

Codiversification of gut microbiota with humans

Taichi A. Suzuki^{1*}, Liam Fitzstevens^{1*}, Victor T. Schmidt¹, Hagay Enav¹, Kelsey Huus¹,
Mirabeau Mbong¹, Bayode R. Adegbite^{2,3}, Jeannot F. Zinsou^{2,3}, Meral Esen^{2,4,5}, Thirumalaisamy
P. Velavan^{2,6}, Ayola A. Adegnika^{2,3,4,7}, Le Huu Song^{6,8}, Timothy D. Spector⁹, Amanda L.
Muehlbauer¹⁰, Nina Marchi¹¹, Ran Blekhman¹², Laure Ségurel^{11,13}, Nicholas D. Youngblut¹,
Peter Kremsner^{2,3,4,5,6}, and Ruth E. Ley^{1,5,#}

*=co-first

correspondence: rley@tuebingen.mpg.de

¹ Max Planck Institute for Developmental Biology, Department of Microbiome Sciences, Tübingen, Germany.

² University of Tübingen, Institute for Tropical Medicine, Tübingen, Germany.

³ Centre de Recherches Médicales de Lambaréné, Lambaréné, Gabon.

⁴ German Center for Infection Research (DZIF), Braunschweig, Germany.

⁵ Cluster of Excellence EXC 2124 Controlling Microbes to Fight Infections, University of Tübingen, Tübingen, Germany.

⁶ Vietnamese German Center for Medical Research (VG-CARE), Hanoi, Vietnam.

⁷ Fondation pour la Recherche Scientifique, Cotonou, Bénin.

⁸ 108 Military Central Hospital, Hanoi, Vietnam.

⁹ Department of Twin Research & Genetic Epidemiology, King's College London, London, United Kingdom.

¹⁰ Department of Ecology, Evolution, and Behavior, University of Minnesota, Minneapolis, MN, USA.

¹¹ Eco-anthropologie, Muséum National d'Histoire Naturelle–CNRS–Université de Paris, Paris, France.

¹² Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, MN, USA.

¹³ Laboratoire de Biométrie et Biologie Evolutive, CNRS - Université Lyon 1, Villeurbanne, France.

Abstract

Some gut microbes have cospeciatiated with hominids, but whether they further codiversified with human populations is unclear. Here, we identify predominant gut microbial species sharing a parallel evolutionary history with human populations. Patterns of strain transfer between populations are generally consistent with an African origin, and suggest long-term vertical transmission over thousands of generations. We show the same strains also faithfully transmit between mothers and their children. Consistent with the development of intimate symbiosis, species with strongest patterns of codiversification have the smallest genomes. This study reveals long-term fidelity of gut microbiota with human populations through transmission among individuals living in close proximity. Dominance of specific strains in different populations is based in part on vertical transmission and they may provide population-specific health benefits.

One-sentence summary: Identification of gut microbes that codiversified with human populations.

1 **Main text**

2 Coevolution, an evolutionary process involving reciprocal selection between two or more
3 species, is more likely to occur when the partners share a parallel evolutionary history, a pattern
4 termed codiversification (1). Codiversification between members of the gut microbiota and their
5 mammalian hosts has been suggested for specific gut bacteria, based on host phylogenies
6 matching bacterial single-gene phylogenies (2–4). Using this approach in wild apes, Moeller et
7 al. (2016) demonstrated that several gut bacterial families codiversified with hominids over the
8 last 15 million years (5). These patterns of codiversification may extend within human
9 populations, as has been shown for one species so far. In a landmark study using multiple locus
10 sequence typing of cultured strains, Falush et al. showed that *Helicobacter pylori*, the stomach
11 bacterium and causative agent of gastric cancer, codiversified with its human hosts,
12 recapitulating patterns of human migration (6).

13 The process of host-microbial codiversification often implies vertical transmission of
14 microbes within related individuals. A recent analysis of captive apes indicates that a shared
15 modern environment could obscure patterns of codiversification, and that geographic separation
16 may be necessary to maintain them (7). Within human populations, specific gut microbes are
17 known to transmit vertically from mothers to children (8, 9), which would allow for their
18 codiversification. Methods that use multiple marker genes for inferring microbial genomes from
19 metagenomes have revealed that a number of commensal gut microbial species also have
20 dominant strains that differ between human populations (10–13). Although environmental
21 factors are sufficient to drive population-level differences in strain dominance, multigenerational
22 vertical transmission may also underlie some of these geographic patterns.

23 Here, we assessed codiversification between gut microbial species and their human hosts
24 by comparing their evolutionary histories. Human- and microbial-genomes from the same

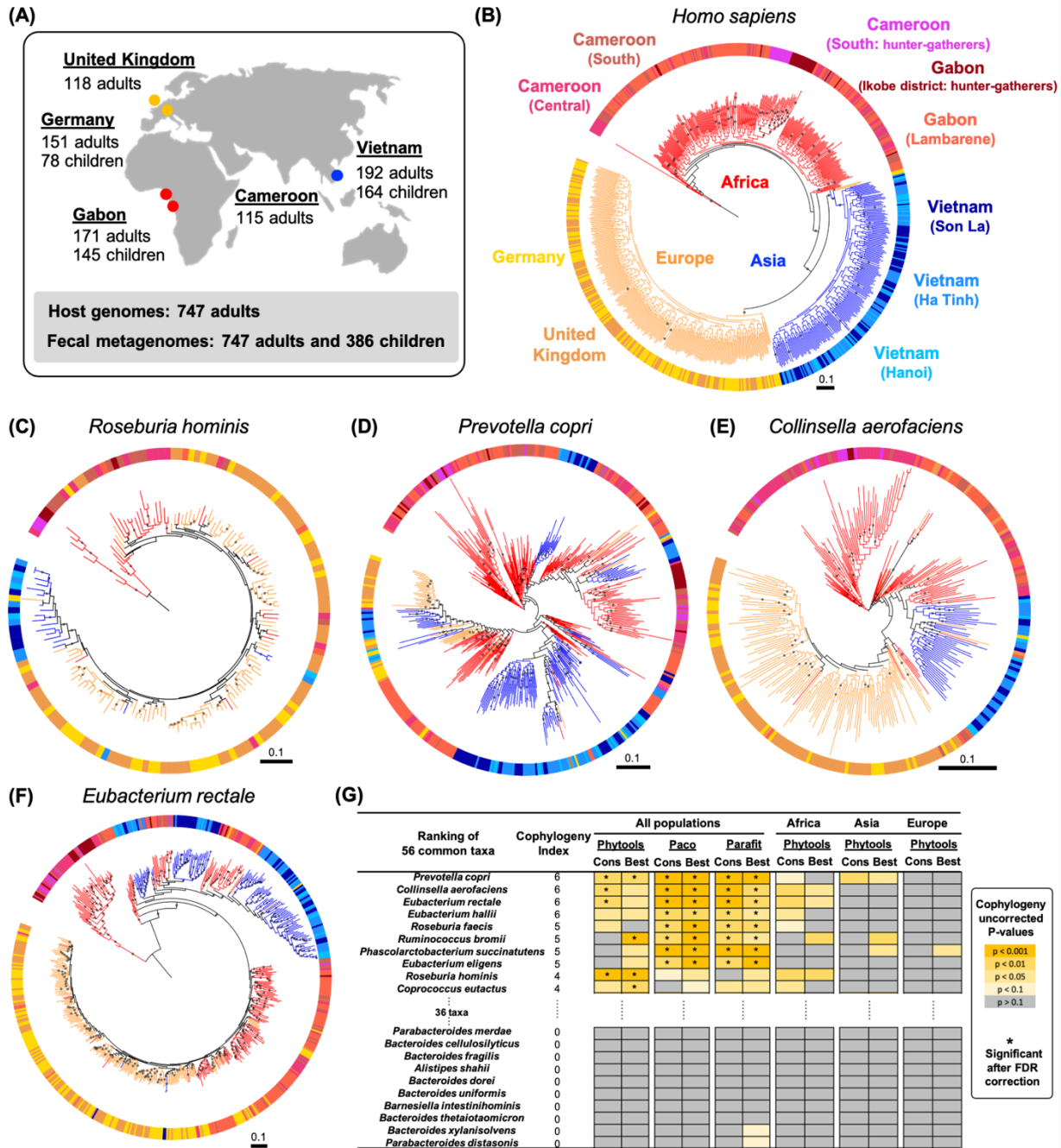
25 individuals allowed us to assess codiversification for a suite of gut microbial species. To also
26 gauge short-term vertical transmission, we characterized strain-sharing within mother-child
27 pairs. After enrolling healthy adult women, with an emphasis on mothers with young children,
28 into the study in three countries (Gabon, Vietnam, Germany), we generated chip-based genotype
29 data for the mothers and fecal metagenomes for mothers and children (Adults/Children for
30 Gabon: 171/144; Vietnam: 192/164; Germany: 151/78). Additionally, we generated human
31 genotype data from Cameroon (n=115) matching to published adult fecal metagenomes (14, 15),
32 and used matching human genomes and fecal metagenomes in adult subjects from the United
33 Kingdom (n=118) (16). Altogether, we assessed codiversification between humans and gut
34 microbial taxa for 747 adults and 386 children, and within generation strain transmission in 386
35 mother-child pairs (Fig. 1A) (Table S1&S2).

36 A maximum likelihood phylogeny of the human hosts based on 32,723 SNPs showed
37 human hosts clustering into three robust major groups matching their geographic origins: Africa
38 (Gabon and Cameroon), Asia (Vietnam), and Europe (Germany and UK) (Fig. 1B). We
39 recovered genomic information for gut microbes from metagenomes and built phylogenies using
40 two methods: (i) species-specific marker genes using StrainPhlan3 (10), and (ii) metagenome-
41 assembled-genomes (MAGs) constructed using PhyloPhlAn (17). Since the recovery of
42 microbial sequences from metagenomes largely depends on a combination of read depth and the
43 relative abundance of taxa (10), for the marker-based phylogenies we focused on 56 species
44 detected in ≥ 100 adults with ≥ 10 individuals per major human group (see Methods and Table
45 S3-S6). MAG-based trees were created for 32/56 taxa in adults. We quantified the degree of
46 codiversification between the host and microbial trees using three different permutation-based
47 methods: Phytools (18), PACo (19), and Parafit (20) (see Methods).

48 These analyses revealed that a subset of the gut microbiota shows significant
49 codiversification with humans. Using our strictest criteria (Phytools using consensus trees),
50 12/56 species assessed show evidence of codiversification (see Methods) (Table S5). Within the
51 12 taxa, four remained significant for codiversification after FDR correction using the most
52 conservative test (Fig. 1C-G). These four taxa include *Prevotella copri* (Fig. 1D) and
53 *Eubacterium rectale* (Fig. 1F), which have previously been shown to exhibit population-specific
54 strain diversity (10–12). For example, Tett et al. suggested that strain diversity of *P. copri* has an
55 African origin before out-of-Africa migration events, and led to population specific strains and
56 genes related to carbohydrate metabolism (11). The two others, *Roseburia hominis* (Fig. 1C) and
57 *Collinsella aerofaciens* (Fig. 1E), are well-known carbohydrate fermenters and short-chain fatty
58 acid producers and are linked to human health (21, 22).

59 To compare the relative degree of host-microbial codiversification among the 56 taxa
60 using all tests, we defined a “Cophylogeny Index (CI)” as the number of uncorrected p-values <
61 0.05 across six tests (see Methods). Eight taxa had CIs of 5 or 6, including several *Eubacterium*
62 spp. (Fig. 1G, Table S5). Intriguingly, *Eubacterium* spp. have been suggested to be under
63 population-specific selection pressures due to their strong biogeographical patterns (10, 13).
64 Other taxa with evidence of codiversification have been linked to host genetic variation,
65 including *Bifidobacterium*, *Collinsella*, *Coprococcus*, and *Methanobrevibacter* (36, 37).
66 Generally, gut microbes with high CIs belonged to the Firmicutes phylum (Fig 2A). In accord,
67 strains within species of Firmicutes are reported as more geographically restricted
68 compared to those of other phyla (7, but see 12). In contrast, taxa with low CIs belonged to the
69 Bacteroidetes phylum, including several species of *Bacteroides* and *Parabacteroides* (Fig. 1G,
70 Fig. 2A Table S5), which are known to survive in environments such as water (23). Applying the

71 same tests to trees constructed with MAGs resulted in consistent results (Table S7). Thus, our
 72 results identify a subset of gut microbes whose strain phylogenies are strongly indicative of
 73 codiversification with their human hosts.



74

Figure 1. Evidence of codiversification between humans and gut microbes. (A) Study design showing sampling locations and sample sizes. (B) A maximum likelihood phylogeny of human hosts based on 32,723 SNPs. (C-F) Maximum likelihood phylogenies based on species-specific marker genes for four bacteria species that showed significant codiversification based on the strictest test (Phytools using consensus trees) after FDR correction. Bootstrap values > 50% are plotted and all phylogenies were rooted at the midpoint. The scale bars show substitutions per site. Colors of branches and outer strips correspond to sampling locations shown in panel B. (G) Codiversification test p-values based on between and within host populations for the 10 top- and bottom-ranking taxa. Cophylogeny index (CI) is based on the number of uncorrected p-values < 0.05 across six tests using all data: three tests (Phytools, Paco, and Parafit) on two types of trees (Cons: majority-rule consensus tree and Best: best maximum likelihood tree). Columns “Africa”, “Asia”, and “Europe” show within-population results in the respective group. Uncorrected p-values that remained significant after FDR correction are indicated by *. The codiversification test results for all 56 taxa with more detailed information can be found in Table S5.

75 We observed a strong negative correlation between CI and genome size (Fig. 2B). Seven
76 out of the eight genera had a negative relationship between CI and genome size independently
77 (Binomial test $p = 0.035$, Fig. S1). Strict vertical transmission of symbionts in other hosts, such
78 as insects, is associated with genome reduction of the microbe resulting from reduced purifying
79 selection (24). Our results provide the first evidence in mammals that a similar process of host
80 adaptation may be occurring through genome reduction.

81 It is well established that modern humans originated in Africa, and therefore long-term
82 vertical transmission of microbial taxa would be expected to result in out-of-Africa patterns. To
83 test for this, we quantified the number and direction of host-region-switch events by applying
84 stochastic character mapping for all 56 microbial phylogenies (marker trees, Fig. 2C-E, Fig.
85 S2A). We found that taxa with high CIs exhibited significantly more transfer events from Africa
86 to other regions, compared to taxa with low CIs ($\rho = 0.40$, $p = 0.003$) (Fig. 2C). Transfer
87 events from Asia to other regions were the same for the top and bottom CI taxa ($\rho = -0.17$, $p =$
88 0.22) (Fig. 2D). In contrast, taxa with low CIs had higher proportions of transfer events from

89 Europe to other regions, compared to high-CI taxa ($\rho = -0.29$, $p = 0.029$) (Fig. 2E). MAG-
90 based trees resulted in similar patterns (Fig. S2B). Overall, the results are consistent with an
91 African origin for high CI taxa.
92

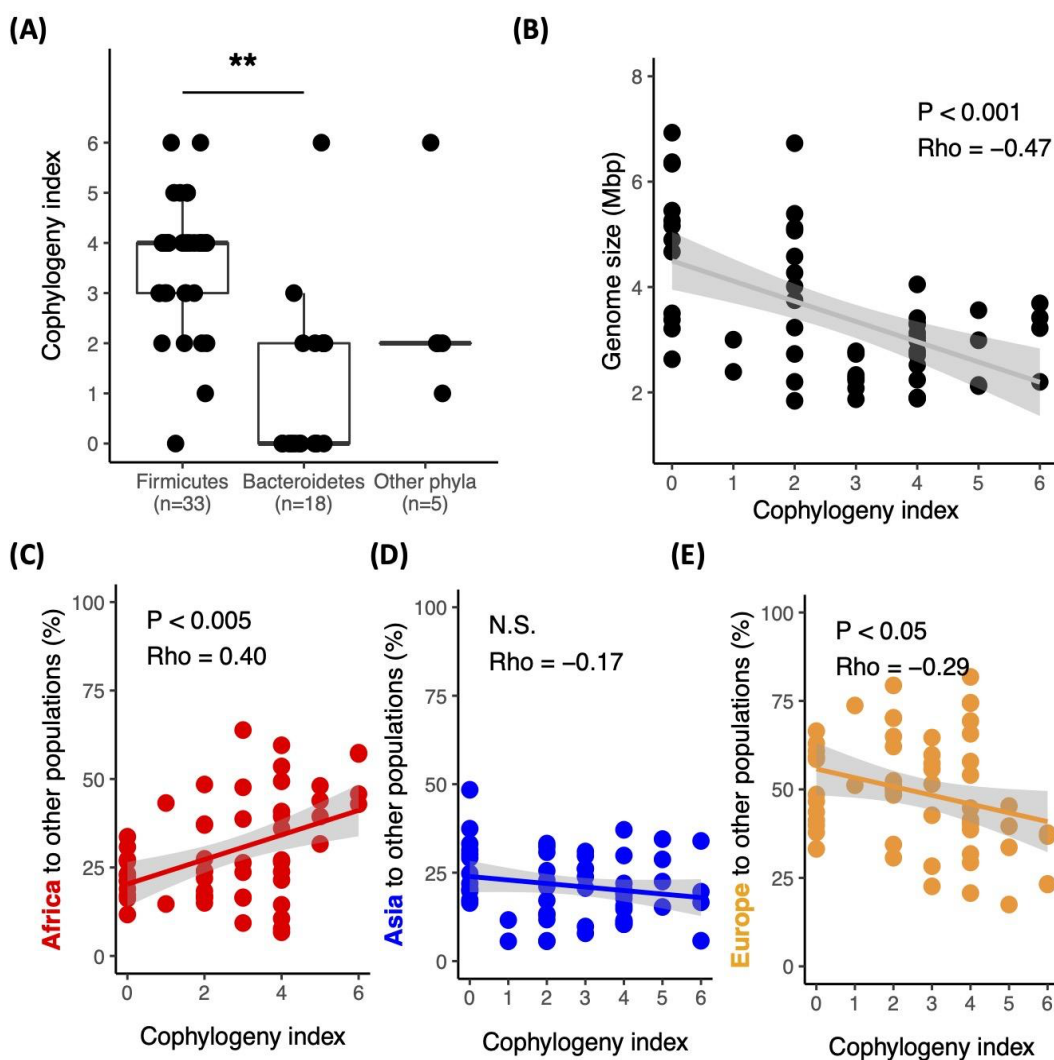


Figure 2. Taxonomy, genome size, and transfer events are related to host-microbial codiversification. (A) CI values by phylum. The P-value is based on the Wilcoxon rank sum test. (B) Correlation between genome size and CI. Correlations between CI and the proportions of transfer events from (C) Africa, (D) Asia, and (E) Europe. The results are based on stochastic character mapping. Each dot represents a microbial species. All correlations are based on Spearman's ρ . The 95% confidence intervals are shaded in gray. ** = $p < 5 \times 10^{-5}$; N.S., non significant ($p > 0.10$).

93 To corroborate the out-of-Africa pattern with independent data, we used 1,219 public
94 human gut metagenomes (Fig. 3A, Table S8&9). Strains derived from the public metagenomes
95 generally clustered with those derived from our samples based on the sampling locations (Fig.
96 3B-F, Fig. S3), and the resulting phylogenies were consistent with those of previous studies (10,
97 11, 13). We observed the same out-of-Africa strain transfer patterns for the top CI-taxa using the
98 public metagenomes with (Fig. 1&3, Fig. S2C) or without the five populations used in the initial
99 analyses (Fig. S2D). We also included primate fecal metagenome data (n=107), and for high CI-
100 species, strains from primate hosts tended to be the most early diverging strains in relation to all
101 human strains (Fig. 3, Fig. S3). In contrast, low-CI taxa had primate strains nested within human
102 strains (Fig. S4). Overall, the results are consistent with human gut microbial taxa originating
103 from ancestors common to hominids (5), and extend this process into human populations as
104 modern humans migrated from Africa.

105 Our inference of strain transfer events is based on patterns within the phylogenies for
106 present-day strains, yet ancient DNA analysis can provide a snapshot directly from the past.
107 Using high-quality ancient MAGs derived for five gut microbial species recovered from 1,000 -
108 2,000 year old paleofeces of Native North American tribes (25), we observed that four out of the
109 five taxa (the exception is *Escherichia coli*) were most closely related to strains from modern
110 East Asians (Fig. S5), consistent with the known migration history of the Americas (26).

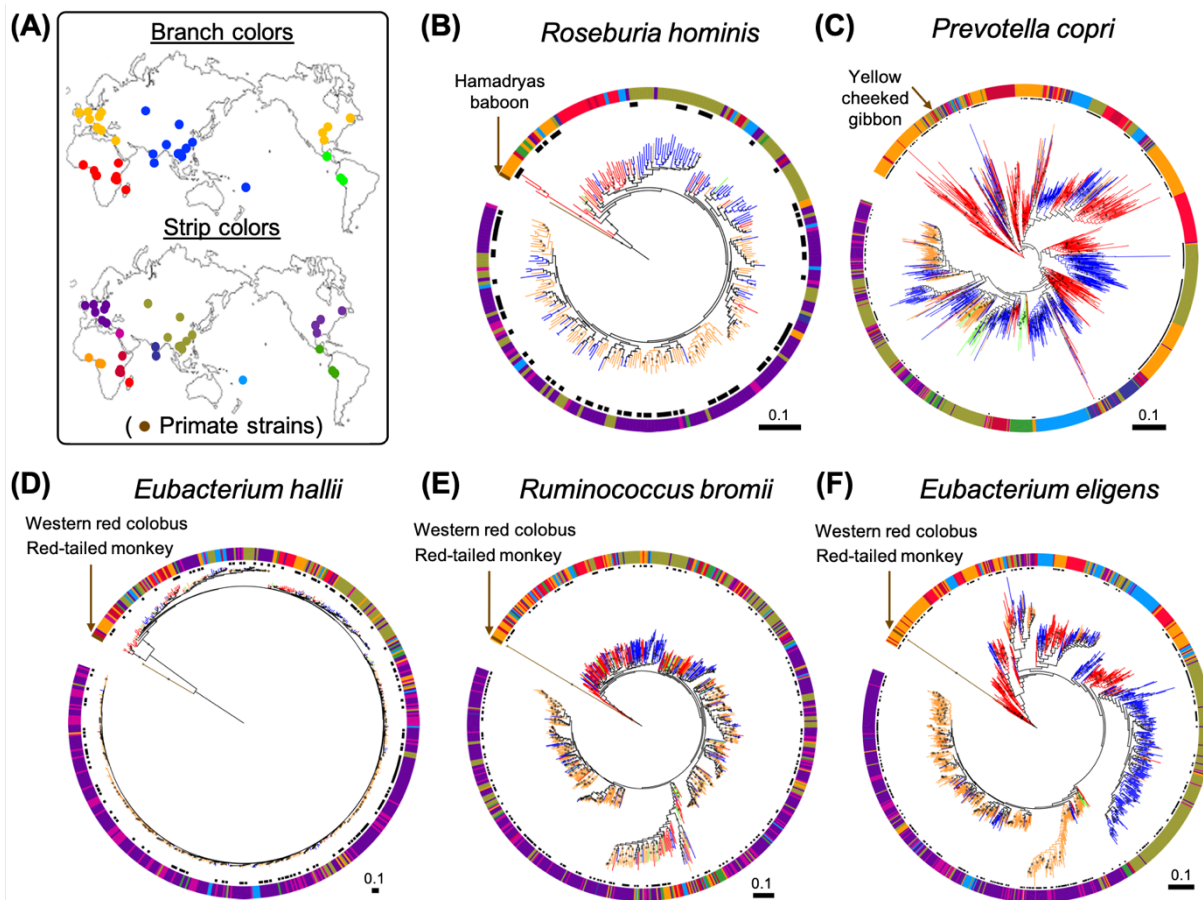


Figure 3. Publicly available fecal metagenomes of humans and primates support the original findings. (A) Map showing the source of fecal metagenomes used in B-F. The colors of the sampling location dots on the map correspond to colors used in B-F for the branches and outer strips. Colors are human genetic structures estimated from sampling locations based on data in Duda and Zrzavy et al. (27) (Table S8&9). (B-F) Maximum likelihood phylogenies of five out of the 10 top-ranking taxa constructed using data generated here (black dots next to the outer strip) combined with data from human and primate public metagenomes. All primate strains in the trees are labeled. Bootstrap values > 50% are plotted. All phylogenies were rooted at the midpoint. The scales are substitutions per site.

111 Next, we tested for evidence of codiversification in the child microbiomes using the
112 mothers' genotypes. *P. copri* showed evidence of codiversification ($p < 0.05$) in both adults (Fig.
113 1D, Table S5) and children (Fig. S6, Table S6). In addition to *P. copri*, we observed
114 *Bifidobacterium* species with high CIs in children. Among the 20 common taxa tested (see

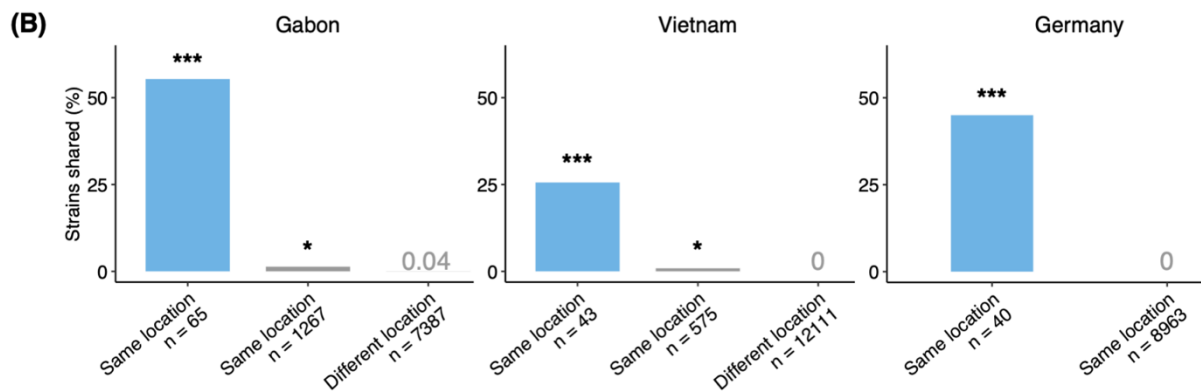
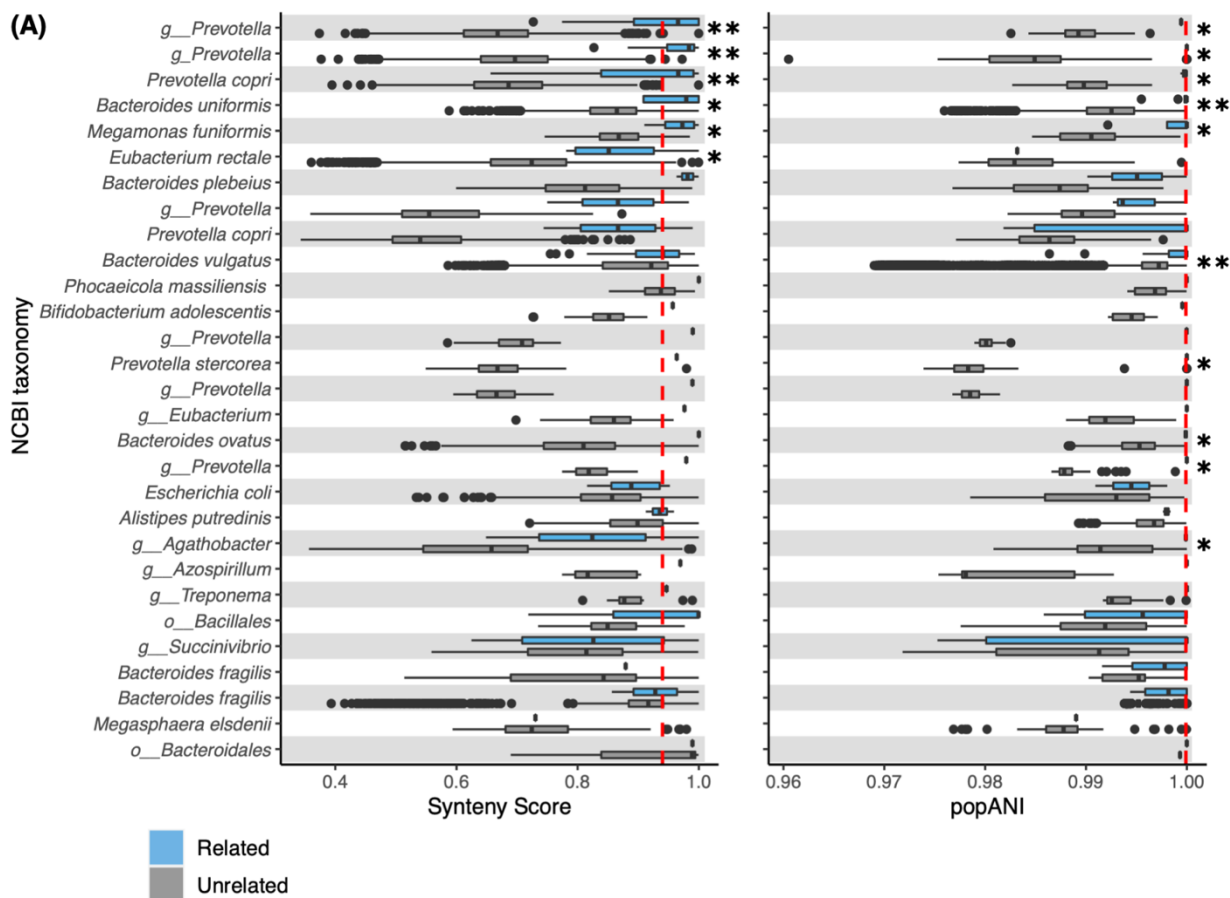
115 Methods), 8/20 showed significance ($p < 0.05$), including all four *Bifidobacterium* species (based
116 on the strictest test), and three remained significant after FDR correction (*B. bifidum*, *B. longum*,
117 and *B. breve*; Fig. S6). Using MAG-based trees, we tested 11 taxa: *B. longum* also showed
118 significance based on the strictest test (Table S10). As with adults, child taxa with highest CIs
119 tended to have transfer events from Africa to the rest of the populations ($p = 0.052$ Fig. S2E).
120 Our results extend those of previous studies, where the bacterial family Bifidobacteriaceae
121 exhibited codiversification with hominid species (5).

122 Signals of codiversification extended within populations for both adults and children.
123 Within-population tests had smaller sample sizes and the signal of codiversification tended to be
124 weaker: none of the p-values remained significant after FDR correction for adults (Fig. 1G,
125 Table S11), and only one taxon remained significant in children (*Bifidobacterium bifidum* within
126 Africa, Table S6). The patterns are nevertheless consistent: many of the same taxa with high-CI
127 also showed evidence of codiversification within populations, particularly for Africa (Fig. 1G).
128 For example, *P. copri* showed evidence of codiversification within-populations in adults (Africa:
129 $p = 0.089$ and Asia: $p = 0.0027$; for children $p = 0.094$). Host population differences in diet likely
130 affect the abundance and diversity of *P. copri* strains (11). Our results between- and within-
131 populations support long-term vertical transmission contributing to the observed pattern as well.

132 Taxa that are transmitted vertically over thousands of generations are expected to be
133 shared in mother-child pairs today. We used two independent approaches to search for shared
134 strains in mother-child pairs: inStrain (28), a tool based on population-level nucleotide strain
135 diversity (popANI), and SynTracker (29), based on the synteny of long-stretches of DNA. We
136 identified bacterial species with strains more similar between mothers and their own children
137 compared to unrelated children. Of the 63 taxa (adult and child taxa combined) assessed for
138 codiversification, 20 were tested for strain-sharing events between related and unrelated mother-

139 child pairs. Of these, notably, strains of *P. copri* and other *Prevotella* spp. were shared between
140 mothers and their children using two independent methods (Fig. 4A, Fig S7&S8). *E. rectale* also
141 showed evidence of mother-child strain sharing (Fig. 4A&S7). These results are consistent with
142 the observed codiversification patterns (Fig. 1, Fig. 3C, Fig. S6).

143 In addition to mother-child strain transmission, we also observed strain sharing between
144 mothers and children who were not their own, particularly for individuals sampled in the same
145 location (Fig. 4B). Higher rates of transmission among socially engaging individuals (that are
146 often genetically related) are known in group-living species such as bees (30), mice (31), and
147 great apes (32, 33), and this can provide signals of codiversification at the population-level but
148 weak signals of vertical transmission between mother-child pairs. Similarly, for taxa where we
149 find no evidence of codiversification but evidence of vertical transmission, such as *Bacteroides*
150 *uniformis* (Fig. 1G, Fig. 4B), this pattern may be due to changes in lifestyle across time (*e.g.*,
151 greater sanitation or isolation reducing group-level transmission).



152

153

154

155

156

Figure 4. Strain sharing assessed within related and unrelated mother-child pairs in Gabon, Vietnam, and Germany. (A) The relatedness of strains per species within related and unrelated mother-child pairs, using SynTracker (left) and inStrain (right). Species listed are those identified in both related and unrelated mother-child pairs. Dashed red lines indicate the pairwise comparison thresholds for strain sharing events (0.96 for synteny; 0.99999 for popANI). The full list of taxa tested in each method is shown in Fig. S7&8. (B) Strain sharing within related and unrelated mother-child pairs by country using inStrain. The percentage of total strain comparisons (“n”) identified as strain sharing events between related mother-child pairs, unrelated mother-child pairs sampled in the same location, and unrelated mother-child pairs sampled in different locations. Stars correspond to q-values (corrected Wilcoxon-Mann-Whitney test for Fig. 4A and corrected hypergeometric test for Fig. 4B; * = $p < 0.05$; ** = $p < 5 \times 10^{-5}$; *** = $p < 5 \times 10^{-25}$; Table S12). Comparisons were calculated with samples collected from 6 locations in Gabon (average inter-location distance = 114km), 4 locations in Vietnam (average inter-location distance = 238km) and 1 location in Germany (Table S13). The phyla comprising the strain sharing events shown here are described in Fig. S9.

157 This work yielded a first list of taxa that codiversified with humans, yet more likely
158 remain to be identified. Indeed, less prevalent and/or low-abundance taxa will have been missed
159 due to low coverage. The taxa most impacted may be those at low relative abundance in any one
160 of the major human groups (*i.e.*, Africa, Asia, and Europe): such taxa would be excluded in this
161 study. In addition, even though we could detect several strains per species per subject using
162 SynTracker and inStrain, we only tested for codiversification for the more dominant strains
163 because marker-based approaches extract the dominant strain per microbial species per
164 individual (*10*), and the MAG-based approach could not distinguish closely related genomes.
165 Targeted methods, such as genome enrichment or isolation of specific microbial taxa, could
166 expand the list of taxa showing patterns of codiversification.

167 In conclusion, we have identified common members of the human gut microbiota that
168 show evidence of codiversification with humans. Our results suggest that these species have
169 transferred within related individuals living in proximity for thousands of generations. These

170 findings extend previous observations of cospeciatiated bacteria with humans from our hominid
171 ancestors: they further codiversified within human populations. Gut microbiota that codiversified
172 with early modern humans as they migrated to populate the world were exposed to new
173 environments and may have participated in the processes of local adaptation in humans (35).
174 Because some of the codiversified taxa have been linked to a variety of diseases in humans (6,
175 39), functional differences among codiversified strains may contribute to discrepancies between
176 studies addressing the role of microbiome in physiology in different populations (12). The list of
177 gut microbial taxa provided here should motivate future studies to investigate the roles of
178 codiversified strains in human health and host-microbial coevolution.

179

References and Notes

- 180 1. N. A. Moran, D. B. Sloan, The hologenome concept: helpful or hollow? *PLoS Biol.* **13**, 1–
181 10 (2015).
- 182 2. M. Groussin, F. Mazel, J. G. Sanders, C. S. Smillie, S. Lavergne, W. Thuiller, E. J. Alm,
183 Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nat.*
184 *Commun.* **8**, 14319 (2017).
- 185 3. N. D. Youngblut, G. H. Reischer, W. Walters, N. Schuster, C. Walzer, G. Stalder, R. E.
186 Ley, A. H. Farnleitner, Host diet and evolutionary history explain different aspects of gut
187 microbiome diversity among vertebrate clades. *Nat. Commun.* **10**, 2200 (2019).
- 188 4. N. D. Youngblut, G. H. Reischer, S. Dauser, C. Walzer, G. Stalder, A. H. Farnleitner, R. E.
189 Ley, Strong influence of vertebrate host phylogeny on gut archaeal diversity. *bioRxiv*
190 (2020), p. 2020.11.10.376293.
- 191 5. A. H. Moeller, A. Caro-Quintero, D. Mjungu, A. V. Georgiev, E. V. Lonsdorf, M. N.
192 Muller, A. E. Pusey, M. Peeters, B. H. Hahn, H. Ochman, Cospeciation of gut microbiota
193 with hominids. *Science.* **353**, 380–382 (2016).
- 194 6. J. C. Atherton, M. J. Blaser, Coadaptation of *Helicobacter pylori* and humans: ancient
195 history, modern implications. *J. Clin. Invest.* **119**, 2475–2487 (2009).
- 196 7. A. H. Nishida, H. Ochman, Captivity and the co-diversification of great ape microbiomes.
197 *Nat. Commun.* **12**, 5632 (2021).
- 198 8. P. Ferretti, E. Pasolli, A. Tett, F. Asnicar, V. Gorfer, S. Fedi, F. Armanini, D. T. Truong, S.
199 Manara, M. Zolfo, F. Beghini, R. Bertorelli, V. De Sanctis, I. Bariletti, R. Canto, R.
200 Clementi, M. Cologna, T. Crifò, G. Cusumano, S. Gottardi, C. Innamorati, C. Masè, D.
201 Postai, D. Savoi, S. Duranti, G. A. Lugli, L. Mancabelli, F. Turroni, C. Ferrario, C. Milani,
202 M. Mangifesta, R. Anzalone, A. Viappiani, M. Yassour, H. Vlamakis, R. Xavier, C. M.
203 Collado, O. Koren, S. Tateo, M. Soffiati, A. Pedrotti, M. Ventura, C. Huttenhower, P. Bork,
204 N. Segata, Mother-to-infant microbial transmission from different body sites shapes the
205 developing infant gut microbiome. *Cell Host Microbe.* **24**, 133–145.e5 (2018).
- 206 9. F. Asnicar, S. Manara, M. Zolfo, D. T. Truong, M. Scholz, F. Armanini, P. Ferretti, V.
207 Gorfer, A. Pedrotti, A. Tett, N. Segata, Studying vertical microbiome transmission from
208 mothers to infants by strain-level metagenomic profiling. *mSystems.* **2** (2017),
209 doi:10.1128/mSystems.00164-16.
- 210 10. D. T. Truong, A. Tett, E. Pasolli, C. Huttenhower, N. Segata, Microbial strain-level
211 population structure and genetic diversity from metagenomes. *Genome Res.* **27**, 626–638
212 (2017).
- 213 11. A. Tett, K. D. Huang, F. Asnicar, H. Fehlner-Peach, E. Pasolli, N. Karcher, F. Armanini, P.
214 Manghi, K. Bonham, M. Zolfo, F. De Filippis, C. Magnabosco, R. Bonneau, J. Lusingu, J.

- 215 Amuasi, K. Reinhard, T. Rattei, F. Boulund, L. Engstrand, A. Zink, M. C. Collado, D. R.
216 Littman, D. Eibach, D. Ercolini, O. Rota-Stabelli, C. Huttenhower, F. Maixner, N. Segata,
217 The *Prevotella copri* complex comprises four distinct clades underrepresented in
218 westernized populations. *Cell Host Microbe*. **26**, 666–679.e7 (2019).
- 219 12. P. I. Costea, L. P. Coelho, S. Sunagawa, R. Munch, J. Huerta-Cepas, K. Forslund, F.
220 Hildebrand, A. Kushugulova, G. Zeller, P. Bork, Subspecies in the global human gut
221 microbiome. *Mol. Syst. Biol.* **13**, 960 (2017).
- 222 13. N. Karcher, E. Pasolli, F. Asnicar, K. D. Huang, A. Tett, S. Manara, F. Armanini, D. Bain,
223 S. H. Duncan, P. Louis, M. Zolfo, P. Manghi, M. Valles-Colomer, R. Raffaetà, O. Rota-
224 Stabelli, M. C. Collado, G. Zeller, D. Falush, F. Maixner, A. W. Walker, C. Huttenhower,
225 N. Segata, Analysis of 1321 *Eubacterium rectale* genomes from metagenomes uncovers
226 complex phylogeographic population structure and subspecies functional adaptations.
227 *Genome Biol.* **21**, 138 (2020).
- 228 14. A. Lokmer, A. Cian, A. Froment, N. Gantois, E. Viscogliosi, M. Chabé, L. Ségurel, Use of
229 shotgun metagenomics for the identification of protozoa in the gut microbiota of healthy
230 individuals from worldwide populations with various industrialization levels. *PLoS One*. **14**,
231 e0211139 (2019).
- 232 15. G. Even, A. Lokmer, J. Rodrigues, C. Audebert, E. Viscogliosi, L. Ségurel, M. Chabé,
233 Changes in the human gut microbiota associated with colonization by *Blastocystis* sp. and
234 *Entamoeba* spp. in non-industrialized populations. *Front. Cell. Infect. Microbiol.* **11**,
235 533528 (2021).
- 236 16. H. Xie, R. Guo, H. Zhong, Q. Feng, Z. Lan, B. Qin, K. J. Ward, M. A. Jackson, Y. Xia, X.
237 Chen, B. Chen, H. Xia, C. Xu, F. Li, X. Xu, J. Y. Al-Aama, H. Yang, J. Wang, K.
238 Kristiansen, J. Wang, C. J. Steves, J. T. Bell, J. Li, T. D. Spector, H. Jia, Shotgun
239 metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut
240 microbiome. *Cell Syst.* **3**, 572–584.e3 (2016).
- 241 17. F. Asnicar, A. M. Thomas, F. Beghini, C. Mengoni, S. Manara, P. Manghi, Q. Zhu, M.
242 Bolzan, F. Cumbo, U. May, J. G. Sanders, M. Zolfo, E. Kopylova, E. Pasolli, R. Knight, S.
243 Mirarab, C. Huttenhower, N. Segata, Precise phylogenetic analysis of microbial isolates and
244 genomes from metagenomes using PhyloPhlAn 3.0. *Nat. Commun.* **11**, 2500 (2020).
- 245 18. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other
246 things): phytools: R package. *Methods Ecol. Evol.* **3**, 217–223 (2012).
- 247 19. J. A. Balbuena, R. Míguez-Lozano, I. Blasco-Costa, PACo: a novel procrustes application
248 to cophylogenetic analysis. *PLoS One*. **8**, e61048 (2013).
- 249 20. P. Legendre, Y. Desdevises, E. Bazin, A statistical test for host–parasite coevolution. *Syst.*
250 *Biol.* **51**, 217–234 (2002).
- 251 21. Z. Tamanai-Shacoori, I. Smida, L. Bousarghin, O. Loreal, V. Meuric, S. B. Fong, M.
252 Bonnaure-Mallet, A. Jolivet-Gougeon, *Roseburia* spp.: a marker of health? *Future*

- 253 *Microbiol.* **12**, 157–170 (2017).
- 254 22. M. Rajilić-Stojanović, W. M. de Vos, The first 1000 cultured species of the human
255 gastrointestinal microbiota. *FEMS Microbiol. Rev.* **38**, 996–1047 (2014).
- 256 23. J. C. Ezeji, D. K. Sarikonda, A. Hopperton, H. L. Erkkila, D. E. Cohen, S. P. Martinez, F.
257 Cominelli, T. Kuwahara, A. E. K. Dichosa, C. E. Good, M. R. Jacobs, M. Khoretchenko, A.
258 Veloo, A. Rodriguez-Palacios, Parabacteroides distasonis: intriguing aerotolerant gut
259 anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in
260 human health. *Gut Microbes.* **13**, 1922241 (2021).
- 261 24. N. A. Moran, G. M. Bennett, The Tiniest Tiny Genomes. *Annu. Rev. Microbiol.* **68**, 195–
262 215 (2014).
- 263 25. M. C. Wibowo, Z. Yang, M. Borry, A. Hübner, K. D. Huang, B. T. Tierney, S. Zimmerman,
264 F. Barajas-Olmos, C. Contreras-Cubas, H. García-Ortiz, A. Martínez-Hernández, J. M.
265 Luber, P. Kirstahler, T. Blohm, F. E. Smiley, R. Arnold, S. A. Ballal, S. J. Pamp, J. Russ, F.
266 Maixner, O. Rota-Stabelli, N. Segata, K. Reinhard, L. Orozco, C. Warinner, M. Snow, S.
267 LeBlanc, A. D. Kostic, Reconstruction of ancient microbial genomes from the human gut.
268 *Nature.* **594**, 234–239 (2021).
- 269 26. T. Goebel, M. R. Waters, D. H. O’Rourke, The late Pleistocene dispersal of modern humans
270 in the Americas. *Science.* **319**, 1497–1502 (2008).
- 271 27. P. Duda, Jan Zrzavý, Human population history revealed by a supertree approach. *Sci. Rep.*
272 **6**, 29890 (2016).
- 273 28. M. R. Olm, A. Crits-Christoph, K. Bouma-Gregson, B. A. Firek, M. J. Morowitz, J. F.
274 Banfield, inStrain profiles population microdiversity from metagenomic data and sensitively
275 detects shared microbial strains. *Nat. Biotechnol.* (2021), doi:10.1038/s41587-020-00797-0.
- 276 29. H. Enay, R. E. Ley, SynTracker: a synteny based tool for tracking microbial strains. *bioRxiv*
277 (2021), p. 2021.10.06.463341.
- 278 30. J. E. Powell, V. G. Martinson, K. Urban-Mead, N. A. Moran, Routes of acquisition of the
279 gut microbiota of the honey bee *Apis mellifera*. *Appl. Environ. Microbiol.* **80**, 7378–7387
280 (2014).
- 281 31. A. H. Moeller, T. A. Suzuki, M. Phifer-Rixey, M. W. Nachman, Transmission modes of the
282 mammalian gut microbiota. *Science.* **362**, 453–457 (2018).
- 283 32. J. Tung, L. B. Barreiro, M. B. Burns, J.-C. Grenier, J. Lynch, L. E. Grieneisen, J. Altmann,
284 S. C. Alberts, R. Blekhnman, E. A. Archie, Social networks predict gut microbiome
285 composition in wild baboons. *Elife.* **4** (2015), doi:10.7554/eLife.05224.
- 286 33. A. H. A. H. Moeller, S. Foerster, M. L. Wilson, A. E. Pusey, B. H. Hahn, H. Ochman,
287 Social behavior shapes the chimpanzee pan-microbiome. *Science Advances.* **2**, e1500997–
288 e1500997 (2016).

- 289 34. N. A. Moran, H. Ochman, T. J. Hammer, Evolutionary and ecological consequences of gut
290 microbial communities. *Annu. Rev. Ecol. Evol. Syst.* **50**, 451–475 (2019).
- 291 35. T. A. Suzuki, R. E. Ley, The role of the microbiota in human genetic adaptation. *Science*.
292 **370** (2020), doi:10.1126/science.aaz6827.
- 293 36. J. K. Goodrich, E. R. Davenport, A. G. Clark, R. E. Ley, The relationship between the
294 human genome and microbiome comes into view. *Annu. Rev. Genet.* **51**, 413–433 (2017).
- 295 37. A. Kurilshikov, C. Medina-Gomez, R. Bacigalupe, D. Radjabzadeh, J. Wang, A. Demirkan,
296 C. I. Le Roy, J. A. Raygoza Garay, C. T. Finnicum, X. Liu, D. V. Zhernakova, M. J.
297 Bonder, T. H. Hansen, F. Frost, M. C. Rühlemann, W. Turpin, J.-Y. Moon, H.-N. Kim, K.
298 Lüll, E. Barkan, S. A. Shah, M. Fornage, J. Szopinska-Tokov, Z. D. Wallen, D. Borisevich,
299 L. Agreus, A. Andreasson, C. Bang, L. Bedrani, J. T. Bell, H. Bisgaard, M. Boehnke, D. I.
300 Boomsma, R. D. Burk, A. Claringbould, K. Croitoru, G. E. Davies, C. M. van Duijn, L.
301 Duijts, G. Falony, J. Fu, A. van der Graaf, T. Hansen, G. Homuth, D. A. Hughes, R. G.
302 Ijzerman, M. A. Jackson, V. W. V. Jaddoe, M. Joossens, T. Jørgensen, D. Keszthelyi, R.
303 Knight, M. Laakso, M. Laudes, L. J. Launer, W. Lieb, A. J. Lusic, A. A. M. Masclee, H. A.
304 Moll, Z. Mujagic, Q. Qibin, D. Rothschild, H. Shin, S. J. Sørensen, C. J. Steves, J. Thorsen,
305 N. J. Timpson, R. Y. Tito, S. Vieira-Silva, U. Völker, H. Völzke, U. Vösa, K. H. Wade, S.
306 Walter, K. Watanabe, S. Weiss, F. U. Weiss, O. Weissbrod, H.-J. Westra, G. Willemsen, H.
307 Payami, D. M. A. E. Jonkers, A. Arias Vasquez, E. J. C. de Geus, K. A. Meyer, J.
308 Stockholm, E. Segal, E. Org, C. Wijmenga, H.-L. Kim, R. C. Kaplan, T. D. Spector, A. G.
309 Uitterlinden, F. Rivadeneira, A. Franke, M. M. Lerch, L. Franke, S. Sanna, M. D’Amato, O.
310 Pedersen, A. D. Paterson, R. Kraaij, J. Raes, A. Zhernakova, Large-scale association
311 analyses identify host factors influencing human gut microbiome composition. *Nat. Genet.*
312 **53**, 156–165 (2021).
- 313 38. M. Groussin, F. Mazel, E. J. Alm, Co-evolution and co-speciation of host-gut bacteria
314 systems. *Cell Host Microbe*. **28**, 12–22 (2020).
- 315 39. T. L. Cover, M. J. Blaser, *Helicobacter pylori* in health and disease. *Gastroenterology*. **136**,
316 1863–1873 (2009).
- 317
- 318
- 319
- 320
- 321
- 322
- 323

324 **Acknowledgements**

325 We thank Nguyen Thu Ha, Anne Pflaiderer, Emily Cosgrove, Andrew Clark, Aleksander Kostic,
326 Marla Taylor, Native American tribes (Mike Bremer, Joseph Aguilar, Juana Charlie, Reyllynne
327 Williams, Barnaby Lewis, Shane Anton, Angela Garcia-Lewis, and Bruce Bernstein), and
328 members of the Department of Microbiome Science. **Funding:** This work was supported by the
329 Max Planck Society. TwinsUK is funded by the Wellcome Trust, Medical Research Council,
330 European Union, Chronic Disease Research Foundation (CDRF), Zoe Global Ltd., and the
331 National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility
332 and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in
333 partnership with King's College London.

334

335 **Competing interests:** The authors declare no conflicts of interests.

336

337 **Data and material availability:** The raw sequence data and MAGs are available from the
338 European Nucleotide Archive under the study accession number PRJEB46788. All sample
339 metadata used in this study are provided in Supplementary information. The code used in data
340 analysis is outlined on GitHub at <https://github.com/leylabmpi/codiversification>. The raw
341 metagenome sequence data are available from the European Nucleotide Archive under the study
342 accession numbers PRJEB47532, PRJEB47543, PRJEB40256, PRJEB9584, PRJEB32731,
343 PRJEB27005, PRJEB30834, and PRJEB46788.