# Cell Host & Microbe

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

Graphical Abstract
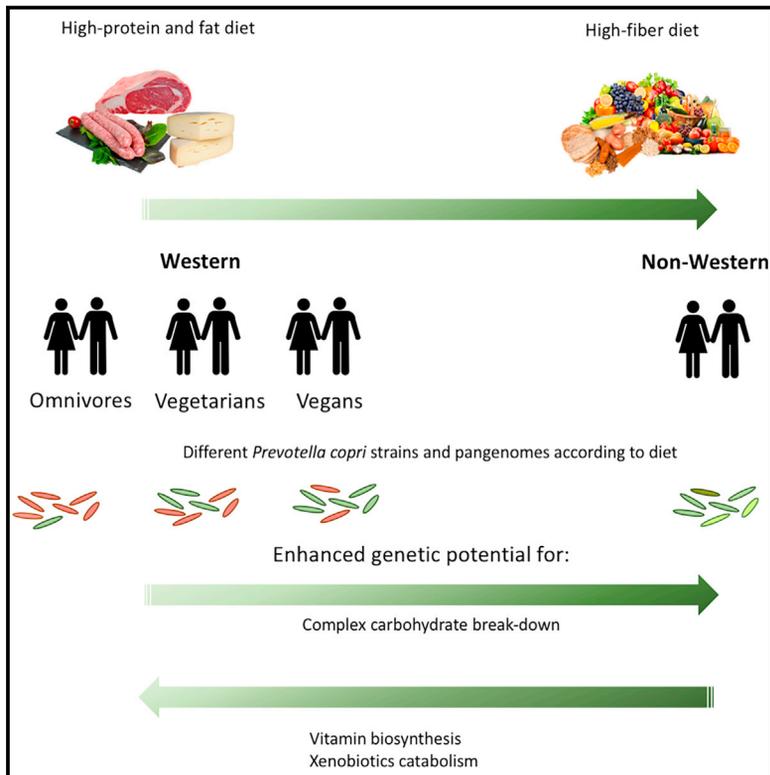
Authors

Francesca De Filippis, Edoardo Pasolli, Adrian Tett, ..., Marco Gobbetti, Nicola Segata, Danilo Ercolini

Correspondence

ercolini@unina.it

In Brief

The gut microbiome includes several strains per species, with high genomic diversity. By examining Italian subjects with varying dietary habits, De Filippis et al. demonstrate that diet may select distinctive *Prevotella copri* strains with distinguishable functions. This diversity may explain subject-specific responses to dietary interventions and variations in human health.

Highlights

- Screening of human gut metagenomes reveals different *Prevotella copri* pangenomes

- Habitual diet and lifestyle can select different *P. copri* strains

- Strains from non-Western subjects show higher potential for complex fiber break-down

- Strains from Western subjects have a higher prevalence of drug metabolism genes

CellPress

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] and Danilo Ercolini[1,2,9,*]

[1]Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy
[2]Task Force on Microbiome Studies, University of Naples Federico II, Naples, 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, Bari, Italy
[6]Department of Food Science, University of Parma, Parma, Italy
[7]Department of Agricultural, Forest and Food Sciences, University of Torino, Turin, Italy
[8]Faculty of Science and Technology, Free University of Bozen, Bozen, Italy
[9]Lead Contact
*Correspondence: ercolini@unina.it
https://doi.org/10.1016/j.chom.2019.01.004

SUMMARY

The role of intestinal *Prevotella* species in human health is controversial, with both positive and negative associations. Strain-level diversity may contribute to discrepancies in genus and species associations with health and disease. We dissected the gut metagenomes of Italians with varying dietary habits, investigating the presence of distinct *Prevotella copri* strains. Fiber-rich diets were linked to *P. copri* types with enhanced potential for carbohydrate catabolism. *P. copri* strains associated with an omnivore diet had a higher prevalence of the *leuB* gene—involved in branched-chain amino acid biosynthesis—a risk factor for glucose intolerance and type 2 diabetes. These *P. copri* pangenomes were compared to existing cohorts, providing evidence of distinct gene repertoires characterizing different *P. copri* populations, with drug metabolism and complex carbohydrate degradation significantly associated with Western and non-Western individuals, respectively. Strain-level *P. copri* diversity in gut microbiomes is affected by diet and should be considered when examining 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, 2017). 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 gut microbiome, and several researches have highlighted that dietary "Westernization"— characterized by higher consumption of high-fat and protein products at the expense of foods rich in fiber—may have caused a loss of microbial diversity, with ultimate repercussions on human health (Segata, 2015; Sonnenburg and Sonnenburg, 2014).

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 linked to 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 external 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 the gut microbiome and human health. Indeed, each microbial genus in the gut includes several species and strains that may harbor 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).
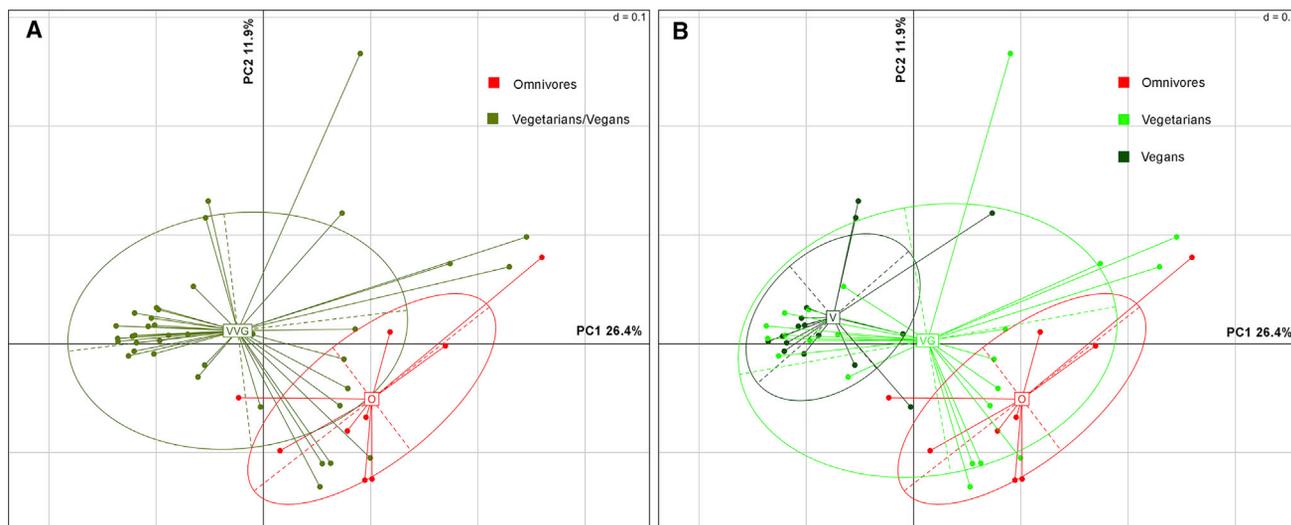
Figure 1. *Prevotella copri* Pangenome Is Associated with Specific Diets
(A) Principal coordinates analysis based on *P. copri* pangenome. Vegetarian and vegan (VVG) subjects are included in the same group.
(B) Same plot as (A), where subjects are colored according to omnivore (O), vegetarian (VG), and vegan (V) diet. Core genes (present in 100% of the samples) were excluded.

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 vegetable-rich diets would suggest that *Prevotella*, as a fiber-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), an improved glucose metabolism (Kovatcheva-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 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 behavior can be most likely explained by a strain-level diversity.

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 their 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 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, and Veiga, 2017; Zmora et al., 2016).

In order to study 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 with 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 for a strain-level analysis in 47 samples (9 O, 22 VG, and 16 V). 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, and 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. The abundance of *P. copri* in our metagenomes ranged from 0 to 83.2% (Figure S1), and it was not significantly associated with diet type (O, V, or VG), 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 Kyoto Encyclopedia of Genes and Genomes (KEGG) functional categories. No associations of *P. copri* pangenome with sex was found by MANOVA (p > 0.05). Principal coordinates analysis (PCoA) clearly separated omnivore from

Table 1. *P. copri* Genes with a Significantly Different Occurrence in Italian Omnivore, Vegetarian, and Vegan Individuals

Comparison of Omnivores, O, versus 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	0.058	29.4	0.0
g000338	3-isopropylmalate dehydrogenase	amino acid metabolism; valine, leucine, and isoleucine biosynthesis	1.1.1.85	0.009	23.5	70.0
g000562	4-amino-4-deoxychorismatel yase	metabolism of cofactors and vitamins; folate biosynthesis	4.1.3.38	0.003	94.1	41.7
g000563	para-aminobenzoate synthetase componentl	metabolism of cofactors and vitamins; folate biosynthesis	2.6.1.85	0.001	94.1	33.3
g000800	alpha-L-fucosidase	glycan biosynthesis and metabolism; other glycan degradation	3.2.1.51	0.006	82.4	25.0
g000807	1,4-beta-xylanase	NA	NA	0.056	94.1	58.3
g000920	pectate lyase	NA	NA	0.0004	76.5	8.3
g000922	alpha-glucosidase	NA	NA	0.003	76.5	16.7
g000924	peptidase S24	NA	NA	0.018	17.7	66.7
g001013	arginase 1	amino acid metabolism; arginine and proline metabolism	3.5.3.1	0.038	88.2	50.0
g001041	phosphoenolpyruvate carboxykinase	carbohydrate metabolism; pyruvate metabolism	4.1.1.49	0.045	82.4	41.7
g001142	rhamnulokinase	carbohydrate metabolism; pentose and glucuronate interconversions	2.7.1.5	0.053	58.8	16.7
g001144	L-rhamnose-proton symport protein (RhaT)	NA	NA	0.053	58.8	16.7
g001203	Putative glycoside hydrolase	NA	NA	0.046	82.4	41.7
g001240	nitroreductase	xenobiotics biodegradation and metabolism; nitrotoluene degradation	NA	0.010	23.5	75.0
g001419	phage associated protein	NA	NA	0.028	35.3	0.0
g001539	acetylxylanesterase	NA	3.1.1.72	0.008	70.6	16.7
g001569	N-acetyl transferase	NA	NA	0.028	35.3	0.0
g002040	RagB/SusD domain protein	NA	NA	0.003	58.3	0.0
g002041	thiol-disulfide isomerase-like thioredoxin	NA	NA	0.005	11.8	66.7
g002054	SusE outer membrane protein	NA	NA	0.019	52.9	8.3
g002058	carbohydrate-binding protein	NA	NA	0.018	66.7	0.0
g002259	nitrate ABC transporter ATPase	NA	NA	0.001	17.7	83.3
g002283	threonine aldolase	amino acid metabolism; glycine, serine, and threonine metabolism	4.1.2.48	0.008	29.4	83.3
g002284	arginase	amino acid metabolism; arginine and proline metabolism	3.5.3.1	0.006	17.7	75.0
g002334	preprotein translocase, SecA subunit	NA	NA	0.001	11.8	75.0
g002397	tripeptidyl aminopeptidase	NA	NA	0.003	5.9	58.3
g002408	RagB/SusD domain protein	NA	NA	0.011	50.0	0.0
g002464	putative phage related protein	NA	NA	0.010	76.5	25.0
g002465	phage uncharacterized protein	NA	NA	0.001	82.4	16.7
g002469	zinc ABC transporter substrate-binding protein	NA	NA	0.029	23.5	66.7
g002508	thiolperoxidase	NA	NA	0.010	23.5	75.0
g003225	thiamine biosynthesis protein ThiH	metabolism of cofactors and vitamins; thiamine metabolism	NA	0.056	5.9	41.7

Table 1. *Continued*

Comparison of Omnivores, O, versus Vegans, V

Gene ID	Gene Name	KEGG Metabolism and Pathway	E.C.	*p Value	Prevalence in V (%)	Prevalence in O (%)
g003319	O-acetylhomoserine (thiol)-lyase	amino acid metabolism; cysteine and methionine metabolism	2.5.1.49	0.001	41.2	100.0
g003320	cystathionine beta-lyase	energy metabolism; sulfur metabolism	4.4.1.8	0.001	41.2	100.0
g003324	mannitol 2-dehydrogenase	carbohydrate metabolism; fructose and mannose metabolism	1.1.1.67	0.010	23.5	75.0

Comparison of Omnivores, O, versus 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	0.030	36.4	0.0
g000271	TonB-dependent receptor	NA	NA	0.036	31.8	0.0
g000562	4-amino-4-deoxychorismatel yase	metabolism of cofactors and vitamins; folate biosynthesis	4.1.3.38	0.026	82.0	41.7
g000859	N-acetylmuramoyl-L-alanine amidase	NA	NA	0.037	100.0	0.0
g001013	arginase 1	amino acid metabolism; arginine and proline metabolism	3.5.3.1	0.012	90.9	50.0
g001419	phage associated protein	NA	NA	0.036	31.8	0.0
g002053	SusD family protein	NA	NA	0.030	59.1	16.7
g002499	Beta-xylosidase, xynB	carbohydrate metabolism; amino sugar and nucleotide sugar metabolism	3.2.1.37	0.042	95.5	66.7
g003320	cystathionine beta-lyase	energy metabolism; sulfur metabolism	4.4.1.8	0.030	63.6	100.0

Comparison of Vegans, V, versus 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	0.024	59.1	94.0
g000800	alpha-L-fucosidase 2	glycan biosynthesis and metabolism; other glycan degradation	3.2.1.51	0.020	40.9	82.4
g003319	O-acetylhomoserine (thiol)-lyase	cysteine and methionine metabolism	2.5.1.49	0.059	72.7	41.2
g003324	mannitol 2-dehydrogenase	carbohydrate metabolism; fructose and mannose metabolism	1.1.1.67	0.023	63.6	23.5
g000800	alpha-L-fucosidase	glycan biosynthesis and metabolism; other glycan degradation	3.2.1.51	0.020	40.9	82.4
g000920	pectate lyase	NA	NA	0.023	36.4	76.5
g001240	nitroreductase	xenobiotics biodegradation and metabolism; nitrotoluene degradation	NA	0.050	59.1	23.5
g001449	putative PTS permease protein	NA	NA	0.024	54.6	17.7
g002259	nitrate ABC transporter ATPase	NA	NA	0.049	50.0	17.7
g002283	threonine aldolase	amino acid metabolism; glycine, serine, and threonine metabolism	4.1.2.48	0.010	72.7	29.4
g002284	arginase	amino acid metabolism; arginine and proline metabolism	3.5.3.1	0.049	50.0	17.7
g002341	glycosyltransferase	NA	NA	0.037	54.6	88.2
g002380	TonB-dependent receptor	NA	NA	0.002	59.1	100.0
g002416	TonB-dependent receptor	NA	NA	0.056	77.3	100.0

Table 1. *Continued*

Comparison of Vegans, V, versus Vegetarians, VG

Gene ID	Gene Name	KEGG Metabolism and Pathway	E.C.	*p Value	Prevalence in VG (%)	Prevalence in V (%)
g002417	glucoside-hydrogenase	carbohydrate metabolism; pentose phosphate pathway	1.1.1.47	0.012	68.2	100.0
g002464	putative phage related protein	NA	NA	0.010	31.8	76.5
g002465	phage uncharacterized protein	NA	NA	0.003	31.8	82.4
g002496	putative bacteriophage integrase	NA	NA	0.010	68.2	23.5
g002301	amino acid carrier protein	NA	NA	0.056	0.0	100.0

NA, not available; VG, vegetarians; V, vegans; O, omnivores; E.C., Enzyme Commission; KEGG, Kyoto Encyclopedia of Genes and Genomes.
*p values were calculated by paired chi-squared test. Occurrence was calculated based on the percentage of samples for each diet group showing the gene.

non-omnivore (V and 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). Compared to O, thirty-six and eight pangenes occurred differentially in *P. copri* pangenomes of V and VG, respectively (p < 0.05; Table 1). Interestingly, V-associated *P. copri* strains showed a higher prevalence of genes involved in complex carbohydrate break-down (Table 1). Vegans showed a 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* pangenes (see STAR Methods). Glycoside hydrolase (GH) and carbohydrate esterase (CE) families were enriched in V compared to O and VG, although only CEs were significantly enriched in V (p < 0.05; Figure S2). In particular, CE7 and CE8, including acetyl xylan esterase and pectin methyl esterase, showed a higher prevalence in vegans. GH5, GH95, and GH127 containing enzymes, involved in complex polysaccharides breakdown, were enriched in V, while GH2, including β-galactosidase, prevailed in O (Table S2). In addition, genes involved in sulfur compound metabolism (cystathionine beta-lyase and O-acetylhomoserine thiol-lyase) were enriched in O compared to V, as well as 3-isopropylmalate dehydrogenase (*leuB*, EC 1.1.1.85), which is involved in branched-chain amino acid (BCAA) biosynthesis. All O, 67% of VG, and 18% of V harbored the *leuB* gene in *P. copri* pangenome. Interestingly, when we divided subjects for presence or absence of *leuB* in the *P. copri* pangenome, we found significantly lower urinary BCAA levels in V and VG individuals not harboring 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 and 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 the other studies clustered together with Italians from this study and Danes from Le Chatelier et al. (2013) (Figure 3B). In particular, 1,368 genes differentiated Western and non-Western subjects (Table S3). Among them, several genes encoding for *SusC* and *SusD* transporters, involved in starch binding, and xylanases, pectinesterases, β-glucosidases, and alpha-amylases, involved in carbohydrate 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, and 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, which are responsible for antibiotic and toxic compound export from the cell, and *DedA* family proteins, which are 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; Schloissnig et al., 2013; Scholz et al., 2016), with possibly different genomic potentials. Each strain may have very specific mechanisms through which it can affect health or respond to dietary patterns. Accordingly, strain-specific microbial determinants were found in virulence (Solheim et al., 2009; Pierce and Bernstein, 2016), drug resistance (Gill et al., 2005) or catabolism (Haiser et al., 2013), nutrient utilization
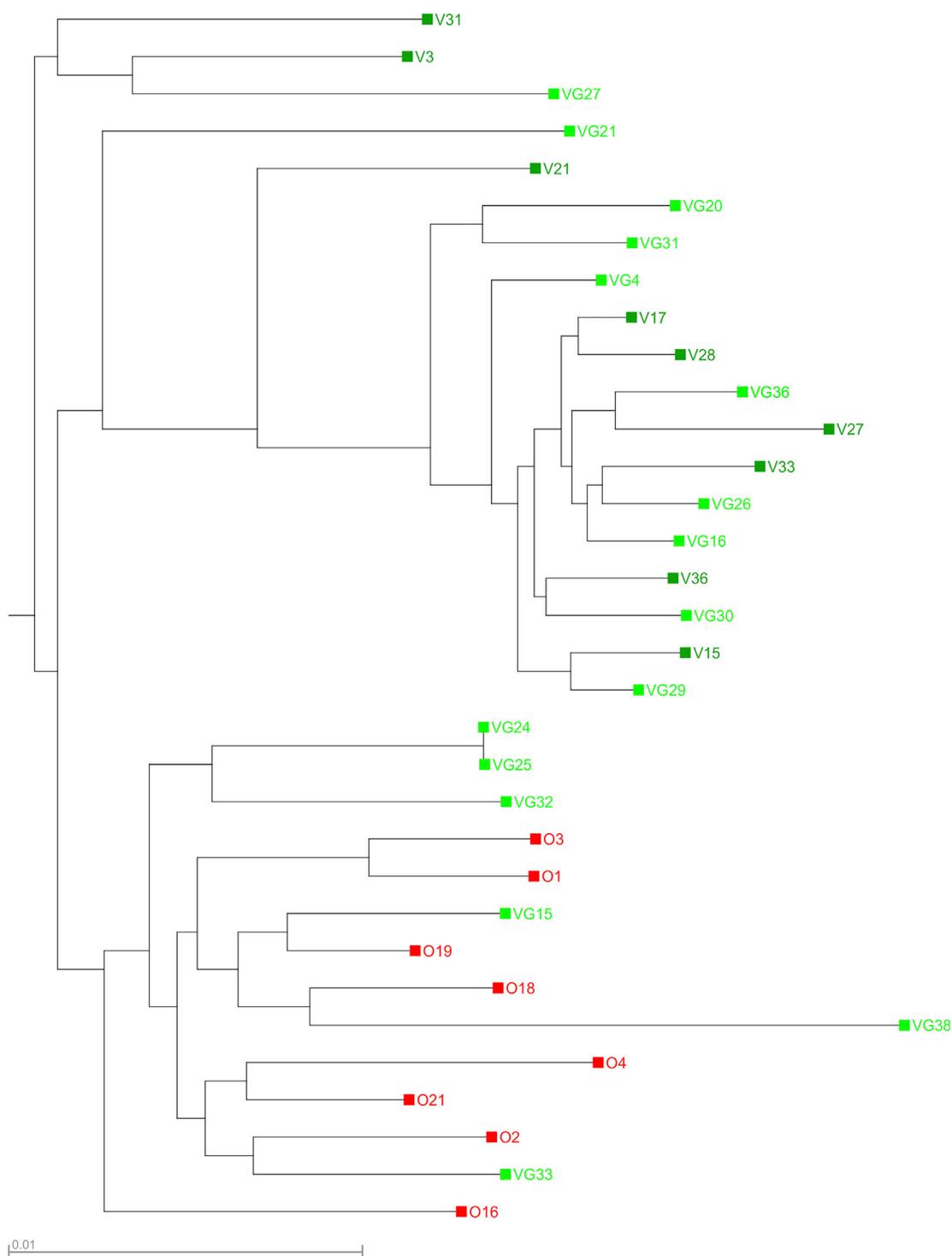
Figure 2. Single-Nucleotide Polymorphisms in *Prevotella copri* Genomes Differentiate Subjects by Diet
Phylogenetic tree built on concatenated *P. copri* genes extracted from assembled metagenomes. Only samples showing at least 2.5 Mb of alignment to *P. copri* DSM 18205 genome were retained.

(Wexler, 2007; Wu et al., 2017), adhesion to the gut epithelium (Hansen et al., 2011), and even induction of obesity (Fei and Zhao, 2013), highlighting the importance of a strain-level dissection to address the functional role of the gut microbiome.

Prevotella defines a clear subtype of the human gut microbiome (Arumugam et al., 2011) and is one of its most abundant members (Falony et al., 2016). However, its role in relation to human health is still unclear (Ley, 2016; Cani, 2018), being
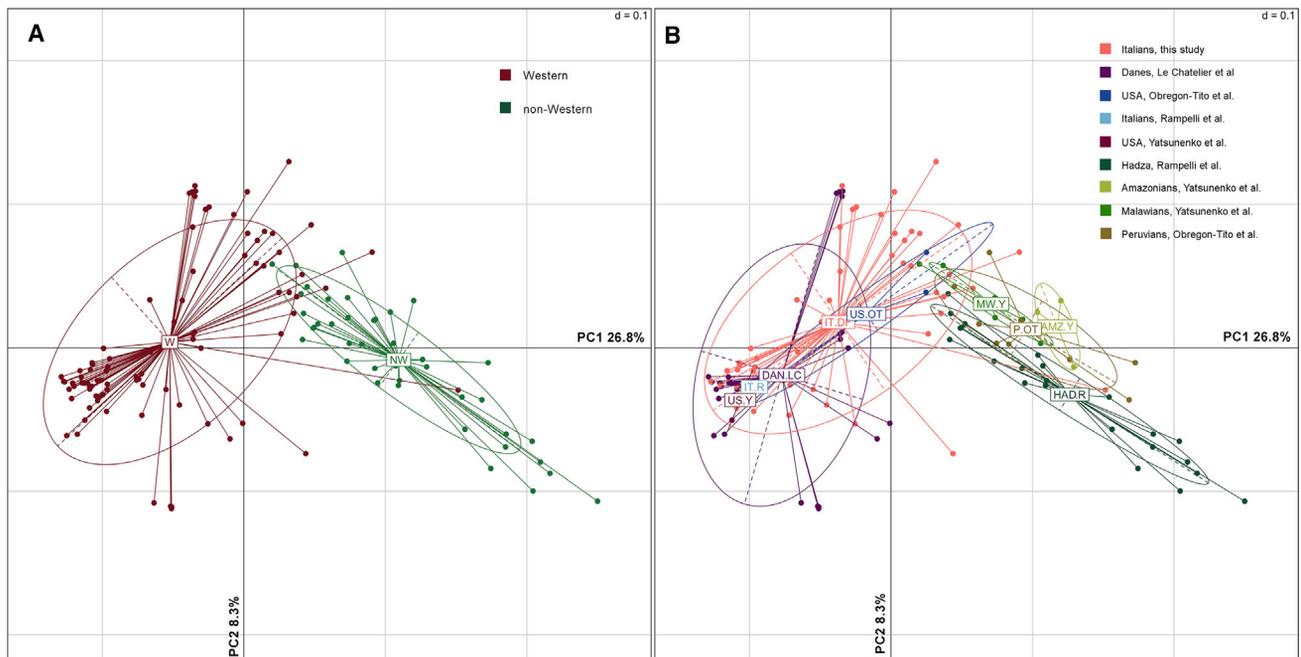
Figure 3. *Prevotella copri* Pangenome Differentiates Western and Non-Western Individuals
(A) Principal coordinates analysis based on *P. copri* pangenome of our cohort and other Western/non-Western cohorts from previous studies. Samples are colored according to the origin.
(B) Same plot as (A), where subjects are both according to the origin and to the reference study. Core genes (present in 100% of the samples) were excluded.

contrastingly associated either with health-promoting (Kovatch-eva-Datchary et al., 2015; De Vadder et al., 2016) or detrimental effects for the host (Lozupone et al., 2014; Maeda et al., 2016; Pedersen et al., 2016; Scher et al., 2013). Thus, further characterization of *P. copri* with strain-level resolution is required to understand if differences in the genomic potentials exist, how they are linked to human health or disease, and whether different dietary styles may select specific *P. copri* strains. Here, we investigated the pangenome of *P. copri* strains from gut metagenome of subjects with different dietary habits. We found a distinct clustering of the subjects based on the sample-specific *P. copri* gene repertoire and driven by diet, suggesting that different dietary habits may select for specific *P. copri* strains, as we had previously speculated based on the results of oligo-typing 16S rRNA gene sequences (De Filippis et al., 2016b). Omnivores with high abundance of *P. copri* oligotypes previously linked with an animal-based diet (P5, P12, and P16; see De Filippis et al., 2016b) consistently harbor a *P. copri* functional potential strongly different from V, supporting the value of oligo-typing in recognizing sub-genus diversity patterns in the gut microbiome (Eren et al., 2013). According to our results, a diet richer in fiber from fruit, vegetables, and legumes may select for *P. copri* strains with higher potential for complex carbohydrate degradation. On the contrary, the *P. copri* strains associated with omnivore diet showed a 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 concentrations 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 Western healthy Danes (Le Chatelier et al., 2013) in the analysis. 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 genes involved in the biosynthesis of vitamin B and folate. Green vegetables are usually rich in folates, 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, which are typically consumed in Western countries. Consistently, a recent study reported loss of *P. copri* genes associated with cellulose,

β-mannans, and xyloglucan degradation in rural Thai upon immigration to the United States (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 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 lifestyles and 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* strain 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, the *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 responses observed to dietary interventions or drug therapies (De Filippis et al., 2018; Zmora et al., 2016; Derrien, and Veiga, 2017). Consequently, inter-personal differences at strain-level exist and should be considered when investigating the role of certain microbial species in disease development. Strain-level biodiversity will also have to be considered in the near future in microbiome-targeted nutritional interventions developed for the prevention or treatment of diseases.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Study Population
- METHOD DETAILS
  - Libraries Preparation and Sequencing
  - Bioinformatics Data Analysis
  - **Prevotella copri** Pangenomics from Short-Reads
  - *Prevotella copri* Genome Reconstruction through Metagenomics Assembly
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical Analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and three tables and can be found with this article online at https://doi.org/10.1016/j.chom.2019.01.004.

WEB RESOURCES

Carbohydrate-Active Enzymes Database, http://www.cazy.org

REFERENCES

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M., et al. (2011). Enterotypes of the human gut microbiome. Nature *473*, 174–180.

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*, 455–477.

Blum, H.E. (2017). The human microbiome. Adv. Med. Sci. *62*, 414–420.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). Blast+: architecture and applications. BMC Bioinformatics *10*, 421.

Cani, P.D. (2018). Human gut microbiomes: hopes, threats and promises. Gut *67*, 1716–1725.

Cox, M.P., Peterson, D.A., and Biggs, P.J. (2010). SolexaQA: at-a-glance quality assessment of Illumina second generation sequencing data. BMC Bioinformatics *11*, 485.

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*, 559–563.

De Angelis, M., Montemurno, E., Vannini, L., Cosola, C., Cavallo, N., Gozzi, G., Maranzano, V., Di Cagno, R., Gobbetti, M., and Gesualdo, L. (2015). Effect of whole-grain barley on the human fecal microbiota and metabolome. Appl. Environ. Microbiol. *81*, 7945–7956.

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*, 1812–1821.

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

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

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and 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*, 14691–14696.

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

Derrien, M., and Veiga, P. (2017). Rethinking diet to aid human–microbe symbiosis. Trends Microbiol. *25*, 100–112.

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

Faith, J.J., Colombel, J.F., and 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*, 633–640.

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

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

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.

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

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

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*, 4599–4606.

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

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

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

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*, 541–546.

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

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*, 61–66.

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

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

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 to arthritis development via activation of autoreactive T cells in the intestine. Arthritis Rheumatol. *68*, 2646–2661.

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

Obregon-Tito, A.J., Tito, R.Y., Metcalf, J., Sankaranarayanan, K., Clemente, J.C., Ursell, L.K., Zech Xu, Z., Van Treuren, W., Knight, R., Gaffney, P.M., et al. (2015). Subsistenc– strategies in traditional societies distinguish gut microbiomes. Nat. Commun. *6*, 6505.

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

Pierce, J.V., and Bernstein, H.D. (2016). Genomic diversity of enterotoxigenic strains of Bacteroides fragilis. PLoS One *11*, e0158171.

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

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

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

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

Segata, N. (2015). Gut microbiome: westernization and disappearance of intestinal diversity. Curr. Biol. *25*, R611–613.

Sharon, G., Sampson, T.R., Geschwind, D.H., and Mazmanian, S.K. (2016). The central nervous system and the gut microbiome. Cell *167*, 915–932.

Solheim, M., Aakra, A., Snipen, L.G., Brede, D.A., and Nes, I.F. (2009). Comparative genomics of *Enterococcus faecalis* from healthy Norwegian infants. BMC Genomics *10*, 194.

Sonnenburg, E.D., and Sonnenburg, J.L. (2014). Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. Cell Metab. *20*, 779–786.

Sonnenburg, J.L., and Bäckhed, F. (2016). Diet-microbiota interactions as moderators of human metabolism. Nature *535*, 56–64.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., and Segata, N. (2015). MetaPhlAn2 for enhanced metagenomics taxonomic profiling. Nat. Methods *12*, 902–903.

Truong, D.T., Tett, A., Pasolli, E., Huttenhower, C., and Segata, N. (2017). Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res. *27*, 626–638.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., and Gordon, J.I. (2007). The human microbiome project. Nature *449*, 804–810.

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 *175*, 962–972.e10.

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

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.

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

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.

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*, 105–108.

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*, 222–227.

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

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*, 565–569.

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

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Healthy Italian adults' fecal samples	De Filippis et al., 2016a	N/A
Healthy Italian adults' fecal samples	This study	N/A
Hadzas' and Italians' fecal samples	Rampelli et al., 2015	N/A
Malawians', Amerindians' and Americans' fecal samples	Yatsunenko et al., 2012	N/A
Peruvians' and Americans' fecal samples	Obregon-Tito et al., 2015	N/A
Danes' fecal samples	Le Chatelier et al., 2013	N/A
Chemicals		
Nextera DNA Library Preparation kit	Illumina	FC-121-1031
Deposited Data		
Gut metagenomes (Italians, n=97)	This study	https://www.ncbi.nlm.nih.gov/sra; NCBI SRA: SRP126540; SRP083099.
Gut metagenomes (Hadza and Italians, n=38)	Rampelli et al., 2015	https://www.ncbi.nlm.nih.gov/sra; NCBI SRA: SRP056480.
Gut metagenomes (Malawian, Amerindians and Americans, n=110)	Yatsunenko et al., 2012	MG-RAST: qiime:621.
Gut metagenomes (Peruvians and Americans, n=58)	Obregon-Tito et al., 2015	https://www.ncbi.nlm.nih.gov/sra; NCBI SRA: PRJNA268964
Gut metagenomes (Danes, n=96)	Le Chatelier et al., 2013	European Nucleotide Archive: ERP003612
Software and Algorithms		
SolexaQA++	Cox et al., 2010	v.3.1.7.1
MetaPhlAn2	Truong et al., 2015	v.2.6
PanPhlAn	Scholz et al., 2016	v.1.0
RAxML	Stamatakis, 2014)	v.8.0
BLAST+	Camacho et al., 2009	v.2.3.0
MEGAN	Huson et al., 2016	v.5.0
SPAdes	Bankevich et al., 2012	v.3.9.0
MetaProdigal	Hyatt et al., 2012	v.2.6.3

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, and the protocol was approved by the Ethics Committees as previously reported (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). Informed consent was obtained from all subjects, the 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 adult subjects (46 males, 51 females): 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 SRA: 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., 2015 – 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., 2009; 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. (2015), 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 (the accession numbers for the sequences reported in this paper are NCBI SRA: SRP126540 and SRP083099).