1 **Title**

2 Functional host-specific adaptation of the intestinal microbiome in hominids

3

4 **Authors**

5 Rühlemann MC[1,2,†], Bang C[1], Gogarten JF[3,4,5,6], Hermes BM[7,8], Groussin M[1], Waschina S[9], Poyet

6 M[8], Ulrich M[4,5], Akoua-Koffi C[10], Deschner T[11], Muyembe-Tamfum JJ[12], Robbins MM[13],

7 Surbeck M[14,15], Wittig RM[16,17], Zuberbühler K[18,19], Baines JF[7,8], Leendertz FH[4,5], Franke A[1,†]

8

9 **Affiliations**

10 1 Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany
11 2 Institute for Medical Microbiology and Hospital Epidemiology, Hannover Medical School
12 Hannover, Germany
13 3 Applied Zoology and Nature Conservation, University of Greifswald, Greifswald, Germany
14 4 Helmholtz Institute for One Health, Helmholtz-Centre for Infection Research (HZI), Greifswald,
15 Germany
16 5 Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute, Berlin, Germany
17 6 Viral Evolution, Robert Koch Institute, Berlin, Germany
18 7 Evolutionary Genomics, Max Planck Institute for Evolutionary Biology, Plön, Germany
19 8 Institute of Experimental Medicine, Kiel University, Kiel, Germany
20 9 Nutriinformatics Research Group, Institute for Human Nutrition and Food Science, Kiel University,
21 Kiel, Germany
22 10 Training and Research Unit for in Medical Sciences, Alassane Ouattara University / University
23 Teaching Hospital of Bouaké, Bouaké, Côte d'Ivoire
24 11 Comparative BioCognition, Institute of Cognitive Science, University of Osnabrück, Osnabrück,
25 Germany
26 12 National Institute for Biomedical Research, National Laboratory of Public Health, Kinshasa,
27 Democratic Republic of the Congo
28 13 Department of Primate Behavior and Evolution, Max Planck Institute for Evolutionary
29 Anthropology, Leipzig, Germany
30 14 Department of Human Evolutionary Biology, Harvard University, Cambridge, Massachusetts,
31 United States of America
32 15 Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
33 16 Institute of Cognitive Sciences, CNRS UMR5229 University Lyon 1, Bron Cedex, France
34 17 Taï Chimpanzee Project, CSRS, Abidjan, Côte d'Ivoire
35 18 Institute of Biology, University of Neuchatel, Neuchatel, Switzerland
36 19 School of Psychology & Neuroscience, University of St Andrews, St Andrews, Scotland (UK)

37

38 † To whom correspondence may be addressed. Email: m.ruehlemann@ikmb.uni-kiel.de or

39 a.franke@mucosa.com

Summary

Fine-scale knowledge of the changes in composition and function of the human gut microbiome compared that of our closest relatives is critical for understanding the evolutionary processes underlying its developmental trajectory. To infer taxonomic and functional changes in the gut microbiome across hominids at different timescales, we performed high-resolution metagenomic-based analyses of the fecal microbiome from over two hundred samples including diverse human populations, as well as wild-living chimpanzees, bonobos, and gorillas. We find human-associated taxa depleted within non-human apes and patterns of host-specific gut microbiota, suggesting the widespread acquisition of novel microbial clades along the evolutionary divergence of hosts. In contrast, we reveal multiple lines of evidence for a pervasive loss of diversity in humans populations in correlation with a high Human Development Index, including evolutionarily conserved clades. Similarly, patterns of co-phylogeny are found to be disrupted in humans. Together with identifying individual microbial taxa and functional adaptations that correlate to host phylogeny, these findings offer new and exciting insights into specific candidates playing a role in the diverging trajectories of the gut microbiome of hominids, demonstrating that repeated horizontal gene transfer and gene loss, as well as the adaptation to transient microaerobic conditions appear to have played a role in the evolution of the human gut microbiome.

Introduction

Human gut microbiome research has demonstrated that numerous factors, including diet, environment, and lifestyle influence the structure of the human gut microbiota, which in turn have profound impacts on human health and disease (Johnson et al., 2019; Lloyd-Price et al., 2019; Vangay et al., 2018). To date, the majority of these studies were conducted on sample collections from high-income countries, however growing efforts to include humans from diverse global populations are underway, thereby providing an additional angle to investigate and evaluate shared and specific microbiome properties across human populations (McCall et al., 2020; Schaan et al., 2021; E. D. Sonnenburg and Sonnenburg, 2019; Vangay et al., 2018). These efforts provided an opportunity to discover signatures of host geography and lifestyle that go beyond conventional differences in diversity parameters in the gut community. For

instance, they revealed elevated rates of horizontal gene transfer (HGT) that correlate with the Human Development Index (HDI, a statistical composite index of indicators encompassing life expectancy, education, and income; (UNDP (United Nations Development Programme), 2022)) of a population, suggesting that gut microbiota constantly acquire new functions in conjunction with host lifestyle changes (Groussin et al., 2021). However, the mechanisms that link changes in gut microbial structure with host behavior and ecology remain largely unexplored.

Additional critical insight into understanding patterns of diversity and composition among human gut communities can be obtained from comparative surveys of the hominid gut microbiota. Humans, chimpanzees, bonobos, and gorillas show increasingly divergent gut microbiota, with more distantly related species exhibiting more divergent community composition (phylosymbiosis; (Brooks et al., 2016)). At the same time, the phylogeny of some of their individual microbial lineages parallels their own phylogeny (codivergence; (Groussin et al., 2020; Suzuki et al., 2022)). Both patterns of phylosymbiosis and codivergence are suggestive of long-term effects of hominid evolution on their communities of symbionts (Groussin et al., 2017; Moeller et al., 2016; Ochman et al., 2010). Notably, results from comparative marker-gene analyses suggest co-diversifying members of the hominid gut microbial communities (both prokaryotic and phage) are lost and replaced with human lineages when animals leave their natural environments and are moved into captivity (Gogarten et al., 2021; Nishida and Ochman, 2021). However, poor taxonomic resolution and a lack of functional characterization precludes a deeper understanding of processes driving these changes.

Functional analyses from shotgun metagenomic data revealed a conserved phylogenetic signal across wild non-human primates (NHP), despite dietary changes over an individual's lifespan and between species, suggesting that the evolution of gut microbiota within wild NHP is partially constrained by host genetics and physiology (Amato et al., 2019). A comparative meta-analysis of microbial functions in NHP and diverse human populations observed a comparable loss of biodiversity in captive NHP and human populations from regions with higher HDI (Manara et al., 2019), supporting previous findings (Moeller et al., 2014). However, overall, the functional characterization of NHP microbiota in previous studies has been limited

to only selected microbial taxa without addressing broader scale functional changes and specific alterations especially in African great apes and humans. As such, robust comparative functional analyses are still needed for a comprehensive understanding of how gut microbiota have evolved with hominids and shaped the current structure and functional capabilities (and deficits) of the human gut microbiome.

To better elucidate host-microbiome interactions in the hominid gut in an evolutionary context, we present a large-scale comparative study of wild non-human apes (NHA) and humans from geographically distinct populations spanning two continents. Functional shotgun metagenomic sequencing was performed on feces samples of wild-living great apes from six African countries, including two gorilla species, three chimpanzee subspecies, and bonobos, and combined with published data from gorillas and chimpanzees from the Republic of Congo (Campbell et al., 2020). Additionally, we sequenced human fecal samples from two African populations (Mossoun et al., 2017) from rural villages of the Taï region in Côte d'Ivoire (HDI_{2021} = 0.550; (UNDP (United Nations Development Programme), 2022)) and the Bandundu region near Salonga National Park, Democratic Republic of the Congo (HDI_{2021} = 0.479), along with samples from Germany (HDI_{2021} = 0.942), and included a published dataset from Denmark (HDI_{2021} = 0.948; (Hansen et al., 2018)) to incorporate varying degrees of HDI. Using this extensive data resource, we created a comprehensive catalog of high-quality prokaryotic genomes assembled from metagenomic data, which we annotated on a taxonomic and functional level. We subsequently explored patterns of diversity and host-specificity for both taxonomic groups and functions, which reveals intriguing patterns associated with human gut microbial communities, convergent functional adaptations across lineages, and the potential mechanisms driving these patterns.

An expanded catalog of microbial genomes from the hominid gut

Using 224 shotgun metagenomic datasets (Suppl. Table S1) from fecal samples of humans (Côte d'Ivoire (CIV), n=12, Dem. Rep. of the Congo (DRC), n=12, Denmark (DK), n=24, and Germany (GER), n=24) and non-human apes, including two gorilla subspecies (*Gorilla gorilla gorilla,* Gabon (GAB), n=8; *Gorilla beringei beringei,* Uganda (UGA), n=11, and Republic of Congo (CG), n=28), three chimpanzee subspecies (*Pan troglodytes verus,* CIV, n=55; *P.t. troglodytes*, GAB, n=11, and CG, n=18; *P.t. schweinfurthii*, UGA, n=12), and bonobos (*Pan*

paniscus, DRC, n=12), we reconstructed a total of 7,700 metagenome-assembled genomes (MAGs) ensuring maximum completeness and low contamination using multiple binning algorithms and dedicated curation and scoring tools ((Rühlemann et al., 2022); see Methods). The most MAGs (quality score > 50%) were reconstructed for the most sampled subgroup of great ape, *P.t. verus* (n=2,182), while the average number of reconstructed MAGs per sample was the highest for *P.t. schweinfurthii* (mean=74.5). Library size / number of sequencing reads was highly correlated with total assembly size ($\rho_{Spearman}$ = 0.644), which in turn was directly correlated with the number of bins recovered for a sample ($\rho_{Spearman}$ = 0.966).

To ensure a comprehensive reference for the analysis, the collection of MAGs was combined with two large collection of microbial reference species reconstructed from human fecal metagenomes (UHGGv2, n = 4,744 isolates and MAGs; (Almeida et al., 2021)), and non-human primate fecal metagenomes (n = 1,295 MAGs; (Manara et al., 2019)), resulting in a total of n = 13,739 genome sequences; MAGs were subsequently clustered into 5,777 species-level genome bins (SGBs; 95% ANI) using stringent criteria (Suppl. Figure S1a and S1b, Supplementary Table S2). Of these, 1,074 SGBs were not previously covered by either of the two large reference sets, mostly originating from NHA samples (n=956, 89.0%; Suppl. Figure S1c). The highest-quality genome sequence in each SGB was chosen as its representative. Overall quality of SGB representative genomes was high (median quality score = 94.1%; Suppl. Figure S1d). SGB representatives were used as comprehensive reference for the estimation of per-sample abundances (Methods, Suppl. Figure S1e, Suppl. Table S3). In total 3,287 SGBs, encompassing 21 bacterial and two archaeal phyla (Suppl. Figure S1b), were found present in the dataset. This newly curated catalog of SGBs from gorillas, bonobos, and chimpanzees increases the number of microbial species genomes previously reconstructed from feces by more than ten-fold, and increases mapping success of fecal metagenomes from NHAs to reference genomes by two- to three-fold (Suppl. Figure S1e; (Manara et al., 2019). As expected, only minor proportions of SGBs from human samples were not previously covered by the included large reference collections, with 5.8%, 2.7 %, 1.8 and 1.7% of novel SGBs found in samples from CIV, DRC, DK, and GER, respectively (Suppl. Figure S1e). For both NHAs and humans, the highest percentages of novel diversity were observed within the phyla Bacteroidota and Spirochaetota, and, to a lesser degree, within Firmicutes and Firmicutes A for NHAs only (Suppl. Figure S1f). Generally, recovered clusters were highly host specific.

While the 7,700 MAGs spanned 1,787 of the final SGBs, only for 48 of these SGBs MAGs were reconstructed from samples of more than a single host genus.

Within sample diversity varied considerably between host (sub-)species and with sequencing depth (Suppl. Figure S1g). The used SGB collection covers large proportions of the diversity found within the human gut (Suppl. Fig. S1e) and Faith's phylogenetic diversity (PD) incorporates SGB relatedness in the diversity calculation, which enables a better estimate of total diversity of a community then simpler richness estimates from taxonomic group abundances. Phylogenetic diversity (PD) at an even mapping depth of 1 million reads per sample showed significantly lower diversity in humans compared to all African great ape hosts ($P_{Wilcoxon} = 1.2 \times 10^{-13}$; Figure 1a). Comparing individual human populations to African great ape hosts revealed that humans from GER and DRC showed lower means than all NHA group (all $P_{Wilcoxon} < 0.05$), while humans from CIV and DK exhibited high variance and were found to have significant lower diversity than all NHA hosts ($P_{Wilcoxon} < 0.05$), except for *G. g. gorilla* and *P.t. troglodytes* ($P_{Wilcoxon} > 0.05$). These two great ape taxa were found to have the lowest library sizes, low numbers of recovered genomes and lowest mapping efficiency. Taken together, this suggests that phylogenetic diversity in these hosts may be biased by lower representation in the reference database and that the reduced diversity could be an artefact of this. Consequently, samples with less than 1 million mapped reads were removed from further analyses. This resulted in the removal of samples from the analysis for the groups *G.b. beringei* (n=1), *G.g. gorilla* ($n_{GAB}=5$, $n_{CG}=1$), *P.t. troglodytes* ($n_{GAB}=4$), *P.t. schweinfurthii* (n=1) and humans from CIV (n=1). Of note, the lowest and highest mean values for within-sample phylogenetic diversity were found for the human subgroups from Germany and Denmark, respectively, the latter exhibiting the only significant differences between human subsets ($P_{Wilcoxon} < 0.05$ vs. DRC and GER), contradicting previous reports of lower alpha diversity generally found in high HDI countries (McCall et al., 2020; Vangay et al., 2018). Additionally, the considerable differences in PD observed between CIV and DRC highlight the diversity found between human populations and the need to better characterize human gut microbiome diversity. Therefore, while total diversity of some host groups was likely not exhaustively sampled, especially considering lower abundant microbial clades, the presented reference collection of high-quality metagenome reconstructed genomes likely represents the current best resource for an in-depth taxonomic and functional assessment of hominid

fecal microbiomes and highlights that some human populations, including those sampled in this analysis, have lost considerable microbial diversity in their guts.

Phylosymbiosis in hominids is strongly supported by community structure

Phylosymbiosis, a pattern in which microbial community divergence parallels that of the hosts, can be a sign of community-level co-evolution of host and microbiota, indicative of host-microbe relationships maintained over evolutionary timescales (Brooks et al., 2016). To investigate such patterns, we used six different measures of beta diversity. Four were based on phylogenetic or taxonomic distance metrics (weighted and unweighted UniFrac, as well as genus level Aitchison and Jaccard distance), and two additional metrics considered the functional capacities of the community (KEGG ortholog (KO) abundance and presence/absence patterns; Suppl. Figure 2). Strong signals for phylosymbiosis were found for Jaccard, and unweighted UniFrac distances ($P < 0.001$; $Q_{Bonferroni} < 0.01$), as well as a less pronounced signal for Aitchison distance and the abundance of functional groups ($P < 0.01$, $Q_{Bonferroni} = 0.0534$ and $Q_{Bonferroni} = 0.0585$, respectively; Figure 1b and c, Suppl. Figure 2, Suppl. Table S4), but not for the distances based on presence and absence of KOs ($P = 0.055$, $Q_{Bonferroni} = 0.33$) and the weighted UniFrac distance ($P = 1$). These results suggest phylosymbiotic divergence in the general microbial structure of hominid microbiota (Jaccard and unweighted UniFrac are based on presence/absence of microbial clades) which in part is paralleled by shifted abundances of distinct taxonomic and functional groups, generally in line with previous observations in other host systems (Brooks et al., 2016). This reduced signal found for taxon-level diversity might be the result of fluctuating clade abundances and functions in response to changing environmental factors, including diet, indicating a high functional plasticity of the hominid gut microbiome in response to immediate influences, while structural changes through acquisition and loss of microbial clades might rather result from longer-term adaptations.

SGBs shared between host groups were analyzed in a permutation-based framework accounting for differences in sequencing depth and group sizes (see Methods for details). We found that especially human-associated SGBs were strongly depleted across multiple NHA groups (Figure 1d), while such signal found in the opposite direction were less pronounced, indicating the widespread acquisition of novel microbial clades in the human intestinal

233 microbiome. A similar pattern was found for *G. b. beringei* and multiple *Pan* subspecies,
234 however not for *G.g. gorilla*. *G.g. gorilla* and *P.t. troglodytes* are sympatric species and were
235 sampled in the same environment in the Republic of Congo and Gabon. The absence of excess
236 strong host-specific signals between these particular taxa might point towards an effects of a
237 shared environment influencing microbiome structure. Human subgroups from Europe and
238 Africa showed strong pairwise SGB-sharing between CIV and DRC, and DK and GER,
239 respectively, however not across geographic regions, suggesting strong connection of the
240 microbiome with environment and lifestyle differences. All human subgroups exhibit strong
241 depletion of SGB-sharing with all other host genera, which resulted in a clear separation of
242 the human microbiota from that of other hominids.
243
244 **Fecal microbiome of European populations marked by loss of evolutionarily conserved core**
245 **microbiota**
246
247 Abundance difference in microbial clades between humans and NHAs and between human
248 communities with differing environments, such as living in rural or urban regions, in regions
249 of the world with lower or higher HDI, can give insights into microbiome-mediated
250 adaptations to environmental changes in the distant and more recent past. We analyzed and
251 compared the abundance profiles of gut microbes between NHAs and humans, including
252 individuals living in rural, lower HDI areas of Africa (CIV and DRC) as well as individuals residing
253 in urban, higher HDI regions within Europe (GER and DK). For all following taxonomic and
254 functional comparisons, we restricted the analysis to human and *Pan* (chimpanzees and
255 bonobos) samples to obtain focused insights into the microbiota divergence since their hosts
256 diverged about 7-8 million years ago (Langergraber et al., 2012).
257
258 A total of 310 microbial genera were included in the analysis, of which 173 were found to be
259 differentially abundant ($Q_{Bonferroni}$<0.05, Figure 1d, Suppl. Table S5) in at least one of these
260 comparisons between human subgroups, or between humans and NHAs, and subsequently
261 sorted into one of four groups. We identified 57 taxa with increased abundances among
262 humans from high HDI regions in Europe, such as *Akkermansia*, *Bacteroides*, and *Alistipes*
263 (Figure 2a and b). We additionally found 39 taxa that are enriched in the two African
264 populations such as *Cryptobacteroides, Prevotella*, and *Succinivibrio* (Figure 2a and b). Overall,

the marker taxa of European microbiomes that we detected are in agreement with previous findings (Jha et al., 2018; Pasolli et al., 2019). Our approach allowed us to identify bacterial taxa that exhibit differential abundance profiles between humans and NHAs that are independent from the human populations (Figure 2b). We found 74 taxa, such as *SIG603* that have increased abundance in NHAs and 50 taxa with increased abundance in humans, of which 15 do not show an association to either of the human subgroups, such as *Coprococcus* and *Agathobacter*. Interestingly, taxa depleted in the microbiome of European individuals compared to humans from Africa are more likely to be also abundant (>0.1%) in the microbiomes of NHAs ($P < 0.001$; Figure 1c), suggesting a loss of evolutionary conserved clades in these populations.

Widespread changes in fecal microbiome function between hosts and across human communities

Taxon-specific changes reflect broad-scale differences between host-groups. A focused analysis of microbial functions can give insights into the specific driving forces of such community-level changes. We performed analysis of abundance differences of 6,340 KEGG orthologs (KOs, (Kanehisa and Goto, 2000)) in NHA (genus *Pan*) vs. human fecal microbiota and humans in European and African societies and found significant abundance differences in 1,092 (17.2 %) and 881 (13.9 %) KOs, respectively ($Q_{Bonferroni}<0.05$; Figure 2a, Suppl. Table S6).

Analysis of higher-level KEGG annotations, including e.g. complete pathways overrepresented among differentially abundant KOs, revealed seven annotations with enrichment of KOs higher abundant in NHAs compared to humans and nine annotations conversely enriched in humans. Sulfur metabolism (map00920) was generally overrepresented in differentially abundant KOs with individual effect directions associating with both groups ($Q_{Bonferroni} < 0.05$; Figure 1j, Suppl. Table S7). NHA associated categories md:M00356, md:M00563 (methanogenesis) and path:map00680 (methane metabolism) clearly indicate a higher abundance of methanogenic archaea. Additionally, we find multiple categories involving ribosomes, including distinct ribosome annotations of archaea, which we could confirm using the pangenome distribution of these KOs (e.g. 92.9% of SGBs with K02866 [large subunit ribosomal protein L10e] belonging to the domain Archaea). In humans, we find enrichments

in categories covering antimicrobial resistance (path:map01501, br:ko01504), bacterial mobility (path:map02030, path:map02040, br:ko02035), biofilm formation (path:map02026) and prokaryotic defense systems (br:ko02048), suggesting a generally higher abundance of virulence genes in human microbiomes, independent of geography. Pangenome distribution of the significantly different KOs additionally confirm, these are not driven by specific microbial clades since they are present across the entire phylogeny.

Enrichments of higher-level KEGG annotations between humans from Africa and Europe were exclusively found for the European subgroup (n=16). These involved vitamin B_{12} / cobalamin biosynthesis (md:M00122, M00924, M00925, and path:map00860) and antimicrobial resistance genes (br:ko01504, path:map01502). Increased antimicrobial functions may be explained by a higher use of antibiotics in human healthcare and extensive animal husbandry, as well as by environmental pollution (Groussin et al., 2021). The increased abundance of vitamin B_{12}-producing microorganisms in the feces of Europeans may be driven by higher dietary intake of meat and dairy products from ruminants, as these food groups contain microorganisms with this metabolic capacity (Watanabe and Bito, 2018). Additionally, we find multiple categories suggesting a community-level shift towards oxidative carbohydrate metabolism, e.g. glycolysis / gluconeogenesis (path:map00010), pyruvate metabolism (path:map00620), phosphotransferase system (path:map02060), and V/A-type ATPase (md:M00159).

Success of taxa associated to European populations relates to oxidative carbohydrate metabolism

Community-level changes in abundances of functional groups give insights into some aspects of adaptations, however, they are overproportionally driven by highly abundant taxonomic groups. We applied pangenome analysis to specifically identify individual genes and functions enriched or depleted within specific microbial clades overrepresented the gut of human individuals residing within Europe in comparison to other human associated taxa. Analyses were restricted to four bacterial families, *Bacteroidaceae, Lachnospiraceae*, *Oscillospiraceae*, and *Ruminococcaceae*, for which sufficient numbers of SGBs (n>=10 in each of both groups) for pangenome analysis were recovered. SGBs in these family represent large proportions of

the overall human microbiome (mean = 53.2%), with no significant differences between the subgroups ($P_{Kruskal\text{-}Wallis}$ = 0.2). The analysis was conducted at the family-level, as higher taxonomic ranks would increase clade-specific functional biases. To account for between-family functional differences, functional differences of SGBs associated with Europeans were first analyzed within microbial families using Fisher's exact test and subsequently subjected to unweighted meta-analysis using Z-scores to leverage shared signals.

We identified 167 enriched and 30 depleted KOs in pangenomes associated with the European populations ($Q_{Meta,Bonferroni}$<0.05; Suppl. Tables S8 and S9). Using higher-level KEGG annotations (modules, pathways, BRITE hierarchies), we found that 11 of these annotations were overrepresented in the dataset ($Q_{Bonferroni}$ < 0.05, Figure 3c, Table 1, Suppl. Table S10). Among these were multiple groups involved in carbohydrate metabolism enriched in taxa associated with Europeans, specifically pointing at aerobic breakdown of sugar molecules for ATP generation (citrate cycle: md:M00009, path:map00020; pentose phosphate pathway: md:M00004; V/A-type ATPase, prokaryote: md:M00159), confirming patterns seen also in the community-level analysis of functional abundances. These signals strongly suggest that the selection for taxa in the gut of humans living within Europe is connected to a diet rich in carbohydrates and potentially the adaptation to transient microaerobic conditions in the gut environment, using oxidative phosphorylation as a mean to release energy from nutrients, which is more efficient than strictly anaerobic fermentation (Jurtshuk, 1996). However, reduced fermentation can impact short-chain fatty acid production, which can, in turn, potentially negatively affect the host's intestinal epithelial cells and metabolism (Deleu et al., 2021). Additionally, we found a pathway connected to the histidine degradation (md:M00045). Especially gut microbial histidine metabolism has been discussed with relevance to human health, as it was shown that an intermediary product of histidine degradation, imidazole propionate, was increased in type 2 diabetic individuals in a large study of almost 2,000 individuals, and that this increase was directly connected to the microbiota and overall unhealthy dietary habits, however independent of dietary histidine intake (Molinaro et al., 2020). The enzyme urocanate hydratase (EC:4.2.1.49; K01712) is responsible for the interconversion between urocanate and imidazole propionate in the histidine degradation pathway. We find this gene significantly enriched in three of the four bacterial families analyzed (P < 0.05, Figure 3d), with a clear trend visible also for

Oscillospiraceae (P = 0.052). This suggests a microbiome-encoded pre-disposition to metabolic disorders in European human communities.

We did not find any higher-level KEGG annotations significantly enriched among the taxa not enriched in Europeans. The relative lower fitness of these taxa may not result from a single mechanism associated to European lifestyles, but rather from multiple selective forces specific to the individual clades.

Convergent host-specific adaptations are found across microbial families

Shared gene gains or losses across multiple microbial clades can indicate a response to specific host intestinal environments, leading to functions being acquired (and selected for) multiple times independently. We performed a pangenome analysis of genera shared between humans and NHAs (n=36) to identify such patterns of convergent adaptation. To control for higher-level clade effects, functional repertoires (KEGG Ontology terms) were compared between SGB pangenomes at the genus-level from NHAs (Suppl. Table S11) and humans; results were then combined in a meta-analysis across genera. In total, 78 KO terms were identified as carrying signatures of convergent adaptation to the respective host group, with 57 of these signatures associating with humans and 21 associating with NHPs ($Q_{Bonferroni}$<0.05; Figure 4a, Suppl. Table S12). Among the human-associated KOs, we found multiple functional groups hinting again at an adaptation to increased oxygen by utilization of oxygen as an electron acceptor within the respiratory chain, such as cytochrome *bd* ubiquinol oxidase subunits (cydA, cydB), as well as adaptation to increased oxidative stress through ferritin (ftnA) and thioredoxin-dependent peroxiredoxin (BCP). Among the KOs enriched in NHP pangenomes we found an outer membrane factor (TC.OMF), a major facilitator superfamily (MFS) transporter (lrmB), 1-epi-valienol-7-phosphate kinase (acbU), and two KOs annotated as polyketide synthases (rhiA, pksN). Products produced by polyketide synthases have diverse functions, including antibiotic activity, virulence and support of symbiotic relationships (Ridley et al., 2008). OMFs and MFS form transmembrane complexes for the transport of a large variety of solutes (Yen et al., 2002), including *e.g.* carbohydrates, metal ions, amino acids, and export of toxic compounds (Davidson and Chen, 2004; Lee et al., 2014). How these potential adaptations relate to NHA hosts is unclear, however, they might indicate adaptation

392 to a diverse diet (Ejby et al., 2016; Tsujikawa et al., 2021) or metabolism of plant-derived
393 xenobiotics (Rodríguez-Daza et al., 2021).
394
395 *Prevotella* represented the largest genus-level clade in the dataset (n=212, 3.7% of all SGBs).
396 While we found this genus across all host species, it is largely decreased in abundance and
397 prevalence in Europeans. We selected this genus for further analysis to elucidate potential
398 functional mechanisms driving the observed patterns. Enrichment analysis revealed 126 KOs
399 with distinct prevalence patterns (Q<0.05; Figure 4b). The most striking difference was found
400 for the cytochrome *bd* ubiquinol oxidases subunits 1 and 2 (cydA and cydB), which were found
401 in 101 of 114 human-associated *Prevotella* SGBs (incl. UHGGv2 genomes), but present in only
402 two out of 72 SGBs from NHAs. It is important to note that these prevalence differences are
403 not driven by single lineages within the *Prevotella* genus that would have distribution ranges
404 restricted to single hosts. Instead, they are observed across multiple sibling clades spanning
405 the entire phylogenetic tree of the taxon (Figure 4c). Cytochrome *bd* oxidases are involved in
406 stress responses, most prominently in transiently microaerobic environments (Giuffrè et al.,
407 2014). Comparison of the *Prevotella* species tree (reconstructed using all SGB representatives
408 recovered from the dataset (n = 184)) and the cydA gene phylogeny exhibit widespread
409 incongruencies between their tree topologies (Figure 4c). We performed a tree reconciliation
410 using a duplication-transfer-loss (DTL; (Kundu and Bansal, 2018)) model between the
411 *Prevotella* and cydA phylogenies, which revealed frequent events of gene transfer ($\overline{T}$ =
412 41.552) between branches (including distant ones) of the *Prevotella* phylogeny and
413 subsequent losses ($\overline{L}$ = 23.52) in the NHA-associated clades. Interestingly, the two cydA-
414 carrying SGBs found in NHAs are phylogenetically distant, however, their cydA genes are
415 highly similar and most likely the result of a transfer from one to the other. The gene transfer
416 events and mappings were robust across 1,000 reconciliations with different starting seeds,
417 with 89.2% of all events and 67.6% of all mappings found with 100% consistency. These results
418 suggest that the enrichment of cytochrome *bd* ubiquinol oxidases observed in humans
419 compared to NHAs are the result of multiple reoccurring events of gene loss and horizontal
420 gene transfer between *Prevotella* clades within the hominid gut.
421
422 **Co-phylogeny is rare in spore-forming microbes and disrupted in humans**
423

Patterns of co-phylogeny between host and microbes can result from close interaction, or even interdependence in extreme cases, and congruent metabolic pathways from co-evolutionary trajectories. Using stringent selection criteria, we subjected 209 subtrees of the SGB phylogeny for co-phylogeny analysis (see Methods). The subtrees spanned 46 families and 945 (28.8%) SGBs present in the dataset in addition to 77 SGBs from the UHGGv2 catalog. We used a Mantel-test based framework and permutation to detect co-phylogeny signals ((Hommola et al., 2009), see Methods for details). When defining co-phylogeny candidates based on a mean P-value < 0.05 across all permutations, 56 of 209 subtrees (26.8%) qualified as exhibiting co-phylogenetic patterns (Suppl. Table S13). These subtrees cover 312 of the 1051 SGBs (30.7%) included in the analysis and 5.84% of the total 5,345 hominid SGBs (excluding the SGBs from Manara et al.). All results and subtrees can be inspected online (https://mruehlemann.shinyapps.io/great_apes_shiny_app). By visually inspecting subtrees with co-phylogeny signals, we find many candidates microbial phylogenies that do not follow host phylogeny, *e.g.,* in a subgroup of the genus *Cryptobacteroides* (Figure 5a). Such signals suggest that co-phylogeny within the *Gorilla* and *Pan* clades can result in statistically significant Mantel tests, despite topological incongruences of human-derived sequences, for which no host sister (sub-)species from the same genus is available.

Human-derived genomes were found in 149 (71.3%) of the tested subtrees, of which 38 (25.5%) were co-phylogeny candidates. Similarly, 18 of 60 (30%) tested subtrees without human-derived representatives exhibited co-phylogeny signals, *e.g.*, *W0P29-013* spp. (Figure 5a). Overall, 21.8% (n = 82 out of 377) of human-derived SGBs in the analysis were found in co-phylogenetic subtrees, which is significantly less compared to 35.7% (n = 231 out of 647) of NHA-derived SGBs ($P_{Fisher} = 2 \times 10^{-6}$). Co-phylogeny signals, defined based on the ratio of SGBs in trees exhibiting co-phylogeny patterns and the total SGBs in the family included in the analysis, differed strongly. Six families showed excessive signals of co-phylogeny and nine families a significant depletion ($Q_{Fisher, FDR} < 0.05$; Figure 5b, Suppl. Table S14). The highest co-phylogeny ratio was found for the family *Dialisteraceae* (Cophyl-Ratio = 100%, Q = 4.61 x 10^{-7}), which in the analysis were entirely represented by 14 SGBs in the genus *Dialister*. *Dialister* are a common, but rather neglected members of the gut microbiota which have been found increased (Vals-Delgado et al., 2021; Zheng et al., 2018) and decreased (Joossens et al., 2011)

with various human diseases, hence their relation to human health remains unclear. However, on species, *Dialister invisus,* was found to be moderately transmissible between human mother-infant pairs and within households in a large meta-analysis (Valles-Colomer et al., 2023). The strongest depletions were found for *e.g.* the families *Lachnospiraceae* and *Treponemataceae*, the former confirming previous results for this clade (Moeller et al., 2016). The latter, *Treponemataceae*, especially the genus *Treponema D,* were found depleted in humans living in Europe and occur in anaerobic sediments (Thingholm et al., 2021), serving as (intermediary) reservoirs for transmission to humans and NHPs, which can disrupt co-phylogenetic signals by constant re-introduction to the community.

When comparing host groups, the proportion of SGBs with co-phylogeny signal is significantly reduced in all human subgroups compared to the NHA hosts (all $Q_{Wilcoxon} < 0.05$; Figure 5c). Further, humans from Germany and Denmark exhibited even lower proportions of co-phylogeny SGBs compared to the two African human populations ($Q_{Wilcoxon} < 0.05$), but not to each other ($P_{Wilcoxon} = 0.68$). The human subgroups from Africa did not differ in their co-phylogeny proportions ($P_{Wilcoxon} = 0.096$). These results suggest a loss of wild great ape associated clades in the intestinal microbiota of humans independent of their geographic origin and the introduction of novel microbial partners with changing environment and lifestyle, confirming again findings from previous analyses.

Signals of cophylogeny suggest a strong association and possible adaptation with the host and the reduction of genome size and gene content are expected patterns connected to this process (Toft and Andersson, 2010) which has previously been shown also for microbes with codivergence patterns in human population (Suzuki et al., 2022). To explore whether these processes could be detected in our dataset, we analyzed genome size and gene count, as well as 43 microbial traits inferred from genome-level annotations for signals in association with co-phylogeny using logistic regression, while controlling for phylogenetic relatedness (see Methods; Suppl. Table S15). Out of 45 analyzed traits, 13 were found significantly depleted in clades exhibiting co-phylogeny signals ($Q_{Bonferroni} < 0.05$; Figure 5d; Suppl. Table S16), including genome size, gene count, capabilities to use multiple simple sugars, and bile susceptibility. Phylum-level analyses confirm this overall trend especially for Actinobacteriota, Firmicutes A, Firmicutes C and Verrucomicrobiota ($P_{Wilcoxon} < 0.05$), while Proteobacteria exhibited an

488 inverse signal (Figure 5d). This inverse signature might be explained by host specific
489 acquisition of genes found *e.g.* for *E. coli* across different and diverse hosts (Tiwari et al.,
490 2023), however these findings needs to be further investigated. A higher susceptibility to bile
491 in clades without co-phylogeny signal clearly suggests that bile tolerance is an adaptation to
492 the Intestinal environment of hominid hosts. The depletion of the utilization of simple dietary
493 sugars (traits: D-Xylose [Figure 5f], D-Mannose, Maltose, Glucose fermenter, Sucrose,
494 Trehalose, L-Arabinose; all $Q < 0.05$) in co-phylogeny clades might suggest an adaptation to
495 host-derived complex carbohydrates or other energy sources in the host-context.
496
497 Only one trait – the Arginie dihydolase pathway, also called Arginine deiminase pathway (ADI)
498 – was found enriched with co-phylogeny signal. Argninie metabolism has been widely
499 discussed in the context of host-microbe interaction (Nüse et al., 2023), and ADI specifically
500 was shown to protect bacteria from acid stress in host-association (Casiano-Colón and
501 Marquis, 1988), but was also shown to modulate host immunity (Ghazisaeedi et al., 2022)
502 and as specifically acquired by Saccharibacteria in the process of colonization of mammals
503 (Tian et al., 2022), confirming its potential association with co-phylogeny in hominids.
504 Previous analyses suggested that spore-forming clades are less likely to be exhibiting co-
505 phylogenic patterns, due to their ability to survive outside of the gut, facilitating dispersal
506 between hosts (Groussin et al., 2020; Hildebrand et al., 2021; Moeller et al., 2016; Nayfach et
507 al., 2016). We did not find any negative correlation between spore-formation ability and co-
508 phylogeny patterns ($Q = 1$, Suppl. Table S16), however the annotation of this trait was
509 restricted to only two phyla (Firmicutes and Firmicutes A) and was also rare within these
510 clades, being found in only 54 SGBs across the dataset. As such, whether these results
511 contradict previous findings cannot be concluded in this analysis and warrants future focused
512 analyses.
513
514 We found 67 out of 157 SGBs in *Bacteroidaceae* within subtrees with cophylogeny signals (Co-
515 pylogeny-Ratio = 43.5%), consistent with previous findings based on *gyrB* amplicon data
516 (Moeller et al., 2016). However, no evidence for strict co-phylogeny was found in
517 *Bifidobacteriaceae* (n_{SGB}=10 in the analysis, none with co-pyhlogeny signals), which is
518 inconsistent with findings from the same report (Moeller et al., 2016). Comparing the
519 phylogeny of metagenome-derived *gyrB* sequences and the GTDB marker-gene phylogeny for

Bifidobacterium spp. revealed clear incongruences between both approaches (Figure 5g), which may explain the differences in the presented analysis and previous findings.

Discussion

Here, we present the largest curated dataset of fecal metagenomes derived from wild African great apes and human populations. For this, we surveyed and reconstructed high-quality microbial genomes from the feces wild non-human apes, including gorillas (*Gorilla gorilla gorilla*; *Gorilla beringei beringei*), chimpanzees (*Pan troglodytes verus*; *P.t. troglodytes*; *P.t. schweinfurthii*), and bonobos (*Pan paniscus*) as well as human populations from Africa and Europe. We identified signals of phylosymbiosis across the included hominids, indicating a conserved evolutionary relationship of microbial communities with their host species. Moreover, by employing a comparative approach, we found extensive changes of microbial taxonomic and functional abundances across the intestinal microbiota of NHAs and humans. Previous studies have pointed to “Western” lifestyles as an important factor influencing the intestinal microbiota in humans. Within our human sample population, we were able to confirm differential signals of prokaryotes associated with the European human populations. Importantly, using a comparative dataset of great ape taxa showed that microbial clades lost in Europeans in comparison to African human populations are also found in wild great ape populations. Thus, we suggest that the loss of these taxa might be regarded as the departure from a natural divergence trajectory since their last shared ancestor, cumulating in a mass extinction event of evolutionary conserved members of hominid-associated gut microbiota. While it is tempting to link these changes to industrialization (as previous studies have done), there are many differences between these human populations (e.g., genomic diversity, diet, exercise, sunlight exposure, exposure to antibiotics, population bottlenecks) and it was certainly not possible with the sampling regime here, to determine the particular factors responsible for the variation observed between the human populations sampled here. Due to logistic constraints, preservation methods for fecal samples from the included hosts and host subgroups differ. While these are expected to influence microbiome composition, previous studies show that individual signature are retained independent of storage methods (Blekhman et al., 2016). Despite this caveat, that fact that considerable variation exists between human populations is notable and highlights the need for much higher resolution sampling of human associated microbial diversity. Similarly, our analysis suggests that there

552 is even more undescribed microbial diversity to be discovered across populations of wild non-
553 human apes.
554
555 In a pangenome analysis to identify individual genes or functions enriched or depleted in
556 genomes of taxa associated with different human populations, we identified numerous
557 functional traits involved in aerobic respiration associated with the European populations in
558 the analysis. We hypothesize that taxa found to be enriched in the fecal samples of humans
559 from Germany and Denmark might have a selective advantage via their clade-independent
560 ability to survive or even utilize aerobic conditions in the intestinal tract. More specifically,
561 we propose that these taxa have undergone convergent adaptation to tolerate high oxygen
562 concentrations. Such aerotolerance could increase microbial fitness, whereby bacteria can
563 withstand high oxygen concentrations to metabolize mucus layers for energy (Zheng et al.,
564 2015). However, the depletion of this mucus by bacteria diminishes an important physical and
565 immunological barrier that protects the human host against microbial assaults and allows for
566 direct interaction between host epithelial cells and microbiota, potentially triggering (auto-
567)immune processes (Costa et al., 2016; Matute et al., 2023). Notably, we showed that the
568 introduction of novel microbes associated with industrialization related to vast differences in
569 the community composition of fecal microbiota in the European populations. Thus, it is
570 possible that susceptibility to intestinal inflammation might be potentiated by specific taxa
571 found in this population. Accordingly, we found increased abundances of well-characterized
572 mucin-degrading taxa, including *Akkermansia* and *Bacteroides*, in the European cohorts.
573 These findings are congruent with previous reports suggesting that there is increased mucus
574 degradation by intestinal microbiota in human populations with direct access to industrial
575 food systems, which may relate to higher incidences of inflammatory bowels diseases
576 observed in developed economies (E. D. Sonnenburg and Sonnenburg, 2019).
577
578 Comparatively, only a few pathways showed conserved enrichment in the opposite direction,
579 suggesting clade specific mechanisms for their loss in some human societies. In particular, we
580 found the taxon *Prevotella* is depleted in German and Danish samples but conserved across
581 hominids, despite representing a diversity of host clades and diets. *Prevotella* is a major
582 determining taxon of one of the human enterotypes, a concept used to define fecal microbial
583 communities (Arumugam et al., 2011). It remains controversial as to whether the human gut

584 microbiome is best classified using such discrete categories, or rather along a dynamic,
585 continuous gradient (Cheng 2019, Knight 2014). Nevertheless, previous reports have shown
586 that individuals who access industrialized food systems (i.e., consume so-called
587 "Westernized" diets) generally display a *Bacteroides*-dominant enterotype. *Bacteroides*-
588 enterotypes have been previously associated with a multitude of intestinal (Vieira-Silva et al.,
589 2019) and extra-intestinal inflammatory diseases (Valles-Colomer et al., 2019). Conversely,
590 individuals who rely on rural and traditional subsistence strategies (i.e., consume plant-rich
591 diets) tend to exhibit a *Prevotella*-dominant enterotype (Vangay et al., 2018). This enterotype
592 is also displayed in about 20% of individuals living within Western societies (Costea et al.,
593 2018). Interestingly, there are conflicting reports concerning *Prevotella* and host health.
594 While it has been shown that *Prevotella* may improve glucose metabolism (Kovatcheva-
595 Datchary et al., 2015), other reports have linked high abundances of *Prevotella* spp. with
596 autoimmune diseases and intestinal inflammation (Iljazovic et al., 2021). While results from
597 model systems have suggested *Prevotella* likely plays a role in autoimmunity, these studies
598 largely relied on mono-colonization of germ-free animals and thus may be biased due to a
599 lack of microbial interaction partners and an aberrant host physiology (Iljazovic et al., 2021).
600 Within human studies, no convincing link between increased *Prevotella* spp. and
601 inflammatory bowel disease has yet been shown (Iljazovic et al., 2021).
602
603 Here, we used an evolutionarily-informed framework to extend the enterotype concept to
604 elucidate the functional dynamics involved in the assembly of the human gut microbiome
605 over evolutionary timescales. Such insights may better inform how changes in the gut
606 microbiome might affect human health. We find the taxon *Prevotella* to be conserved across
607 the sampled hominids. Moreover, the sheer diversity of *Prevotella* displayed across all
608 hominid clades clearly suggests an evolutionary conservation and long-standing interaction
609 of this microbial clade with the host, as further revealed by host-specific microbial functions
610 identified in the metagenomic pangenome analysis. In other words, we find the *Prevotella*
611 clade to be an integral member of the intestinal microbial community of all hominids.
612 Therefore, we propose that the *Prevotella*-enterotype represents an evolutionary ancestral
613 community state for the human gut microbiome. Rather than a discrete enterotype, a
614 reduced abundance and diversity of *Prevotella* may better regarded a key biomarker for
615 disease risk (Gorvitovskaia et al., 2016) or for microbiota insufficiency syndrome (J. L.

Sonnenburg and Sonnenburg, 2019), which seems to be partly driven by changes associated with Western lifestyles. Additional research and large-scale strain collections for *Prevotella* are needed for an in-depth analysis and evaluation of this diverse taxonomic group with regard to host health and its role in inflammation. Such research must consider *Prevotella spp.* as members of a complex consortium of interacting microorganisms and as, we argue, a potential target for pre- and probiotic intervention in chronic inflammatory disorders.

Lastly, we leveraged our catalog of high-quality metagenome-assembled genomes from hominid fecal samples together with existing data to investigate co-phylogenetic patterns across the sampled hosts. Overall, co-phylogeny showed highly clade-specific enrichments and depletions. In addition, human-derived MAGs were found significantly less often among co-phylogenetic groups than MAGs from NHAs. Since we included human-derived data from global reference datasets (Almeida et al., 2021), this effect is unlikely to be an artifact of non-exhaustive coverage of human microbiome members. We found several microbial traits depleted among bacteria with cophylogenetic patterns. Among these were, as expected, reduced genome sizes and gene counts, as well as susceptibility to bile and utilization of multiple simple carbohydrates. We hypothesize that these depletions mirror the specialization of microbes to colonize the hominid gut and utilize host-derived complex carbohydrates.

Our study has limitations. The co-phylogeny analysis relies heavily on genome-sequences recovered from shotgun metagenomic sequencing (MAGs), which are potentially contaminated and incomplete, which could bias tree structure and thus, co-phylogeny estimates, and also the can potentially under- (or over-)estimate the functional capacities of recovered microbial genomes. To address the potential shortcomings of MAGs, we implemented stringent data processing pipelines and quality control to achieve the high-quality MAGs and a host-specific pan-genome based functional annotation framework incorporating information from multiple MAGs per species representative to reduce potential genome gaps (see Methods). Additionally, the commonly used estimates of divergence times of the hominid hosts included in the analysis set the timeframe of the split from a shared ancestor to 8-19 million years ago (Langergraber et al., 2012; Scally et al., 2012). Bacterial speciation events happen in the timeframe of 10-100 million years, or longer (Marin et al.,

2017; McDonald and Currie, 2017; Ochman et al., 1999), and thus, co-phylogeny in hominids is expected to be observed within microbial species or possibly genera. In the presented dataset, species-level sharing of MAGs between host genera was low (1.68%; n = 30 out of 1,787 reconstructed SGBs, not including Manara *et al*. and UHGGv2). SGBs were defined on 95% average nucleotide identity, a measure generally regarded as appropriate (Jain et al., 2018), but it is nevertheless prone to clade specific biases, potentially further influenced by altered speciation dynamics in association with (evolutionary) changes in host lifestyle (Lawrence and Retchless, 2009), *i.e.* previously demonstrated increases in horizontal gene transfer (HGT) within individual microbiomes (Groussin et al., 2021). Accounting for these potential biases, we relaxed the threshold to a shared genus-level annotations for subgroups to be included in the co-phylogeny testing, while keeping the number of tested sub-phylogenies to a minimum through the definition of stringent inclusion criteria (see Methods). Despite these considerations, the observation of signals of co-phylogeny across hominids is supported by a robust statistical framework.

Additional limitations stem from the focus on humans and African great apes. While the comparisons between these host-clades provide a framework for the in-depth investigation of (evolutionary) rather recent adaptations and between-host divergences, they potentially neglect that could be revealed by broader-scale investigations, such as the previously described convergence of the human gut microbiota towards that of cercopithecines (Amato et al., 2019). However, our analyses show the impact of the unique trajectories taken by the intestinal microbiomes of *Pan* and *Homo* since their last common ancestor.

Our work here lays the foundation for the analysis of disease-associated changes in the human intestinal microbiome in an evolutionarily informed framework, thereby allowing researchers to evaluate microbiome-associated inflammatory disorders from a point of view that considers both proximal and evolutionary influences. Future investigations should consider in-depth analysis of horizontal gene transfer events within or even between primate hosts to shed further light on also cross-species dynamics and transition of microbes. Such analyses however require either microbial isolate genomes or at least long-read sequencing data to increase confidence in detection events. Additionally, time series data for host groups sharing the same habitat, *e.g. G. g. gorilla* and *P. t. troglodytes*, could give additional insights

into cross-species sharing dynamics which cannot be appropriately elucidated based on single-timepoint data.

In summary, we present an in-depth taxonomic and functional description and analysis of hominid-associated fecal communities spanning about ten million of years of evolution and host-microbiome interactions in the gut of humans and African great apes. Western lifestyle and maybe more precisely industrialization associated changes in human gut microbiota have been previously suggested as a driver of microbiome insufficiency syndrome, whereby an incompatibility between quickly adapting microbiota and slowly evolving host genes leads to chronic inflammatory diseases such as metabolic syndrome, type 2 diabetes, and inflammatory bowel disease (E. D. Sonnenburg and Sonnenburg, 2019; Wallenborn and Vonaesch, 2022). Thus, a comparative analysis of human and NHA intestinal microbiota that considers evolutionary forces as presented herein provides a powerful platform to advance our understanding of human-associated microbiota and guide the development of personalized, targeted interventions to prevent and treat chronic inflammatory disorders.

Methods

Ethics & Inclusion statement

Ethical approval for work on human samples was obtained from the Local Ethics Committee Germany, Kiel (reference number A156/03), the Ivorian ethics commission (Comité national d'éthique et de la recherche [CNER], permit number 101 10/MSHP/CNER/P) and the Congolese ethics commission (Comité d'Éthique, Ministère de l'Enseignement Supérieur et Universiaire, permit number ESO/CE/018/11). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1975 Helsinki declaration and its later amendments or comparable ethical standards. Sampling of wild-living great apes and human populations in Africa were granted by: Bwindi Impenetrable Forest National Park, Uganda (*Gorilla beringei beringi, Pan troglodytes schweinfurthii*): the Uganda National Council for Science and Technology and the Uganda Wildlife Authority; Kokolopori Bonobo Reserve and Bandundu region, Democratic Republic of the Congo (*Pan paniscus,* Human): the Ministere de Recherche Scientifique et Technologie, Democratic Republic of the Congo; Loango National Park, Gabon (*Gorilla gorilla gorilla, Pan troglodytes troglodytes*): the Agence Nationale des Parcs Nationaux, the Centre National de la Recherche Scientifique et Technique of Gabon; Taï National Park and region, Côte d'Ivoire (*Pan troglodytes verus*, Human): the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, the Ministère des Eaux et Fôrets in Côte d'Ivoire, and the Office Ivoirien des Parcs et Réserves. Researchers from CIV and DRC contributing to the conducted research and fulfilling the authorship criteria were included as co-authors. Research at sites in Africa was conducted in collaboration with local partners as stated in the acknowledgements section, granted by local authorities and in agreement with local policies. Feces from wild-living, habituated animals were collected after defecation without interfering with the animals. Research at great ape sites was increasingly performed following the IUCN guidelines to minimize disease risk for great apes. We did not stratify or correct for sex or gender effects in the analysis. Our analyses focus on the comparison of gut metagenomes from either distinct hominid species or between human subgroups from populations with differences in human development index. We expect that the effects of sex and/or gender are negligible in this context and these factors have not been explored in the current analysis.

Fecal Sampling, DNA extraction and data generation

Sampling of wild-living great apes and human populations in Africa were conducted at: Bwindi Impenetrable Forest National Park, Uganda (*Gorilla beringei beringi, Pan troglodytes schweinfurthii*); Kokolopori Bonobo Reserve and villages adjacent to Salonga-Sud National Park, Democratic Republic of the Congo (*Pan paniscus,* Human); Loango National Park, Gabon (*Gorilla gorilla gorilla, Pan troglodytes troglodytes*); Taï National Park and adjacent villages, Côte d'Ivoire (*Pan troglodytes verus*, Human). Sampling procedures for collecting feces from humans (n=48) and wild non-human primates (n=109) have been previously described (Gogarten et al., 2021). Briefly, fecal samples were collected immediately after defecation, and, depending on the local infrastructure, either stored in RNAlater and frozen at -20°C or stored in a cryotube, cooled in a thermos until return to the field laboratory, and subsequently snap frozen in liquid nitrogen. Appropriate government permits and permission to conduct research on wild primates were granted by the relevant authorities (see Acknowledgments for site-specific details). Human fecal samples from the Democratic Republic of Congo (n = 12) were stored in RNAlater and frozen at -20°C. Human fecal samples from Côte d'Ivoire (n=12) were stored in a cryotube, cooled in a thermos until return to the field laboratory and subsequently snap frozen in liquid nitrogen. Human fecal samples from Germany were collected at home by the participant in standard fecal collection tubes, mailed to the study center, and stored at -80°C. DNA extraction from fecal samples was performed from 200mg of stool transferred to 0.70mm Garnet Bead tubes (Qiagen) with 1.1 mL ASL buffer, followed by bead beating in a SpeedMill PLUS (Analytik Jena AG) for 45 s at 50 Hz. Samples were heated to 95°C for 5 min and centrifuged, retaining 200 µl of the supernatant for DNA extraction with the QIAamp DNA Stool Mini Kit (Qiagen) automated on a QIAcube system (Qiagen) according to the manufacturer's protocol. DNA quality was assessed by Qubit and Genomic DNA ScreenTape (Agilent). Illumina Nextera DNA Library Preparation Kit was used to construct shotgun metagenomic libraries, and subsequently sequenced with either 2 × 125 bp reads on a HiSeq 2500 platform or with 2 × 150 bp reads on a NovaSeq 4000 machine (Illumina).

Data processing, assembly and metagenomic binning

Raw sequencing FastQ files were quality controlled and preprocessed using the BBMap software suite ("BBMap," n.d.). Host reads were removed using bbmap.sh. A masked human

reference database ("Introducing RemoveHuman," 2014) and a lenient mapping threshold of 95% identity was used to account for a broader host range to also capture host contamination from the *Pan* and *Gorilla* host. Metagenomic contigs were assembled with metaSPAdes and contigs >= 2000 bp were retained (Bankevich et al., 2012). Reads were mapped to the contigs of the respective samples using Minimap2 (Li, 2018), converted to BAM files with Samtools (Li et al., 2009) and used to estimate per-contig mapping depth with the jgi_summarize_bam_contig__depths binary from the MetaBAT2 binning tool (Kang et al., 2019). Contig binning for individual samples was performed with MetaBAT2 (Kang et al., 2019), MaxBin2 (Wu et al., 2016) and CONCOCT (Alneberg et al., 2014). In addition, the VAMB binning tool (Nissen et al., 2021) was used on a cross-mapping catalog of the merged contigs from all samples within each host group. Individual binning results were refined using MAGScoT (Rühlemann et al., 2022) to acquire high quality metagenome-assembled genomes (MAGs) for each sample. Clustering of MAGs to species-level genome bins (SGBs) was performed with dRep (Olm et al., 2017) in a multi-step approach to control for inflated SGBs due to low MAG quality. First, MAGs were dereplicated to 97% similarity within each host group, choosing the MAG with the highest score (calculated by MAGScoT based on completeness and contamination) as cluster representative. High and good quality representatives (score > =0.7) from all host groups together with representative sequences from the UHGG v2 were then clustered into 95% SGBs using dRep, again selecting the highest quality MAG as representatives. Medium quality (scores between 0.5 and 0.7) 97% representatives from previous clustering step were then compared to SGB representatives using fastANI (Jain et al., 2018), assigning MAGs with high similarity (>=95%) to the respective SGB. Medium quality 97% representatives without hits to the high quality SGB library were then clustered into 95% SGBs and added to the catalog in the case of at least two genomes in the cluster, discarding singleton clusters. The final catalog of SGB representatives was used to quantify contig abundances in all samples using Salmon in metagenome mode (Patro et al., 2017). Taxonomic annotations were performed using the GTDBtk (v2.1) and GTDB release 207v2 (Chaumeil et al., 2022; Parks et al., 2022). For SGBs without genus- and/or species-level assignments, the SGB ID was used as taxonomic label. GTDBtk marker gene alignments were used to generate a phylogenetic tree of all SGB representatives using the respective "infer" function of the GTDBtk. All data processing scripts are available online: https://github.com/mruehlemann/greatapes_mgx_scripts

792

793 **Pangenome catalog creation, annotation and analysis**

794 All MAGs underwent calling of coding sequences using prodigal (v2.6.3) (Hyatt et al., 2010).
795 Protein sequences were clustered based on 95% similarity using MMseqs (Hauser et al., 2016;
796 Steinegger and Söding, 2017) and annotated using the emapper.py script of the eggNOG-
797 mapper v2 (Cantalapiedra et al., 2021) annotation tool with the eggNOG 5.0 reference
798 database (Huerta-Cepas et al., 2019). MAG level functional profiles based on KEGG Ortholog
799 annotations were collapsed into SGB-level pangenomes for each host genus (*Homo, Gorilla*,
800 and *Pan*). In the case that no MAGs of an SGB were recovered from a given host genus,
801 functional profiles were inferred from MAGs across the other host groups, accounting for
802 host-specificities in the inferred accessory genomes/functions by considering a function to be
803 present if it was present in all host-specific pangenomes of the respective SGB with MAGs
804 recovered from the metagenomic data.

805

806 **Calculation of microbial clade and functional abundances**

807 All downstream data processing and statistical calculations were performed in R v4.2 (R Core
808 Team, 2022) and using the tidyverse library (Wickham et al., 2019). Per-sample contig
809 abundances for the SGB representatives from Salmon were used to estimate SGB
810 abundances. Salmon output includes total mapped reads per contig and mapping reads
811 adjusted for library size and total sequencing depth as transcripts per million (TPM), a
812 measure from the transcriptomics field which can be directly transferred to metagenomic
813 libraries. Individual contig coverages were calculated from the number of mapped reads and
814 the effective lengths of the Salmon mapping output, considering contigs with > 10% coverage
815 as present. An SGB was considered present when at least 20% of its total length was in contigs
816 marked as “present” and if at least 1,000 total reads and 250 TPM mapped to it. Final SGB
817 abundances were calculated as TPM, calculated from the reads mapping to the SGBs present
818 in the respective sample, thus representing a normalized abundance across all samples.
819 Combining SGB abundances with taxonomic assignments, domain- to species-level
820 abundances were calculated as cumulative TPM abundances within the respective taxonomic
821 bins. Rarefactions were calculated based on 5-fold repeated subsampling of contig level
822 mapped reads at 100k, 250k, 500k, 1M, 2.5M, 5M, and 10M reads, followed by TPM
823 calculations as described above. By rarefying reads and not TPM we realistically simulate

824 sampling effects introduced by low coverage and low abundances of SGBs affecting especially
825 samples with small library sizes. Community level functional profiles were calculated be
826 multiplying TPM abundances of SGBs with the respective host-genus specific functional
827 profiles (presence of KEGG orthologs [KOs]) of the SGBs and summarizing the per-SGB values
828 into a sample-level abundance of functional annotations. Ultimately, functional abundances
829 of individual KOs represent the cumulative TPM abundance of SGBs carrying the respective
830 KO.
831
832 **Alpha and beta diversity**
833 Faith's phylogenetic diversity (PD) (Faith, 1992) was used as measure of alpha diversity,
834 calculated from the phylogenetic tree based on GTDBtk marker genes using the pd() function
835 of the picante package for R (Kembel et al., 2010). Genus-level increase of PD from novel SGBs
836 was calculated from the differences of PDs with and without novel SGB annotated as the
837 respective genus. Sample level PDs were calculated from the SGB presence/absence patterns.
838 Beta diversity was assessed as unweighted and weighted UniFrac distances (Lozupone et al.,
839 2011) using the UniFrac() function of the phyloseq package for R (McMurdie and Holmes,
840 2013) and SGB abundances and the phylogenetic tree based on GTDBtk marker genes as
841 input. Aitchison distance (Aitchison, 1982) was calculated from the CLR-transformed genus-
842 level TPM abundances obtained from the clr() function from the compositions package for R
843 (van den Boogaart and Tolosana-Delgado, 2008) and adding a pseudocount of 1 to all
844 abundances, setting all CLR-transformed abundances below zero to zero. Jaccard distances
845 (Jaccard, 1912) were also calculated on genus-level presence/absence patterns using the
846 vegdist() function from the vegan package for R (Oksanen et al., 2022). Genus-level
847 abundances were chosen for Aitchison and Jaccard distance, as SGBs are highly host-specific,
848 thus would lead to high beta-diversities simply due to host exclusive SGBs, grouping at genus-
849 level prevents from this and UniFrac distances use phylogenetic relations between SGBs.
850 Beta-diversity on functional abundances were calculated from the Euclidian distances of the
851 log-transformed KO abundances adding a pseudocount of 1 to avoid undefined values.
852 Presence-absence values of KOs were treated in the same way as described above and using
853 Jaccard distance to infer pairwise distances.
854
855 **Cross-host sharing analysis**

Permutation-based analysis of excessive and reduced sharing of SGBs between host groups were based on the mean SGB abundances of the five rarefaction of 1M mapped reads to account for differences in library depth impacting SGB richness and per-group sample sizes. For each host group, 100-fold sampling of five samples from this group were drawn and the SGBs found in the host were analyzed for their presence in five random samples of each of the other host groups, calculating the relative amount of shared SGBs as $Rel_{shared} = n_{SGB,shared} / n_{SGB,host}$. The mean of all 100 samplings was used as relative sharing coefficient for all host pairs in both directions. Excess and reduced sharing was analyzed by 1000-fold drawing of five random samples accounting for differences of host groups and the repetition of above calculations for relative sharing with all host groups. P-values were calculated from the proportions of random samplings exceeding/falling below the true sharing coefficients.

Phylosymbiosis analysis

Phylosymbiosis was assessed using five measures for community level diversity, unweighted and weighted UniFrac, genus-level Aitchison, and Jaccard distances, as well as KEGG ortholog (KO) abundance based Aitchison distance, and following the approach described in Brooks et al. (2016). Briefly, host group differences were used to infer microbiome dendrograms by UPGMA clustering. Branch support was calculated from 1000-fold jackknife sampling. Robinson-Foulds distances between microbiome trees and host phylogeny were calculated using the RF.dist() function from the phangorn package for R (Schliep, 2011). Significance of phylosymbiosis was assessed using the host phylogeny and 100,000 random trees as comparison for the microbiome trees. Tanglegrams were created with the ggtree and cowplot packages for R (Yu et al., 2017).

Assessment of between-group abundance differences

Taxonomic abundance differences between Humans and NHAs, as well as between humans living outside and within industrialized systems were based on CLR-transformed abundances to account for the compositionality of microbiome data (Gloor et al., 2017). Included in the calculations were all genera with a prevalence > 20% and relative abundance (before CLR-transformation) of > 0.1% in at least of the host groups and all KO abundances with a prevalence > 20% and CLR-transformed abundance of > 1 in at least one of the host groups.

KO abundances were filtered accordingly and subsequently (log+1)-transformed to achieve a less skewed distribution. Log-transformation was chosen, as CLR-transformation assumes compositionality of the data, which is – unlike for taxonomic abundances - not fulfilled for functional abundances. Abundance differences were assessed in a linear regression in R (R Core Team, 2022) using abundances as dependent variable and human/NHA and European/African dichotomies as explanatory variables in a single model for each taxon (or function), assessing associations with all groups at once, the model was defined as lm(abundance ~ Human + European). P-values were calculated from the t-values of the resulting models using the summary.lm() function. Log-fold differences were calculated using group mean abundances and a pseudocount of 0.01. P-values were adjusted for multiple testing using Bonferroni correction. Features with significant ($Q < 0.05$) positive association with NHAs were grouped as "NHA associated". Features associating with geographic differences (Europe/Africa) were grouped into the respective group they were positively associated with. Remaining genera with significant differences between humans and NHAs, but not with a particular subgroup were grouped as "human associated". Genera without abundance differences in any of these comparisons were grouped as "unchanged" or "other".

Functional pangenome differences between groups

Pangenome catalogs of human-associated SGBs were compared within microbial families between SGBs in taxonomic groups found enriched in European communities compared to other human-associated taxa, independent of a strong association with geography. KEGG Ontology (KO) term annotations were used as functional groups and their prevalence differences between groups were assessed using Fisher's exact test. Per-family effect sizes (Z-Scores) of KOs were calculated from P_{Fisher}-values and the direction of the effects which were assessed using the log2 of the ratio of prevalences in the two groups and a pseudo count of 0.01. The sum of the Z-Scores were added and divided by the square-root of the total number of families the respective KOs were found in to obtain a Z_{Meta} for each KO term, used to calculated P_{Meta}. P_{Meta}-values were adjusted for multiple testing using Holm-correction. KO terms with $Q < 0.05$ and present in at least two of the microbial families in the analysis were considered as functions with differential prevalence. A similar approach was employed to assess functional differences between NHP- and Human-associated SGBs, however in this case, SGB pangenome differences were compared on genus level and the meta-analysis was

919 performed combining signals from all genera, and specifically across the genera within
920 particular phyla.
921
922 **Tree reconciliation analysis**
923 Proteins from the representative SGBs of the genus *Prevotella* annotated with the annotation
924 "cydA" (cytochrome *bd* oxidase subunit 1) as "Preferred name" in the emapper/eggNOG
925 annotation were extracted from the unclustered protein sequence catalog. The same
926 procedure was followed for *Paraprevotella clara,* which was included as an outgroup.
927 Incomplete cydA sequences were removed using a length threshold of 200. Protein sequences
928 were aligned using Clustal Omega (Sievers et al., 2011). The alignment was used to
929 reconstruct the phylogenetic tree using IQTREE2 (Minh et al., 2020) and a automatic model
930 selection, which resulted in an LG+F+R8 model to be chosen as best-fit model according to
931 the Bayesian information criterion (BIC). Branch support values were calculated using UFBoot
932 (Minh et al., 2013) and performing SH-aLRT test (Guindon et al., 2010). Alignments of GTDBtk
933 marker protein sequences for *Prevotella* SGBs and *Paraprevotella clara* were used to
934 reconstruct a genome-level species phylogeny in the same respective way as described above
935 for the cydA sequences (BIC best-fit: LG+I+I+R5). Very low confident branches (< 60%
936 bootstrap support) in the cydA phylogeny were resolved together with the species tree using
937 the OptResolutions supplementary program of the RANGER-DTL 2.0 software (Bansal et al.,
938 2018) resulting in 495 equally probable trees with optimized duplication-transfer-loss costs
939 using default values (duplication: 2, loss: 1, transfer: 3). A randomly chosen output tree was
940 using in the reconciliation analysis with the species tree in RANGER-DTL 2.0 using default
941 values and 1,000 random starting seeds in parallel (Tange, 2011) to assess robustness.
942 Resulting sampling outcomes were summarized using the AggregateRanger tool of the
943 RANGER-DTL 2.0 software package.
944
945 **Co-phylogeny analysis**
946 Host phylogenetic trees were obtained from the 10kTrees website ((Arnold et al., 2010);
947 https://10ktrees.nunn-lab.org/). To assure high quality microbial phylogenies for the co-
948 phylogeny analysis, family-level maximum-likelihood trees were reconstructed from the
949 GTDBtk marker gene alignments with the IQTREE2 software (Minh et al., 2020) and a WAG
950 model including a random SGB outside the respective families as outgroups. Family level trees

were rooted and for each SGB traced from tip to root to identify for each SGB the smallest subtree which covered 4, 5, 6, and 7 host groups. Combining information from all SGBs, the overall set of smallest trees to be included in the co-phylogeny analysis were identified, discarding subtrees for which the inclusion criterion was fulfilled already for a smaller tree starting from a different tip. In addition, subtrees spanning more than a single genus were excluded from the analysis, as divergence times of microbial genera predate divergence of the included hosts (Ochman et al., 1999). For all subgroups included in the analysis, maximum-likelihood distances and trees using a WAG model in IQTREE2 were inferred from the marker gene alignment of all MAGs assigned to the SGBs in the respective subgroups. Co-phylogeny of the subgroup was assessed by randomly selecting one MAG per host, calculating congruence with the host tree by Robinson-Foulds metric and by Mantel-test (Hommola et al., 2009). Tiplabels were permutated 999-fold and P-values calculated. This process starting from the random selection of one MAG per host was repeated 999 times to obtain final P-values. Family-level co-phylogeny ratios were calculated based on the ratio of SGBs within subtrees with co-phylogeny signal and total SGBs in the respective family that were included in the analysis. Enrichment of co-phylogeny for each microbial family was calculated by using Fisher's exact test on the SGBs in the analysis dividing them into four groups based on family membership and being in a subtree with co-phylogeny signal. All P-values were adjusted using FDR correction.

Correlation of microbial phenotypes with cophylogeny signals

SGB representative genome sequences were analyzed using the Traitar tool (Weimann et al., 2016) to infer up to 67 microbial traits. A total of 43 inferred traits present in more than 50 and less than 1017-50 = 967 of the 1,017 SGBs included in the cophylogeny analysis were analyzed for their association with cophylogeny signals. Using the R package lme4qtl (Ziyatdinov et al., 2018), for each of the traits a mixed logistic regression model was fitted across for the 1,017 SGBs, using signal of cophylogeny (binary trait) as dependent variable and the trait as bineray fixed effect explanatory variable, accounting for phylogenetic and taxonomic relatedness between SGBs by including a relatedness matrix and phylum-level categories as random effects in the model and using a binomal function with probit as link. The relationship matrix was calculated by using the cophenetic distance matrix from the SGB phylogeny, scaled to values between 0 and 1 by dividing by its largeste distance. Accordingly,

genome size in megabases and gene counts derived from the number of genes in the prodigal output were included as fixed effect continuous traits. Effect sizes and P-values of the individual models were taken from the summary() function in R. P-values were adjusted for multiple testing by Bonferroni correction. Exemplary phylum-level differences in D-xylose utilization and gene counts between SGBs with and without cophylogeny signals were calculated using non-parametric Fisher's exact and Wilcoxon rank sum test, respectively.

Data availability

All metagenomic sequencing data is available via the NCBI BioProject accession IDs PRJNA692042, PRJNA539933, and PRJNA491335. The collection of 7,700 metagenome-assembled genomes has been deposited in the European Nucleotide Archive, Accession: XXXXX.

Code availability

All code to process sequencing files to generate the presented results and manuscript figures is available via https://github.com/mruehlemann/greatapes_mgx_scripts.

References

Aitchison, J., 1982. The Statistical Analysis of Compositional Data. Journal of the Royal Statistical Society: Series B (Methodological) 44, 139–160. https://doi.org/10.1111/j.2517-6161.1982.tb01195.x

Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z.J., Pollard, K.S., Sakharova, E., Parks, D.H., Hugenholtz, P., Segata, N., Kyrpides, N.C., Finn, R.D., 2021. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat Biotechnol 39, 105–114. https://doi.org/10.1038/s41587-020-0603-3

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., Quince, C., 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11, 1144–1146. https://doi.org/10.1038/nmeth.3103

Amato, K.R., Mallott, E.K., McDonald, D., Dominy, N.J., Goldberg, T., Lambert, J.E., Swedell, L., Metcalf, J.L., Gomez, A., Britton, G.A.O., Stumpf, R.M., Leigh, S.R., Knight, R., 2019. Convergence of human and Old World monkey gut microbiomes demonstrates the importance of human ecology over phylogeny. Genome Biology 20, 201. https://doi.org/10.1186/s13059-019-1807-z

Arnold, C., Matthews, L.J., Nunn, C.L., 2010. The 10kTrees website: A new online resource for primate phylogeny. Evolutionary Anthropology: Issues, News, and Reviews 19, 114–118. https://doi.org/10.1002/evan.20251

1020 Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes,
1021 G.R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez,
1022 L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K.,
1023 Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain,
1024 J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G.,
1025 Wang, J., Guarner, F., Pedersen, O., de Vos, W.M., Brunak, S., Doré, J., MetaHIT
1026 Consortium, Antolín, M., Artiguenave, F., Blottiere, H.M., Almeida, M., Brechot, C.,
1027 Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner,
1028 K.U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-
1029 Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K.,
1030 Maguin, E., Mérieux, A., Melo Minardi, R., M'rini, C., Muller, J., Oozeer, R., Parkhill, J.,
1031 Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K.,
1032 Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J.,
1033 Ehrlich, S.D., Bork, P., 2011. Enterotypes of the human gut microbiome. Nature 473,
1034 174–180. https://doi.org/10.1038/nature09944
1035 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M.,
1036 Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N.,
1037 Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A New Genome Assembly
1038 Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19, 455–477.
1039 https://doi.org/10.1089/cmb.2012.0021
1040 Bansal, M.S., Kellis, M., Kordi, M., Kundu, S., 2018. RANGER-DTL 2.0: rigorous reconstruction
1041 of gene-family evolution by duplication, transfer and loss. Bioinformatics 34, 3214–
1042 3216. https://doi.org/10.1093/bioinformatics/bty314
1043 BBMap [WWW Document], n.d. . SourceForge. URL
1044 https://sourceforge.net/projects/bbmap/ (accessed 12.1.22).
1045 Blekhman, R., Tang, K., Archie, E.A., Barreiro, L.B., Johnson, Z.P., Wilson, M.E., Kohn, J.,
1046 Yuan, M.L., Gesquiere, L., Grieneisen, L.E., Tung, J., 2016. Common methods for fecal
1047 sample storage in field studies yield consistent signatures of individual identity in
1048 microbiome sequencing data. Sci Rep 6, 31519. https://doi.org/10.1038/srep31519
1049 Brooks, A.W., Kohl, K.D., Brucker, R.M., Opstal, E.J. van, Bordenstein, S.R., 2016.
1050 Phylosymbiosis: Relationships and Functional Effects of Microbial Communities
1051 across Host Evolutionary History. PLOS Biology 14, e2000225.
1052 https://doi.org/10.1371/journal.pbio.2000225
1053 Campbell, T.P., Sun, X., Patel, V.H., Sanz, C., Morgan, D., Dantas, G., 2020. The microbiome
1054 and resistome of chimpanzees, gorillas, and humans across host lifestyle and
1055 geography. ISME J 14, 1584–1599. https://doi.org/10.1038/s41396-020-0634-2
1056 Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021.
1057 eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain
1058 Prediction at the Metagenomic Scale. Molecular Biology and Evolution 38, 5825–
1059 5829. https://doi.org/10.1093/molbev/msab293
1060 Casiano-Colón, A., Marquis, R.E., 1988. Role of the arginine deiminase system in protecting
1061 oral bacteria and an enzymatic basis for acid tolerance. Appl Environ Microbiol 54,
1062 1318–1324.
1063 Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., Parks, D.H., 2022. GTDB-Tk v2: memory
1064 friendly classification with the genome taxonomy database. Bioinformatics 38, 5315–
1065 5316. https://doi.org/10.1093/bioinformatics/btac672

Costa, F.R.C., Françozo, M.C.S., de Oliveira, G.G., Ignacio, A., Castoldi, A., Zamboni, D.S., Ramos, S.G., Câmara, N.O., de Zoete, M.R., Palm, N.W., Flavell, R.A., Silva, J.S., Carlos, D., 2016. Gut microbiota translocation to the pancreatic lymph nodes triggers NOD2 activation and contributes to T1D onset. J Exp Med 213, 1223–1239. https://doi.org/10.1084/jem.20150744

Costea, P.I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M.J., Bushman, F.D., de Vos, W.M., Ehrlich, S.D., Fraser, C.M., Hattori, M., Huttenhower, C., Jeffery, I.B., Knights, D., Lewis, J.D., Ley, R.E., Ochman, H., O'Toole, P.W., Quince, C., Relman, D.A., Shanahan, F., Sunagawa, S., Wang, J., Weinstock, G.M., Wu, G.D., Zeller, G., Zhao, L., Raes, J., Knight, R., Bork, P., 2018. Enterotypes in the landscape of gut microbial community composition. Nat Microbiol 3, 8–16. https://doi.org/10.1038/s41564-017-0072-8

Davidson, A.L., Chen, J., 2004. ATP-binding cassette transporters in bacteria. Annu Rev Biochem 73, 241–268. https://doi.org/10.1146/annurev.biochem.73.011303.073626

Deleu, S., Machiels, K., Raes, J., Verbeke, K., Vermeire, S., 2021. Short chain fatty acids and its producing organisms: An overlooked therapy for IBD? EBioMedicine 66, 103293. https://doi.org/10.1016/j.ebiom.2021.103293

Ejby, M., Fredslund, F., Andersen, J.M., Vujičić Žagar, A., Henriksen, J.R., Andersen, T.L., Svensson, B., Slotboom, D.J., Abou Hachem, M., 2016. An ATP Binding Cassette Transporter Mediates the Uptake of α-(1,6)-Linked Dietary Oligosaccharides in Bifidobacterium and Correlates with Competitive Growth on These Substrates. J Biol Chem 291, 20220–20231. https://doi.org/10.1074/jbc.M116.746529

Faith, D.P., 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation 61, 1–10. https://doi.org/10.1016/0006-3207(92)91201-3

Ghazisaeedi, F., Meens, J., Hansche, B., Maurischat, S., Schwerk, P., Goethe, R., Wieler, L.H., Fulde, M., Tedin, K., 2022. A virulence factor as a therapeutic: the probiotic Enterococcus faecium SF68 arginine deiminase inhibits innate immune signaling pathways. Gut Microbes 14, 2106105. https://doi.org/10.1080/19490976.2022.2106105

Giuffrè, A., Borisov, V.B., Arese, M., Sarti, P., Forte, E., 2014. Cytochrome bd oxidase and bacterial tolerance to oxidative and nitrosative stress. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 18th European Bioenergetics Conference 2014 Lisbon, Portugal 1837, 1178–1187. https://doi.org/10.1016/j.bbabio.2014.01.016

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. Frontiers in Microbiology 8.

Gogarten, J.F., Rühlemann, M., Archie, E., Tung, J., Akoua-Koffi, C., Bang, C., Deschner, T., Muyembe-Tamfun, J.-J., Robbins, M.M., Schubert, G., Surbeck, M., Wittig, R.M., Zuberbühler, K., Baines, J.F., Franke, A., Leendertz, F.H., Calvignac-Spencer, S., 2021. Primate phageomes are structured by superhost phylogeny and environment. Proceedings of the National Academy of Sciences 118, e2013535118. https://doi.org/10.1073/pnas.2013535118

Gorvitovskaia, A., Holmes, S.P., Huse, S.M., 2016. Interpreting Prevotella and Bacteroides as biomarkers of diet and lifestyle. Microbiome 4, 15. https://doi.org/10.1186/s40168-016-0160-7

Groussin, M., Mazel, F., Alm, E.J., 2020. Co-evolution and Co-speciation of Host-Gut Bacteria Systems. Cell Host & Microbe 28, 12–22. https://doi.org/10.1016/j.chom.2020.06.013

1113 Groussin, M., Mazel, F., Sanders, J.G., Smillie, C.S., Lavergne, S., Thuiller, W., Alm, E.J., 2017.
1114 Unraveling the processes shaping mammalian gut microbiomes over evolutionary
1115 time. Nat Commun 8, 14319. https://doi.org/10.1038/ncomms14319
1116 Groussin, M., Poyet, M., Sistiaga, A., Kearney, S.M., Moniz, K., Noel, M., Hooker, J., Gibbons,
1117 S.M., Segurel, L., Froment, A., Mohamed, R.S., Fezeu, A., Juimo, V.A., Lafosse, S.,
1118 Tabe, F.E., Girard, C., Iqaluk, D., Nguyen, L.T.T., Shapiro, B.J., Lehtimäki, J.,
1119 Ruokolainen, L., Kettunen, P.P., Vatanen, T., Sigwazi, S., Mabulla, A., Domínguez-
1120 Rodrigo, M., Nartey, Y.A., Agyei-Nkansah, A., Duah, A., Awuku, Y.A., Valles, K.A.,
1121 Asibey, S.O., Afihene, M.Y., Roberts, L.R., Plymoth, A., Onyekwere, C.A., Summons,
1122 R.E., Xavier, R.J., Alm, E.J., 2021. Elevated rates of horizontal gene transfer in the
1123 industrialized human microbiome. Cell 184, 2053-2067.e18.
1124 https://doi.org/10.1016/j.cell.2021.02.052
1125 Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New
1126 Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing
1127 the Performance of PhyML 3.0. Systematic Biology 59, 307–321.
1128 https://doi.org/10.1093/sysbio/syq010
1129 Hansen, L.B.S., Roager, H.M., Søndertoft, N.B., Gøbel, R.J., Kristensen, M., Vallès-Colomer,
1130 M., Vieira-Silva, S., Ibrügger, S., Lind, M.V., Mærkedahl, R.B., Bahl, M.I., Madsen,
1131 M.L., Havelund, J., Falony, G., Tetens, I., Nielsen, T., Allin, K.H., Frandsen, H.L.,
1132 Hartmann, B., Holst, J.J., Sparholt, M.H., Holck, J., Blennow, A., Moll, J.M., Meyer,
1133 A.S., Hoppe, C., Poulsen, J.H., Carvalho, V., Sagnelli, D., Dalgaard, M.D., Christensen,
1134 A.F., Lydolph, M.C., Ross, A.B., Villas-Bôas, S., Brix, S., Sicheritz-Pontén, T., Buschard,
1135 K., Linneberg, A., Rumessen, J.J., Ekstrøm, C.T., Ritz, C., Kristiansen, K., Nielsen, H.B.,
1136 Vestergaard, H., Færgeman, N.J., Raes, J., Frøkiær, H., Hansen, T., Lauritzen, L.,
1137 Gupta, R., Licht, T.R., Pedersen, O., 2018. A low-gluten diet induces changes in the
1138 intestinal microbiome of healthy Danish adults. Nat Commun 9, 4630.
1139 https://doi.org/10.1038/s41467-018-07019-x
1140 Hauser, M., Steinegger, M., Söding, J., 2016. MMseqs software suite for fast and deep
1141 clustering and searching of large protein sequence sets. Bioinformatics 32, 1323–
1142 1330. https://doi.org/10.1093/bioinformatics/btw006
1143 Hildebrand, F., Gossmann, T.I., Frioux, C., Özkurt, E., Myers, P.N., Ferretti, P., Kuhn, M.,
1144 Bahram, M., Nielsen, H.B., Bork, P., 2021. Dispersal strategies shape persistence and
1145 evolution of human gut bacteria. Cell Host & Microbe 29, 1167-1176.e9.
1146 https://doi.org/10.1016/j.chom.2021.05.008
1147 Hommola, K., Smith, J.E., Qiu, Y., Gilks, W.R., 2009. A permutation test of host-parasite
1148 cospeciation. Mol Biol Evol 26, 1457–1468. https://doi.org/10.1093/molbev/msp062
1149 Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H.,
1150 Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., Bork, P., 2019.
1151 eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology
1152 resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47, D309–
1153 D314. https://doi.org/10.1093/nar/gky1085
1154 Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal:
1155 prokaryotic gene recognition and translation initiation site identification. BMC
1156 Bioinformatics 11, 119. https://doi.org/10.1186/1471-2105-11-119
1157 Iljazovic, A., Amend, L., Galvez, E.J.C., de Oliveira, R., Strowig, T., 2021. Modulation of
1158 inflammatory responses by gastrointestinal Prevotella spp. – From associations to

functional studies. International Journal of Medical Microbiology 311, 151472. https://doi.org/10.1016/j.ijmm.2021.151472

Introducing RemoveHuman: Human Contaminant Removal [WWW Document], 2014. . SEQanswers. URL https://www.seqanswers.com/forum/bioinformatics/bioinformatics-aa/37175-introducing-removehuman-human-contaminant-removal (accessed 8.11.23).

Jaccard, P., 1912. The Distribution of the Flora in the Alpine Zone.1. New Phytologist 11, 37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., Aluru, S., 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9, 5114. https://doi.org/10.1038/s41467-018-07641-9

Jha, A.R., Davenport, E.R., Gautam, Y., Bhandari, D., Tandukar, S., Ng, K.M., Fragiadakis, G.K., Holmes, S., Gautam, G.P., Leach, J., Sherchand, J.B., Bustamante, C.D., Sonnenburg, J.L., 2018. Gut microbiome transition across a lifestyle gradient in Himalaya. PLoS Biol 16, e2005396. https://doi.org/10.1371/journal.pbio.2005396

Johnson, A.J., Vangay, P., Al-Ghalith, G.A., Hillmann, B.M., Ward, T.L., Shields-Cutler, R.R., Kim, A.D., Shmagel, A.K., Syed, A.N., Walter, J., Menon, R., Koecher, K., Knights, D., 2019. Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. Cell Host & Microbe 25, 789-802.e5. https://doi.org/10.1016/j.chom.2019.05.005

Joossens, M., Huys, G., Cnockaert, M., Preter, V.D., Verbeke, K., Rutgeerts, P., Vandamme, P., Vermeire, S., 2011. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. Gut 60, 631–637. https://doi.org/10.1136/gut.2010.223263

Jurtshuk, P., 1996. Bacterial Metabolism, in: Baron, S. (Ed.), Medical Microbiology. University of Texas Medical Branch at Galveston, Galveston (TX).

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28, 27–30. https://doi.org/10.1093/nar/28.1.27

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z., 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7, e7359. https://doi.org/10.7717/peerj.7359

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., Webb, C.O., 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26, 1463–1464. https://doi.org/10.1093/bioinformatics/btq166

Kovatcheva-Datchary, P., Nilsson, A., Akrami, R., Lee, Y.S., De Vadder, F., Arora, T., Hallen, A., Martens, E., Björck, I., Bäckhed, F., 2015. Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella. Cell Metab 22, 971–982. https://doi.org/10.1016/j.cmet.2015.10.001

Kundu, S., Bansal, M.S., 2018. On the impact of uncertain gene tree rooting on duplication-transfer-loss reconciliation. BMC Bioinformatics 19, 290. https://doi.org/10.1186/s12859-018-2269-0

Langergraber, K.E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., Inoue, E., Inoue-Muruyama, M., Mitani, J.C., Muller, M.N., Robbins, M.M., Schubert, G., Stoinski, T.S., Viola, B., Watts, D., Wittig, R.M., Wrangham, R.W., Zuberbühler, K., Pääbo, S., Vigilant, L., 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proc Natl Acad Sci U S A 109, 15716–15721. https://doi.org/10.1073/pnas.1211740109

Lawrence, J.G., Retchless, A.C., 2009. The Interplay of Homologous Recombination and Horizontal Gene Transfer in Bacterial Speciation, in: Gogarten, M.B., Gogarten, J.P., Olendzenski, L.C. (Eds.), Horizontal Gene Transfer: Genomes in Flux, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 29–53. https://doi.org/10.1007/978-1-60327-853-9_3

Lee, J., Yang, J., Zhitnitsky, D., Lewinson, O., Rees, D., 2014. Structural and functional characterization of a heavy metal detoxifying ABC transporter (997.2). The FASEB Journal 28, 997.2. https://doi.org/10.1096/fasebj.28.1_supplement.997.2

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lloyd-Price, J., Arze, C., Ananthakrishnan, A.N., Schirmer, M., Avila-Pacheco, J., Poon, T.W., Andrews, E., Ajami, N.J., Bonham, K.S., Brislawn, C.J., Casero, D., Courtney, H., Gonzalez, A., Graeber, T.G., Hall, A.B., Lake, K., Landers, C.J., Mallick, H., Plichta, D.R., Prasad, M., Rahnavard, G., Sauk, J., Shungin, D., Vázquez-Baeza, Y., White, R.A., Braun, J., Denson, L.A., Jansson, J.K., Knight, R., Kugathasan, S., McGovern, D.P.B., Petrosino, J.F., Stappenbeck, T.S., Winter, H.S., Clish, C.B., Franzosa, E.A., Vlamakis, H., Xavier, R.J., Huttenhower, C., 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569, 655–662. https://doi.org/10.1038/s41586-019-1237-9

Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., Knight, R., 2011. UniFrac: an effective distance metric for microbial community comparison. ISME J 5, 169–172. https://doi.org/10.1038/ismej.2010.133

Manara, S., Asnicar, F., Beghini, F., Bazzani, D., Cumbo, F., Zolfo, M., Nigro, E., Karcher, N., Manghi, P., Metzger, M.I., Pasolli, E., Segata, N., 2019. Microbial genomes from non-human primate gut metagenomes expand the primate-associated bacterial tree of life with over 1000 novel species. Genome Biology 20, 299. https://doi.org/10.1186/s13059-019-1923-9

Marin, J., Battistuzzi, F.U., Brown, A.C., Hedges, S.B., 2017. The Timetree of Prokaryotes: New Insights into Their Evolution and Speciation. Molecular Biology and Evolution 34, 437–446. https://doi.org/10.1093/molbev/msw245

Matute, J.D., Duan, J., Flak, M.B., Griebel, P., Tascon-Arcila, J.A., Doms, S., Hanley, T., Antanaviciute, A., Gundrum, J., Mark Welch, J.L., Sit, B., Abtahi, S., Fuhler, G.M., Grootjans, J., Tran, F., Stengel, S.T., White, J.R., Krupka, N., Haller, D., Clare, S., Lawley, T.D., Kaser, A., Simmons, A., Glickman, J.N., Bry, L., Rosenstiel, P., Borisy, G., Waldor, M.K., Baines, J.F., Turner, J.R., Blumberg, R.S., 2023. Intelectin-1 binds and alters the localization of the mucus barrier-modifying bacterium Akkermansia muciniphila. J Exp Med 220, e20211938. https://doi.org/10.1084/jem.20211938

McCall, L.-I., Callewaert, C., Zhu, Q., Song, S.J., Bouslimani, A., Minich, J.J., Ernst, M., Ruiz-Calderon, J.F., Cavallin, H., Pereira, H.S., Novoselac, A., Hernandez, J., Rios, R., Branch, O.H., Blaser, M.J., Paulino, L.C., Dorrestein, P.C., Knight, R., Dominguez-Bello, M.G., 2020. Home chemical and microbial transitions across urbanization. Nature microbiology 5, 108. https://doi.org/10.1038/s41564-019-0593-4

McDonald, B.R., Currie, C.R., 2017. Lateral Gene Transfer Dynamics in the Ancient Bacterial Genus Streptomyces. mBio 8, e00644-17. https://doi.org/10.1128/mBio.00644-17

McMurdie, P.J., Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8, e61217. https://doi.org/10.1371/journal.pone.0061217

Minh, B.Q., Nguyen, M.A.T., von Haeseler, A., 2013. Ultrafast Approximation for Phylogenetic Bootstrap. Molecular Biology and Evolution 30, 1188–1195. https://doi.org/10.1093/molbev/mst024

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015

Moeller, A.H., Caro-Quintero, A., Mjungu, D., Georgiev, A.V., Lonsdorf, E.V., Muller, M.N., Pusey, A.E., Peeters, M., Hahn, B.H., Ochman, H., 2016. Cospeciation of gut microbiota with hominids. Science 353, 380–382. https://doi.org/10.1126/science.aaf3951

Moeller, A.H., Li, Y., Mpoudi Ngole, E., Ahuka-Mundeke, S., Lonsdorf, E.V., Pusey, A.E., Peeters, M., Hahn, B.H., Ochman, H., 2014. Rapid changes in the gut microbiome during human evolution. Proceedings of the National Academy of Sciences 111, 16431–16435. https://doi.org/10.1073/pnas.1419136111

Molinaro, A., Bel Lassen, P., Henricsson, M., Wu, H., Adriouch, S., Belda, E., Chakaroun, R., Nielsen, T., Bergh, P.-O., Rouault, C., André, S., Marquet, F., Andreelli, F., Salem, J.-E., Assmann, K., Bastard, J.-P., Forslund, S., Le Chatelier, E., Falony, G., Pons, N., Prifti, E., Quinquis, B., Roume, H., Vieira-Silva, S., Hansen, T.H., Pedersen, H.K., Lewinter, C., Sønderskov, N.B., Køber, L., Vestergaard, H., Hansen, T., Zucker, J.-D., Galan, P., Dumas, M.-E., Raes, J., Oppert, J.-M., Letunic, I., Nielsen, J., Bork, P., Ehrlich, S.D., Stumvoll, M., Pedersen, O., Aron-Wisnewsky, J., Clément, K., Bäckhed, F., 2020. Imidazole propionate is increased in diabetes and associated with dietary patterns and altered microbial ecology. Nat Commun 11, 5881. https://doi.org/10.1038/s41467-020-19589-w

Mossoun, A., Calvignac-Spencer, S., Anoh, A.E., Pauly, M.S., Driscoll, D.A., Michel, A.O., Nazaire, L.G., Pfister, S., Sabwe, P., Thiesen, U., Vogler, B.R., Wiersma, L., Muyembe-Tamfum, J.-J., Karhemere, S., Akoua-Koffi, C., Couacy-Hymann, E., Fruth, B., Wittig, R.M., Leendertz, F.H., Schubert, G., 2017. Bushmeat Hunting and Zoonotic Transmission of Simian T-Lymphotropic Virus 1 in Tropical West and Central Africa. Journal of Virology 91, 10.1128/jvi.02479-16. https://doi.org/10.1128/jvi.02479-16

Nayfach, S., Rodriguez-Mueller, B., Garud, N., Pollard, K.S., 2016. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res 26, 1612–1625. https://doi.org/10.1101/gr.201863.115

Nishida, A.H., Ochman, H., 2021. Captivity and the co-diversification of great ape microbiomes. Nat Commun 12, 5632. https://doi.org/10.1038/s41467-021-25732-y

Nissen, J.N., Johansen, J., Allesøe, R.L., Sønderby, C.K., Armenteros, J.J.A., Grønbech, C.H., Jensen, L.J., Nielsen, H.B., Petersen, T.N., Winther, O., Rasmussen, S., 2021. Improved metagenome binning and assembly using deep variational autoencoders. Nat Biotechnol 39, 555–560. https://doi.org/10.1038/s41587-020-00777-4

Nüse, B., Holland, T., Rauh, M., Gerlach, R.G., Mattner, J., 2023. L-arginine metabolism as pivotal interface of mutual host–microbe interactions in the gut. Gut Microbes 15, 2222961. https://doi.org/10.1080/19490976.2023.2222961

Ochman, H., Elwyn, S., Moran, N.A., 1999. Calibrating bacterial evolution. Proceedings of the National Academy of Sciences 96, 12638–12643. https://doi.org/10.1073/pnas.96.22.12638

Ochman, H., Worobey, M., Kuo, C.-H., Ndjango, J.-B.N., Peeters, M., Hahn, B.H., Hugenholtz, P., 2010. Evolutionary Relationships of Wild Hominids Recapitulated by Gut Microbial Communities. PLOS Biology 8, e1000546. https://doi.org/10.1371/journal.pbio.1000546

Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M.D., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D., Ouellette, M.-H., Cunha, E.R., Smith, T., Stier, A., Braak, C.J.F.T., Weedon, J., 2022. vegan: Community Ecology Package.

Olm, M.R., Brown, C.T., Brooks, B., Banfield, J.F., 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 11, 2864–2868. https://doi.org/10.1038/ismej.2017.126

Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., Hugenholtz, P., 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res 50, D785–D794. https://doi.org/10.1093/nar/gkab776

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M.C., Rice, B.L., DuLong, C., Morgan, X.C., Golden, C.D., Quince, C., Huttenhower, C., Segata, N., 2019. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell 176, 649-662.e20. https://doi.org/10.1016/j.cell.2019.01.001

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nat Methods 14, 417–419. https://doi.org/10.1038/nmeth.4197

R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ridley, C.P., Lee, H.Y., Khosla, C., 2008. Evolution of polyketide synthases in bacteria. Proceedings of the National Academy of Sciences 105, 4595–4600. https://doi.org/10.1073/pnas.0710107105

Rodríguez-Daza, M.C., Pulido-Mateos, E.C., Lupien-Meilleur, J., Guyonnet, D., Desjardins, Y., Roy, D., 2021. Polyphenol-Mediated Gut Microbiota Modulation: Toward Prebiotics and Further. Frontiers in Nutrition 8.

Rühlemann, M.C., Wacker, E.M., Ellinghaus, D., Franke, A., 2022. MAGScoT - a fast, lightweight, and accurate bin-refinement tool. Bioinformatics btac694. https://doi.org/10.1093/bioinformatics/btac694

Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T., McCarthy, S., Montgomery, S.H., Schwalie, P.C., Tang, Y.A., Ward, M.C., Xue, Y., Yngvadottir, B., Alkan, C., Andersen,

L.N., Ayub, Q., Ball, E.V., Beal, K., Bradley, B.J., Chen, Y., Clee, C.M., Fitzgerald, S., Graves, T.A., Gu, Y., Heath, P., Heger, A., Karakoc, E., Kolb-Kokocinski, A., Laird, G.K., Lunter, G., Meader, S., Mort, M., Mullikin, J.C., Munch, K., O'Connor, T.D., Phillips, A.D., Prado-Martinez, J., Rogers, A.S., Sajjadian, S., Schmidt, D., Shaw, K., Simpson, J.T., Stenson, P.D., Turner, D.J., Vigilant, L., Vilella, A.J., Whitener, W., Zhu, B., Cooper, D.N., de Jong, P., Dermitzakis, E.T., Eichler, E.E., Flicek, P., Goldman, N., Mundy, N.I., Ning, Z., Odom, D.T., Ponting, C.P., Quail, M.A., Ryder, O.A., Searle, S.M., Warren, W.C., Wilson, R.K., Schierup, M.H., Rogers, J., Tyler-Smith, C., Durbin, R., 2012. Insights into hominid evolution from the gorilla genome sequence. Nature 483, 169–175. https://doi.org/10.1038/nature10842

Schaan, A.P., Sarquis, D., Cavalcante, G.C., Magalhães, L., Sacuena, E.R.P., Costa, J., Fonseca, D., Mello, V.J., Guerreiro, J.F., Ribeiro-dos-Santos, Â., 2021. The structure of Brazilian Amazonian gut microbiomes in the process of urbanisation. npj Biofilms Microbiomes 7, 1–12. https://doi.org/10.1038/s41522-021-00237-0

Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27, 592–593. https://doi.org/10.1093/bioinformatics/btq706

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7, 539. https://doi.org/10.1038/msb.2011.75

Sonnenburg, E.D., Sonnenburg, J.L., 2019. The ancestral and industrialized gut microbiota and implications for human health. Nat Rev Microbiol 17, 383–390. https://doi.org/10.1038/s41579-019-0191-8

Sonnenburg, J.L., Sonnenburg, E.D., 2019. Vulnerability of the industrialized microbiota. Science 366, eaaw9255. https://doi.org/10.1126/science.aaw9255

Steinegger, M., Söding, J., 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 35, 1026–1028. https://doi.org/10.1038/nbt.3988

Suzuki, T.A., Fitzstevens, J.L., Schmidt, V.T., Enav, H., Huus, K.E., Mbong Ngwese, M., Grießhammer, A., Pfleiderer, A., Adegbite, B.R., Zinsou, J.F., Esen, M., Velavan, T.P., Adegnika, A.A., Song, L.H., Spector, T.D., Muehlbauer, A.L., Marchi, N., Kang, H., Maier, L., Blekhman, R., Ségurel, L., Ko, G., Youngblut, N.D., Kremsner, P., Ley, R.E., 2022. Codiversification of gut microbiota with humans. Science 377, 1328–1332. https://doi.org/10.1126/science.abm7759

Tange, O., 2011. GNU Parallel - The Command-Line Power Tool. ;login: The USENIX Magazine 36, 42–47.

Thingholm, L.B., Bang, C., Rühlemann, M.C., Starke, A., Sicks, F., Kaspari, V., Jandowsky, A., Frölich, K., Ismer, G., Bernhard, A., Bombis, C., Struve, B., Rausch, P., Franke, A., 2021. Ecology impacts the decrease of Spirochaetes and Prevotella in the fecal gut microbiota of urban humans. BMC microbiology 21. https://doi.org/10.1186/s12866-021-02337-5

Tian, J., Utter, D.R., Cen, L., Dong, P.-T., Shi, W., Bor, B., Qin, M., McLean, J.S., He, X., 2022. Acquisition of the arginine deiminase system benefits epiparasitic Saccharibacteria and their host bacteria in a mammalian niche environment. Proc Natl Acad Sci U S A 119, e2114909119. https://doi.org/10.1073/pnas.2114909119

Tiwari, S.K., van der Putten, B.C.L., Fuchs, T.M., Vinh, T.N., Bootsma, M., Oldenkamp, R., La Ragione, R., Matamoros, S., Hoa, N.T., Berens, C., Leng, J., Álvarez, J., Ferrandis-Vila,

M., Ritchie, J.M., Fruth, A., Schwarz, S., Domínguez, L., Ugarte-Ruiz, M., Bethe, A., Huber, C., Johanns, V., Stamm, I., Wieler, L.H., Ewers, C., Fivian-Hughes, A., Schmidt, H., Menge, C., Semmler, T., Schultsz, C., 2023. Genome-wide association reveals host-specific genomic traits in Escherichia coli. BMC Biology 21, 76. https://doi.org/10.1186/s12915-023-01562-w

Toft, C., Andersson, S.G.E., 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. Nat Rev Genet 11, 465–475. https://doi.org/10.1038/nrg2798

Tsujikawa, Y., Ishikawa, S., Sakane, I., Yoshida, K., Osawa, R., 2021. Identification of genes encoding a novel ABC transporter in Lactobacillus delbrueckii for inulin polymers uptake. Sci Rep 11, 16007. https://doi.org/10.1038/s41598-021-95356-1

UNDP (United Nations Development Programme), 2022. Human Development Report 2021-22. UNDP (United Nations Development Programme).

Valles-Colomer, M., Blanco-Míguez, A., Manghi, P., Asnicar, F., Dubois, L., Golzato, D., Armanini, F., Cumbo, F., Huang, K.D., Manara, S., Masetti, G., Pinto, F., Piperni, E., Punčochář, M., Ricci, L., Zolfo, M., Farrant, O., Goncalves, A., Selma-Royo, M., Binetti, A.G., Becerra, J.E., Han, B., Lusingu, J., Amuasi, J., Amoroso, L., Visconti, A., Steves, C.M., Falchi, M., Filosi, M., Tett, A., Last, A., Xu, Q., Qin, N., Qin, H., May, J., Eibach, D., Corrias, M.V., Ponzoni, M., Pasolli, E., Spector, T.D., Domenici, E., Collado, M.C., Segata, N., 2023. The person-to-person transmission landscape of the gut and oral microbiomes. Nature 614, 125–135. https://doi.org/10.1038/s41586-022-05620-1

Valles-Colomer, M., Falony, G., Darzi, Y., Tigchelaar, E.F., Wang, J., Tito, R.Y., Schiweck, C., Kurilshikov, A., Joossens, M., Wijmenga, C., Claes, S., Van Oudenhove, L., Zhernakova, A., Vieira-Silva, S., Raes, J., 2019. The neuroactive potential of the human gut microbiota in quality of life and depression. Nat Microbiol 4, 623–632. https://doi.org/10.1038/s41564-018-0337-x

Vals-Delgado, C., Alcala-Diaz, J.F., Molina-Abril, H., Roncero-Ramos, I., Caspers, M.P.M., Schuren, F.H.J., Van den Broek, T.J., Luque, R., Perez-Martinez, P., Katsiki, N., Delgado-Lista, J., Ordovas, J.M., van Ommen, B., Camargo, A., Lopez-Miranda, J., 2021. An altered microbiota pattern precedes Type 2 diabetes mellitus development: From the CORDIOPREV study. J Adv Res 35, 99–108. https://doi.org/10.1016/j.jare.2021.05.001

van den Boogaart, K.G., Tolosana-Delgado, R., 2008. “compositions”: A unified R package to analyze compositional data. Computers & Geosciences 34, 320–338. https://doi.org/10.1016/j.cageo.2006.11.017

Vangay, P., Johnson, A.J., Ward, T.L., Al-Ghalith, G.A., Shields-Cutler, R.R., Hillmann, B.M., Lucas, S.K., Beura, L.K., Thompson, E.A., Till, L.M., Batres, R., Paw, B., Pergament, S.L., Saenyakul, P., Xiong, M., Kim, A.D., Kim, G., Masopust, D., Martens, E.C., Angkurawaranon, C., McGready, R., Kashyap, P.C., Culhane-Pera, K.A., Knights, D., 2018. US Immigration Westernizes the Human Gut Microbiome. Cell 175, 962-972.e10. https://doi.org/10.1016/j.cell.2018.10.029

Vieira-Silva, S., Sabino, J., Valles-Colomer, M., Falony, G., Kathagen, G., Caenepeel, C., Cleynen, I., van der Merwe, S., Vermeire, S., Raes, J., 2019. Quantitative microbiome profiling disentangles inflammation- and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses. Nat Microbiol 4, 1826–1831. https://doi.org/10.1038/s41564-019-0483-9

Wallenborn, J.T., Vonaesch, P., 2022. Intestinal microbiota research from a global perspective. Gastroenterology Report 10, goac010. https://doi.org/10.1093/gastro/goac010

Watanabe, F., Bito, T., 2018. Vitamin B12 sources and microbial interaction. Exp Biol Med (Maywood) 243, 148–158. https://doi.org/10.1177/1535370217746612

Weimann, A., Mooren, K., Frank, J., Pope, P.B., Bremges, A., McHardy, A.C., 2016. From Genomes to Phenotypes: Traitar, the Microbial Trait Analyzer. mSystems 1, 10.1128/msystems.00101-16. https://doi.org/10.1128/msystems.00101-16

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. Journal of Open Source Software 4, 1686. https://doi.org/10.21105/joss.01686

Wu, Y.-W., Simmons, B.A., Singer, S.W., 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32, 605–607. https://doi.org/10.1093/bioinformatics/btv638

Yen, M.-R., Peabody, C.R., Partovi, S.M., Zhai, Y., Tseng, Y.-H., Saier, M.H., 2002. Protein-translocating outer membrane porins of Gram-negative bacteria. Biochimica et Biophysica Acta (BBA) - Biomembranes 1562, 6–31. https://doi.org/10.1016/S0005-2736(02)00359-0

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.-Y., 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8, 28–36. https://doi.org/10.1111/2041-210X.12628

Zheng, L., Kelly, C.J., Colgan, S.P., 2015. Physiologic hypoxia and oxygen homeostasis in the healthy intestine. A Review in the Theme: Cellular Responses to Hypoxia. Am J Physiol Cell Physiol 309, C350–C360. https://doi.org/10.1152/ajpcell.00191.2015

Zheng, P., Li, Z., Zhou, Z., 2018. Gut microbiome in type 1 diabetes: A comprehensive review. Diabetes/Metabolism Research and Reviews 34, e3043. https://doi.org/10.1002/dmrr.3043

Ziyatdinov, A., Vázquez-Santiago, M., Brunel, H., Martinez-Perez, A., Aschard, H., Soria, J.M., 2018. lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. BMC Bioinformatics 19, 68. https://doi.org/10.1186/s12859-018-2057-x

Acknowledgements

We thank the IKMB Microbiome and NGS laboratories for excellent technical support. We thank Sébastien Calvignac-Spencer and Grit Schubert for their valuable contributions to the analysis and the manuscript. This study is an outcome of the Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Center 1182 “Origin and Function of Metaorganisms” (https://www.metaorganism-research.com, Project number: 261376515, Projects A2 and Z2) and received infrastructure support from the DFG Excellence Cluster 2167

"Precision Medicine in Chronic Inflammation" (PMI; Project number: 390884018). JFG was supported by the DFG grant “The ecology and evolution of primate phageomes” (GO 3443/1-1: Project number: 453352748) and the DFG Research Group “Sociality and Health in Primates” (FOR2136; CA 1108/3-1; Project number: 244372499). Research at Kokolopori was supported by the Max Planck Society and Harvard University. Research at Ozouga and Taï National Park was supported by the Max Planck Society. Research at Loango was supported by US Fish and Wildlife Service, Tusk Trust, Berggorilla Regenwald Direkthilfe, and the Max Planck Society. Research at Bwindi was supported by the Max Planck Society. The survey of the Bwindi mountain gorillas was funded by the International Gorilla Conservation Programme coalition members with supplemental funding from the Wildlife Conservation Society. We are extremely grateful to the organizations that facilitated work at these field sites for sample collection: Bwindi - The mountain gorilla survey was conducted by the Uganda Wildlife Authority, l’Institut Congolais pour la Conservation de la Nature, the Rwanda Development Board, the International Gorilla Conservation Programme, the Max Planck Institute for Evolutionary Anthropology, Conservation Through Public Health, the Mountain Gorilla Veterinary Project, the Institute for Tropical Forest Conservation, and The Dian Fossey Gorilla Fund; Kokolopori - Vie Sauvage, the Bonobo Conservation Initiative; Loango - Agence Nationale des Parcs Nationaux, the Centre National de la Recherche Scientifique et Technique of Gabon; Taï National Park - the Centre Suisse de Recherches Scientifiques en Côte d’Ivoire and the staff members of the Taï Chimpanzee Project. We thank Taylor Hermes for support in the writing of the manuscript.

Author contributions

J.F.B., F.H.L., and A.F. designed the research project; C.A.-K., C.B., T.D., J.-J.M.-T., M.M.R., M.S., R.M.W., K.Z., A.F., and F.H.L. built and maintained research infrastructure, performed field research, and provided materials. M.R., C.B., J.F.G, M.G., S.W., M.P., J.F.B., F.H.L., and A.F. analyzed data. M.R., B.M.H., C.B., J.F.G., M.G., M.P., M.U., J.F.B., F.H.L., and A.F. wrote the manuscript. All authors read, edited, and approved the submitted version of the manuscript.

Competing Interests

All authors declare no conflicts of interest.

Figures

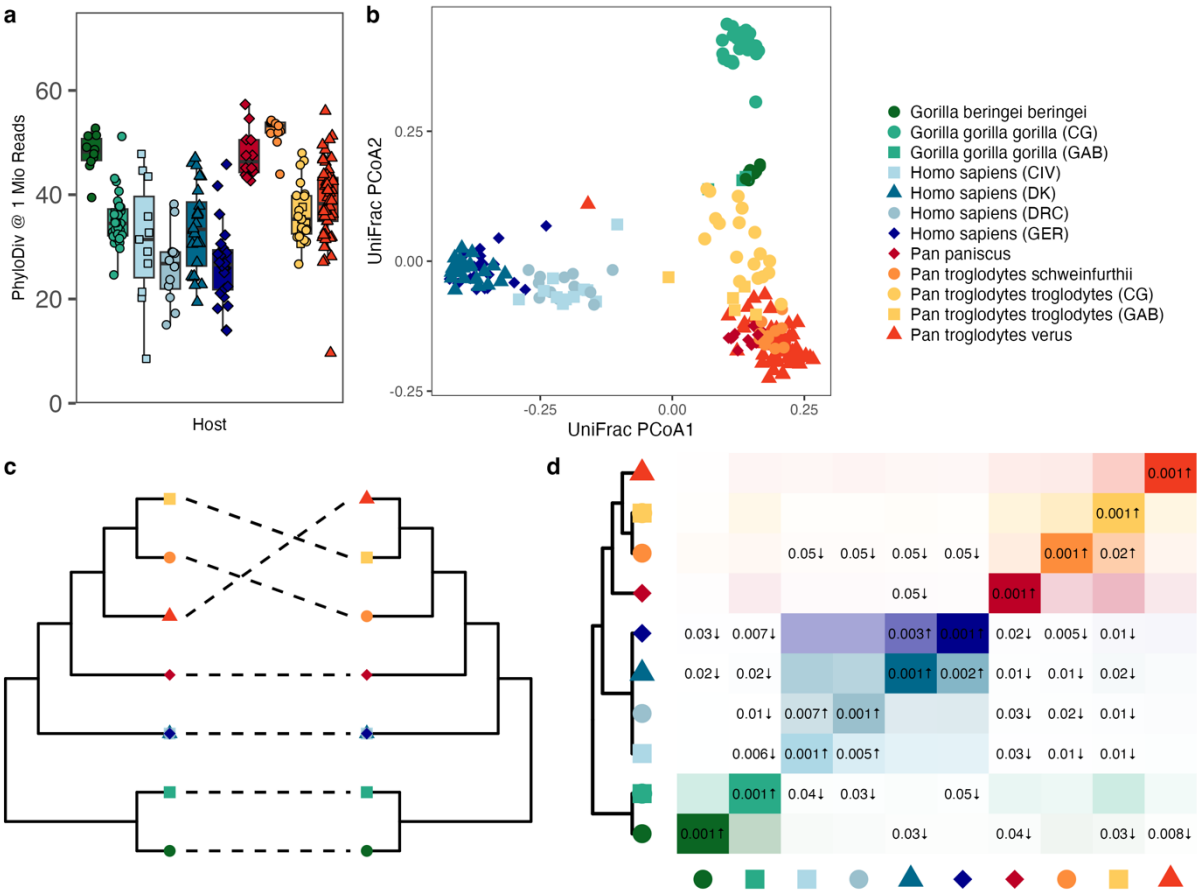

Figure 1: Community-level specificities of human and NHA fecal microbiomes. (a) Phylogenetic Diversity across host groups at a sampling depth of 1 Mio. mapped reads per sample. (b) Ordination of unweighted UniFrac distances of all samples, colored by host subgroups. (c) Tanglegram of host (left) and microbiome (right) trees, the latter based on unweighted UniFrac distances. (d) SGB sharing coefficients between host group. Rows represent reference host groups; columns represent the groups with which they share overlap. Numbers in the tiles are P-values from the analysis for enrichment (#) and depletion ($) in the reference group. *Boxplots in this and other plots show the following elements: center line: median, box limits: upper and lower quartile; whiskers: 1.5 x interquartile ranges.*

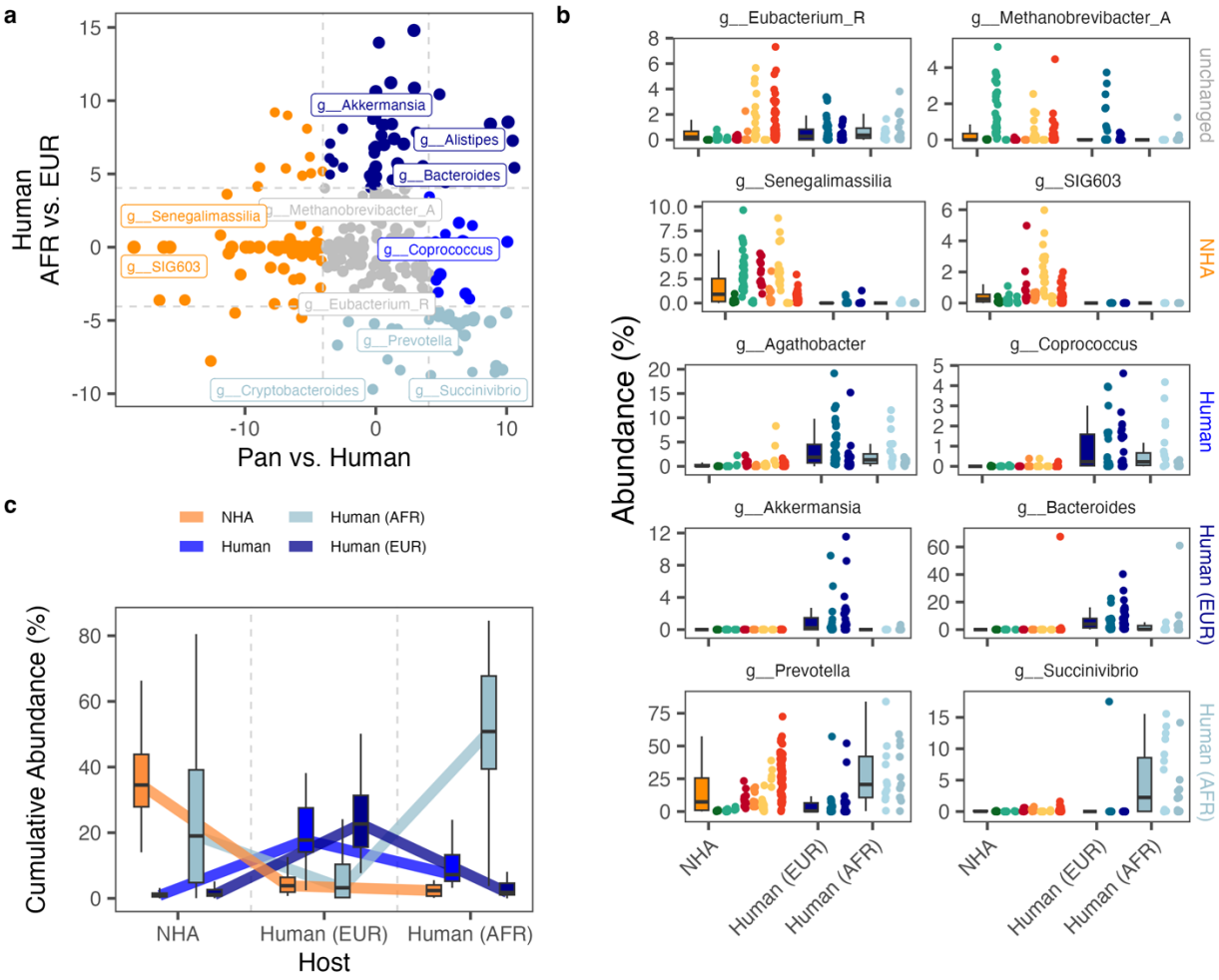

Figure 2: Taxonomic differences the microbiota of humans and NHAs: (a) Effect sizes (t-values from univariate linear regression) of abundance differences of genera in fecal samples of NHAs and humans (x-axis) and humans within African and European populations separated as well (y-axis). Points are colored according to the association groups with European (dark blue) and African human populations (light blue), or to indicate enrichment in NHAs (orange) or humans (bright blue). Taxonomic groups not found to be associated with any of the groups (all Q > 0.05) are shown in grey. Horizontal and vertical lines depict the t-value threshold (|t-value| > 4.04) for statistical significance after Bonferroni-correction. (b) Per-sample and host group abundances of selected genera found with unchanged abundances across all groups (top), increased abundance in NHAs or humans (rows 2 and 3), or in humans from Africa or Europe (rows 4 and 5). Points are colored according to host genus: Gorilla = greens, Pan = reds, oranges, and yellows, and human = blues. (c) Cumulative abundance trajectories of taxa associated with human communities and NHAs. Shown are the per-sample cumulative abundances within each host group, grouped based on a taxon's association with either NHAs, all humans, or one of the human population subgroups.

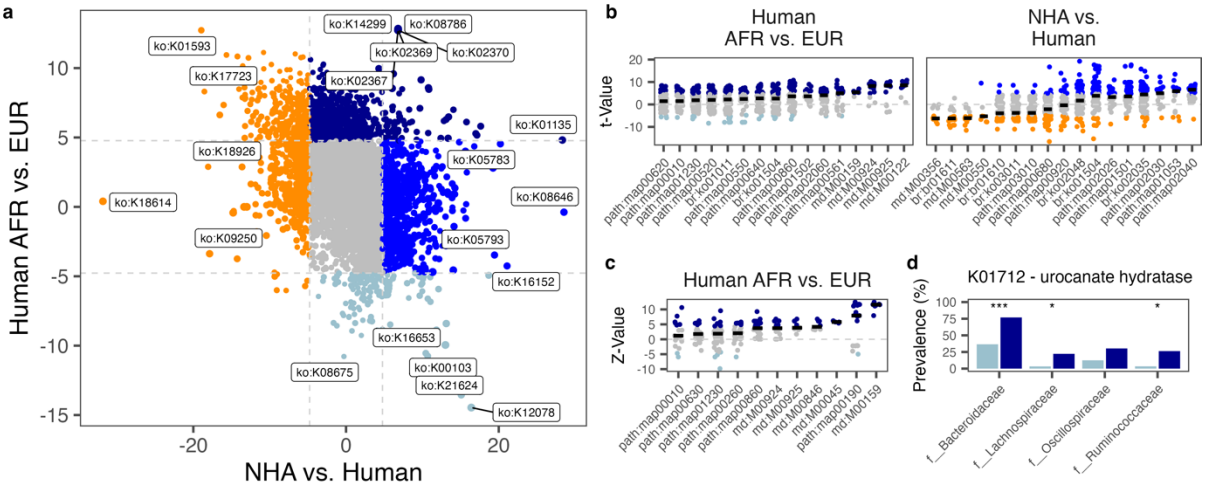

Figure 3: Functional differences in the microbiota of humans and NHAs: (a) Effect sizes (t-values from univariate linear regression) of the abundance differences of KEGG orthologs in fecal samples of NHAs and humans (x-axis) and humans within African and European populations separated as well (y-axis). Points are colored according to the association groups with European (dark blue) and African human populations (light blue), or according to general enrichment in NHAs (orange) or humans (mid-blue). Taxonomic groups not found associated with any of the groups (all Q > 0.05) are shown in grey. Horizontal and vertical lines depict the t-value threshold (|t-value| > 4.77) for statistical significance after Bonferroni-correction. (b) KEGG ortholog (KO) effect sizes for differential abundance between African and European human population associated taxa. Shown KOs are ordered into functional higher-level KEGG categories that were found enriched (QFisher<0.05) among KOs with significantly different abundances between groups. Horizontal bars indicate median t-values of all KOs in a KEGG category as an estimate for the direction of the enrichment. (c) KOs found enriched in the pangenomes of humans from Africa or Europe across four bacterial families shared across continents. Shown are the Z-values of the fixed-effects meta-analysis of KOs for enrichment across microbial families. KOs are sorted into functional higher-level KEGG categories that were found enriched (QFisher<0.05) among KOs with significantly different prevalence (QMeta<0.05) between African and European pangenomes. Horizontal bars depict median ZMeta-values of KOs within higher-level KEGG categories. (d) Prevalence of the urocanate hydratase gene (K01712) in clades found higher abundant in humans from Europe across four microbial families. Stars indicate per-clade differences in gene prevalence: *p < 0.05, **p< 0.01, ***, p<0.001.

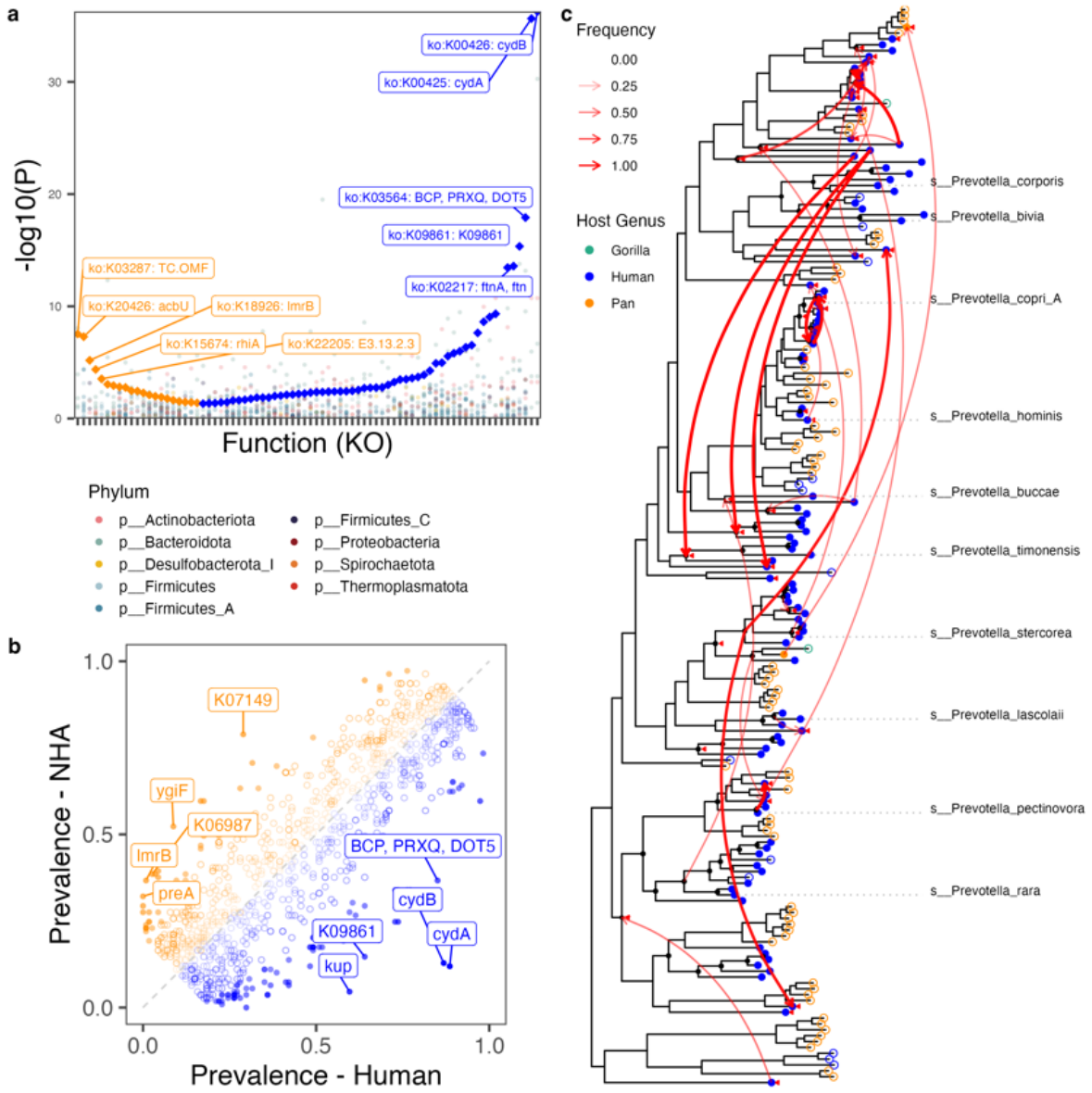

Figure 4: Cross-microbial clade functional associations with NHA and human hosts. (a) KO terms consistently enriched in human- (blue diamonds) and NHA-associated (orange diamonds) taxonomic clades. Individual genus-level P_{Fisher}-values are shown in points colored by phylum. Shown P-values are unadjusted for taxon-level tests and Bonferroni-corrected for multiple testing for the meta-analysis (b) Prevalence patterns of KO terms in human- (blue) and NPH-associated (orange) Prevotella SGBs. Filled shapes represent KOs with significant (Q<0.05) differences in the statistical test. (c) Results of the tree reconciliation analysis for the cydA gene in Prevotella SGBs found in humans (blue) and NHAs (Pan: orange; Gorilla: green) demonstrate a history of frequent transfer events (red arrows) across 1000 reconciliations with random seeds. Filled and empty shapes represent cydA-positive and -negative SGBs, respectively. Arrows are weighted by frequency. Red triangles mark nodes that were identified as gene transfer recipients with > 50% frequency independent of the donor node. Black circles mark speciation events with > 50% frequency. The ten highest abundant Prevotella species with established names are shown for orientation. The Prevotella tree was rooted using Paraprevotella clara as the outgroup (not shown).

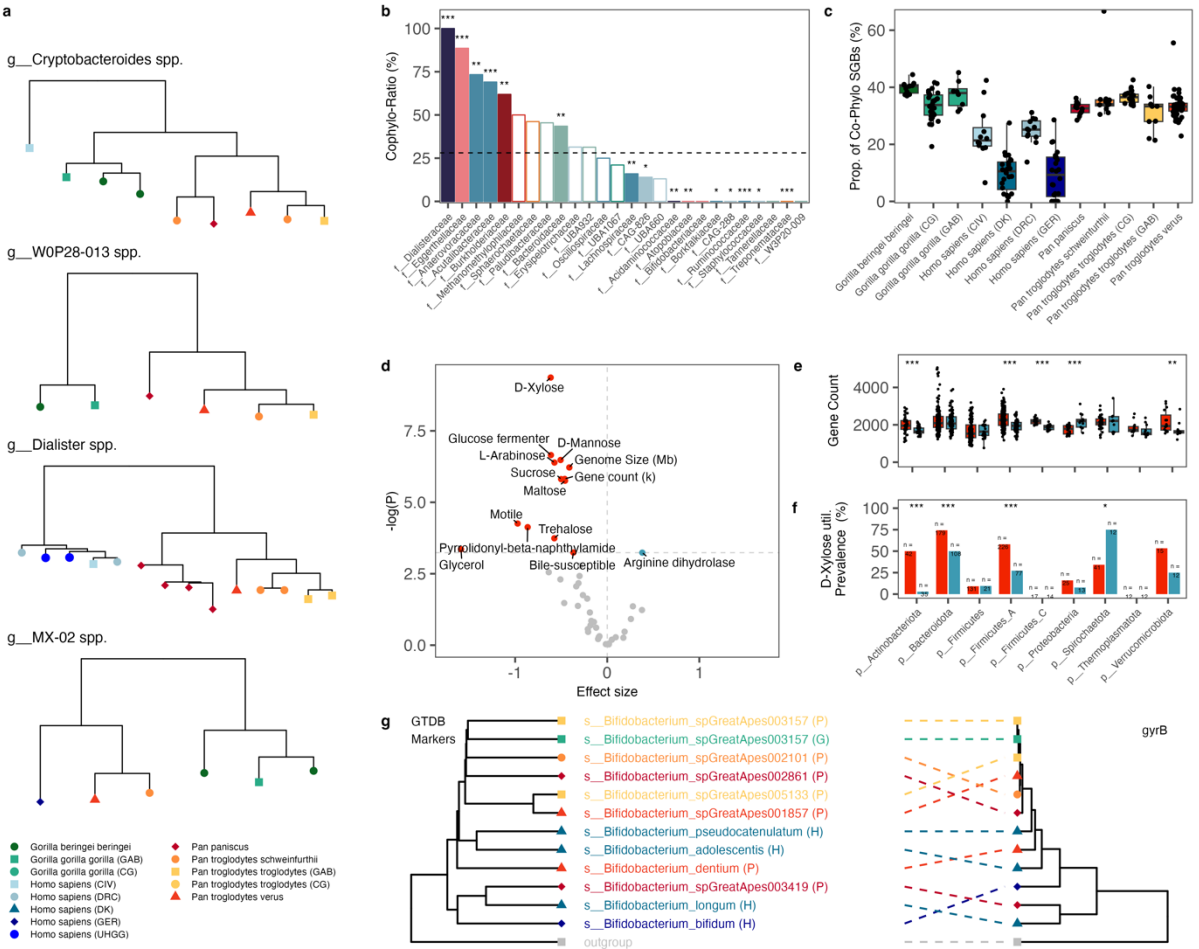

Figure 5: Cophylogeny across humans and non-human African great apes. (a) Subtree phylogenies of groups with significant results in the Mantel-test based analysis for co-phylogeny. Tip colors and shapes correspond to the host subgroups. Trees were rooted on a randomly selected outgroup from a related family (not shown). (b) Enrichment and depletion of co-phylogeny patterns across microbial families with at least 10 SGBs in the analysis. Bars are colored by phylum, corresponding to the colors in Figure 1. Filled bars denote significant (Q<0.05) enrichment and depletion. The dashed line represents the average co-phylogeny ratio across all SGBs. (c) Per-sample proportion of SGBs with co-phylogeny patterns across and colored by host subgroups. (d) Enrichment (blue) and depletion (red) of 43 in-silico inferred microbial traits, genome size and gene count in association with cophylogeny signals. Effect sizes and P-values from mixed effects logistic regression accounting for phylogenetic relatedness of SGBs. The horizontal line marks the threshold of significant Bonferroni-adjusted P-values. (e) SGB-level gene counts across nine phyla, grouped by the presence (blue) and absence (red) of a cophylogeny signal. Within-phylum differences were assessed by two-sided Wilcoxon rank-sum test. (f) Prevalence of inferred D-Xylose utilization by SGBs across phyla, grouped by the presence (blue) and absence (red) of a cophylogeny signal. Within-phylum differences in prevalence were assessed using a two-sided Fisher-test. (g) Tanglegram of Bifidobacterium maximum-likelihood phylogenies based on 120 GTDB marker genes (left) and gyrB sequence (right). Tip colors and shapes correspond to the host subgroups. Across all panels, stars indicate level of significance: * P < 0.05, ** P < 0.01, *** P < 0.001.

Tables

Table1: KEGG functional groups with significant enrichment (Q < 0.05) in the gut microbiome of humans living in Europe.

KEGG ID	Name	# of KOs	Mean Z-Score	P-value	Q-value
md:M00159	V/A-type ATPase, prokaryotes	9	11.19	3.23E-16	2.64E-13
path:map00190	Oxidative phosphorylation	223	5.83	1.39E-08	1.13E-05
md:M00924	Cobalamin biosynthesis, anaerobic, uroporphyrinogen III => sirohydrochlorin => cobyrinate a,c-diamide	22	3.59	2.38E-08	1.94E-05
path:map00860	Porphyrin metabolism	139	3.51	5.08E-08	4.15E-05
md:M00045	Histidine degradation, histidine => N-formiminoglutamate => glutamate	8	5.79	4.37E-07	3.57E-04
path:map00010	Glycolysis / Gluconeogenesis	106	1.81	5.21E-06	4.25E-03
md:M00846	Siroheme biosynthesis, glutamyl-tRNA => siroheme	16	4.64	5.99E-06	4.90E-03
md:M00925	Cobalamin biosynthesis, aerobic, uroporphyrinogen III => precorrin 2 => cobyrinate a,c-diamide	17	4.36	1.07E-05	8.78E-03
path:map00630	Glyoxylate and dicarboxylate metabolism	101	1.94	3.60E-05	2.94E-02
path:map00260	Glycine, serine and threonine metabolism	109	1.74	5.87E-05	4.80E-02
path:map01230	Biosynthesis of amino acids	238	1.41	6.05E-05	4.94E-02