Distinct Genetic and Functional Traits of Human Intestinal Prevotella copri Strains Are Associated with Different Habitual Diets

Distinct genetic and functional traits of human intestinal *Prevotella copri* strains are associated with different habitual diets

Francesca De Filippis[1,2], Edoardo Pasolli[1,3], Adrian Tett[3], Sonia Tarallo[4], Alessio Naccarati[4], Maria De Angelis[5], Erasmo Neviani[6], Luca Cocolin[7], Marco Gobbetti[8], Nicola Segata[3], Danilo Ercolini[1,2,*†]

[1]Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy
[2]Task Force on Microbiome Studies, University of Naples Federico II, Italy
[3]Centre for Integrative Biology, University of Trento, Italy
[4]Italian Institute for Genomic Medicine, Torino, Italy
[5]Department of Soil, Plant and Food Science, University of Bari Aldo Moro, Italy
[6]Department of Food Science, University of Parma, Italy
[7]Department of Agricultural, Forest and Food Sciences, University of Torino, Italy
[8]Faculty of Science and Technology, Free University of Bozen, Italy
[†] Lead contact

*Corresponding author
Department of Agricultural Sciences, Division of Microbiology, University of Naples Federico II, Via Università 100, 80055 Portici, Italy
Phone: +39 081 2539449. Fax: +39 081 2539407.
E-mail: ercolini@unina.it.

Summary

The role of intestinal *Prevotella* species in human health is controversial, since it has been associated with either potential positive or negative effects for the host. Strain-level diversity might be the cause of such discrepancies in genus and species associations with health and disease. We dissected the gut metagenomes of Italian subjects with different dietary habits investigating the possible presence of distinct *Prevotella copri* strains. Fibre-rich diets were linked to *P. copri* types with enhanced potential for carbohydrates catabolism, while *P. copri* strains associated with an omnivore diet had higher prevalence of *leuB* gene, involved in branched-chain amino acid biosynthesis, a known risk factor in the development of glucose intolerance and type-2 diabetes. The newly profiled *P. copri* pangenomes were compared to those of previously characterized cohorts, providing evidence of distinct gene repertoires characterizing different populations, with drug metabolism and complex carbohydrates degradation significantly associated with Western and non-Western individuals, respectively. Strain-level diversity of *P. copri* in our gut microbiome can be shaped by diet and should be taken into account when considering host-microbe associations.

Introduction

The gut microbiome plays a key role in human well-being, performing important metabolic functions, such as the biosynthesis of vitamins or the breakdown of indigestible compounds, and interacting with the host through the production of beneficial or detrimental metabolites (De Filippis et al., 2018; Derrien and Veiga, 2016). Indeed, an imbalance among the microbial organisms inhabiting our gut (commonly referred to as dysbiosis) has been linked with the pathogenesis of both intestinal and extra-intestinal diseases, including neurological disorders, obesity, atherosclerosis, inflammatory bowel disease and cancer (Marchesi et al., 2016; Sharon et al., 2016; Blum, 2017).

In healthy adults, the gut microbiome may be influenced by many extrinsic factors, among which diet may be considered one of the most important (Zhernakova et al., 2016; Falony et al., 2016; Sonnenburg and Bäckhed, 2016). Habitual diet shapes the species-level composition and abundances of the gut microbiome and several researches highlighted that a dietary "Westernization", characterized by higher consumption of high-fat and protein products at the expense of foods rich in fibre could have caused a loss of microbial diversity, with ultimate repercussions on human health (Segata, 2015; Sonnenburg and Sonnenburg, 2014).

The gut microbiome is dominated by two major bacterial phyla: Firmicutes and Bacteroidetes (Lozupone et al., 2012; Falony et al., 2016; Zhernakova et al., 2016). Among the Bacteroidetes, two genera prevail — *Bacteroides* and *Prevotella* – and while *Bacteroides* species are highly prevalent, they are usually dominated by *Prevotella* when this genus is present (Falony et al., 2016; Arumugam et al., 2011). Higher abundance of *Prevotella* was traditionally associated with the consumption of an agrarian-type diet, rich in fruit and vegetables, while the abundance of *Bacteroides* is usually associated with high-fat and protein rich diets (David et al., 2014; Wu et al., 2011).

In the past decades, metagenomics deeply increased our knowledge on the role of the gut microbiome and how it is influenced by environmental factors. Nevertheless, our knowledge often relies on a genus or species-level taxonomic assignment that, although useful, may not be sufficient for a comprehensive understanding of the complex inter-connections between gut microbiome and human

health. Indeed, each microbial genus in the gut is represented by several species and strains, that may harbour substantial differences in their genomes. Such inter- and intra-species variation endows each species and even each strain with potentially distinct functional capacities (Faith et al., 2015, Greenblum et al., 2015; Lloyd-Price et al., 2017; Schloissnig et al., 2013; Scholz et al., 2016; Wu et al., 2017; Zhang and Zhao, 2016).

The role of *Prevotella* spp. in the human gut microbiome is controversial and deserves further exploration (Ley, 2016; Cani, 2018). Its usual connection with agrarian and vegetables-rich diets would suggest that *Prevotella*, as a fibre-degrader, is an indicator of a microbiome associated with a healthy status. Indeed, it was recognized as positively associated with the production of health-promoting compounds such as short-chain fatty acids (De Filippis et al., 2016a; De Filippo et al., 2010), with an improved glucose metabolism (Kovachenka-Datchary et al., 2015; De Vadder et al., 2016) or an overall anti-inflammatory effect (De Angelis et al., 2015; Vitaglione et al., 2015). Nevertheless, some studies also highlighted an association of *P. copri* with inflammatory conditions (Lozupone et al., 2014; Maeda et al., 2016; Scher et al., 2013), as well as with insulin resistance and glucose intolerance (Pedersen et al., 2016). Consistently, it has been recently brought to our attention that *P. copri* represents one of the clearest cases of dissimilar associations with either health or disease (Cani, 2018), and such behaviour can be most likely explained by a strain-level diversity.

Recent findings suggested that dietary habits (omnivore vs. vegetarian/vegan diets) may select for potentially different *Prevotella copri* strains in the gut microbiome (De Filippis et al., 2016b). Therefore, the current associations of *Prevotella* with the host may represent an oversimplification that does not consider the wide diversity possibly existing among different *P. copri* strains. The vastly different genomic repertoires of *P. copri* strains may help to explain some of the differences observed across individuals in the metabolic responses to diet (Ley, 2016; Truong et al., 2017; Cani, 2018). Any attempt to assess the influence of the gut microbiome on human health or disease must acknowledge that many relevant functions may well be strain-specific and therefore strain-level dissection of metagenomics data can be crucial to demonstrate a causative role of the gut microbiome

in the balance between health and disease. In particular, the response to different dietary regimens or to nutritional interventions may be strain-dependent, and therefore unpredictable in the current scenario of genus- or species-scale resolution, complicating the possibility of microbiome-targeted dietary interventions (De Filippis et al., 2018; Derrien et al., 2016; Zmora et al., 2016).

In order to study more in-depth the association between diet and strain-level determinants in the microbiome, we sequenced the gut metagenome of healthy Italian adults with different habitual diets and carried out a strain-level analysis of *P. copri* to explore the possible diet-driven selection of specific strains and functions. Moreover, we compared the overall *P. copri* functional potential of our Italian subjects with previously studied non-Westernized cohorts.

Results

Strain-level differences of *P. copri* are associated to habitual diet

We analyzed the gut metagenome of 97 Italian omnivores (O, n=23), vegetarians (VG, n= 38) and vegans (V, n=36). The relative abundance of *P. copri* was high enough in 47 samples for a strain-level analysis (9 omnivores, 22 vegetarians, 16 vegans). The average age of the 47 subjects was 40.8 ± 8.9, 42.1 ± 7.6 and 40.1 ± 12.3 years and Body Mass Index (BMI) was 24.5 ± 4.5, 21.9 ± 3.0, 21.8 ± 3.7 Kg/m^2 for O, VG and V, respectively. No significant difference in age and BMI was detected by pair-wise Wilcoxon tests ($p>0.05$). Fifty-three subjects were part of a larger cohort previously characterized (De Filippis et al., 2016a), while 44 subjects belonged to a newly recruited cohort. Dietary habits and main demographics are reported in Table S1. At the species level, the abundance of *P. copri* in our metagenomes ranged from 0 to 83.2% (Figure S1). The abundance of *P. copri* was not significantly associated with diet type (omnivorous, vegan or vegetarian), as determined by Multivariate Analysis of Variance (MANOVA) based on Bray Curtis' dissimilarity matrix. To test the hypothesis that strain-level structures could be associated with diet, we characterized the strain-specific *P. copri* functional potential by pangenome profiling using PanPhlAn and grouping orthologous genes into KEGG functional categories. Principal Coordinates Analysis (PCoA) clearly

separated omnivore from non-omnivore (V/VG) subjects (Figure 1A) based on the *P. copri* gene repertoire. Moreover, by further distinguishing V and VG subjects, we observed a gradient of separation from vegans to omnivores (Figure 1B). Thirty-six and eight pan-genes occurred differentially in *P. copri* pangenomes of V and VG compared to O, respectively ($p<0.05$; Table 1). Interestingly, V-associated *P. copri* strains showed higher prevalence of genes involved in complex carbohydrates break-down (Table 1). Vegans showed higher prevalence of genes identified as acetylxylan esterase, pectate lyase, alpha-L-fucosidase, 1,4 beta-xylanase, phosphoenolpyruvate carboxykinase and several carbohydrate transporters (*susD* family). As confirmation, we also used the CAZy database (Lombard et al., 2014; http://www.cazy.org/) for the identification of *P. copri* pan-genes (see STAR Methods). Glycoside hydrolase (GH) and carbohydrate esterase (CE) families were enriched in V compared to O and VG, although only CE were significantly enriched in V compared to both O and VG ($p<0.05$; Figure S2). In particular, CE7 and CE8, including acetyl xylan esterase and pectin methyl esterase, showed higher prevalence in vegans. GH5, GH95 and GH127, containing enzymes involved in complex polysaccharides break-down, were enriched in vegans, while GH2, including β-galactosidase, prevailed in omnivores (Table S2). In addition, genes involved in sulphur compounds metabolism (cystathionine beta-lyase, O-acetylhomoserine thiol-lyase) were enriched in O compared to V, as well as 3-isopropylmalate dehydrogenase (*leuB*, EC 1.1.1.85), involved in branched-chain amino acids (BCAA) biosynthesis. All omnivores, 67% of VG and 18% of V harboured the *leuB* gene in *P. copri* pangenome. Interestingly, when we divided subjects for presence/absence of *leuB* in the *P. copri* pangenome, we found significantly lower urinary BCAA levels in V/VG individuals not harbouring the *P. copri leuB* gene ($p<0.05$; Figure S3).

In order to confirm the results obtained by reference-based computational profiling, we assembled the metagenomes into contigs and extracted those belonging to *P. copri* (see STAR Methods). Core genes identified in the assemblies were aligned and used to build a phylogenetic tree. Although only part of the samples had > 2.5 Mb total alignment to *P. copri* genome (due to assembly/coverage

limitations), results still showed a sharp separation of *P. copri* strains present in O and V, while VG subjects were separated in the two groups (Figure 2).

Comparison of *P. copri* functional potential between Western and non-Western cohorts

We also compared the *P. copri* gene repertoire of our cohort with Western and non-Western populations from previously published studies (Rampelli et al., 2015; Le Chatelier et al., 2013; Obregon-Tito et al., 2015; Yatsunenko et al., 2012). The functional potential of *P. copri* strains present in Western and non-Western populations was different and we could identify two main clusters, separated by subject-origin (Figure 3A). Interestingly, the few American and Italian controls from Obregon-Tito et al. (2015), Yatsunenko et al. (2012) and Rampelli et al. (2015) clustered together with Italians from this study and Danes from Le Chatelier et al. (2013; Figure 3B). In particular, 1368 genes differentiated Western and non-Western subjects (Table S3). Among them, several genes encoding for *SusC* and *SusD* transporters, involved in starch binding, xylanases, pectin esterases, β-glucosidases and alpha-amylases, involved in carbohydrates catabolism were enriched in non-Western subjects. Conversely, proteases and genes related to the biosynthesis of several vitamins of the B group (B1, B2, B5, B6) and folate prevailed in Western individuals. Finally, the *P. copri* pangenome of Western subjects showed higher prevalence of genes encoding for *TolC* and MATE family (Multi-drug efflux transporters) proteins, responsible for antibiotics and toxic compounds export from the cell, and *DedA* family proteins, membrane proteins possibly involved in drug resistance.

Discussion

Our understanding of gut microbial communities is usually limited to a genus- or species-level description that, although useful, is still a simplification in a complex *consortium* of strains (Lloyd-Price et al., 2017; Truong et al., 2017). Indeed, inter-individual differences in the type and number of strains present in the healthy human microbiome exist (Faith et al., 2015; Hansen et al., 2011;

169 Schloissnig et al., 2013; Scholz et al., 2016), with possibly different genomic potentials. Each strain
170 may have very specific mechanisms through which it can affect health or respond to dietary patterns.
171 Accordingly, strain-specific microbial determinants were found in virulence (Solheim et al., 2009;
172 Pierce and Bernstein, 2016), drug resistance (Gill et al., 2005) or catabolism (Haiser et al., 2013),
173 nutrient utilization (Siezen et al., 2010; Wexler, 2007; Wu et al., 2017), adhesion to the gut epithelium
174 (Hansen et al., 2011) and even induction of obesity (Fei and Zhao, 2013), highlighting the importance
175 of a strain-level dissection to address the functional role of the gut microbiome.
176 *Prevotella* defines a clear subtype of the human gut microbiome (Arumugam et al., 2011) and it is
177 one of its most abundant members (Falony et al., 2016). However, its role in relation to human health
178 is still unclear (Ley, 2016; Cani, 2018), being contrastingly associated either with health-promoting
179 (Kovachenka-Datchary et al., 2015; De Vadder et al., 2016) or detrimental effects for the host
180 (Lozupone et al., 2014; Maeda et al., 2016; Pedersen et al., 2016; Scher et al., 2013). Thus, a further
181 characterization of *P. copri* with strain-level resolution is required to understand if differences in the
182 genomic potentials exist, how they are linked to human health or disease, and whether different
183 dietary styles may select specific *P. copri* strains. Here, we investigated the pangenome of *P. copri*
184 strains from gut metagenome of subjects with different dietary habits.
185 We found a distinct clustering of the subjects based on the samples-specific *P. copri* gene repertoire
186 and driven by diet, suggesting that different dietary habits may select for specific *P. copri* strains, as
187 we had previously speculated based on the results of the oligotyping of 16S rRNA gene sequences
188 (De Filippis et al., 2016b). Omnivores with high abundance of *P. copri* oligotypes previously linked
189 with an animal-based diet (P5, P12, P16; see De Filippis et al., 2016b) consistently harbor a *P. copri*
190 functional potential strongly different from V, supporting the value of oligotyping in recognizing sub-
191 genus diversity patterns in the gut microbiome (Eren et al., 2013). According to our results, a diet
192 richer in fibre from fruit, vegetables and legumes may select for *P. copri* strains with higher potential
193 for complex carbohydrates degradation. On the contrary, the *P. copri* strains associated with
194 omnivore diet showed higher prevalence of genes related to BCAA biosynthesis and subjects

harboring these genes in their microbiome consistently showed higher BCAA urinary levels. Accordingly, Pedersen et al. (2016) found high plasma concentration of BCAA in subjects showing higher levels of *P. copri* in their gut microbiome and associated *P. copri* abundance with the development of insulin resistance, a forerunner of type-2 diabetes. Moreover, gut metagenomes of Western individuals were found to be enriched in genes related to BCAA biosynthesis compared to non-westernized populations (Rampelli et al., 2015). Our results suggest that specific *P. copri* strains, possibly selected by diet, may contribute to BCAA biosynthesis. We also compared *P. copri* strains of our cohort with those found in non-Westernized populations, still adhering to a hunting-gathering (Hadza, Amerindians) or agriculturalist (Tunapuco, Malawian) subsistence pattern (closely resembling those of our ancestors), consuming primarily a plant-based diet, heavily relying on fibrous tubers and vegetables (Rampelli et al., 2015; Obregon-Tito et al., 2015; Yatsunenko et al., 2012). In addition, we included in the analysis Western healthy Danes from a previously studied cohort (Le Chatelier et al., 2013). The totally different dietary habits and lifestyles of the Western and non-Western cohorts probably drive a consistent selection of different *P. copri*. Indeed, *P. copri* strains from non-Western populations are well-adapted to rescue energy from a wide range of complex plant polysaccharides, as demonstrated by the enrichment in genes encoding for enzymes acting on starch, xylans, pectins and polygalacturonans. On the contrary, the *P. copri* pangenome of Western subjects was enriched in proteases, possibly reflecting a diet richer in proteins, as well as in genes involved in the biosynthesis of vitamin B and folate. Folate are usually rich in green vegetables, while vitamins B1 and B2 are present in a wide range of food products, including cereal bran. Therefore, these are consistently reduced in refined cereals, typically consumed in Western countries. Consistently, a recent study reported loss of *P. copri* genes associated with cellulose, β-mannans and xyloglucans degradation in rural Thai (Vangay et al., 2018). Our results support the existence of a strain-level selection that probably took place during our evolutionary history, in response to the different diets. Nevertheless, we have to point out that other factors besides diet might have their influence in shaping such strain diversity (e.g. different lifestyle, host genetics). Interestingly, Italian V and VG clustered

with Western populations, suggesting that a Western plant-based diet is still not effective in establishing a *P. copri* strains *consortium* typical of a traditional agrarian diet and supporting the existence of a geographically-based distribution of different strain patterns (Truong et al., 2017). Finally, *P. copri* gene repertoire in Western individuals was enriched in genes associated with drug metabolism and antibiotic resistance, pointing to the dramatic impact of the widespread contact with xenobiotics in the Western world on our intestinal microbial counterpart.

Our results provide evidence of a strain-level diversity in our gut microbiome, with possible evolutionary implications, that can be driven at least by diet beyond other environmental factors. Recent studies suggested that gut microbiome features may be implicated in the different response observed to dietary interventions or drug therapies (De Filippis et al., 2018; Zmora et al., 2016; Derrien et al., 2016). Consequently, inter-personal differences at strain-level exist and should be considered when investigating the role of certain microbial species in disease development. The strain-level biodiversity will have to be also considered in the near future in microbiome-targeted nutritional interventions developed for the prevention or treatment of diseases.

Acknowledgements

This study was supported by the Italian Ministry of University and Research (MIUR) program PRIN 2010-2011. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

S.T. was supported by Fondazione Umberto Veronesi “Post-doctoral fellowship Years 2017 and 2018”.

Author contributions

Conceptualization, D.E. and N.S.; Formal Analysis, F.D.F., E.P. and A.T.; Investigation, F.D.F., E.P. and A.T.; Resources, M.D.A., A.N., S.T., E.N., L.C., N.S., M.G. and D.E.; Writing–Original Draft, F.D.F. and D.E.; Writing–Review & Editing, F.D.F., N.S., E.P., A.T. and D.E.; Funding Acquisition, N.S. and D.E.

Declaration of interests

The authors declare no competing interests

References

1. Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, D., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M., et al. (2011). Enterotypes of the human gut microbiome. Nature 473(7346), 174-180.
2. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19(5), 455-477.
3. Blum, H.E. (2017). The human microbiome. Adv. Med. Sci. 62, 414-420.
4. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L. (2008). BLAST+: architecture and applications. BMC Bioinformatics 10, 421.
5. Cani, P.D. (2018). Human gut microbiomes: hopes, threats and promises. Gut, in press. doi:10.1136/gutjnl-2018-316723.
6. Cox, M.P., Peterson, D.A., Biggs, P.J. (2010). SolexaQA: at-a-glance quality assessment of Illumina second generation sequencing data. BMC Bioinformatics 11, 485.
7. David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature 505(7484), 559-563.
8. De Angelis, M., Montemurno, E., Vannini, L., Cosola, C., Cavallo, N., Gozzi, G., Maranzano, V., Di Cagno, R., Gobbetti, M., Gesualdo, L. (2015). Effect of whole-grain barley on the human fecal microbiota and metabolome. Appl. Environ. Microbiol. 81(22), 7945-7956.

9. De Filippis, F., Pellegrini, N., Laghi, L., Gobbetti, M., Ercolini, D. (2016b). Unusual sub-genus associations of faecal *Prevotella* and *Bacteroides* with specific dietary patterns. Microbiome 4(1), 57.

10. De Filippis, F., Pellegrini, N., Vannini, L., Jeffery, I.B., La Storia, A., Laghi, L., Serrazanetti, D.I., Di Cagno, R., Ferrocino, I., Lazzi, C., et al. (2016a). High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. Gut 65(11), 1812-1821.

11. De Filippis, F., Vitaglione, P., Cuomo, R., Berni Canani, R., Ercolini, D. (2018). Dietary interventions to modulate the gut microbiome: how far away are we from precision medicine. Inflamm. Bowel Dis. 24(10), 2142-2154.

12. De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc. Natl. Acad. Sci. U.S.A. 107(33), 14691-14696.

13. De Vadder, F., Kovatcheva-Datchary, P., Zitoun, C., Duchampt, A., Bäckhed, F., Mithieux, G. (2016). Microbiota-produced succinate improves glucose homeostasis via intestinal gluconeogenesis. Cell Metab. 24(1), 151-157.

14. Derrien, M., Veiga, P. (2016). Rethinking diet to aid human–microbe symbiosis. Trends Microbiol. 25(2), 100-112.

15. Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G., Sogin, M.L. (2013). Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol. Evol. 4(12).

16. Faith, J.J., Colombel, J.F., Gordon, J.I. (2015). Identifying strains that contribute to complex diseases through the study of microbial inheritance. Proc. Natl. Acad. Sci. U.S.A. 112(3), 633-640

17. Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurlishikov, A., Bonder, M.J., Valles-Colomer, M., Vandeputte, D., et al. (2016). Population-level analysis of gut microbiome variation. Science 352 (6285), 560-564.

18. Fei, N., Zhao, L. (2013). An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. ISME J. 7(4), 880–884.

19. Gill, S.R., Fouts, D.E., Archer, G.L., Mongodin, E.F., Deboy, R.T., Ravel, J., Paulsen, I.T., Kolonay, J.F., Brinkac, L., Beanan, M., et al. (2005). Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. J. Bacteriol. 187, 2426–2438.

20. Greenblum, S., Carr, R., Borenstein, E. (2015). Extensive strain-level copy-number variation across human gut microbiome species. Cell 160, 583–594.

21. Haiser, H.J., Gootenberg, D.B., Chatman, K., Sirasani, G., Balskus, E.P., Turnbaugh, P.J. (2013). Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. Science 341(6143), 295-298.

22. Hansen, E.E., Lozupone, C.A., Rey, F.E., Wu, M., Guruge, J.L., Narra, A., Goodfellow, J., Zaneveld, J.R., McDonald, D.T., Goodrich, J.A., et al. (2011). Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. Proc. Natl. Acad. Sci. U.S.A. 108(Suppl), 4599–4606.

23. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh, H.J., Tappu, R. (2016). MEGAN Community Edition – Interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput. Biol. 12(6), e1004957.

24. Hyatt, D., LoCascio, P.F., Hauser, L.J., Uberbacher, E.C. (2012). Gene and translation initiation site prediction in metagenomic sequences. Bioinformatics 28(17), 2223-2230.

25. Kovachenka-Datchary, P., Nilsson, A., Akrami, R., Lee, Y.S., De Vadder, F., Arora, T., Hallen, A., Martens, E., Björck, I., Bäckhed, F. (2015). Dietary fiber-induced improvement in glucose metabolism is associated with increased abundance of *Prevotella*. Cell Metab. 22(6), 971-982.

26. Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.M., Kennedy, S., et al. (2013). Richness of human gut microbiome correlates with metabolic markers. Nature 500(7464), 541-546.

27. Ley, R.E. (2016). Gut microbiota in 2015: *Prevotella* in the gut: choose carefully. Nat. Rev. Gastroenterol. Hepatol. 13(2), 69-70.

28. Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A.B., Brady, A., Creasy, H.H., McCracken, C., Giglio, M.G., et al. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. Nature 550(7674), 61-66.

29. Lombard, V., Golaconda Ramulu, H., Drula, E, Coutinho, P.M., Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, d490–d495.

30. Lozupone, C.A., Rhodes, M.E., Neff, C.P., Fontenot, A.P., Campbell, T.B., Palmer, B.E. (2014). HIV-induced alteration in gut microbiota: driving factors, consequences, and effects of antiretroviral therapy. Gut Microbes 5, 562–570.

31. Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. Nature 489(7415), 220-230.

32. Maeda, Y., Kurakawa, T., Umemoto, E., Motooka, D., Ito, Y., Gotoh, K., Hirota, K., Matsushita, M., Furuta, Y., Narazaki, M., et al. (2016). Dysbiosis contributes

346 to arthritis development via activation of autoreactive T cells in the intestine. Arthritis
347 Rheumatol. 68(11), 2646-2661.

348 33. Marchesi, J.R., Adams, D.H., Fava, F., Hermes, G.D.A., Hirschfield, G.M., Hold, G., Quraishi,
349 M.N., Kinross, J., Smidt, H., Tuohy, K.M., et al. (2016). The gut microbiota and host health:
350 a new clinical frontier. Gut 65(2), 330-339.

351 34. Obregon-Tito, A.J., Tito, R.Y., Metcalf, J., Sankaranarayanan, K., Clemente, J.C., Ursell,
352 L.K., Zech Xu, Z., Van Treunen, W., Knight, R., Gaffney, P.M., et al. (2015). Subsistence
353 strategies in traditional societies distinguish gut microbiome. Nat. Commun. 6, 6505.

354 35. Pedersen, H.K., Gudmundsdottir, V., Nielsen, H.B., Hyotylainen, T., Nielsen, T., Jensen,
355 B.A., Forslund, K., Hildebrand, F., Prifti, E., Falony, G., et al. 2016. Human gut microbes
356 impact host serum metabolome and insulin sensitivity. Nature 535(7612), 376-381.

357 36. Pierce, J.V., Bernstein, H.D. (2016). Genomic diversity of enterotoxigenic strains
358 of *Bacteroides fragilis*. PLoS ONE 11(6), e0158171.

359 37. Rampelli, S., Schnorr, S.L., Consolandi, C., Turroni, S., Severgnini, M., Peano, C., Brigidi,
360 P., Crittenden, A.N., Henry, A.G., Candela, M. (2015). Metagenome sequencing of the Hadza
361 hunter-gatherer gut microbiota. Curr. Biol. 25(13), 1682-1693.

362 38. Scher, J.U., Sczesnak, A., Longman, R.S., Segata, N., Ubeda, C., Bielski, C., Rostron, T.,
363 Cerundolo, V., Pamer, E.G., Abramson, S.B., et al. 2013. Expansion of intestinal *Prevotella*
364 *copri* correlates with enhanced susceptibility to arthritis. Elife 2, e01202.

365 39. Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., Waller, A.,
366 Mende, D.R., Kultima, J.R., Martin, J., et al. (2013). Genomic variation landscape of the
367 human gut microbiome. Nature 493(7430), 45-50.

368 40. Scholz, M., Ward, D.V., Pasolli, E., Tolio, T., Zolfo, M., Asnicar, F., Truong, D.T., Tett, A.,
369 Morrow, A.L., Segata, N. 2016. Strain-level microbial epidemiology and population
370 genomics from shotgun metagenomics. Nat. Methods 13(5), 435-438.

41. Segata, N. (2015). Gut microbiome: Westernization and disappearance of intestinal diversity. Curr. Biol. 25(14), R611-613.
42. Sharon, G., Sampson, T.R., Geschwind, D.H., Mazmanian, SK. (2016). The central nervous system and the gut microbiome. Cell 167(4), 915-932.
43. Siezen, R.J., Tzeneva, V.A., Castioni, A., Wels, M., Phan, H.T., Rademaker, J.L., Starrenburg, M.J., Kleerebezem, M., Molenaar, D., van Hylckama Vlieg, J.E. (2010). Phenotypic and genomic diversity of *Lactobacillus plantarum* strains isolated from various environmental niches. Environ. Microbiol. 12(3), 758–773
44. Solheim, M., Aakra, A., Snipen, L.G., Brede, D.A., Nes, I.F. (2009). Comparative genomics of *Enterococcus faecalis* from healthy Norwegian infants. BMC Genomics 10, 194.
45. Sonnenburg, E.D., Sonnenburg, J.L. (2014). Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. Cell Metab. 20(5), 779-786.
46. Sonnenburg, J.L., Bäckhed, F. (2016). Diet-microbiota interactions as moderators of human metabolism. Nature 535(7610), 56-64.
47. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9), 1312-1313.
48. Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., Segata, N. (2015). MetaPhlAn2 for enhanced metagenomics taxonomic profiling. Nat. Methods 12(10), 902-903.
49. Truong, D.T., Tett, A., Pasolli, E., Huttenhower, C., Segata, N. (2017). Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res. 27(4), 626-638.
50. Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C., Knight, R., Gordon, J.I. (2007). The human microbiome project: exploring the microbial part of ourselves in a changing world. Nature 449(7164), 804-810.

51. Vangay, P., Abigail, J., Johnson, J., Ward, T.L., Al-Ghalith, G.A., Shields-Cutler, R.R., Hillmann, B.M., Lucas, S.K., Beura, L.K., Thompson, E.A., et al. (2018). US immigration Westernizes the human gut microbiome. Cell 178, 962-972.

52. Vineis, P., Riboli, E. (2009). The EPIC study: an update. Recent Results Cancer Res. 181, 63-70.

53. Vitaglione, P., Mennella, I., Ferracane, R., Rivellese, A.A., Giacco, R., Ercolini, D., Gibbons, S.M., La Storia, A., Gilbert, J.A., Jonnalagadda, S., et al. (2015). Whole-grain wheat consumption reduces inflammation in a randomized controlled trial on overweight and obese subjects with unhealthy dietary and lifestyle behaviors: role of polyphenols bound to cereal dietary fiber. Am. J. Clin. Nutr. 101, 251-261

54. Wexler, H.M. (2007). *Bacteroides*: the good, the bad, and the nitty-gritty. Clin. Microbiol. Rev. 20(4), 593–621.

55. Wu, G., Zhang, C., Wu, H., Wang, R., Shen, J., Wang, L., Zhao, Y., Pang, X., Zhang, X., Zhao, L., et al. (2017). Genomic microdiversity of *Bifidobacterium pseudocatenulatum* underlying differential strain-level responses to dietary carbohydrate intervention. mBio 8, e02348-16.

56. Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. Science 334(6052), 105-108.

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

58. Zhang, C., Zhao, L. (2016). Strain-level dissection of the contribution of the gut microbiome to human metabolic disease. Genome Med. 8(1), 41.

59. Zhernakova, A., Kurilshikov, A., Bonder, M.J., Tigchelaar, E.F., Schirmer, M., Vatanen, T.,

Mujagic, Z., Vila, A.V., Falony, G., Vieira-Silva, S., et al. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science 352(6285), 565-569.

60. Zmora, N., Zeevi, D., Korem, T., Segal, E., Elinav, E. (2016). Taking it personally: personalized utilization of the human microbiome in health and disease. Cell Host Microbe 19(1), 12-20.

Figure legends

Figure 1. *Prevotella copri* pangenome is associated with specific diets. Panel A: Principal Coordinates Analysis based on *P. copri* pangenome. Vegetarian and vegan (VVG) subjects are included in the same group. Panel B: same plot as panel A, where subjects are coloured according to omnivore (O), vegetarian (VG) and vegan (V) diet. Core genes (present in 100% of the samples) were excluded.

Figure 2. Single Nucleotide Polymorphisms (SNPs) in *Prevotella copri* genomes differentiate subjects by diet. Phylogenetic tree built on concatenated *P. copri* genes extracted from assembled metagenomes. Only samples showing at least 2.5 Mb of alignment to *P. copri* DSM 18205 genome were retained.

Figure 3. *Prevotella copri* pangenome differentiates Western and non-Western individuals. Panel A: Principal Coordinates Analysis based on *P. copri* pangenome of our cohort and other Western/non-Western cohorts from previous studies. Samples are coloured according to the origin. Panel B: same plot as panel A, where subjects are both according to the origin and to the reference study. Core genes (present in 100% of the samples) were excluded.

445

446 **Table 1.** *P. copri* genes with a significantly different occurrence in Italian omnivore, vegetarian and
447 vegan individuals.
448

Comparison of Omnivores (O) vs Vegans (V)						
Gene ID	**Gene name**	**KEGG metabolism and pathway**	**E.C.**	**p-value**	**Prevalence in V**	**Prevalence in O**
g000271	TonB-dependent receptor	NA	NA	0.05878	29.41%	0%
g000338	3-isopropylmalate dehydrogenase	Amino Acid Metabolism; Valine, leucine and isoleucine biosynthesis	1.1.1.85	0.00953	23.53%	70%
g000562	4-amino-4-deoxychorismatel yase	Metabolism of Cofactors and Vitamins; Folate biosynthesis	4.1.3.38	0.003252	94.12%	41.67%
g000563	para-aminobenzoate synthetase componentI	Metabolism of Cofactors and Vitamins; Folate biosynthesis	2.6.1.85	0.000862	94.12%	33.33%
g000800	alpha-L-fucosidase	Glycan Biosynthesis and Metabolism; Other glycan degradation	3.2.1.51	0.006041	82.35%	25%
g000807	1,4-beta-xylanase	NA	NA	0.05634	94.12%	58.33%
g000920	pectate lyase	NA	NA	0.000484	76.47%	8.33%
g000922	alpha-glucosidase	NA	NA	0.002509	76.47%	16.67%
g000924	peptidase S24	NA	NA	0.01773	17.65%	66.67%
g001013	arginase 1	Amino acid metabolism; Arginine and proline metabolism	3.5.3.1	0.03819	88.24%	50%
g001041	phosphoenolpyruvate carboxykinase	Carbohydrate Metabolism; Pyruvate metabolism	4.1.1.49	0.04597	82.35%	41.67%
g001142	rhamnulokinase	Carbohydrate Metabolism; Pentose and glucuronate interconversions	2.7.1.5	0.05348	58.82%	16.67%
g001144	L-rhamnose-proton symport protein (RhaT)	NA	NA	0.05348	58.82%	16.67%
g001203	Putative glycoside hydrolase	NA	NA	0.04597	82.35%	41.67%
g001240	nitroreductase	Xenobiotics biodegradation and metabolism; Nitrotoluene degradation	NA	0.00953	23.53%	75%
g001419	phage associated protein	NA	NA	0.028	35.29%	0%
g001539	acetylxylanesterase	NA	3.1.1.72	0.007775	70.59%	16.67%
g001569	N-acetyl transferase	NA	NA	0.028	35.29%	0%
g002040	RagB/SusD domain protein	NA	NA	0.003252	58.33%	0
g002041	thiol-disulfide isomerase-like thioredoxin	NA	NA	0.004522	11.76%	66.67%
g002054	SusE outer membrane protein	NA	NA	0.01909	52.94%	8.33%

g002058	carbohydrate-binding protein	NA	NA	0.01773	66.67%	0
g002259	nitrate ABC transporter ATPase	NA	NA	0.000720	17.65%	83.33%
g002283	threonine aldolase	Amino acid metabolism; Glycine. serine and threonine metabolism	4.1.2.48	0.007775	29.41%	83.33%
g002284	arginase	Amino acid metabolism; Arginine and proline metabolism	3.5.3.1	0.006041	17.65%	75%
g002334	preprotein translocase, SecA subunit	NA	NA	0.001255	11.76%	75%
g002397	tripeptidyl aminopeptidase	NA	NA	0.003252	5.88%	58.33%
g002408	RagB/SusD domain protein	NA	NA	0.01057	50%	0
g002464	putative phage related protein	NA	NA	0.00953	76.47%	25%
g002465	phage uncharacterized protein	NA	NA	0.000720	82.35%	16.67%
g002469	zinc ABC transporter substrate-binding protein	NA	NA	0.02874	23.53%	66.67%
g002508	thiolperoxidase	NA	NA	0.00953	23.53%	75%
g003225	thiamine biosynthesis protein ThiH	Metabolism of Cofactors and Vitamins; Thiamine metabolism	NA	0.05634	5.88%	41.67%
g003319	O-acetylhomoserine (thiol)-lyase	Amino acid metabolism; Cysteine and methionine metabolism	2.5.1.49	0.001161	41.18%	100%
g003320	cystathionine beta-lyase	Energy Metabolism; Sulfur metabolism	4.4.1.8	0.001161	41.18%	100%
g003324	mannitol 2-dehydrogenase	Carbohydrate Metabolism; Fructose and mannose metabolism	1.1.1.67	0.00953	23.53%	75%

Comparison of Omnivores (O) vs Vegetarians (VG)						
Gene ID	**Gene name**	**KEGG metabolism and pathway**	**E.C.**	**p-value**	**Prevalence in VG**	**Prevalence in O**
g000046	alpha-L-fucosidase	Glycan Biosynthesis and Metabolism; Other glycan degradation	3.2.1.51	0.03036	36.36%	0%
g000271	TonB-dependent receptor	NA	NA	0.03563	31.82%	0%
g000562	4-amino-4-deoxychorismatel yase	Metabolism of Cofactors and Vitamins; Folate biosynthesis	4.1.3.38	0.02556	82%	41.67%
g000859	N-acetylmuramoyl-L-alanine amidase	NA	NA	0.03676	100%	0%
g001013	arginase 1	Amino acid metabolism; Arginine and proline metabolism	3.5.3.1	0.01274	90.91%	50%
g001419	phage associated protein	NA	NA	0.03563	31.82%	0%
g002053	SusD family protein	NA	NA	0.02968	59.09%	16.67%

g002499	Beta-xylosidase, xynB	Carbohydrate Metabolism; Amino sugar and nucleotide sugar metabolism	3.2.1.37	0.04198	95.45%	66.67%
g003320	cystathionine beta-lyase	Energy Metabolism; Sulfur metabolism	4.4.1.8	0.03036	63.64%	100%

Comparison of Vegans (V) vs Vegetarians (VG)						
Gene ID	**Gene name**	**KEGG metabolism and pathway**	**E.C.**	**p-value**	**Prevalence in VG**	**Prevalence in V**
g000563	para-aminobenzoate synthetase component I	Metabolism of Cofactors and Vitamins; Folate biosynthesis	2.6.1.85	0.02402	59.09%	94%
g000800	alpha-L-fucosidase 2	Glycan Biosynthesis and Metabolism; Other glycan degradation	3.2.1.51	0.02015	40.91%	82.35%
g003319	O-acetylhomoserine (thiol)-lyase	Cysteine and methionine metabolism	2.5.1.49	0.05862	72.73%	41.18%
g003324	mannitol 2-dehydrogenase	Carbohydrate Metabolism; Fructose and mannose metabolism	1.1.1.67	0.02273	63.64%	23.53%
g000800	alpha-L-fucosidase	Glycan Biosynthesis and Metabolism; Other glycan degradation	3.2.1.51	0.02015	40.91%	82.35%
g000920	pectate lyase	NA	NA	0.02273	36.36%	76.47%
g001240	nitroreductase	Xenobiotics biodegradation and metabolism; Nitrotoluene degradation	NA	0.0496	59.09%	23.53%
g001449	putative PTS permease protein	NA	NA	0.02441	54.55%	17.65%
g002259	nitrate ABC transporter ATPase	NA	NA	0.04901	50%	17.65%
g002283	threonine aldolase	Amino acid metabolism; Glycine. serine and threonine metabolism	4.1.2.48	0.01052	72.73%	29.41%
g002284	arginase	Amino acid metabolism; Arginine and proline metabolism	3.5.3.1	0.04901	50%	17.65%
g002341	glycosyltransferase	NA	NA	0.03651	54.55%	88.24%
g002380	TonB-dependent receptor	NA	NA	0.002462	59.09%	100%
g002416	TonB-dependent receptor	NA	NA	0.05649	77.27%	100%
g002417	glucoside-hydrogenase	Carbohydrate Metabolism; Pentose phosphate pathway	1.1.1.47	0.01235	68.18%	100%
g002464	putative phage related protein	NA	NA	0.00953	31.82%	76.47%
g002465	phage uncharacterized protein	NA	NA	0.003118	31.82%	82.35%
g002496	putative bacteriophage integrase	NA	NA	0.00953	68.18%	23.53%
g002301	aminoacid carrier protein	NA	NA	0.05649	0.00%	100%

449 VG, vegetarians; V, vegans; O, omnivores.
450 NA, not available.
451 p-values were calculated by paired Chi-squared test.
452 Occurrence was calculated based on the percentage of samples for each diet group showing the gene.

453

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Danilo Ercolini (ercolini@unina.it).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study population

Study population included 53 subjects coming from our previously characterized cohort (De Filippis et al., 2016a). Moreover, we added 44 metagenomes from a newly recruited cohort study in Turin (Italy). Subjects were healthy volunteers recruited between May 2017-July 2018, in collaboration with the Italian Society of Vegetarian Nutrition (http://www.scienzavegetariana.it/). All subjects filled a validated, self-administered food-frequency questionnaire assessing the usual diet, together with lifestyle and personal history data, in accordance to the EPIC study standards (Vineis and Riboli, 2009), where specific questions for vegan and vegetarian dietary regimes were included. All subjects were 18-60 years old, following the declared dietary regime for at least one year. The following exclusion criteria were applied: supplementation with prebiotics or probiotics, consumption of antibiotics in the previous 3 months. pregnancy and lactation. intestinal pathologies (Crohn's disease, chronic ulcerative colitis, bacterial overgrowth syndrome, constipation, celiac disease, Irritable Bowel Syndrome, colorectal cancer), and other pathologies (type I or type II diabetes, cardiovascular or cerebrovascular diseases, concomitant neoplastic diseases, neurodegenerative disease, rheumatoid arthritis, allergies). This protocol was approved by the Ethics Committee of Azienda Ospedaliera "SS. Antonio e Biagio e C. Arrigo" of Alessandria, Italy (protocol number Colorectal miRNA_CEC2014).

The final dataset included 97 subjects: omnivores, n=23; vegetarians, n=38; vegans, n=36. The main characteristics of the cohort are reported in Table S1. Metagenomes sequenced in this study are available at the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI, accession numbers: SRP083099; SRP126540).

We further complemented the metagenomes of our cohort with those of previously published cohorts from Western and non-Western populations: 27 Hadza (traditional hunter-gatherer population from East-Africa) and 11 Italians (Rampelli et al., 2015 - NCBI SRA SRP056480); 23 Malawian, 21 Amerindians and 66 Americans (Yatsunenko et al., 2012 – MG-RAST qiime:621); 24 Matses (hunter-gatherer population from the Peruvian Amazon), 12 Tunapuco (traditional agricultural community from the Andean highlands) and 22 Americans (Obregon-Tito et al., 2016 – NCBI SRA PRJNA268964); 96 healthy, normal-weight Danes (Le Chatelier et al., 2013 – EBI European Nucleotide Archive ERP003612).

Only 79 metagenomes (i.e., 2 Italians and 23 Hadza from Rampelli et al.; 10 Tunapuco and 3 Americans from Obregon-Tito et al.; 7 Malawian, 3 Amerindians and 2 Americans from Yatsunenko et al.; 29 Danes from Le Chatelier et al.) from the different cohorts were retained in the analysis due to low abundance of *P. copri* in the other samples.

Urinary levels of the branched-chain amino acids (BCAA) leucine and isoleucine available for some of the Italian subjects were retrieved from the results reported in the original study (De Filippis et al., 2016a).

METHOD DETAILS

Libraries preparation and sequencing

Libraries were prepared using the Nextera DNA Library Preparation kit (Illumina) and sequenced on an Illumina HiSeq platform (leading to 40,552,111 ±9,650,536 reads/sample).

Bioinformatics data analysis

Read filtering

Host contamination was removed using the human sequence removal procedure from the Human Microbiome Project (Turnbaugh et al., 2007). Raw reads were quality-trimmed (Phred score < 25) and reads shorter than 60 bp were discarded with the SolexaQA++ software (Cox et al., 2010). Number of reads/sample resulting after filtering is reported in Table S1.

Taxonomic profiling

Taxonomic profiling was carried out by using MetaPhlAn2 (version 2.6; Truong et al., 2015).

Prevotella copri *pangenomics from short-reads*

Strain-level analysis was performed through a gene-content-based profiling using PanPhlAn (Scholz et al., 2016). *P. copri* pangenomes from all the samples were inferred using PanPhlAn (Scholz et al., 2016), with parameters --min_coverage 1, --left_max 1.70, and --right_min 0.30. This led to 50 samples left in the Italian dataset (9 omnivores, 22 vegetarians, 19 vegans). Pangenome representative sequences were extracted from the PanPhlAn *P. copri* database and aligned to the NCBI Non-Redundant (NR) database using BLASTx (version 2.3.0, Camacho et al., 2008; e-value cutoff of $1e^{-5}$, requiring a hit to display > 90% of identity over at least 30% of the query length) for performing functional annotation. Functional annotation in the KEGG database was performed through MEGAN 6 Ultimate Edition (Huson et al., 2016). Carbohydrates Active Enzymes were identified using the CAZy database (Lombard et al., 2014; http://www.cazy.org/).

Finally, we compared *P. copri* pangenomes of our samples with Western and non-Western subjects from Rampelli et al. (2015). Obregon-Tito et al. (2016), Yatsunenko et al. (2012), Le Chatelier et al. (2013). Seventy-nine metagenomes (i.e., 2 Italians and 23 Hadza from Rampelli et al.; 10 Tunapuco and 3 Americans from Obregon-Tito et al.; 7 Malawian, 3 Amerindians and 2 Americans from Yatsunenko et al.; 29 Danes from Le Chatelier et al.) were retained in the analysis by considering the aforementioned PanPhlAn parameters. Differential occurrence of pangenes between O, VG and V groups or between Western and non-Western groups was determined by Chi-squared test. carried out in the R environment (*chisq.test* function in the package *MASS*).

Prevotella copri *genome reconstruction through metagenomics assembly*

High-quality reads were assembled with SPAdes 3.9.0 using the --meta option (Bankevich et al., 2012) and kmer length from 21 to 91bp. Resulting contigs (> 1000bp length) were then aligned to the *P. copri* reference genome (DSM 18205) with BLASTn, using an e-value cutoff of $1e^{-5}$, requiring a hit to display > 90% of identity over at least 30% of the query length. ORFs were called on the

resulting contigs with the automated gene prediction pipeline MetaProdigal (version 2.6.3; Hyatt et al., 2012). Assembly results and alignment length are reported in Table S1. Since the *P. copri* reference genome is about 3.5 Mb, samples with less than 2.5 Mb of alignment length were excluded from further analysis, which resulted in a total of 33 reconstructed genomes. This generated a catalogue of 26,104 ORFs spanning 154 core (present in 100% of the strains) and soft core (95-99% of the strains) genes. Core genes were concatenated, aligned and processed with RAxML (version 8; Stamatakis, 2014) to generate a phylogenetic tree, visualized by using Archaeopterix (https://sites.google.com/site/cmzmasek/home/software/archaeopteryx).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical Analysis

Statistical analyses were carried out in the R environment. Chi-squared test was carried out using the *chisq.test* function in the package *MASS*. Principal Coordinates Analysis (*dudi.pco* function in *made4* package) was carried out on a distance matrix calculated on Bray Curtis's distance (*vegdist* function in package *vegan*). Multivariate Analysis of Variance (MANOVA, *adonis* function in package *vegan*) was carried out on Bray Curtis' dissimilarity matrix to test the overall difference in pangenome composition among diet groups or between Western and non-Western subjects. Pair-wise Wilcoxon-Mann-Withney (*pairwise.wilcox.test* function in package *base*) test was used to test differential abundance of *Prevotella copri*, BCAA or CAZymes hits. If not specified, p-value < 0.05 was considered statistically significant.

DATA AND SOFTWARE AVAILABILITY

Metagenomes produced in this study are available at the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI, Accession numbers: SRP126540 and SRP083099).

Supplemental item titles

Table S1. Samples analysed in this study. Related to STAR Methods.

Table S2. *P. copri* CAZymes with a significantly different occurrence in Italian omnivores (O), vegetarians (VG) and vegans (V). Related to Figure 1.

Table S3. *P. copri* genes with a significantly different occurrence in Western and non-Western cohorts (p<0.05). Related to Figure 4.

Figure S1. Abundance of *Prevotella copri* found in the gut metagenomes. Results are expressed as % on total reads, normalized at 20 million reads/sample. Samples are coloured according to the diet type. Related to Figure 1.

Figure S2. Box plots showing the number of hits to the CAZy Carbohydrate Esterase family in the *P. copri* pangenome of omnivores (O), vegetarians (VG) and vegans (V). Boxes represent the interquartile range (IQR) between the first and third quartiles, and the line inside represents the median (2nd quartile). Whiskers denote the lowest and the highest values within 1.5 IQR from the first and third quartiles, respectively. Different letters on the top of the boxes indicate significantly different values (p<0.05). Related to Figure 1.

Figure S3. Occurrence of *P. copri leuB* gene is linked to different BCAA urinary levels. Box plots showing the abundance of urinary branched-chain amino acids (BCAA), reported as sum of leucine and isoleucine. The subjects are grouped according to diet type and to the presence of *leuB* gene in *P. copri* pangenome. Red, omnivores; light green, vegans/vegetarians positive for *leuB*; dark green, vegans/vegetarians negative for *leuB*. Different letters on the top of the boxes indicate significantly different values (Wilcoxon-Mann-Withney, p<0.01). For a definition of box plot refer to Figure S2. Related to Figure 1.