



Supplementary Materials for
**Gut microbiome heritability is nearly universal
but environmentally contingent**

Laura Grieneisen*, Mauna Dasari, Trevor J. Gould, Johannes R. Björk, Jean-Christophe Grenier,
Vania Yotova, David Jansen, Neil Gottel, Jacob B. Gordon, Niki H. Learn,
Laurence R. Gesquiere, Tim L. Wango, Raphael S. Mututua, J. Kinyua Warutere,
Long'ida Siodi, Jack A. Gilbert, Luis B. Barreiro, Susan C. Alberts, Jenny Tung†*,
Elizabeth A. Archie†*, Ran Blekhman†*

†These authors contributed equally to this work.

*Corresponding author. Email: lgrienei@umn.edu (L.G.);
blekhman@umn.edu (R.B.); earchie@nd.edu (E.A.A.);
jenny.tung@duke.edu (J.T.)

Published 9 July 2021, *Science* **373**, 181 (2021)
DOI: 10.1126/science.aba5483

This PDF file includes:

Materials and Methods
Figs. S1 to S18
Captions for Tables S1 to S15
References

Other Supplementary Material for this manuscript includes the following:
(available at science.sciencemag.org/content/373/6551/181/suppl/DC1)

Tables S1 to S15 (Excel)
MDAR Reproducibility Checklist (PDF)

Materials and Methods

1.1 Study system and metadata collection

The Amboseli Baboon Research Project (ABRP) collects continuous data on the demography, social interactions, ecology, and diet of hundreds of individual baboons in the Amboseli savannah ecosystem in East Africa (2°40'S, 37°15'E, 1100 m altitude; (38)) (**Fig. 1B and 1C**). The population is primarily composed of yellow baboons (*Papio cynocephalus*) with some admixture from nearby anubis baboon (*Papio anubis*) populations (note that prior research in our population finds no link between host hybrid ancestry and microbiome composition (39)). At any given time, ABRP studies approximately 300 baboons living in 4 to 6 social groups (14); ten social groups were observed during the 14-year sampling period in this study (**Fig. 1B**). Each group is visited 3 to 4 times per week, year-round, by experienced field observers who recognize all baboon group members on sight using morphological characteristics (38). During each monitoring visit, observers census the group and record detailed information on the group members' locations, social interactions, feeding patterns, and resource use, as well as collect fecal samples. We included data on baboon age, sex, kinship, social group membership, social behavior, diet, and the Amboseli environment in our implementations of the animal model to estimate heritability (**tables S4-S6**). Below we describe how data on these variables were collected.

Season and hydrological year. Amboseli is a highly seasonal ecosystem, with a long dry season from June to October (“dry season”) during which there is no rain and a highly variable wet season from November to May (“wet season”) during which rainfall is sometimes intense and sometimes very low. Hydrological year is calculated from November – October, to correspond with the start of the long rainy season in November (40).

Rainfall. Almost no rain falls during the dry season. In the remaining 7 months (November–May), the amount of rain varies widely between years (yearly average = 340 mm; range = 141–757 mm) (40). Because rainfall is inconsistent across months within a wet season, we modeled the effects of total rainfall in the 30 days prior to sample collection (following (31)).

Social group membership and size. Social group membership was known from censuses of all group members collected on each visit to each social group. This information was used to calculate the number of baboons, of all ages and both sexes, belonging to each group on the date of sample collection.

Age. For most baboon subjects, age is known to within a few days' error (n = 466 of the 585 individuals in our data set). For the remaining 119 subjects (e.g. males born in a social group not under long-term observation who later immigrated into the study population), birth dates are estimated using well-defined body metrics, including body size and carriage, tooth wear, coat condition, and comparison to baboons of known age (41). Baboons have well-defined life stages: “infants” nurse until ~70 weeks of age after which they are categorized as “juveniles.” The

juvenile period ends at sexual maturity (median age in Amboseli = ~4.5 years for females and ~5.7 years for males) after which individuals are considered reproductively mature adults (although they may not have achieved all social markers of adulthood at that point) (14, 15). Excluding individuals with estimated birthdates did not qualitatively change the age heritability analysis results (**Methods Section 1.12**).

Kinship and maternal identity. For every baboon born in a study group, maternity is known from observation. Paternity is determined based on genotyping from either blood or fecal-extracted DNA using up to 14 highly polymorphic microsatellite loci (42, 43). Using the resulting pedigree data, a kinship matrix for all individuals in the data set was created in R using the *pedantics* package (44). Out of the 585 baboons in our final data set, 497 individuals (n = 14,604 samples) have known maternities, and 464 individuals (n = 13,663 samples) have known paternities (**fig. S1**).

A kinship matrix for all individuals in the population was used to calculate the average pedigree relatedness within each social group for each of the 14 hydrological years with microbiome samples (n=77 social group years). Specifically, within-group mean and standard deviation of pedigree relatedness was calculated for each of the 77 social group years. To describe overall patterns of within group relatedness, we then reported a grand median for pedigree relatedness and a grand median for standard deviation of relatedness values across group years in the main text.

Diet. Data on diet composition is collected during 10-minute random order focal animal samples of adult female and juvenile behavior (45). During each focal animal sample, activity (e.g. feeding, walking, resting) is recorded at 1-minute intervals; when feeding is observed, the food type is recorded. Food items are classified by species and part (e.g. *Ramphicarpa montana* blossoms). Following (18, 19, 31), diet was estimated using feeding observations on all adult females and juveniles in the subject's social group in the 30 days prior to the sample collection date. Although individual-level diet information would be ideal, our focal sampling data are sparse on the individual level, and as a result, we cannot use an individual's focal animal samples to infer individual variation in diet over time-scales shorter than several months. However, social group-level diet is likely a sufficient representation of individual-level diet in the Amboseli population for three reasons. First, each baboon social group travels together throughout the day, stopping to forage together in the same location. For example, group membership predicts dietary composition for pregnant females, but dominance rank, parity, and fetal sex does not (46). Indeed, such group foraging is proposed to be adaptive as younger members can learn which foods to eat, and infant baboons have been observed consuming the same diets as their mothers (47). Second, although individual baboons do have dietary preferences (48), available food choices are determined by season and rainfall. For example, time spent foraging is more strongly predicted by environmental conditions than social group membership (49). Finally, several studies that have intensively sampled individual-level diets over short time scales have found differences in the quantity but not composition of food consumed. For example, pregnant and lactating females spent more time eating and consume more calories than cycling females in

the same social groups, but did not differ in types of nutrient intake (50). Low-ranking animals are interrupted during foraging bouts by other baboons more often than high ranking animals, but interruption rate was not dependent on food quality, suggesting that higher-ranking animals are not preferentially taking certain food items (51). Taken together, these studies suggest that calculating diet at the level of social group is an adequate representation of individual-level diet.

To determine diet composition, we classified the baboons' primary foods into 14 categories (**table S1**). Diet category definitions were based on previous baboon microbiome studies, which collapsed plants based on parts eaten (blossom, bark, etc.) (18, 19). Because values collected during the 1-minute intervals within each focal sample are autocorrelated, we generated 1,000 random subsets of 1 value per focal sample, after limiting focal samples to only points where the observed activity was feeding. For each individual, we then averaged the abundance of each food category to calculate the mean relative contribution of each food category to their diet. To create a diet composition variable, we then used these data to create an individual-based dietary PCA using *labdsv*, such that each microbiome sample from a baboon with social group-level diet data is associated with a value for each of PCs 1-13 of the diet data (**table S2**). To create a diet diversity variable for the same samples, we used the dietary relative abundance data to calculate a Shannon's H index, following (31). We determined that at least 15 focal samples were needed to make sure that diet profiles were stable, leaving us with a final data set of 16,234 samples with associated diet data (**Fig. 1C**).

To calculate age-based differences in diet, we used a slightly different approach, as detailed in **Methods Section 1.12**.

1.2 Creating microbiome profiles

Sample collection and storage. Fecal samples were collected opportunistically from individually recognized baboons within 15 minutes of defecation, using a wooden stick and waxed paper cup (38). The defecation time and date, together with the individual's name, were noted. The fecal sample was homogenized by stirring, and an aliquot of the sample was placed into a 20 mL vial with 11 mL 95% ethanol. Samples were transported from the field back to the Amboseli field camp and stored in the field at ~25°C. At 2-week intervals, these samples were transported to the University of Nairobi where they were freeze-dried and sifted following (52). The freeze-dried samples were stored at -20°C until shipment to the United States, where they were stored at -20°C. The original purpose for fecal sample collection was to characterize the concentration of steroid hormone metabolites (e.g., (53–58)). Our analyses used excess freeze-dried fecal powder that did not undergo hormone extraction. Past work has found that microbiome samples processed following this pipeline yield comparable results to other microbiome preservation methods (59).

DNA extraction and sequencing. DNA was extracted using the MoBio and QIAGEN PowerSoil kit for 96-well plates (60, 61). DNA was extracted from ~0.05 g of freeze-dried fecal powder for each sample. Extraction proceeded according to the manufacturer's instructions, with some modifications (61): (i) we increased the amount of PowerBead solution to 950 µL/well to

increase the hydration of the freeze-dried samples, and (ii) after the addition of PowerBead solution and lysis buffer C1, we incubated the plates at 60°C for 10 minutes.

16S rRNA amplicon libraries were constructed following the Earth Microbiome Project's (EMP) standard protocols (62). Specifically, we used polymerase chain reaction (PCR) to amplify a ~390 bp-long fragment that encompassed the V4 region of the 16S rRNA gene using primers 515F - 806R (63, 64). Amplicons were quantified with Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher/Invitrogen cat. no. P11496). Equal amounts of amplicon DNA from each sample (70 ng) were pooled into a single tube. The pooled amplicons were cleaned using AMPure XP beads (Beckman Coulter). The cleaned amplicon pool was quantified using a Qubit fluorometer. Libraries were sequenced on the Illumina HiSeq 2500 using the Rapid Run mode (2 lanes per run). Sequences were single indexed on the forward primer and 12 bp Golay barcoded.

Sequence processing and initial quality filtering. Sequences were processed using the Illumina demultiplexing protocol and DADA2 pipeline (65). To maintain quality control, we added steps to prevent barcode misassignment, resolve singleton amplicon sequence variants (i.e., sequence variants that only appear in one sample), and account for technical variation.

Samples were demultiplexed using the standard *bcl2fastq* function. To prevent barcode misassignment, we included the additional optional argument *--create-fastq-for-index-reads*, which generates additional fastq files for index reads, allowing us to filter sample fastq reads by their corresponding index sequence (66). Sequences in the index fastq files with a maximum error estimate (*maxEE*) < 0.1 were retained using DADA2's *filterAndTrim(...)* function (65). Sequences in the sample fastq files were only retained if their corresponding index read passed the quality filter. Finally, *cutadapt* with *minimum length=150* bases was used for Illumina library adapter removal and to remove reads that were too short for further analysis (67).

We merged read pairs and performed initial quality filtering using the default DADA2 protocol for large data sets (68) with the following exceptions: *FilterAndTrim(truncLen=c(240,160), maxN=0, maxEE=2, rm.phix=TRUE, truncQ=11)*. DADA2 processing was performed on each HiSeq sequencing run of two lanes each (*n* = 8 runs), and then runs were combined (*mergeSequenceTables(...)*). Potentially chimeric reads were removed (*removeBimeraDenovo(...)*).

Because DADA2 was run separately on each HiSeq sequencing run, we pooled all of the samples and used *LearnErrors(...)* to resolve any singleton ASVs across runs, which could reflect errors. If the pipeline was run without pooling, there were 45,350 ASVs in the final data set, but pooling decreased the number of ASVs to 36,349. Lastly, we performed taxonomic identification with *IdTaxa(...)*, available in the DECIPHER package (69), against the Silva reference database SILVA_SSU_r132_March2018.RData (70). The numbers of reads remaining at each stage in this processing pipeline are provided in **table S3**.

Additional quality filtering: We designed our study to include rigorous quality control checks on the processed data. First, to ensure that older samples had not degraded, we tested if sample age predicted DNA concentration. We found no relationship between how long a sample had been stored and its DNA concentration (linear model; *b* = 2.0×10^{-04} , *p* = 0.064) (**fig. S2**).

Second, batch effects are of great concern in microbiome studies (71–73). To ameliorate potential batch effects in our data, we pooled samples as part of the DADA2 pipeline to resolve singleton ASVs, as discussed above. To further test for batch effects, we sequenced technical replicates (i.e., the same sample extracted and prepared for sequencing on multiple plates, and sequenced on multiple lanes/flowcells). There were 182 technical replicates from 30 samples, spread over 40 96-well library prep plates (range = 3 - 8 replicates / sample, median = 6). Technical replicates from different plates clustered with each other rather than with their plate, indicating that true biological differences between samples are larger than plate-based batch effects (**fig. S3**). Third, we included blank sequencing wells (i.e. DNA extraction blanks). Blank wells had significantly less DNA than the sample wells (t-test; $t = -50$, $p < 2.2 \times 10^{-16}$). We removed samples that had DNA extraction concentrations $< 4X$ the blank on that sample's plate, as well as samples with $< 1,000$ reads. Finally, we removed blank samples and technical replicates, as well as sequences found in only 1 sample, leaving us with a data set of 17,167 samples and 10,720 ASVs. Over 87% of samples were retained after all of our quality filtering checks. The number of reads per sample in this data set ranged from 1,021 to 477,241 with a median of 50,220 (**fig. S4**). Note that the final data set used in this paper was smaller ($N = 16,234$ samples from 585 individuals) because diet information was not available for some samples.

1.3 Defining gut microbiome phenotypes

We estimated the heritability of 1,034 gut microbiome “phenotypes”. These phenotypes included: **(i)** 7 measures of overall microbiome composition (*community phenotypes*: ASV richness, ASV Shannon’s H index, and the first five principal coordinates (PCs) of a Bray-Curtis dissimilarity matrix); **(ii)** 283 *single-taxon phenotypes* representing the relative abundance of individual microbiome taxa—from ASVs through phyla—that were found in $>50\%$ of samples; **(iii)** 744 *presence/absence phenotypes* coded as a 1 or 0 to represent the presence (1) or absence (0) of single taxa at the ASV through phylum levels found in 10% to 90% of samples. Beyond these 1,034 phenotypes, for some analyses we aggregated the 290 *community* and *single-taxon phenotypes* into **(iv)** 100 *collapsed phenotypes*. Details on how each set of phenotypes was constructed are as follows:

(i) Community phenotypes. In total, there were 7 community phenotypes: two measures of alpha diversity and five principal coordinates of a Bray-Curtis dissimilarity matrix. We calculated two measures of microbial alpha diversity based on the final ASV table. ASV richness was calculated for each sample as the number of unique ASVs found in the sample. ASV Shannon’s H was calculated using “metric = Shannon” in the *vegan* package (74) (**fig. S5**). To measure overall microbiome composition, we next normalized the ASV table using a CSS transformation (*metagenomeSeq*; (75)) and calculated a pairwise Bray-Curtis dissimilarity matrix between all pairs of samples using *vegdist()* in the *vegan* package. We then calculated sample-level projections onto the first 5 coordinates of a Principal Coordinates Analysis (PCoA) on the Bray-Curtis matrix using *labdsv* (76). The first 5 coordinates explain 48.5% of the variance in

microbiome composition between samples (PC1 = 19%, PC2 = 14.8%, PC3 = 5.8%, PC4 = 5.3%, PC5 = 3.6%).

(ii) Single-taxon phenotypes. In total, there were 283 single-taxon phenotypes, which reflect the relative abundance of individual ASVs and higher-level taxonomic groupings, from genera through phyla. To calculate the single-taxon phenotypes, we used the ASV table that had not been CSS transformed. We aggregated by taxonomic level (phylum, class, order, family, genus, ASV) and rescaled each sample to 100 to get the percent abundance per sample per taxonomic level. For taxa that were aggregated to a taxonomic level above the ASV (i.e. genera through phyla), we removed those that were unassigned or that matched chloroplasts or mitochondria. This filtering left 1,102 aggregated, assigned taxa (genus through phylum levels) and 10,635 individual ASVs. We then filtered this set to retain taxa that were found in at least 50% of samples, which produced 144 higher-level taxonomic groupings (genus through phylum; **fig. S6**) and 139 ASVs (144 non-ASV taxa and 139 ASVs = 283 taxa).

To test if our heritability results could be due to data set properties, we ran heritability models on two additional single-taxon data set transformations. These accounted for compositionality of microbiome data (centered log ratio transformation) and phylogenetic relatedness between taxa (phylogenetic balances transformation).

Centered log ratio (CLR) transformation. The centered log ratio (CLR) transformation accounts for the compositionality of microbiome data by computing the log-ratio of each feature to the sample mean, making it robust to differences in read depth across samples (77). We limited the ASV counts table to the 283 ASV through phylum phenotypes found in >50% of samples and added 0.65 as a pseudocount to all values (following (78)) before calculating the CLR for each of the 283 microbiome phenotypes using the R package *compositions* (79).

Phylogenetic balances (PhILR) transformation. PhILR is a **Phylogenetically** aware transformation of **Iso**metric **L**og-**R**atios that incorporates both the microbiome's compositionality and the phylogenetic relationships between taxa (78). Specifically, a PhILR balance represents the log-ratio of the geometric mean abundance of the two groups of taxa that split at a given node of the inputted phylogenetic tree. We constructed a phylogenetic tree of the 139 ASV sequences found in >50% of samples and used the PhILR transformation in the *philr* R package to calculate phylogenetic balances (for 139 sequences, there are n=138 nodes, resulting in 138 balances in the tree; (78)).

(iii) Presence/absence phenotypes. We estimated the heritability of the presence/absence of individual taxa at the ASV through phylum level, limited to those found in 10% to 90% of samples (n = 744; **fig. S9A, S9B**).

(iv) Finally, for a subset of analyses (those comparing heritability differences across seasons, years, age windows, and sampling depths, as well as the coefficient of variation analyses), we ran analyses on 100 *collapsed phenotypes*. These collapsed phenotypes included the 7 community-level phenotypes and, following common practice in prior studies (1, 2, 6), 93 single-taxon phenotypes from the genus to phylum levels (we excluded ASVs, which are rarely

comparable across studies) that were collapsed across phylogenetically nested taxonomic levels. For example, the relative abundance of the family Christensenellaceae was nearly completely driven by, and 99% correlated with, the relative abundance of the genus *Christensenellaceae R-7 group* nested within it, so only *Christensenellaceae R-7 group* was included in the set of collapsed phenotypes. We determined collapsed phenotypes by running pairwise Pearson correlations between all 144 non-ASV taxa found in >50% of samples using the `rcorr` function in the *Hmisc* package (80). Of these, 86 nested taxon pairs were correlated at >0.97 (fig. S14). We consolidated these taxa, such that only the lowest taxonomic level was used (e.g. we used the genus if it was highly correlated with the family), leaving the 93 collapsed traits in table S13. In models that use collapsed phenotypes, we present heritability estimates for the lowest taxonomic level of the collapsed taxa.

1.4 Response variable comparison

To test if heritability estimates for taxa were consistent across transformations, and to determine directionality, we ran Pearson's correlations and paired t-tests comparing heritability estimates for single-taxon phenotypes to their corresponding CLR transformations and presence/absence heritability estimates. We found that heritability estimates for taxa were significantly correlated across transformations, and the heritability of relative abundance yielded the most conservative heritability estimates of any transformation (fig. S9D, S9E). Presence/absence heritability estimates for $n = 204$ taxa found in 50-90% of samples were correlated with the relative abundance of those same taxa (Pearson's $R=0.68$, $p=3.2 \times 10^{-29}$) but were significantly higher for presence/absence phenotypes (paired t-test $p=2.2 \times 10^{-27}$). CLR heritability estimates were highly correlated with relative abundance estimates ($n = 283$ taxa in >50% of samples; Pearson's $R = 0.82$, $p = 2.3 \times 10^{-69}$), and CLR estimates were also significantly higher than relative abundance estimates (paired t-test; $p = 2.7 \times 10^{-27}$). Taken together, these analyses reveal that both common and rare taxa demonstrate high rates of heritability, and that these rates are robust to compositionality.

We also tested if prevalence and heritability of presence/absence phenotypes were positively correlated. We found that taxa present in more samples tended to have higher heritability estimates (among the $n = 744$ taxa found in 10-90% of samples; Pearson's $R=0.28$, $p=2.3 \times 10^{-15}$; fig. S9C).

Finally, Bray-Curtis PC1 (the most heritable single-taxon and community phenotype) was also correlated with the relative abundance of the top 3 most heritable individual taxa (an ASV belonging to *Prevotella 9*, $R = 0.50$, $h^2=0.20$; the genus *Prevotella 2*, $R = 0.40$, $h^2=0.17$; and an ASV belonging to Betaproteobacteriales, $R = 0.36$, $h^2=0.16$; all $p < 1.1 \times 10^{-13}$).

1.5 Estimating heritability using the "animal model"

Introduction to the animal model. To estimate the heritability of each microbial phenotype, we used the narrow-sense definition of heritability (h^2): that is, the amount of total phenotypic variance of a trait explained by additive genetic effects (excluding the effects of dominance, epistasis, and other non-linear interactions) (25). We used the `asreml()` function in the R package *ASReml* to run the 'animal model', a linear mixed effects model that partitions variance between random effects after controlling for fixed effects using a restricted maximum

likelihood (REML) approach (23). The animal model is a standard method for calculating the heritability of physical and behavioral traits in pedigreed livestock (81–85) and wild animal populations (86–89). In livestock, the animal model has been used, for example, to determine the heritability of traits including litter size of piglets (84), weaning weight of cattle (81), quality of sheep wool (85), cannibalism in hens (83), and spawning date in farmed salmon (82). In wild populations, the animal model has been used to understand environmental and genetic variation in complex traits, including plasticity in laying date in wild gulls (86), sexually antagonistic selection in red deer (88), age at first and last reproduction in swans (89), and body mass in relation to climate change in gulls (87). Many of the wild populations analyzed with the animal model to date are long-term study systems with decades of individual-level information on behaviors and physical traits, similar to the Amboseli baboons (reviewed in (24)).

Choosing random effects for the animal model. The animal model partitions variance between the model random effects and categorizes variance that cannot be explained by the random effects as residual environmental variance. **Tables S4** and **S5** describe the fixed and random effects we included in each model. Our random effects included: (i) individual identity, to account for repeated sampling from each individual (i.e. non-additive interindividual variation; also referred to as “permanent environment effects”); (ii) maternal identity, to estimate maternal effects (both genetic and environmental); (iii) DNA extraction plate, to account for batch effects; and (iv) the additive genetic effects for each individual, on the trait of interest (sometimes referred to as “genetic merit” or “breeding value” (24)), which is estimated from the expected covariance in additive genetic effects between relatives in a pedigree and is the key parameter for estimating heritability.

Choosing fixed effects for the animal model. Fixed effects in the animal model serve two purposes (12, 25, 26). First, they help control for sampling bias that arises from the fact that individuals are sampled at different ages and in different environments, which in turn affects microbiome phenotype values. Second, fixed effects help account for shared environments, diets, and behaviors among relatives; failure to do so could inflate heritability estimates because variance attributable to similar environments and behaviors (or direct transmission between related individuals in the same environment) could be incorrectly attributed to additive genetic variance.

The fixed effects that we included in the model were those that have also been shown to influence microbiome composition in past studies and/or are likely to have more similar values between related individuals (references for each trait below; **fig. S8, S9; table S4**). To account for diet composition, we included 13 principal components of dietary variation as described above (e.g. (90, 91)). To account for host traits, we included host age at the time of sampling and host sex (e.g. (92, 93)). To account for temporal autocorrelation in the data and fluctuations in the physical environment, we included collection date as a running count of days through the data set (starting at 1 on the earliest date of collection, with a maximum of 4,899 days), collection month (as a factor), rainfall in the 30 days prior to sample collection, and hydrological year (as a factor; e.g. (31, 94)). To account for the host’s social environment, we included social

group membership and social group size (i.e. number of baboon group members) (e.g., (18, 19, 92)). Finally, we included post-PCR DNA concentration and sequencing read count to control for technical effects. Each fixed effect improved the majority of model fits (mean=65%) for the 100 *collapsed phenotypes* (Wald's test).

Model input. The basic animal model structure (24) is

$$y = Xb + Z_u + Z_m + Z_p + Z_i + e$$

where y is a vector of values for the microbiome phenotype of interest; b is the fixed effects vector; X is a matrix of observed values for the model covariates, for all samples in the data set; e is a vector of residual errors; and the Z terms are the random effects: Z_u represents additive genetic effects, Z_m represents maternal identity, Z_p represents extraction plate, and Z_i represents individual identity.

Our model input for testing the heritability of single-taxon and community phenotypes (Figs. 2A-C) was as follows, where the identity of the individual in the data set (`baboon_id`; see **table S4** for the codes used for each variable) is linked to the inverse relatedness matrix using the `ginverse` argument, `ped(baboon_id,var=T,init=1)` is the additive genetic variance term, `ide(baboon_id,var=T,init=1)` is the individual identity term, and `na.method.X="include"` means that rows with missing values are still included in the model (i.e. if maternal identity is unknown; see **Methods Section 1.15**). For the *presence/absence phenotypes*, we ran the same model, but additionally specified `family = asreml.binomial(link = 'logit')` and `maxiter = 20` (to account for longer convergence times).

```
#create the inverse of the additive genetic relationship matrix from
the pedigree data, ped2
ainv<-asreml.Ainverse(ped2)$ginv

#run the model
  modely <- asreml(fixed=asin(sqrt(taxon_relative_abundance))~ 1 +
collection_date + readcount + age + sex + rain_month_mm + social_group
+ group_size + diet_PC1 + diet_PC2 + diet_PC3 + diet_PC4 + diet_PC5 +
diet_PC6 + diet_PC7 + diet_PC8 + diet_PC9 + diet_PC10 + diet_PC11 +
diet_PC12 + diet_PC13 + post_pcr_dna_ng + hydrological_year +
collection_month
, random= ~ped(baboon_id,var=T,init=1) + plate + MOTHER +
ide(baboon_id,var=T,init=1)
, ginverse=list(baboon_id=ainv)
, na.method.X="include", na.method.Y="omit")
```

How the animal model calculates heritability. We estimated heritability following the convention in human genetics and plant and animal breeding, in which the total phenotypic variance (the denominator of the heritability ratio) is estimated after correcting for known fixed

effects (12, 26, 27). Importantly, this allows us to exclude the contribution of purely technical factors, such as batch and DNA concentration, from our estimate of phenotypic variance. It also allows us to exclude the contribution of demographic variables like sex and age. This approach is similar to how h^2 estimates generated for traits like human height, where the typical estimate (~0.7-0.8) is generated after adjusting for the fixed effect contributions of sex, age, and sometimes cohort or other secular trends (95, 96).

Specifically, the animal model output from *asreml* includes a table of the variance partitioned to each random effect after controlling for fixed effects (**tables S4-S10**). To calculate the heritability (h^2) of a phenotype, we divided the additive genetic variance by the sum of all the estimated variance components (i.e. phenotypic variance, also called V_p) (25):

$$h^2 = \frac{V_{\text{Additive genetic}}}{(V_{\text{Additive genetic}} + V_{\text{Individual}} + V_{\text{Maternal identity}} + V_{\text{Technical}} + V_{\text{Residual}})}$$

Choosing response variables for the animal model. For our main analyses, we ran the animal model on all 16,234 samples using the fixed effects and random effects described above and in **table S4** and **S5**. We ran the model separately for each set of response variables (see **Methods Section 1.3** for details of these phenotypes; see **tables S6-S9** for model outputs):

- 7 community phenotypes
- 283 single-taxon phenotypes; using the arcsin square-root transformed relative abundances of taxa. We also ran a second set of models on the centered log-ratio transformed counts for the same 283 phenotypes, and a third set on the PhILR balances of n=139 ASVs found in >50% of samples;
- 744 presence/absence phenotypes
- 100 collapsed phenotypes

Hypothesis testing. To determine if a single-taxon or community phenotype was significantly heritable, we ran the full animal model, and then ran a reduced model that excluded the additive genetic effects term. We used a likelihood ratio test to determine if the full model was a significantly better fit to the data than the reduced model following (25) (see code below). We corrected for multiple testing within each set of phenotypes by calculating a 10% false discovery rate (FDR) using the approach of Benjamini and Hochberg (97) (model outputs shown in **table S6**).

```
#code for reduced model for testing the significance of the additive genetic variance term
```

```
#run the full model, as described in Methods Section 1.5
```

```
modely <- asreml(fixed=asin(sqrt(taxa))~ 1 + collection_date +
readcount + age + sex + rain_month_mm + social_group + group_size +
diet_PC1 + diet_PC2 + diet_PC3 + diet_PC4 + diet_PC5 + diet_PC6 +
```

```

diet_PC7 + diet_PC8 + diet_PC9 + diet_PC10 + diet_PC11 + diet_PC12 +
diet_PC13 + post_pcr_dna_ng + hydrological_year + collection_month
, random= ~ped(baboon_id,var=T,init=1) + plate + MOTHER +
ide(baboon_id,var=T,init=1)
, ginverse=list(baboon_id=ainv)
, na.method.X="include", na.method.Y="omit")

#run the reduced model
modely_no_pedigree <- asreml(fixed=asin(sqrt(taxa))~ 1 +
collection_date + readcount + age + sex + rain_month_mm + social_group
+ group_size + diet_PC1 + diet_PC2 + diet_PC3 + diet_PC4 + diet_PC5 +
diet_PC6 + diet_PC7 + diet_PC8 + diet_PC9 + diet_PC10 + diet_PC11 +
diet_PC12 + diet_PC13 + post_pcr_dna_ng + hydrological_year +
collection_month
, random= ~plate + MOTHER + ide(baboon_id,var=T,init=1)
, ginverse=list(baboon_id=ainv)
, na.method.X="include", na.method.Y="omit")

#calculate the p-value of the likelihood ratio test comparing models
LRT_p <- 1-pchisq(2*(modely$loglik - modely_no_pedigree$loglik),1)

```

FDR threshold. In the main text, we report heritable phenotypes using an FDR of 10%, which led us to infer that gut microbiome phenotypes are near-universally heritable (97% of single-taxon and community phenotypes). This inference is robust to a more stringent FDR threshold: at 1% FDR, 93% of phenotypes remained significantly heritable.

Permutation test of heritability estimate size. To determine if heritability estimates were significantly higher than expected by chance, we randomly permuted the baboon identities 1000 times and re-ran the heritability model for each permutation for each of the 100 collapsed phenotypes. We then plotted the observed heritability estimate for each phenotype within the distribution of the permuted estimates and found that for 97 / 100 phenotypes, the observed heritability estimate was substantially higher than all 1000 permutations ($p < 0.001$; **fig. S10**). For 3 phenotypes, the heritability estimate was on the tail end of the distribution, but still higher than expected by chance; the genus *Lachnospiraceae FCS020 group*, $p = 0.005$; the class Bacilli, $p = 0.003$; and the genus *Senegalimassilia*, $p = 0.001$. Bacilli and *Lachnospiraceae FCS020 group* are not significantly heritable in the animal models, and *Senegalimassilia* is significantly heritable at $h^2 = 0.082$. Taken together, the permutation results suggest that our heritability estimates are higher than expected by chance.

Heritability significance for presence/absence phenotypes. There are two substantial differences in interpreting heritability using a binomial model. First, because our binomial model used the logit link function, the link variance must be included in the denominator of the heritability calculation (98) rather than the residual variance (VR, which is outputted as a dummy

variable). The link variance for the logit link is 3.29 (99). As such, the heritability calculation for presence/absence models is:

$$h^2 = \frac{V_{Additive\ genetic}}{(V_{Additive\ genetic} + V_{Individual} + V_{Maternal\ identity} + V_{Technical} + 3.29)}$$

Second, likelihood ratio tests, which were used for model selection in the other heritability models, cannot be used for the binomial animal model (23). To determine if presence/absence phenotypes were significantly heritable, we therefore used two alternative approaches. First, we classified a phenotype as heritable if the V_A component made a positive (non-zero) contribution to the model and if the standard error bars did not overlap with 0. 704 / 744 phenotypes (95%) met these criteria (model outputs show in **table S7**). Second, to determine if heritability estimates were significantly higher than expected by chance, we randomly permuted the baboon identities 20 times and re-ran the heritability model for each permutation for each of the 744 microbial phenotypes. 720 phenotypes had heritability estimates higher than expected by chance ($p < 0.05$), and the 24 phenotypes that did not have heritability estimates higher than expected by chance all also failed the first criterion for non-zero heritability. Given that the first approach yielded a more conservative estimate of the number of heritable phenotypes than the second approach, we used the first approach to define significantly heritable phenotypes.

1.6 Testing for phylogenetic effects on heritability estimates

To test for a phylogenetic effect on taxon heritability, we constructed an ultrametric phylogenetic tree of the 139 ASV sequences found in >50% of samples using the *ape* and *phangorn* packages in R (100, 101). We tested for an overall signal of phylogenetic relatedness on the heritability of these 139 ASVs using standard metrics of phylogenetic signal (Moran's I, Blomberg's K, and Pagel's lambda) in the *phylosignal* R package (102). There was an overall signal of phylogenetic relatedness on Moran's I and Pagel's lambda ($I = 0.0996$, $p = 0.001$; $\lambda = 0.73$, $p = 0.001$) but not on Blomberg's K ($K = 0.051$, $p = 0.09$). The Pagel's lambda for microbial heritability estimates is in a similar range to the Pagel's lambda for social and ecological traits across primate species, including the number of females in a social group (0.742), day range (0.743), and midpoint of latitudinal range (0.733) (103). To determine which taxa were driving the phylogenetic signal, we ran local Moran's I in *phylosignal* (**fig. S11**).

1.7 Correlating heritability estimates for microbial taxa found in humans and baboons

We compared the heritability of baboon microbiome phenotypes to heritability estimates of human microbiome phenotypes obtained from 7 human data sets in 5 studies (one study, (1), analyzed 3 data sets) (1, 2, 4–6, 32, 33). We note that this comparison is likely conservative, as different methods were used to assign microbial taxonomy and measure heritability across the human studies and between the human studies and our study (1, 2, 4–6). There were 32 taxa at the genus through phylum level that were heritable in the current baboon study and at least 1

human data set (based on the significance criteria applied by the study authors; **table S12** and **Fig. 2D**). To test whether heritability estimates for these 32 taxa were correlated between baboons and humans, we ran a linear mixed effects model with heritability estimates in humans as the response variable, heritability estimates in baboons as the fixed effect, and the identity of the human data set as a random effect. This analysis revealed that heritability estimates were positively correlated in baboons and humans ($b = 0.91$, $p = 0.014$). This pattern was also supported when we ran a Pearson's correlation on the same data set, as reported in the main text, using the average heritability of a taxon across the human studies in which it was detected as heritable. We found that h^2 in baboons was significantly correlated with this mean human h^2 value (Pearson's $R = 0.52$, $p = 0.002$).

1.8 Co-occurrence network analysis

To test if heritable taxa in baboons demonstrated similar co-occurrence networks to heritable taxa in humans, we ran SparCC on 31 significantly heritable microbial families found in > 50% of samples (104), following (1). SparCC cannot account for multiple samples from the same individual, so we ran SparCC for 100 permutations on the mean taxon abundance within each baboon individual for each family. We did not find the same clustering patterns as (1), but we did find that a taxon's heritability was positively correlated with its number of significant network connections (Pearson's $R = 0.58$, $p = 0.0006$; **fig. S13A, S13B**).

1.9 Comparing microbiome heritability estimates to other heritable traits in primates

To place our results in a broader evolutionary context, we collected heritability estimates for other traits in nonhuman primates and created a qualitative visualization (**fig. S12**). A full list of traits, studies, and primate species is shown in **table S10**. There are no published studies on the heritability of complex traits in wild baboons, so instead we focus on 180 heritable traits drawn from 19 studies across captive baboons and bonobos, and in free-ranging macaques (105–123). We classified the traits as:

1. *Free-ranging primate behaviors*. These include 4 studies measuring the heritability of 37 social behaviors in free-ranging macaques (*Macaca mulatta*).
2. *Zoo primate behaviors*. These include 1 study measuring the heritability of 10 social behaviors in zoo bonobos (*Pan paniscus*).
3. *Baboon physiology*. These include 11 studies measuring the heritability of 84 physical traits in captive baboons.
4. *Captive baboon behaviors*. These include 1 study measuring the heritability of 66 behaviors in captive baboons.
5. *Baboon life history*. These include 2 studies measuring the heritability of 2 traits in captive baboons.

1.10 Seasonal effects on heritability estimates

Amboseli is a highly seasonal ecosystem. To test the effects of season on heritability estimates, the data set was split into dry season ($n = 6,650$ samples from 543 individuals) and

wet season (n = 9,584 samples from 536 individuals). Within each season, the animal model was run on each of the 100 collapsed phenotypes, using the same fixed and random effects as in the full model, as well as the same procedure to correct for multiple testing (see **Methods Section 1.5**). 93 phenotypes were significantly heritable in the dry season, 92 phenotypes were significantly heritable in the wet season, and 89 were significantly heritable in both (10% FDR).

Environmental diversity may explain seasonal effects. Although we control for diet in our models, individuals who eat diverse diets may be more exposed to other aspects of season-dependent environmental diversity that our model does not capture. To test this idea, we stratified the data in terms of dietary Shannon's H and applied the animal model to the 50% most diverse diet samples, using the same procedure as for the full model. 72% of the samples in the high diet diversity group are from the wet season. We then repeated the models on the 50% least diverse diet samples, and ran a paired t-test on the heritability estimates between the two model groups (**fig. S15B**). If samples are split into tertiles, a similar pattern emerges: 77% of samples in the most diverse tertile are from the wet season, and heritability estimates decreased from the highest to lowest dietary diversity groups (**fig. S15C**).

Seasonal differences in phenotypic variance. We tested if seasonal differences in heritability estimates were due to higher total phenotypic variance (V_p) in the wet season than in the dry season. We found significantly higher V_p for microbial phenotypes in the wet season versus the dry season (paired t-test; $t = 4.3$, $p = 4.2 \times 10^{-5}$; **fig. S15A**). For the purpose of this analysis, we excluded ASV richness and ASV Shannon's H, as V_p for these two phenotypes was several orders of magnitude higher than for other microbial phenotypes. If the V_p values for ASV richness and ASV Shannon's H are scaled to a similar range to the other V_p values by dividing by 100,000, the paired t-test yields nearly identical results ($t = 4.3$, $p\text{-value} = 3.8 \times 10^{-5}$).

1.11 Variation in heritability over time

To test if heritability estimates varied over time, we ran the animal model on the 15 most heritable collapsed microbial phenotypes in hydrological years with ≥ 150 baboons and ≥ 1000 samples. The hydrological years included 2002 - 2005, 2007 - 2009, and 2011-2012 (see **table S11** for sample sizes and **table S5** for fixed and random effects included in each model).

1.12 Effects of host age on microbiome heritability estimates

Defining age windows. To test how host age affected heritability estimates, we ran the animal model on sliding windows of host age, randomly subsampled to 2000 samples from 200 individuals per window (n = 14 windows). Each sliding window encompassed 3 years, and increased by 1-year increments, with one exception: the final window encompassed animals aged 13 - 27 years, due to small sample sizes in the oldest age classes (**table S14**). We chose these parameters because sample size affects heritability estimates and the majority of baboons sampled were between the ages of 1 and 14 years. Using 3-year overlapping windows allowed us to sample as deeply and evenly as possible while still capturing age variation, although we note that some windows have < 2000 samples or < 200 individuals because there were insufficient

numbers of samples available for that window, such that the distribution of values is 2000 +/-284 (median +/- SD) samples from 200 +/- 18 individuals (see **table S14**).

Running age models. Within each of the 14 age windows described above, the animal model was run on each of the 100 collapsed phenotypes, using the same fixed and random effects as in the full model (**table S5**), as well as the same procedure to correct for multiple testing (see **Methods Section 1.5**). We used two approaches to test the relationship between heritability and host age. First, we tested for a global change in heritability with age. We ran a linear mixed model with the heritability estimates of all 100 collapsed phenotypes across all 14 age windows as the response variable; age window, number of individuals, and sample size as fixed effects; and phenotype identity as a random effect. This analysis revealed a global increase in microbiome heritability estimates with host age ($b = 0.0021$, $p = 0.00027$).

Second, we tested which of the 100 collapsed phenotypes changed with age. To do so, we ran a set of linear models, with the heritability estimates of each of the 100 collapsed phenotypes as the response variables and age window as the fixed effect. This analysis revealed that 32 phenotypes changed in heritability with age, with 29/32 increasing with age (**Fig. 3E, 3F**).

Age window sample size effects on heritability estimates. The age window analysis included ~2000 samples and ~200 individuals per age window. We obtained qualitatively similar results if sample sizes were decreased to ~1500 samples from 150 individuals per window (Pearson correlation for significant heritability estimates; $R = 0.82$, $p < 0.001$). Further, the ~2000 sample, 200 individual data set also yielded repeatable results across 3 random subsamples (Pearson correlation for significant heritability estimates; $R = 0.92 - 0.94$; $p < 0.001$). On average, 63% of the available samples within each age window (and 78% of the available individuals) were used within each age window.

Age differences in variance components. We tested if age differences in heritability estimates were due to decreasing total phenotypic variance (V_p) with age, increased additive genetic variance (V_A), or decreased residual environmental variance (V_R). Parallel to our tests of age-related changes in heritability, we ran 3 linear mixed models with total V_p , V_A , and V_R for all 100 collapsed phenotypes across all 14 age windows as the respective response variables, age window, number of individuals, and sample size as fixed effects, and phenotype identity as a random effect. These analyses revealed no change in V_p with host age ($b = -2.0 \times 10^{-6}$, $p = 0.19$), an increase in V_A ($b = 1.7 \times 10^{-5}$, $p = 0.0085$), and a decrease in V_R ($b = -2.9 \times 10^{-5}$, $p = 0.00015$). For the purpose of these analyses, we excluded ASV richness and ASV Shannon's H, as V_p for these two phenotypes was several orders of magnitude higher than for other microbial phenotypes. We then limited the data set to the 29 phenotypes that showed increased heritability with age and ran a separate linear model for each of the 29 phenotypes with V_p as the response variable and age window as the fixed effect. This analysis revealed V_p changed with age for 22/29 phenotypes, and V_p increased with age for 17 of these 22 phenotypes.

Age differences in taxa relative abundance. We tested if age differences in heritability estimates could be related to age-related changes in taxon relative abundances or overall compositional phenotypes, but did not find a clear relationship. We ran a linear mixed model for each of the 100 collapsed phenotypes with the phenotype as a response variable; age, sex, collection date, month, hydrological year, read count, post-PCR DNA concentration, rainfall, social group, social group size, and the 13 diet PCs as fixed effects; and maternal identity, sequencing plate and individual identity as random effects. After correcting for multiple testing, age was a significant predictor of taxon relative abundance in 46/100 models. Of these 46, 21 showed a positive relationship with age (corrected $p < 0.05$). There was no correlation between the age effect estimate from the linear mixed models and phenotype heritability (Pearson's $R = -0.0573$; $p = 0.57$).

Birth date accuracy. To test if including individuals with estimated rather than known birthdates affected the heritability age window results, we subset the data set to 12,699 samples from 466 individuals with known birthdates and repeated the age window heritability analysis (**Methods Section 1.12**). We obtained qualitatively similar results: there was a global increase in heritability with age for windows with sufficient sample size (for $n = 8$ age windows with >1000 samples from >150 individuals per window; linear mixed model controlling for taxon, $b = 0.0013$, $p = 0.020$).

Calculating age-specific dietary diversity. To test if the increase in microbiome heritability with age could be explained by differences in feeding behaviors, we calculated age-specific dietary diversity. We calculated dietary diversity separately for the dry and wet seasons, as there are substantial seasonal effects on diet. To calculate age-specific dietary diversity, we used the same data set that we used to calculate diet diversity for the whole data set: feeding observations during focal samples on all adult females and juveniles observed by the ABRP during the duration of sample collection (not only on individuals for whom we collected microbiome data). Food items were then grouped into categories (**table S1**).

To quantify age-specific diet, we calculated diet for individual baboons (not at the level of the social group, as in the primary analysis of heritability; note that this approach limits our analysis to females and juveniles only, who are the primary targets of focal sampling in Amboseli). For each individual with diet information, we split diet observations from that individual into dry or wet season and grouped diet observations into the same overlapping age windows used for other age window analyses (e.g., ages 1-3 years, 2-4 years, and so on). We limited this data set to individual-age window combinations that contained data from at least 15 focal samples with feeding points. We then calculated the average abundance of each food category for each baboon, in each age window, in the wet and dry seasons separately, based on 1,000 random subsets of 1 point per focal sample. Dietary diversity was calculated as a Shannon's H index of these food category relative abundances, per individual baboon per age window per season. For example, a dry season dietary diversity data point for this analysis would capture a summary of diet based on the recorded consumption of foods by individual X, when she was Y to Y+3 years old, across every dry season during that age window.

For the wet season, there were 1,726 dietary age window points from 348 baboons. We ran a linear mixed model with age window (numeric; using the window starting age, Y) as a fixed effect and baboon identity as a random effect to test if dietary diversity decreased with age in the wet season, and found that it did ($b = -0.016$, $p = 1.6 \times 10^{-24}$; **fig. S16A**). The number of dietary age window points decreased with age, so we repeated the analysis after randomly subsetting the data to the same number of points per age window ($n = 57$; 798 points total), and found that diversity still significantly decreased with age ($b = -0.0074$, $p = 0.00027$). We repeated the analysis in the dry season, which had 1,686 dietary age window points from 345 baboons, and found qualitatively similar results. Specifically, in the dry season, diet diversity decreased with age ($b = -1.2 \times 10^{-2}$, $p = 2.07 \times 10^{-11}$; **fig. S16B**), and this relationship held if the data set was structured so each window contained the same number of points ($n = 60$; 840 points total; $b = -0.0099$, $p = 4.4 \times 10^{-5}$).

Calculating age-specific grooming patterns. To test if the increase in microbiome heritability with age could be explained by differences in grooming behaviors, we calculated age-specific grooming partner diversity. We calculated grooming partner diversity separately for males and females, as there are sex differences in baboon grooming behaviors, and limited the data set to individuals >4 years in age, as younger individuals primarily groom with their mothers (124). For the 270 females and 315 males in our data set, we pulled all of their grooming interactions for each year of their life (i.e. age >4-5 years, >5-6 years, etc.). Within each year, we calculated an individual's grooming partner richness as the total number of individuals they were observed to groom with, and grooming partner diversity as the grooming Shannon's H index. We then ran linear mixed models separately for each sex with the grooming metric for each baboon individual as the response variable; age in years (using the end point of each age window: i.e., 2 represents the window >1 year old – 2 years old) as the fixed effect; and social group memberships over the age-year of interest, hydrological years over the age-year of interest (e.g., a factor where "2002-2003" represents an individual whose age >1 year – 2 years old includes the hydrological years 2002 and 2003) as fixed effects, and individual identity as the random effects (results for females shown in **fig. S16C; S16D**). We found that grooming diversity decreased with age in females for both metrics (grooming Shannon's H: $b = -0.025$, $p = 1.9 \times 10^{-28}$; social partner diversity with age: $b = -0.35$, $p = 1.4 \times 10^{-19}$).

1.13 Individual effects on microbiome phenotype variation

Coefficient of variation (CV) versus heritability. To test if phenotypes that exhibit the highest heritability estimates are also more consistent within individual hosts over time, we first limited the data set to the 357 individuals with >10 samples each ($n = 15,238$ samples). For each of the 93 collapsed single-taxon phenotypes (i.e., excluding the 7 community phenotypes), we used the *raster* package to calculate the coefficient of variation (CV) of taxon relative abundance within each individual, and then calculated the mean and standard deviation of CV across all individuals (125). We likewise calculated CV for the 7 compositional phenotypes, and for the 5 Bray-Curtis PC compositional phenotypes (i.e. Bray-Curtis PC1 – PC5) we rescaled values to a 0 – 100 scale. We ran a Pearson's correlation to test if mean CV predicted heritability (**fig. S17A**).

Repeatability. We additionally tested if highly heritable phenotypes were more repeatable than less heritable or nonheritable microbial phenotypes. A phenotype’s repeatability captures the degree to which phenotypes vary between individuals due to both unspecified genetic and non-genetic factors and is an upper bound for heritability (25). The formula for repeatability is:

$$\text{repeatability} = \frac{V_{\text{Additive genetic}} + V_{\text{Individual}}}{(V_{\text{Additive genetic}} + V_{\text{Individual}} + V_{\text{Maternal identity}} + V_{\text{Technical}} + V_{\text{Residual}})}$$

We found that repeatability was strongly correlated with heritability across all microbial phenotypes, as $V_{\text{Individual}}$ was very low for all phenotypes ($R = 0.95$, $p = 6.7 \times 10^{-62}$; **fig. S17B**).

1.14 Testing how the number of samples per host affects the prevalence of significantly heritable taxa and heritability estimates

Testing the effects of cross-sectional sampling designs. Human studies on the microbiome currently rely on cross-sectional designs, with only one sample per subject. To test how longitudinal sampling affected our ability to detect heritable phenotypes, we subset our data set to 100 randomly selected cross-sectional data sets with only one sample per host. We identified significantly heritable collapsed phenotypes in each subset to calculate the average proportion of significantly heritable microbiome phenotypes across all 100 subsets. We then repeated this procedure for random subsets with 2, 5, 10, and 20 samples per host (**Fig. 4B**; **table S12**).

Testing the effects of sample size. Human studies on microbiome heritability often have < 2,000 samples (1, 4–6). To test how sample size affected our ability to estimate heritability, we subset our data set to $n = 13$ sampling depths. Sampling depths ranged from 100 to 14,000 samples, to encompass a broad range of sample sizes (**Fig. 4C, 4D**; **fig. S18**). We ran heritability models for the 100 collapsed phenotypes for 100 random subsets at each of the 13 sampling depths. Within each sampling depth, there was no correlation between the number of hosts represented in a subset and the heritability estimates (**table S15**). For each random subset, we ran the animal model using the same parameters as for the full data set, as well as the same procedure to correct for multiple testing (see **Methods Section 1.5**).

Testing the effects of sample overlap. To test if increasing concordance of h^2 estimates with increasing sample size is simply a function of increased overlap with the full data set, we randomly split the data set into two non-overlapping subsets of 8,117 samples and ran heritability models. We found highly concordant estimates between the subsets ($R = 0.92$, $p = 9.5 \times 10^{-44}$).

1.15 Robustness of model selection

We performed multiple tests to evaluate the robustness of our conclusions. Specifically, we tested if assortative mating, maternal effects, grooming networks, or technical effects could explain our heritability estimate results. We ran the following checks:

Assortative mating effects on heritability estimates. Assortative mating can inflate heritability estimates (126), and baboons are known to display some degree of assortative mating by rank and genetic ancestry (127). To demonstrate that assortative mating is unlikely to be inflating our heritability calculations, we tested if pairs of individuals in our data set who are parents (i.e., have had an offspring together) have more similar microbiomes to each other than the non-parent female-male pairs in the data set. We calculated the mean lifetime microbiome Bray-Curtis dissimilarity between all female-male pairs of individual baboons in our data set and ran a Mantel test on that matrix to test for a correlation with a 1-0 matrix indicating if individual female-male pairs were parents or not parents (**fig. S7B**).

Maternal effects on heritability estimates. Maternal identity is not known for some individuals in our data set. To ensure that our results were robust to error in the maternal effects estimates, we performed several checks. First, we limited the data set to individuals with known mothers ($n = 13,780$ samples). For the 100 collapsed phenotypes, we then ran 4 animal model iterations that used the same fixed effects as the models that included the full data set ($n = 16,234$ samples), but made the following changes to the random effects components: Model 0 did not include the additive genetic effects term or maternal identity, Model P+M included both the additive genetic effects term and maternal identity, Model P included the additive genetic effects term but not maternal identity, and Model M included maternal identity but not the additive genetic effects term. Models were compared using likelihood ratio tests and a 10% FDR for each analysis. If the additive genetic effects term was included, adding maternal identity improved 0 / 100 models (Model P = Model P+M). If the additive genetic effects term was not included, adding maternal identity improved 70 / 100 models (Model 0 < Model M). If maternal identity was included, adding the additive genetic effects term improved 95 / 100 models (Model M < Model P+M). Further, heritability estimates did not significantly change if maternal identity was included versus excluded as a random effect (paired t-test, $p = 0.16$; **fig. S7A**).

If the full data set is used (16,234 samples), the heritability estimates do not significantly change when maternal identity is included as a random effect in the model (paired t-test, $p = 0.35$; **fig. S7A**; note that the animal model requires 2+ offspring per mother to adequately partition maternal variance (25)). If we subset to the 439 baboons ($n = 12,784$ samples) with at least 1 maternal sibling in the data set, we still found that adding maternal identity did not improve any models, suggesting that maternal effects are unlikely to explain the heritability estimates we observe in our current analyses.

Social network effects on yearly heritability estimates. Previous work in Amboseli has shown that close grooming partners have more similar microbiomes than individuals who groom less often (18, 128). Because grooming networks are also positively correlated with pedigree relatedness in our data set (Mantel test per hydrological year; $R = 0.44 - 0.475$, $p < 0.001$), heritability estimates could be inflated because relatedness also captures information about grooming behavior (i.e., positive gene-environment correlation). Our data set is not fine-grained enough at short time scales to test social network effects in the same manner as previous studies

in Amboseli, which collected microbiome samples from all adult group members within ~1 month's time and thus could calculate the same year-long grooming network for all samples (18, 128). Instead, to assess the possibility that heritability estimates could be influenced by grooming behaviors, we included a grooming network matrix over one hydrological year as the variance-covariance of an environmental random effect in the animal model, following (129)'s approach for running multiple-matrix animal models in *asreml*. This analysis also limited the data set to individuals for whom we had data on grooming relationships for the full hydrological year in question. Below, we describe how grooming data were collected, how yearly grooming networks were calculated, and the results from including grooming networks in the animal model.

Grooming interactions are recorded as part of representative interaction sampling during all group monitoring visits (38). During representative interaction sampling, observers move through the group as they perform focal animal samples on a randomized rotation of animals. The observer remains with each subject for 10 minutes and records all observations of physical contact for all individuals within their line of sight (including grooming interactions observed for animals who are not the subject of the focal sample). After 10 minutes, the observer moves to a new subject, ensuring representative sampling of the whole social group. For each grooming interaction, the observer records the identity of the groomer and the individual being groomed.

We used these grooming interaction data to calculate a network of grooming bond strength, a dyadic grooming metric following (18). Because baboon social networks are dynamic, we calculated a separate grooming network for each social group for each hydrological year in the data set, limited to years with at least 150 baboons who were resident for the full year and had at least 1000 total microbiome samples collected over the year (years 2002, 2003, 2007-2009). Continuous group residency was defined as being a member of the same group for at least 360 days of a calendar year, excluding years with group fissions or other data gaps that make assigning residency difficult. Baboons only groom other members of their social group, so within each hydrological year, separate grooming networks were constructed for each social group and then concatenated such that all between-group grooming scores were 0. The number of grooms between pairs of continuous group residents within each social group within each year was counted (i.e. group-years) and scaled to the highest number of interactions within each group-year such that the highest value in each group-year was 1. Hence, all grooming bonds were represented as a proportional strength relative to the strongest bond in their group in the hydrological year of interest.

We ran the animal model with the grooming network matrix for the 100 collapsed phenotypes, limited to the 5 years that matched the criteria described above (5 years * 100 phenotypes = 500 models run). We found that adding grooming networks to models *without* the pedigree matrix improved 0/500 models after correcting for multiple testing (Model 0 = Model G). Further, adding grooming networks to models *with* the pedigree matrix improved 0/500 models after correcting for multiple testing (Model P = Model P + G). In contrast, adding pedigree data to models that included grooming networks improved 145/500 models after correcting for multiple testing (LRT FDR $p < 0.1$; **fig. S7**; Model G < Model G + P). When grooming is included in yearly models, the heritability estimates show a slight decrease of 0.0051 on average, suggesting that the effect of grooming network, if any, is much smaller than the effect of genotype sharing. Consistent with this interpretation, we found that grooming

network effects explained, on average, 1.2% of the variation in the above-mentioned 145/500 significantly heritable microbiome phenotypes models, compared to 14.6% explained by additive genetic variance. These results indicate that pedigree relatedness is much more likely to drive our results than grooming networks. We note that because our current study is longitudinal, and microbiome samples from grooming partners were often collected weeks or months apart (versus all microbiome samples collected in the same ~1 month as in previous studies), the time between samples may be too great to detect strong social effects on microbiome composition.

Excluding variance explained by technical effects in calculating total phenotypic variance (V_P). Variance components in the animal model generally reflect aspects of an individual's environment and kin relationships. We included a technical component, DNA extraction plate, in our calculation of V_P because it explained substantial variance in some microbial phenotypes. To rule out the possibility that the inclusion of this technical effect influenced our interpretation of h^2 in the full or seasonal models, we conducted the following tests.

First, we tested if excluding $V_{\text{Technical}}$ affected our overall heritability estimates in the full (16,234 sample) data set for the 100 collapsed phenotypes. If heritability was calculated using the below formula, h^2 estimates were highly correlated with those estimated when $V_{\text{Technical}}$ was included ($R = 0.98$, $p < 2.2 \times 10^{-16}$). Excluding $V_{\text{Technical}}$ only increased h^2 estimates by 0.005 on average.

$$h^2 = \frac{V_{\text{Additive genetic}}}{(V_{\text{Additive genetic}} + V_{\text{Individual}} + V_{\text{Maternal identity}} + V_{\text{Residual}})}$$

Second, we tested if excluding $V_{\text{Technical}}$ affected seasonal differences in V_P . We found that wet season V_P was still significantly higher than dry season V_P if $V_{\text{Technical}}$ was excluded (paired t-test; $p = 1.5 \times 10^{-4}$).

Supplementary Figures

Table of Contents

- Fig. S1.** Pedigree relationships, organized by birth year.
- Fig. S2.** Sample age does not predict DNA concentration (ng/uL).
- Fig. S3.** Technical replicates cluster with each other rather than with their sequencing plate.
- Fig. S4.** Number of reads per sample in the 16,234 samples in the full dataset.
- Fig. S5.** Distribution of alpha diversity across samples.
- Fig. S6.** Prevalence and abundance of 1,102 microbiome taxa that were identified at the genus through phylum levels (ASVs are excluded).
- Fig. S7.** Heritability estimates are robust to maternal effects, grooming-based social networks, and assortative mating.
- Fig. S8.** Ordination plots illustrating the effects of (A) host sex, (B) season, (C) social group membership, and (D) host identity, without accounting for time.
- Fig. S9.** Heritability estimates are robust to data transformations, and heritability is consistently higher for presence/absence phenotypes than single-taxon phenotypes.
- Fig. S10.** Distribution of permuted heritability estimates versus observed heritability estimates.
- Fig. S11.** Evolutionary relationships between taxa predict their heritability estimates.
- Fig. S12.** Heritability estimates for non-microbiome primate traits.
- Fig. S13.** A co-occurrence network analysis (SparCC) run for 31 microbial families found in >50% of samples.
- Fig. S14.** To make sure that heritability analyses were not conducted on redundant, phylogenetically nested taxa, we identified 100 *collapsed phenotypes*.
- Fig. S15.** Seasonal differences in heritability estimates likely reflect differences in environmental diversity.
- Fig. S16.** Diet and grooming partner diversity decrease with baboon age.
- Fig. S17.** Microbiome phenotypes vary substantially over individual host's lifetimes.
- Fig. S18.** Heritability estimates vary depending on sampling depth.

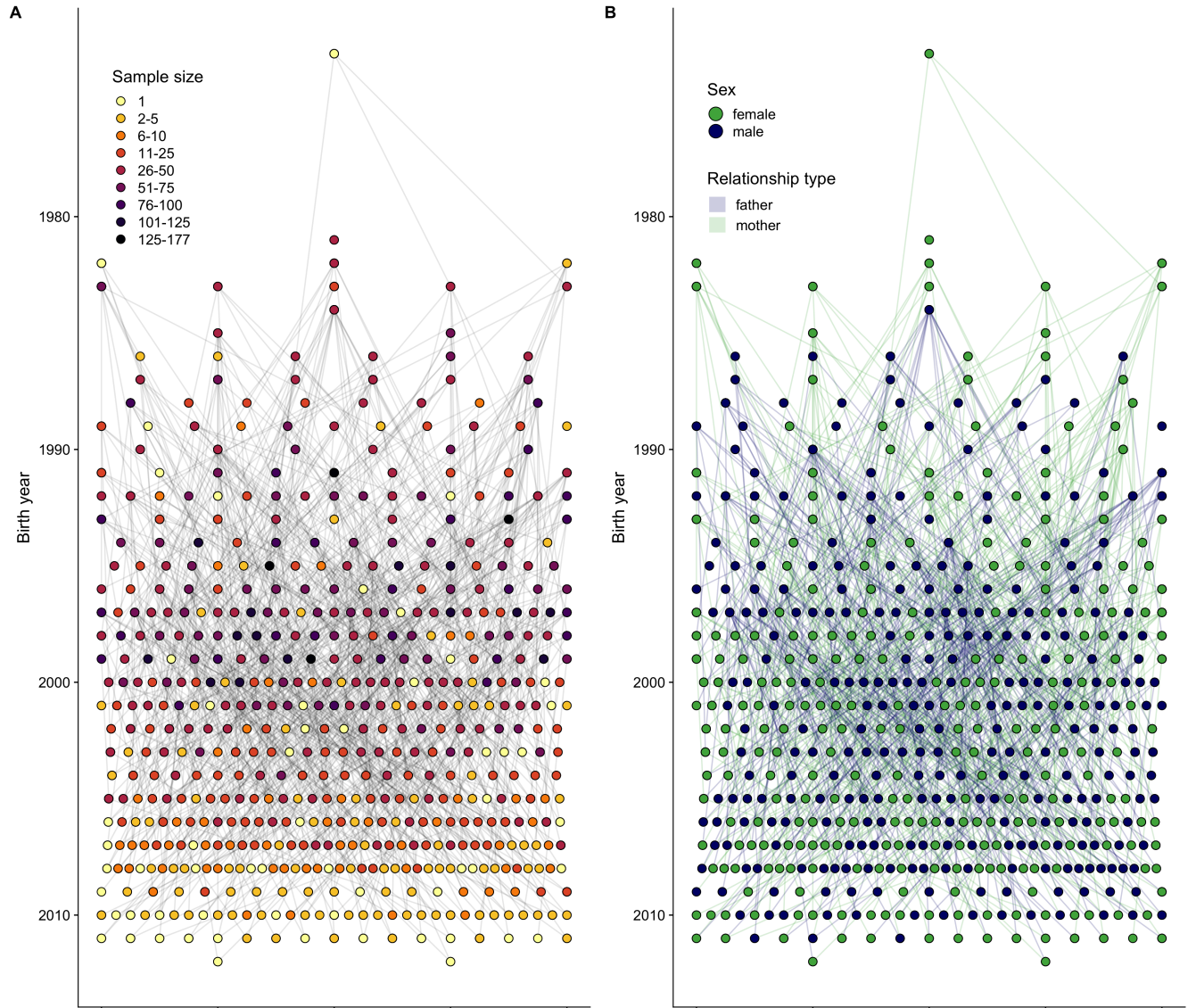


Fig. S1. Pedigree relationships, organized by birth year. In (A) and (B), each dot represents one of the 585 individual baboons in the data set, and lines indicate maternal and paternal parent-offspring relationships. (A) Colors indicate the number of microbiome samples analyzed from each individual; (B) Colors indicate the sex and relationship type.

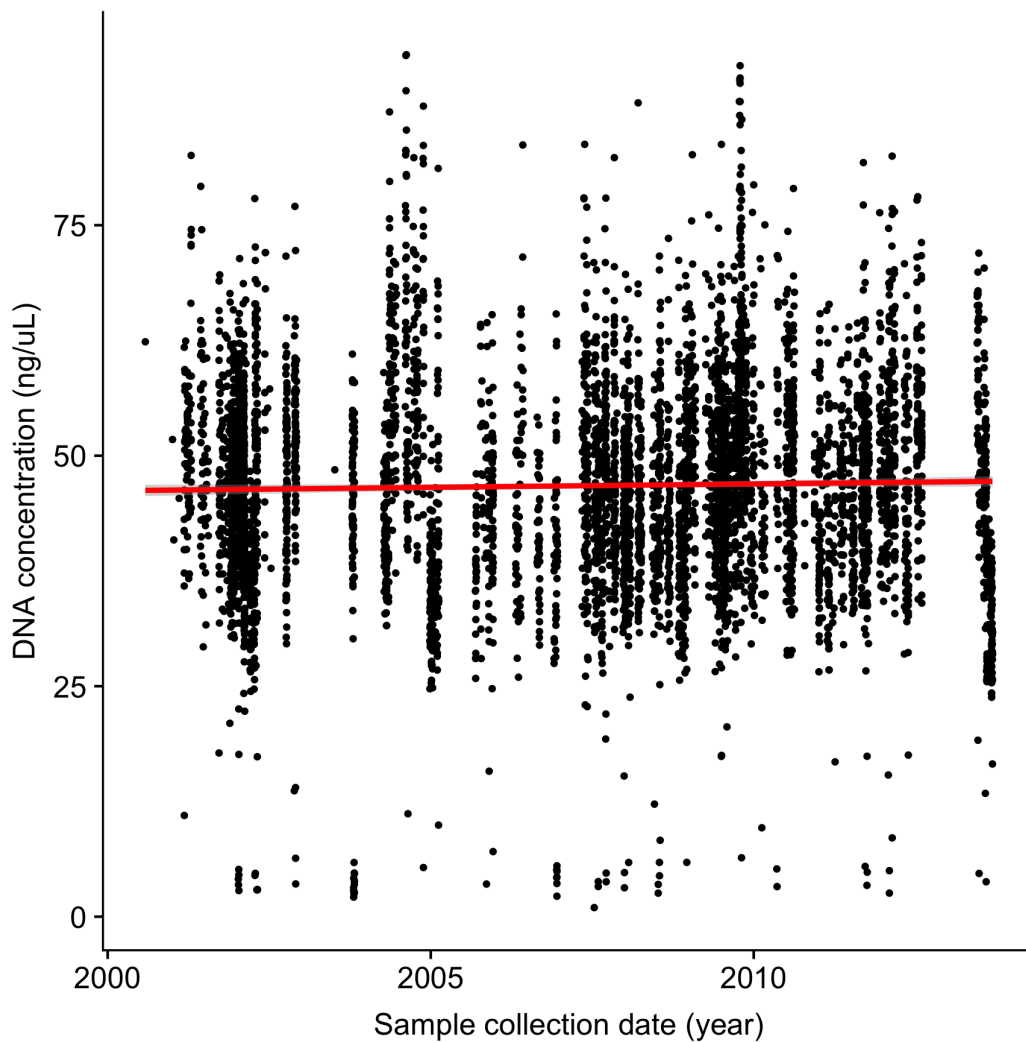


Fig. S2. Sample age does not predict DNA concentration (ng/uL). We found no relationship between storage time for each sample (i.e., collection date) and DNA concentration, indicating that older samples were not substantially more degraded ($b = 2.0 \times 10^{-4}$, $p = 0.064$).

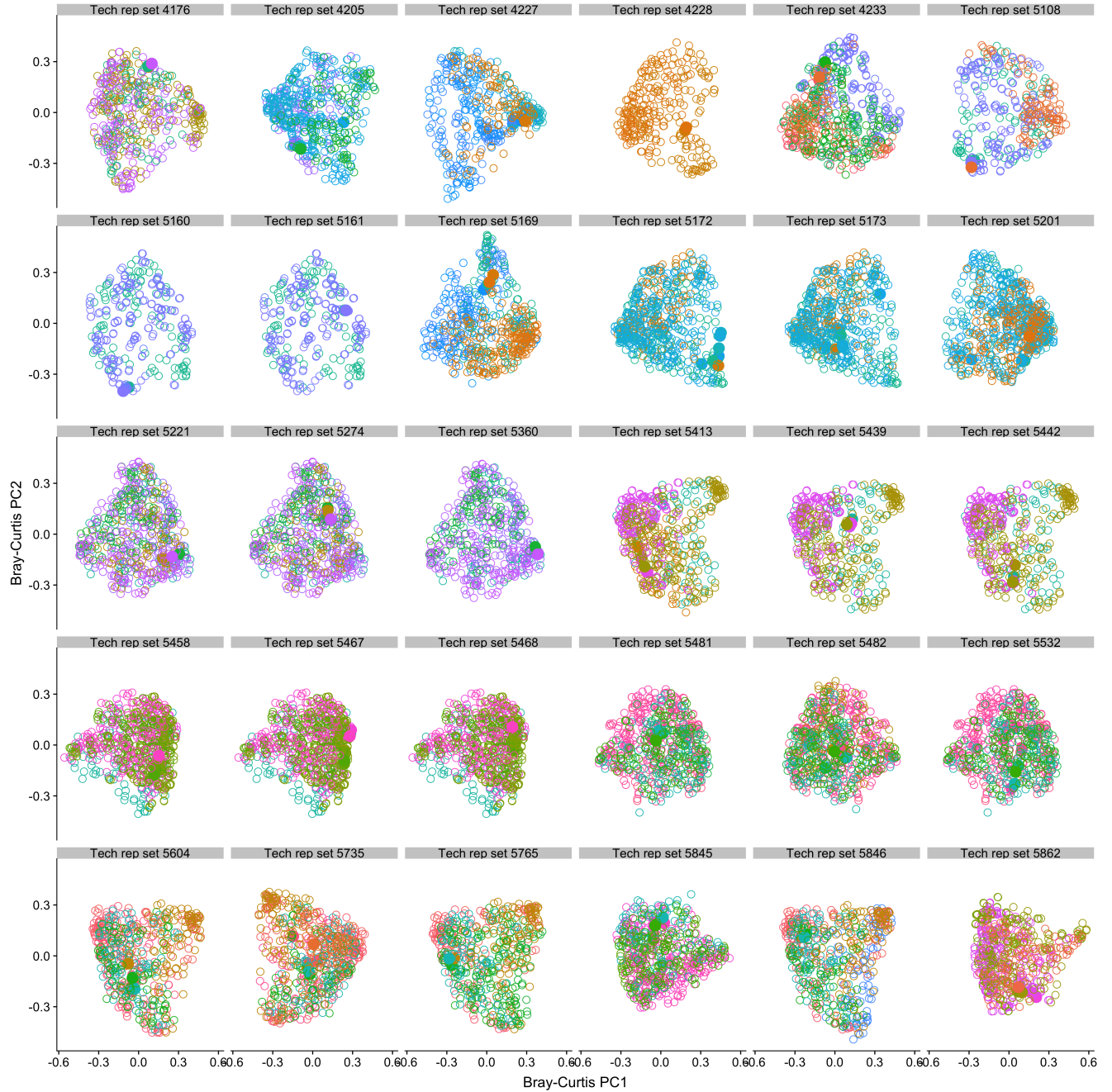


Fig. S3. Technical replicates cluster with each other rather than with their sequencing plate. Ordination plots show the first two principal coordinates of a Bray-Curtis dissimilarity matrix. Each plot depicts a set of technical replicates (i.e., the same fecal sample subject to multiple extractions on different DNA extraction plates; facet labels are the replicate set identity number). Filled circles indicate the technical replicates and open circles indicate the rest of the samples on the sequencing plate. For each plot, the Bray-Curtis matrix is calculated only from the plates on which the technical replicates are located. Colors are sequencing plates.

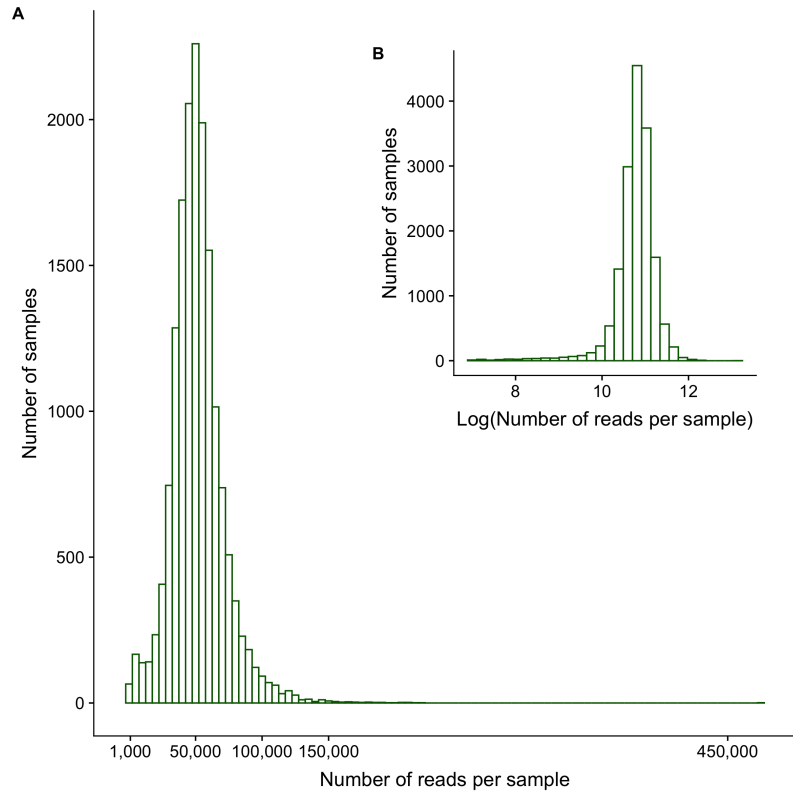


Fig. S4. Number of reads per sample in the 16,234 samples in the full dataset, plotted on (A) a normal scale and (B) a log scale.

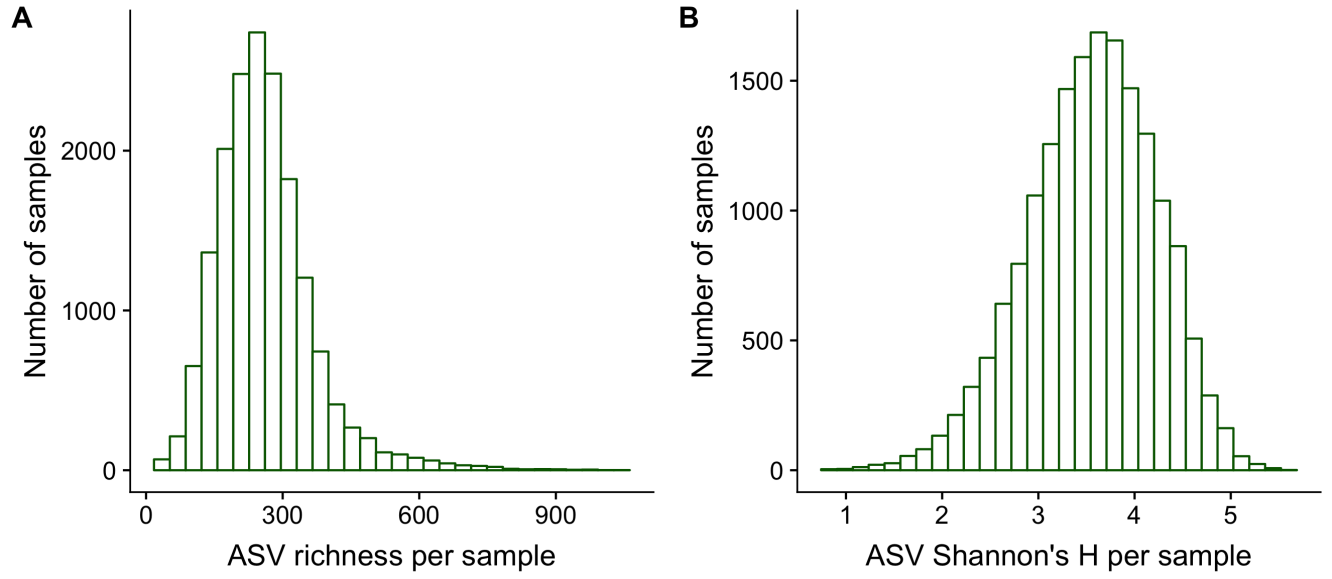


Fig. S5. Distribution of alpha diversity across samples. (A) Distribution of ASV richness. (B) Distribution of ASV Shannon's H index.

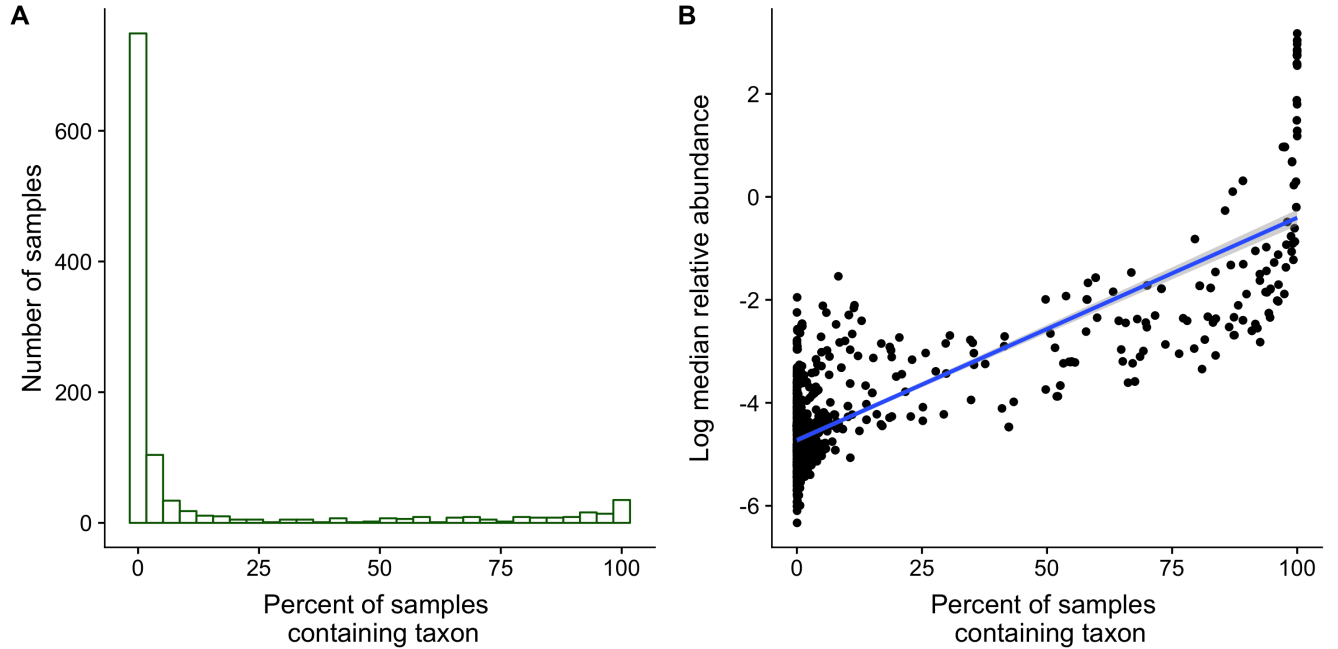


Fig. S6. Prevalence and abundance of 1,102 microbiome taxa that were identified at the genus through phylum levels (ASVs are excluded). (A) Prevalence of all 1,102 identified taxa. 75% of these taxa were found in <5% of samples. 144 of these taxa were found in >50% of samples and comprise part of the 283 single-taxon phenotypes (the remaining 139 taxa are ASVs, which are not plotted here). (B) The prevalence and average per-sample abundance were correlated across all 1,102 taxa (Pearson's $R = 0.43$, $p < 0.001$). Y-axis values were calculated only for those samples in which the taxon was detected. The average per-sample relative abundance of most taxa was low: 75% of taxa comprised 0.03% of the reads in a sample on average.

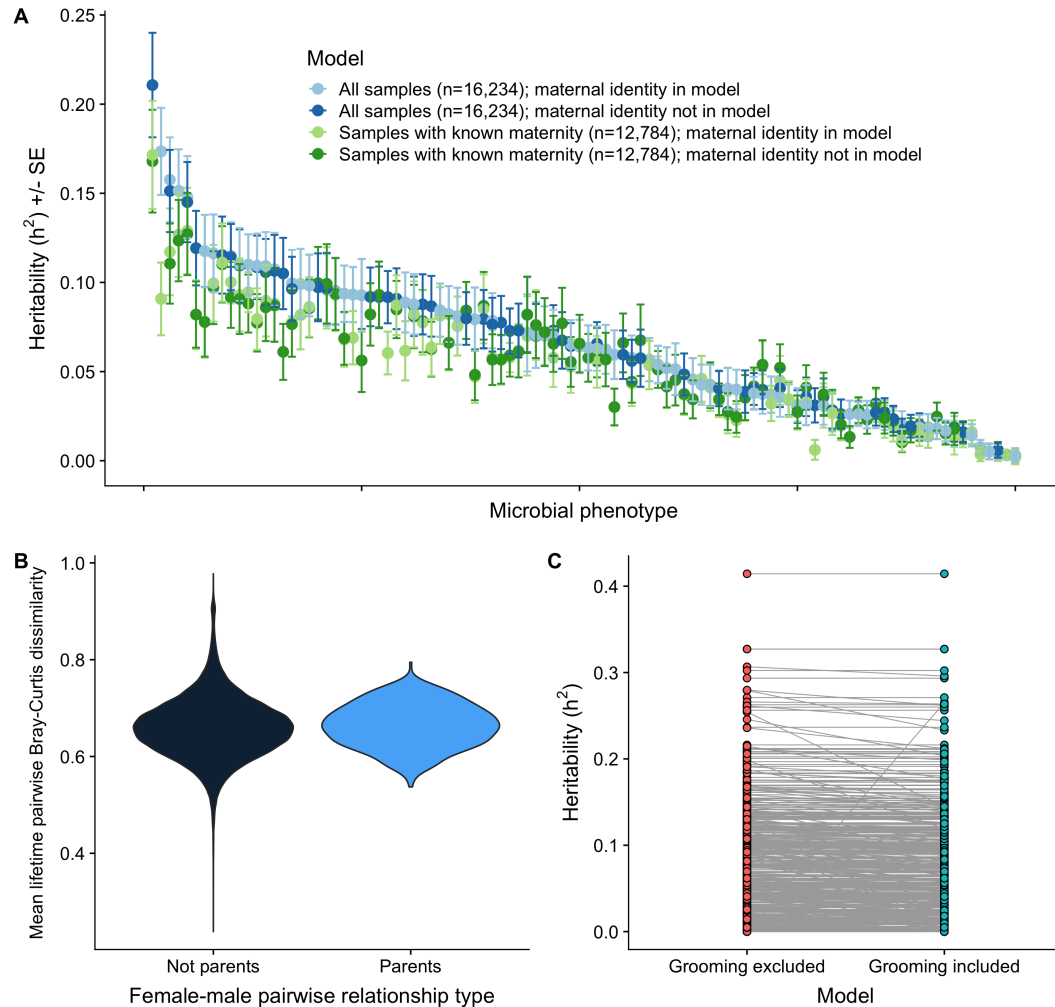


Fig. S7. Heritability estimates are robust to maternal effects, grooming-based social networks, and assortative mating. (A) Heritability estimates do not significantly change if maternal identity is included or excluded from models. This pattern holds true if the dataset is limited to individuals with known maternal identity (green), or if the full dataset is used (blue). The x-axis is ordered by heritability estimate size for the $n=16,234$ samples with maternal identity included. (B) Assortative mating does not predict microbiome similarity. Specifically, pairs of individuals who were parents do not have more similar lifetime microbiome compositions than female-male pairs of individuals who were not parents (Mantel $R = 0.004$, $p = 0.22$). (C) Grooming effects do not significantly change heritability estimates for the 100 collapsed phenotypes (see **table S5**). Specifically, adding a grooming network matrix to a subset of year-long models (i.e. the 5 years with at least 150 individual baboons who had been resident for the entire year and at least 1000 total samples; See **Supplement Section 1.15**) did not significantly improve any models after correcting for multiple testing. Adding a grooming network matrix compared to not accounting for grooming relationships decreased heritability estimates by only 0.0051 on average.

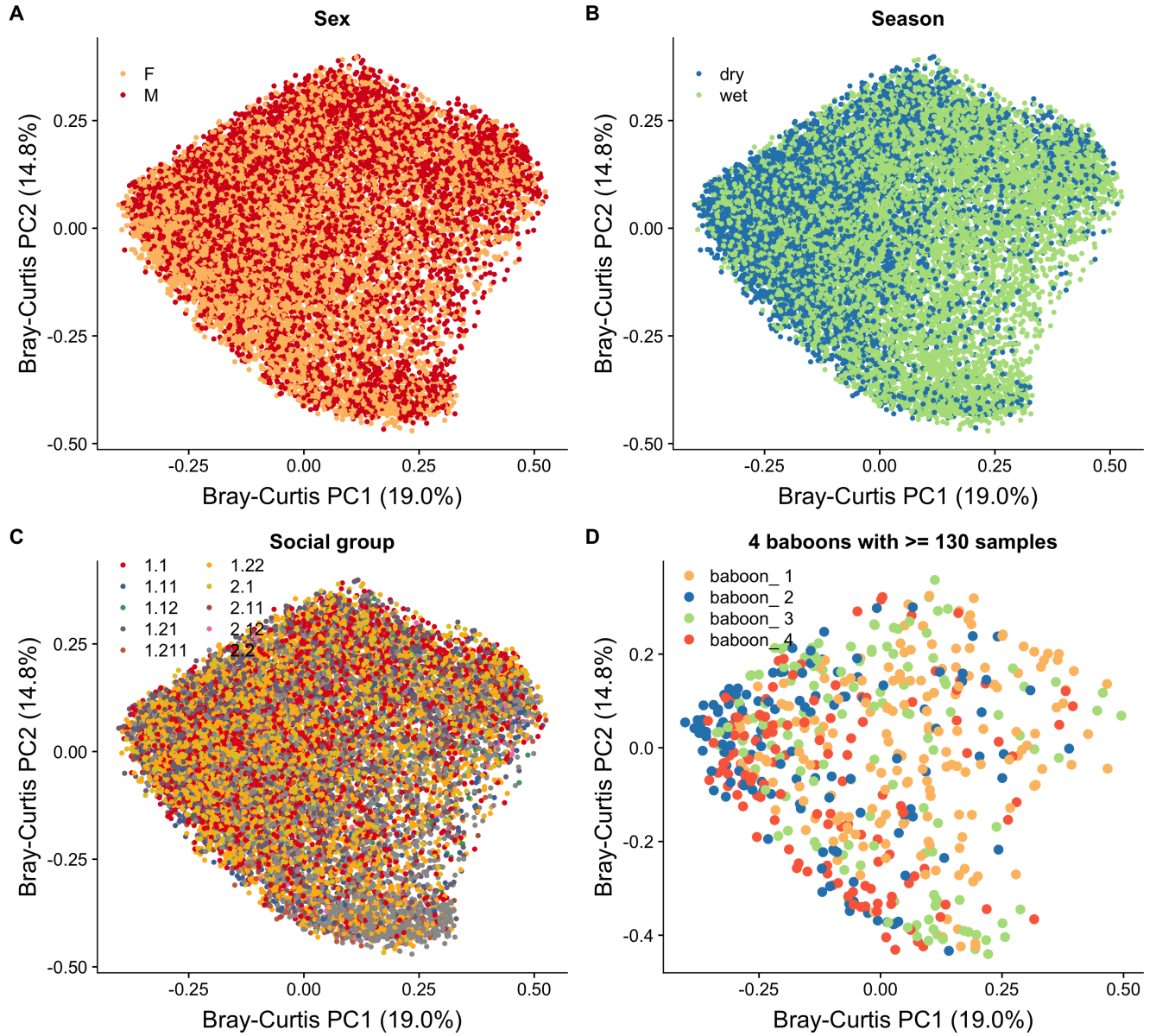


Fig. S8. Ordination plots illustrating the effects of (A) host sex, (B) season, (C) social group membership, and (D) host identity, without accounting for time. Plots A-C show all 16,234 samples spanning 14 years; each point is a sample; axes represent the first two principal coordinates of a Bray-Curtis dissimilarity matrix. Colors indicate categories within each fixed effect.

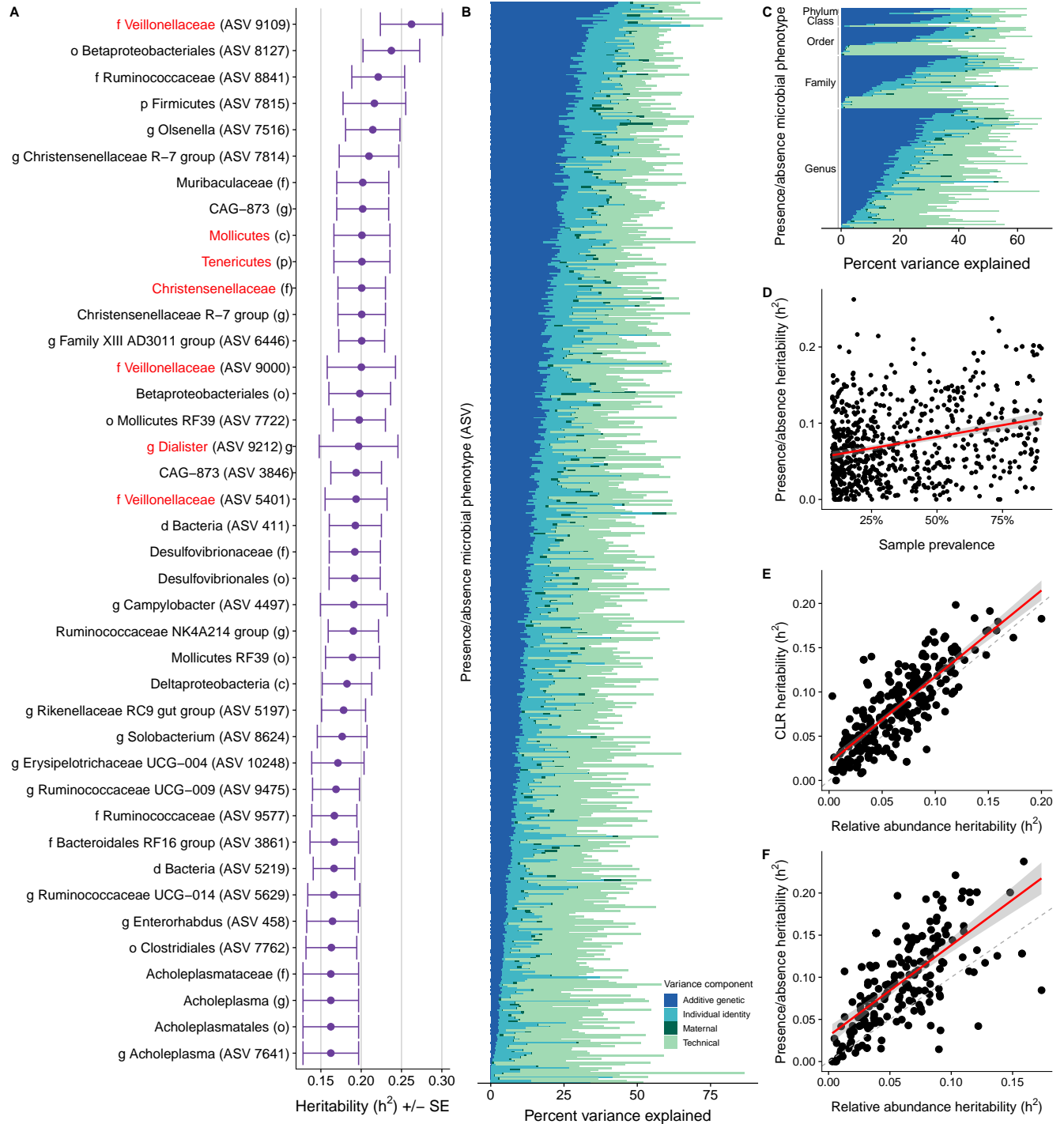


Fig. S9. Heritability estimates are robust to data transformations, and heritability is consistently higher for presence/absence phenotypes than single-taxon phenotypes. (A) Heritability estimates for the top 40 most heritable presence/absence phenotypes, parallel to Fig. 2A in the main text, which shows parallel results for single-taxon phenotypes. Phenotypes that have previously been shown to be heritable in humans are shown in red text (1,2,4-6). (B/C) In the 744 presence/absence phenotypes (divided into (B) ASVs and (C) taxa at the genus through phylum levels; the legend in B applies to both panels), the amount of variance explained by additive genetic variance was significantly higher than that explained by host identity or maternal effects. The y-axis is ordered by heritability within each taxonomic level, as given in **table S7**. This panel shows the same information for presence/absence phenotypes as **Fig. 2C** shows for single-taxon phenotypes. (D) The

prevalence of a given presence/absence phenotype is positively correlated with its heritability estimate ($n = 744$ phenotypes; Pearson's $R = 0.38$, $p = 2.3 \times 10^{-15}$). (E) Heritability estimates are correlated for the CLR (centered log ratio) and relative abundance transformations of the 283 single-taxon phenotypes (Pearson's $R = 0.82$, $p = 2.3 \times 10^{-69}$). (E) Heritability estimates are correlated but higher for presence/absence phenotypes than for single-taxon phenotypes for taxa in both model sets ($n=203$ taxa at the ASV through phylum levels found in 50 - 90% of samples; Pearson's $R = 0.68$, $p=3.2 \times 10^{-29}$). In D and E, the red line represents the linear fit and the dashed grey line represents $x=y$.



Fig. S10. Distribution of permuted heritability estimates versus observed heritability estimates. To determine if the observed heritability (h^2) estimates were different than those obtained by chance, we randomly permuted the baboon identities 1000 times and re-ran the heritability models for the 100 collapsed phenotypes. Each facet is a microbial phenotype. Red stars indicate observed heritability estimates and black histograms indicate the distribution of 1000 permuted heritability estimates.

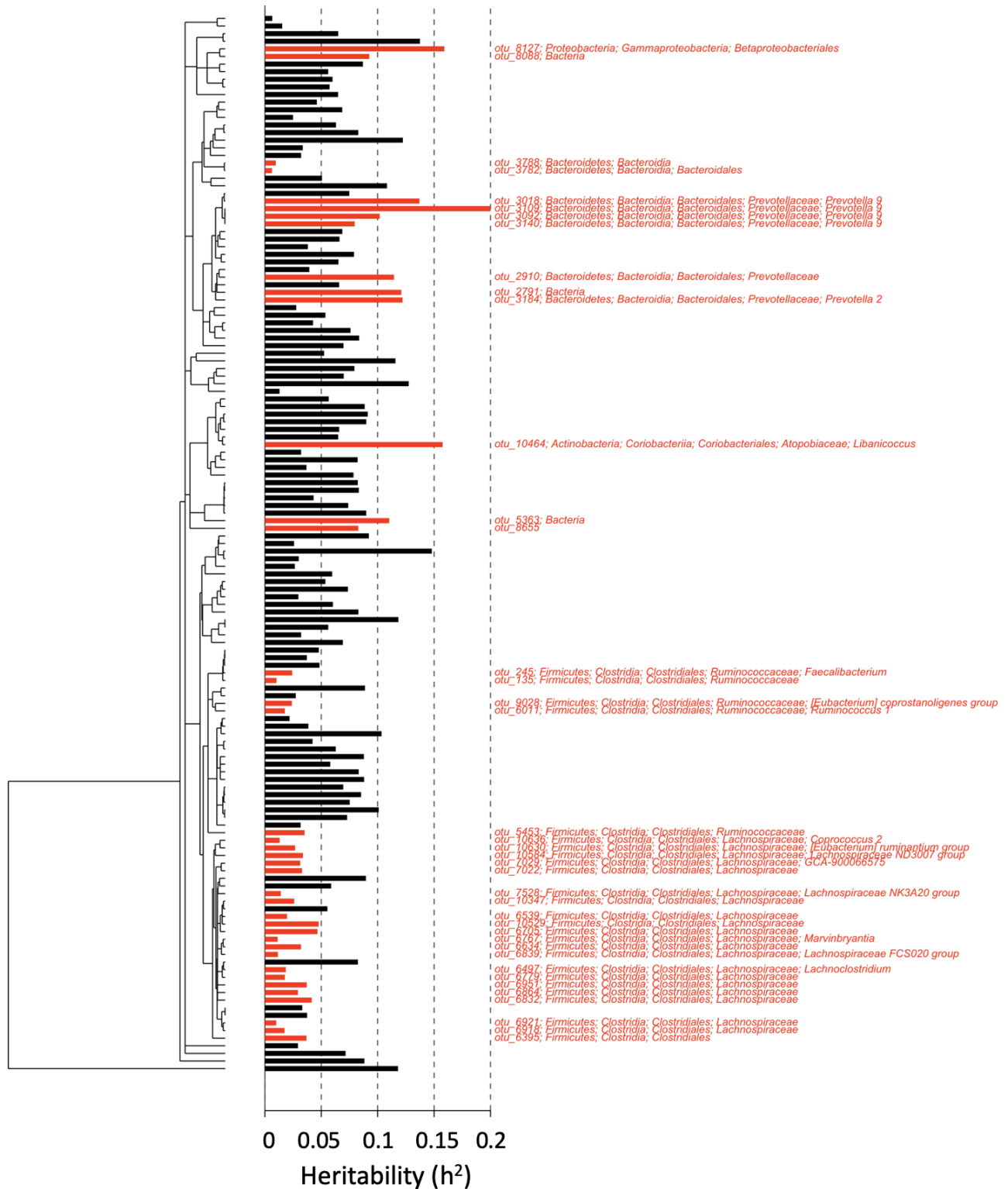


Fig. S11. Evolutionary relationships between taxa predict their heritability estimates. The tree was constructed from $n=139$ ASVs found in $>50\%$ of microbiome samples. Red lines indicate ASVs whose relative abundances have significantly phylogenetically structured heritability (Local Moran's $I p < 0.05$), and red text describes their taxonomic classification.

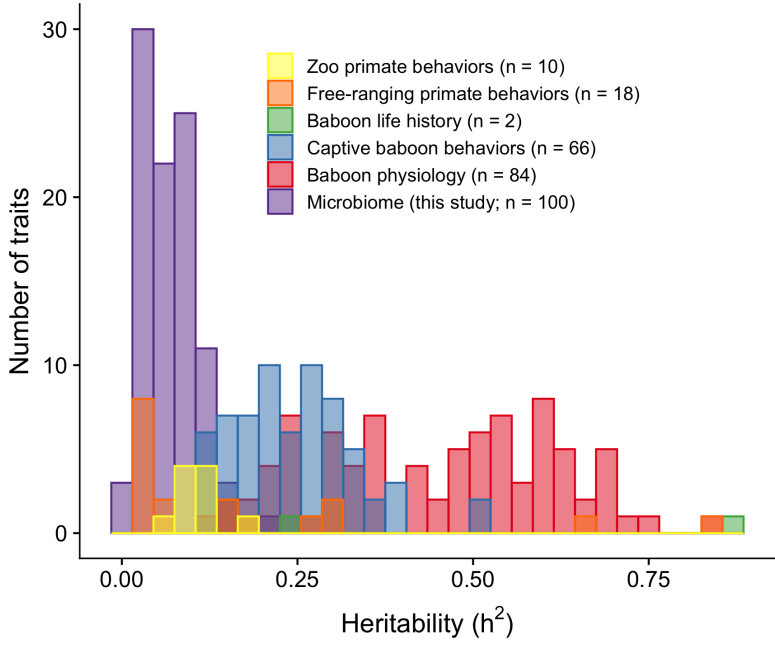


Fig. S12. Heritability estimates for non-microbiome primate traits. Heritability estimates for baboon microbiome traits in the current study (100 collapsed phenotypes shown) are similar to heritability estimates for complex behavioral traits in free-ranging and zoo-housed primates and, as expected, lower than heritability estimates for physiological traits in baboons (see full list of traits and studies in **table S10**)

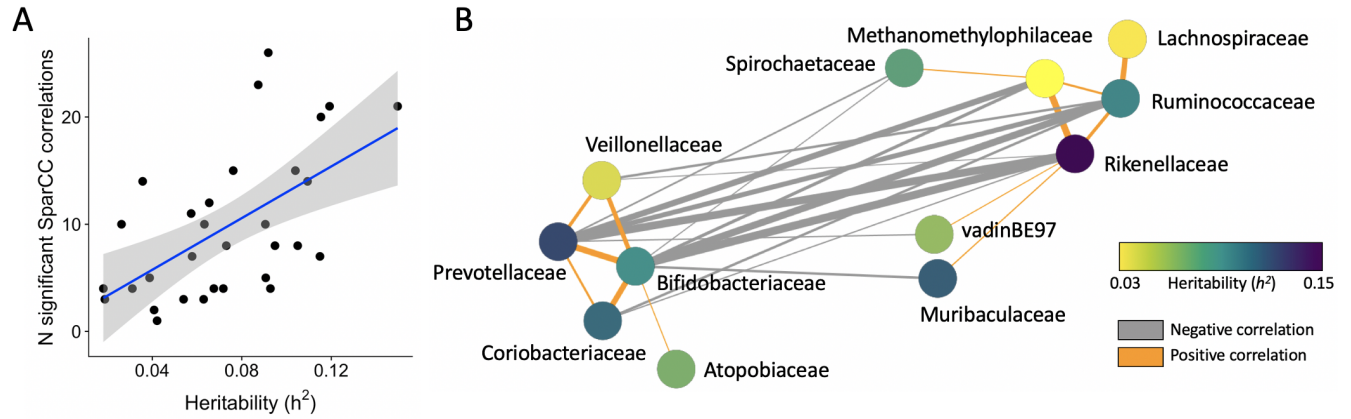


Fig. S13. A co-occurrence network analysis (SparCC) run for 31 microbial families found in >50% of samples. (A) There is a positive correlation between a family's heritability and co-occurrence network connectivity, such that more connected taxa were also more heritable (Pearson's $R = 0.58$, $p = 0.006$). (B) Heritable families in baboons do not form discrete clusters.

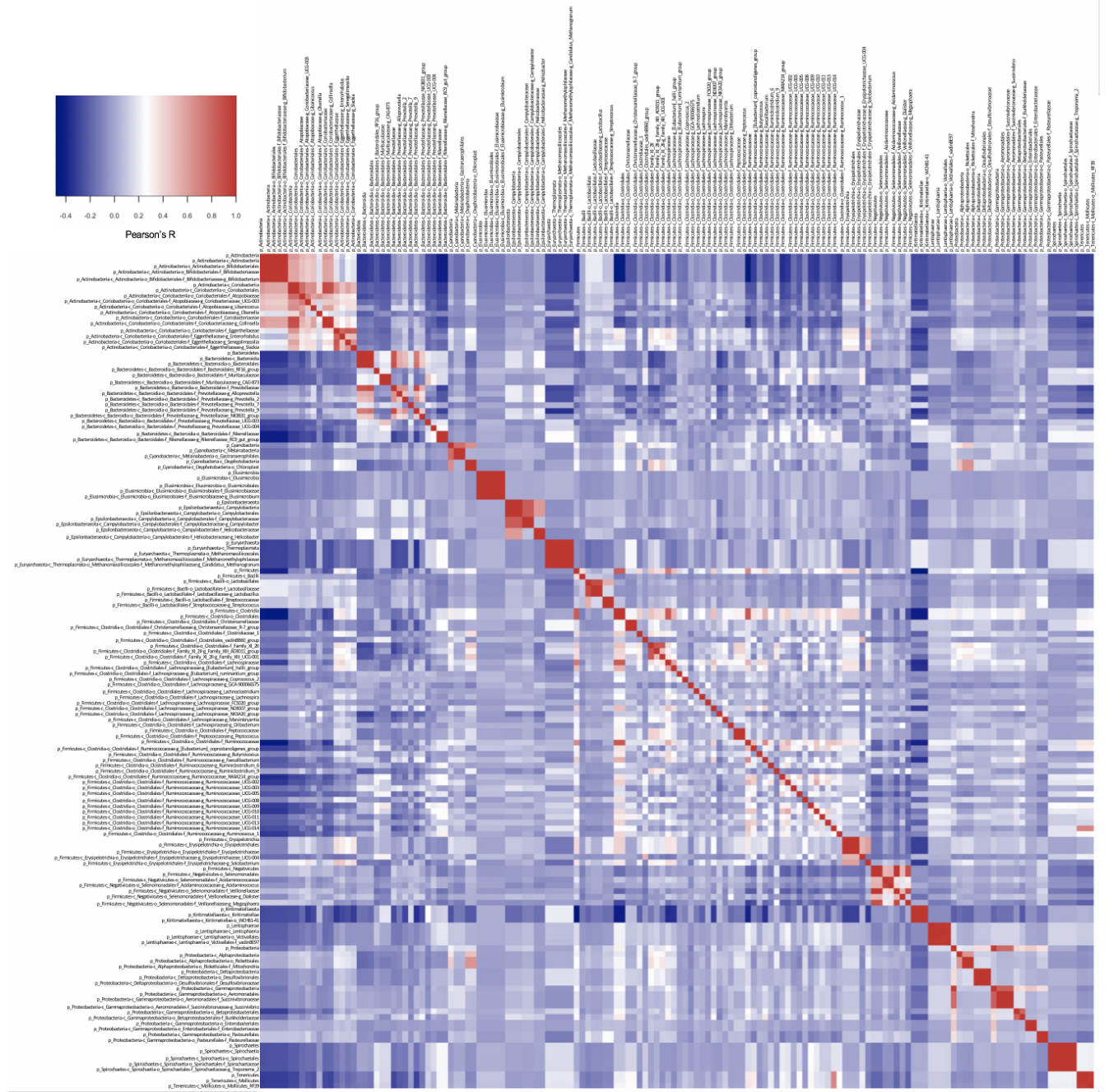


Fig. S14. Phylogenetically nested taxa were collapsed into 100 collapsed phenotypes. To identify these phenotypes, we calculated the pairwise Pearson correlations between the relative abundance of 144 taxa found in >50% of samples (excluding ASVs; i.e. the single-taxon phenotypes at the genus through phylum levels). This heat map shows the pairwise Pearson correlations between these 144 taxa. Of these, 86 nested taxon pairs were correlated at >0.97 (e.g. the genus *Christensenellaceae_R-7_group* is nested within the family Christensenellaceae). Some pairs overlapped with each other (e.g. the family Veillonellaceae within the order Selenomonadales, and the order Selenomonadales within the class Negativicutes; see **table S13** for details). For these nested taxa, the lowest level taxon was used as the representative taxon for the 93 single taxa in the 100 collapsed phenotype dataset (the remaining 7 phenotypes in this data set are community phenotypes). Cells are ordered by taxonomy. Colors indicate correlation between taxon relative abundances, with values ranging from Pearson's R = -0.47 (dark blue) to 1.0 (self; dark red).

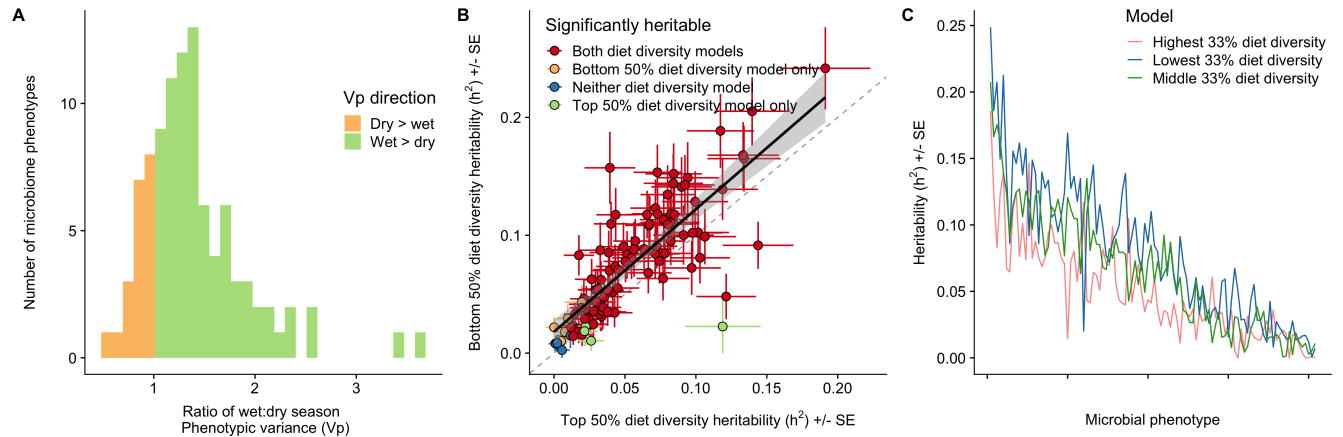


Fig. S15. Seasonal differences in heritability estimates likely reflect differences in environmental diversity. (A) Total phenotypic variance is higher in the wet season than in the dry season for most of the 100 collapsed phenotypes. (B) Heritability estimates are systematically higher for samples with low dietary diversity. This is shown by the correlation between heritability estimates for the 100 collapsed phenotypes measured in samples with high (top 50%) diet diversity, as measured by a dietary Shannon's H index, in the preceding 30 days (x-axis) compared to low (bottom 50%) diet diversity in the preceding 30 days (y-axis). The grey, dashed line represents $x=y$, data points represent individual taxa in the collapsed phenotype data set, and colors indicate whether the taxon was significantly heritable in a given set of models. (C) The differences shown in (B) are consistent when diet diversity is divided into three levels instead of two (as in (B)). Phenotypes are ordered by heritability estimates in the full dataset.

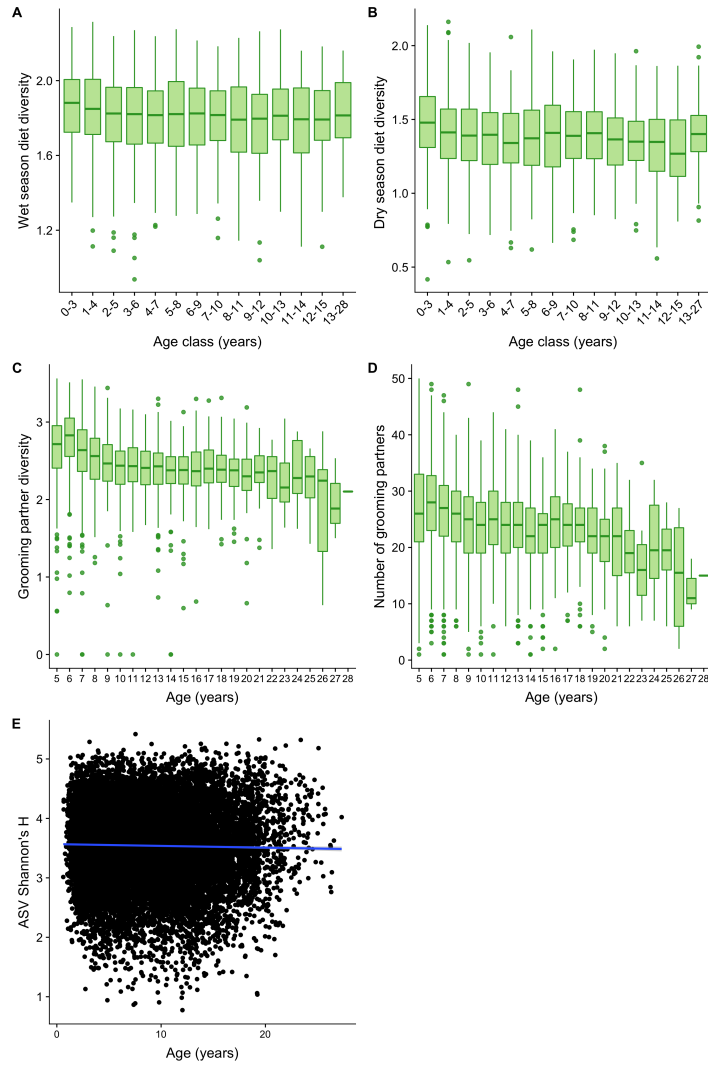


Fig. S16. Diet and grooming partner diversity decrease with baboon age. (A) Wet season diet (linear mixed model controlling for baboon identity; wet season $b = -1.6 \times 10^{-2}$, $p = 1.6 \times 10^{-24}$). (B) Dry season diet ($b = -1.2 \times 10^{-2}$, $p = 2.0 \times 10^{-11}$). Each box plot shows the distribution of individual dietary Shannon's H index values for all feeding observations by one individual baboon within the corresponding age window. (C) Grooming partner Shannon's H diversity in female baboons (linear mixed model controlling for baboon identity, social group membership, and sampling years; $b = -0.025$, $p = 1.9 \times 10^{-28}$). (D) Number of grooming partners in female baboons ($b = -0.35$, $p = 1.4 \times 10^{-19}$). (E) ASV Shannon's H decreases slightly with age (linear mixed model; $b = -0.0063$, $p = 0.024$).

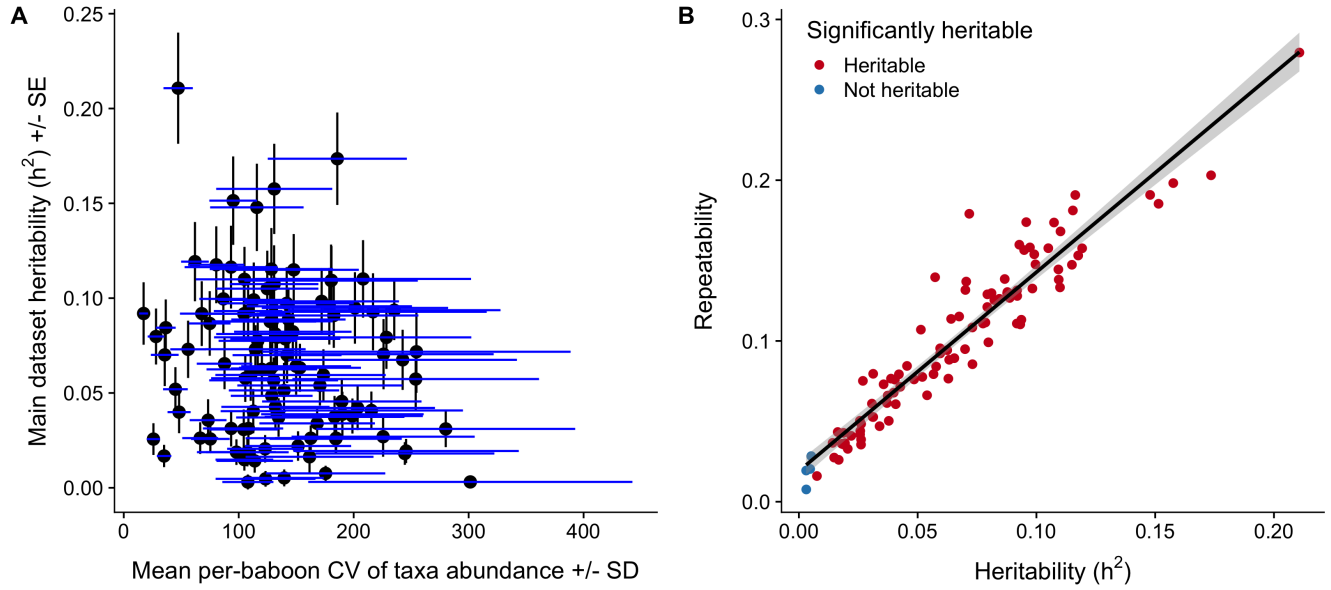


Fig. S17. Microbiome phenotypes vary substantially over individual host's lifetimes. (A) The heritability estimates of single-taxon phenotypes are not correlated with the mean lifetime coefficient of variation (CV) for those phenotypes ($R = -0.14$, $p = 0.15$). CV was calculated by limiting the dataset to baboons with >10 samples each ($n=357$ baboons; 15,238 total samples), then calculating the CV within each baboon individual. The mean and standard deviation for CV was calculated across individuals for each phenotype. (B) Heritability and repeatability estimates are highly correlated because individual identity explains a low, but consistent amount of phenotypic variance. Each point is one of the 100 collapsed phenotypes, and the line represents a linear fit ($R = 0.95$, $p = 6.7 \times 10^{-62}$). Colors indicate heritability in the 100 collapsed phenotypes (red is significantly heritable, blue is not heritable, 10% FDR).



Fig. S18. Heritability estimates vary depending on sampling depth. Each facet is one of the 100 collapsed phenotypes. Sampling depth is shown on the x-axis. The range of heritability estimates from 100 subsamples at each sampling depth is shown on the y-axis. Colors indicate the percentile (e.g., dark green represents the middle 50% of permutations).

Supplementary Tables

Table S1. Dietary food items in each diet category.

Table S2. Principal components (PCs) of dietary composition reflect seasonal tradeoffs in food availability. We collapsed diet composition into dietary PCs by running a principal components analysis (PCA) on the relative abundance of 14 dietary food categories (Table S1). The loadings column indicates the two foods with the two highest loading scores on the dietary PC.

Table S3. Sequencing reads at each step in the sample processing pipeline.

Table S4. A list of the fixed and random effects used in the ‘animal model’ as implemented in ASReml-R v3 to produce the results shown in Figure 2 in the main text. Diet PCs are defined in Table S2. See Table S5 for a list of which fixed and random effects were included in each iteration of the animal model.

Table S5. List of fixed and random effects included in the heritability models for each analysis and the corresponding figures arising from these analyses. See Supplement for a description of each model. Response phenotypes are defined in the main text.

Table S6. Heritability model outputs for the single taxon phenotypes (n=283) and community phenotypes (n=7). See Supplement Section 1.3 for how phenotypes were determined. Taxonomic levels are indicated by letters; p indicates phylum, c indicates class, o indicates order, f indicates family, g indicates genus, and ASV indicates ASV. Heritability estimates are from the 16,234 sample models. Phenotypes are ordered by descending heritability within phenotype description.

Table S7. Heritability model outputs for the presence/absence phenotypes (n=744). See Supplement Section 1.3 for how phenotypes were determined. Taxonomic levels are indicated by letters; p indicates phylum, c indicates class, o indicates order, f indicates family, g indicates genus, and ASV indicates ASV. Heritability estimates are from the 16,234 sample models. Phenotypes are ordered by descending heritability. Note that likelihood ratio tests, which were used for model selection in the single-taxon and community phenotype heritability models, cannot be used for the binomial animal model. We therefore classified a presence/absence phenotype as heritable if the additive genetic variance (VA) component made a positive (non-zero) contribution to the model and if the standard error bars did not overlap with 0 (“Heritability status” column).

Table S8. Heritability model outputs for the centered-log ratio (CLR)-transformed single-taxon phenotypes (n=283). See Supplement Section 1.3 for how phenotypes were determined. Taxonomic levels are indicated by letters; p indicates phylum, c indicates class, o indicates order, f indicates family, g indicates genus, and ASV indicates ASV. Heritability estimates are from the 16,234 sample models. Phenotypes are ordered by descending heritability.

Table S9. Heritability model outputs for the PhILR-transformed single-taxon phenotypes (n=139 ASVS transformed to 138 PhILR balances). See Supplement Section 1.3 for how phenotypes were determined. Taxonomic levels are indicated by letters; p indicates phylum, c indicates class, o indicates order, f indicates family, g indicates genus, and ASV indicates ASV. Heritability estimates are from the 16,234 sample models. Phenotypes are ordered by descending heritability.

Table S10. Heritability estimates for primate traits presented in Figure S12.

Table S11. Sample sizes for yearly models and for yearly grooming network models. Note that only years with >150 baboons and >1000 samples were used for yearly models, and only years with >150 baboons who were full-year residents and >1000 samples were used for grooming network models. Models were run on the top 15 most heritable collapsed phenotypes. Results are shown in Figures 3A, S7C.

Table S12. Heritability estimates and percent heritable taxa in 7 human studies and across varying sampling depths of the current baboon study. Note that the ‘percent of heritable phenotypes’ and ‘mean heritability’ for the studies marked ‘Current study subsample’ are based on running 100 random subsamples.

Table S13. Single-taxon collapsed phenotypes (n=93) and their corresponding phylogenetically nested, correlated single-taxon phenotypes found in 50% of the samples (n=283). See Supplement Section 1.3 for how

collapsed phenotypes were calculated. Taxonomic levels are indicated by single letters; p indicates phylum, c indicates class, o indicates order, f indicates family, and g indicates genus. Heritability estimates are from the 16,234 sample models.

Table S14. Sample sizes in heritability age window models. Models were run on the top 15 most heritable collapsed phenotypes from the full model (16,234 samples). Results are shown in Figures 3E, 3F.

Table S15. For the sampling depth set of models (N samples), the number of baboons in each random subset (N baboons) does not correlate with the heritability estimates for that subset (Pearson's R, Pearson's p-value) for any sampling depth except for the lowest one. 100 permutations were run at each sampling depth. Models were run on the 100 collapsed phenotypes. Results are shown in Figures 4C, 4D.

References and Notes

1. J. K. Goodrich, J. L. Waters, A. C. Poole, J. L. Sutter, O. Koren, R. Blekhman, M. Beaumont, W. Van Treuren, R. Knight, J. T. Bell, T. D. Spector, A. G. Clark, R. E. Ley, Human genetics shape the gut microbiome. *Cell* **159**, 789–799 (2014). [doi:10.1016/j.cell.2014.09.053](https://doi.org/10.1016/j.cell.2014.09.053) [Medline](#)
2. J. K. Goodrich, E. R. Davenport, M. Beaumont, M. A. Jackson, R. Knight, C. Ober, T. D. Spector, J. T. Bell, A. G. Clark, R. E. Ley, Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* **19**, 731–743 (2016). [doi:10.1016/j.chom.2016.04.017](https://doi.org/10.1016/j.chom.2016.04.017) [Medline](#)
3. R. Blekhman, J. K. Goodrich, K. Huang, Q. Sun, R. Bukowski, J. T. Bell, T. D. Spector, A. Keinan, R. E. Ley, D. Gevers, A. G. Clark, Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.* **16**, 191 (2015). [doi:10.1186/s13059-015-0759-1](https://doi.org/10.1186/s13059-015-0759-1) [Medline](#)
4. E. R. Davenport, D. A. Cusanovich, K. Michelini, L. B. Barreiro, C. Ober, Y. Gilad, Genome-wide association studies of the human gut microbiota. *PLOS ONE* **10**, e0140301 (2015). [doi:10.1371/journal.pone.0140301](https://doi.org/10.1371/journal.pone.0140301) [Medline](#)
5. M. Y. Lim, H. J. You, H. S. Yoon, B. Kwon, J. Y. Lee, S. Lee, Y.-M. Song, K. Lee, J. Sung, G. Ko, The effect of heritability and host genetics on the gut microbiota and metabolic syndrome. *Gut* **66**, 1031–1038 (2017). [doi:10.1136/gutjnl-2015-311326](https://doi.org/10.1136/gutjnl-2015-311326) [Medline](#)
6. W. Turpin, O. Espin-Garcia, W. Xu, M. S. Silverberg, D. Kevans, M. I. Smith, D. S. Guttman, A. Griffiths, R. Panaccione, A. Otley, L. Xu, K. Shestopaloff, G. Moreno-Hagelsieb, A. D. Paterson, K. Croitoru; GEM Project Research Consortium, Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat. Genet.* **48**, 1413–1417 (2016). [doi:10.1038/ng.3693](https://doi.org/10.1038/ng.3693) [Medline](#)
7. D. Rothschild, O. Weissbrod, E. Barkan, A. Kurilshikov, T. Korem, D. Zeevi, P. I. Costea, A. Godneva, I. N. Kalka, N. Bar, S. Shilo, D. Lador, A. V. Vila, N. Zmora, M. Pevsner-Fischer, D. Israeli, N. Kosower, G. Malka, B. C. Wolf, T. Avnit-Sagi, M. Lotan-Pompan, A. Weinberger, Z. Halpern, S. Carmi, J. Fu, C. Wijmenga, A. Zhernakova, E. Elinav, E. Segal, Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018). [doi:10.1038/nature25973](https://doi.org/10.1038/nature25973) [Medline](#)
8. A. Kurilshikov, C. Medina-Gomez, R. Bacigalupe, D. Radjabzadeh, J. Wang, A. Demirkan, C. I. Le Roy, J. A. Raygoza Garay, C. T. Finnicum, X. Liu, D. V. Zhernakova, M. J. Bonder, T. H. Hansen, F. Frost, M. C. Rühlemann, W. Turpin, J.-Y. Moon, H.-N. Kim, K. Lüll, E. Barkan, S. A. Shah, M. Fornage, J. Szopinska-Tokov, Z. D. Wallen, D. Borisevich, L. Agreus, A. Andreasson, C. Bang, L. Bedrani, J. T. Bell, H. Bisgaard, M. Boehnke, D. I. Boomsma, R. D. Burk, A. Claringbould, K. Croitoru, G. E. Davies, C. M. van Duijn, L. Duijts, G. Falony, J. Fu, A. van der Graaf, T. Hansen, G. Homuth, D. A. Hughes, R. G. Ijzerman, M. A. Jackson, V. W. V. Jaddoe, M. Joossens, T. Jørgensen, D. Keszthelyi, R. Knight, M. Laakso, M. Laudes, L. J. Launer, W. Lieb, A. J. Lysis, A. A. M. Masclee, H. A. Moll, Z. Mujagic, Q. Qibin, D. Rothschild, H. Shin, S. J. Sørensen, C. J. Steves, J. Thorsen, N. J. Timpson, R. Y. Tito, S. Vieira-Silva, U. Völker, H. Völzke, U. Vösa, K. H. Wade, S. Walter, K. Watanabe, S. Weiss, F. U. Weiss, O. Weissbrod, H.-J. Westra, G. Willemsen, H. Payami, D. M. A. E. Jonkers, A. Arias Vasquez, E. J. C. de

- Geus, K. A. Meyer, J. Stokholm, E. Segal, E. Org, C. Wijmenga, H.-L. Kim, R. C. Kaplan, T. D. Spector, A. G. Uitterlinden, F. Rivadeneira, A. Franke, M. M. Lerch, L. Franke, S. Sanna, M. D'Amato, O. Pedersen, A. D. Paterson, R. Kraaij, J. Raes, A. Zhernakova, Large-scale association analyses identify host factors influencing human gut microbiome composition. *bioRxiv* 173724 [Preprint]. 16 December 2020. <https://doi.org/10.1101/2020.06.26.173724>.
9. J. K. Goodrich, E. R. Davenport, A. G. Clark, R. E. Ley, The relationship between the human genome and microbiome comes into view. *Annu. Rev. Genet.* **51**, 413–433 (2017). [doi:10.1146/annurev-genet-110711-155532](https://doi.org/10.1146/annurev-genet-110711-155532) [Medline](#)
 10. S. Lax, D. P. Smith, J. Hampton-Marcell, S. M. Owens, K. M. Handley, N. M. Scott, S. M. Gibbons, P. Larsen, B. D. Shogan, S. Weiss, J. L. Metcalf, L. K. Ursell, Y. Vázquez-Baeza, W. Van Treuren, N. A. Hasan, M. K. Gibson, R. Colwell, G. Dantas, R. Knight, J. A. Gilbert, Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science* **345**, 1048–1052 (2014). [doi:10.1126/science.1254529](https://doi.org/10.1126/science.1254529) [Medline](#)
 11. B. H. Schlomann, R. Parthasarathy, Timescales of gut microbiome dynamics. *Curr. Opin. Microbiol.* **50**, 56–63 (2019). [doi:10.1016/j.mib.2019.09.011](https://doi.org/10.1016/j.mib.2019.09.011) [Medline](#)
 12. P. M. Visscher, W. G. Hill, N. R. Wray, Heritability in the genomics era—Concepts and misconceptions. *Nat. Rev. Genet.* **9**, 255–266 (2008). [doi:10.1038/nrg2322](https://doi.org/10.1038/nrg2322) [Medline](#)
 13. T. Ge, C.-Y. Chen, B. M. Neale, M. R. Sabuncu, J. W. Smoller, Phenome-wide heritability analysis of the UK Biobank. *PLOS Genet.* **13**, e1006711 (2017). [doi:10.1371/journal.pgen.1006711](https://doi.org/10.1371/journal.pgen.1006711) [Medline](#)
 14. S. C. Alberts, J. Altmann, “The Amboseli Baboon Research Project: 40 years of continuity and change,” in *Long-Term Field Studies of Primates*, P. M. Kappeler, D. P. Watts, Eds. (Springer, 2012), pp. 261–287.
 15. A. M. Bronikowski, M. Cords, S. C. Alberts, J. Altmann, D. K. Brockman, L. M. Fedigan, A. Pusey, T. Stoinski, K. B. Strier, W. F. Morris, Female and male life tables for seven wild primate species. *Sci. Data* **3**, 160006 (2016). [doi:10.1038/sdata.2016.6](https://doi.org/10.1038/sdata.2016.6) [Medline](#)
 16. A. C. Markham, V. Guttal, S. C. Alberts, J. Altmann, When good neighbors don't need fences: Temporal landscape partitioning among baboon social groups. *Behav. Ecol. Sociobiol.* **67**, 875–884 (2013). [doi:10.1007/s00265-013-1510-0](https://doi.org/10.1007/s00265-013-1510-0) [Medline](#)
 17. Materials and methods are available as supplementary materials.
 18. J. Tung, L. B. Barreiro, M. B. Burns, J.-C. Grenier, J. Lynch, L. E. Grieneisen, J. Altmann, S. C. Alberts, R. Blekman, E. A. Archie, Social networks predict gut microbiome composition in wild baboons. *eLife* **4**, e05224 (2015). [doi:10.7554/eLife.05224](https://doi.org/10.7554/eLife.05224) [Medline](#)
 19. L. E. Grieneisen, J. Livermore, S. Alberts, J. Tung, E. A. Archie, Group Living and Male Dispersal Predict the Core Gut Microbiome in Wild Baboons, Group Living and Male Dispersal Predict the Core Gut Microbiome in Wild Baboons. *Integr. Comp. Biol.* **57**, 770–785 (2017). [doi:10.1093/icb/ix046](https://doi.org/10.1093/icb/ix046) [Medline](#)
 20. K. Berer, L. A. Gerdes, E. Cekanaviciute, X. Jia, L. Xiao, Z. Xia, C. Liu, L. Klotz, U. Stauffer, S. E. Baranzini, T. Kümpfel, R. Hohlfeld, G. Krishnamoorthy, H. Wekerle, Gut microbiota from multiple sclerosis patients enables spontaneous autoimmune

- encephalomyelitis in mice. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10719–10724 (2017).
[doi:10.1073/pnas.1711233114](https://doi.org/10.1073/pnas.1711233114) [Medline](#)
21. A. E. Mann, F. Mazel, M. A. Lemay, E. Morien, V. Billy, M. Kowalewski, A. Di Fiore, A. Link, T. L. Goldberg, S. Tecot, A. L. Baden, A. Gomez, M. L. Sauter, F. P. Cuozzo, G. A. O. Rice, N. J. Dominy, R. Stumpf, R. J. Lewis, L. Swedell, K. Amato, L. Wegener Parfrey, Biodiversity of protists and nematodes in the wild nonhuman primate gut. *ISME J.* **14**, 609–622 (2020). [doi:10.1038/s41396-019-0551-4](https://doi.org/10.1038/s41396-019-0551-4) [Medline](#)
 22. W. A. Walters, Z. Jin, N. Youngblut, J. G. Wallace, J. Sutter, W. Zhang, A. González-Peña, J. Peiffer, O. Koren, Q. Shi, R. Knight, T. Glavina Del Rio, S. G. Tringe, E. S. Buckler, J. L. Dangl, R. E. Ley, Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7368–7373 (2018).
[doi:10.1073/pnas.1800918115](https://doi.org/10.1073/pnas.1800918115) [Medline](#)
 23. A. R. Gilmour, “ASREML for testing fixed effects and estimating multiple trait variance components,” in Proceedings of the Association for the Advancement of Animal Breeding and Genetics (AABG, 1997), vol. 12, pp. 386–390.
 24. L. E. B. Kruuk, Estimating genetic parameters in natural populations using the “animal model”. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**, 873–890 (2004).
[doi:10.1098/rstb.2003.1437](https://doi.org/10.1098/rstb.2003.1437) [Medline](#)
 25. A. J. Wilson, D. Réale, M. N. Clements, M. M. Morrissey, E. Postma, C. A. Walling, L. E. B. Kruuk, D. H. Nussey, An ecologist’s guide to the animal model. *J. Anim. Ecol.* **79**, 13–26 (2010). [doi:10.1111/j.1365-2656.2009.01639.x](https://doi.org/10.1111/j.1365-2656.2009.01639.x) [Medline](#)
 26. A. J. Wilson, Why h^2 does not always equal V_A/V_P ? *J. Evol. Biol.* **21**, 647–650 (2008).
[doi:10.1111/j.1420-9101.2008.01500.x](https://doi.org/10.1111/j.1420-9101.2008.01500.x) [Medline](#)
 27. L. E. B. Kruuk, J. D. Hadfield, How to separate genetic and environmental causes of similarity between relatives. *J. Evol. Biol.* **20**, 1890–1903 (2007). [doi:10.1111/j.1420-9101.2007.01377.x](https://doi.org/10.1111/j.1420-9101.2007.01377.x) [Medline](#)
 28. N. Barban, R. Jansen, R. de Vlaming, A. Vaez, J. J. Mandemakers, F. C. Tropf, X. Shen, J. F. Wilson, D. I. Chasman, I. M. Nolte, V. Tragante, S. W. van der Laan, J. R. B. Perry, A. Kong, T. S. Ahluwalia, E. Albrecht, L. Yerges-Armstrong, G. Atzmon, K. Auro, K. Ayers, A. Bakshi, D. Ben-Avraham, K. Berger, A. Bergman, L. Bertram, L. F. Bielak, G. Bjornsdottir, M. J. Bonder, L. Broer, M. Bui, C. Barbieri, A. Cavadino, J. E. Chavarro, C. Turman, M. P. Concas, H. J. Cordell, G. Davies, P. Eibich, N. Eriksson, T. Esko, J. Eriksson, F. Falahi, J. F. Felix, M. A. Fontana, L. Franke, I. Gandin, A. J. Gaskins, C. Gieger, E. P. Gunderson, X. Guo, C. Hayward, C. He, E. Hofer, H. Huang, P. K. Joshi, S. Kanoni, R. Karlsson, S. Kiechl, A. Kifley, A. Kluttig, P. Kraft, V. Lagou, C. Lecoeur, J. Lahti, R. Li-Gao, P. A. Lind, T. Liu, E. Makalic, C. Mamasoula, L. Matteson, H. Mbarek, P. F. McArdle, G. McMahon, S. F. W. Meddens, E. Mihailov, M. Miller, S. A. Missmer, C. Monnereau, P. J. van der Most, R. Myhre, M. A. Nalls, T. Ntile, I. P. Kalafati, E. Porcu, I. Prokopenko, K. B. Rajan, J. Rich-Edwards, C. A. Rietveld, A. Robino, L. M. Rose, R. Rueedi, K. A. Ryan, Y. Saba, D. Schmidt, J. A. Smith, L. Stolk, E. Streeten, A. Tönjes, G. Thorleifsson, S. Ulivi, J. Wedenoja, J. Wellmann, P. Willeit, J. Yao, L. Yengo, J. H. Zhao, W. Zhao, D. V. Zhernakova, N. Amin, H. Andrews, B. Balkau, N. Barzilai, S. Bergmann, G. Biino, H. Bisgaard, K. Bønnelykke, D. I. Boomsma, J. E.

- Buring, H. Campbell, S. Cappellani, M. Ciullo, S. R. Cox, F. Cucca, D. Toniolo, G. Davey-Smith, I. J. Deary, G. Dedoussis, P. Deloukas, C. M. van Duijn, E. J. C. de Geus, J. G. Eriksson, D. A. Evans, J. D. Faul, C. F. Sala, P. Froguel, P. Gasparini, G. Girotto, H.-J. Grabe, K. H. Greiser, P. J. F. Groenen, H. G. de Haan, J. Haerting, T. B. Harris, A. C. Heath, K. Heikkilä, A. Hofman, G. Homuth, E. G. Holliday, J. Hopper, E. Hyppönen, B. Jacobsson, V. W. V. Jaddoe, M. Johannesson, A. Jugessur, M. Kähönen, E. Kajantie, S. L. R. Kardia, B. Keavney, I. Kolcic, P. Koponen, P. Kovacs, F. Kronenberg, Z. Kutalik, M. La Bianca, G. Lachance, W. G. Iacono, S. Lai, T. Lehtimäki, D. C. Liewald, C. M. Lindgren, Y. Liu, R. Luben, M. Lucht, R. Luoto, P. Magnus, P. K. E. Magnusson, N. G. Martin, M. McGue, R. McQuillan, S. E. Medland, C. Meisinger, D. Mellström, A. Metspalu, M. Traglia, L. Milani, P. Mitchell, G. W. Montgomery, D. Mook-Kanamori, R. de Mutsert, E. A. Nohr, C. Ohlsson, J. Olsen, K. K. Ong, L. Paternoster, A. Pattie, B. W. J. H. Penninx, M. Perola, P. A. Peyser, M. Pirastu, O. Polasek, C. Power, J. Kaprio, L. J. Raffel, K. Rääkkönen, O. Raitakari, P. M. Ridker, S. M. Ring, K. Roll, I. Rudan, D. Ruggiero, D. Rujescu, V. Salomaa, D. Schlessinger, H. Schmidt, R. Schmidt, N. Schupf, J. Smit, R. Sorice, T. D. Spector, J. M. Starr, D. Stöckl, K. Strauch, M. Stumvoll, M. A. Swertz, U. Thorsteinsdottir, A. R. Thurik, N. J. Timpson, J. Y. Tung, A. G. Uitterlinden, S. Vaccargiu, J. Viikari, V. Vitart, H. Völzke, P. Vollenweider, D. Vuckovic, J. Waage, G. G. Wagner, J. J. Wang, N. J. Wareham, D. R. Weir, G. Willemsen, J. Willeit, A. F. Wright, K. T. Zondervan, K. Stefansson, R. F. Krueger, J. J. Lee, D. J. Benjamin, D. Cesarini, P. D. Koellinger, M. den Hoed, H. Snieder, M. C. Mills; BIOS Consortium; LifeLines Cohort Study, Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48**, 1462–1472 (2016). [doi:10.1038/ng.3698](https://doi.org/10.1038/ng.3698) [Medline](#)
29. D. Cesarini, P. M. Visscher, Genetics and educational attainment. *NPJ Sci. Learn.* **2**, 4 (2017). [doi:10.1038/s41539-017-0005-6](https://doi.org/10.1038/s41539-017-0005-6) [Medline](#)
30. A. P. Hendry, D. J. Schoen, M. E. Wolak, J. M. Reid, The Contemporary Evolution of Fitness. *Annu. Rev. Ecol. Evol. Syst.* **49**, 457–476 (2018). [doi:10.1146/annurev-ecolsys-110617-062358](https://doi.org/10.1146/annurev-ecolsys-110617-062358)
31. T. Ren, L. E. Grieneisen, S. C. Alberts, E. A. Archie, M. Wu, Development, diet and dynamism: Longitudinal and cross-sectional predictors of gut microbial communities in wild baboons. *Environ. Microbiol.* **18**, 1312–1325 (2016). [doi:10.1111/1462-2920.12852](https://doi.org/10.1111/1462-2920.12852) [Medline](#)
32. T. Yatsunenko, F. E. Rey, M. J. Manary, I. Trehan, M. G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, R. N. Baldassano, A. P. Anokhin, A. C. Heath, B. Warner, J. Reeder, J. Kuczynski, J. G. Caporaso, C. A. Lozupone, C. Lauber, J. C. Clemente, D. Knights, R. Knight, J. I. Gordon, Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012). [doi:10.1038/nature11053](https://doi.org/10.1038/nature11053) [Medline](#)
33. P. J. Turnbaugh, M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, M. L. Sogin, W. J. Jones, B. A. Roe, J. P. Affourtit, M. Egholm, B. Henrissat, A. C. Heath, R. Knight, J. I. Gordon, A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009). [doi:10.1038/nature07540](https://doi.org/10.1038/nature07540) [Medline](#)
34. J. L. Waters, R. E. Ley, The human gut bacteria Christensenellaceae are widespread, heritable, and associated with health. *BMC Biol.* **17**, 83 (2019). [doi:10.1186/s12915-019-0699-4](https://doi.org/10.1186/s12915-019-0699-4) [Medline](#)

35. R. Plomin, I. J. Deary, Genetics and intelligence differences: Five special findings. *Mol. Psychiatry* **20**, 98–108 (2015). [doi:10.1038/mp.2014.105](https://doi.org/10.1038/mp.2014.105) [Medline](#)
36. A. Gonzalez, J. A. Navas-Molina, T. Kosciolk, D. McDonald, Y. Vázquez-Baeza, G. Ackermann, J. DeReus, S. Janssen, A. D. Swafford, S. B. Orchanian, J. G. Sanders, J. Shorenstein, H. Holste, S. Petrus, A. Robbins-Pianka, C. J. Brislawn, M. Wang, J. R. Rideout, E. Bolyen, M. Dillon, J. G. Caporaso, P. C. Dorrestein, R. Knight, Qiita: Rapid, web-enabled microbiome meta-analysis. *Nat. Methods* **15**, 796–798 (2018). [doi:10.1038/s41592-018-0141-9](https://doi.org/10.1038/s41592-018-0141-9) [Medline](#)
37. L. Grieneisen, M. Dasari, T. J. Gould, J. R. Björk, J.-C. Grenier, V. Yotova, D. Jansen, N. Gottel, J. B. Gordon, N. H. Learn, L. R. Gesquiere, T. L. Wango, R. S. Mututua, J. K. Warutere, L. Siodi, J. A. Gilbert, L. B. Barreiro, S. C. Alberts, J. Tung, E. A. Archie, R. Blekhman, Data for: Gut microbiome heritability is nearly universal but environmentally contingent, Zenodo (2021); <https://doi.org/10.5281/zenodo.4662081>.
38. S. C. Alberts, J. Altmann, E. A. Archie, J. Tung, “Monitoring guide for the Amboseli Baboon Research Project: Protocols for long-term monitoring and data collection” (Amboseli Baboon Research Project, 2011); http://princeton.edu/baboon/monitoring_guide.html.
39. L. E. Grieneisen, M. J. E. Charpentier, S. C. Alberts, R. Blekhman, G. Bradburd, J. Tung, E. A. Archie, Genes, geology and germs: Gut microbiota across a primate hybrid zone are explained by site soil properties, not host species. *Proc. Biol. Sci.* **286**, 20190431 (2019). [doi:10.1098/rspb.2019.0431](https://doi.org/10.1098/rspb.2019.0431) [Medline](#)
40. S. C. Alberts, J. Hollister-Smith, R. S. Mututua, S. N. Sayialel, P. M. Muruthi, J. K. Altmann, “Seasonality and long term change in a savannah environment,” in *Primate Seasonality: Implications for Human Evolution*, D. K. Brockman, C. P. van Schaik, Eds. (2005), pp. 157–196.
41. S. C. Alberts, J. Altmann, Balancing costs and opportunities: Dispersal in male baboons. *Am. Nat.* **145**, 279–306 (1995). [doi:10.1086/285740](https://doi.org/10.1086/285740)
42. J. C. Buchan, S. C. Alberts, J. B. Silk, J. Altmann, True paternal care in a multi-male primate society. *Nature* **425**, 179–181 (2003). [doi:10.1038/nature01866](https://doi.org/10.1038/nature01866) [Medline](#)
43. S. C. Alberts, J. C. Buchan, J. Altmann, Sexual selection in wild baboons: From mating opportunities to paternity success. *Anim. Behav.* **72**, 1177–1196 (2006). [doi:10.1016/j.anbehav.2006.05.001](https://doi.org/10.1016/j.anbehav.2006.05.001)
44. M. B. Morrissey, A. J. Wilson, pedantics: An r package for pedigree-based genetic simulation and pedigree manipulation, characterization and viewing. *Mol. Ecol. Resour.* **10**, 711–719 (2010). [doi:10.1111/j.1755-0998.2009.02817.x](https://doi.org/10.1111/j.1755-0998.2009.02817.x) [Medline](#)
45. J. Altmann, Observational study of behavior: Sampling methods. *Behaviour* **49**, 227–266 (1974). [doi:10.1163/156853974X00534](https://doi.org/10.1163/156853974X00534) [Medline](#)
46. J. B. Silk, Activities and feeding behavior of free-ranging pregnant baboons. *Int. J. Primatol.* **8**, 593–613 (1987). [doi:10.1007/BF02735779](https://doi.org/10.1007/BF02735779)
47. S. A. Altmann, Fallback foods, eclectic omnivores, and the packaging problem. *Am. J. Phys. Anthropol.* **140**, 615–629 (2009). [doi:10.1002/ajpa.21097](https://doi.org/10.1002/ajpa.21097) [Medline](#)
48. S. A. Altmann, *Foraging for Survival: Yearling Baboons in Africa* (Univ. of Chicago Press,

- 1998).
49. A. M. Bronikowski, J. Altmann, Foraging in a variable environment: Weather patterns and the behavioral ecology of baboons. *Behav. Ecol. Sociobiol.* **39**, 11–25 (1996). [doi:10.1007/s002650050262](https://doi.org/10.1007/s002650050262)
 50. P. Muruthi, J. Altmann, S. Altmann, Resource base, parity, and reproductive condition affect females' feeding time and nutrient intake within and between groups of a baboon population. *Oecologia* **87**, 467–472 (1991). [doi:10.1007/BF00320408](https://doi.org/10.1007/BF00320408) [Medline](#)
 51. J. M. Shopland, Food quality, spatial deployment, and the intensity of feeding interference in yellow baboons (*Papio cynocephalus*). *Behav. Ecol. Sociobiol.* **21**, 149–156 (1987). [doi:10.1007/BF00303204](https://doi.org/10.1007/BF00303204)
 52. L. Gesquiere, J. Beehner, M. Khan, J. Lynch, E. Fox, J. Altmann, “Amboseli Baboon Research Project: Laboratory procedures for fecal hormones” (Princeton Univ. Press, 2017); https://amboselibaboons.nd.edu/assets/243149/abrp_lab_procedures_for_fecal_hormones_updated_21_jul_2017.pdf.
 53. P. O. Onyango, L. R. Gesquiere, J. Altmann, S. C. Alberts, Puberty and dispersal in a wild primate population. *Horm. Behav.* **64**, 240–249 (2013). [doi:10.1016/j.yhbeh.2013.02.014](https://doi.org/10.1016/j.yhbeh.2013.02.014) [Medline](#)
 54. L. R. Gesquiere, N. H. Learn, M. C. M. Simao, P. O. Onyango, S. C. Alberts, J. Altmann, Life at the top: Rank and stress in wild male baboons. *Science* **333**, 357–360 (2011). [doi:10.1126/science.1207120](https://doi.org/10.1126/science.1207120) [Medline](#)
 55. J. C. Beehner, L. Gesquiere, R. M. Seyfarth, D. L. Cheney, S. C. Alberts, J. Altmann, Testosterone related to age and life-history stages in male baboons and geladas. *Horm. Behav.* **56**, 472–480 (2009). [doi:10.1016/j.yhbeh.2009.08.005](https://doi.org/10.1016/j.yhbeh.2009.08.005) [Medline](#)
 56. L. R. Gesquiere, P. O. Onyango, S. C. Alberts, J. Altmann, Endocrinology of year-round reproduction in a highly seasonal habitat: Environmental variability in testosterone and glucocorticoids in baboon males. *Am. J. Phys. Anthropol.* **144**, 169–176 (2011). [doi:10.1002/ajpa.21374](https://doi.org/10.1002/ajpa.21374) [Medline](#)
 57. L. R. Gesquiere, J. Altmann, M. Z. Khan, J. Couret, J. C. Yu, C. S. Endres, J. W. Lynch, P. Ogola, E. A. Fox, S. C. Alberts, E. O. Wango, Coming of age: Steroid hormones of wild immature baboons (*Papio cynocephalus*). *Am. J. Primatol.* **67**, 83–100 (2005). [doi:10.1002/ajp.20171](https://doi.org/10.1002/ajp.20171) [Medline](#)
 58. P. O. Onyango, L. R. Gesquiere, J. Altmann, S. C. Alberts, Testosterone positively associated with both male mating effort and paternal behavior in Savanna baboons (*Papio cynocephalus*). *Horm. Behav.* **63**, 430–436 (2013). [doi:10.1016/j.yhbeh.2012.11.014](https://doi.org/10.1016/j.yhbeh.2012.11.014) [Medline](#)
 59. R. Blekhman, K. Tang, E. A. Archie, L. B. Barreiro, Z. P. Johnson, M. E. Wilson, J. Kohn, M. L. Yuan, L. Gesquiere, L. E. Grieneisen, J. Tung, Common methods for fecal sample storage in field studies yield consistent signatures of individual identity in microbiome sequencing data. *Sci. Rep.* **6**, 31519 (2016). [doi:10.1038/srep31519](https://doi.org/10.1038/srep31519) [Medline](#)
 60. MoBiol, “PowerSoil DNA isolation kit” (MoBiol, YEAR);

- <https://mobio.com/media/wysiwyg/pdfs/protocols/12888.pdf>.
61. QIAGEN, “MO BIO’s PowerSoil-htp 96 well soil DNA kit handbook” (QIAGEN, YEAR); <https://www.qiagen.com/us/resources/resourcedetail?id=fd3fa52e-3a66-4d55-a9cd-7ed20ea046d9&lang=en>.
 62. J. A. Gilbert, J. K. Jansson, R. Knight, The Earth Microbiome project: Successes and aspirations. *BMC Biol.* **12**, 69 (2014). [doi:10.1186/s12915-014-0069-1](https://doi.org/10.1186/s12915-014-0069-1) [Medline](#)
 63. J. G. Caporaso, C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, R. Knight, Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* **108** (Suppl 1), 4516–4522 (2011). [doi:10.1073/pnas.1000080107](https://doi.org/10.1073/pnas.1000080107) [Medline](#)
 64. J. G. Caporaso, C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J. A. Gilbert, G. Smith, R. Knight, Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012). [doi:10.1038/ismej.2012.8](https://doi.org/10.1038/ismej.2012.8) [Medline](#)
 65. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, S. P. Holmes, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016). [doi:10.1038/nmeth.3869](https://doi.org/10.1038/nmeth.3869) [Medline](#)
 66. Illumina, “bcl2fastq2 conversion software v2.20” (Illumina, 2019); https://emea.support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/bcl2fastq/bcl2fastq2-v2-20-software-guide-15051736-03.pdf.
 67. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011). [doi:10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200)
 68. B. Callahan, “A DADA2 workflow for big data (1.4 or later)” (GitHub, YEAR); <https://benjjneb.github.io/dada2/bigdata.html>.
 69. E. S. Wright, L. S. Yilmaz, D. R. Noguera, DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Appl. Environ. Microbiol.* **78**, 717–725 (2012). [doi:10.1128/AEM.06516-11](https://doi.org/10.1128/AEM.06516-11) [Medline](#)
 70. C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F. O. Glöckner, The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41** (D1), D590–D596 (2013). [doi:10.1093/nar/gks1219](https://doi.org/10.1093/nar/gks1219) [Medline](#)
 71. W. W. B. Goh, W. Wang, L. Wong, Why batch effects matter in omics data, and how to avoid them. *Trends Biotechnol.* **35**, 498–507 (2017). [doi:10.1016/j.tibtech.2017.02.012](https://doi.org/10.1016/j.tibtech.2017.02.012) [Medline](#)
 72. J. T. Leek, R. B. Scharpf, H. C. Bravo, D. Simcha, B. Langmead, W. E. Johnson, D. Geman, K. Baggerly, R. A. Irizarry, Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **11**, 733–739 (2010). [doi:10.1038/nrg2825](https://doi.org/10.1038/nrg2825) [Medline](#)
 73. P. D. Schloss, D. Gevers, S. L. Westcott, Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLOS ONE* **6**, e27310 (2011).

- [doi:10.1371/journal.pone.0027310](https://doi.org/10.1371/journal.pone.0027310) [Medline](#)
74. J. Oksanen, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O’Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, H. H. Wagner, “Vegan: Community ecology package” (R package version 2.0-10, 2015); URL.
75. J. N. Paulson, M. Pop, H. C. Bravo, “metagenomeSeq: Statistical analysis for sparse high-throughput sequencing” (Bioconductor package version 1.6.0, 2013); <http://cbcb.umd.edu/software/metagenomeSeq>.
76. D. W. Roberts, “labdsv: Ordination and multivariate analysis for ecology” (R package version 1.0-5, 2007); <http://CRAN.R-project.org/package=labdsv>.
77. G. B. Gloor, J. M. Macklaim, V. Pawlowsky-Glahn, J. J. Egozcue, Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* **8**, 2224 (2017). [doi:10.3389/fmicb.2017.02224](https://doi.org/10.3389/fmicb.2017.02224) [Medline](#)
78. J. D. Silverman, A. D. Washburne, S. Mukherjee, L. A. David, A phylogenetic transform enhances analysis of compositional microbiota data. *eLife* **6**, e21887 (2017). [doi:10.7554/eLife.21887](https://doi.org/10.7554/eLife.21887) [Medline](#)
79. K. G. van den Boogaart, R. Tolosana-Delgado, “compositions”: A unified R package to analyze compositional data. *Comput. Geosci.* **34**, 320–338 (2008). [doi:10.1016/j.cageo.2006.11.017](https://doi.org/10.1016/j.cageo.2006.11.017)
80. F. E. Harrell Jr, M. C. Dupont, “The Hmisc package” (R package version 3.3, 2006); <https://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf>.
81. B. A. Pico, F. W. C. Naser, J. B. Van Wyk, Genetic parameters for growth traits in South African Brahman cattle. *S. Afr. J. Anim. Sci.* **34**, 44–46 (2004). [doi:10.4314/sajas.v34i6.3827](https://doi.org/10.4314/sajas.v34i6.3827)
82. R. Neira, N. F. Díaz, G. A. E. Gall, J. A. Gallardo, J. P. Lhorente, A. Alert, Genetic improvement in coho salmon (*Oncorhynchus kisutch*). II: Selection response for early spawning date. *Aquaculture* **257**, 1–8 (2006). [doi:10.1016/j.aquaculture.2006.03.001](https://doi.org/10.1016/j.aquaculture.2006.03.001)
83. E. D. Ellen, V. Ducrocq, B. J. Ducro, R. F. Veerkamp, P. Bijma, Genetic parameters for social effects on survival in cannibalistic layers: Combining survival analysis and a linear animal model. *Genet. Sel. Evol.* **42**, 27 (2010). [doi:10.1186/1297-9686-42-27](https://doi.org/10.1186/1297-9686-42-27) [Medline](#)
84. R. E. Crump, C. S. Haley, R. Thompson, J. Mercer, Individual animal model estimates of genetic parameters for reproduction traits of Landrace pigs performance tested in a commercial nucleus herd. *Anim. Sci.* **65**, 285–290 (1997). [doi:10.1017/S1357729800016593](https://doi.org/10.1017/S1357729800016593)
85. E. Safari, N. M. Fogarty, A. R. Gilmour, A review of genetic parameter estimates for wool, growth, meat and reproduction traits in sheep. *Livest. Prod. Sci.* **92**, 271–289 (2005). [doi:10.1016/j.livprodsci.2004.09.003](https://doi.org/10.1016/j.livprodsci.2004.09.003)
86. J. E. Brommer, K. Rattiste, A. J. Wilson, Exploring plasticity in the wild: Laying date-temperature reaction norms in the common gull *Larus canus*. *Proc. Biol. Sci.* **275**, 687–693 (2008). [doi:10.1098/rspb.2007.0951](https://doi.org/10.1098/rspb.2007.0951) [Medline](#)
87. C. Teplitsky, J. A. Mills, J. S. Alho, J. W. Yarrall, J. Merilä, Bergmann’s rule and climate

- change revisited: Disentangling environmental and genetic responses in a wild bird population. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13492–13496 (2008). [doi:10.1073/pnas.0800999105](https://doi.org/10.1073/pnas.0800999105) [Medline](#)
88. K. Foerster, T. Coulson, B. C. Sheldon, J. M. Pemberton, T. H. Clutton-Brock, L. E. B. Kruuk, Sexually antagonistic genetic variation for fitness in red deer. *Nature* **447**, 1107–1110 (2007). [doi:10.1038/nature05912](https://doi.org/10.1038/nature05912) [Medline](#)
89. A. Charmantier, C. Perrins, R. H. McCleery, B. C. Sheldon, Quantitative genetics of age at reproduction in wild swans: Support for antagonistic pleiotropy models of senescence. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6587–6592 (2006). [doi:10.1073/pnas.0511123103](https://doi.org/10.1073/pnas.0511123103) [Medline](#)
90. A. L. Hicks, K. J. Lee, M. Couto-Rodriguez, J. Patel, R. Sinha, C. Guo, S. H. Olson, A. Seimon, T. A. Seimon, A. U. Ondzie, W. B. Karesh, P. Reed, K. N. Cameron, W. I. Lipkin, B. L. Williams, Gut microbiomes of wild great apes fluctuate seasonally in response to diet. *Nat. Commun.* **9**, 1786 (2018). [doi:10.1038/s41467-018-04204-w](https://doi.org/10.1038/s41467-018-04204-w) [Medline](#)
91. L. A. David, C. F. Maurice, R. N. Carmody, D. B. Gootenberg, J. E. Button, B. E. Wolfe, A. V. Ling, A. S. Devlin, Y. Varma, M. A. Fischbach, S. B. Biddinger, R. J. Dutton, P. J. Turnbaugh, Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014). [doi:10.1038/nature12820](https://doi.org/10.1038/nature12820) [Medline](#)
92. G. Bennett, M. Malone, M. L. Sauter, F. P. Cuzzo, B. White, K. E. Nelson, R. M. Stumpf, R. Knight, S. R. Leigh, K. R. Amato, Host age, social group, and habitat type influence the gut microbiota of wild ring-tailed lemurs (*Lemur catta*). *Am. J. Primatol.* **78**, 883–892 (2016). [doi:10.1002/ajp.22555](https://doi.org/10.1002/ajp.22555) [Medline](#)
93. J. G. M. Markle, D. N. Frank, S. Mortin-Toth, C. E. Robertson, L. M. Feazel, U. Rolle-Kampczyk, M. von Bergen, K. D. McCoy, A. J. Macpherson, J. S. Danska, Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science* **339**, 1084–1088 (2013). [doi:10.1126/science.1233521](https://doi.org/10.1126/science.1233521) [Medline](#)
94. C. F. Maurice, S. C. Knowles, J. Ladau, K. S. Pollard, A. Fenton, A. B. Pedersen, P. J. Turnbaugh, Marked seasonal variation in the wild mouse gut microbiota. *ISME J.* **9**, 2423–2434 (2015). [doi:10.1038/ismej.2015.53](https://doi.org/10.1038/ismej.2015.53) [Medline](#)
95. J. Yang, B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, P. M. Visscher, Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010). [doi:10.1038/ng.608](https://doi.org/10.1038/ng.608) [Medline](#)
96. A. Jelenkovic, Y.-M. Hur, R. Sund, Y. Yokoyama, S. H. Siribaddana, M. Hotopf, A. Sumathipala, F. Rijdsdijk, Q. Tan, D. Zhang, Z. Pang, S. Aaltonen, K. Heikkilä, S. Y. Öncel, F. Aliev, E. Rebato, A. D. Tarnoki, D. L. Tarnoki, K. Christensen, A. Skytthe, K. O. Kyvik, J. L. Silberg, L. J. Eaves, H. H. Maes, T. L. Cutler, J. L. Hopper, J. R. Ordoñana, J. F. Sánchez-Romera, L. Colodro-Conde, W. Cozen, A. E. Hwang, T. M. Mack, J. Sung, Y.-M. Song, S. Yang, K. Lee, C. E. Franz, W. S. Kremen, M. J. Lyons, A. Busjahn, T. L. Nelson, K. E. Whitfield, C. Kandler, K. L. Jang, M. Gatz, D. A. Butler, M. A. Stazi, C. Fagnani, C. D’Ippolito, G. E. Duncan, D. Buchwald, C. A. Derom, R. F. Vlietinck, R. J. F. Loos, N. G. Martin, S. E. Medland, G. W. Montgomery, H.-U. Jeong,

- G. E. Swan, R. Krasnow, P. K. E. Magnusson, N. L. Pedersen, A. K. Dahl-Aslan, T. A. McAdams, T. C. Eley, A. M. Gregory, P. Tynelius, L. A. Baker, C. Tuvblad, G. Bayasgalan, D. Narandalai, P. Lichtenstein, T. D. Spector, M. Mangino, G. Lachance, M. Bartels, T. C. E. M. van Beijsterveldt, G. Willemsen, S. A. Burt, K. L. Klump, J. R. Harris, I. Brandt, T. S. Nilsen, R. F. Krueger, M. McGue, S. Pahlen, R. P. Corley, J. V. Hjelmborg, J. H. Goldberg, Y. Iwatani, M. Watanabe, C. Honda, F. Inui, F. Rasmussen, B. M. Huibregtse, D. I. Boomsma, T. I. A. Sørensen, J. Kaprio, K. Silventoinen, Genetic and environmental influences on adult human height across birth cohorts from 1886 to 1994. *eLife* **5**, e20320 (2016). [doi:10.7554/eLife.20320](https://doi.org/10.7554/eLife.20320) [Medline](#)
97. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
98. D. G. Butler, B. R. Cullis, A. R. Gilmour, B. J. Gogel, R. Thompson, “ASReml-R reference manual version 4” (The State of Queensland, Department of Primary Industries and Fisheries, 2009); <https://asreml.kb.vsnr.co.uk/wp-content/uploads/sites/3/2018/02/ASReml-R-Reference-Manual-4.pdf>.
99. P. de Villemereuil, Quantitative genetic methods depending on the nature of the phenotypic trait. *Ann. N. Y. Acad. Sci.* **1422**, 29–47 (2018). [doi:10.1111/nyas.13571](https://doi.org/10.1111/nyas.13571) [Medline](#)
100. K. P. Schliep, phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011). [doi:10.1093/bioinformatics/btq706](https://doi.org/10.1093/bioinformatics/btq706) [Medline](#)
101. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004). [doi:10.1093/bioinformatics/btg412](https://doi.org/10.1093/bioinformatics/btg412) [Medline](#)
102. F. Keck, F. Rimet, A. Bouchez, A. Franc, phylsignal: An R package to measure, test, and explore the phylogenetic signal. *Ecol. Evol.* **6**, 2774–2780 (2016). [doi:10.1002/ece3.2051](https://doi.org/10.1002/ece3.2051) [Medline](#)
103. J. M. Kamilar, N. Cooper, Phylogenetic signal in primate behaviour, ecology and life history. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120341 (2013). [doi:10.1098/rstb.2012.0341](https://doi.org/10.1098/rstb.2012.0341) [Medline](#)
104. J. Friedman, E. J. Alm, Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, e1002687 (2012). [doi:10.1371/journal.pcbi.1002687](https://doi.org/10.1371/journal.pcbi.1002687) [Medline](#)
105. L. J. N. Brent, S. Semple, A. Maclarnon, A. Ruiz-Lambides, J. Gonzalez-Martinez, M. L. Platt, Personality traits in rhesus macaques (*Macaca mulatta*) are heritable but do not predict reproductive output. *Int. J. Primatol.* **35**, 188–209 (2014). [doi:10.1007/s10764-013-9724-6](https://doi.org/10.1007/s10764-013-9724-6) [Medline](#)
106. S. Madlon-Kay, M. J. Montague, L. J. N. Brent, S. Ellis, B. Zhong, N. Snyder-Mackler, J. E. Horvath, J. H. P. Skene, M. L. Platt, Weak effects of common genetic variation in oxytocin and vasopressin receptor genes on rhesus macaque social behavior. *Am. J. Primatol.* **80**, e22873 (2018). [doi:10.1002/ajp.22873](https://doi.org/10.1002/ajp.22873) [Medline](#)
107. G. E. Blomquist, L. J. N. Brent, Applying quantitative genetic methods to primate social behavior. *Int. J. Primatol.* **35**, 108–128 (2014). [doi:10.1007/s10764-013-9709-5](https://doi.org/10.1007/s10764-013-9709-5) [Medline](#)
108. L. J. N. Brent, S. R. Heilbronner, J. E. Horvath, J. Gonzalez-Martinez, A. Ruiz-Lambides,

- A. G. Robinson, J. H. P. Skene, M. L. Platt, Genetic origins of social networks in rhesus macaques. *Sci. Rep.* **3**, 1042 (2013). [doi:10.1038/srep01042](https://doi.org/10.1038/srep01042) [Medline](#)
109. N. Staes, A. Weiss, P. Helsen, M. Korody, M. Eens, J. M. G. Stevens, Bonobo personality traits are heritable and associated with vasopressin receptor gene 1a variation. *Sci. Rep.* **6**, 38193 (2016). [doi:10.1038/srep38193](https://doi.org/10.1038/srep38193) [Medline](#)
110. L. M. Havill, M. R. Allen, T. L. Bredbenner, D. B. Burr, D. P. Nicolella, C. H. Turner, D. M. Warren, M. C. Mahaney, Heritability of lumbar trabecular bone mechanical properties in baboons. *Bone* **46**, 835–840 (2010). [doi:10.1016/j.bone.2009.11.002](https://doi.org/10.1016/j.bone.2009.11.002) [Medline](#)
111. J. L. Joganic, K. E. Willmore, J. T. Richtsmeier, K. M. Weiss, M. C. Mahaney, J. Rogers, J. M. Cheverud, Additive genetic variation in the craniofacial skeleton of baboons (genus *Papio*) and its relationship to body and cranial size. *Am. J. Phys. Anthropol.* **165**, 269–285 (2018). [doi:10.1002/ajpa.23349](https://doi.org/10.1002/ajpa.23349) [Medline](#)
112. L. J. Martin, M. C. Mahaney, A. M. Bronikowski, K. D. Carey, B. Dyke, A. G. Comuzzie, Lifespan in captive baboons is heritable. *Mech. Ageing Dev.* **123**, 1461–1467 (2002). [doi:10.1016/S0047-6374\(02\)00083-0](https://doi.org/10.1016/S0047-6374(02)00083-0) [Medline](#)
113. S. Williams-Blangero, J. Blangero, Heritability of age at first birth in captive olive baboons. *Am. J. Primatol.* **37**, 233–239 (1995). [doi:10.1002/ajp.1350370305](https://doi.org/10.1002/ajp.1350370305) [Medline](#)
114. L. J. Hlusko, K. M. Weiss, M. C. Mahaney, Statistical genetic comparison of two techniques for assessing molar crown size in pedigreed baboons. *Am. J. Phys. Anthropol.* **117**, 182–189 (2002). [doi:10.1002/ajpa.10022](https://doi.org/10.1002/ajpa.10022) [Medline](#)
115. D. D. Miley, M. H. Baumgartner, J. M. Cheverud, C. C. Roseman, J. Rogers, D. E. McLeod, E. Reyes, C. F. Hildebolt, Heritability of alveolar bone loss from periodontal disease in a baboon population: A pilot study. *J. Periodontol.* **82**, 575–580 (2011). [doi:10.1902/jop.2010.100189](https://doi.org/10.1902/jop.2010.100189) [Medline](#)
116. H. L. Hansen, T. L. Bredbenner, D. P. Nicolella, M. C. Mahaney, L. M. Havill, Cross-sectional geometry of the femoral midshaft in baboons is heritable. *Bone* **45**, 892–897 (2009). [doi:10.1016/j.bone.2009.05.028](https://doi.org/10.1016/j.bone.2009.05.028) [Medline](#)
117. V. S. Voruganti, M. E. Tejero, J. M. Proffitt, S. A. Cole, L. A. Cox, M. C. Mahaney, J. A. Rogers, J. H. Freeland-Graves, A. G. Comuzzie, Characterization of ghrelin in pedigreed baboons: Evidence for heritability and pleiotropy. *Obesity (Silver Spring)* **16**, 804–810 (2008). [doi:10.1038/oby.2007.107](https://doi.org/10.1038/oby.2007.107) [Medline](#)
118. J. Rogers, L. J. Martin, A. G. Comuzzie, J. J. Mann, S. B. Manuck, M. Leland, J. R. Kaplan, Genetics of monoamine metabolites in baboons: Overlapping sets of genes influence levels of 5-hydroxyindolacetic acid, 3-hydroxy-4-methoxyphenylglycol, and homovanillic acid. *Biol. Psychiatry* **55**, 739–744 (2004). [doi:10.1016/j.biopsych.2003.12.017](https://doi.org/10.1016/j.biopsych.2003.12.017) [Medline](#)
119. S. A. Cole, L. J. Martin, K. W. Peebles, M. M. Leland, K. Rice, J. L. VandeBerg, J. Blangero, A. G. Comuzzie, Genetics of leptin expression in baboons. *Int. J. Obes. Relat. Metab. Disord.* **27**, 778–783 (2003). [doi:10.1038/sj.ijo.0802310](https://doi.org/10.1038/sj.ijo.0802310) [Medline](#)
120. C. M. Kammerer, Effects of sex, age, weight, and heredity on blood pressure in baboons. *Am. J. Hum. Biol.* **7**, 149–158 (1995). [doi:10.1002/ajhb.1310070203](https://doi.org/10.1002/ajhb.1310070203) [Medline](#)

121. Z. Johnson, L. Brent, J. C. Alvarenga, A. G. Comuzzie, W. Shelledy, S. Ramirez, L. Cox, M. C. Mahaney, Y.-Y. Huang, J. J. Mann, J. R. Kaplan, J. Rogers, Genetic influences on response to novel objects and dimensions of personality in *Papio baboons*. *Behav. Genet.* **45**, 215–227 (2015). [doi:10.1007/s10519-014-9702-6](https://doi.org/10.1007/s10519-014-9702-6) [Medline](#)
122. A. G. Comuzzie, S. A. Cole, L. Martin, K. D. Carey, M. C. Mahaney, J. Blangero, J. L. VandeBerg, The baboon as a nonhuman primate model for the study of the genetics of obesity. *Obes. Res.* **11**, 75–80 (2003). [doi:10.1038/oby.2003.12](https://doi.org/10.1038/oby.2003.12) [Medline](#)
123. C. E. Jaquish, T. Dyer, S. Williams-Blangero, B. Dyke, M. Leland, J. Blangero, Genetics of adult body mass and maintenance of adult body mass in captive baboons (*Papio hamadryas* subspecies). *Am. J. Primatol.* **42**, 281–288 (1997). [doi:10.1002/\(SICI\)1098-2345\(1997\)42:4<281:AID-AJP3>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1098-2345(1997)42:4<281:AID-AJP3>3.0.CO;2-T) [Medline](#)
124. E. A. Archie, J. Tung, M. Clark, J. Altmann, S. C. Alberts, Social affiliation matters: Both same-sex and opposite-sex relationships predict survival in wild female baboons. *Proc. Biol. Sci.* **281**, 20141261 (2014). [Medline](#)
125. R. J. Hijmans, “raster” (R package 3.4-10, 2015); <https://cran.r-project.org/web/packages/raster/index.html>.
126. J. G. Ruby, K. M. Wright, K. A. Rand, A. Kermany, K. Noto, D. Curtis, N. Varner, D. Garrigan, D. Slinkov, I. Dorfman, J. M. Granka, J. Byrnes, N. Myres, C. Ball, Estimates of the heritability of human longevity are substantially inflated due to assortative mating. *Genetics* **210**, 1109–1124 (2018). [doi:10.1534/genetics.118.301613](https://doi.org/10.1534/genetics.118.301613) [Medline](#)
127. J. Tung, M. J. Charpentier, S. Mukherjee, J. Altmann, S. C. Alberts, Genetic effects on mating success and partner choice in a social mammal. *Am. Nat.* **180**, 113–129 (2012). [doi:10.1086/665993](https://doi.org/10.1086/665993) [Medline](#)
128. K. R. Amato, S. Van Belle, A. Di Fiore, A. Estrada, R. Stumpf, B. White, K. E. Nelson, R. Knight, S. R. Leigh, Patterns in gut microbiota similarity associated with degree of sociality among sex classes of a neotropical primate. *Microb. Ecol.* **74**, 250–258 (2017). [doi:10.1007/s00248-017-0938-6](https://doi.org/10.1007/s00248-017-0938-6) [Medline](#)
129. C. E. Thomson, I. S. Winney, O. C. Salles, B. Pujol, A guide to using a multiple-matrix animal model to disentangle genetic and nongenetic causes of phenotypic variance. *PLOS ONE* **13**, e0197720 (2018). [doi:10.1371/journal.pone.0197720](https://doi.org/10.1371/journal.pone.0197720) [Medline](#)