Title

Comparative metagenomics reveals host-specific functional adaptation of intestinal microbiota across hominids

Authors

Rühlemann MC[1,2], Bang C[1], Gogarten JF[3,4,5,6], Hermes BM[7,8], Groussin M[1], Waschina S[9], Poyet M[8], Ulrich M[4,5], Akoua-Koffi C[10], Deschner T[11], Muyembe-Tamfum JJ[12], Robbins MM[13], Surbeck M[14,15], Wittig RM[16,17], Zuberbühler K[18], Baines JF[7,8], Leendertz FH[4,5], Franke A[1]

Affiliations

1 Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany
2 Institute for Medical Microbiology and Hospital Epidemiology, Hannover Medical School Hannover, Germany
3 Applied Zoology and Nature Conservation, University of Greifswald, Greifswald, Germany
4 Helmholtz Institute for One Health, Greifswald, Germany
5 Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute, Berlin, Germany
6 Viral Evolution, Robert Koch Institute, Berlin, Germany
7 Evolutionary Genomics, Max Planck Institute for Evolutionary Biology, Plön, Germany
8 Institute of Experimental Medicine, Kiel University, Kiel, Germany
9 Nutriinformatics Research Group, Institute for Human Nutrition and Food Science, Kiel University, Kiel, Germany
10 Training and Research Unit for in Medical Sciences, Alassane Ouattara University / University Teaching Hospital of Bouaké, Bouaké, Côte d'Ivoire
11 Comparative BioCognition, Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany
12 National Institute for Biomedical Research, National Laboratory of Public Health, Kinshasa, Democratic Republic of the Congo
13 Department of Primate Behavior and Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
14 Department of Human Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States of America
15 Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
16 Institute of Cognitive Sciences, CNRS UMR5229 University Lyon 1, Bron Cedex, France
17 Taï Chimpanzee Project, CSRS, Abidjan, Côte d'Ivoire
18 Institute of Biology, University of Neuchatel, Neuchatel, Switzerland

Summary

Characterizing trajectories of the composition and function of hominid gut microbiota across diverse environments and host species can help reveal specific properties of the human microbiota, with possible implications for host evolution and health. Using shotgun metagenomic sequencing, we investigated taxonomic and functional diversity in the gut microbiota of wild-living great apes, including two gorilla subspecies (*Gorilla gorilla gorilla*, *Gorilla beringei beringei*), three chimpanzee subspecies (*Pan troglodytes verus*, *P.t. troglodytes*, *P.t. schweinfurthii*), and bonobos (*Pan paniscus*), together with human samples from Africa and Europe. We identified microbial taxonomic and functional adaptations convergent with host phylogeny at both the community and microbial genomic levels. We could show that repeated horizontal gene transfer and gene loss are processes involved in these adaptations. We hypothesize, that these adaptation processes and changes in the microbiome predispose the host to chronic inflammatory disorders, such as type 2 diabetes via altered histidine metabolism and inflammatory bowel disease indicated by adaptation of microbes to aerobic conditions. Additionally, we find multiple lines of evidence suggesting a widespread loss of microbial diversity and evolutionary conserved clades in the human microbiota, especially in the European population. Lastly, we observed patterns consistent with codivergence of hosts and microbes, particularly for the bacterial family *Dialisteraceae*, though we find that overall, co-phylogeny patterns are frequently disrupted in humans.

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, has 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 have been conducted on sample collections from high-income countries and only recently, has interest shifted to include humans from diverse global populations, 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 recent efforts provided an opportunity to discover microbiome signatures of host geography and lifestyle that go beyond differences in

microbiota diversity structures. For instance, they revealed elevated rates of horizontal gene transfer (HGT) that correlate with the Human Development Index (HDI) 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 bacterial 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 nonhuman 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

102 such, robust comparative functional analyses are still needed for a comprehensive
103 understanding of how gut microbiota have evolved with hominids and shaped the current
104 structure and functional capabilities (and deficits) of the human gut microbiome.
105
106 Here, we present a large-scale comparative study of wild NHP and humans from
107 geographically distinct populations spanning two continents to better elucidate host-
108 microbiome interactions in the hominid gut, within an evolutionary context. To functionally
109 characterize hominid gut microbial communities, we performed shotgun metagenomic
110 sequencing from the feces of wild-living great apes, including two gorilla subspecies (*Gorilla*
111 *gorilla gorilla,* Gabon (GAB), n=8; *Gorilla beringei beringei,* Uganda (UGA), n=11), three
112 chimpanzee subspecies (*Pan troglodytes verus,* Côte d'Ivoire (CIV), n=55; *P.t. troglodytes*,
113 GAB, n=11; *P.t. schweinfurthii*, UGA, n=12), and bonobos (*Pan paniscus,* Dem. Rep. of the
114 Congo (DRC), n=12). Further, we included published metagenomic data from samples of *G.*
115 *g. gorilla* (n=28) and *P.t. troglodytes* (n=18) from the Republic of Congo (CG) (Campbell et
116 al., 2020). We additionally sequenced human fecal samples from a rural, remote village from
117 Côte d'Ivoire (n=12) and a rural community from the Democratic Republic of Congo (n=12),
118 jointly representing a spectrum of African populations with lower HDI; samples from a
119 German population were included in our analyses as a representative of populations with
120 high HDI.
121
122 Using this extensive data resource, we created a comprehensive catalog of high-quality
123 prokaryotic genomes assembled from metagenomic data, which we annotated on a
124 taxonomic and functional level. Using this catalog we explored patterns of diversity, host-
125 specificities in abundances of taxonomic groups and functions, and phylogenetic parallels
126 between host and microbiota on a community level, but also for individual microbial clades.
127 These analyses reveal intriguing patterns associated with human gut microbial communities,
128 convergent functional adaptations across lineages and potential mechanisms driving these.
129
130 **An expanded catalog of microbial genomes from the hominid gut**
131 From 203 fecal samples of humans and non-human primates with shotgun metagenomic
132 data, we reconstructed a total of 7,506 metagenome-assembled genomes (MAGs) ensuring
133 maximum completeness and low contamination using multiple binning algorithms and

134 dedicated curation and scoring tools ((Rühlemann et al., 2022); see Methods). The most
135 MAGs (quality score > 50%) were reconstructed for the most sampled subgroup of great
136 ape, *P.t. verus* (n=2,431), while the per sample on average the highest number of
137 reconstructed MAGs was the highest for *P.t. schweinfurthii* (mean=74.8). Library size /
138 number of sequencing reads was highly correlated with total assembly size ($\rho_{Spearman}$ =
139 0.644), which in turn was directly correlated with the number of bins recovered for a sample
140 ($\rho_{Spearman}$ = 0.966).
141
142 Together with two large collection of reference species reconstructed from human fecal
143 metagenomes (UHGGv2; (Almeida et al., 2021)), and non-human primate fecal
144 metagenomes(Manara et al., 2019), MAGs were clustered into 5,648 species-level genome
145 bins (SGBs; 95% ANI) using stringent criteria (Supplementary Table 1). These included 1,047
146 novel SGBs not covered by any of the two large references, which mostly originated from
147 NHP samples (n=954, 91.1%; Suppl. Figure 1b). The highest-quality MAG in each SGB was
148 chosen as representative. Overall quality of SGB representative genomes was high (median
149 quality score = 94.5%; Suppl. Figure 1c). SGB representatives were used as comprehensive
150 reference for the estimation of per-sample abundances (Methods, Suppl. Figure 1d, Suppl.
151 Table 2). In total 2,934 SGBs, encompassing 21 bacterial and two archaeal phyla (Suppl.
152 Figure 1a), were found present in the dataset. This newly curated catalog of SGBs from
153 gorillas, bonobos, and chimpanzees increases the number of microbial species genomes
154 previously reconstructed from feces by more than ten-fold, and increases mapping success
155 of fecal metagenomes from NHPs to reference genomes by two- to three-fold (Suppl. Figure
156 1d; (Manara et al., 2019). Novelty in human fecal samples was comparably low, with 5.6%,
157 2.3 %, and 1.1% of novel SGBs found in samples from CIV, DRC, and GER, respectively (Suppl.
158 Figure 1d). For both NHPs and humans, the highest percentages of novel diversity were
159 observed within the phyla Bacteroidota and Spirochaetota, and, to a lesser degree, within
160 Firmicutes and Firmicutes A for NHPs only (Suppl. Figure 1e). Generally, recovered clusters
161 were highly host specific. While the 7,506 MAGs spanned 1,759 of the final SGBs, only for 48
162 of these SGBs MAGs were reconstructed from samples of more than a single host genus.
163
164 Within sample diversity varied considerably between host (sub-)species and with
165 sequencing depth (Suppl. Figure 1f). The used SGB collection covers large proportions of the

166 diversity found in the human gut (Suppl. Fig. 1d) and Faith's PD incorporates SGB
167 relatedness in the diversity calculation, which enables a better estimate of total diversity of
168 a community then simpler richness estimates from taxonomic group abundances.
169 Phylogenetic diversity (PD) at an even mapping depth of 1 Mio. reads per sample showed
170 significantly lower diversity in Humans compared to all African great ape hosts ($P_{Wilcoxon} = 6.9$
171 $x\ 10^{-14}$; Figure 1a). Comparing individual groups, human from GER and DRC showed lower
172 means than all NHP group (all $P_{Wilcoxon} < 0.05$), while Humans from CIV exhibited high
173 variance and were found to have significant lower diversity than all NHP hosts ($P_{Wilcoxon} <$
174 0.05), except for *G. g. gorilla* and *P.t. troglodytes* ($P_{Wilcoxon} > 0.05$). These two great ape taxa
175 were found to have the lowest library sizes, low numbers of recovered genomes and lowest
176 mapping efficiency. Taken together, this suggests that phylogenetic diversity in these hosts
177 may be biased by lower representation in the reference database and that the reduced
178 diversity could be an artefact of this. Consequently, samples with less than 1 Million
179 mapped reads were removed from further analyses. This resulted in the removal of samples
180 from the analysis for the groups *G.b. beringei* (n=1), *G.g. gorilla* (n_{GAB}=5, n_{CG}=1), *P.t.*
181 *troglodytes* (n_{GAB}=4), *P.t. schweinfurthii* (n=1) and humans from CIV (n=1). Of note, lowest
182 mean values for within-sample phylogenetic diversity were found for the human subgroup
183 from Germany, which however were not significantly different from that of humans from
184 African communities (all $P_{Wilcoxon} > 0.05$), contradicting previous reports (McCall et al., 2020;
185 Vangay et al., 2018). Additionally, the considerable differences in PD observed between CIV
186 and DRC highlight the diversity found between human populations and the need to better
187 characterize human diversity. Therefore, while total diversity of some host groups might not
188 be exhaustively sampled, especially considering lower abundant microbial clades, the
189 presented reference collection of high-quality metagenome reconstructed genomes likely
190 represents the current best resource for an in-depth taxonomic and functional assessment
191 of hominid fecal microbiomes and highlights that some human populations, including those
192 sampled in this analysis, have lost considerable microbial diversity in their guts.
193

Microbial taxonomy, but not function, supports phylosymbiosis in hominids

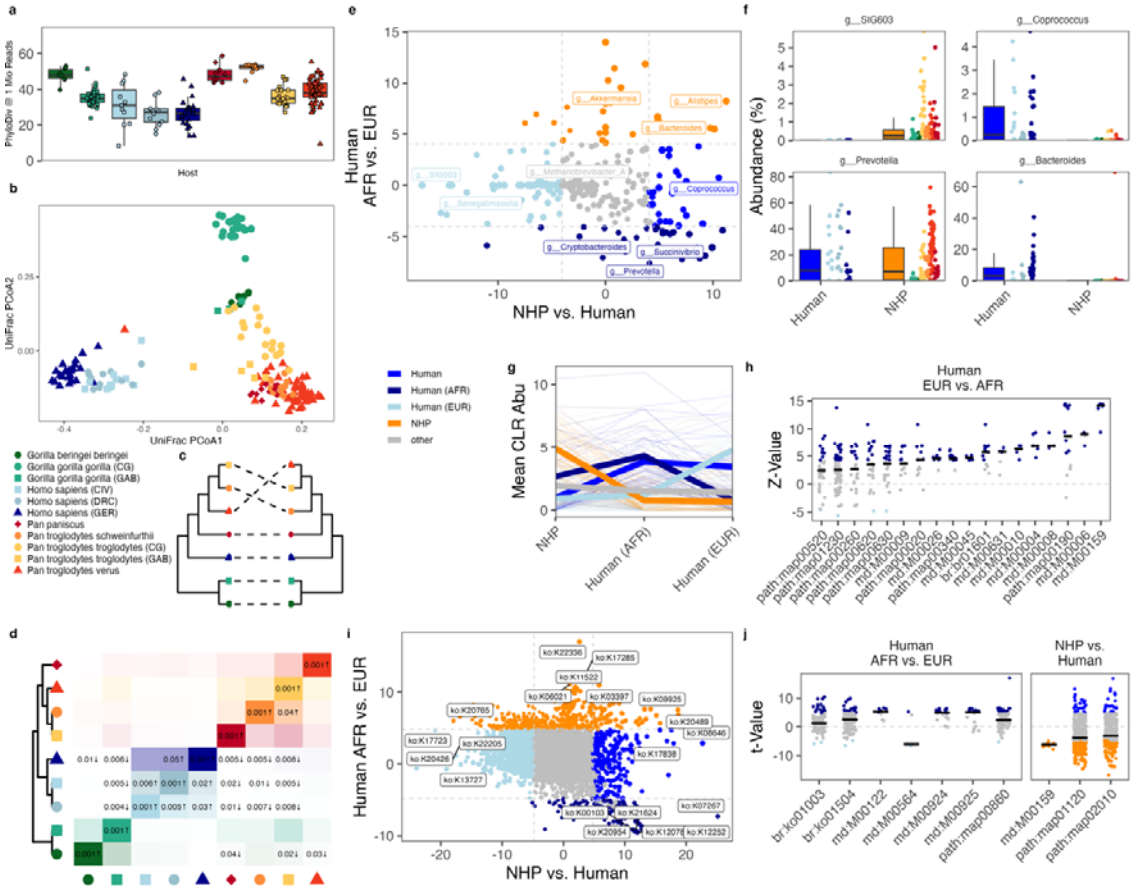

Figure 1: Functional and taxonomic specificities of human and NHP fecal microbiome. (a) Phylogenetic Diversity across host groups at a sampling depth of 1 Mio. mapped reads per sample. (b) Ordination of unweighted UniFrac distances of all samples, colored by host subgroups. (c) Tanglegram of host (left) and microbiome (right) trees, the latter based on unweighted UniFrac distances. (d) SGB sharing coefficients between host group. Rows represent reference host groups, columns the groups they are sharing overlap with. Numbers in the tiles are P-values from the analysis for enrichment (↑) and depletion (↓) in the reference group. *(e) Effect sizes of the abundance differences of genera in fecal samples of NHPs and humans (x-axis) and humans with African and European populations separated as well (y-axis). Points are colored as belonging to the association groups with European (dark blue) and African human populations (light blue), or as being generally enriched in NHPs (orange) or humans (mid-blue). (f) Per-sample and host group abundances of selected genera. (g) Abundance trajectories of taxa associated with human communities and NHPs (h) KEGG ortholog (KO) effect sizes in the test for differential abundance between African and European human population associated taxa. Shown are KOs in functional KEGG categories enriched among with differential abundance (Q<0.1). (i) Effect sizes (t-values) of the community-level abundance differences of KOs in fecal samples of NHPs and humans (x-axis) and humans from African and European (y-axis) (j) Effect sizes of community level KO abundance differences within KEGG modules significantly enriched in the NHP vs. human African and European population comparison. Grey points indicate non-significant values (Q>0.05).*

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 indicating a close relationship 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), two on functional capacities of the community (KEGG

217 ortholog (KO) abundance and presence/absence patterns; Supplementary Figure 2). Strong
218 signals for phylosymbiosis were found for Aitchison, Jaccard and unweighted UniFrac
219 distances (all $P < 0.001$, Figure 1b and c, Suppl. Figure 2, Suppl. Table 3), but not for the
220 function-based distances and the weighted UniFrac distance (all $P > 0.25$). These results
221 suggest phylosymbiotic divergence in the general structure of hominid microbiota (Jaccard
222 and unweighted UniFrac are based on presence/absence of microbial clades), while
223 abundance patterns of distinct microbial clades and functions do not seem to follow this
224 trend, which is in line with previous observations in other host systems (Brooks et al., 2016).
225 This pattern might be explained as the result of fluctuating clade abundances and functions
226 in response to changing environmental factors, including diet, indicating a high functional
227 plasticity of the hominid gut microbiome in response to immediate influences, while
228 structural changes through acquisition and loss of microbial clades are rather the result of
229 longer-term adaptations.
230
231 Assessing shared SGBs between host groups, we found that especially human-associated
232 SGBs were strongly depleted across multiple NHP groups (Figure 1d), while no such signal
233 was found in the opposite direction, suggesting the widespread acquisition of novel
234 microbial clades in the human intestinal microbiome. A similar pattern was found for *G. b.*
235 *beringei* and multiple *Pan* subspecies, however not for *G.g. gorilla*. *G.g. gorilla* and *P.t.*
236 *troglodytes* are sympatric species and were sampled in the same environment in the
237 Republic of Congo and Gabon. The absence of excess strong host-specific signals between
238 these groups might suggest effects of shared environmental factors influencing microbiome
239 structure. SGBs from human samples from lower HDI populations were enriched in all
240 human groups, indicating a shared conserved core microbiota, however no
241 overrepresentation of SGBs from higher HDI populations were found in the two other
242 groups, suggesting that specific SGBs were introduced to the human gut in association with
243 lifestyle changes.
244
245 **Fecal microbiome of high HDI populations marked by loss of evolutionarily conserved core**
246 **microbiota**
247 Abundance difference in microbial clades between humans and NHPs 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 far and more recent past. We analyzed and
251 compared the abundance profiles of gut microbes between NHPs and humans, including
252 individuals living in rural, lower HDI areas of CIV and DRC as well as individuals residing in
253 urban, higher HDI regions within GER. For all following taxonomic and functional
254 comparisons, we restricted the analysis to human and *Pan* (chimpanzees and bonobos)
255 samples to obtain focused insights into the microbiota divergence since their hosts diverged
256 about 7-8 million years ago (Langergraber et al., 2012).
257
258 A total of 287 microbial genera were included in the analysis, of which 148 were found to be
259 differentially abundant ($Q_{Bonferroni}$<0.05, Figure 1d, Suppl. Table 4) in at least one of these
260 comparisons between human subgroups, or between humans and NHPs, and subsequently
261 sorted into one of four groups. We identified 30 taxa with increased abundances among
262 humans from Germany, such as *Akkermansia*, *Bacteroides*, and *Alistipes* (Figure 1d and e).
263 We additionally found 30 taxa that are enriched in the two African populations such as
264 *Cryptobacteroides, Prevotella*, and *Succinivibrio* (Figure 1e and f). Overall, the marker taxa of
265 European microbiomes that we detected are in agreement with previous findings (Jha et al.,
266 2018; Pasolli et al., 2019). Our approach allowed us to identify bacterial taxa that exhibit
267 differential abundance profiles between humans and NHPs that are independent from the
268 human population considered (Figure 1f). We found 63 taxa, such as *SIG603* that have
269 increased abundance in NHPs and 25 taxa, such as *Coprococcus*, with increased abundance
270 in humans. Interestingly, taxa depleted in the microbiome of European individuals are more
271 likely to be also abundant (>0.1%) in the microbiota of NHPs ($P < 0.001$; Figure 1g),
272 suggesting a loss of evolutionary conserved clades in these populations.
273
274 **Widespread changes in fecal microbiome function between hosts and across human**
275 **communities**
276 Taxon-specific changes reflect broad-scale differences between host-groups. A focused
277 analysis of microbial functions can give insights into the specific driving forces of such
278 community-level changes. We performed, analysis of abundance differences of 6,179 KEGG
279 orthologs (KOs, (Kanehisa and Goto, 2000)) in NHP (genus *Pan*) vs. human fecal microbiota
280 and humans in European and African societies and found significant abundance differences

281 in 2,030 (32.9 %) and 607 (9.82 %) KOs, respectively ($Q_{Bonferroni}$<0.05; Figure 1i, Suppl. Table
282 5). For the NHP vs. human comparisons, the KOs with the lowest Q-values included K20426
283 (acbU, 1-epi-valienol-7-phosphate kinase; Q = 6.03 x 10^{-48}) and K17723 (preA;
284 dihydropyrimidine dehydrogenase (NAD+) subunit PraA; Q = 2.25 x 10^{-41}) with higher
285 abundances in NHPs and K07267 (oprB; porin; Q = 2.11 x 10^{-50}) and K08646 (MEP; peptidyl-
286 Lys metalloendopeptidase; Q = 3.08 x 10^{-45}) with higher abundances in humans. K22336
287 (bfrB, bacterioferritin B; Q = 3.43 x 10^{-32}) and K17285 (SELENBP1, methanethiol oxidase; Q =
288 8.21 x 10^{-21}) were strongly increased in Germans. The strongest signals for humans with
289 African lifestyles were found for K12252 (aruH; arginine:pyruvate transaminase; Q = 1.01 x
290 10^{-12}) and K12078 (ptlA; type IV secretion system protein PtlA; Q = 1.24 x 10^{-12}).
291
292 Analysis of higher-level KEGG annotations including e.g. complete pathways
293 overrepresented among differentially abundant KOs revealed three annotations with
294 enrichment in NHPs compared to humans, and seven annotations differentiating humans in
295 Europe and Africa (Q_{FDR} < 0.05; Figure 1j, Suppl. Table 6). The enriched KEGG categories in
296 NHPs were "V/A-type ATPase, prokaryotes" and "ABC tansporters". ABC transporters are a
297 large family of ATP-dependent translocases (n_{KO}=377) with broad substrate spectra,
298 including *e.g.* carbohydrates, metal ions, amino acids, that are involved in nutrient uptake
299 (Davies et al., 2021) and export of toxic compounds (Davidson and Chen, 2004; Lee et al.,
300 2014). The signal is driven by a broad range of ABC transporters found with increase
301 abundance in NHPs (Suppl. Table 6), which might indicate adaptation to a diverse diet (Ejby
302 et al., 2016; Tsujikawa et al., 2021) or metabolism of plant-derived xenobiotics (Rodríguez-
303 Daza et al., 2021). Additionally, we found an enrichment of terms in "Microbial metabolism
304 in diverse environments" (map01120), an extensive pathway that serves as an umbrella
305 category for KEGG metabolic modules; thus, this signal can be interpreted as pointing to
306 substantial changes in the abundance of non-specific metabolic functions between NHPs
307 and humans.
308
309 Signals driven by differences between human subgroups included the KEGG module
310 "Helicobacter pylori pathogenicity signature, cagA pathogenicity island" (md:M00564),
311 driven by increased functional abundances in the fecal communities of humans living in
312 Africa. Individuals infected with CagA-positive *H.* pylori are at increased risk to develop

313 gastric cancer (Hatakeyama, 2014). Curiously, *H.pylori* was not found in the dataset,
314 however the KOs annotated for this pathway are found in the genus *Succinivibrio,* belonging
315 to the same phylum as *Helicobacter* (Proteobacteria), which is depleted in Europeans.
316 Whether *Succinivibrio* in the gut can under certain circumstances exhibit a higher risk for
317 other cancer types needs to be investigated.
318
319 In the European population, Vitamin B_{12} / Cobalamin biosynthesis (md:M00122, M00924,
320 M00925, and path:map00860) and "antimicrobial resistance genes" (br:ko01504) were
321 found to be enriched. Increased antimicrobial functions may be explained by a higher use of
322 antibiotics in human healthcare and animal husbandry, as well as by environmental
323 pollution (Groussin et al., 2021). The increased abundance of Vitamin B_{12}-producing
324 microorganisms in the feces of Europeans may be driven by higher dietary intake of meat
325 and dairy products from ruminants (Watanabe and Bito, 2018). These associations warrant
326 further experimental studies to identify their driving factors and to understand their
327 implications on the gut ecosystem.
328
329 **Success of taxa associated to European populations relates to oxidative carbohydrate**
330 **metabolism**
331 To elucidate the functional basis of the success of taxa in the human gut of individuals
332 residing within in Europe, we analyzed differences in the functional capacities of these in
333 comparison to other human associated taxa either enriched in African societies or without
334 association to this dichotomy. Compared to community-level functional analyses used to
335 identify higher-level functional patterns, pangenome analysis can specifically identify
336 individual genes and functions enriched or depleted in association with a trait of interest
337 within a specific microbial clade. Analyses were restricted to five bacterial families,
338 *Acidaminococcaceae, Bacteroidaceae, Lachnospiraceae*, *Oscillospiraceae*, and
339 *Ruminococcaceae*, for which sufficient numbers of SGBs (n>=20) for pangenome analysis
340 were recovered. SGBs in these family represent large proportions of the overall human
341 microbiome (mean = 61.7%), with no significant differences between the subgroups ($P_{Kruskal-}$
342 $_{Wallis}$ = 0.8). The analysis was conducted at the family-level, as higher taxonomic ranks would
343 increase clade-specific functional biases. To account for between-family functional
344 differences, functional differences of SGBs associated with Europeans were first analyzed

345 within families using Fisher's exact test and subsequently subjected to unweighted meta-
346 analysis using Z-scores to leverage shared signals.
347
348 We identified 276 enriched and 50 depleted KOs in pangenomes associated with the
349 European population ($Q_{Bonferroni}$<0.05; Suppl. Tables 7 and 8). Then, by using higher-level
350 KEGG annotations (modules, pathways, BRITE hierarchies), we found that 18 of these
351 annotations were overrepresented in the dataset ($Q_{Bonferroni}$ < 0.05, Figure 1h, Table 1, Suppl.
352 Table 9).
353 Among these were multiple groups involved in carbohydrate metabolism enriched in taxa
354 enriched in Europeans, specifically pointing at aerobic breakdown of sugar molecules for
355 ATP generation (Citrate cycle: md:M00009, path:map00020; Pentose phosphate pathway:
356 md:M00004; V/A-type ATPase, prokaryote: md:M00159). These signals strongly suggest that
357 the selection for taxa in the gut of humans living in Europe is connected to the adaptation to
358 transient microaerobic conditions in the gut environment, and to the use of oxidative
359 phosphorylation as a mean to release energy from nutrients, which is more efficient than
360 strictly anaerobic fermentation (Jurtshuk, 1996). However, reduced fermentation can
361 impact short-chain fatty acid production, which can, in turn, potentially negatively affect the
362 host's intestinal epithelial cells and metabolism (Deleu et al., 2021). Additionally, we found
363 multiple pathways connected to the metabolism of amino acids (Histidine metabolism:
364 path:map00340, md:M00026, md:M00045; Glycine, serine and threonine metabolism:
365 path:map00260). Especially gut microbial histidine metabolism has been discussed with
366 relevance to human health, as it was shown that an intermediary product of histidine
367 degradation, imidazole propionate, was increased in type 2 diabetic individuals in a large
368 study of almost 2,000 individuals, and that this increase was directly connected to the
369 microbiota and overall unhealthy dietary habits, however independent of dietary histidine
370 intake (Molinaro et al., 2020). This suggests a microbiome-driven pre-disposition to
371 metabolic disorders in European human communities.
372
373 We did not find any higher-level KEGG annotations significantly enriched among the taxa
374 not enriched in Europeans. the relative lower fitness of these taxa may not result from a
375 single mechanism associated to European lifestyles, but rather from multiple selective
376 forces specific to the individual clades.

Convergent host-specific adaptations are found across microbial families

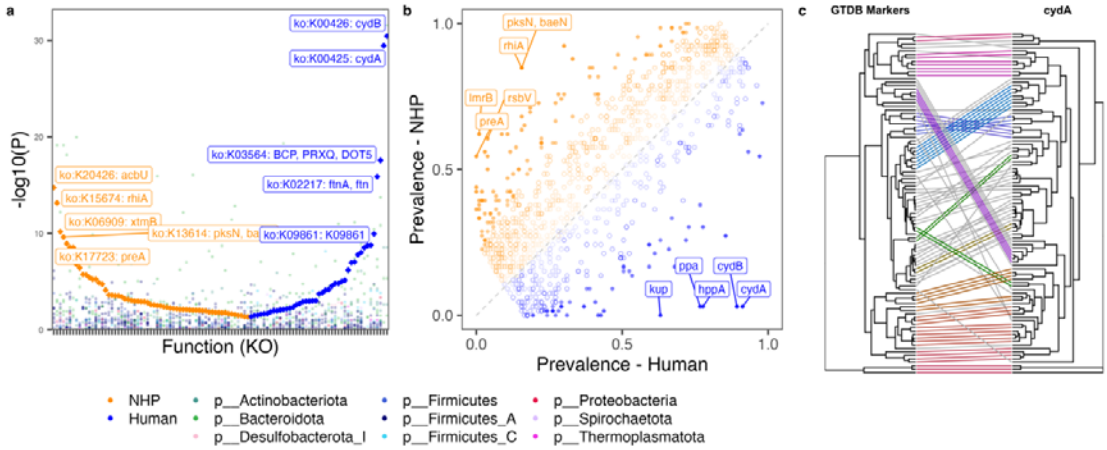

Figure 2: (a) KO terms consistently enriched in human- (blue diamonds) and NHP-associated (orange diamonds). Individual genus-level results are shown in points, colored by phylum. (b) Prevalence patterns of KO terms in human- (blue) and NPH-associated (orange) Prevotella SGBs. Filled shapes represent KOs with significant (Q<0.05) differences in the statistical test. (c) Tanglegram of phylogenetic tree of cydA-positive Prevotella and cydA-gene tree. Colored lines highlight common subtrees. Both trees were rooted using Paraprevotella as outgroup.

Shared gene gains or losses across multiple microbial clades can indicate convergent adaptations to specific host intestinal environments. We performed a pangenome analysis of genera shared between humans and NHPs (n=37) to identify such patterns. To control for higher-level clade effects, functional repertoires (KEGG Ontology terms) were compared at the genus-level between SGBs from NHPs (Suppl. Table 10) and humans and then combined in a meta-analysis across genera. In total, 123 KO terms were identified as carrying signatures of convergent adaptation to the respective host group, with 44 of these signatures associating with humans and 79 associating with NHPs ($Q_{Bonferroni}$<0.05; Figure 2a, Suppl. Table S10). 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 1-epi-valienol-7-phosphate kinase (acbU), phosphoserine phosphatase (rsbU/P), two KOs annotated as polyketide synthases (rhiA, pksN) and a phage terminase (xtmB). Products produced by polyketide synthases have diverse functions, including antibiotic activity, virulence and support of symbiotic relationships (Ridley et al., 2008).

403 *Prevotella* represented the largest genus-level clade in the dataset (n=113, 5.2% of all SGBs).
404 While we found this genus across all host species, it is largely decreased in Europeans. Thus,
405 we selected this genus for further analysis to elucidate potential functional mechanisms
406 driving the observed patterns. Enrichment analysis revealed 55 KOs with distinct prevalence
407 patterns (Q<0.05; Figure 2b). The most striking difference was found for the cytochrome *bd*
408 ubiquinol oxidases subunits 1 and 2 (cydA and cydB), which were ubiquitous in all 31
409 human-associated *Prevotella* SGBs, but present in only two out of 62 SGBs from NHPs. It is
410 important to note that these prevalence differences are not driven by single lineages within
411 the *Prevotella* genus that would have distribution ranges restricted to single hosts. Instead,
412 they are observed across multiple sibling clades spanning the entire phylogenetic tree of the
413 taxon. Cytochrome *bd* oxidases are involved in stress responses, most prominently in
414 transiently microaerobic environments (Giuffrè et al., 2014). Comparison of the *Prevotella*
415 species tree (reconstructed using all SGB representatives recovered from the dataset (n =
416 191; including UHGGv2)) and the cydA gene phylogeny exhibit widespread incongruencies
417 between the tree topologies (Figure 2c). We performed a tree reconciliation using a
418 duplication-transfer-loss (DTL; (Kundu and Bansal, 2018)) model between the *Prevotella* and
419 cydA phylogenies, which revealed a high rate of gene transfer ($\bar{T}$ = 32.227) and loss ($\bar{L}$ =
420 35.769) driving the evolution of the cydA gene family. The gene transfer events and
421 mappings were robust across 1,000 reconciliations with different starting seeds, with 94% of
422 all events and 88% of all mappings found with 100% consistency. These results suggest that
423 the enrichment of cytochrome *bd* ubiquinol oxidases observed in humans compared to
424 NHPs are the result of multiple individual events of gene loss and horizontal gene transfer
425 within the hominid gut.
426
427 **Co-phylogeny is rare in spore-forming microbes and disrupted in humans**
428

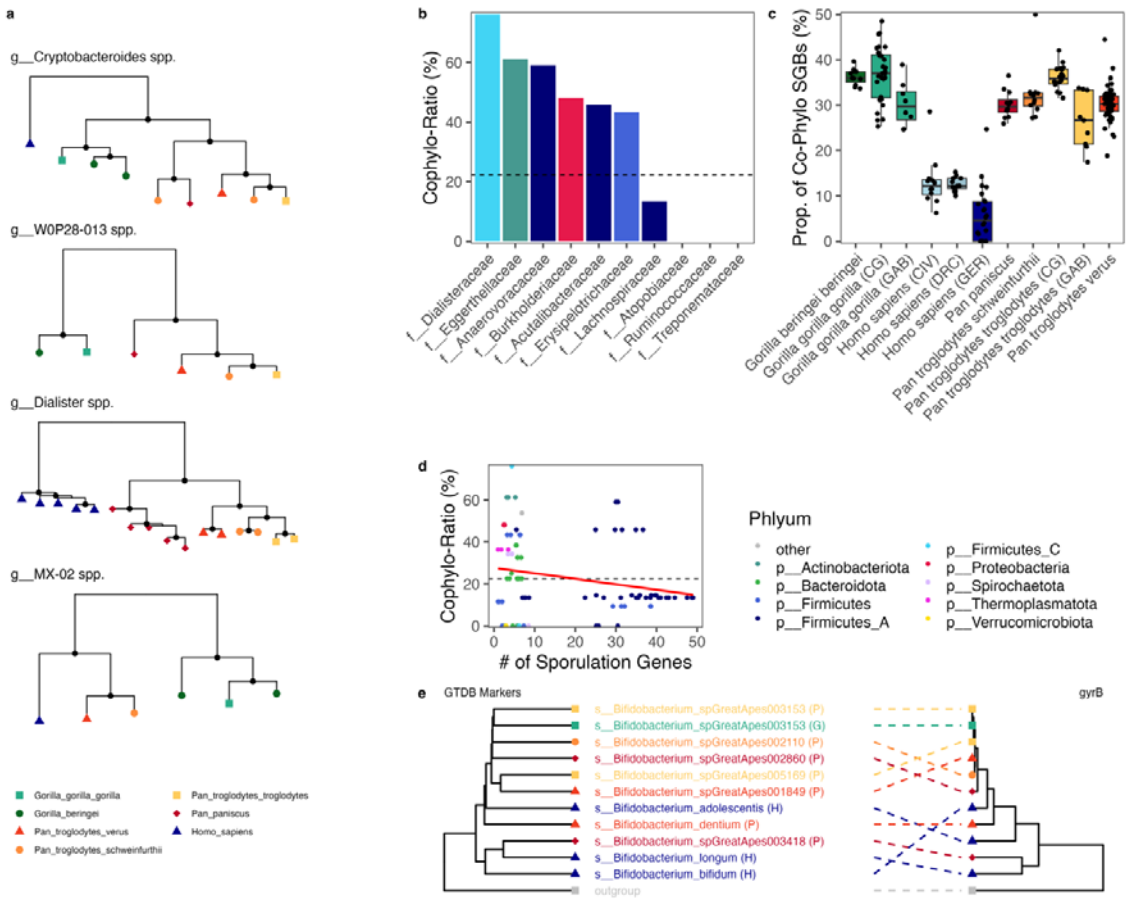

Figure 3: (a) Subtree phylogenies of groups with significant results in the Mantel-test based analysis for co-phylogeny. Tip colors and shapes correspond to the host subgroups.Trees were rooted on a randomly selected outgroup from a related family (not shown). (b) Microbial families with significant (Q < 0.05) enrichment and depletion of co-phylogeny patterns. Bars are colored by phylum, corresponding to the colors in Figure 1; 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) Correlation between co-phylogeny ratio and mean number of sporulation genes. Points represent genera with >= 10 SGBs in the co-phylogeny analysis and are colored by taxonomic phylum. (e) 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.

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 222 subtrees of the SGB phylogeny for co-phylogeny analysis (see Methods). The subtrees spanned 45 families and 961 (32.8%) SGBs present in the dataset in addition to 90 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, 53 of 222 subtrees (23.9%) qualified as exhibiting co-phylogenetic patterns (Suppl. Table S11). These

449 subtrees cover 303 of the 1051 SGBs (28.8%) included in the analysis and 5.7% of the total
450 5,316 hominid SGBs (excluding the SGBs from Manara et al.). All results and subtrees can be
451 inspected online (https://mruehlemann.shinyapps.io/great_apes_shiny_app). By visually
452 inspecting subtrees with co-phylogeny signals, we find many candidates microbial
453 phylogenies that do not follow host phylogeny, *e.g.,* in a subgroup of the genus
454 *Cryptobacteroides* (Figure 3a). Such signals suggest that co-phylogeny within the *Gorilla* and
455 *Pan* clades can result in statistically significant Mantel tests, despite topological
456 incongruences of human-derived sequences, for which no host sister (sub-)species from the
457 same genus is available.
458
459 Human-derived genomes were found in 160 (72.1%) of the tested subtrees, of which 35
460 (21.9%) were co-phylogeny candidates. In comparison, 18 of 62 (29%) tested subtrees
461 without human-derived representatives exhibited co-phylogeny signals, *e.g.*, *W0P29-013*
462 spp. (Figure 3a). Overall, 16.3% (n = 66 out of 404) of human-derived SGBs in the analysis
463 were found in co-phylogenetic subtrees, which is significantly less compared to 34.6% (n =
464 227 out of 656) of NHP-derived SGBs ($P_{Fisher} = 6 \times 10^{-11}$). Co-phylogeny signals, defined based
465 on the ratio of SGBs in trees exhibiting co-phylogeny patterns and the total SGBs in the
466 family, differed strongly. Six families showed excessive signals of co-phylogeny and four
467 families a significant depletion ($Q_{Fisher, FDR} < 0.05$; Figure 3b, Suppl. Table S12). The strongest
468 co-phylogeny ratio was found for the family *Dialisteraceae* (Cophyl-Ratio = 61.22%, Q = 4.44
469 x 10^{-7}), which in the analysis were entirely represented by members of the genus *Dialister*.
470 *Dialister* are a common, but rather neglected members of the gut microbiota which have
471 been found increased (Vals-Delgado et al., 2021; Zheng et al., 2018) and decreased
472 (Joossens et al., 2011) with various human diseases, hence their relation to human health
473 remains unclear. However, on species, *Dialister invisus,* was found to be moderately
474 transmissible between human mother-infant pairs and within households in a large meta-
475 analysis (Valles-Colomer et al., 2023). The strongest depletions were found for *e.g.* the
476 families *Lachnospiraceae* and *Treponemataceae*, the former confirming previous results for
477 this clade (Moeller et al., 2016). The latter, *Treponemataceae*, especially the genus
478 *Treponema D,* were found depleted in humans living in Germany and occur in anaerobic
479 sediments (Thingholm et al., 2021), serving as (intermediary) reservoirs for transmission to

480 humans and NHPs, which can disrupt co-phylogenetic signals by constant re-introduction to
481 the community.
482
483 When comparing host groups, the proportion of SGBs with co-phylogeny signal is
484 significantly reduced in all human subgroups compared to the NHP hosts (all $Q_{Wilcoxon} < 0.05$;
485 Figure 3c). Further, humans from Germany exhibited even lower proportions of co-
486 phylogeny SGBs compared to the two African human populations ($Q_{Wilcoxon} < 0.05$). The
487 human subgroups from Africa did not differ in their co-phylogeny proportions ($P_{Wilcoxon}$ =
488 0.583). These results clearly suggest a loss of wild great ape associated clades with
489 "humanization" of the intestinal microbiota and the introduction of novel microbial partners
490 with changing environment and lifestyle, confirming again findings from previous analyses.
491
492 Previous analyses suggested that spore-forming clades are less likely to be exhibiting co-
493 phylogenic patterns, due to their ability to survive outside of the gut, facilitating dispersal
494 between hosts (Groussin et al., 2020; Moeller et al., 2016). We find a negative correlation
495 between the number of sporulation-associated genes and family-level cophylogeny-ratio (P
496 $< 1.4 \times 10^{-4}$, Figure 3d, Suppl. Table S13), confirming this hypothesis, also supporting
497 previous findings of increased direct host-to-host transmission of non-spore forming
498 microbial clades in baboons (Tung et al., 2015).
499
500 Although not enriched, we found multiple (n=10 out of 28) subgroups in *Bacteroidaceae*
501 with co-phylogeny signals, consistent with previous findings based on *gyrB* amplicon data
502 (Moeller et al., 2016). However, no evidence for strict co-phylogeny was found in
503 *Bifidobacteriaceae*, which is inconsistent with findings from the same previous report
504 (Moeller et al., 2016). Comparing the phylogeny of metagenome-derived *gyrB* sequences
505 and the GTDB marker-gene phylogeny for *Bifidobacterium spp.* revealed clear incongruences
506 between both approaches (Figure 3e), which can likely explain the differences in the
507 presented analysis and previous findings.
508
509 **Discussion**
510 Here, we present the largest curated dataset of fecal metagenomes derived from wild
511 African great apes and human populations. For this, we surveyed and reconstructed high-

512 quality microbial genomes from the feces wild nonhuman primates, including gorillas
513 (*Gorilla gorilla gorilla*; *Gorilla beringei beringei*), chimpanzees (*Pan troglodytes verus*; *P.t.*
514 *troglodytes*; *P.t. schweinfurthii*), and bonobos (*Pan paniscus*) as well as human populations
515 from Africa and Germany. We identified signals of phylosymbiosis across the included
516 hominids, indicating a conserved evolutionary relationship of microbial communities with
517 their host species. Moreover, by employing a comparative approach, we found extensive
518 changes of microbial taxonomic and functional abundances across the intestinal microbiota
519 of NHPs and humans. Previous studies have pointed to Western lifestyles as an important
520 factor influencing the intestinal microbiota in humans. Within our human sample
521 population, we were able to confirm differential signals of prokaryotes associated with the
522 German human population. Importantly, using a comparative dataset of great ape taxa
523 showed that microbial clades lost in the German population in comparison to African human
524 populations are also found in wild great ape populations. Thus, we suggest that the loss of
525 these taxa might be regarded as a mass extinction event of evolutionary conserved
526 members of hominid-associated gut microbiota. While it is tempting to link these changes to
527 industrialization (as previous studies have done), there are many differences between these
528 human populations (e.g., genomic diversity, diet, exercise, sunlight exposure, exposure to
529 antibiotics, population bottlenecks) and it was certainly not possible with the sampling
530 regime here, to determine the particular factors responsible for the variation observed
531 between the human populations sampled here. Despite this caveat, that fact that
532 considerable variation exists between human populations is notable and highlights the need
533 for much higher resolution sampling of human associated microbial diversity.
534
535 In a pangenome analysis to identify individual genes or functions enriched or depleted in
536 genomes of taxa associated with different human populations, we identified numerous
537 functional traits involved in aerobic respiration associated with the German population in
538 the analysis. We hypothesize that taxa found to be enriched in the fecal samples of humans
539 from Germany might have a selective advantage via their clade-independent ability to
540 survive or even utilize aerobic conditions in the intestinal tract. More specifically, we
541 propose that these taxa have undergone convergent adaptation to tolerate high oxygen
542 concentrations. Such aerotolerance could increase microbial fitness, whereby bacteria can
543 withstand high oxygen concentrations to metabolize mucus layers for energy (Zheng et al.,

544 2015). However, the depletion of this mucus by bacteria diminishes an important physical
545 and immunological barrier that protects the human host against microbial assaults and
546 allows for direct interaction between host epithelial cells and microbiota, potentially
547 triggering (auto-)immune processes (Costa et al., 2016; Matute et al., 2023). Notably, we
548 showed that the introduction of novel microbes associated with industrialization related to
549 vast differences in the community assembly of fecal microbiota in the German population.
550 Thus, it is possible that susceptibility to intestinal inflammation might be potentiated by
551 specific taxa found in this population. Accordingly, we found increased abundances of well-
552 characterized mucin-degrading taxa, including *Akkermansia* and *Bacteroides*, in the German
553 cohort. These findings are congruent with previous reports suggesting that there is
554 increased mucus degradation by intestinal microbiota in human populations with direct
555 access to industrial food systems, which may relate to higher incidences of inflammatory
556 bowels diseases observed in developed economies (E. D. Sonnenburg and Sonnenburg,
557 2019).
558
559 Comparatively, only a few pathways showed conserved enrichment in the opposite
560 direction, suggesting clade specific mechanisms for their loss in some human societies. In
561 particular, we found the taxon *Prevotella* is depleted in Germans but conserved across
562 hominids, despite representing a diversity of host clades and diets. *Prevotella* is a major
563 determining taxon of one of the human enterotypes, a concept used to define fecal
564 microbial communities (Arumugam et al., 2011). It remains controversial as to whether the
565 human gut microbiome is best classified using such discrete categories, or rather along a
566 dynamic, continuous gradient (Cheng 2019, Knight 2014). Nevertheless, previous reports
567 have shown that individuals who access industrialized food systems (i.e., consume so-called
568 "Westernized" diets) generally display a *Bacteroides*-dominant enterotype. *Bacteroides*-
569 enterotypes have been previously associated with a multitude of intestinal (Vieira-Silva et
570 al., 2019) and extra-intestinal inflammatory diseases (Valles-Colomer et al., 2019).
571 Conversely, individuals who rely on rural and traditional subsistence strategies (i.e.,
572 consume plant-rich diets) tend to exhibit a *Prevotella*-dominant enterotype (Vangay et al.,
573 2018). This enterotype is also displayed in about 20% of individuals living within Western
574 societies (Costea et al., 2018). Interestingly, there are conflicting reports concerning
575 *Prevotella* and host health. While it has been shown that *Prevotella* may improve glucose

576 metabolism (Kovatcheva-Datchary et al., 2015), other reports have linked high abundances
577 of *Prevotella* spp. with autoimmune diseases and intestinal inflammation (Iljazovic et al.,
578 2021). While results from model systems have suggested *Prevotella* likely plays a role in
579 autoimmunity, these studies largely relied on mono-colonization of germ-free animals and
580 thus may be biased due to a lack of microbial interaction partners and an aberrant host
581 physiology (Iljazovic et al., 2021). Within human studies, no convincing link between
582 increased *Prevotella* spp. and inflammatory bowel disease has yet been shown (Iljazovic et
583 al., 2021).
584
585 Here, we used an evolutionarily-informed framework to extend the enterotype concept to
586 elucidate the functional dynamics involved in the assembly of the human gut microbiome
587 over evolutionary timescales. Such insights may better inform how changes in the gut
588 microbiome might affect human health. We find the taxon *Prevotella* to be conserved across
589 the sampled hominids. Moreover, the sheer diversity of *Prevotella* displayed across all
590 hominid clades clearly suggests an evolutionary conservation and long-standing interaction
591 of this microbial clade with the host, as further revealed by host-specific microbial functions
592 identified in the metagenomic pangenome analysis. In other words, we find the *Prevotella*
593 clade to be an integral member of the intestinal microbial community of all hominids.
594 Therefore, we propose that the *Prevotella*-enterotype represents an evolutionary ancestral
595 community state for the human gut microbiome. Rather than a discrete enterotype, a
596 reduced abundance and diversity of *Prevotella* may better regarded a key biomarker for
597 disease risk (Gorvitovskaia et al., 2016) or for microbiota insufficiency syndrome (J. L.
598 Sonnenburg and Sonnenburg, 2019), which seems to be partly driven by changes associated
599 with Western lifestyles. Additional research and large-scale strain collections for *Prevotella*
600 are needed for an in-depth analysis and evaluation of this diverse taxonomic group with
601 regard to host health and its role in inflammation. Such research must consider *Prevotella*
602 *spp.* as members of a complex consortium of interacting microorganisms and as, we argue, a
603 potential target for pre- and probiotic intervention in chronic inflammatory disorders.
604
605 Lastly, we leveraged our catalog of high-quality metagenome-assembled genomes from
606 hominid fecal samples together with existing data to investigate co-phylogenetic patterns
607 across the sampled hosts. Overall, co-phylogeny showed highly clade-specific enrichments

608 and depletions. In addition, human-derived MAGs were found significantly less often among
609 co-phylogenetic groups than MAGs from NHPs. Since we included human-derived data from
610 global reference datasets (Almeida et al., 2021), this effect is unlikely to be an artifact of
611 non-exhaustive coverage of human microbiome members. Additionally, we provide
612 evidence for the previously proposed effect of sporulation capabilities as a mechanism for
613 prokaryotes to become independent of direct host-to-host transmission, disrupting co-
614 phylogeny patterns, potentially through repeated environmental uptake of spore-forming
615 bacteria (Groussin et al., 2020).
616
617 Our study has limitations. The co-phylogeny analysis relies heavily on genome-sequences
618 recovered from shotgun metagenomic sequencing (MAGs), which are potentially
619 contaminated and incomplete, which could bias tree structure and thus, co-phylogeny
620 estimates, and also the can potentially under- (or over-)estimate the functional capacities of
621 recovered microbial genomes. To address the potential shortcomings of MAGs, we
622 implemented stringent data processing pipelines and quality control to achieve the high-
623 quality MAGs and a host-specific pan-genome based functional annotation framework
624 incorporating information from multiple MAGs per species representative to reduce
625 potential genome gaps (see Methods). Additionally, the commonly used estimates of
626 divergence times of the hominid hosts included in the analysis set the timeframe of the split
627 from a shared ancestor to 8-19 million years ago (Langergraber et al., 2012; Scally et al.,
628 2012). Bacterial speciation events happen in the timeframe of 10-100 million years, or
629 longer (Marin et al., 2017; McDonald and Currie, 2017; Ochman et al., 1999), and thus, co-
630 phylogeny in hominids is expected to be observed within microbial species or possibly
631 genera. In the presented dataset, species-level sharing of MAGs between host genera was
632 low (2.73%; n = 48 out of 1,759 newly reconstructed SGBs, not including Manara *et al*. and
633 UHGGv2). SGBs were defined on 95% average nucleotide identity, a measure generally
634 regarded as appropriate (Jain et al., 2018), but it is nevertheless prone to clade specific
635 biases, potentially further influenced by altered speciation dynamics in association with
636 (evolutionary) changes in host lifestyle (Lawrence and Retchless, 2009), *i.e.* previously
637 demonstrated increases in horizontal gene transfer (HGT) within individual microbiomes
638 (Groussin et al., 2021). Accounting for these potential biases, we relaxed the threshold to a
639 shared genus-level annotations for subgroups to be included in the co-phylogeny testing,

640 while keeping the number of tested sub-phylogenies to a minimum through the definition of
641 stringent inclusion criteria (see Methods). Despite these considerations, the observation of
642 signals of co-phylogeny across hominids is supported by a robust statistical framework.
643
644 Our work here lays the foundation for the analysis of disease-associated changes in the
645 human intestinal microbiome in an evolutionarily informed framework, thereby allowing
646 researchers to evaluate microbiome-associated inflammatory disorders from a point of view
647 that considers both proximal and evolutionary influences. Future investigations should
648 consider in-depth analysis of horizontal gene transfer events within or even between
649 primate hosts to shed further light on also cross-species dynamics and transition of
650 microbes. Such analyses however require either microbial isolate genomes or at least long-
651 read sequencing data to increase confidence in detection events. Additionally, time series
652 data for host groups sharing the same habitat, *e.g. G. g. gorilla* and *P. t. troglodytes*, could
653 give additional insights into cross-species sharing dynamics which cannot be appropriately
654 elucidated based on single-timepoint data.
655
656 In summary, we present an in-depth taxonomic and functional description and analysis of
657 hominid-associated fecal communities spanning about ten millions of years of evolution and
658 host-microbiome interactions in the gut of humans and African great apes. Western lifestyle
659 and maybe more precisely industrialization associated changes in human gut microbiota
660 have been previously suggested as a driver of microbiome insufficiency syndrome, whereby
661 an incompatibility between quickly adapting microbiota and slowly evolving host genes
662 leads to chronic inflammatory diseases such as metabolic syndrome, type 2 diabetes, and
663 inflammatory bowel disease (E. D. Sonnenburg and Sonnenburg, 2019; Wallenborn and
664 Vonaesch, 2022). Thus, a comparative analysis of human and NHP intestinal microbiota that
665 considers evolutionary forces as presented herein provides a powerful platform to advance
666 our understanding of human-associated microbiota and guide the development of
667 personalized, targeted interventions to prevent and treat chronic inflammatory disorders.
668

Tables

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

KEGG ID	Name	# of KOs	Mean Z-Score	P-value	Q-Value
path:map01230	Biosynthesis of amino acids	238	2.43	2.80E-17	2.29E-14
path:map00340	Histidine metabolism	47	4.69	3.92E-16	3.20E-13
md:M00159	V/A-type ATPase, prokaryotes	9	13.61	5.08E-15	4.15E-12
md:M00026	Histidine biosynthesis, PRPP => histidine	20	4.56	3.05E-13	2.49E-10
path:map00630	Glyoxylate and dicarboxylate metabolism	101	3.27	3.26E-11	2.66E-08
br:br01601	Enzymes of 2-oxocarboxylic acid metabolism	60	5.39	1.33E-08	1.09E-05
md:M00010	Citrate cycle, first carbon oxidation, oxaloacetate => 2-oxoglutarate	6	5.98	3.68E-07	3.01E-04
md:M00004	Pentose phosphate pathway (Pentose phosphate cycle)	15	7.29	8.40E-07	6.87E-04
md:M00045	Histidine degradation, histidine => N-formiminoglutamate => glutamate	8	4.7	1.68E-06	1.38E-03
md:M00631	D-Galacturonate degradation (bacteria), D-galacturonate => pyruvate + D-glyceraldehyde 3P	8	4.92	1.68E-06	1.38E-03
path:map00520	Amino sugar and nucleotide sugar metabolism	156	2.38	4.74E-06	3.87E-03
path:map00620	Pyruvate metabolism	132	3.41	6.86E-06	5.60E-03
path:map00260	Glycine, serine and threonine metabolism	109	2.54	9.95E-06	8.13E-03
md:M00009	Citrate cycle (TCA cycle, Krebs cycle)	42	3.81	1.37E-05	1.12E-02
path:map00020	Citrate cycle (TCA cycle)	66	4.35	1.94E-05	1.59E-02
md:M00008	Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate	5	7.59	1.97E-05	1.61E-02
path:map00190	Oxidative phosphorylation	223	8.71	2.74E-05	2.24E-02
md:M00006	Pentose phosphate pathway, oxidative phase, glucose 6P => ribulo se 5P	6	8.44	3.90E-05	3.19E-02

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 Collaborative Research Center 1182, Origin and Function of Metaorganisms (https://www.metaorganism-research.com, no. SFB1182, Projects A2 and Z2) and received infrastructure support from the DFG Excellence Cluster 2167 "Precision Medicine in Chronic Inflammation" (PMI). JFG was supported by the DFG grant “The ecology and evolution of primate phageomes” (GO 3443/1-1: Project number: 453352748) and the Deutsche Forschungsgemeinschaft (DFG) Research Group “Sociality and Health in Primates” (FOR2136; CA 1108/3-1). 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 and granted permission for sample collection: Bwindi—The mountain gorilla survey was conducted by the Uganda Wildlife Authority, l'Institut Congolais pour la Conservation de la Nature, the Rwanda Development Board, the International Gorilla Conservation Programme, the Max Planck Institute for Evolutionary Anthropology, Conservation Through Public Health, the Mountain Gorilla Veterinary Project, the Institute for Tropical Forest Conservation, and The Dian Fossey Gorilla Fund and was conducted in compliance with the regulations of and permission of the Uganda National Council for Science and Technology and the Uganda Wildlife Authority; Kokolopori—permission was granted through the Ministere de Recherche Scientifique et Technologie, Democratic Republic of the Congo, and work was supported by Vie Sauvage, the Bonobo Conservation Initiative; Loango—permission was granted by the Agence Nationale des Parcs Nationaux, the Centre National de la Recherche Scientifique et Technique of Gabon; Taï National Park—permission was granted by the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, the Ministère des Eaux et Fôrets in Côte d'Ivoire, and the Office Ivoirien des Parcs et Réserves, and work was supported by the Centre Suisse de Recherches Scientifiques en Côte d'Ivoire and the staff members of the Taï Chimpanzee Project. Ethical approval for work on human samples was obtained from the Local Ethics Committee Germany, Kiel (reference number A156/03), and we thank the national and local health authorities in Côte d'Ivoire and Democratic Republic of Congo as well as the according ethics commission for granting permission for this work. We thank Taylor Hermes for support in the writing of the manuscript.

Data and code availability

716 All metagenomic sequencing data is available via the NCBI BioProject accession IDs
717 PRJNA692042 and PRJNA539933. All code to process sequencing files to generate the
718 presented results is available via https://github.com/mruehlemann/greatapes_mgx_scripts.
719
720 **Conflicts of interest**
721 All authors declare no conflicts of interest.
722
723 **Material and Methods**
724 **Fecal Sampling, DNA extraction and data generation**
725 Sampling procedures for collecting feces from humans (n=48) and wild non-human primates
726 (n=109) have been previously described (Gogarten et al., 2021). Briefly, fecal samples were
727 collected immediately after defecation, and, depending on the local infrastructure, either
728 stored in RNAlater and frozen at -20°C or stored in a cryotube, cooled in a thermos until
729 return to the field laboratory, and subsequently snap frozen in liquid nitrogen. Appropriate
730 government permits and permission to conduct research on wild primates were granted by
731 the relevant authorities (see Acknowledgments for site-specific details). Human fecal
732 samples from the Democratic Republic of Congo (n = 12) were stored in RNAlater and frozen
733 at -20°C. Human fecal samples from Côte d'Ivoire (n=12) were stored in a cryotube, cooled
734 in a thermos until return to the field laboratory and subsequently snap frozen in liquid
735 nitrogen. Human fecal samples from Germany were collected at home by the participant in
736 standard fecal collection tubes, mailed to the study center, and stored at -80°C. Ethical
737 approval was obtained from the Local Ethics Committee (Kiel, Germany; reference number
738 A156/03). DNA extraction from fecal samples was performed from 200mg of stool
739 transferred to 0.70mm Garnet Bead tubes (Qiagen) with 1.1 mL ASL buffer, followed by
740 bead beating in a SpeedMill PLUS (Analytik Jena AG) for 45 s at 50 Hz. Samples were heated
741 to 95°C for 5 min and centrifuged, retaining 200 µl of the supernatant for DNA extraction
742 with the QIAamp DNA Stool Mini Kit (Qiagen) automated on a QIAcube system (Qiagen)
743 according to the manufacturer's protocol. DNA quality was assessed by Qubit and Genomic
744 DNA ScreenTape (Agilent). Illumina Nextera DNA Library Preparation Kit was used to
745 construct shotgun metagenomic libraries, and subsequently sequenced with either 2 × 125
746 bp reads on a HiSeq 2500 platform or with 2 × 150 bp reads on a HiSeq 4000 machine
747 (Illumina).

748

749

750 **Data processing, assembly and metagenomic binning**

751 Raw sequencing FastQ files were quality controlled and preprocessed using the BBMap
752 software suite ("BBMap," n.d.). Host reads were removed using a human reference
753 database and a lenient threshold of 95% identity to account for a broader host range using
754 bbmap.sh. Metagenomic contigs were assembled with metaSPAdes and contigs >= 2000 bp
755 were retained (Bankevich et al., 2012). Reads were mapped to the contigs of the respective
756 samples using Minimap2 (Li, 2018), converted to BAM files with Samtools (Li et al., 2009)
757 and used to estimate per-contig mapping depth with the
758 jgi_summarize_bam_contig__depths binary from the MetaBAT2 binning tool (Kang et al.,
759 2019). Contig binning for individual samples was performed with MetaBAT2 (Kang et al.,
760 2019), MaxBin2 (Wu et al., 2016) and CONCOCT (Alneberg et al., 2014). In addition, the
761 VAMB binning tool (Nissen et al., 2021) was used on a cross-mapping catalog of the merged
762 contigs from all samples within each host group. Individual binning results were refined
763 using MAGScoT (Rühlemann et al., 2022) to acquire high quality metagenome-assembled
764 genomes (MAGs) for each sample. Clustering of MAGs to species-level genome bins (SGBs)
765 was performed with dRep (Olm et al., 2017) in a multi-step approach to control for inflated
766 SGBs due to low MAG quality. First, MAGs were dereplicated to 97% similarity within each
767 host group, choosing the MAG with the highest score (calculated by MAGScoT based on
768 completeness and contamination) as cluster representative. High and good quality
769 representatives (score > =0.7) from all host groups together with representative sequences
770 from the UHGG v2 were then clustered into 95% SGBs using dRep, again selecting the
771 highest quality MAG as representatives. Medium quality (scores between 0.5 and 0.7) 97%
772 representatives from previous clustering step were then compared to SGB representatives
773 using fastANI (Jain et al., 2018), assigning MAGs with high similarity (>=95%) to the
774 respective SGB. Medium quality 97% representatives without hits to the high quality SGB
775 library were then clustered into 95% SGBs and added to the catalog in the case of at least
776 two genomes in the cluster, discarding singleton clusters. The final catalog of SGB
777 representatives was used to quantify contig abundances in all samples using Salmon in
778 metagenome mode (Patro et al., 2017). Taxonomic annotations were performed using the
779 GTDBtk (v2.1) and GTDB release 207v2 (Chaumeil et al., 2022; Parks et al., 2022). For SGBs

780 without genus- and/or species-level assignments, the SGB ID was used as taxonomic label.
781 GTDBtk marker gene alignments were used to generate a phylogenetic tree of all SGB
782 representatives using the respective "infer" function of the GTDBtk. All data processing
783 scripts are available online: https://github.com/mruehlemann/greatapes_mgx_scripts
784
785 **Pangenome catalog creation, annotation and analysis**
786 All MAGs underwent calling of coding sequences using prodigal (v2.6.3) (Hyatt et al., 2010).
787 Protein sequences were clustered based on 95% similarity using MMseqs (Hauser et al.,
788 2016; Steinegger and Söding, 2017) and annotated using the emapper.py script of the
789 eggNOG-mapper v2 (Cantalapiedra et al., 2021) annotation tool with the eggNOG 5.0
790 reference database (Huerta-Cepas et al., 2019). MAG level functional profiles based on
791 KEGG Ortholog annotations were collapsed into SGB-level pangenomes for each host genus
792 (*Homo, Gorilla*, and *Pan*). In the case that no MAGs of an SGB were recovered from a given
793 host genus, functional profiles were inferred from MAGs across the other host groups,
794 accounting for host-specificities in the inferred accessory genomes/functions by considering
795 a function to be present if it was present in all host-specific pangenomes of the respective
796 SGB with MAGs recovered from the metagenomic data.
797
798 **Calculation of microbial clade and functional abundances**
799 All downstream data processing and statistical calculations were performed in R v4.2 (R
800 Core Team, 2022) and using the tidyverse library (Wickham et al., 2019). Per-sample contig
801 abundances for the SGB representatives from Salmon were used to estimate SGB
802 abundances. Salmon output includes total mapped reads per contig and mapping reads
803 adjusted for library size and total sequencing depth as transcripts per million (TPM), a
804 measure from the transcriptomics field which can be directly transferred to metagenomic
805 libraries. Individual contig coverages were calculated from the number of mapped reads and
806 the effective lengths of the Salmon mapping output, considering contigs with > 10%
807 coverage as present. An SGB was considered present when at least 20% of its total length
808 was in contigs marked as "present" and if at least 1,000 total reads and 250 TPM mapped to
809 it. Final SGB abundances were calculated as TPM, calculated from the reads mapping to the
810 SGBs present in the respective sample, thus representing a normalized abundance across all
811 samples. Combining SGB abundances with taxonomic assignments, domain- to species-level

812 abundances were calculated as cumulative TPM abundances within the respective
813 taxonomic bins. Rarefactions were calculated based on 5-fold repeated subsampling of
814 contig level mapped reads at 100k, 250k, 500k, 1M, 2.5M, 5M, and 10M reads, followed by
815 TPM calculations as described above. By rarefying reads and not TPM we realistically
816 simulate sampling effects introduced by low coverage and low abundances of SGBs affecting
817 especially samples with small library sizes. Community level functional profiles were
818 calculated be multiplying TPM abundances of SGBs with the respective host-genus specific
819 functional profiles of the SGBs and summarizing the per-SGB values into a sample-level
820 abundance of functional annotations.
821
822 **Alpha and beta diversity**
823 Faith's phylogenetic diversity (PD) (Faith, 1992) was used as measure of alpha diversity,
824 calculated from the phylogenetic tree based on GTDBtk marker genes using the pd()
825 function of the picante package for R (Kembel et al., 2010). Genus-level increase of PD from
826 novel SGBs was calculated from the differences of PDs with and without novel SGB
827 annotated as the respective genus. Sample level PDs were calculated from the SGB
828 presence/absence patterns. Beta diversity was assessed as unweighted and weighted
829 UniFrac distances (Lozupone et al., 2011) using the UniFrac() function of the phyloseq
830 package for R (McMurdie and Holmes, 2013) and SGB abundances and the phylogenetic tree
831 based on GTDBtk marker genes as input. Aitchison distance (Aitchison, 1982) was calculated
832 from the CLR-transformed genus-level TPM abundances obtained from the clr() function
833 from the compositions package for R (van den Boogaart and Tolosana-Delgado, 2008) and
834 adding a pseudocount of 1 to all abundances, setting all CLR-transformed abundances below
835 zero to zero. Jaccard distances (Jaccard, 1912) were also calculated on genus-level
836 presence/absence patterns using the vegdist() function from the vegan package for R
837 (Oksanen et al., 2022). Genus-level abundances were chosen for Aitchison and Jaccard
838 distance, as SGBs are highly host-specific, thus would lead to high beta-diversities simply
839 due to host exclusive SGBs, grouping at genus-level prevents from this and UniFrac distances
840 use phylogenetic relations between SGBs.
841
842 **Cross-host sharing analysis**

843 Analysis of excessive and reduced sharing of SGBs between host groups were based on the
844 mean SGB abundances of the five rarefaction of 1M mapped reads to account for
845 differences in library depth impacting SGB richness. For each host group, 100-fold sampling
846 of five samples from this group were drawn and the SGBs found in the host were analyzed
847 for their presence in five random samples of each of the other host groups, calculating the
848 relative amount of shared SGBs. The mean of all 100 samplings was used as relative sharing
849 coefficient for all host pairs in both directions. Excess and reduced sharing was analyzed by
850 1000-fold drawing of five random samples accounting for differences of host groups and the
851 repetition of above calculations for relative sharing with all host groups. P-values were
852 calculated from the proportions of random samplings exceeding/falling below the true
853 sharing coefficients.
854
855 **Phylosymbiosis analysis**
856 Phylosymbiosis was assessed using five measures for community level diversity, unweighted
857 and weighted UniFrac, genus-level Aitchison, and Jaccard distances, as well as KEGG
858 ortholog (KO) abundance based Aitchison distance, and following the approach described in
859 Brooks et al. (2016). Briefly, host group differences were used to infer microbiome
860 dendrograms by UPGMA clustering. Branch support was calculated from 1000-fold jackknife
861 sampling. Robinson-Foulds distances between microbiome trees and host phylogeny were
862 calculated using the RF.dist() function from the phangorn package for R (Schliep, 2011).
863 Significance of phylosymbiosis was assessed using the host phylogeny and 100,000 random
864 trees as comparison for the microbiome trees. Tanglegrams were created with the ggtree
865 and cowplot packages for R (Yu et al., 2017).
866
867 **Assessment of between-group abundance differences**
868 Taxonomic and functional abundance differences between Humans and NHPs, as well as
869 between humans living outside and within industrialized systems were based on CLR-
870 transformed abundances to account for the compositionality of microbiome data (Gloor et
871 al., 2017). Included in the calculations were all genera with a prevalence > 20% and relative
872 abundance (before CLR-transformation) of > 0.1% in at least of the host groups and all KO
873 abundances with a prevalence > 20% and CLR-transformed abundance of > 1 in at least of
874 the host groups. Abundance differences were assessed using linear regression. Log-fold

875 differences were calculated using group mean abundances and a pseudocount of 0.01. P-
876 values were adjusted for multiple testing using Bonferroni correction. Genera with
877 significant differences with lifestyle were grouped into BloSSUM (Bloom or Selected in
878 Societies of Urbanization/Modernization) and VANISH (Volatile and/or Associated
879 Negatively with Industrialized Societies of Humans) taxa. Remaining genera with significant
880 differences between humans and NHPs were grouped according to these patterns. Genera
881 without abundance differences in any of these comparisons were grouped as "other".
882 Accordingly, significantly different KO abundances were grouped into association with
883 human subgroups or humans vs. NHPs depending on the statistical significance and
884 direction of the association.
885
886 **Functional pangenome differences between groups**
887 Pangenome catalogs of human-associated SGBs were compared within microbial families
888 between SGBs in BloSSUM and non-BloSSUM taxa, the latter comprising all VANISH taxa and
889 taxa with higher abundance in humans compared to NHPs, however not associated with the
890 industrialization gradient. KEGG Ontology (KO) term annotations were used as functional
891 groups and their prevalence differences between groups were assessed using Fisher's exact
892 test. Effect sizes were calculated using the log2 of the ratio of prevalences in the two groups
893 and a pseudo count of 0.01. Per-family Z-Scores of KOs were calculated from P-values and
894 the direction of the effects. The sum of the Z-Scores were added and divided by the square-
895 root of the total number of families the respective KOs were found in to obtain a Z_{Meta} for
896 each KO term, used to calculated P_{Meta}. P_{Meta}-values were adjusted for multiple testing using
897 Holm-correction. KO terms with $Q < 0.05$ and present in at least two of the microbial
898 families in the analysis were considered as functions with differential prevalence. A similar
899 approach was employed to assess functional differences between NHP- and Human-
900 associated SGBs, however in this case, SGB pangenome differences were compared on
901 genus level and the meta-analysis was performed combining signals from all genera, and
902 specifically across the genera within particular phyla.
903
904 **Tree reconciliation analysis**
905 Proteins from the representative SGBs of the genus *Prevotella* annotated with the
906 annotation "cydA" (cytochrome *bd* oxidase subunit 1) as "Preferred name" in the

907 emapper/eggNOG annotation were extracted from the unclustered protein sequence
908 catalog. The same procedure was followed for *Paraprevotella,* which was included as an
909 outgroup. Incomplete cydA sequences were removed using a length threshold of 200.
910 Protein sequences were aligned using Clustal Omega (Sievers et al., 2011). The alignment
911 was used to reconstruct the phylogenetic tree using IQTREE2 (Minh et al., 2020) and a WAG
912 model (Whelan and Goldman, 2001), calculating branch support using UFBoot (Minh et al.,
913 2013) and performing SH-aLRT test (Guindon et al., 2010). Alignments of GTDBtk marker
914 protein sequences for *Prevotella* and *Paraprevotella* SGBs were used to reconstruct a
915 genome-level species phylogeny in the same respective way as described above for the cydA
916 sequences. Very low confident branches (< 60% bootstrap support) in the cydA phylogeny
917 were resolved together with the species tree using the OptResolutions supplementary
918 program of the RANGER-DTL 2.0 software (Bansal et al., 2018) resulting in 495 equally
919 probable trees with optimized duplication-transfer-loss costs using default values
920 (duplication: 2, loss: 1, transfer: 3). A randomly chosen output tree was using in the
921 reconciliation analysis with the species tree in RANGER-DTL 2.0 using default values and
922 1,000 random starting seeds in parallel (Tange, 2011) to assess robustness. Resulting
923 sampling outcomes were summarized using the AggregateRanger tool of the RANGER-DTL
924 2.0 software package.
925
926 **Co-phylogeny analysis**
927 Host phylogenetic trees were obtained from the 10kTrees website ((Arnold et al., 2010);
928 https://10ktrees.nunn-lab.org/). To assure high quality microbial phylogenies for the co-
929 phylogeny analysis, family-level maximum-likelihood trees were reconstructed from the
930 GTDBtk marker gene alignments with the IQTREE2 software (Minh et al., 2020) and a WAG
931 model including a random SGB outside the respective families as outgroups. Family level
932 trees were rooted and for each SGB traced from tip to root to identify for each SGB the
933 smallest subtree which covered 4, 5, 6, and 7 host groups. Combining information from all
934 SGBs, the overall set of smallest trees to be included in the co-phylogeny analysis were
935 identified, discarding subtrees for which the inclusion criterion was fulfilled already for a
936 smaller tree starting from a different tip. In addition, subtrees spanning more than a single
937 genus were excluded from the analysis, as divergence times of microbial genera predate
938 divergence of the included hosts (Ochman et al., 1999). For all subgroups included in the

939 analysis, maximum-likelihood distances and trees using a WAG model in IQTREE2 were
940 inferred from the marker gene alignment of all MAGs assigned to the SGBs in the respective
941 subgroups. Co-phylogeny of the subgroup was assessed by randomly selecting one MAG per
942 host, calculating congruence with the host tree by Robinson-Foulds metric and by Mantel-
943 test (Hommola et al., 2009). Tiplabels were permutated 999-fold and P-values calculated.
944 This process starting from the random selection of one MAG per host was repeated 999
945 times to obtain final P-values. Family-level co-phylogeny ratios were calculated based on the
946 ratio of SGBs within subtrees with co-phylogeny signal and total SGBs in the respective
947 family that were included in the analysis. Enrichment of co-phylogeny for each microbial
948 family was calculated by using Fisher's exact test on the SGBs in the analysis dividing them
949 into four groups based on family membership and being in a subtree with co-phylogeny
950 signal. All P-values were adjusted using FDR correction.
951
952 **Correlation between co-phylogeny ratio and sporulation genes**
953 Genes were selected using the KEGG Ontology annotations, selecting KOs with the words
954 "spore" or "sporulation" in their description. This resulted in a total of 149 unique KOs. The
955 presence of these KOs in an SGB was assessed based on the pangenome annotations (see
956 above). The number of KOs was summed up for each genome and genus-level averages
957 were regressed against the family-level co-phylogeny ratios. The P-value of the correlation
958 was calculated from the two-sided t-statistic calculated from regression coefficient and
959 standard error.
960
961 **References**

962 Aitchison, J., 1982. The Statistical Analysis of Compositional Data. Journal of the Royal
963 Statistical Society: Series B (Methodological) 44, 139–160.
964 https://doi.org/10.1111/j.2517-6161.1982.tb01195.x
965 Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z.J., Pollard, K.S.,
966 Sakharova, E., Parks, D.H., Hugenholtz, P., Segata, N., Kyrpides, N.C., Finn, R.D., 2021.
967 A unified catalog of 204,938 reference genomes from the human gut microbiome.
968 Nat Biotechnol 39, 105–114. https://doi.org/10.1038/s41587-020-0603-3
969 Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman,
970 N.J., Andersson, A.F., Quince, C., 2014. Binning metagenomic contigs by coverage
971 and composition. Nat Methods 11, 1144–1146. https://doi.org/10.1038/nmeth.3103
972 Amato, K.R., Mallott, E.K., McDonald, D., Dominy, N.J., Goldberg, T., Lambert, J.E., Swedell,
973 L., Metcalf, J.L., Gomez, A., Britton, G.A.O., Stumpf, R.M., Leigh, S.R., Knight, R., 2019.
974 Convergence of human and Old World monkey gut microbiomes demonstrates the

975 importance of human ecology over phylogeny. Genome Biology 20, 201.
976 https://doi.org/10.1186/s13059-019-1807-z
977 Arnold, C., Matthews, L.J., Nunn, C.L., 2010. The 10kTrees website: A new online resource
978 for primate phylogeny. Evolutionary Anthropology: Issues, News, and Reviews 19,
979 114–118. https://doi.org/10.1002/evan.20251
980 Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes,
981 G.R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez,
982 L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K.,
983 Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain,
984 J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G.,
985 Wang, J., Guarner, F., Pedersen, O., de Vos, W.M., Brunak, S., Doré, J., MetaHIT
986 Consortium, Antolín, M., Artiguenave, F., Blottiere, H.M., Almeida, M., Brechot, C.,
987 Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner,
988 K.U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-
989 Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K.,
990 Maguin, E., Mérieux, A., Melo Minardi, R., M'rini, C., Muller, J., Oozeer, R., Parkhill, J.,
991 Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K.,
992 Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J.,
993 Ehrlich, S.D., Bork, P., 2011. Enterotypes of the human gut microbiome. Nature 473,
994 174–180. https://doi.org/10.1038/nature09944
995 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M.,
996 Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N.,
997 Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A New Genome Assembly
998 Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19, 455–477.
999 https://doi.org/10.1089/cmb.2012.0021
1000 Bansal, M.S., Kellis, M., Kordi, M., Kundu, S., 2018. RANGER-DTL 2.0: rigorous reconstruction
1001 of gene-family evolution by duplication, transfer and loss. Bioinformatics 34, 3214–
1002 3216. https://doi.org/10.1093/bioinformatics/bty314
1003 BBMap [WWW Document], n.d. . SourceForge. URL
1004 https://sourceforge.net/projects/bbmap/ (accessed 12.1.22).
1005 Brooks, A.W., Kohl, K.D., Brucker, R.M., Opstal, E.J. van, Bordenstein, S.R., 2016.
1006 Phylosymbiosis: Relationships and Functional Effects of Microbial Communities
1007 across Host Evolutionary History. PLOS Biology 14, e2000225.
1008 https://doi.org/10.1371/journal.pbio.2000225
1009 Campbell, T.P., Sun, X., Patel, V.H., Sanz, C., Morgan, D., Dantas, G., 2020. The microbiome
1010 and resistome of chimpanzees, gorillas, and humans across host lifestyle and
1011 geography. ISME J 14, 1584–1599. https://doi.org/10.1038/s41396-020-0634-2
1012 Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021.
1013 eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain
1014 Prediction at the Metagenomic Scale. Molecular Biology and Evolution 38, 5825–
1015 5829. https://doi.org/10.1093/molbev/msab293
1016 Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., Parks, D.H., 2022. GTDB-Tk v2: memory
1017 friendly classification with the genome taxonomy database. Bioinformatics 38, 5315–
1018 5316. https://doi.org/10.1093/bioinformatics/btac672
1019 Costa, F.R.C., Françozo, M.C.S., de Oliveira, G.G., Ignacio, A., Castoldi, A., Zamboni, D.S.,
1020 Ramos, S.G., Câmara, N.O., de Zoete, M.R., Palm, N.W., Flavell, R.A., Silva, J.S., Carlos,
1021 D., 2016. Gut microbiota translocation to the pancreatic lymph nodes triggers NOD2

1022 activation and contributes to T1D onset. J Exp Med 213, 1223–1239.
1023 https://doi.org/10.1084/jem.20150744
1024 Costea, P.I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M.J., Bushman, F.D., de Vos,
1025 W.M., Ehrlich, S.D., Fraser, C.M., Hattori, M., Huttenhower, C., Jeffery, I.B., Knights,
1026 D., Lewis, J.D., Ley, R.E., Ochman, H., O'Toole, P.W., Quince, C., Relman, D.A.,
1027 Shanahan, F., Sunagawa, S., Wang, J., Weinstock, G.M., Wu, G.D., Zeller, G., Zhao, L.,
1028 Raes, J., Knight, R., Bork, P., 2018. Enterotypes in the landscape of gut microbial
1029 community composition. Nat Microbiol 3, 8–16. https://doi.org/10.1038/s41564-
1030 017-0072-8
1031 Davidson, A.L., Chen, J., 2004. ATP-binding cassette transporters in bacteria. Annu Rev
1032 Biochem 73, 241–268. https://doi.org/10.1146/annurev.biochem.73.011303.073626
1033 Davies, J.S., Currie, M.J., Wright, J.D., Newton-Vesty, M.C., North, R.A., Mace, P.D., Allison,
1034 J.R., Dobson, R.C.J., 2021. Selective Nutrient Transport in Bacteria: Multicomponent
1035 Transporter Systems Reign Supreme. Frontiers in Molecular Biosciences 8.
1036 Deleu, S., Machiels, K., Raes, J., Verbeke, K., Vermeire, S., 2021. Short chain fatty acids and
1037 its producing organisms: An overlooked therapy for IBD? EBioMedicine 66, 103293.
1038 https://doi.org/10.1016/j.ebiom.2021.103293
1039 Ejby, M., Fredslund, F., Andersen, J.M., Vujičić Žagar, A., Henriksen, J.R., Andersen, T.L.,
1040 Svensson, B., Slotboom, D.J., Abou Hachem, M., 2016. An ATP Binding Cassette
1041 Transporter Mediates the Uptake of α-(1,6)-Linked Dietary Oligosaccharides in
1042 Bifidobacterium and Correlates with Competitive Growth on These Substrates. J Biol
1043 Chem 291, 20220–20231. https://doi.org/10.1074/jbc.M116.746529
1044 Faith, D.P., 1992. Conservation evaluation and phylogenetic diversity. Biological
1045 Conservation 61, 1–10. https://doi.org/10.1016/0006-3207(92)91201-3
1046 Giuffrè, A., Borisov, V.B., Arese, M., Sarti, P., Forte, E., 2014. Cytochrome bd oxidase and
1047 bacterial tolerance to oxidative and nitrosative stress. Biochimica et Biophysica Acta
1048 (BBA) - Bioenergetics, 18th European Bioenergetics Conference 2014 Lisbon,
1049 Portugal 1837, 1178–1187. https://doi.org/10.1016/j.bbabio.2014.01.016
1050 Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. Microbiome Datasets
1051 Are Compositional: And This Is Not Optional. Frontiers in Microbiology 8.
1052 Gogarten, J.F., Rühlemann, M., Archie, E., Tung, J., Akoua-Koffi, C., Bang, C., Deschner, T.,
1053 Muyembe-Tamfun, J.-J., Robbins, M.M., Schubert, G., Surbeck, M., Wittig, R.M.,
1054 Zuberbühler, K., Baines, J.F., Franke, A., Leendertz, F.H., Calvignac-Spencer, S., 2021.
1055 Primate phageomes are structured by superhost phylogeny and environment.
1056 Proceedings of the National Academy of Sciences 118, e2013535118.
1057 https://doi.org/10.1073/pnas.2013535118
1058 Gorvitovskaia, A., Holmes, S.P., Huse, S.M., 2016. Interpreting Prevotella and Bacteroides as
1059 biomarkers of diet and lifestyle. Microbiome 4, 15. https://doi.org/10.1186/s40168-
1060 016-0160-7
1061 Groussin, M., Mazel, F., Alm, E.J., 2020. Co-evolution and Co-speciation of Host-Gut Bacteria
1062 Systems. Cell Host & Microbe 28, 12–22.
1063 https://doi.org/10.1016/j.chom.2020.06.013
1064 Groussin, M., Mazel, F., Sanders, J.G., Smillie, C.S., Lavergne, S., Thuiller, W., Alm, E.J., 2017.
1065 Unraveling the processes shaping mammalian gut microbiomes over evolutionary
1066 time. Nat Commun 8, 14319. https://doi.org/10.1038/ncomms14319
1067 Groussin, M., Poyet, M., Sistiaga, A., Kearney, S.M., Moniz, K., Noel, M., Hooker, J., Gibbons,
1068 S.M., Segurel, L., Froment, A., Mohamed, R.S., Fezeu, A., Juimo, V.A., Lafosse, S.,

1069 Tabe, F.E., Girard, C., Iqaluk, D., Nguyen, L.T.T., Shapiro, B.J., Lehtimäki, J.,
1070 Ruokolainen, L., Kettunen, P.P., Vatanen, T., Sigwazi, S., Mabulla, A., Domínguez-
1071 Rodrigo, M., Nartey, Y.A., Agyei-Nkansah, A., Duah, A., Awuku, Y.A., Valles, K.A.,
1072 Asibey, S.O., Afihene, M.Y., Roberts, L.R., Plymoth, A., Onyekwere, C.A., Summons,
1073 R.E., Xavier, R.J., Alm, E.J., 2021. Elevated rates of horizontal gene transfer in the
1074 industrialized human microbiome. Cell 184, 2053-2067.e18.
1075 https://doi.org/10.1016/j.cell.2021.02.052
1076 Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New
1077 Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing
1078 the Performance of PhyML 3.0. Systematic Biology 59, 307–321.
1079 https://doi.org/10.1093/sysbio/syq010
1080 Hatakeyama, M., 2014. Helicobacter pylori CagA and gastric cancer: a paradigm for hit-and-
1081 run carcinogenesis. Cell Host Microbe 15, 306–316.
1082 https://doi.org/10.1016/j.chom.2014.02.008
1083 Hauser, M., Steinegger, M., Söding, J., 2016. MMseqs software suite for fast and deep
1084 clustering and searching of large protein sequence sets. Bioinformatics 32, 1323–
1085 1330. https://doi.org/10.1093/bioinformatics/btw006
1086 Hommola, K., Smith, J.E., Qiu, Y., Gilks, W.R., 2009. A permutation test of host-parasite
1087 cospeciation. Mol Biol Evol 26, 1457–1468. https://doi.org/10.1093/molbev/msp062
1088 Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H.,
1089 Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., Bork, P., 2019.
1090 eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology
1091 resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47, D309–
1092 D314. https://doi.org/10.1093/nar/gky1085
1093 Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal:
1094 prokaryotic gene recognition and translation initiation site identification. BMC
1095 Bioinformatics 11, 119. https://doi.org/10.1186/1471-2105-11-119
1096 Iljazovic, A., Amend, L., Galvez, E.J.C., de Oliveira, R., Strowig, T., 2021. Modulation of
1097 inflammatory responses by gastrointestinal Prevotella spp. – From associations to
1098 functional studies. International Journal of Medical Microbiology 311, 151472.
1099 https://doi.org/10.1016/j.ijmm.2021.151472
1100 Jaccard, P., 1912. The Distribution of the Flora in the Alpine Zone.1. New Phytologist 11, 37–
1101 50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x
1102 Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., Aluru, S., 2018. High
1103 throughput ANI analysis of 90K prokaryotic genomes reveals clear species
1104 boundaries. Nat Commun 9, 5114. https://doi.org/10.1038/s41467-018-07641-9
1105 Jha, A.R., Davenport, E.R., Gautam, Y., Bhandari, D., Tandukar, S., Ng, K.M., Fragiadakis, G.K.,
1106 Holmes, S., Gautam, G.P., Leach, J., Sherchand, J.B., Bustamante, C.D., Sonnenburg,
1107 J.L., 2018. Gut microbiome transition across a lifestyle gradient in Himalaya. PLoS
1108 Biol 16, e2005396. https://doi.org/10.1371/journal.pbio.2005396
1109 Johnson, A.J., Vangay, P., Al-Ghalith, G.A., Hillmann, B.M., Ward, T.L., Shields-Cutler, R.R.,
1110 Kim, A.D., Shmagel, A.K., Syed, A.N., Walter, J., Menon, R., Koecher, K., Knights, D.,
1111 2019. Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans.
1112 Cell Host & Microbe 25, 789-802.e5. https://doi.org/10.1016/j.chom.2019.05.005
1113 Joossens, M., Huys, G., Cnockaert, M., Preter, V.D., Verbeke, K., Rutgeerts, P., Vandamme,
1114 P., Vermeire, S., 2011. Dysbiosis of the faecal microbiota in patients with Crohn's

1115 disease and their unaffected relatives. Gut 60, 631–637.
1116 https://doi.org/10.1136/gut.2010.223263
1117 Jurtshuk, P., 1996. Bacterial Metabolism, in: Baron, S. (Ed.), Medical Microbiology.
1118 University of Texas Medical Branch at Galveston, Galveston (TX).
1119 Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic
1120 Acids Res 28, 27–30. https://doi.org/10.1093/nar/28.1.27
1121 Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z., 2019. MetaBAT 2: an
1122 adaptive binning algorithm for robust and efficient genome reconstruction from
1123 metagenome assemblies. PeerJ 7, e7359. https://doi.org/10.7717/peerj.7359
1124 Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D.,
1125 Blomberg, S.P., Webb, C.O., 2010. Picante: R tools for integrating phylogenies and
1126 ecology. Bioinformatics 26, 1463–1464.
1127 https://doi.org/10.1093/bioinformatics/btq166
1128 Kovatcheva-Datchary, P., Nilsson, A., Akrami, R., Lee, Y.S., De Vadder, F., Arora, T., Hallen, A.,
1129 Martens, E., Björck, I., Bäckhed, F., 2015. Dietary Fiber-Induced Improvement in
1130 Glucose Metabolism Is Associated with Increased Abundance of Prevotella. Cell
1131 Metab 22, 971–982. https://doi.org/10.1016/j.cmet.2015.10.001
1132 Kundu, S., Bansal, M.S., 2018. On the impact of uncertain gene tree rooting on duplication-
1133 transfer-loss reconciliation. BMC Bioinformatics 19, 290.
1134 https://doi.org/10.1186/s12859-018-2269-0
1135 Langergraber, K.E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., Inoue, E.,
1136 Inoue-Muruyama, M., Mitani, J.C., Muller, M.N., Robbins, M.M., Schubert, G.,
1137 Stoinski, T.S., Viola, B., Watts, D., Wittig, R.M., Wrangham, R.W., Zuberbühler, K.,
1138 Pääbo, S., Vigilant, L., 2012. Generation times in wild chimpanzees and gorillas
1139 suggest earlier divergence times in great ape and human evolution. Proc Natl Acad
1140 Sci U S A 109, 15716–15721. https://doi.org/10.1073/pnas.1211740109
1141 Lawrence, J.G., Retchless, A.C., 2009. The Interplay of Homologous Recombination and
1142 Horizontal Gene Transfer in Bacterial Speciation, in: Gogarten, M.B., Gogarten, J.P.,
1143 Olendzenski, L.C. (Eds.), Horizontal Gene Transfer: Genomes in Flux, Methods in
1144 Molecular Biology. Humana Press, Totowa, NJ, pp. 29–53.
1145 https://doi.org/10.1007/978-1-60327-853-9_3
1146 Lee, J., Yang, J., Zhitnitsky, D., Lewinson, O., Rees, D., 2014. Structural and functional
1147 characterization of a heavy metal detoxifying ABC transporter (997.2). The FASEB
1148 Journal 28, 997.2. https://doi.org/10.1096/fasebj.28.1_supplement.997.2
1149 Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34,
1150 3094–3100. https://doi.org/10.1093/bioinformatics/bty191
1151 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
1152 Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence
1153 Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.
1154 https://doi.org/10.1093/bioinformatics/btp352
1155 Lloyd-Price, J., Arze, C., Ananthakrishnan, A.N., Schirmer, M., Avila-Pacheco, J., Poon, T.W.,
1156 Andrews, E., Ajami, N.J., Bonham, K.S., Brislawn, C.J., Casero, D., Courtney, H.,
1157 Gonzalez, A., Graeber, T.G., Hall, A.B., Lake, K., Landers, C.J., Mallick, H., Plichta, D.R.,
1158 Prasad, M., Rahnavard, G., Sauk, J., Shungin, D., Vázquez-Baeza, Y., White, R.A.,
1159 Braun, J., Denson, L.A., Jansson, J.K., Knight, R., Kugathasan, S., McGovern, D.P.B.,
1160 Petrosino, J.F., Stappenbeck, T.S., Winter, H.S., Clish, C.B., Franzosa, E.A., Vlamakis,
1161 H., Xavier, R.J., Huttenhower, C., 2019. Multi-omics of the gut microbial ecosystem in

1162 inflammatory bowel diseases. Nature 569, 655–662.
1163 https://doi.org/10.1038/s41586-019-1237-9
1164 Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., Knight, R., 2011. UniFrac: an effective
1165 distance metric for microbial community comparison. ISME J 5, 169–172.
1166 https://doi.org/10.1038/ismej.2010.133
1167 Manara, S., Asnicar, F., Beghini, F., Bazzani, D., Cumbo, F., Zolfo, M., Nigro, E., Karcher, N.,
1168 Manghi, P., Metzger, M.I., Pasolli, E., Segata, N., 2019. Microbial genomes from non-
1169 human primate gut metagenomes expand the primate-associated bacterial tree of
1170 life with over 1000 novel species. Genome Biology 20, 299.
1171 https://doi.org/10.1186/s13059-019-1923-9
1172 Marin, J., Battistuzzi, F.U., Brown, A.C., Hedges, S.B., 2017. The Timetree of Prokaryotes:
1173 New Insights into Their Evolution and Speciation. Molecular Biology and Evolution
1174 34, 437–446. https://doi.org/10.1093/molbev/msw245
1175 Matute, J.D., Duan, J., Flak, M.B., Griebel, P., Tascon-Arcila, J.A., Doms, S., Hanley, T.,
1176 Antanaviciute, A., Gundrum, J., Mark Welch, J.L., Sit, B., Abtahi, S., Fuhler, G.M.,
1177 Grootjans, J., Tran, F., Stengel, S.T., White, J.R., Krupka, N., Haller, D., Clare, S.,
1178 Lawley, T.D., Kaser, A., Simmons, A., Glickman, J.N., Bry, L., Rosenstiel, P., Borisy, G.,
1179 Waldor, M.K., Baines, J.F., Turner, J.R., Blumberg, R.S., 2023. Intelectin-1 binds and
1180 alters the localization of the mucus barrier-modifying bacterium Akkermansia
1181 muciniphila. J Exp Med 220, e20211938. https://doi.org/10.1084/jem.20211938
1182 McCall, L.-I., Callewaert, C., Zhu, Q., Song, S.J., Bouslimani, A., Minich, J.J., Ernst, M., Ruiz-
1183 Calderon, J.F., Cavallin, H., Pereira, H.S., Novoselac, A., Hernandez, J., Rios, R.,
1184 Branch, O.H., Blaser, M.J., Paulino, L.C., Dorrestein, P.C., Knight, R., Dominguez-Bello,
1185 M.G., 2020. Home chemical and microbial transitions across urbanization. Nature
1186 microbiology 5, 108. https://doi.org/10.1038/s41564-019-0593-4
1187 McDonald, B.R., Currie, C.R., 2017. Lateral Gene Transfer Dynamics in the Ancient Bacterial
1188 Genus Streptomyces. mBio 8, e00644-17. https://doi.org/10.1128/mBio.00644-17
1189 McMurdie, P.J., Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive
1190 Analysis and Graphics of Microbiome Census Data. PLOS ONE 8, e61217.
1191 https://doi.org/10.1371/journal.pone.0061217
1192 Minh, B.Q., Nguyen, M.A.T., von Haeseler, A., 2013. Ultrafast Approximation for
1193 Phylogenetic Bootstrap. Molecular Biology and Evolution 30, 1188–1195.
1194 https://doi.org/10.1093/molbev/mst024
1195 Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A.,
1196 Lanfear, R., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
1197 Inference in the Genomic Era. Molecular Biology and Evolution 37, 1530–1534.
1198 https://doi.org/10.1093/molbev/msaa015
1199 Moeller, A.H., Caro-Quintero, A., Mjungu, D., Georgiev, A.V., Lonsdorf, E.V., Muller, M.N.,
1200 Pusey, A.E., Peeters, M., Hahn, B.H., Ochman, H., 2016. Cospeciation of gut
1201 microbiota with hominids. Science 353, 380–382.
1202 https://doi.org/10.1126/science.aaf3951
1203 Moeller, A.H., Li, Y., Mpoudi Ngole, E., Ahuka-Mundeke, S., Lonsdorf, E.V., Pusey, A.E.,
1204 Peeters, M., Hahn, B.H., Ochman, H., 2014. Rapid changes in the gut microbiome
1205 during human evolution. Proceedings of the National Academy of Sciences 111,
1206 16431–16435. https://doi.org/10.1073/pnas.1419136111
1207 Molinaro, A., Bel Lassen, P., Henricsson, M., Wu, H., Adriouch, S., Belda, E., Chakaroun, R.,
1208 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

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

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.

1255 Ridley, C.P., Lee, H.Y., Khosla, C., 2008. Evolution of polyketide synthases in bacteria.
1256 Proceedings of the National Academy of Sciences 105, 4595–4600.
1257 https://doi.org/10.1073/pnas.0710107105
1258 Rodríguez-Daza, M.C., Pulido-Mateos, E.C., Lupien-Meilleur, J., Guyonnet, D., Desjardins, Y.,
1259 Roy, D., 2021. Polyphenol-Mediated Gut Microbiota Modulation: Toward Prebiotics
1260 and Further. Frontiers in Nutrition 8.
1261 Rühlemann, M.C., Wacker, E.M., Ellinghaus, D., Franke, A., 2022. MAGScoT - a fast,
1262 lightweight, and accurate bin-refinement tool. Bioinformatics btac694.
1263 https://doi.org/10.1093/bioinformatics/btac694
1264 Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A.,
1265 Lappalainen, T., Mailund, T., Marques-Bonet, T., McCarthy, S., Montgomery, S.H.,
1266 Schwalie, P.C., Tang, Y.A., Ward, M.C., Xue, Y., Yngvadottir, B., Alkan, C., Andersen,
1267 L.N., Ayub, Q., Ball, E.V., Beal, K., Bradley, B.J., Chen, Y., Clee, C.M., Fitzgerald, S.,
1268 Graves, T.A., Gu, Y., Heath, P., Heger, A., Karakoc, E., Kolb-Kokocinski, A., Laird, G.K.,
1269 Lunter, G., Meader, S., Mort, M., Mullikin, J.C., Munch, K., O'Connor, T.D., Phillips,
1270 A.D., Prado-Martinez, J., Rogers, A.S., Sajjadian, S., Schmidt, D., Shaw, K., Simpson,
1271 J.T., Stenson, P.D., Turner, D.J., Vigilant, L., Vilella, A.J., Whitener, W., Zhu, B.,
1272 Cooper, D.N., de Jong, P., Dermitzakis, E.T., Eichler, E.E., Flicek, P., Goldman, N.,
1273 Mundy, N.I., Ning, Z., Odom, D.T., Ponting, C.P., Quail, M.A., Ryder, O.A., Searle,
1274 S.M., Warren, W.C., Wilson, R.K., Schierup, M.H., Rogers, J., Tyler-Smith, C., Durbin,
1275 R., 2012. Insights into hominid evolution from the gorilla genome sequence. Nature
1276 483, 169–175. https://doi.org/10.1038/nature10842
1277 Schaan, A.P., Sarquis, D., Cavalcante, G.C., Magalhães, L., Sacuena, E.R.P., Costa, J., Fonseca,
1278 D., Mello, V.J., Guerreiro, J.F., Ribeiro-dos-Santos, Â., 2021. The structure of Brazilian
1279 Amazonian gut microbiomes in the process of urbanisation. npj Biofilms
1280 Microbiomes 7, 1–12. https://doi.org/10.1038/s41522-021-00237-0
1281 Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27, 592–593.
1282 https://doi.org/10.1093/bioinformatics/btq706
1283 Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H.,
1284 Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable
1285 generation of high-quality protein multiple sequence alignments using Clustal
1286 Omega. Mol Syst Biol 7, 539. https://doi.org/10.1038/msb.2011.75
1287 Sonnenburg, E.D., Sonnenburg, J.L., 2019. The ancestral and industrialized gut microbiota
1288 and implications for human health. Nat Rev Microbiol 17, 383–390.
1289 https://doi.org/10.1038/s41579-019-0191-8
1290 Sonnenburg, J.L., Sonnenburg, E.D., 2019. Vulnerability of the industrialized microbiota.
1291 Science 366, eaaw9255. https://doi.org/10.1126/science.aaw9255
1292 Steinegger, M., Söding, J., 2017. MMseqs2 enables sensitive protein sequence searching for
1293 the analysis of massive data sets. Nat Biotechnol 35, 1026–1028.
1294 https://doi.org/10.1038/nbt.3988
1295 Suzuki, T.A., Fitzstevens, J.L., Schmidt, V.T., Enav, H., Huus, K.E., Mbong Ngwese, M.,
1296 Grießhammer, A., Pfleiderer, A., Adegbite, B.R., Zinsou, J.F., Esen, M., Velavan, T.P.,
1297 Adegnika, A.A., Song, L.H., Spector, T.D., Muehlbauer, A.L., Marchi, N., Kang, H.,
1298 Maier, L., Blekhman, R., Ségurel, L., Ko, G., Youngblut, N.D., Kremsner, P., Ley, R.E.,
1299 2022. Codiversification of gut microbiota with humans. Science 377, 1328–1332.
1300 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

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

Tung, J., Barreiro, L.B., Burns, M.B., Grenier, J.-C., Lynch, J., Grieneisen, L.E., Altmann, J., Alberts, S.C., Blekhman, R., Archie, E.A., 2015. Social networks predict gut microbiome composition in wild baboons. eLife 4, e05224. https://doi.org/10.7554/eLife.05224

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

Whelan, S., Goldman, N., 2001. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. Molecular Biology and Evolution 18, 691–699. https://doi.org/10.1093/oxfordjournals.molbev.a003851

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

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