# Metagenomic evidence for a healthy human blood microbiome

Cedric C.S. Tan

Department of Structural and Molecular Biology

University College London

Gower Street, London, WC1E 6BT, United Kingdom

# Graphical abstract

# Summary (150)

150 words

# Introduction (553)

Human blood has traditionally been considered to be sterile, where the occasional entry of microorganisms is believed to cause bloodstream infections (BSIs) associated with high mortality and morbidity (Goto and Al-Hasan, 2013). However, there have been an increasing number of studies reporting bacterial DNA or even culturable bacteria in the blood of apparently ‘healthy’ individuals (*i.e.* with no clinical symptoms of disease) (reviewed by Castillo *et al.* (2019)). These studies sought to investigate two main questions: is there a blood microbiome? If there is, what does the microbial profile look like? To do so, most studies have employed culture-based or sequencing approaches. The former approach involves the use of microbiological cultures of blood samples; the latter involves either targeted 16S marker-gene, or non-targeted ‘shotgun’ metagenomic sequencing of microbial DNA extracted from blood. Both types of approaches have their advantages and limitations. Culture-based methods demonstrate the viability of detected microorganisms. However, they cannot detect ‘unculturable’ microorganisms and have a notoriously low sensitivity, especially when the bacterial load is low (Benjamin and Wagner, 2007). Additionally, the microorganisms detected are largely biased by the type of culture medium used. This approach therefore prevents a comprehensive characterisation of the breadth of blood microorganisms.

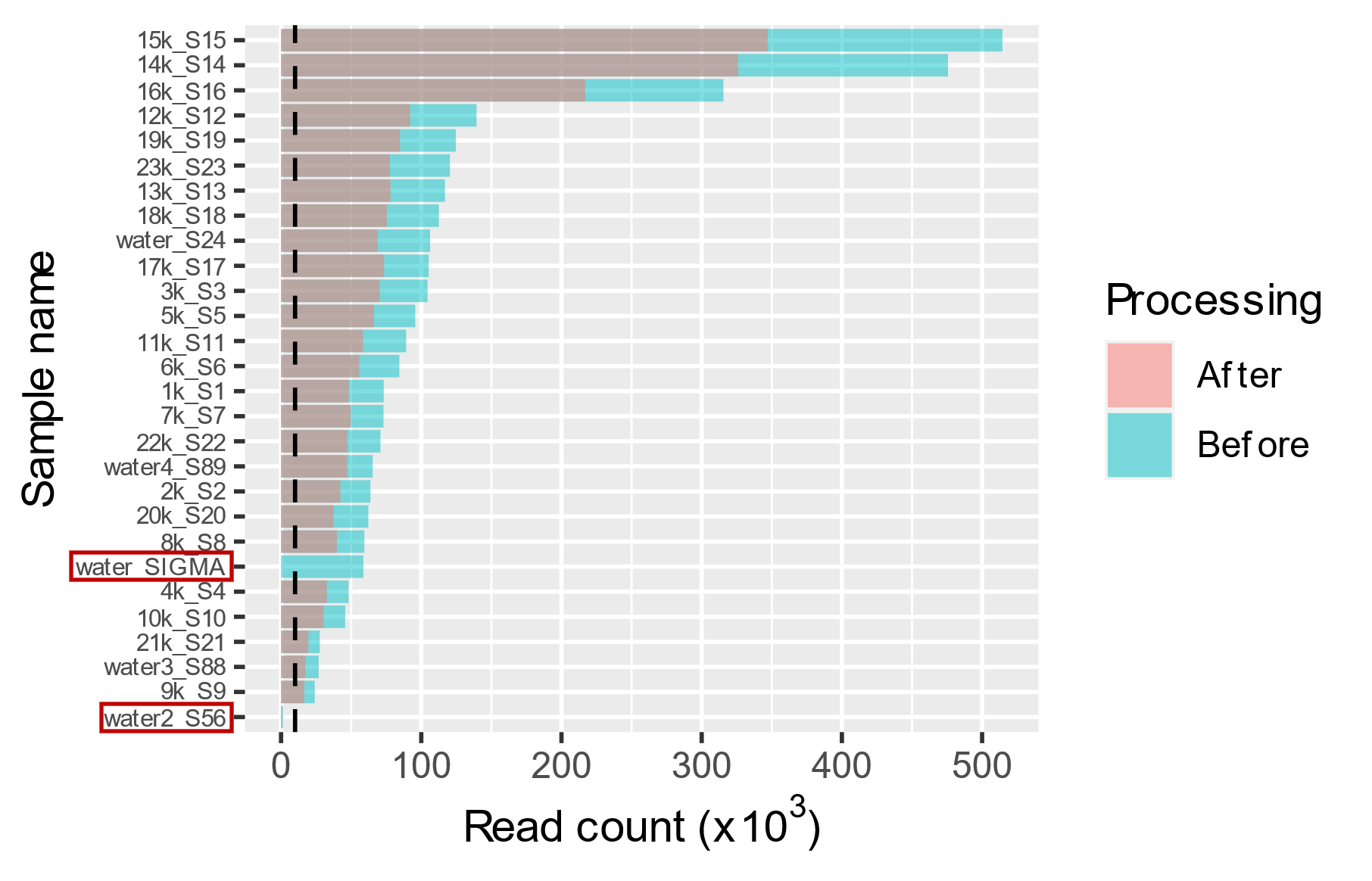
On the other hand, metagenomic sequencing can detect unculturable or low abundance microorganisms, and is therefore a more sensitive approach for profiling the blood microbiome. The main trade-off for this increased sensitivity is that sequencing is highly prone to environmental contamination. Contaminant DNA from laboratory reagents, referred to as the ‘kitome’, can be introduced during sample processing (Salter *et al.*, 2014; Weiss *et al.*, 2014; Glassing *et al.*, 2016). Alternatively, contaminant DNA may be introduced when using non-sterile equipment or poor aseptic technique during sample collection. These contaminating sequences are known to skew biological inferences of metagenomic sequencing data, particularly for sample types with low microbial biomass such as blood (Eisenhofer *et al.*, 2019). Delineating contaminant DNA sequences from those endogenous to blood is therefore crucial in lending weight to arguments for a blood microbiome. However, as Castillo *et al.* (2019) noted, experimental controls for addressing the contamination problem are often omitted in blood microbiome studies. Even when controls are included, they are often only used for quantifying the overall level of contamination introduced by the experimental protocols used (Païssé *et al.*, 2016; Gosiewski *et al.*, 2017). Such controls should ideally be used to identify and remove putative contaminant taxa before making any biological inference of the microbial profiles observed.

In this study, 16S marker-gene metagenomic sequencing (henceforth ‘16S sequencing’) was performed on microbial DNA extracted from the blood of 23 clinically healthy individuals to characterise the hypothetical human blood microbiome. To address the problem of contamination, five negative controls were also sequenced in the same sequencing run, and their taxonomic profiles compared at the genus rank. These controls had sterile water in place of DNA extracted from blood and were used to account for contaminant sequences introduced during library preparation and sequencing. Indeed, since the microbial taxa observed in the negative controls are assumed to be contaminants, these taxa can be selectively omitted to produce a more accurate depiction of the microbial composition in blood. The results suggest that the number and relative abundance of microbial taxa detected differed between the two sample types. This implies that the microbial sequences found in healthy human blood are more likely to originate from endogenous bacteria rather than from contamination sources during library preparation, providing support for the existence of a human blood microbiome.

# Results (1151)

## 16S metagenomic sequencing

Sequencing of the hypervariable region spanning V3-V4 of the bacterial 16S ribosomal RNA (rRNA) gene was peformed to characterise the microbial composition of 23 healthy human blood samples and five negative water controls. A total of 3,205,568 paired-end reads were generated for all 28 samples in a single sequencing run, and 2,120,549 reads were retained after all read processing steps. The distribution of reads per sample before and after these processing steps (**Figure 1**) suggests that most samples had a reasonable sequencing depth and overall read quality. Two negative controls with less than 10,000 reads are considered potentially problematic samples but were not excluded from further analysis to avoid introducing additional biases.



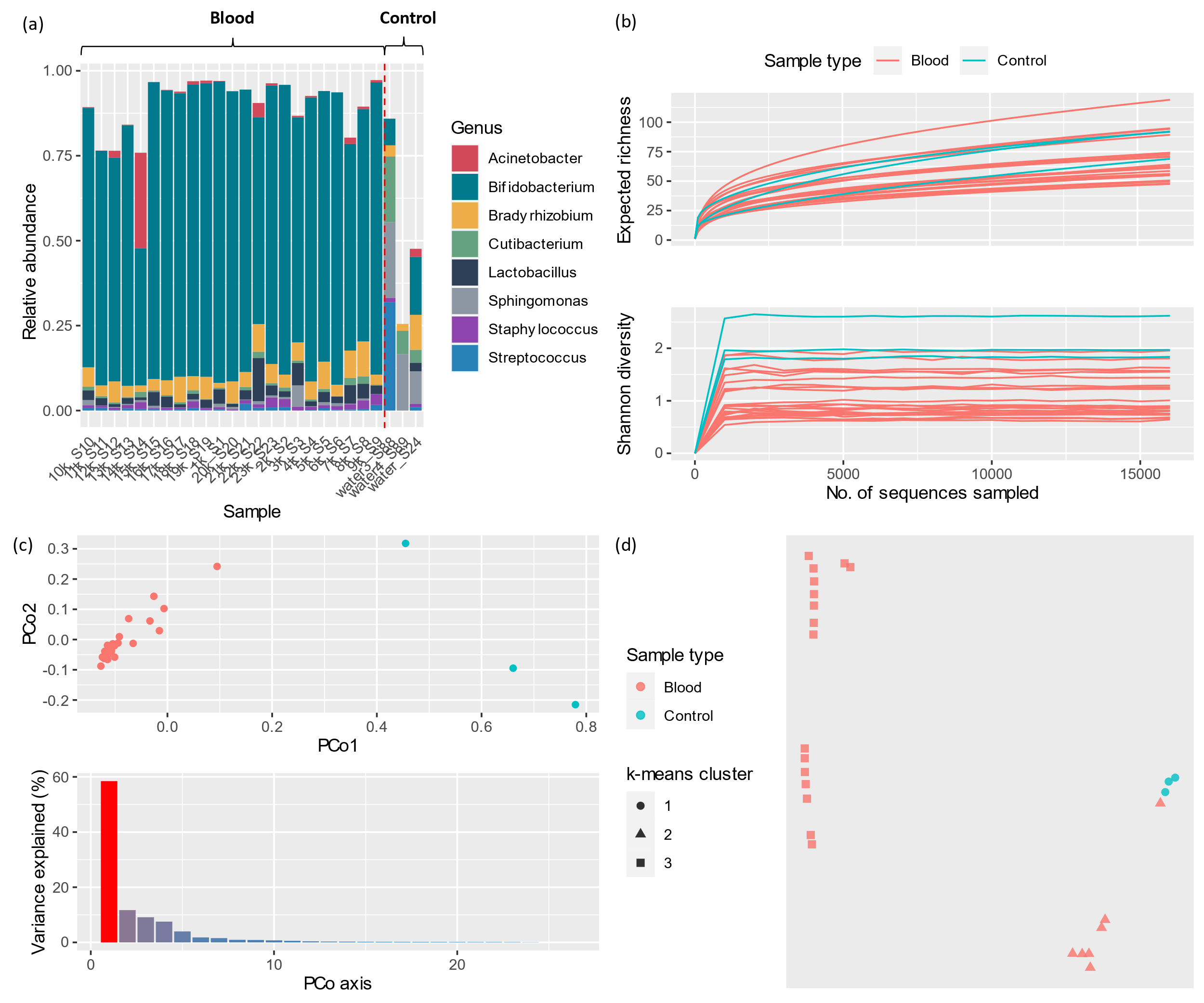
**Figure 1. Per-sample distribution of reads.** The read counts of samples before and after all read processing steps were extracted from the respective FASTQ files and visualised. The black dashed line indicates the read count threshold of 10,000 reads. Negative controls that were excluded are denoted by red boxes.

## Microbial composition of healthy human blood is distinct from that of negative controls

Taxonomic classification of the processed reads for each sample was performed using *Kraken 2* (Wood, Lu and Langmead, 2019). This strategy was used instead of conventional 16S taxonomic classification pipelines such as *QIIME 2* (Bolyen *et al.*, 2019) since *Kraken 2* was shown to be faster and more accurate (Lu and Salzberg, 2020). Additionally, since there is less uncertainty in taxonomic assignments at the genus than at the species level (Simon *et al.*, 2019), the taxonomic assignments were only used to calculate the relative abundance of genera for downstream analyses. The relative abundance of the eight most abundant genera in blood samples and in controls were visually distinct (**Figure 2a**). For example, *Bifidobacterium* predominates in blood samples, but not in the negative controls. To determine if the diversity of microbial sequences differed between the sample types, the genus richness and Shannon index (Shannon, 1948) for each sample was compared. Rarefaction was performed to account for the different sequencing depths for each sample (**Figure 1**). Blood samples and controls had comparable genus richness, but blood samples had a visibly lower Shannon index (**Figure 2b**). The Shannon index is a measure of both the number of genera (*i.e.* richness) and how equal the relative abundance of genera are (*i.e.* evenness). A higher Shannon index indicates a higher richness and/or evenness. Since the genus richness for the two sample types were comparable, the lower Shannon index for blood samples indicate a lower evenness, which is likely driven by the excess of *Bifidobacterium* sequences.

To further determine whether the microbial composition of the two sample types could be distinguished, ordination using principal coordinates analysis (PCoA) was performed. The first principal coordinate axis separated the two sample types clearly (**Figure 2c**; upper panel) and accounted for a major proportion of variation in the relative abundance data (58.5%; lower panel). Additionally, *t*-distributed stochastic neighbour embedding (*t*-SNE) was used to identify groups of samples with similar microbial compositions in an unsupervised fashion. In this context, *t*-SNE projects the relative genera abundance for each sample on a two-dimensional space, where samples with more similar microbial compositions are brought closer together, and those with dissimilar compositions further apart. Echoing the results of the PCoA, blood samples could be spatially separated from the negative controls (**Figure 2d**). The results of the two analyses collectively suggest that the two sample types have distinct microbial compositions. Interestingly, the blood samples appear to form more than one cluster on the *t*-SNE projection (**Figure 2d**). *k*-means clustering, which can classify samples with similar microbial compositions into groups, confirmed that blood samples could be classified into two groups, in addition to a separate negative control group. This suggests that the hypothetical healthy blood microbiome may be categorised into different subtypes.

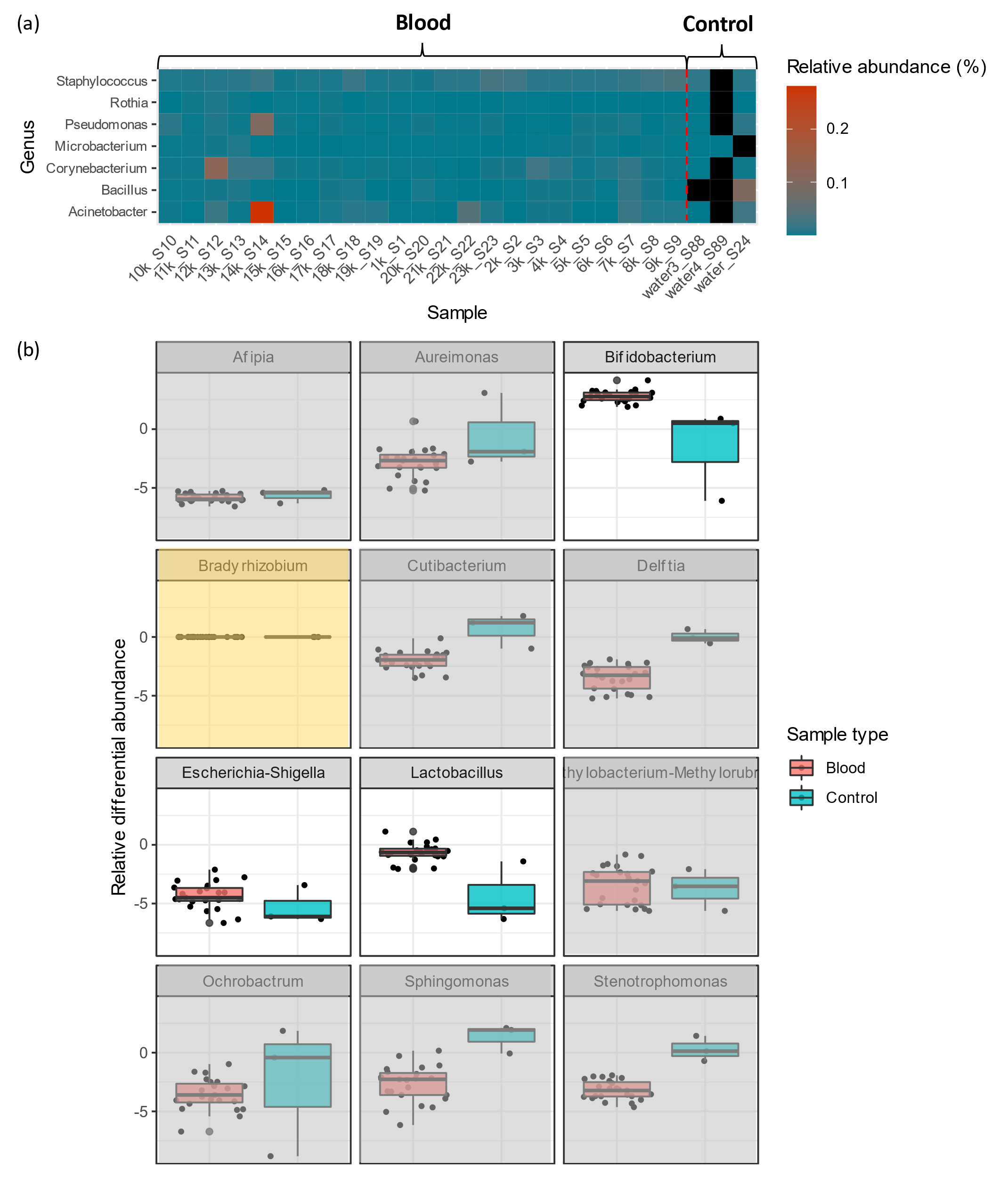
Taken together, the results indicate that the number and composition of microbial genera in blood samples can be distinguished from that in the negative controls. This implies that the microbial sequences in blood are not solely due to contamination from sequencing reagents, suggesting the presence of DNA from potentially viable bacteria in healthy human blood.



**Figure 2. Microbial composition and diversity of healthy blood and negative controls.** (a) Relative abundance of the eight most abundant genera (ranked by their median values across all samples) are visualised in a stacked bar-plot. (b) Rarefaction plots of genus richness and Shannon index. (c) Upper panel: PCoA plot of all samples on the first two PCo axes. Each point represents a single sample after ordination using PCoA. Lower panel: Percentage of variance explained by each PCo axis. (d) *t-*SNE plot projecting the relative genera abundance of samples on a two-dimensional space. Each point represents a single sample after ordination using *t*-SNE. The shape of each point represents the clusters identified using the *k*-means clustering algorithm.

## Candidate members of the human blood microbiome

Since all samples were processed and sequenced in the same batch, the taxonomic assignments for the negative controls could be used to identify potential microbiome-associated genera while excluding putative contaminants. Firstly, seven genera were found in all blood samples, but absent in at least one negative control (**Figure 3a**). Species within these genera present as preliminary targets for members of the blood microbiome. Notably, some genera are present in all samples, but may also be part of the human blood microbiome. One way to determine whether such genera are microbiome-associated or contaminants is to compare their relative abundance in blood and in controls (Salter *et al.*, 2014). This is because taxa that are present in a true microbial community are likely to be highly abundant in microbiome samples and less abundant in negative controls. Conversely, contaminant taxa are likely to be highly abundant in controls but not in samples. Therefore, the measured relative abundance of taxa is a good proxy for selecting candidate members of the human blood microbiome. However, if there is a microbial community in blood, blood samples should have a higher microbial load than negative controls. This differing microbial load is known to confound comparisons of relative abundance (Morton *et al.*, 2019). As such, relative differential abundance (RDA), where the relative abundance of a particular taxon is normalised by a reference taxon in the same sample (see Methods), was compared. RDA controls for the differing microbial loads in samples when comparing relative abundances. Using *Bradyrhizobium* as the reference genus, the RDA of genera that were found in all samples were computed. Of the twelve genera that were present in all samples, only *Bifidobacterium*, *Lactobacillus* and *Escherichia-Shigella* appeared to be more abundant in blood than in controls (**Figure 3b**). These genera are therefore candidate members of the human blood microbiome.



**Figure 3.** (a) Heatmap showing the relative abundance of genera that are found in all blood samples but not in all negative controls. Each tile represents the relative abundance of a single genus for a particular sample. Black tiles indicate a relative abundance of zero. (b) Boxplots of RDA for genera that were present in all samples. Each point represents a single sample. Panels that are not likely candidate members of the human blood microbiome are shaded grey, and the reference frame is shaded yellow.

# Discussion

## The healthy human blood microbiome

This study demonstrates that many of the microbial sequences detected in healthy human blood are not likely to originate from sequencing reagents and may be associated with an active microbial community. Moreover, *in silico* analysis of blood and negative controls allowed the careful identification of potential members of the hypothetical blood microbiome, which are prime targets for further experimental validation. Some of these genera have been identified in human blood samples previously. For example, *Acinetobacter*, *Staphylococcus,* *Pseudomonas* and *Bacillus* species were isolated from at least one healthy blood sample via microbiological culture, with *S*. *epidermis* found in 23% of samples (Damgaard *et al.*, 2015). This indicates that live microbial cells from said genera can indeed be found in healthy human blood. Additionally, in another metagenomic sequencing study involving 30 healthy French blood donors, *Acinetobacter*, *Escherichia/Shigella*, *Corynebacterium*, *Pseudomonas*, and *Staphylococcus* sequences were also found (Païssé *et al.*, 2016). Given that different experimental protocols were used, and that the samples were from a different cohort, these findings serve as independent validation of our results.

Separately, our results suggest that the gut and mouth may be potential sources of the bacteria found in blood. For example, *Bifidobacterium*, *Lactobacillus* and *Rothia* are often found in the human oral cavity (Marsh, 2000; Hojo *et al.*, 2007) and gastrointestinal tract (Reuter, 2001; Turroni *et al.*, 2014). The detection of these genera in blood suggests that they may have translocated from other bodily sites via breaches in the mucosal barrier. In fact, the translocation of bacteria across the intestinal wall has been previously described (Berg, 1999). In particular, *Bifidobacterium breve* has even been used as an orally administered vehicle for delivering anti-cancer therapeutic agents to tumour cells (Cronin *et al.*, 2010). Since bacteria can penetrate mucosal barriers and bloodstream is a nutrient rich environment, it seems plausible that bacterial cells from other bodily sites can colonise and establish a microbial community in blood.

-Microbiome subtypes. Reference other examples. Emphasise preliminary https://www.nature.com/articles/s41598-019-44012-w#Sec2

Other studies have also identified

-16S so bacteria only. There is also virome (<https://www.nature.com/articles/s41598-020-72808-8>) (near 100% prevalence of anelloviridae in humans worldwide Abe, K., Inami, T., Asano, K., Miyoshi, C., Masaki, N., Hayashi, S., Ishikawa, K., Takebe, Y., Win, K. M., and El-Zayadi, A. R. (1999) TT virus infection is widespread in the general populations from different geographic regions. *J. Clin. Microbiol.* **37**, 2703–2705

## Clinical implications of the blood microbiome

PAMPs

Additionally, *Acinetobacter* *Baumanii* sequences were found in 23% of type II diabetic patients (Perera *et al.*, 2021). The presence of thesesequences was positively correlated with proinflammatory cytokines and associated with chronic inflammation. While this finding does not involve healthy indivudals

## Limitations and suggestions

This study focused on contamination arising from reagents during library preparation and sequencing. However, contaminant microbial sequences may be introduced at any step along the process from sample collection to sequence. Ideally, appropriate negative controls for each sample processing step should be included. For example, Moriyama *et al.* (2008) used saline water exposed to povidone-iodine skin as a negative control for phlebotomy-associated contamination. Additionally, DNA extraction can be performed on skin-exposed saline water to control for extraction-associated contamination. Alternatively, sterile water can be used in place of blood during DNA extraction. Comparisons of microbial taxonomic profiles obtained from these different controls would allow us to track contamination introduced by the different sample processing steps. This then permits more stringent filtering of contaminant taxa from the observed taxonomic profiles for blood. Separately, other strategies to computationally reduce the effect of contamination on observed taxonomic profiles may be employed. For example, dilutions of blood samples with DNA-free water to yield samples with multiple DNA concentrations post-extraction could be performed. In such a study design, taxa whose observed relative abundance is inversely correlated to the sample DNA concentration can be considered contaminants (Jervis-Bardy *et al.*, 2015). Such alternative strategies can be used in tandem with the contamination analysis presented in this study for more stringent contaminant removal.

Separately, metagenomic sequencing measures the DNA molecules present and not live microbial cells. Cell-free DNA molecules entering the bloodstream from other bodily sites such as the mouth or gut may therefore be an alternative explanation for the taxonomic profiles observed in the blood samples. Additionally, even if the detected DNA sequences reflect the presence of live cells in blood, it remains unclear whether these microbial cells are transient or persistent. This is because the current study design offers only a snapshot of the microbial sequences present. Given these limitations, it must be stressed that further experimental validation is necessary to prove the existence of a hypothetical blood microbiome. One possible strategy is to probe the microbial mRNA profile in blood. For example, if metagenomic DNA sequencing identified *Escherica coli* in blood, we expect to also observe *E. coli*-specific mRNA species. Concordance between DNA and RNA sequencing results therefore provides stronger evidence for an active microbial community in blood. Additionally, sequencing of blood samples from the same individuals across different time points (*i.e.* a longitudinal study) may shed light on the persistence of detected microorganisms. Such a study design also permits modelling of how the population numbers of microbial taxa changes with time, offering insight into their underlying ecological relationships and interactions (Faust and Raes, 2012). Agreement of these findings with known microbial interactions would provide additional evidence for an active community.

Finally, a limitation of this study is the small sample size. The number of controls (*n* = 3) was much lower relative to blood samples (*n* = 23), resulting in highly unequal group sizes. Statistical tests are not meaningful in this context and were therefore avoided to prevent over-interpreting the data. While the addition of more negative controls would have permitted formal statistical testing, this was arguably unnecessary. Indeed, all analyses presented in this study, except that for rarefied genus richness, concordantly suggest an obvious difference between the microbial taxonomic profiles of the blood and control groups (**Figure 2**).

Irrespective of these limitations, our results nonetheless demonstrate the importance of considering the full polymicrobial component of sepsis and suggest that a metagenomics-based approach may provide biological and clinical insights supporting the future development of rapid diagnostic tools.

In this study, we demonstrate the promise of a metagenomics-based approach to sepsis. Our results provide evidence that septic infections should be considered as polymicrobial in nature, comprising multiple co-occurring pathogens indicative of disease. Our findings thus pave the way for more microbial-focused models of sepsis, with long run potential to inform early detection, clinical interventions and improve patient outcomes.

# Methods

## Sample collection

Phlebotomies were carried out on 13 female and 10 male volunteers (median age = 59). These individuals had no clinical symptoms of bloodstream infections and had normal levels of the C-reactive protein inflammatory marker. All volunteers were not undergoing antibiotic treatment prior to sample collection. Blood samples were collected in 4ml Vacutainer K3E (Becton Dickinson) tubes.

## DNA extraction

Microbial DNA was extracted from blood using the Blood Mini kit (A&A Biotechnology), as described previously (Gosiewski, Szała, *et al.*, 2014). The concentration and purity of extracted DNA was determined using NanoDrop (Thermo Scientific). Sterile water was used as a substitute for the extracted DNA in the negative water controls.

## 16S rRNA library preparation and sequencing

Paired-end sequencing of the hypervariable V3-V4 regions of 16S rRNA gene from extracted DNA was performed on the Illumina MiSeq System. Library preparation and sequencing was carried out in accordance with the manufacturer’s instructions (<https://tinyurl.com/ujr5j45b>). Briefly, the hypervariable V3-V4 regions were amplified via a nested polymerase chain reaction (PCR) procedure, as described previously (Gosiewski, Jurkiewicz-Badacz, *et al.*, 2014). PCR was carried out using the KAPA HiFi HotStart ReadyMix kit (KAPA BIOSYSTEMS) on a C100 thermal cycler (BioRad). The final DNA concentration of amplicon libraries for each sample was quantified using PicoGreen (Life Technologies). The libraries were then normalised to the same concentration and pooled to yield a single sequencing library. The sequences of all primers and adapters used are shown in Table 1. Sequencing was performed using the MiSeq Reagent Kit v3 (600 cycles) with a 20% PhiX spike-in as an internal control.

## **Sequencing read processing**

Raw sequencing reads were demultiplexed using bcl2fastq (v2.17.1.14; default parameters). Primers and sequencing adapters were removed, and low-quality bases at the start and end of reads were trimmed using *Cutadapt* (v2.6) (Martin, 2011). Read pairs were then merged using *ea-utils* (v1.1.2.779) (Aronesty, 2013). For read pairs where merging failed, only forward unmerged reads were retained. Reads with a Q-score below 20 removed using *BBMap* (v38.18) (Bushnell, 2014). Samples with less than 10,000 reads were excluded from further analysis. Read quality assessments before and after processing was performed using the *fastqcr* package (v0.1.2) (Andrews, 2010) in *R*.

## Taxonomic classification

*Kraken 2* (v2.0.7-beta; default parameters) (Wood, Lu and Langmead, 2019) was used to assign each processed read to a particular taxon based on the *Kraken 2*-built *SILVA* database (v138) (Quast *et al.*, 2012). All unclassified reads were excluded from the analyses. The taxonomic assignments for all reads were aggregated to generate a relative abundance matrix using a custom *Python* script written by Tan *et al.* (2020). In this matrix, samples are represented as rows and genera as columns. Each element equals the number of reads assigned to a particular genus for a single sample divided by the total number of reads for said sample.

## Microbial diversity

The rarefaction analysis of genus richness was performed using the *rarefy* function from the *Vegan* package (v2.5.7) in *R*. Rarefaction analysis of Shannon index was performed using a custom *R* script and the *diversity* function from the same package.

## Ordination and clustering

The Bray-Curtis distance matrix was computed from the relative abundance matrix using the *vegdist* function in *Vegan*. PCoA, *t*-SNE and *k*-means was computed using the *pcoa*, *Rtsne* and *kmeans* functions in the *ape* (v5.4.1) (Paradis and Schliep, 2019), *Rtsne* (v0.15) (Krijthe, van der Maaten and Krijthe, 2018) and *stats* (v3.6.0) packages in *R*, respectively. The Barnes-Hut implementation of *t*-SNE (Van Der Maaten, 2014) was used with the perplexity parameter set to two. The ‘centers’ and ‘nstart’ parameters in *kmeans* was set to three and 25 respectively.

## Relative differential abundance

As suggested previously (Morton *et al.*, 2019), RDA was computed using the equation, , where *xtarget*and *xreference* is the relative abundance of the genus in question and the reference genus, respectively. *Bradyrhizobium* was selected as the reference genus since it is the most common sequencing contaminant in metagenomic sequencing studies (Laurence, Hatzis and Brash, 2014) and therefore not likely to be part of the blood microbiome. This is important because the RDA of a reference taxon always equals one and therefore cannot be interpreted meaningfully.

# Data and code availability

All analyses or data processing was performed in *R* v3.6.0 or *Python* v3.6.13. Relevant source code is hosted on GitHub (<https://github.com/cednotsed/BIOC0023_dissertation>). The raw sequence data for this study was obtained from the study by Gosiewski *et al.* (2017).

# References

Andrews, S. (2010) ‘FastQC: a quality control tool for high throughput sequence data’. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Aronesty, E. (2013) ‘Comparison of sequencing utility programs’, *The open bioinformatics journal*, 7(1).

Benjamin, R. J. and Wagner, S. J. (2007) ‘The residual risk of sepsis: modeling the effect of concentration on bacterial detection in two‐bottle culture systems and an estimation of false‐negative culture rates’, *Transfusion*. Wiley Online Library, 47(8), pp. 1381–1389.

Berg, R. D. (1999) ‘Bacterial translocation from the gastrointestinal tract’, *Mechanisms in the pathogenesis of enteric diseases 2*. Springer, pp. 11–30.

Bolyen, E. *et al.* (2019) ‘Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2’, *Nature biotechnology*. Nature Publishing Group, 37(8), pp. 852–857.

Bushnell, B. (2014) *BBMap: a fast, accurate, splice-aware aligner*. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

Castillo, D. J. *et al.* (2019) ‘The healthy human blood microbiome: Fact or fiction?’, *Frontiers in cellular and infection microbiology*. Frontiers, 9, p. 148.

Cronin, M. *et al.* (2010) ‘Orally administered bifidobacteria as vehicles for delivery of agents to systemic tumors’, *Molecular Therapy*. Elsevier, 18(7), pp. 1397–1407.

Damgaard, C. *et al.* (2015) ‘Viable bacteria associated with red blood cells and plasma in freshly drawn blood donations’, *PLoS One*. Public Library of Science, 10(3), p. e0120826.

Eisenhofer, R. *et al.* (2019) ‘Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations’, *Trends in Microbiology*, 27(2), pp. 105–117. doi: https://doi.org/10.1016/j.tim.2018.11.003.

Faust, K. and Raes, J. (2012) ‘Microbial interactions: from networks to models’, *Nature Reviews Microbiology*. Nature Publishing Group, 10(8), pp. 538–550.

Glassing, A. *et al.* (2016) ‘Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples’, *Gut pathogens*. Springer, 8(1), p. 24.

Gosiewski, T., Jurkiewicz-Badacz, D., *et al.* (2014) ‘A novel, nested, multiplex, real-time PCR for detection of bacteria and fungi in blood’, *BMC microbiology*. Springer, 14(1), pp. 1–7.

Gosiewski, T., Szała, L., *et al.* (2014) ‘Comparison of methods for isolation of bacterial and fungal DNA from human blood’, *Current microbiology*. Springer, 68(2), pp. 149–155.

Gosiewski, T. *et al.* (2017) ‘Comprehensive detection and identification of bacterial DNA in the blood of patients with sepsis and healthy volunteers using next-generation sequencing method-the observation of DNAemia’, *European Journal of Clinical Microbiology & Infectious Diseases*. Springer, 36(2), pp. 329–336.

Goto, M. and Al-Hasan, M. N. (2013) ‘Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe’, *Clinical Microbiology and Infection*. Elsevier, 19(6), pp. 501–509.

Hojo, K. *et al.* (2007) ‘Distribution of salivary Lactobacillus and Bifidobacterium species in periodontal health and disease’, *Bioscience, biotechnology, and biochemistry*. Taylor & Francis, 71(1), pp. 152–157.

Jervis-Bardy, J. *et al.* (2015) ‘Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data’, *Microbiome*. Springer, 3(1), p. 19.

Krijthe, J., van der Maaten, L. and Krijthe, M. J. (2018) ‘Package “Rtsne”’. GitHub.

Laurence, M., Hatzis, C. and Brash, D. E. (2014) ‘Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes’, *PloS one*. Public Library of Science, 9(5), p. e97876.

Lu, J. and Salzberg, S. L. (2020) ‘Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2’, *Microbiome*, 8(1), p. 124. doi: 10.1186/s40168-020-00900-2.

Van Der Maaten, L. (2014) ‘Accelerating t-SNE using tree-based algorithms’, *The Journal of Machine Learning Research*. JMLR. org, 15(1), pp. 3221–3245.

Marsh, P. D. (2000) ‘Role of the oral microflora in health’, *Microbial Ecology in Health and Disease*. Taylor & Francis, 12(3), pp. 130–137.

Martin, M. (2011) ‘Cutadapt removes adapter sequences from high-throughput sequencing reads’, *EMBnet. journal*, 17(1), pp. 10–12.

Moriyama, K. *et al.* (2008) ‘Polymerase chain reaction detection of bacterial 16S rRNA gene in human blood’, *Microbiology and immunology*. Wiley Online Library, 52(7), pp. 375–382.

Morton, J. T. *et al.* (2019) ‘Establishing microbial composition measurement standards with reference frames’, *Nature Communications*, 10(1), p. 2719. doi: 10.1038/s41467-019-10656-5.

Païssé, S. *et al.* (2016) ‘Comprehensive description of blood microbiome from healthy donors assessed by 16 S targeted metagenomic sequencing’, *Transfusion*. Wiley Online Library, 56(5), pp. 1138–1147.

Paradis, E. and Schliep, K. (2019) ‘ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R’, *Bioinformatics*. Oxford University Press, 35(3), pp. 526–528.

Perera, D. *et al.* (2021) ‘Impaired host response and the presence of Acinetobacter baumannii in the serum microbiome of type-II diabetic patients’, *Iscience*. Elsevier, 24(1), p. 101941.

Quast, C. *et al.* (2012) ‘The SILVA ribosomal RNA gene database project: improved data processing and web-based tools’, *Nucleic acids research*. Oxford University Press, 41(D1), pp. D590–D596.

Reuter, G. (2001) ‘The Lactobacillus and Bifidobacterium microflora of the human intestine: composition and succession’, *Current issues in intestinal microbiology*, 2(2), pp. 43–53.

Salter, S. J. *et al.* (2014) ‘Reagent and laboratory contamination can critically impact sequence-based microbiome analyses’, *BMC biology*. Springer, 12(1), p. 87.

Shannon, C. E. (1948) ‘A mathematical theory of communication’, *The Bell system technical journal*. Nokia Bell Labs, 27(3), pp. 379–423.

Simon, H. Y. *et al.* (2019) ‘Benchmarking metagenomics tools for taxonomic classification’, *Cell*. Elsevier, 178(4), pp. 779–794.

Tan, C. C. S. *et al.* (2020) ‘Metagenomic evidence for a polymicrobial signature of sepsis’, *bioRxiv*. Cold Spring Harbor Laboratory.

Turroni, F. *et al.* (2014) ‘Bifidobacterium bifidum as an example of a specialized human gut commensal’, *Frontiers in microbiology*. Frontiers, 5, p. 437.

Weiss, S. *et al.* (2014) ‘Tracking down the sources of experimental contamination in microbiome studies’, *Genome biology*. BioMed Central, 15(12), p. 564.

Wood, D. E., Lu, J. and Langmead, B. (2019) ‘Improved metagenomic analysis with Kraken 2’, *Genome biology*. Springer, 20(1), p. 257.