Experimental tests challenge the evidence of a healthy human blood microbiome

Leandro Di Gloria[1] (iD), Simone Baldi[2], Lavinia Curini[3], Sara Bertorello[2], Giulia Nannini[2], Francesco Cei[2], Elena Niccolai[2], Matteo Ramazzotti[1] (iD) and Amedeo Amedei[2,4]

1 Department of Experimental and Clinical Biomedical Sciences, University of Florence, Italy
2 Department of Experimental and Clinical Medicine, University of Florence, Italy
3 Cardiovascular Tissue Engineering Research Unit – Centro Cardiologico Monzino, IRCCS, Italy
4 Network of Immunity in Infection, Malignancy and Autoimmunity (NIIMA), Universal Scientific Education and Research Network (USERN), Florence, Italy

The advent of next-generation sequencing (NGS) technologies has made it possible to investigate microbial communities in various environments, including different sites within the human body. Therefore, the previously established belief of the sterile nature of several body sites, including human blood, has now been challenged. However, metagenomics investigation of areas with an anticipated low microbial biomass may be susceptible to misinterpretation. Here, we critically evaluate the results of 16S targeted amplicon sequencing performed on total DNA collected from healthy donors' blood samples while incorporating specific negative controls aimed at addressing potential bias to supplement and strengthen the research in this area. We prepared negative controls by increasing the initial DNA quantity through sequences that can be recognized and subsequently discarded. We found that only three organisms were sporadically present among the samples, and this was mostly attributable to bacteria ubiquitously present in laboratory reagents. Despite not fully confirming or denying the existence of healthy blood microbiota, our results suggest that living bacteria, or at least their residual DNA sequences, are not a common feature of human blood in healthy people. Finally, our study poses relevant questions on the design of controls in this research area that must be considered in order to avoid misinterpreted results that appear to contaminate current high-throughput research.

Introduction

The high diversity within the prokaryotic domain allows bacteria and archaea to thrive in almost every environment of the world, including other organisms. Notably, bacteria inhabit many human tissues forming symbiotic relationship with the host, constituting the so-called microbiome, a crucial entity considered on par with human organs [1]. Under physiological conditions, our immune system and physiological barriers such as the gut-vascular and the blood–brain barriers prevent the microbial colonization of districts that should remain sterile in healthy subjects. This understanding has prevailed until recent years when advancement in next-generation sequencing (NGS) has enabled a cost-effective molecular survey of microbes

Abbreviations
DMEM, Dulbecco's modified eagle medium; ENC, *Escherichia coli* negative control; FBS, fetal bovine serum; HBM, healthy blood microbiota; NGS, next-generation sequencing; OTC, off-target control; PBS, phosphate-buffered saline; PCoA, principal coordinate analysis.

in body habitats, challenging established beliefs. Among these faiths, there is the view that the blood of healthy subjects is a sterile environment [2].

Recent NGS technologies have revealed the presence of bacterial genera such as *Pseudomonas, Propionibacterium, Halomonas, Streptomyces, Sphingomonas, Staphylococcus, Leifsonia, Bacillus, Flavobacteria, Serratia* etc. in healthy blood samples [2–5].

However, authors highlighted potential source of contamination and procedural errors in these studies. As early as 2001, Nikkari *et al.* [6] documented the presence of 16S DNA in healthy blood samples and empathized that bacteria detected may derive from used reagents or incorrect sampling. In detail, microbes from skin flora including *Cutibacter, Streptococcus, Sphingomonas*, and *Bacillus* genera are well recognized to contaminate blood cultures, with the venepuncture process introducing skin microbes into the samples [7–9].

Moreover, due to the high sensitivity of NGS sequencing to noises and contaminants [5,6,10], it has been established that the presence of contaminants, referred to as "kitome", is unavoidable in both DNA extraction and PCR kits, especially when targeting low microbial biomass environments such as blood [10–15].

The sequencing procedure can itself contribute to ecological diversity inflation in samples, introducing potential issue like "index hopping", polymerase errors and contamination from residual sequences deriving from prior sequencing runs [13,16]. Furthermore, PCR-based analysis of low bacterial biomass tissues involves off-target amplification of the host DNA due to the overwhelming prevalence of human cells [17].

Despite these challenges, many authors have hypothesized that DNA reads obtained from sequencing healthy blood samples may belong to bacteria characteristic of this human district, proposing the existence of a "human healthy blood microbiota" (HBM). Remarkably, in 2022, Khan and colleagues defined the existence of a blood microbiota as "already established", speculating that bacteria from organs, especially the gut, may translocate into the circulatory system without epithelial impairment using a yet unknown mechanism [18].

According to current literature [2], the potential HBM is mainly composed by Proteobacteria. The out-of-gut origin of blood microbes contrasts with the well-established prevalence of Firmicutes, Bacteroidetes and Actinomycetes in the gut [18,19]. Other authors suggest that a more likely source of blood bacteria may be the skin or the oral mucosa [2,18]. Conversely, in 2016, Santiago and colleagues explored the serum microbiota composition in cirrhotic patients and reported that 69%

of the identified bacterial sequences in their negative controls belonged to the Proteobacteria phylum [5], mimicking the profile often attributed to the healthy blood microbiota. In 2023, Tan and collaborators raised a strong counterpoint, reporting no common species among 9770 healthy human blood sequences collected from databases and concluding that hypothetical blood microbiota members may be sporadic microbes that transiently migrate in the bloodstream [20]. This would not be the first occurrence of contrasting results regarding the existence of a microbiota in an environment conventionally defined as sterile. For instance, in 2019 de Goffau *et al.* [21] showed that there was no evidence to support the recently hypothesized human placenta microbiome and that almost all of the related signals sourced from contaminations.

Regardless, it is crucial to consider that the detected DNA sequences may originate from destroyed bacteria rather than viable ones, with relevance for the concept of microbiota as a community of living organism that interact among themselves and with the host in defining the so-called holobiont [22].

In addition, in this intricate scenario adopting different DNA extraction protocols could further complicate this research, impacting the quality, quantity and purity of extracted DNA or its preservation, subsequently affecting the detectable bacterial sequences in the blood [15,18,23–25].

Given the complexities mentioned above, delving into this challenging topic requires many precautions, notably the inclusion of negative controls in the study design. However, this undertaking is not always straightforward, particularly with negative controls composed of pure sterile water, which are often challenging to sequence. In fact, obtaining such traditional negative controls for this study has not been successful either. This difficulty hampers our ability to finely discern which reads may be influenced by the several biases described.

So, the primary aim of this study was to investigate the microbial composition of healthy human blood, employing custom controls that may permit the discrimination between contaminants and potential blood commensals. The ultimate goal was to provide insights into the potential existence and characteristics of HBM, while investigating the similarities between its profile and monitored artifacts sourced from the sequencing itself. To take into account potential peculiarities which may arise from using a certain kit, we extracted the DNA of each enrolled subject and control employing two different DNA extraction kits, as described in the Materials and methods section. Including two samples for each subject has also the purpose of inspecting their reciprocal similarities and

differences. In fact, on this matter, if we assume that a characteristic bacterial community exists, then we expect such intra-subject differences to be minimal when compared to the inter-subject difference, or at least being explicable by recurrent patterns in line with being sourced by using different kits. Notwithstanding, our main aim is not to draw conclusions about the two kits performances but to explore and challenge the HBM hypotheses in more than a way, while also considering the eventuality that a hypothetical blood microbe may be detectable by only a certain kit. Therefore, although a performance comparison cannot be avoided by reporting the results, any related statistics have been computed. In this regard, we underline that this study has not been designed to employ statistics beyond that required to process the DNA reads, in order to avoid tricky conclusions when addressing the following simple (yet challenging) question: "how far can the supposed HBM microbiota profile be mimicked by sheer artifacts?". Delving into the noises of this type of analysis is crucial to pursue the researches on this topic, because confirming the presence of commensal bacteria or their DNA fragments in healthy blood would valid the hypothesis regarding their interaction with the immune system, potentially unveiling new valuable biomarkers.

Results

Percentage of reads retained after quality and abundance filtering

The sequencing of all samples was successful, except the failure of the DNA sample extracted with the QIAamp® DNA Microbiome Kit. A total of 690 228 reads has been obtained of which 451 628 (65.4%) originated from samples extracted with the QIAamp® DNA Microbiome Kit and 238 600 (34.6%) from samples extracted with the DNeasy® Blood & Tissue Kit.

From a further check, the obtained data showed different sequencing depths among the samples although they have been sequenced in the same lane and run (Fig. 1). Indeed, the negative control samples consisting in DNA from *Escherichia coli* pure culture (ENC, see Materials and methods), extracted with the two kits feature the largest number of reads, amounting to 56.3% of the total obtained reads. Excluding ENC samples, the majority of acquired reads were discarded
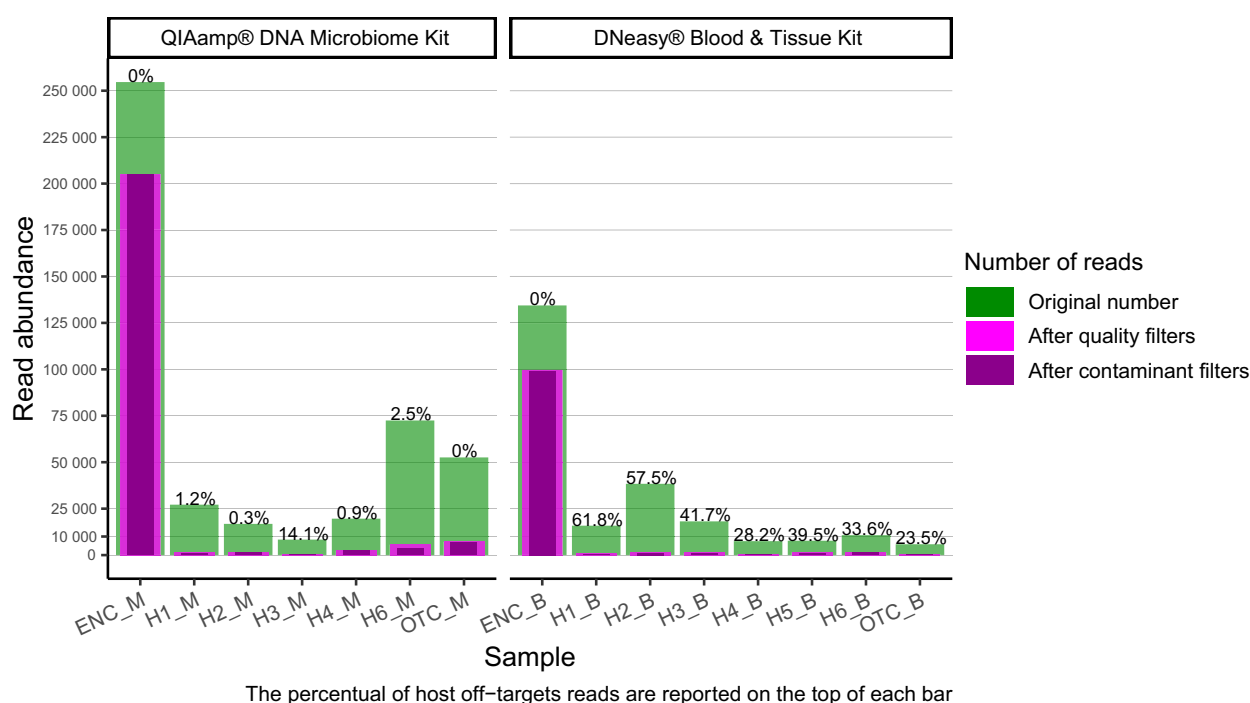

The percentual of host off−targets reads are reported on the top of each bar

Fig. 1. Bar plots displaying the abundances of reads in samples as a whole (green), after quality filters (magenta) and after contaminant removal (dark magenta). The percentage of removed off-target reads is reported at the top of each bar. The healthy samples are labeled as "H" while the controls are labeled as "ENC" or "OTC" (see Materials and methods). The labels' suffixes "_M" and "_B" represent the extraction conducted with either the QIAamp® DNA Microbiome Kit or DNeasy® Blood (M) & Tissue Kit (B) respectively. The *Y*-axis is scaled to enhance the visibility of lower read abundances.

during the quality filtering steps, leaving an average of 7.5% of sequences for subsequent analyses. Specifically, samples processed with QIAamp® DNA Microbiome Kit exhibited a lower loss of reads after filtering compared to samples processed with DNeasy® Blood & Tissue Kit (9.3% vs 4.0%). The abundance of human off-targets DNA was significantly lower in samples extracted though QIAamp® DNA Microbiome Kit (3.8% of total sequences) compared to the samples extracted with the DNeasy® Blood & Tissue Kit (40.8% of total sequences).

The additional filters based on relative abundances and prevalence had a limited impact on decreasing the reads' number, excluding approximately the 0.2% of the original reads (Fig. 1).

Although the majority of reads have been discarded through the aforementioned filters, each sample appears to be saturated (Fig. 2). Therefore, it is conceivable that increasing the sequencing depth would not lead to the identification of further bacteria.

Taxa distribution in the whole dataset

After filtering, the dataset retained a total of 25 genera (Table 1), among which 23 successfully classified at genus rank and two classified only at family rank. All the genera were members of 6 phyla, namely Proteobacteria, Actinobacteriota, Bacteroidota, Firmicutes, Verrucomicrobiota, and Campilobacterota phyla. Anyway, the complete list of genera and samples in which they were found is provided in Table 1. Of note, each genus has been identified in at least one sample, using both DNA extraction kits, with the exception of *Flavobacterium* and *Lachnospiraceae_ND3007* which were
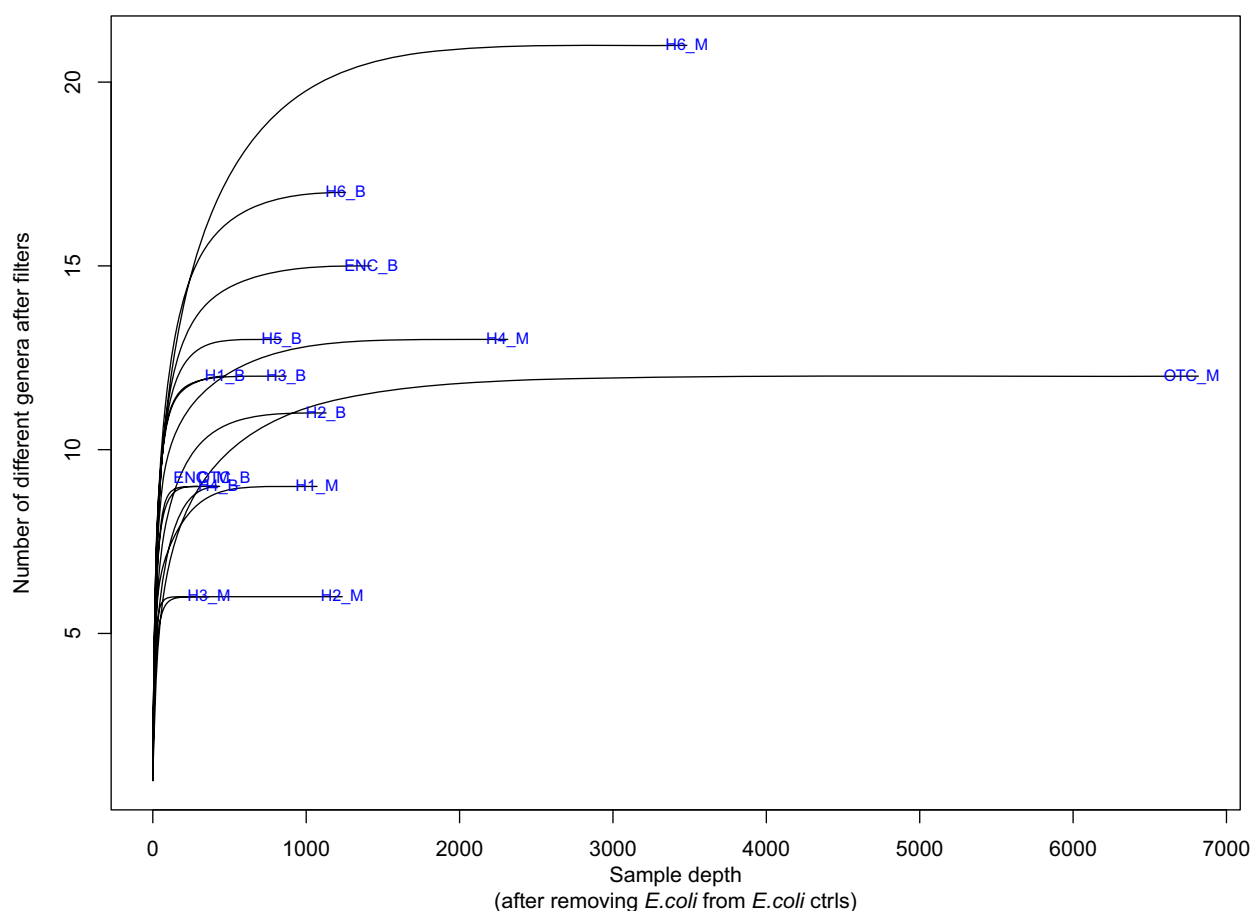

Fig. 2. Saturation analysis on genera performed after removing *Escherichia coli* from the controls where it was used as spike. The *X*-axis represents the number of remaining reads after the filters (the "sample depth"), while the *Y*-axis indicates the number of distinct bacterial genera observed at each sampling depth. Each sample is visualized as a separate line in the plot, showing how the observed genus richness varies with different levels of sequencing depth. The healthy samples are labeled as "H" while the controls are labeled as "ENC" or "OTC" (see Materials and methods). The labels' suffixes "_M" and "_B" represent the extraction conducted with either the QIAamp® DNA Microbiome Kit or DNeasy® Blood (M) & Tissue Kit (B) respectively.

Table 1. Table of all identified bacterial taxa. The presence is confirmed when a particular genus is found in at least one sample within the groups, denoted by an asterisk (*) in the corresponding column. The four groups are: C-DBT (controls processed with DNeasy® Blood & Tissue Kit), C-MIC (controls processed with QIAamp® Microbiome kit), B-DBT (blood samples processed with DNeasy® Blood & Tissue Kit) blood samples and B-MIC (blood samples processed with QIAamp® Microbiome kit). The taxa reported as potential contaminant in previous published papers (references listed in method section) are denoted by an asterisk in the column "KNOWN".

Phylum	Genera	C-MIC	C-DBT	B-MIC	B-DBT	KNOWN
Proteobacteria	*Escherichia-Shigella*	*	*	*	*	*
Proteobacteria	*Burkholderia-Caballeronia*	*	*	*	*	*
Actinobacteriota	*Leifsonia*	*	*	*	*	*
Proteobacteria	*Sphingomonas*	*	*	*	*	*
Actinobacteriota	*Cutibacterium*	*	*	*	*	*
Bacteroidota	*Bacteroides*	*	*	*	*	*
Firmicutes	*Clostridia_UCG-014*	*		*	*	*
Proteobacteria	*Methylobacterium-Methylorubrum*	*	*	*	*	*
Verrucomicrobiota	*Akkermansia*	*		*	*	
Firmicutes	*Ruminococcus*		*	*	*	*
Proteobacteria	Family *Xanthobacteraceae*	*	*	*	*	*
Campilobacterota	*Helicobacter*		*	*	*	
Proteobacteria	*Pseudomonas*	*		*	*	*
Actinobacteriota	*Bifidobacterium*	*	*	*	*	
Bacteroidota	*Flavobacterium*		*		*	*
Firmicutes	*Lactobacillus*			*	*	*
Firmicutes	*Phascolarctobacterium*			*	*	
Firmicutes	*Roseburia*		*	*	*	
Firmicutes	*Lachnospiraceae_ND3007*		*		*	*
Bacteroidota	*Alistipes*		*	*	*	
Bacteroidota	*Prevotella*	*		*	*	*
Firmicutes	*Christensenellaceae_R-7*		*	*	*	
Firmicutes	Family *Lachnospiraceae*		*	*	*	*
Firmicutes	*Subdoligranulum*	*	*	*	*	
Firmicutes	*Coprococcus*			*	*	*

detected only in samples extracted with DNeasy® Blood & Tissue Kit. However, nearly every genus was also detected in the negative controls, as discussed in more detail below. In addition, almost every genus is potential contaminants according to our literature research.

Comparison of control and blood sample microbial profiles

The most abundant phyla in blood samples were Proteobacteria (60.67%), Actinobacteriota (16.37%), Firmicutes (10.42%), Actinobacteriota (16.37%), and Verrucomicrobiota (2.01%) (Fig. 3A) while the five most represented genera were *Burkholderia-Caballeronia-Paraburkholderia* (51.94%), *Leifsonia* (13.87%), *Sphingomonas* (5.53%), *Bacteroides* (5.56%), and *Clostridia UCG-014* (3.18%) (Fig. 3B).

Apart from their abundances, each of these taxa's traces were found in control samples as well. Specifically, the abundances in the control consisting in eukaryote off-target DNA (OTC) extracted with DNeasy® Blood & Tissue Kit, closely resemble those observed in blood samples. Furthermore, upon removing *E. coli* reads from the control samples (used as spike signal) the relative abundances of the contaminant bacteria become more apparent (Fig. 4A,B). Consequently, the abundances in negative controls closely resembled those observed in blood samples (Fig. 4A,B).

We further compared samples' profiles through a principal coordinate analysis (PCoA) which revealed dissimilarity between the controls and the blood samples. Only the OTC sample extracted with DNeasy® Blood & Tissue Kit clustered with blood samples (Fig. 5A). The same analysis after excluding *E. coli* from those controls showed that three out of four controls closely resemble blood samples, strongly suggesting a similar abundance profile. The only exception was the OTC sample extracted with the DNA Microbiome kit, which exhibited a distinct profile, positioning in the region that was originally occupied by *E. coli* controls (Fig. 5B). Moreover, with the exclusion of OTC samples, the two extraction kits did not lead to notably different profiles for each sample pair. When considering samples derived from the same
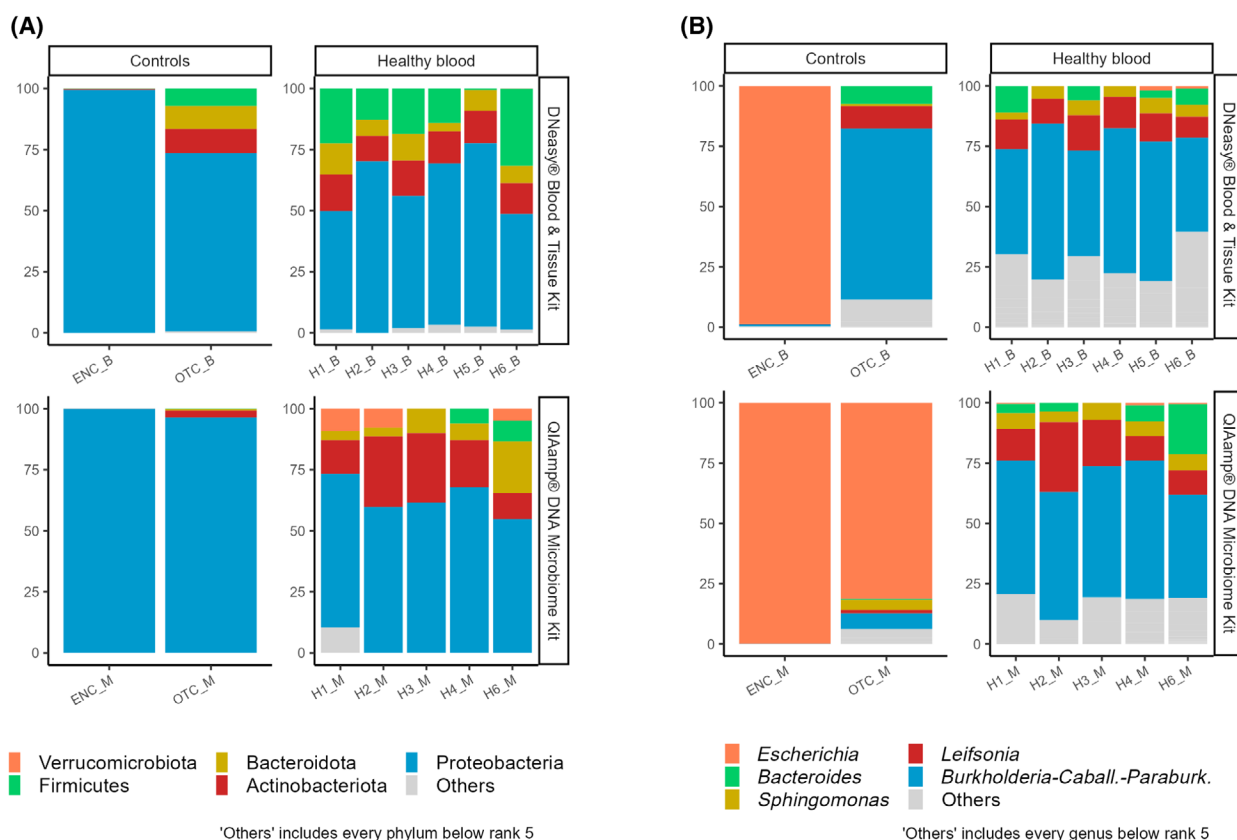
Fig. 3. Comparison of percentage abundance of the five most abundant phyla (A) and genera (B) in negative controls and samples before removing *Escherichia coli* from the controls where it was used as spike. "Others" includes every taxon below rank 5. The healthy samples are labeled as "H" while the controls are labeled as "ENC" or "OTC" (see Materials and methods). The labels' suffixes "_M" and "_B" represent the extraction conducted with either the QIAamp® DNA Microbiome Kit or DNeasy® Blood (M) & Tissue Kit (B) respectively.

subject, some pairs exhibit greater dissimilarity, while others appear more similar, seemingly without a discernible pattern (Fig. 5A,B). Since the OTC sample extracted with QIAamp® DNA Microbiome kit showed an unexpected excess of *E. coli* reads we monitored its relationship with other samples by removing *E. coli* associated reads. As expected, this simple *E. coli* purging brought this sample in agreement with other samples, reasonably indicating a specific contamination during processing (Fig. 5C).

Investigating the unexpected behavior of the mis-positioned OTC mice control that clustered in the *E. coli* region, we observed that the large majority of the reads in this control were actually dominated by *E. coli* reads. The removal of such reads, reasonably deriving from cross-contamination by *E. coli* samples, led in fact to a greater decrease of the two main coordinates (indicating less variation among samples) and the positioning of the contaminated OTC much closer to all other samples that eventually appear as ensemble of unresolved individuals in the community.

Three genera, *Coprococcus*, *Lactobacillus*, and *Phascolarctobacterium*, were exclusively identified in blood samples and absent in negative controls (Fig. 6A). More specifically, each of these three genera was found in lower abundance in no more than three samples from different donors and from samples processed with different extraction kits (Fig. 6B). Furthermore, we also consider informative to highlight that the low cumulative abundances of these bacteria required an important edit to the *Y*-axis of this figure (explained in the related caption) in order to allow their visualization.

Discussion

The ongoing discussion on the sterility of healthy blood has significant implications for the upcoming research, especially in the physiology field. Nevertheless, our study underscores that exploring this topic is a complex journey fraught with challenges stemming from both technical and environmental sources, introducing noise and contaminants.
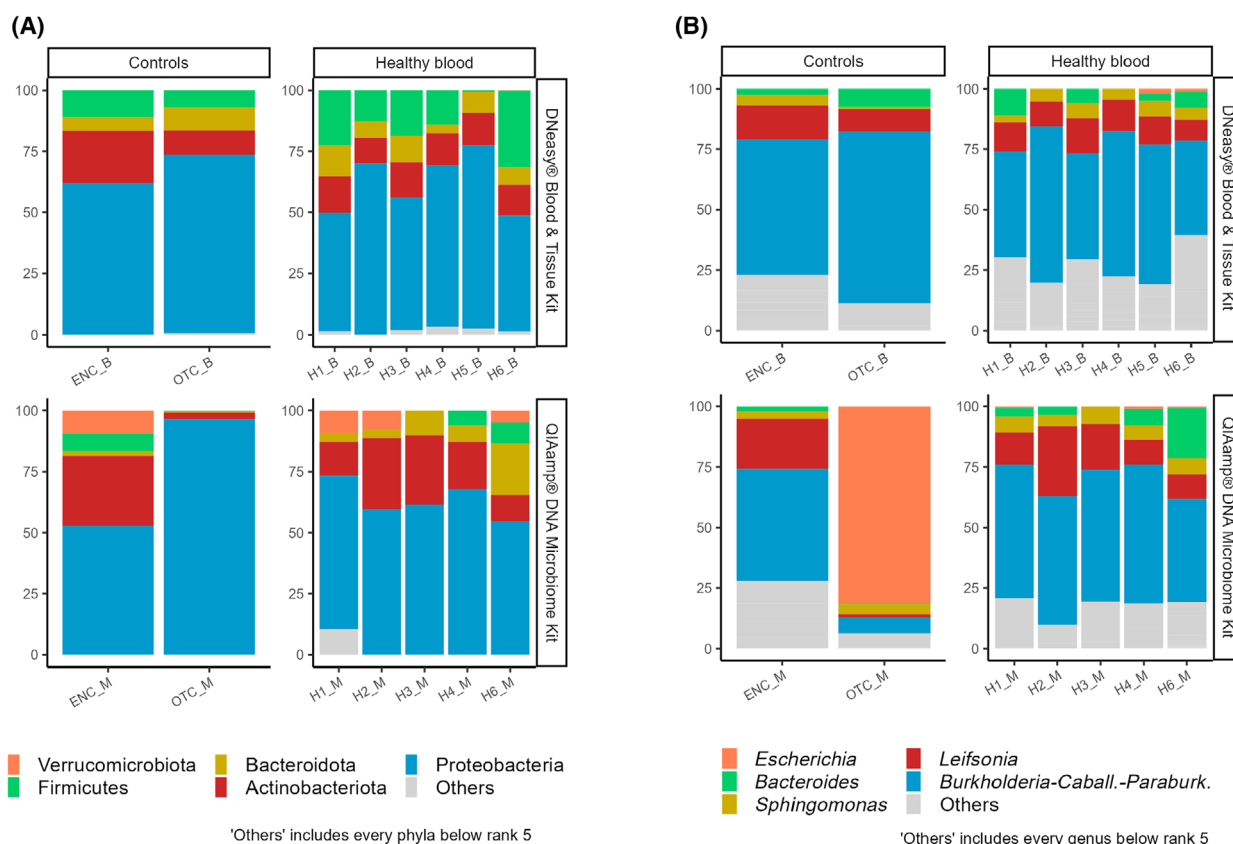
Fig. 4. Comparison of percent abundance of the five most abundant phyla (A) and genera (B) in negative controls and samples after removing *Escherichia coli* from the controls where it was used as spike, thereby enabling a clearer comparison with the blood samples, differently from the unadjusted representation in Fig. 3. "Others" includes every taxon below rank 5. The healthy samples are labeled as "H" while the controls are labeled as "ENC" or "OTC" (see Materials and methods). The labels' suffixes "_M" and "_B" represent the extraction conducted with either the QIAamp® DNA Microbiome Kit or DNeasy® Blood (M) & Tissue Kit (B) respectively.

We explored the presence of a potential blood microbiome in samples from six healthy subjects, utilizing two distinct DNA extraction kits: one specifically designed for extracting microbial DNA from blood and the other intended to eliminate human DNA while isolating microbial DNA. In addition, we introduced two control samples rich of known microbial DNA for an easier identification and exclusion: one comprising mice DNA (off-target control), and another enriched with *E. coli* DNA. In our experience, negative controls consisting of pure sterile water are often challenging to sequence due to the insufficient total amount of DNA detected after PCR, especially given the cycle number characteristic of the conventional Illumina protocol. While this outcome validates the overall sterility of the workflow, it does not allow us to discern which reads may stem from the various biases described. To overcome this limitation, we hypothesize the possibility of "enhancing" these negative controls by increasing the initial DNA quantity

through sequences that can be recognized and subsequently discarded.

Firstly, we noted an unsuccessful sequencing attempt for one healthy blood sample processed with the QIAamp® DNA Microbiome Kit. However, all other samples, including the corresponding sample from the same subject processed with the DNeasy® Blood & Tissue Kit, and even the negative controls, were successfully sequenced. Nevertheless, we view this issue as a valuable clue regarding the estimable microbiota richness, at least for this particular subject.

Overall, the samples obtained using the DNeasy® Blood & Tissue Kit feature a high number of off-target sequences despite the specificity of the used primers, implying an extremely low bacterial biomass, if any. Conversely, the samples obtained through the QIAamp® DNA Microbiome Kit, designed to remove host DNA, were nearly devoid of host sequences. However, irrespective of the DNA extraction kit, only a limited number of sequences for each sample passed

the quality filters during the FASTQ processing, except for the *E. coli* controls which underwent processing with no substantial loss of reads. This outcome confirms both the unusual behavior of the blood sample sequences and the processing effectiveness.

Similar challenges were also encountered during the processing of the OTC samples. Indeed, Glassing and colleagues reported that when sequencing DNA extracted from blood samples with the MoBio Power-Max® Soil DNA Isolation Kit, they obtained around 2000 sequences [11]. However, despite the significantly greater depth of sequencing, only 25% of these sequences were identified as prokaryotes [11]. Additionally, the authors noted that only minor traces of the genera *Anaerostipes*, *Mogibacterium*, *Subdoligranulum*, *Halocella* and *Sphingobium* were exclusively present in their blood samples and not in blank controls, as per their abundance filters.

Among these genera only reads from *Subdoligranulum* have been identified in our blood samples as well as in our controls. Specifically, in our dataset, after applying the abundance filters, only 25 genera were detected, despite the permissive thresholds used. In agreement with previous reports, the most abundant phylum that we have observed in blood samples is Proteobacteria. Given the speculations about the potential sources of bacteria in healthy blood, it is noteworthy that this phylum has been reported as the most abundant in the healthy lungs [26]. However, it is also frequently the most abundant in blank controls [5,14], a pattern consistent with our control samples.

Notably, the most abundant genera identified in blood samples were found also in each control, exhibiting the same abundances profile, particularly in the mice DNA control processed with DNeasy® Blood & Tissue Kit and in the *E. coli* controls after discarding the spike signal. Probably explanations of these results include the possibility that these bacteria originate from an environmental contamination. These remnants could than been sequenced alongside the off-target DNA. However, we observed a distinct microbial profile in the mice DNA control processed with the QIAamp® DNA Microbiome Kit. We assume that the variability in controls processed with different kits could be attributed to (a) differences in the "kitome" of the two extraction kits; (b) stochastic amplification of extremely low abundant DNA templates [27,28]; (c) different contaminants present in various samples within a single sequencing lane [29].

Finally, to further explore the distinctive features distinguishing healthy blood samples from the controls, we checked which bacterial DNA have been exclusively sequenced in the blood samples.

This analysis revealed that the genera *Coprococcus*, *Lactobacillus* and *Phascolarctobacterium* have been identified only in blood samples, regardless of the extraction kit used, but only in a limited number of specimens. This outcome can be attributed to their actual presence in the healthy blood of only some individual or, alternatively, to a shared portion of the kitome present in both extraction kits or introduced by the PCR reagents. Notably, these sequences were produced in some samples in our dataset, with the exception of the controls.

In detail, both *Lactobacillus* and *Coprococcus* are recognized as common bacteria of intestinal flora and as possible contaminants from the DNA extraction kit [11,30]. Meanwhile, *Phascolarctobacterium* is also an intestinal commensal genus, not reported as known kit contaminants, but identified in traces in only two out six subjects. It is relevant to know that many contaminants are often associated to gastrointestinal tract or skin [11] whereas the community profile of our samples doesn't align with either of those environments. Moreover, as far as we know, *Phascolarctobacterium* has not been identified in any other research on healthy blood microbiota. Conversely, *Coprococcus* was reported as a potential blood bacterium by Jagare and colleagues. However, its abundance overcame the threshold chosen by the authors only in the blood of patients with gut disease, not in healthy subjects and negative controls, where it was also detected [31].

For these reasons, we cannot conclusively assert that these bacterial traces are common features of human blood. On the other hand, it is still relevant to underline that defining a contaminants list by just comparing the genera names between analyzed samples and controls it is not a faultless method, for example, because some taxon may be characteristic of both the sample types [32]. In addition, phylogenetic clades can be missed depending on the employed 16S primers having these not a truly "universal" coverage [33,34]; thus, we can assume that a hypothetical healthy blood common bacterium is missing in our dataset. In light of this, it is in our view that, despite the undoubted usefulness of highlighting certain potential taxa, a metagenomics approach performs at its finest when the overall profile is analyzed. Accordingly, hereby we emphasize the substantial profile overlap between blood sample and controls. The most abundant genera in our dataset are constantly present in every blood sample, and each of these taxa is detected in every control and also reported as common contaminants originating from reagents [10,11,30]. Furthermore, also their relative abundances are very similar between samples and controls. We consider this outcome
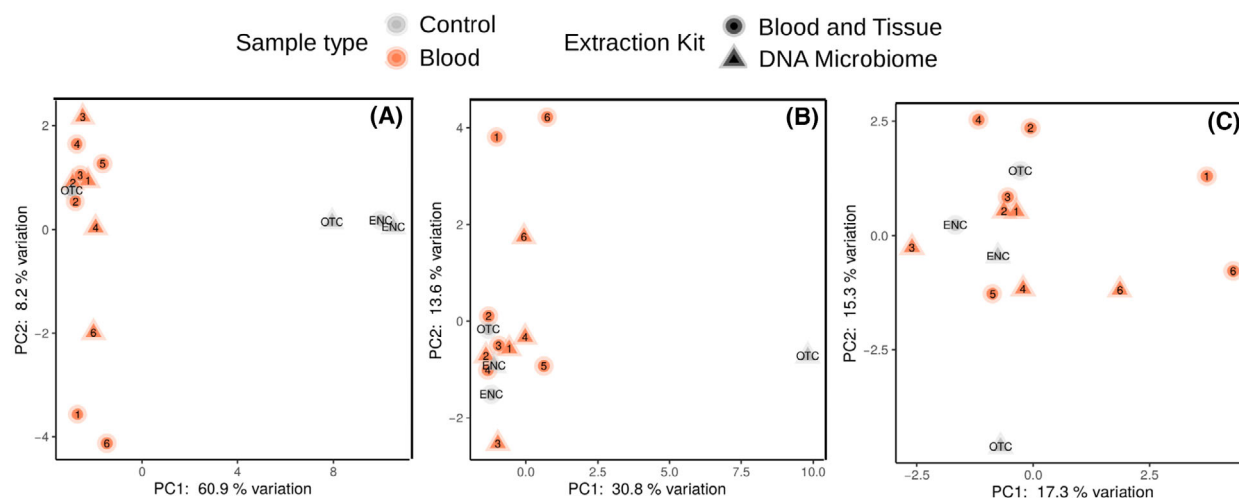
Fig. 5. Principal Coordinate Analysis (PCoA) representing the computed Hellinger distances among samples. The label on each point shows the type of control samples or the identity of the 6 healthy donors. (A) Non-processed samples. (B) Samples after removing *Escherichia coli* specific reads from the ENC. (C) Samples after removing *E. coli* specific reads from all the controls. OTC (off-target control): genomic DNA form C2C12 cultured mouse cells. ENC (*E. coli* negative control): genomic DNA from *E. coli*.
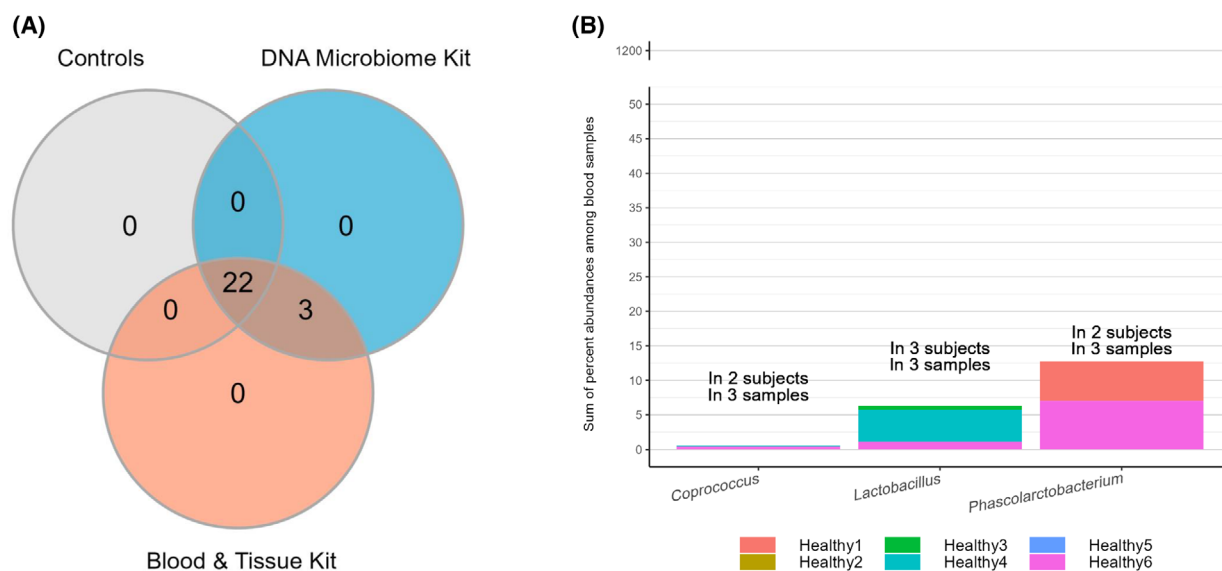

Fig. 6. Taxa distribution. (A) Venn diagram illustrating the number of genera shared among controls (gray) and blood samples obtained using extraction kits (coral, blue). (B) Stacked bar plot representing the cumulative percentage abundance of the three genera exclusive to the healthy blood group. Abundance is calculated by summing up the percentage abundance across each sample (12 samples, totalling 1200%). The *Y*-axis breaks between 50 and 1200 have been hidden to improve the visibility of lower values.

remarkable because these genera alone constitute the majority of our dataset (about the 75% of the total abundance for each sample, excluding the OTC control extracted with QIAamp® DNA Microbiome kit). Even considering the high sensitivity of PCR amplification, the prevalence of well-known contaminants as the most abundant genera implies an abundance of the hypothetical true healthy blood bacteria close or equal

to zero. Although it is conceivable that some bacteria or related DNA sequence in laboratory kits may overlap with those found in the healthy blood, it seems improbable that this holds true for each of them. A slight difference between few samples and the controls has been observed in the PCoA plot which is computed considering every genus besides the most abundant ones; however, this difference is still negligible

when compared to the inter-subject differences. Even justifying this latter result by taking into account a hypothetic major bias sourced by using different kits, we consider the proximity between most of the samples and the controls to be meaningful. At this regard, it is also important to note that the first two principal components of the PCoA have not explained most of the data variability by themselves and that the blood samples have not been positioned in a distinct cluster, despite the employed controls are sheer monitored artifacts sourced from a theoretical "empty" environment. Nevertheless, we do not rule out the possibility that some living bacteria, or at least traces of them, may sporadically and transiently be present in bloodstream of healthy subjects. However, such occurrence does not match with an established ecological community such as the microbiota.

Conclusions

Our analyses and methodologies strongly support the relevance of considering issues and risks when exploring the healthy blood environment or similar ones through the 16S rRNA NGS. Based on our data, we cannot definitively affirm or deny the existence of healthy blood bacteriota. However, the described results cast relevant doubts on the notion that certain bacteria, or their residual DNA sequences, are a common and distinctive feature of the healthy human blood. Consequently, we firmly emphasize that, at current state of knowledge, further, extremely careful research is needed before asserting the existence of a healthy blood microbiota.

Materials and methods

Sample collection and DNA extraction

In this study, we collected EDTA-anticoagulated whole blood from six young and healthy Italian volunteers (three male and three female), ranged in age from 25 to 35 years. Including both the genders has been designed to prevent missing results about hypothetical characteristic bacteria featured only in a certain physiology during the profiling of the main elements of the HBM, and it has not the purpose to represent the entire population due to both variability and rare occurrences that clearly would require many more samples and a different study design. The whole collection and extraction procedures have been performed with careful attention to avoid every source of external contaminations. For each subject, the first mL of blood has been discarded to reduce the risk of contamination from skin microbes. Once collected, the samples have been immediately processed using gloves, pipettes and biological hoods thoroughly cleaned with bleach and ethanol.

In addition, two different types of controls, prepared by inflating the initial DNA quantity through sequences that can be recognized and subsequently discarded, have been included. Specifically, our controls included (a) the off-target DNA control (OTC) from C2C12 cells (immortalized mouse myoblast cell line) cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with fetal bovine serum (FBS) 10%, penicillin and streptomycin and washed three times in phosphate buffered saline (PBS) solution before being collected; (b) the microbial DNA from a *E. coli* negative control (ENC) extracted from an overnight, kanamicin selected LB culture of kanamycin resistant *E. coli strain K12* at 1 : 1000 inoculum. In particular, the ENCs play a role also as positive controls to test the overall success of the library preparation and sequencing before discarding their spikes, due to the assured traces of actual bacterial DNA.

For both samples and controls, we used two different DNA extraction kits, namely the DNeasy® Blood & Tissue Kit (Qiagen, Hilden, Germany) (specifically designed for the purification of total DNA from the blood environment) and the QIAamp® DNA Microbiome Kit (Qiagen) (designed for enhancing the purification of DNA from intact bacterial cells though lysis of host cells and enzymatic digestion of free DNA prior to the lysis of bacteria), following manufacturer instructions. Both kits were unsealed for the first time specifically for this research work. Hence, a total of 16 DNA samples has been collected from 6 subjects and 2 controls.

16S sequencing and bioinformatics analysis

The extracted DNA was sent to IGA Technology Services (Udine, Italy) for 16S amplicon paired-end sequencing (2 × 300 cycles, 50 000 reads) on the MiSeq Illumina platform, according to the Illumina 16S Metagenomic Sequencing Library Preparation protocol. In particular, the V3–V4 hypervariable region has been amplified using the primer pair 341F (CCTACGGGNGGCWGCAG) and 805R (GACTACNVGGGTWTCTAATCC) [35].

The demultiplexed sequence reads were processed in the QIIME2 2022.8 environment [36]. Briefly, the sequencing primers and the reads without primers were removed using the CUTADAPT tool. DADA2 was used to perform paired-end reads filtering, merging and chimeras removal steps after trimming low quality nucleotides from both forward and reverse reads. Hence, ASVs (amplicon sequence variants) were generated and the taxonomic assignments were performed through Vsearch using the SILVA SSU database (release 138). Every sequence unassigned at the domain taxonomic level or associated to chloroplasts or mitochondria according to SILVA has been discarded. Moreover, every cross-amplified host DNA has been

identified and removed by aligning the ASVs to GRCm39 (murine reference genome) and GRCh38.p14 (human reference genome) [37] through BOWTIE2 2.2.5 [38]. Finally, every genus with a mean relative abundance less than 0.01% (computed considering also the host DNA in the total library size) has been removed to further avoid probable contaminants [11,32]. Such threshold has been chosen according to literature research, considering the nature of the samples (potentially low biomass environments) and prioritizing the accuracy over the sensitivity in order to decrease the risk of mislabeling sequences as contaminants [32]. In addition, every genus found in less than 3 out of 12 blood samples has been considered as a potential contaminant or at least as far from being a "common" healthy blood bacteria and accordingly discarded. The analysis of bacterial communities was performed in R 4.3.0. The packages PHYLOSEQ 1.44.0 [39], VEGAN 2.6-4 [40] and GGPLOT2 3.4.2 [41] were used to plot data and results. A rarefaction analysis on genera was performed on every sample using the function rarecurve (step 100 reads), further processed to highlight saturated samples (arbitrarily defined as samples with a final slope in the rarefaction curve with an increment in genus number per reads < 1e-4). The most abundant bacteria have been defined according to average percentage abundance among the samples. The taxa have been highlighted as potential contaminants in Table 1 whereas reported as such or featured in negative controls in the hereby listed literature [9,10,14,30,42]. Principal Coordinate Analyses (PCoAs) were performed using the Hellinger distance on Hellinger transformed genera abundances to address the compositional nature of the data [43].

Finally, a Venn diagram has been used to display the genus shared between ENCs and blood samples. Further details about the processing of reads and the subsequent bacterial community analysis are reported as a publicly available scripts (see Data availability statement section).

Ethics approval and consents

This study was performed in compliance with standards set by the Local Ethic Committee Area Vasta Centro (CEAVC 13725_bio, April 18 2023) and the Declaration of Helsinki. All methods were carried out in accordance with relevant guidelines and regulations. All participants gave their informed written consent.

Strength and limitations of this study

Although a modest number of subjects have been recruited for this study, we assert the reliability of our conclusions as they were derived by identifying consistent patterns and presences that are expected to be characteristic of a particular environment, rather than relying solely on statistical comparisons with a larger sample size. To strengthen our observations, the present data are in agreement with findings frequently reported in other larger-scale studies. Of course, expanding the number of samples and controls would strengthen our conclusion. However, taking into account the above-mentioned agreement with other studies, enrolling further subjects would be useful mostly to infer rare occurrences, while the primary aim of this paper is to examine the hypothesis of HBM by simply comparing its main abundances with controls. Of note, we have highlighted a concurrence in the profiles of our controls and the healthy blood samples, a finding that we consider highly informative regardless the number of samples.

Furthermore, our study paves the way to designing a more targeted procedural approach in this research area or, at least, aims to underscore the need for precautions that must be adopted when exploring such topics through NGS. Implementing relative abundance filters is advisable to avoid contaminants and sequencing errors in low bacterial biomass environments but, inevitably, arbitrary threshold values are applied. In this regard, we settled our thresholds after an accurate investigation of the literature attempting to achieve permissive yet efficient filtering strategies. Finally, it is important to point out that we purposely did not perform *in silico* decontaminations based on the comparison with the controls. This choice was made to focus on evaluating results achievable by sequencing the blood samples and the kitome itself, without incorporating statistics whenever possible. Alternative approaches, such as the use of decontamination algorithms, may change the estimated taxonomic profile of the blood samples. However, this study has not been designed to use such methods, which would perform better with more samples and not DNA inflated negative controls. In all cases, the raw reads in FASTQ format are released as publicly available (see below) to allow any re-analysis with different settings.

project "problem-driven" to be carried out through the establishment of public-private partnerships within the framework of the PNR (ex D. M. 737/2021).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

AA, MR, EN, and LDG made substantial contributions to the conception and design of the work. MR, AA, and LDG contributed to the design of the work. SBa, SBe, GN, LC, LDG, and FC were involved in the acquisition and analysis of data. LDG, EN, MR, SBa, and AA contributed to the interpretation of data. EN, LDG, and SBa drafted the work, and AA and MR revised it. All authors have approved the submitted version (and any substantially modified version that involves the author's contribution to the study) and have agreed to be personally accountable for their own contributions. They also agree to ensure that questions related to the accuracy or integrity of any part of the work, even those in which they were not personally involved, are appropriately investigated, resolved, and documented in the literature.

Peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/febs.17362.

Data availability statement

The sequencing reads for samples used in this study have been deposited in the NCBI GEO (Gene Expression Omnibus) database under the accession code GSE254843. Analysis scripts for the bioinformatics processing of reads and for the microbial community analyses are freely available at https://github.com/LeandroD94/Papers/tree/main/2024_seeking_blood_microbiota.

References

1 Najmanova L, Videnska P & Cahova M (2022) Healthy microbiome – a mere idea or a sound concept? *Physiol Res* **71**, 719–738.

2 Castillo DJ, Rifkin RF, Cowan DA & Potgieter M (2019) The healthy human blood microbiome: fact or fiction? *Front Cell Infect Microbiol* **9**, 148.

3 Damgaard C, Magnussen K, Enevold C, Nilsson M, Tolker-Nielsen T, Holmstrup P & Nielsen CH (2015) Viable bacteria associated with red blood cells and plasma in freshly drawn blood donations. *PLoS One* **10**, e0120826.

4 Dinakaran V, Rathinavel A, Pushpanathan M, Sivakumar R, Gunasekaran P & Rajendhran J (2014) Elevated levels of circulating DNA in cardiovascular disease patients: metagenomic profiling of microbiome in the circulation. *PLoS One* **9**, e105221.

5 Santiago A, Pozuelo M, Poca M, Gely C, Nieto JC, Torras X, Roman E, Campos D, Sarrabayrouse G, Vidal S *et al.* (2016) Alteration of the serum microbiome composition in cirrhotic patients with ascites. *Sci Rep* **6**, 25001.

6 Nikkari S, McLaughlin IJ, Bi W, Dodge DE & Relman DA (2001) Does blood of healthy subjects contain bacterial ribosomal DNA? *J Clin Microbiol* **39**, 1956–1959.

7 Moriyama K, Ando C, Tashiro K, Kuhara S, Okamura S, Nakano S, Takagi Y, Miki T, Nakashima Y & Hirakawa H (2008) Polymerase chain reaction detection of bacterial 16S rRNA gene in human blood. *Microbiol Immunol* **52**, 375–382.

8 Ramirez Galleymore P & Gordon Sahuquillo M (2019) Antisepsis for blood culture extraction. Blood culture contamination rate. *Med Intensiva (Engl Ed)* **43**(Suppl 1), 31–34.

9 Opota O, Croxatto A, Prod'hom G & Greub G (2015) Blood culture-based diagnosis of bacteraemia: state of the art. *Clin Microbiol Infect* **21**, 313–322.

10 Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ & Walker AW (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* **12**, 87.

11 Glassing A, Dowd SE, Galandiuk S, Davis B & Chiodini RJ (2016) Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog* **8**, 24.

12 Stinson LF, Keelan JA & Payne MS (2019) Identification and removal of contaminating microbial DNA from PCR reagents: impact on low-biomass microbiome analyses. *Lett Appl Microbiol* **68**, 2–8.

13 Nearing JT, Comeau AM & Langille MGI (2021) Identifying biases and their potential solutions in human microbiome studies. *Microbiome* **9**, 113.

14 Olomu IN, Pena-Cortes LC, Long RA, Vyas A, Krichevskiy O, Luellwitz R, Singh P & Mulks MH (2020) Elimination of "kitome" and "splashome" contamination results in lack of detection of a unique placental microbiome. *BMC Microbiol* **20**, 157.

15 Biesbroek G, Sanders EA, Roeselers G, Wang X, Caspers MP, Trzciński K, Bogaert D & Keijser BJ (2012) Deep sequencing analyses of low density microbial communities: working at the boundary of accurate microbiota detection. *PLoS One* **7**, e32942.

16 Degnan PH & Ochman H (2012) Illumina-based analysis of microbial community diversity. *ISME J* **6**, 183–194.

17 Walker SP, Barrett M, Hogan G, Flores Bueso Y, Claesson MJ & Tangney M (2020) Non-specific amplification of human DNA is a major challenge for 16S rRNA gene sequence analysis. *Sci Rep* **10**, 16356.

18 Khan I, Khan I, Jianye Z, Xiaohua Z, Khan M, Hilal MG, Kakakhel MA, Mehmood A, Lizhe A & Zhiqiang L (2022) Exploring blood microbial communities and their influence on human cardiovascular disease. *J Clin Lab Anal* **36**, e24354.

19 Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A & Mele MC (2019) What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms* **7**, 14.

20 Tan CCS, Ko KKK, Chen H, Liu J, Loh M, SG10K_Health Consortium, Chia M & Nagarajan N (2023) No evidence for a common blood microbiome based on a population study of 9,770 healthy humans. *Nat Microbiol* **8**, 973–985.

21 de Goffau MC, Lager S, Sovio U, Gaccioli F, Cook E, Peacock SJ, Parkhill J, Charnock-Jones DS & Smith GCS (2019) Human placenta has no microbiome but can contain potential pathogens. *Nature* **572**, 329–334.

22 Bordenstein SR & Theis KR (2015) Host biology in light of the microbiome: ten principles of Holobionts and Hologenomes. *PLoS Biol* **13**, e1002226.

23 Lam NY, Rainer TH, Chiu RW & Lo YM (2004) EDTA is a better anticoagulant than heparin or citrate for delayed blood processing for plasma DNA analysis. *Clin Chem* **50**, 256–257.

24 Foley C, O'Farrelly C & Meade KG (2011) Technical note: comparative analyses of the quality and yield of genomic DNA from invasive and noninvasive, automated and manual extraction methods. *J Dairy Sci* **94**, 3159–3165.

25 Kotikalapudi R & Patel RK (2015) Comparative study of the influence of EDTA and sodium heparin on long term storage of cattle DNA. *Cell J* **17**, 181–186.

26 Bingula R, Filaire E, Molnar I, Delmas E, Berthon JY, Vasson MP, Bernalier-Donadille A & Filaire M (2020) Characterisation of microbiota in saliva, bronchoalveolar lavage fluid, non-malignant, peritumoural and tumour tissue in non-small cell lung cancer patients: a cross-sectional clinical trial. *Respir Res* **21**, 129.

27 van Oorschot RA, Ballantyne KN & Mitchell RJ (2010) Forensic trace DNA: a review. *Investig Genet* **1**, 14.

28 Buckleton J (2009) Validation issues around DNA typing of low level DNA. *Forensic Sci Int Genet* **3**, 255–260.

29 Lusk RW (2014) Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS One* **9**, e110808.

30 Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R & Weyrich LS (2019) Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol* **27**, 105–117.

31 Jagare L, Rozenberga M, Silamikelis I, Ansone L, Elbere I, Briviba M, Megnis K, Konrade I, Birka I, Straume Z *et al.* (2023) Metatranscriptome analysis of blood in healthy individuals and irritable bowel syndrome patients. *J Med Microbiol* **72**, doi: 10.1099/jmm.0.001719

32 Karstens L, Asquith M, Davin S, Fair D, Gregory WT, Wolfe AJ, Braun J & McWeeney S (2019) Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments. *mSystems* **4**, e00290-19.

33 Soergel DA, Dey N, Knight R & Brenner SE (2012) Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J* **6**, 1440–1444.

34 Mazzoli L, Munz G, Lotti T & Ramazzotti M (2020) A novel universal primer pair for prokaryotes with improved performances for anammox containing communities. *Sci Rep* **10**, 15648.

35 Albertsen M, Karst SM, Ziegler AS, Kirkegaard RH & Nielsen PH (2015) Back to basics–the influence of DNA extraction and primer choice on phylogenetic analysis of activated sludge communities. *PLoS One* **10**, e0132783.

36 Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F *et al.* (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852–857.

37 The Genome Reference Consortium. https://www.ncbi.nlm.nih.gov/grc

38 Langmead B & Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. *Nat Methods* **9**, 357–359.

39 McMurdie PJ & Holmes S (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217.

40 Oksanen J, Simpson GL, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E *et al.* (2017) vegan: Community ecology package. https://cran.r-project.org/web/packages/vegan/vegan.pdf

41 Hadley W (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, New York, NY.

42 Piro VC & Renard BY (2022) Contamination detection and microbiome exploration with GRIMER. *GigaScience* **12**, giad017.

43 Legendre P & Legendre L (2012) Chapter 7 – ecological resemblance. In Developments in Environmental Modelling (Legendre P & Legendre L, eds), pp. 265–335. Elsevier, Amsterdam, The Netherlands.