, *8*(11), e1000546.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Shannon, C. E., and Weaver, W. (1949). "The mathematical theory of communication". University of Illinois Press, Champaign, Illinois.

Shigeno, Y., Toyama, M., Nakamura, M., Niimi, K., Takahashi, E., and Benno, Y. (2018). Comparison of gut microbiota composition between laboratory−bred marmosets (*Callithrix jacchus*) with chronic diarrhea and healthy animals using terminal restriction fragment length polymorphism analysis. *Microbiology and Immunology*, *62*(11), 702-710.

Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H., and Kuijper, E.J. (2016). *Clostridium difficile* infection. *Nature Reviews Disease Primers*, *2*(1), 1-20.

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., Prill, R. J., Tripathi, A., Gibbons, S. M., Ackermann, G., and Navas-Molina, J. A. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, *551*(7681), 457-463.

Tsukayama, P., Boolchandani, M., Patel, S., Pehrsson, E. C., Gibson, M. K., Chiou, K. L., Jolly, C. J., Rogers, J., Phillips-Conroy, J. E., and Dantas, G. (2018). Characterization of wild and captive baboon gut microbiota and their antibiotic resistomes. *Msystems*, *3*(3).

Uenishi, G., Fujita, S., Ohashi, G., Kato, A., Yamauchi, S., Matsuzawa, T., and Ushida, K. (2007). Molecular analyses of the intestinal microbiota of chimpanzees in the wild and in captivity. *American Journal of Primatology*, *69*(4), 367-376.

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., and Heath, A. C. (2012). Human gut microbiome viewed across age and geography. *Nature*, *486*(7402), 222-227.

[dataset] Houtz, J., Sanders J., Denice A., Moeller A.H. ; 2021; Predictable and host-species specific humanization of the gut microbiota in captive primates; European Nucleotide Archive PRJEB44013; qiitaID:12140

Figure 1. Humanization of captive NHP gut microbiotas. (A) Plot shows the first and second Principal Coordinates of the Binary Sorensen-Dice dissimilarities among wild primates, captive primates, and humans. Shapes represent fecal samples collected from individual primates and are colored based on host species identity as indicated by the inset. Circles unite groups of shapes corresponding to wild, captive, or human samples. (B) Boxplots show maxima, minima, interquartile ranges, and medians of BSD dissimilarities between primates (wild or captive) and humans. Boxes are colored by primate species as in (A). Significant differences between means based on permutation tests are indicated by asterisks; *** p = 0.001.

Figure 2. Differential abundance of microbial genera between captive and wild NHP individuals. Volcano plots show the ANCOM test statistic (y-axis) and the centered logarithmic transformed (CLR) log-fold difference in relative abundance between captive and wild individuals for each microbial genus. Dots represent microbial genera colored by captive NHP population based on Figure 1, and each plot shows results for a single captive primate population: captive chimpanzee (USA2) (A), captive chimpanzee (USA1) (B), captive gorilla (C), captive douc (USA) (D), and captive douc (SGP) (E). Dashed line in each plot represents W statistic cutoff of 0.9. Statistics for all ASVs for each pair of captive and wild NHP populations are presented in Table S3. Venn diagrams (F and G) show overlap of the sets of differentially abundant microbial genera between pairs of captive NHP populations, represented by circles colored by captive NHP population as in Figure 1 and in the key. In (F), the number of genera shared by each pair of captive NHP populations is shown above lines connecting cartoons colored as in Figure 1. In (F) and (G), numbers in non-overlapping portions of circles represent differentially abundant microbial genera detected in both NHP populations but differentially abundant in only one NHP population. Numbers in overlapping portions of circles represent microbial genera that were differentially abundant in both NHP populations. Significance of overlap for each comparison is shown by asterisks: *** p-value < 0.001; NS not significant.

Figure 3. Host Specificity Scores reveal microbial genus-specific humanization of captive NHP gut microbiotas. Shown are examples of two microbial genera that display evidence of humanization (A) or retention from the wild (B) in captive chimpanzees. Boxplots show the maximum, minimum, interquartile range, and median of BSD dissimilarities between wild chimpanzees and captive chimpanzees or between humans and captive chimpanzees. The logarithm of the Host Specificity Score, calculated as the ratio of the means of boxplots within each panel, was <0 for *Collinsella* (A) and >0 for vadinCA11 (B), indicating humanization and retention from the wild, respectively. Boxes are colored by primate species as in Figure 1. Significant differences between means based on permutation tests are indicated with asterisks; *** p = 0.001.

Figure 4. Humanization of gut microbial genera in captive NHPs is reproducible and host-species specific. Plots show relationships of microbial genus-level Host Specificity Scores (HSSs) between pairs of captive NHP populations. Each point represents the HSSs of a bacterial genus in two captive NHP populations (rows and columns), and lines indicate best-fit linear regressions. Cartoons represent NHP species and are colored as in Figure 1. Tree connecting cartoons shows phylogenetic relationships among NHPs. Rows and columns correspond to NHP species, with each plot showing a regression between HSSs for a pair of NHP species. Boxes are colored based on R^2 of regression, with $R^2 > 0.8$ colored dark red, $0.8 > R^2 > 0.1$ colored light red, and $R^2 < 0.1$ colored gray. R^2, Spearman rho, and corresponding p-values are reported in the top left corner of each plot. Significance for non-zero slope of regression line (based on Pearson correlation) is indicated in the bottom right corner of each plot; NS not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.