



OPEN

A proposal for the reference intervals of the Italian microbiota “scaffold” in healthy adults

Davide Sisti^{1,10}, Valerio Paziienza^{2,10}, Fabio Piccini^{3,10}, Barbara Citterio¹, Wally Baffone¹, Sabrina Donati Zeppa¹, Francesca Biavasco⁴, Emilia Prospero⁵, Antonio De Luca⁶, Marco Artico⁷, Samanta Taurone⁷, Andrea Minelli¹, Francesco Perri², Elena Binda⁸, Riccardo Pracella⁸, Riccardo Santolini⁹, Stefano Amatori^{1✉}, Piero Sestili¹, Marco B. L. Rocchi^{1,11} & Pietro Gobbi^{1,11}

Numerous factors, ranging from genetics, age, lifestyle, and dietary habits to local environments, contribute to the heterogeneity of the microbiota in humans. Understanding the variability of a “healthy microbiota” is a major challenge in scientific research. The gut microbiota profiles of 148 healthy Italian volunteers were examined by 16S rRNA gene sequencing to determine the range and diversity of taxonomic compositions in the gut microbiota of healthy populations. Possible driving factors were evaluated through a detailed anamnestic questionnaire. Microbiota reference intervals were also calculated. A “scaffold” of a healthy Italian gut microbiota composition was identified. Differences in relative quantitative ratios of microbiota composition were detected in two clusters: a bigger cluster (C_2), which included 124 subjects, was characterized by more people from the northern Italian regions, who habitually practised more physical activity and with fewer dietary restrictions. Species richness and diversity were significantly higher in this cluster (C_2) than in the other one (C_1) (C_1 : 146.67 ± 43.67 ; C_2 : 198.17 ± 48.47 ; $F = 23.40$; $P < 0.001$ and C_1 : 16.88 ± 8.66 ; C_2 : 35.01 ± 13.40 ; $F = 40.50$; $P < 0.001$, respectively). The main contribution of the present study was the identification of the existence of a primary healthy microbiological framework that is only marginally affected by variations. Taken together, our data help to contextualize studies on population-specific variations, including marginal aspects, in human microbiota composition. Such variations must be related to the primary framework of a healthy microbiota and providing this perspective could help scientists to better design experimental plans and develop strategies for precision tailored microbiota modulation.

The microbial communities colonizing different body districts comprise trillions of microorganisms that perform vital functions and play a role in keeping us healthy¹. The human gut microbiota is composed of bacteria, archaea, fungi, protozoans, and viruses, particularly bacteriophages, probably due to the prevalence of bacteria in this environment². The microbiota can be viewed as a community composed of autochthonous or resident microorganisms and allochthonous or transient microorganisms³. Bacteria, mostly anaerobic, are predominant in this environment and consist of two main phylotypes: *Bacteroidetes*, including the genera *Prevotella* and *Bacteroides*, and *Firmicutes*, including *Clostridium* clusters and members of *Eubacterium*, *Faecalibacterium*, *Roseburia*, and *Ruminococcus*. On the other hand, *Proteobacteria*, *Actinobacteria*, *Fusobacteria*, and *Verrucomicrobia* phyla are present in relatively small numbers⁴. Various roles are attributed to the microbial community, including immune system maintenance, vitamin production, digestion, energy homeostasis, angiogenesis, metabolite synthesis, and the maintenance of intestinal barrier integrity⁵. The gut microbiota develops in children between the ages of one

¹Department of Biomolecular Sciences, University of Urbino Carlo Bo, Piazza Rinascimento 7, 61029 Urbino, Italy. ²Division of Gastroenterology, IRCCS “Casa Sollievo della Sofferenza” Hospital, 71013 San Giovanni Rotondo, Italy. ³Italian Microbiome Project, 35100 Padua, Italy. ⁴Department of Life and Environmental Sciences, Università Politecnica delle Marche, 60121 Ancona, Italy. ⁵Department of Biomedical Sciences, Università Politecnica delle Marche, 60121 Ancona, Italy. ⁶Department of Mental and Physical Health and Preventive Medicine, Section of Human Anatomy, University of Campania “Luigi Vanvitelli”, 80138 Naples, Italy. ⁷Department of Sense Organs, La Sapienza University, 00185 Rome, Italy. ⁸Cancer Stem Cells Unit ISBReMIT, IRCCS “Casa Sollievo della Sofferenza” Hospital, 71013 San Giovanni Rotondo, Italy. ⁹Department of Humanities, University of Urbino Carlo Bo, 61029 Urbino, Italy. ¹⁰These authors contributed equally: Davide Sisti, Valerio Paziienza and Fabio Piccini. ¹¹These authors jointly supervised this work: Marco B. L. Rocchi and Pietro Gobbi. ✉email: stefano.amatori1@uniurb.it

and three and remains relatively stable throughout life. During the transition from childhood to adulthood, the genera *Bifidobacteria* decrease, while *Bacteroidetes* increase, affecting the gut's metabolic activity and health. Several studies have reported that *Bifidobacteria* are beneficial, playing a role in protecting the gut epithelium, while *Firmicutes*, particularly *Clostridia* and *Enterobacteriaceae*, whose numbers increase in elderly subjects, are considered detrimental⁶.

Although microbiota evolves throughout the lifetime, it is now recognized that one-third of the gut microbiota is common to most people, whereas the remaining two-thirds are specific to each individual. In particular, at lower taxonomic levels (i.e., species), the microbiota is influenced by several individual factors, such as type of delivery at birth and the method of infant feeding, the use or abuse of medications, especially antibiotics, diet, supplements, lifestyle habits (smoking, physical activity) etc^{1,7,8}. Recent studies have highlighted emerging differences in microbiota composition even in population cohorts with similar genetic and cultural backgrounds^{9–11}. The balance between the two most important phyla found in the gut, *Firmicutes* and *Bacteroidetes*, is essential to maintaining homeostasis in the host and, consequently, health.

Considering the central role played by intestinal microbiota in modulating several pathological disorders and response to treatments, investigating their population-specific variations may lead to findings that contribute to enhancing the benefits of existing diagnostic and therapeutic strategies^{12–14}. Identifying and classifying specific sets of microbiota features that promote health is an essential first step to correcting microbial configurations implicated in disease. In fact, although attention is already being focused on how to manipulate microbiota to improve health status, there is a consensus within the scientific community that studying the factors constituting the normal ranges of these features in healthy populations is of fundamental importance¹⁵. Despite several important papers have been published on the microbiota composition of Italians suffering from specific diseases or as a changing ecosystem, few data are available on the microbiota of the healthy population^{1,9,16}. Hence, our study aimed to define the reference intervals of the gut microbiota of a sample of Italian subjects with relatively homogeneous physiological features.

Results

Samples of 148 participants (M: 69, F: 79; age: 39.8 ± 16.8 years; height: 164.4 ± 18.3 cm; weight: 61.2 ± 17.5 kg; body mass index: 22.0 ± 4.1 kg/m²) were collected and analysed. Of the 148 subjects, 22 were under 18 years old, 16 were smokers, 35 were ex-smokers, and 97 were non-smokers. In addition, 104 practised sports. Regarding diet, 93 participants had a typical Mediterranean diet, 15 were vegetarian, 4 were vegan, 4 were on a paleo diet, and the remaining 32 reported following another type of diet. Finally, 15 subjects reported avoiding eating certain foods for dietary or ethical reasons, and 27 had other kinds of sporadic food restrictions. The sample size was large enough for the aim of the paper. Indeed, the total number of genera found in the gut pan-microbiota increased with the number of samples, as shown in Supplementary Fig. S1.

Subject clustering. The analysis performed using the elbow and silhouette methods showed that the best clustering solution was the one with two clusters of subjects; both methods yielded the same results (Supplementary Fig. S2). The two clusters are clearly defined: Cluster 1 (C₁) comprises 24 subjects, while Cluster 2 (C₂) includes 124 subjects. Differences were observed between the two clusters in terms of phyla ($F=72.42$), families ($F=9.43$), and genera ($F=4.94$) (PERMANOVA, Bray–Curtis dissimilarity index, Bonferroni corrected; $P<0.001$ for all comparisons). Heatmap was used to represent the abundance of the most prevalent phyla. *Firmicutes* and *Bacteroidetes* together constituted more than 85% of the total phyla abundance. The difference between the two clusters, C₁ (Fig. 1 above) and C₂ (Fig. 1 below), is evident: *Bacteroidetes* were prevalent in C₁, while *Firmicutes* were prevalent in C₂.

General features of the population. The subjects in the two clusters had similar characteristics, as was shown by the comparison reported in Table 1. No significant differences were found between the two clusters in terms of gender, body mass index, diet type, smoking habits, or alcohol consumption. Regarding food restrictions, no significant difference was found for lactose "intolerance" (subject's self-definition, in the absence of any clinical diagnosis) between the two clusters. By contrast, a three-fold increase was found for other food restrictions in the subjects in C₁ (41.7% C₁ vs 13.7% C₂, $P=0.003$). The macro-region of origin (north vs centre vs south) also showed a difference in the distribution between the two clusters, with C₂ being comprising almost a doubled proportion of subjects living in northern Italy with respect to C₁ (62.1% vs 33.3%, respectively), and a lower percentage of subjects living in central Italy (29.8% vs 54.2%, respectively) ($P=0.03$). Finally, physical activity habits differed between the two clusters, with a higher proportion of subjects in C₂ who declared practising more than 4 h/week of physical activity with respect to C₁ (26.6% vs 8.3%, respectively; $P=0.04$).

The microbial diversity and richness of the two clusters. In addition to the presence of food restrictions in C₂, the cluster is also characterized by higher species richness (OTU number: C₁ = 146.67 ± 43.67 ; C₂ = 198.17 ± 48.47 ; $F_{(1,146)} = 23.40$; $P<0.001$) and diversity (Shannon effective number of species: C₁ = 16.88 ± 8.66 ; C₂ = 35.01 ± 13.40 ; $F_{(1,146)} = 40.50$; $P<0.001$) when compared to C₁.

Microbial community composition. The microbiota profile of 148 Italian volunteers were characterized at the taxonomic level (i.e., phylum, family, and genus; Table 2). The taxonomic assignment of the V3-V4 hypervariable region of the 16S rRNA gene showed *Firmicutes* (46.5%) and *Bacteroidetes* (43.2%) to be the predominant phyla. *Proteobacteria* (6.2%) were less abundant, while the remaining phyla were rarely detected, constituting altogether 4.2% of the total population. The analysis of the two clusters shows a higher percentage of *Bacteroidetes* (57.5% vs 40.4%) and a lower percentage of *Firmicutes* (26.9% vs 50.2%) in C₁ compared to

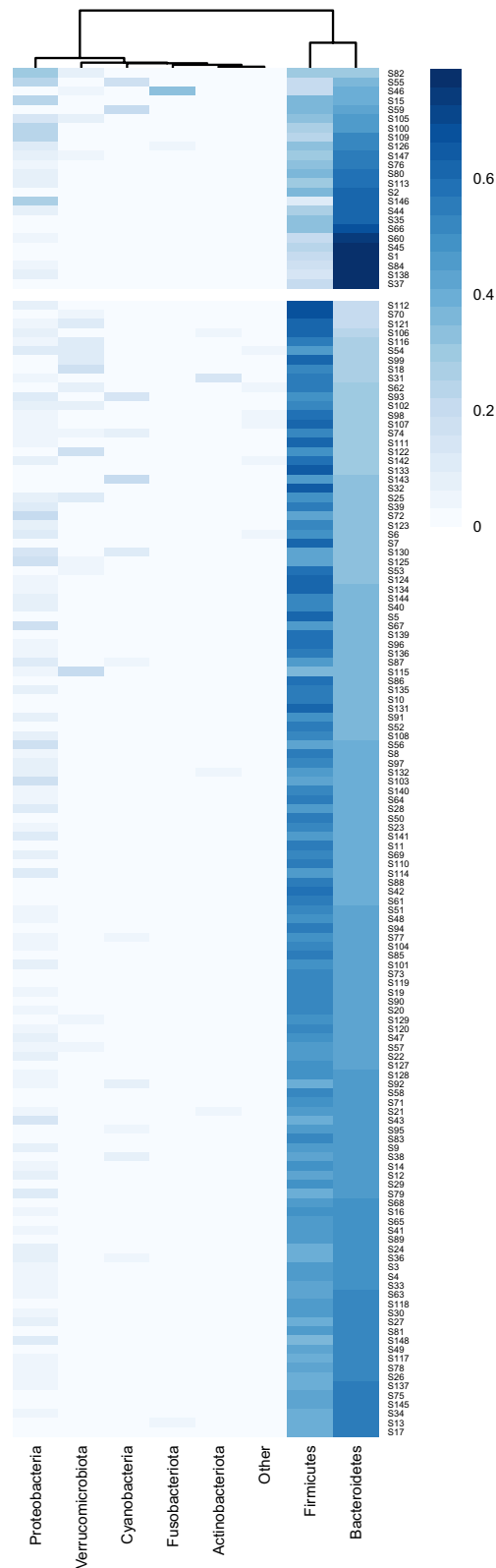


Figure 1. Heatmap of the most represented phyla. Each column represents a single phylum, with each row representing a different subject/sample. Bray–Curtis distance was used as a clustering method. Two different clusters of subjects were selected as the best solution (see above).

	Cluster 1 (N=24)	Cluster 2 (N=124)	p (Cramer's V)
Gender			
Male	12 (50.0%)	57 (46.0%)	0.72 (0.03)
Female	12 (50.0%)	67 (54.0%)	
Age			
< 18	4 (16.7%)	18 (14.5%)	0.68 (0.08)
18–65	20 (83.3%)	101 (81.5%)	
> 65	0 (0.0%)	5 (4.0%)	
Body mass index (kg/m²)			
< 18.5	4 (16.7%)	18 (14.5%)	1.00 (0.02)
18.5–25	15 (62.5%)	80 (64.5%)	
> 25	5 (20.8%)	26 (21.0%)	
Physical activity			
< 1 h/week	7 (29.2%)	45 (36.3%)	0.04 (0.21)
1–4 h/week	15 (62.5%)	46 (37.1%)	
> 4 h/week	2 (8.3%)	33 (26.6%)	
Region of residence			
North	8 (33.3%)	77 (62.1%)	0.03 (0.22)
Central	13 (54.2%)	37 (29.8%)	
South	3 (12.5%)	10 (8.1%)	
Type of residence			
City	15 (62.5%)	90 (72.6%)	0.33 (0.08)
Countryside	9 (37.5%)	34 (27.4%)	
Smoker			
Yes	4 (16.7%)	12 (9.7%)	0.38 (0.11)
Ex-smoker	7 (29.2%)	28 (22.6%)	
No	13 (54.2%)	84 (67.7%)	
Food restrictions			
Lactose restriction			
Yes	4 (16.7%)	11 (8.9%)	0.27 (0.10)
No	20 (83.3%)	113 (91.1%)	
Other food restrictions			
Yes	10 (41.7%)	17 (13.7%)	0.003 (0.27)
No	14 (58.3%)	107 (86.3%)	
Diet			
Mediterranean	12 (50.0%)	81 (65.3%)	0.29 (0.13)
Vegetarian/Vegan	5 (20.8%)	14 (11.3%)	
Other	7 (29.2%)	29 (23.4%)	
Breast feeding			
Yes	21 (87.5%)	102 (82.3%)	0.58 (0.05)
No	3 (12.5%)	22 (17.7%)	
Animals			
Yes	14 (58.3%)	55 (44.4%)	0.21 (0.10)
No	10 (41.7%)	69 (55.6%)	
Alcohol (last 2 days)			
Yes	9 (37.5%)	43 (34.7%)	0.79 (0.02)
No	15 (62.5%)	81 (65.3%)	

Table 1. The sociodemographic characteristics of the subjects in the two clusters. Significant values are in bold.

C_2 ; hence, there was a higher *Firmicutes/Bacteroidetes* ratio in C_2 than in C_1 , as shown in Fig. 2 (*F/B* Ratio: C_1 0.51 ± 0.22 , C_2 1.33 ± 0.48 ; Mann–Whitney $U = 73.0$; $P < 0.001$).

At the family level, *Bacteroidaceae* (25.9%), *Lachnospiraceae* (21.8%), *Prevotellaceae* (10.5%), *Ruminococcaceae* (9.9%), and *Oscillospiraceae* (2.5%) were the most representative among the over two-hundred families that were detected. Among these, significant differences were detected between the two clusters for *Lachnospiraceae*, *Ruminococcaceae*, and *Oscillospiraceae*, all the three being more abundant in C_2 with respect to C_1 . Among the genera, *Bacteroides* (25.9%), *Prevotella* (8.0%), and *Faecalibacterium* (4.1%) showed the highest percentages of abundance.

	Total	Cluster 1	Cluster 2
Phyla			
<i>P. Firmicutes</i>	46.45% (18.05–66.11)	26.99% (3.75–37.38)	50.21% (36.53–66.01)
<i>P. Bacteroidetes</i>	43.17% (20.94–89.29)	57.55% (21.76–80.83)	40.38% (17.8–56.85)
<i>P. Proteobacteria</i>	6.15% (0.56–30.17)		
<i>P. Verrucomicrobiota</i>	1.6% (0–17.41)		
<i>P. Cyanobacteria</i>	1.16% (0–21.28)		
<i>P. Actinobacteriota</i>	0.6% (0–4.74)	0.3% (0–3.16)	0.66% (0–4.83)
<i>P. Fusobacteriota</i>	0.29% (0–2.45)		
<i>P. Euryarchaeota</i>	0.24% (0–4.27)		
<i>P. Desulfobacterota</i>	0.17% (0–1.59)		
Families			
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Bacteroidaceae</i>	25.93% (1.09–67.09)		
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae</i>	21.8% (3.22–41.47)	13.55% (3.32–23.15)	23.39% (4.42–41.24)
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Prevotellaceae</i>	10.45% (0–73.71)		
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Ruminococcaceae</i>	9.96% (0.56–27.93)	6.08% (0–13.42)	10.71% (0.32–28.56)
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Oscillospiraceae</i>	2.53% (0.01–15.98)	0.98% (0–5.43)	2.83% (0.12–15.49)
<i>P. Proteobacteria; C. Gammaproteobacteria; O. Burkholderiales; F. Sutterellaceae</i>	2.26% (0–10.88)		
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Tannerellaceae</i>	2.21% (0–12.91)		
<i>P. Proteobacteria; C. Alphaproteobacteria; O. Rhodospirillales; F. uncultured</i>	2.18% (0–17.76)		
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Rikenellaceae</i>	2.16% (0–9.91)		
<i>P. Firmicutes; C. Clostridia; O. Clostridia_UCG-014; F. Clostridia_UCG-014</i>	1.94% (0–13.12)	0.64% (0–4.63)	2.19% (0–13.12)
<i>P. Verrucomicrobiota; C. Verrucomicrobiae; O. Verrucomicrobiales; F. Akkermansiaceae</i>	1.51% (0–17.41)		
<i>P. Firmicutes; C. Negativicutes; O. Acidaminococcales; F. Acidaminococcaceae</i>	1.28% (0–8.02)		
<i>P. Proteobacteria; C. Gammaproteobacteria; O. Enterobacteriales; F. Enterobacteriaceae</i>	1.24% (0–31.06)		
<i>P. Cyanobacteria; C. Vampirivibrionia; O. Gastranaerophilales; F. Gastranaerophilales</i>	1.15% (0–21.28)		
<i>P. Firmicutes; C. Clostridia; O. Clostridia_vadinBB60_group; F. Clostridia_vadinBB60_group</i>	1.15% (0–8.73)	0.24% (0–1.85)	1.32% (0–9.76)
<i>P. Firmicutes; C. Clostridia; O. Christensenellales; F. Christensenellaceae</i>	1.1% (0–7.76)	0.3% (0–6.38)	1.25% (0–8.44)
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Barnesiellaceae</i>	1.03% (0–5.14)		
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. [Eubacterium]_coprostanoligenes_group</i>	0.88% (0–5.58)	0.39% (0–2.72)	0.97% (0–6.59)
<i>P. Firmicutes; C. Negativicutes; O. Veillonellales-Selenomonadales; F. Veillonellaceae</i>	0.87% (0–6.53)		
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Marinifilaceae</i>	0.75% (0–2.38)		
<i>P. Firmicutes; C. Bacilli; O. Erysipelotrichales; F. Erysipelatoclostridiaceae</i>	0.74% (0.02–5.48)	0.28% (0–1.88)	0.83% (0.01–8.81)
<i>P. Firmicutes; C. Clostridia; O. Monoglobales; F. Monoglobaceae</i>	0.7% (0–3.66)		
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. UCG-010</i>	0.68% (0–5.74)	0.3% (0–9.27)	0.76% (0–6.48)
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Muribaculaceae</i>	0.56% (0–6.46)		
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Butyrivibrionaceae</i>	0.48% (0.02–2.08)	0.24% (0–1.07)	0.53% (0.01–2.04)
<i>P. Actinobacteriota; C. Actinobacteria; O. Bifidobacteriales; F. Bifidobacteriaceae</i>	0.48% (0–3.67)	0.22% (0–2.71)	0.53% (0–4.24)
<i>P. Firmicutes; C. Bacilli; O. Erysipelotrichales; F. Erysipelotrichaceae</i>	0.36% (0–3.93)		
<i>P. Firmicutes; C. Bacilli; O. Izemoplasmatales; F. Izemoplasmatales</i>	0.35% (0–4.88)	0.02% (0–0.44)	0.41% (0–6.95)
<i>P. Firmicutes; C. Clostridia; O. Peptostreptococcales-Tissierellales; F. Peptostreptococcaceae</i>	0.3% (0–1.96)		
<i>P. Fusobacteriota; C. Fusobacteriia; O. Fusobacteriales; F. Fusobacteriaceae</i>	0.29% (0–2.45)		
<i>P. Proteobacteria; C. Gammaproteobacteria; O. Pasteurellales; F. Pasteurellaceae</i>	0.26% (0–2.86)		
<i>P. Euryarchaeota; C. Methanobacteria; O. Methanobacteriales; F. Methanobacteriaceae</i>	0.24% (0–4.19)		
<i>P. Firmicutes; C. Clostridia; O. Clostridiales; F. Clostridiaceae</i>	0.23% (0–1.49)		
<i>P. Firmicutes; C. Bacilli; O. Acholeplasmatales; F. Acholeplasmataceae</i>	0.23% (0–4.75)		
<i>P. Firmicutes; C. Bacilli; O. Lactobacillales; F. Streptococcaceae</i>	0.19% (0–1.65)		
<i>P. Firmicutes; C. Bacilli; O. RF39; F. RF39</i>	0.18% (0–2.99)	0.02% (0–0.41)	0.21% (0–3.06)
<i>P. Desulfobacterota; C. Desulfovibrionia; O. Desulfovibrionales; F. Desulfovibrionaceae</i>	0.17% (0–1.57)		
<i>P. Proteobacteria; C. Gammaproteobacteria; O. Aeromonadales; F. Succinivibrionaceae</i>	0.15% (0–0.05)		
<i>P. Firmicutes; C. Clostridia; O. Peptostreptococcales-Tissierellales; F. Anaerovoracaceae</i>	0.14% (0–0.93)		
Genera			
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Bacteroidaceae; G. Bacteroides</i>	25.93% (1.09–67.09)		
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Prevotellaceae; G. Prevotella</i>	7.98% (0–70.33)		
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Ruminococcaceae; G. Faecalibacterium</i>	4.07% (0.34–20.59)		
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Unclassified</i>	2.61% (0.08–7.66)	1.51% (0–4.36)	2.82% (0.31–7.83)
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Roseburia</i>	2.58% (0–12.04)		

Continued

	Total	Cluster 1	Cluster 2
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Lachnospira</i>	2.57% (0–15.38)		
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Tannerellaceae; G. Parabacteroides</i>	2.21% (0–12.78)		
<i>P. Proteobacteria; C. Alphaproteobacteria; O. Rhodospirillales; F. uncultured; G. uncultured</i>	2.18% (0–17.76)		
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Blautia</i>	2.02% (0.31–7.38)		
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Rikenellaceae; G. Alistipes</i>	1.97% (0–9.89)		
<i>P. Firmicutes; C. Clostridia; O. Clostridia_UCG-014; F. Clostridia_UCG-014; G. Clostridia_UCG-014</i>	1.94% (0–13.28)	0.64% (0–4.63)	2.19% (0–13.12)
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Ruminococcaceae; G. Ruminococcus</i>	1.9% (0–12.56)	1.1% (0–12.59)	2.05% (0–16.77)
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Lachnospiraceae_NK4A136_group</i>	1.83% (0–11.57)	0.97% (0–7.16)	2% (0–11.99)
<i>P. Verrucomicrobiota; C. Verrucomicrobiae; O. Verrucomicrobiales; F. Akkermansiaceae; G. Akkermansia</i>	1.51% (0–17.41)		
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. [Eubacterium]_eligens_group</i>	1.42% (0–8.07)	0.75% (0–4.08)	1.55% (0–8.3)
<i>P. Proteobacteria; C. Gammaproteobacteria; O. Burkholderiales; F. Sutterellaceae; G. Sutterella</i>	1.36% (0–6.15)		
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Agathobacter</i>	1.16% (0–6.4)		
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Coprococcus</i>	1.15% (0–7.53)	0.47% (0–2.96)	1.28% (0–7.98)
<i>P. Cyanobacteria; C. Vampirivibrionia; O. Gastranaerophilales; F. Gastranaerophilales; G. Gastranaerophilales</i>	1.15% (0–21.28)		
<i>P. Firmicutes; C. Negativicutes; O. Acidaminococcales; F. Acidaminococcaceae; G. Phascolarctobacterium</i>	1.15% (0–8.55)		
<i>P. Firmicutes; C. Clostridia; O. Clostridia_vadinBB60_group; F. Clostridia_vadinBB60_group; G. Clostridia_vadinBB60_group</i>	1.15% (0–8.73)	0.24% (0–1.85)	1.32% (0–9.76)
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Ruminococcaceae; G. [Eubacterium]_siraeum_group</i>	1.13% (0–9.32)		
<i>P. Proteobacteria; C. Gammaproteobacteria; O. Enterobacteriales; F. Enterobacteriaceae; G. Escherichia-Shigella</i>	1.13% (0–28.31)		
<i>P. Firmicutes; C. Clostridia; O. Christensenellales; F. Christensenellaceae; G. Christensenellaceae_R-7_group</i>	1.08% (0–7.63)	0.29% (0–6.35)	1.23% (0–8.17)
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Prevotellaceae; G. Prevotellaceae_NK3B31_group</i>	1.07% (0–18.99)		
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Oscillospiraceae; G. UCG-002</i>	1.04% (0–9.18)	0.3% (0–2.77)	1.18% (0–11.44)
<i>P. Proteobacteria; C. Gammaproteobacteria; O. Burkholderiales; F. Sutterellaceae; G. Parasutterella</i>	0.91% (0–12.06)		
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. [Eubacterium]_coprostanoligenes_group; G. [Eubacterium]_coprostanoligenes_group</i>	0.88% (0–5.58)	0.39% (0–2.72)	0.97% (0–6.59)
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Ruminococcaceae; G. Subdoligranulum</i>	0.84% (0–4.84)	0.33% (0–3.53)	0.94% (0–4.88)
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Barnesiellaceae; G. Barnesiella</i>	0.79% (0–3.99)		
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Fusicatenibacter</i>	0.73% (0–4.78)	0.31% (0–3.19)	0.82% (0.01–5.18)
<i>P. Firmicutes; C. Negativicutes; O. Veillonellales-Selenomonadales; F. Veillonellaceae; G. Dialister</i>	0.7% (0–6.14)		
<i>P. Firmicutes; C. Clostridia; O. Monoglobales; F. Monoglobaceae; G. Monoglobus</i>	0.7% (0–3.65)		
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. UCG-010; G. UCG-010</i>	0.68% (0–5.74)	0.3% (0–9.27)	0.76% (0–6.5)
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Ruminococcaceae; G. Unclassified</i>	0.68% (0–5.09)		
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Ruminococcaceae; G. CAG-352</i>	0.65% (0–7.56)		
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Prevotellaceae; G. Alloprevotella</i>	0.62% (0–16.06)		
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. [Eubacterium]_ruminantium_group</i>	0.59% (0–8.19)		
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Anaerostipes</i>	0.55% (0–3.16)	0.26% (0–0.69)	0.61% (0.01–3.43)
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Muribaculaceae; G. Muribaculaceae</i>	0.55% (0–6.47)		
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Dorea</i>	0.54% (0–1.88)		
<i>P. Actinobacteriota; C. Actinobacteria; O. Bifidobacteriales; F. Bifidobacteriaceae; G. Bifidobacterium</i>	0.48% (0–3.64)	0.22% (0–2.71)	0.53% (0–4.24)
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Lachnoclostridium</i>	0.46% (0–2.32)		
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Marinifilaceae; G. Odoribacter</i>	0.46% (0–1.76)		
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Prevotellaceae; G. Paraprevotella</i>	0.46% (0–3.83)		
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Oscillospiraceae; G. UCG-005</i>	0.45% (0–2.71)	0.17% (0–3.99)	0.5% (0–2.64)
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Butyricicoccaceae; G. Butyricoccus</i>	0.43% (0–2.16)	0.24% (0–1.07)	0.47% (0–2.19)
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Ruminococcaceae; G. uncultured</i>	0.39% (0–2.45)		
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Lachnospiraceae_UCG-001</i>	0.39% (0–3.69)	0.14% (0–1.19)	0.44% (0–3.78)
<i>P. Firmicutes; C. Bacilli; O. Erysipelotrichales; F. Erysipelatoclostridiaceae; G. Erysipelotrichaceae_UCG-003</i>	0.38% (0–2.04)	0.18% (0–1.96)	0.42% (0–2)
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Lachnospiraceae_UCG-003</i>	0.37% (0–10.84)		
<i>P. Firmicutes; C. Bacilli; O. Izemoplasmatales; F. Izemoplasmatales; G. Izemoplasmatales</i>	0.35% (0–4.86)	0.02% (0–0.44)	0.41% (0–6.95)
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. [Eubacterium]_xylanophilum_group</i>	0.35% (0–1.58)	0.1% (0–1.35)	0.4% (0–1.62)
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Oscillospiraceae; G. UCG-003</i>	0.33% (0–1.4)	0.14% (0–1.49)	0.36% (0–1.52)
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. Butyrivibrio</i>	0.32% (0–15.4)		
<i>P. Firmicutes; C. Clostridia; O. Lachnospirales; F. Lachnospiraceae; G. [Eubacterium]_hallii_group</i>	0.3% (0–0.97)	0.16% (0–0.86)	0.32% (0.01–1.1)
<i>P. Fusobacteriota; C. Fusobacteriia; O. Fusobacteriales; F. Fusobacteriaceae; G. Fusobacterium</i>	0.29% (0–2.45)		
<i>P. Bacteroidetes; C. Bacteroidia; O. Bacteroidales; F. Marinifilaceae; G. Butyricimonas</i>	0.28% (0–1.17)		
<i>P. Firmicutes; C. Clostridia; O. Oscillospirales; F. Oscillospiraceae; G. NK4A214_group</i>	0.27% (0–3)	0.04% (0–0.32)	0.31% (0–4.19)
Continued			

	Total	Cluster 1	Cluster 2
<i>P: Firmicutes; C: Clostridia; O: Lachnospirales; F: Lachnospiraceae; G: Tyzzerella</i>	0.27% (0–4.34)		
<i>P: Firmicutes; C: Clostridia; O: Lachnospirales; F: Lachnospiraceae; G: [Ruminococcus]_torques_group</i>	0.26% (0.01–2)		
<i>P: Proteobacteria; C: Gammaproteobacteria; O: Pasteurellales; F: Pasteurellaceae; G: Haemophilus</i>	0.25% (0–2.78)		
<i>P: Firmicutes; C: Clostridia; O: Lachnospirales; F: Lachnospiraceae; G: [Eubacterium]_ventriosum_group</i>	0.24% (0–1.32)	0.1% (0–1.35)	0.27% (0–1.34)
<i>P: Firmicutes; C: Bacilli; O: Acholeplasmatales; F: Acholeplasmataceae; G: Anaeroplasmata</i>	0.23% (0–4.75)		
<i>P: Firmicutes; C: Clostridia; O: Clostridiales; F: Clostridiaceae; G: Clostridium_sensu_stricto_1</i>	0.23% (0–1.5)		
<i>P: Firmicutes; C: Bacilli; O: Erysipelotrichales; F: Erysipelotrichaceae; G: Holdemanella</i>	0.23% (0–2.68)		
<i>P: Euryarchaeota; C: Methanobacteria; O: Methanobacteriales; F: Methanobacteriaceae; G: Methanobrevibacter</i>	0.22% (0–4.14)		
<i>P: Firmicutes; C: Clostridia; O: Lachnospirales; F: Lachnospiraceae; G: Lachnospiraceae_UCG-004</i>	0.19% (0–1.05)		
<i>P: Firmicutes; C: Bacilli; O: Erysipelotrichales; F: Erysipelotrichaceae; G: Asteroleplasma</i>	0.19% (0–5.16)		
<i>P: Firmicutes; C: Bacilli; O: Lactobacillales; F: Streptococcaceae; G: Streptococcus</i>	0.18% (0–1.66)		
<i>P: Bacteroidetes; C: Bacteroidia; O: Bacteroidales; F: Rikenellaceae; G: Rikenellaceae_RC9_gut_group</i>	0.18% (0–2.95)		
<i>P: Firmicutes; C: Bacilli; O: RF39; F: RF39; G: RF39</i>	0.18% (0–2.92)	0.02% (0–0.41)	0.21% (0–3.06)
<i>P: Firmicutes; C: Clostridia; O: Lachnospirales; F: Lachnospiraceae; G: Lachnospiraceae_ND3007_group</i>	0.17% (0–0.79)	0.06% (0–0.37)	0.19% (0–0.94)
<i>P: Firmicutes; C: Clostridia; O: Lachnospirales; F: Lachnospiraceae; G: Lachnospiraceae_UCG-010</i>	0.17% (0–0.87)	0.08% (0–0.98)	0.18% (0–1.02)
<i>P: Firmicutes; C: Clostridia; O: Oscillospirales; F: Ruminococcaceae; G: Incertae_Sedis</i>	0.15% (0–1.14)	0.08% (0–0.51)	0.17% (0–1.26)
<i>P: Bacteroidetes; C: Bacteroidia; O: Bacteroidales; F: Prevotellaceae; G: uncultured</i>	0.15% (0–3.18)		
<i>P: Firmicutes; C: Clostridia; O: Lachnospirales; F: Lachnospiraceae; G: CAG-56</i>	0.15% (0–1.07)		
<i>P: Firmicutes; C: Clostridia; O: Peptostreptococcales-Tissierellales; F: Peptostreptococcaceae; G: Romboutsia</i>	0.15% (0–1.1)	0.07% (0–0.49)	0.16% (0–1.42)
<i>P: Bacteroidetes; C: Bacteroidia; O: Bacteroidales; F: Prevotellaceae; G: Prevotellaceae_UCG-001</i>	0.14% (0–3.37)		
<i>P: Proteobacteria; C: Gammaproteobacteria; O: Aeromonadales; F: Succinivibrionaceae; G: Succinivibrio</i>	0.14% (0–0)		
<i>P: Bacteroidetes; C: Bacteroidia; O: Bacteroidales; F: Barnesiellaceae; G: uncultured</i>	0.12% (0–1.67)		
<i>P: Firmicutes; C: Negativicutes; O: Acidaminococcales; F: Acidaminococcaceae; G: Acidaminococcus</i>	0.12% (0–2.86)		
<i>P: Firmicutes; C: Clostridia; O: Oscillospirales; F: Oscillospiraceae; G: Colidextribacter</i>	0.12% (0–0.59)		
<i>P: Bacteroidetes; C: Bacteroidia; O: Bacteroidales; F: Barnesiellaceae; G: Coprobacter</i>	0.11% (0–0.92)		
<i>P: Firmicutes; C: Clostridia; O: Oscillospirales; F: Oscillospiraceae; G: Oscillibacter</i>	0.11% (0–0.58)		
<i>P: Firmicutes; C: Clostridia; O: Peptostreptococcales-Tissierellales; F: Peptostreptococcaceae; G: Intestinibacter</i>	0.11% (0–1.03)		
<i>P: Firmicutes; C: Clostridia; O: Lachnospirales; F: Lachnospiraceae; G: [Ruminococcus]_gavvreauii_group</i>	0.1% (0–0.66)		

Table 2. Mean, upper and lower limits of 95% reference intervals are reported for phyla, families, and genera with abundance > 0.1%. The reference intervals are reported for C₁ and C₂ if the abundance between the two clusters was significantly different (Mann–Whitney test with false discovery rate). *P* phylum, *C* class, *O* order, *F* family, *G* genus.

In Fig. 3, non-parametric correlation matrices are reported for all genera that showed an abundance greater than 0.5%, both for C₁ (Fig. 3, upper plot) and C₂ (Fig. 3, lower plot). In C₁, some genera showed a strong negative correlation, as for *Prevotella* with *Bacteroides*, as well as *Sutterella* with *Lachnospira*, *Monoglobus*, and *Parasutterella*. On the contrary, some genera were found to be positively co-graduated. Overall, correlations detected in C₁ should be interpreted with caution, as this cluster only comprises 24 subjects. As can be noted, C₂ showed weaker correlations among genera with respect to C₁, except for strongly negative values between genus *Bacteroides* and genus *Clostridia* UCG-014, and again, between *Bacteroides* and *Prevotella*.

Discussion

Defining a healthy microbiota has become a major challenge for scientists. Although variations in the microbial community associated with a wide range of pathologies, from gastrointestinal disorders, autoimmune diseases, and cancer to mood disorders, have been documented, the composition and functional characteristics of a healthy microbiota have yet to be fully elucidated^{12,13,17–20}. Defining the composition and functional characteristics of a healthy microbiota is so challenging because the microbiota is strongly influenced by a wide range of factors, including genetics; the mode of delivery at birth and the method of infant feeding; the use or abuse of medications and supplements, especially antibiotics, diet, lifestyle habits (smoking, physical activity), etc^{1,7,8}. Recent studies have highlighted differences in microbiota composition even in population cohorts with similar genetic and cultural backgrounds^{9–11}.

This study aimed to characterize the gut microbiota composition of healthy volunteers from Italy. The analysis of faecal samples provided information on the microbial composition that was in line with previously reported results²¹. The phyla *Firmicutes* and *Bacteroidetes* are indeed closely related to a healthy profile as well as a low number of species belonging to *Proteobacteria* phyla due to the anaerobic conditions in the colon²². Nevertheless, by k-means clustering, we were able to identify two main groups, C₁ and C₂, with the latter characterized by higher richness, diversity, and *Firmicutes/Bacteroidetes* ratio.

It is well known that most bacterial genera of the human gut microbiota belong to the phyla *Firmicutes* or *Bacteroidetes*, which account for about 90% of intestinal resident microorganisms²³. *Firmicutes*, sub-grouped in

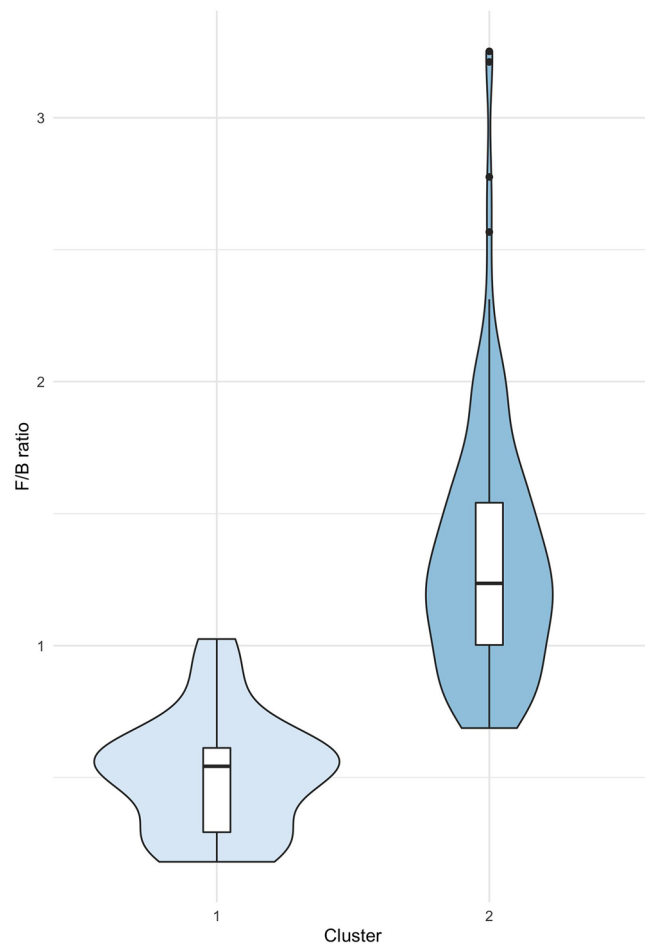


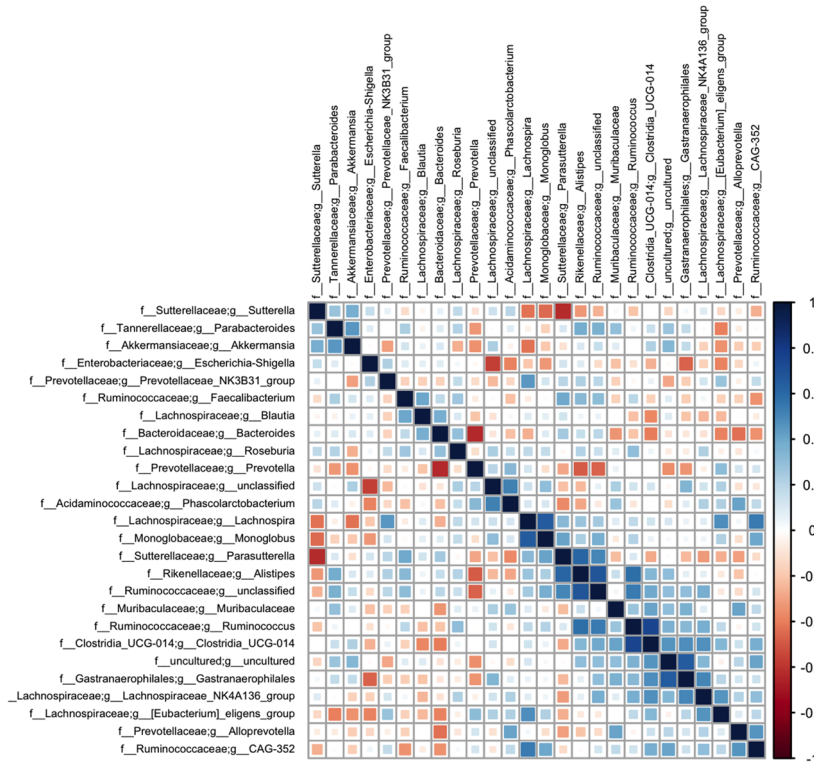
Figure 2. Violin plots of the *Firmicutes/Bacteroidetes* ratio in the two clusters.

Clostridium coccooides (*Clostridium* cluster XIVa) and *Clostridium leptum* (*Clostridium* cluster IV), are responsible for assimilating carbohydrates and animal fat, which are associated with the onset of obesity^{24,25}. Among *Bacteroidetes*, the two prevalent genera in the human colon are *Bacteroides* and *Prevotella*; the former is highly associated with the consumption of animal proteins, amino acids, and saturated fats, which are typical components of the Western diet, and the latter with the consumption of complex carbohydrates and simple sugars, which are important components of vegetarian diets^{26,27}.

Although several studies have attempted to define the composition of a healthy microbiota, such a definition remains elusive due to the many intrinsic and extrinsic factors influencing the gut ecosystem²⁸. Nishijima et al.²⁹ compared the compositions of the gut microbiomes of people from twelve different countries worldwide, showing great variations in the microbiome structure and function in healthy adults from different countries. At the genus level, the relative abundance of *Bacteroides* reported herein for the Italian population was similar to that reported for the United States, Canada, and Spain; similarly, the relative abundance of *Prevotella* was analogous to that which was reported for Canada, Denmark, Spain, and Russia. It can be noted that the microbiota composition found in this study in the healthy Italian population is similar to the composition reported in other countries with a predominantly Caucasian ethnicity²⁹. The additional value of our study compared to similar investigations consists of quantifying reference intervals, which could have a direct application in diagnostics. Moreover, De Filippo et al. reported a significant enrichment in *Bacteroidetes* and a depletion of *Firmicutes* in African children whose diet was based on cereals, legumes, and vegetables and rich in carbohydrates, fibre, and non-animal protein³⁰. The bacteria belonging to the genus *Bacteroides* are known to produce short-chain fatty acids (SCFA) and may thus contribute to preventing gut inflammation. Accordingly, multiple studies have reported an association between inflammatory bowel disease (IBD) and the flora disequilibrium of *Bacteroides*^{31–34}.

In our study, *Firmicutes* showed a lower abundance than *Bacteroidetes* in C₁. The greater relative abundance of *Bacteroidetes* suggests that in the intestines of those subjects, there may be a lower number of bacterial species favouring the onset of metabolic diseases such as those belonging to *Firmicutes*. In this regard, we also observed that the participants in this study, grouped according to their food habits and, to a lesser degree, according to their geographical origins, are mostly included in C₂. Indeed, a significantly higher proportion of participants in C₁ reported having food restrictions, mainly related to the consumption of dried and fresh vegetables. A limited consumption, or even worse, the exclusion of fresh vegetables and their derivatives from the diet, can induce alteration in the gut microbiota composition, leading to a reduction of fibre-degrading bacteria, able to produce

C₁



C₂

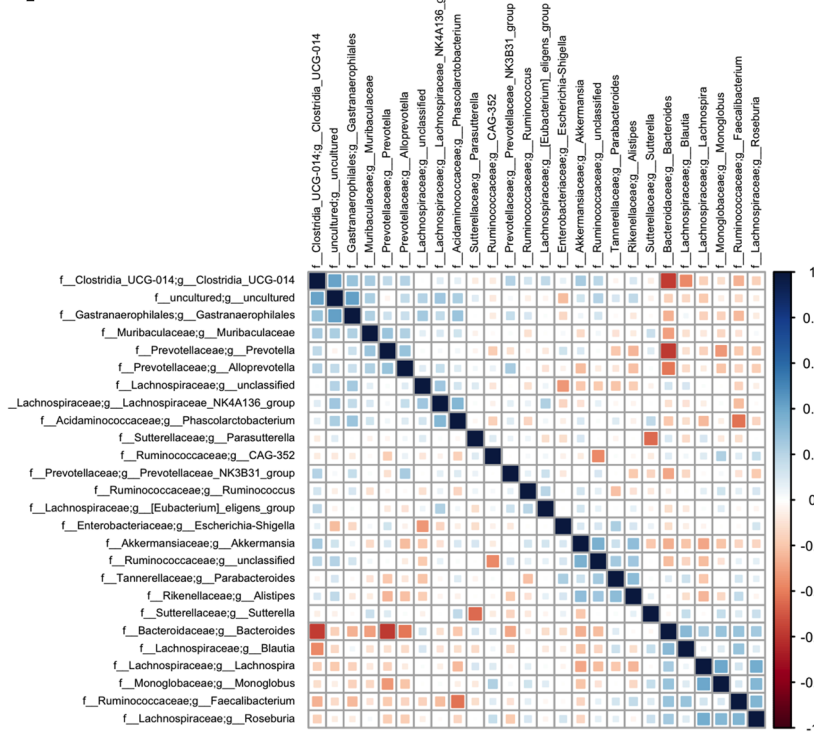


Figure 3. Spearman correlation plots of genera (cut off: relative abundance 0.5%) in the two clusters.

greater amounts of SCFA^{35,36}. Further, the heterogeneity of vegetables consumed has been positively correlated to the microbial alpha-diversity³⁷. In addition, a significant difference in the distribution of the subjects among

the Italian macro-regions (north, centre, south) was detected, with participants from the northern regions being more represented in the C₂ and participants from the central regions being prevalent in the C₁. This result is in accordance with those previously reported by Fontana and colleagues⁹, who detected differences in the gut microbiota composition of people pertaining to three different regions of Italy (one for each macro-area). Further, regular physical activity seems to be a discerning factor between the two clusters, with people who practice a higher volume of physical exercise (i.e., more than 4 h per week) being more present in the C₂. This result is also coherent with a higher *Firmicutes/Bacteroidetes* ratio in this cluster, as this ratio was previously associated with higher cardiorespiratory fitness and athletic status³⁸. This result is in accordance with other studies in the literature: Clarke et al.³⁹ found a higher *Firmicutes/Bacteroidetes* ratio in a sample of rugby players compared to overweight controls; Huang et al.⁴⁰ reported an increase in this ratio after six weeks of exercise and dietary restriction in obese adolescents; Donati Zeppa et al.⁴¹ found an increase in the *Firmicutes/Bacteroidetes* ratio after nine weeks of high-intensity interval training in healthy males.

In conclusion, this study further supports the significance of *Firmicutes* and *Bacteroidetes* and their ratio as a scaffold of a microbiologically healthy gut and, consequently, of the body's wellbeing, similarly to a conventional human structural organ. Reference intervals at every taxonomic level can be used as a reference for verifying whether a single subject or sample belongs to the microbiological community defined in this study. Reference intervals here reported refer to a sample of subjects who do not self-report clear disease symptoms; however, it will be interesting to use these reference intervals for future studies to verify whether samples from patients with specific diseases may or may not have a different microbial community. In this sense, it will then be possible to define the sensitivity and specificity of the reported reference intervals according to specific pathologies. This can be done either at the univariate level (e.g., genus by genus) to verify whether each specific abundance is included within the reference interval or using a multivariate approach. In the latter case, the non-parametric approach of the Mahalanobis distance can be used. In this approach, the ranks of the abundances are used rather than the single absolute values⁴².

The association between the richness and diversity of gut microbiota and health has been demonstrated by Rinninella et al.¹, although it appears difficult to identify a unique optimal gut microbiota composition. The main contribution of the present study is to help identify the existence, within the healthy Italian population, of a commonly distributed, constantly present, principal microbiological pattern, thus suggesting the presence of a sort of microbiological framework or scaffolding. In our view, this finding reinforces the concept that the human intestinal microbiota, with its morpho-functional and pathophysiological aspects, represents a real organ. Indeed, like anatomical organs, the intestinal microbiota may have individual variability; however, such variability must not substantially alter the fundamental framework of a healthy microbiota to ensure its correct functioning. Possible differences in the gut microbiota composition, diversity, and richness among individuals with the same ethnicity, residing in different Italian regions, or with different lifestyles only marginally affect the composition of the two main microbiological clusters identified in the present study. Taken together, our data highlight the significance of studies on population-specific variations in human microbiota composition. Nevertheless, at the same time, the present investigation underscores the need for variability studies to be able to consider even minimal variations in the intestinal microbiological population, given that, at least in the healthy population, there is a significant and reproducible presence of well-defined groups of bacteria, which represent a constant scaffold, or framework. In the next future, it will be crucial to share these data coming from different research groups in order to implement the "normal" range values or to build an algorithm capable of translating the composition of the microbiota associated with diseases states and of suggesting any dietary, pharmacological or lifestyle interventions in order to recover the state of eubiosis. Moreover, integrating these data with metabolomics and genetic variants could improve patient management. Again, this approach to studying the intestinal microbiota calls to mind studies focusing on human structural organs. Indeed, in our view, such an approach to the microbiota could help scientists to better design experimental plans and set up strategies based on precision tailored microbiota engineering.

Methods

Participants. A total of 148 control subjects from 17 Italian regions were recruited by the medical board, some of whose members were contributing authors of this work. The subjects, 69 males and 79 females, ranging in age from 23 to 57, were recruited from different Italian universities under the supervision of the University of Urbino Carlo Bo (Ethics Committee approval no. 34_2021). All the participating institutions followed the same pre-analytical and analytical procedures. All the subjects agreed to participate according to the ethical guidelines of the 2013 Declaration of Helsinki and signed written informed consent. The volunteer participants were selected to create a model of the Italian adult Caucasian population adequately represented in terms of gender, age, geographical origin, and place of residence (city or countryside) and that falls within the criteria of WHO definition of a "healthy" state of "complete physical, mental, and social well-being, not merely the absence of disease or infirmity". In detail, the medical board evaluated each subject's complete medical history in order to exclude those who did not meet the study's inclusion criteria. The following subjects were excluded: those being treated with antibiotics or other drugs, those consuming probiotics, and those having a known history of inflammatory bowel disease, systemic disease, other autoimmune, metabolic, or psychiatric disorders or cancer. A questionnaire was then administered to each participant to collect the following information: body mass index, dietary habits, contact with farm animals or pets, smoking and physical activity habits, alcohol consumption, breastfeeding). Dietary patterns were classified into the Mediterranean, vegetarian/vegan, and others; the routine use of probiotics was also assessed. Furthermore, lactose and other food restrictions were evaluated, i.e., voluntary limited consumption or exclusion of specific food groups (mainly dried or fresh vegetables and derivatives).

Sample collection and DNA extraction. Samples were collected over two years, from fall 2017 to spring 2019. Fresh stool samples were collected within tubes containing a DNA stabilization buffer (Canvax Biotech) from each participant. In order to reduce any possible bias, pre-analytical and analytical procedures were performed at only one centre, according to our previously published study⁹. QIAamp DNA Stool Mini Kit (Qiagen, Milan, Italy) was utilized to perform total DNA extraction starting from 250 μ L of each sample following the manufacturer's protocol. Once collected in the stabilizing liquid, samples were processed according to standardized times, usually not exceeding 5 days from the withdrawal; when was not possible to process the samples immediately at the arrival in the laboratory, they were stored at -80° until processing, after assessing DNA concentration and purity.

16S rRNA gene sequence data processing. The Illumina 16S Metagenomic Sequencing Library Preparation for high-throughput sequencing was performed as follows: 12.5 ng of each DNA extract was employed for the amplification of the V3–V4 hypervariable regions of the bacterial 16S ribosomal RNA (rRNA) gene, using the following primers with Illumina adapters (underlined):

Forward primer (341F):

5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG

Reverse primer (785R):

5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC

As reported in Klindworth et al.⁴³, Agencourt AMPure XP beads (Beckman Coulter, Milan, Italy) were used to purify PCR amplicons. The amplicons were then used for a second PCR in order to barcode the libraries using the Illumina dual-index system (Nextera XT Index Kit, Illumina Inc., San Diego, CA, USA) necessary for multiplexing. Following a second purification step, the eluted DNA products were quantified using the Qubit dsDNA BR Kit assay, diluted to 4 nM and pooled. The purified DNA products were then subjected to an additional PCR to attach dual Illumina indices (Nextera XT Index Kit, Illumina Inc., San Diego, CA, USA) necessary for multiplexing. Paired-end sequencing (2×300 cycles) was carried out using an Illumina MiSeq instrument (Illumina Inc.) according to the manufacturer's instructions. Sequences were demultiplexed based on index sequences, and FASTQ files were generated. FASTQ raw sequencing data were imported into QIIME2 v.2021.2⁴⁴ environments, and then Illumina primers were removed using q2-cutadapt plugin in trim-paired mode⁴⁵. Trimmed sequences were denoised in paired-end mode using q2-dada2 plugin⁴⁶. The assignment of taxonomy to amplicon sequence variants (ASVs) was performed with q2-feature-classifier plugin⁴⁷ against the pre-trained Naïve Bayes classifier SILVA 138 99% operational taxonomic units (OTUs) full-length sequence dataset⁴⁹.

Statistical analyses. The pan-microbiota (total observed richness in all samples) was determined in subsets of increasing size composed of randomly chosen samples (250 repetitions for each sample size). A collector's curve, i.e., the total number of observed genera with increasing numbers of samples collected, was subsequently calculated (chronological order) (10 repetitions for each sample size), according to Falony et al.⁵⁰. Once the sample's representativeness was checked, and it was noted whether the abundances showed very dispersed values, the presence of any homogeneous subgroups within the sample was verified. The *vegdist* function (*vegan* R package) was used to calculate Bray–Curtis distance, and the *kmeans* function was used to create the clusters⁵¹. The elbow and silhouette methods were used for determining the optimal clusters. Permutational multivariate analysis of variance (PERMANOVA) was performed on the Bray–Curtis distance matrix to determine if the gut microbiota structure differed between the two clusters, considering phyla, families, and genera. The *adonis2* function of the *vegan* R package was used. A heatmap was built as a graphical representation of the most abundant (representing 99% of total abundance) phyla using the *pheatmap* R package⁵². A chi-square test, or Fisher exact test when at least one class had $n < 5$, was used to test differences in microbiota composition and participant characteristics between the two clusters. The results are presented with $p(\chi^2)$ and Cramer's V. V values should be interpreted as > 0.5 = high association, 0.3 to 0.5 = moderate association, 0.1 to 0.3 = low association, 0 to 0.1 = little or no association. Richness (OTUs number) and Shannon's effective number were calculated using the *vegan* R package. A non-parametric method was used to calculate the reference intervals related to phyla, families, and genera. The 90% confidence intervals relative to the 95% lower and upper limits of the reference intervals were calculated using the bootstrap method according to the NCCLS Guidance Document C28A2⁵³. The *referenceIntervals* R package was used⁵⁴. The significance of differences in the abundance of phyla, families, and genera between clusters was tested using the Mann–Whitney test: *P* values from all statistical tests were adjusted for multiple comparisons within each taxonomic level, controlling the False Discovery Rate (FDR) (*FSA* R package) at level 0.05 using the Benjamini–Hochberg step-up procedure⁵⁵. A graphical display of a non-parametric correlation matrix, based on Spearman's R, ordered according to hierarchical clustering, was obtained using the *corrplot* R package. All the analyses were conducted using Microsoft Excel 16, Prism 8 (GraphPad Software, San Diego, CA), and R Studio 3.6.2.

Ethics approval. The study was approved by the Urbino University Ethics Committee (approval number 34_2021).

Data availability

Additional data is available in the supplementary material, and the datasets generated during the current study will be available upon request to the corresponding author.

Received: 20 September 2021; Accepted: 22 February 2022

Published online: 10 March 2022

References

- Rinninella, E. *et al.* Food components and dietary habits: Keys for a healthy gut microbiota composition. *Nutrients* <https://doi.org/10.3390/nu11102393> (2019).
- Illiano, P., Brambilla, R. & Parolini, C. The mutual interplay of gut microbiota, diet and human disease. *FEBS J.* **287**, 833–855. <https://doi.org/10.1111/febs.15217> (2020).
- Ventura, M. *et al.* Microbial diversity in the human intestine and novel insights from metagenomics. *Front. Biosci. (Landmark Ed.)* **14**, 3214–3221. <https://doi.org/10.2741/3445> (2009).
- Grenham, S., Clarke, G., Cryan, J. F. & Dinan, T. G. Brain-gut-microbe communication in health and disease. *Front. Physiol.* **2**, 94. <https://doi.org/10.3389/fphys.2011.00094> (2011).
- Yadav, M., Verma, M. K. & Chauhan, N. S. A review of metabolic potential of human gut microbiome in human nutrition. *Arch. Microbiol.* **200**, 203–217. <https://doi.org/10.1007/s00203-017-1459-x> (2018).
- Rahayu, E. S. *et al.* Gut microbiota profile in healthy Indonesians. *World J. Gastroenterol.* **25**, 1478–1491. <https://doi.org/10.3748/wjg.v25.i12.1478> (2019).
- Donati Zeppa, S. *et al.* Mutual interactions among exercise, sport supplements and microbiota. *Nutrients* <https://doi.org/10.3390/nu12010017> (2019).
- Klement, R. J. & Pазienza, V. Impact of different types of diet on gut microbiota profiles and cancer prevention and treatment. *Medicina (Kaunas)* <https://doi.org/10.3390/medicina55040084> (2019).
- Fontana, A. *et al.* Gut microbiota profiles differ among individuals depending on their region of origin: An Italian pilot study. *Int. J. Environ. Res. Public Health* <https://doi.org/10.3390/ijerph16214065> (2019).
- He, Y. *et al.* Author Correction: Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat. Med.* **24**, 1940. <https://doi.org/10.1038/s41591-018-0219-z> (2018).
- Senghor, B., Sokhna, C., Ruimy, R. & Lagier, J. C. Gut microbiota diversity according to dietary habits and geographical provenance. *Hum. Microbiome J.* **7**–8, 1–9. <https://doi.org/10.1016/j.humic.2018.01.001> (2018).
- Panebianco, C., Potenza, A., Andriulli, A. & Pазienza, V. Exploring the microbiota to better understand gastrointestinal cancers physiology. *Clin. Chem. Lab. Med.* **56**, 1400–1412. <https://doi.org/10.1515/cclm-2017-1163> (2018).
- Picchianti-Diamanti, A. *et al.* Analysis of gut microbiota in rheumatoid arthritis patients: Disease-related dysbiosis and modifications induced by etanercept. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms19102938> (2018).
- Wang, B., Yao, M., Lv, L., Ling, Z. & Li, L. The human microbiota in health and disease. *Engineering* **3**, 71–82. <https://doi.org/10.1016/J.ENG.2017.01.008> (2017).
- Parfrey, L. W. & Knight, R. Spatial and temporal variability of the human microbiota. *Clin. Microbiol. Infect.* **18**(Suppl 4), 8–11. <https://doi.org/10.1111/j.1469-0691.2012.03861.x> (2012).
- Rinninella, E. *et al.* What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms* <https://doi.org/10.3390/microorganisms7010014> (2019).
- De Luca, F. & Shoenfeld, Y. The microbiome in autoimmune diseases. *Clin. Exp. Immunol.* **195**, 74–85. <https://doi.org/10.1111/cei.13158> (2019).
- Nagao-Kitamoto, H., Kitamoto, S., Kuffa, P. & Kamada, N. Pathogenic role of the gut microbiota in gastrointestinal diseases. *Intest. Res.* **14**, 127–138. <https://doi.org/10.5217/ir.2016.14.2.127> (2016).
- Flowers, S. A., Ward, K. M. & Clark, C. T. The gut microbiome in bipolar disorder and pharmacotherapy management. *Neuropsychobiology* **79**, 43–49. <https://doi.org/10.1159/000504496> (2020).
- Huang, T. T. *et al.* Current understanding of gut microbiota in mood disorders: An update of human studies. *Front. Genet.* **10**, 98. <https://doi.org/10.3389/fgene.2019.00098> (2019).
- Tang, W. H. W., Backhed, F., Landmesser, U. & Hazen, S. L. Intestinal microbiota in cardiovascular health and disease: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* **73**, 2089–2105. <https://doi.org/10.1016/j.jacc.2019.03.024> (2019).
- Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638. <https://doi.org/10.1126/science.1110591> (2005).
- Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180. <https://doi.org/10.1038/nature09944> (2011).
- Gallardo-Becerra, L. *et al.* Metatranscriptomic analysis to define the Secrebiome, and 16S rRNA profiling of the gut microbiome in obesity and metabolic syndrome of Mexican children. *Microb. Cell Fact.* **19**, 61. <https://doi.org/10.1186/s12934-020-01319-y> (2020).
- Gerard, P. Gut microbiota and obesity. *Cell. Mol. Life Sci.* **73**, 147–162. <https://doi.org/10.1007/s00018-015-2061-5> (2016).
- Cooper, P. *et al.* Patent human infections with the whipworm, *Trichuris trichiura*, are not associated with alterations in the faecal microbiota. *PLoS ONE* **8**, e76573. <https://doi.org/10.1371/journal.pone.0076573> (2013).
- Koeth, R. A. *et al.* Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* **19**, 576–585. <https://doi.org/10.1038/nm.3145> (2013).
- Aya, V., Florez, A., Perez, L. & Ramirez, J. D. Association between physical activity and changes in intestinal microbiota composition: A systematic review. *PLoS ONE* **16**, e0247039. <https://doi.org/10.1371/journal.pone.0247039> (2021).
- Nishijima, S. *et al.* The gut microbiome of healthy Japanese and its microbial and functional uniqueness. *DNA Res.* **23**, 125–133. <https://doi.org/10.1093/dnares/dsw002> (2016).
- De Filippo, C. *et al.* 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. <https://doi.org/10.1073/pnas.1005963107> (2010).
- de Alencar, H. Jr. *et al.* The relationship between the commensal microbiota levels and Crohn's disease activity. *JGH Open* **4**, 784–789. <https://doi.org/10.1002/jgh3.12338> (2020).
- Mazzarella, G. *et al.* Pathogenic role of associated adherent-invasive *Escherichia coli* in Crohn's disease. *J. Cell. Physiol.* **232**, 2860–2868. <https://doi.org/10.1002/jcp.25717> (2017).
- Pittayanon, R. *et al.* Differences in gut microbiota in patients with vs without inflammatory bowel diseases: A systematic review. *Gastroenterology* **158**, 930e931–946e931. <https://doi.org/10.1053/j.gastro.2019.11.294> (2020).
- Perna, A. *et al.* Adherent-invasive *Escherichia coli* (AIEC): Cause or consequence of inflammation, dysbiosis, and rupture of cellular joints in patients with IBD? *J. Cell. Physiol.* **235**, 5041–5049. <https://doi.org/10.1002/jcp.29430> (2020).
- Merra, G. *et al.* Influence of Mediterranean diet on human gut microbiota. *Nutrients* <https://doi.org/10.3390/nu13010007> (2020).
- McDonald, D. *et al.* American gut: An open platform for citizen science microbiome research. *mSystems* <https://doi.org/10.1128/mSystems.00031-18> (2018).
- Johnson, A. J. *et al.* Daily sampling reveals personalized diet-microbiome associations in humans. *Cell Host Microbe* **25**, 789e785–802e785. <https://doi.org/10.1016/j.chom.2019.05.005> (2019).
- Han, M. *et al.* Stratification of athletes' gut microbiota: The multifaceted hubs associated with dietary factors, physical characteristics and performance. *Gut Microbes* **12**, 1–18. <https://doi.org/10.1080/19490976.2020.1842991> (2020).
- Clarke, S. F. *et al.* Exercise and associated dietary extremes impact on gut microbial diversity. *Gut* **63**, 1913–1920. <https://doi.org/10.1136/gutjnl-2013-306541> (2014).
- Huang, J. *et al.* Six-week exercise training with dietary restriction improves central hemodynamics associated with altered gut microbiota in adolescents with obesity. *Front. Endocrinol. (Lausanne)* **11**, 569085. <https://doi.org/10.3389/fendo.2020.569085> (2020).

41. Donati Zeppa, S. *et al.* Nine weeks of high-intensity indoor cycling training induced changes in the microbiota composition in non-athlete healthy male college students. *J. Int. Soc. Sports Nutr.* **18**, 74. <https://doi.org/10.1186/s12970-021-00471-z> (2021).
42. Nanna, M. J. Hotelling's T² vs. the rank transform with real Likert data. *J. Mod. Appl. Stat. Methods* **1**, 83–99. <https://doi.org/10.22237/jmasm/1020255180> (2002).
43. Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucl. Acids Res.* **41**, e1. <https://doi.org/10.1093/nar/gks808> (2013).
44. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857. <https://doi.org/10.1038/s41587-019-0209-9> (2019).
45. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12. <https://doi.org/10.14806/ej.17.1.200> (2011).
46. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583. <https://doi.org/10.1038/nmeth.3869> (2016).
47. Bokulich, N. A. *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90. <https://doi.org/10.1186/s40168-018-0470-z> (2018).
48. Bokulich, N. A. *et al.* Bokulich-lab/RESCRIPt: 2021.2.0. Zenodo <https://doi.org/10.5281/zenodo.4569379> (2021).
49. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucl. Acids Res.* **41**, D590–D596. <https://doi.org/10.1093/nar/gks1219> (2013).
50. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560–564. <https://doi.org/10.1126/science.aad3503> (2016).
51. *vegan: Community Ecology Package* (2008).
52. *heatmap: Pretty Heatmaps* (2015).
53. NCCLS. *How to Define and Determine Reference Intervals in the Clinical Laboratory; Approved Guideline—Second Edition.* NCCLS Document C28-A2, vol. 20 (2000).
54. *referenceIntervals: Reference Intervals* (2020).
55. *FSA: Fisheries Stock Analysis* (2021).

Acknowledgements

The research project has been partially supported by Carlo Bo University of Urbino research grants. Author V.P. is supported by Grant from “Associazione Italiana per la Ricerca sul Cancro” (IG2019–ID23006). We thank Gandino Mencarelli for his technical support.

Author contributions

D.S., Fa.P., M.B.L.R., P.S. and P.G. conceived the study and supervised the project. Fa.P. and P.G. collected samples. V.P., S.D.Z., E.B., Fr.P. and R.P. performed DNA extraction and sequencing. D.S., S.A. and M.B.L.R. analyzed the data. D.S., V.P., B.C., W.B., F.B., E.P., A.D.L., M.A., S.T., A.M., R.S., S.A., P.S. and P.G. wrote the original manuscript. All authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-08000-x>.

Correspondence and requests for materials should be addressed to S.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022