UNIVERSITY OF TURKU

Circulating bacterial DNA and its taxonomic characterisation

Institute of Biomedicine
MDP in Biomedical Sciences, Drug Discovery and Development
Master's thesis

Author:
Heidi Aarnio

13.5.2025
Turku

Master's thesis

Subject: Drug Discovery and Development
Author(s): Heidi Aarnio
Title: Circulating bacterial DNA and its taxonomic characterisation
Supervisor(s): Docent Arno Hänninen, PhD Teemu Kallonen
Number of pages: 56 pages
Date: 13.5.2025

The role of human microbiome has been broadly investigated in disease and health. It has been shown that bacterial DNA circulate in the blood during various ailments. The presence of this circulating bacterial DNA (cbDNA) has been linked to onset and prognosis of various diseases. In addition to disease, cbDNA has been recently located from the blood of healthy individuals as a part of healthy blood microbiome (HBM). After the discovery of HBM, studies comparing the taxonomic composition of cbDNA between different patient populations and healthy volunteers have been conducted, showing the effect of cbDNA in multiple disease settings.

This thesis is a precursory study to the role of cbDNA in systemic lupus erythematosus, a multiorgan rheumatic autoimmune disease with broad range of symptoms, focusing on characterising the taxonomy of HBM in different blood fractions; whole blood (WB), buffy coat (BC) and peripheral blood mononuclear cells (PBMCs). In this study DNA was extracted from blood samples, followed by 16S rRNA PCR and Illumina MiSeq sequencing. The data was processed in Galaxy using modified Mothur standard operating protocol.

We found Actinobacteria to be the most common phylum in the PBMCs, BC and WB of healthy donors followed by Proteobacteria, Firmicutes and Bacteroidetes. We also found that different genera have preferential localisation within the blood fractions. The composition of bacteria phyla in the samples resembled skin microbiome. HBM was found to be diverse, dynamic, and individual. The methods used do however have some limitations and more research is needed.

Key words: 16S rRNA PCR, cbDNA, Healthy blood microbiota

Table of contents

1 Introduction

1.1 Systemic lupus erythematosus

Systemic lupus erythematosus (SLE) is a multisystemic rheumatic autoimmune disease with diverse phenotype and broad range of clinical symptoms. SLE is characterized by autoantibody production that can lead to organ damage in varying systems in the body. (Ameer et al., 2022; Justiz Vaillant et al., 2023)

1.1.1 Epidemiology, prevalence, and incidence

Globally SLE is estimated to affect 3.42 million people with incidence of 0.34 million new cases annually (Tian et al., 2023). The prevalence and incidence of SLE vary significantly between sexes, populations, and ethnic groups. Women between the ages of 20-30 are predominantly at risk of developing SLE with 9 to 1 female to male ratio. However, SLE in men is often more severe. The risk of SLE development is also present in elderly and paediatric populations. The disease tends to be more severe in children with early age of onset. Later age of onset is usually connected to milder disease. The risk of SLE decreases significantly after menopause. (Justiz Vaillant et al., 2023)

While epidemiological data is not available in all countries, high-income countries have higher incidence of SLE compared to low-income countries with the highest incidence in Poland, USA, Barbados, and China (Tian et al., 2023). In USA, African Americans have a higher risk and earlier onset of SLE compared to Caucasian population. Asian population also has increased risk of SLE compared to Caucasians. (Ameer et al., 2022; Justiz Vaillant et al., 2023)

1.1.2 Aetiology

Multiple different genetic, environmental, endocrine, and immunological factors are associated with the development of SLE (Akhil et al., 2023). Despite the effort, the aetiology of SLE is still largely unknown. Over 130 genes affecting either poly- or monogenic SLE have been discovered. Most notably deficiencies in complement components C1q, C1r and C1s are associated with over 90% risk of SLE, C4 and C2 regions with risks of 50% and 20%, respectively. These mutations, however, are rare and explain only a fraction of SLE cases. (Justiz Vaillant et al., 2023) SLE is also suggested to have some association with the X chromosome. This is supported by the fact that in addition to the 9 to 1 ratio of female to male

patients, the risk of developing SLE increases 14-fold in patients with 47 XXY or Klinefelter syndrome. (Weckerle and Niewold, 2011) Environmental factors associated with SLE include sun exposure, exposure to certain drugs (drug-induced lupus) and tobacco smoke. Immunological factors such as viral infections are also suspected to play a role as Epstein-Barr virus antibodies are more frequent in SLE patients than in controls. (Akhil et al., 2023; Justiz Vaillant et al., 2023) The leading endocrine etiologic factor of SLE is the influence of female sex hormones, although the mechanism is unknown (Weckerle and Niewold, 2011).

1.1.3 Pathogenesis

Multiple pathologic immunological pathways have been recognised behind the development of SLE, but the mechanism and cause behind autoimmunity activation remains unclear. The continuous B- and T-cell activation, caused by immune system exposure to self-antigens, leads to cytokine release and complement activation that causes autoantibody production and leads to organ damage. The pathogenic function of these autoantibodies eventually causes more B- and T-cell activation leading to self-sustained loop of autoimmunity. The whole process in complex and involves many immunological pathways, many of which are not fully understood. (Ameer et al., 2022; Akhil et al., 2023; Justiz Vaillant et al., 2023)

1.1.4 Clinical symptoms

The range of clinical symptoms caused by SLE is vast and symptoms can present in most organs. Severity of symptoms also vary from mild mucocutaneus changes to severe multiorgan failures and central nervous system deficits. (Ameer et al., 2022) Most common symptoms include constitutional, mucocutaneus and musculoskeletal manifestations occurring in 80-90% of patients. These include a wide array of different symptoms such as fever, weight loss, fatigue, different rashes and skin lesions, arthralgia, arthritis and many more. Hematologic manifestations are also common with anaemia being the most common and present in about 50% of patients. SLE can also manifest in the nervous-, renal-, pulmonary-, cardiovascular- and gastrointestinal systems with various symptoms and severity. Pregnancy complications and ocular symptoms are also common. (Justiz Vaillant et al., 2023)

1.1.5 Diagnosis, biomarkers, and current treatment

The diagnosis of SLE can be difficult. Several different classification criteria have been developed over the years. Currently the most used classification criteria are the “Classification

Criteria for Systemic Lupus Erythematosus" by European League Against Rheumatism and the American College of Rheumatology published in 2019 (Aringer et al., 2019). Diagnosis is based on clinical symptoms together with laboratory test results in domain immunology. The presence of antinuclear antibodies is prerequisite for SLE diagnosis. In the clinical setup, diagnosis can be difficult to make highlighting the importance of ruling out other options while making SLE diagnosis. Antinuclear antibodies are considered the hallmark of SLE and are always present, however, they are not inclusive to SLE. Other biomarkers include anti-double-stranded DNA, present in 60-70% of patients with 95% specificity, anti-Ro, anti-La, Anti-Smith antibodies and common markers for inflammation. (Ameer et al., 2022)

The treatment of SLE depends heavily on the symptoms, severity and the organ systems involved. Aim of the treatment is to prevent damage to organs and cause remission of the disease. Despite the treatment SLE still causes significant mortality and morbidity risk. The prognosis is improved by early diagnosis and treatment. With current treatment, the survival rate during first ten years after diagnosis is 85-90%. Most common causes of death in SLE are infections, renal disease, and cardiovascular events. (Justiz Vaillant et al., 2023)

1.2 Human microbiome, dysbiosis and bacterial translocation

The human microbiome and its connection to disease has received increasing attention in the recent years. The human microbiome consists of taxa from bacterial, viral, fungal, and archaeal sources in exterior and interior locations around the human body (Castillo et al., 2019). Bacteria present in the human microbiome amounts to about 0.2 kg of the human body mass (Sender et al., 2016b). The number of bacteria in the body is often cited to be 10-fold compared to human cells. In reality, the relative abundance alters due to excretion of the intestinal bacterial flora caused by defecation and is closer to 1:1 ratio (Sender et al., 2016a; b). The microbiota lives in symbiosis with multiple different niches around the body having impact in health and disease (Khan et al., 2022).

The most complex and prominent bacterial niche in the body is in the gut. Gut microbiota has many roles in physiological functions including digestion, moreover, it plays an important role in metabolism and immune function (Goraya et al., 2022).While gut microbiome differs between individuals and alters between the compartments of the gastrointestinal tract, it mostly consists of *Firmicutes, Bacteroidetes, Actinobacteria, Fusobacteria, Proteobacteria* and *Verrucomicrobia* phyla out of which *Bacteroidetes* and *Firmicutes* are the most prominent. The most common genera in gut within the *Bacteroidetes* phylum are *Bacteroides*

and *Prevotella*. Within the *Firmicutes* phylum on the other hand, *Clostridium, Eubacterium* and *Ruminococcus* are the most common. (Adak and Khan, 2019)

In addition to gut, oral cavity, skin, and vagina are abundant microbial niches. Oral microbiota consists mainly of anaerobic gram-negative and gram-positive bacteria within the phyla *Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Proteobacteria, Saccharibacteria* and *Spirochaetes* (Goraya et al., 2022; Baker et al., 2024). Disease conditions of the mouth, such as caries and periodontitis have major effect on the oral microbiota (Goraya et al., 2022). Skin microbiota, on the other hand, is mainly comprised of *Actinobacteria, Firmicutes, Bacteroidetes and Proteobacteria* phyla. The abundance of genera depends on the area of the skin with *Staphylococcus* and *Corynebacterium* being the most common genera in moist areas of the skin. (Grice and Segre, 2011) Lastly, the vaginal microbiota that is dominated by the *Lactobacilli* genera, which notably functions to prevent the colonisation of pathogens within the vagina (Chee et al., 2020). The possible microbial niches of enclosed compartments of the human body, including circulatory compartments such as cerebrospinal fluid and blood, are less explored but have received cumulative interest as of late (Goraya et al., 2022).

In recent years, the role of human microbiome in disease and health has been broadly investigated. Relations between the disturbance of the microbiota in different compartments of the body, especially the gut, and many diseases, including type 1 diabetes and rheumatoid arthritis, have been shown (Mu et al., 2015). This dysbiosis has also been recognised in the case of SLE (Pan et al., 2021). In addition to dysbiosis, bacterial translocation from one niche to other, primarily in the gut-oral axis, has been recognised in many diseases including inflammatory bowel disease, breast cancer and rheumatoid arthritis (Jin et al., 2022).

1.3 Healthy blood microbiome (HBM)

1.3.1 Discovery of circulating bacterial DNA (cbDNA) in healthy blood

Bacteria or parts of them, including DNA, have been shown to circulate in the bloodstream during various ailments. The role of circulating bacterial DNA (cbDNA) in disease is still mostly unknown, however, the levels of cbDNA have been linked to systemic inflammation and immune activation in peritoneal dialysis patients and in common variable immunodeficiency (Kwan et al., 2013; Ho et al., 2021). Additively, the presence of cbDNA,

the species variety of cbDNA located from the blood of prostate and breast cancer patients have been shown to have prognostic value to disease progression (Xiao et al., 2021).

More recently, cbDNA has been located from the blood of healthy individuals in addition to disease settings (Nikkari et al., 2001; Païssé et al., 2016; Panaiotov et al., 2018; Velmurugan et al., 2020; Whittle et al., 2018). Bacterial presence in the blood has been previously considered as a sign of infection or sepsis. The notion of bacteria population in niches previously considered sterile, such as blood, is fairly new and still of then contested. (Nikkari et al., 2001) The discovered quantity of cbDNA located in the healthy blood, also known as the healthy blood microbiome (HBM), is considerably lower compared to sepsis (Païssé et al., 2016). Because of the low cbDNA quantity, the samples are extremely prone to external contamination. This has been one of the reasons why the notion of HBM has received criticism. Besides contamination, other arguments against the notion of HBM have included transient bacteraemia, erythrocyte derived microparticles, membrane vesicle artefacts or blood protein aggregates interpretated as cbDNA (Martel et al., 2017; Mitchell et al., 2016).

The first evidence of HBM was discovered in 1969 when bacterial L-forms were discovered from healthy blood (Tedeschi et al., 1969). Since then, multiple studies characterising the HBM have been conducted (Castillo et al., 2019). In addition to humans, HBM has been discovered from birds and domesticated animals. Multiple independent studies with different methods have reported similar results of the HBM profile at the phyla level consolidating the presence of cbDNA in healthy blood. (Goraya et al., 2022)

The HBM is reported to be highly dynamic in terms of composition and abundance, with significant changes over time and between individuals (Païssé et al., 2016). The composition can be affected by multiple factors including age, environment and lifestyle factors (Goraya et al., 2022). The bacteria present in blood is suggested to be dormant and thus unable to cause infection or sepsis. Instead, they are speculated to have an important role in normal physiology and immunology. (Potgieter et al., 2015)

1.3.2 cbDNA in different blood fractions

Different fractions of blood have been reported to contain varying amounts of cbDNA. Whole blood (WB), red blood cells (RBCs), buffy coat (BC), peripheral blood mononuclear cells (PBMCs) and plasma have been investigated. (Païssé et al., 2016; Yamaguchi et al., 2013) BC consists of lymphocytes (T-, B- and NK-cells), monocytes, granulocytes (neutrophils,

eosinophils, and basophils) and platelets, while PBMCs include lymphocytes and monocytes. Monocytes can differentiate to macrophages and dendritic cells. The BC and PBMC fractions can be isolated from blood samples with traditional centrifugation and density gradient centrifugation correspondingly. The investigation of the role of leukocytes in HBM and cbDNA present in blood during various diseases is one of the current lanes of human microbiota research.

94% of the WB bacterial load is located in the BC, 6% in the RBCs and >1% in plasma (Païssé et al., 2016). Some bacteria, including *Chlamydia* and *Staphylococcus* species, have been found in PBMCs (Yamaguchi et al., 2013). The profile of bacteria in different blood fractions stay constant regardless of sex (Païssé et al., 2016). It is possible that certain bacteria localise to different blood fractions due to preferential living conditions within these niches. Another theory is that the bacteria are being endocytosed and transported within the circulation by these cells. Blood has bacteriostatic and bactericidal components; hence the cells of the blood might provide better environment for the bacteria than plasma. The theorised dormant state of the bacteria might be another reason for their survival in circulation. (Païssé et al., 2016; Goraya et al., 2022; Panaiotov et al., 2018)

Different blood fractions also have different properties for elimination of bacteria. During overnight hold, WB samples spiked with *Staphylococcus* did not show degradation, while *Escherichia* and *Klebsiella* were eliminated from the samples (Taha et al., 2016). Equivalent results have been made from BC with *Enterobacter, Escherichia, Pseudomonas, Serratia* and *Yersinia* being eliminated from WB and BC after 6 or 8 hours hold at room temperature (RT). *Bacillus*, *Propionibacterium* and *Staphylococcus* species showed no degradation. (Mohr et al., 2006) In another study, the degradation of *C. pneumoniae* in PBMCs was investigated. Reportedly *C. pneumoniae* did not grow within the PBMC fraction before elimination, on the other hand some growth was observed in PBMC derived macrophages before elimination (Wolf et al., 2005). This raises the question whether PBMCs and BC might have varied immunodegradation properties due to their different leucocyte composition. Interestingly, antigens released from the *C. pneumoniae* were still located from the samples multiple days after the infection (Wolf et al., 2005). Similar transient bacteraemia events could impact the results when studying the cbDNA composition HBM.

1.3.3 Phyla and genera associated with HBM

The characterised taxonomy of cbDNA varies between blood fractions and is affected by external factors for example age, diet, and stress. The results reported in different studies have some variation, but most of them report the core HBM to consist of mostly *Proteobacteria* phyla (80-87 %), followed by *Actinobacteria* (7-10 %), *Firmicutes* (3-6 %), and *Bacteroidetes* (2.5-3 %). Some bacterial genera seem to be more present in certain fractions of blood, with *Sphingobacteria* observed primarily in BC and plasma, *Flavobacteria* and *Fusobacteria* primarily in RBCs and *Clostridia* in plasma and RBCs. (Panaiotov et al., 2018; Païssé et al., 2016; Whittle et al., 2018) *Staphylococcus, Bacilli* and *Micrococcus* species have also been discovered from healthy WB (Damgaard et al., 2015).

The phylum *Proteobacteria*, or *Pseudomonadota*, consists of gram-negative bacteria. *Proteobacteria* are often credited of gut dysbiosis with them being the most unstable of the gut phyla and their bloom often being a sign of dysbiosis. (Kumar and Kumar, 2022) Notable genera within the *Proteobacteria* phylum are *Paracoccus, Pseudomonas* and *Moraxella*. The family *Enterobacteriaceae* including genera *Escherichia* and *Shigella* is also part of the *Proteobacteria* phylum.

Actinobacteria (also known as *Actinomycetota*) are gram-positive bacteria that produce secondary metabolites, notably anticancer compounds, that could have valuable therapeutic applications (Pongen et al., 2023). Genera of *Actinobacteria* include *Micrococcus, Dietzia, Corynebacterium* and *Cutibacterium* all of which are known to inhabit the skin microbiome, with *Corynebacterium* being a common contaminant of clinical samples (Tauch et al., 2016). Notable *Actinobacteria* inhabiting the gut microbiome include *Bifidobacteria.*

Firmicutes (*Bacilliota*) is a phylum of gram-positive bacteria predominantly present in the gut. Elevated presence of *Firmicutes* in the gut is associated with obesity. The weight gain effect is promoted by certain members of the *Lactobacillus* genus. (Cao et al., 2019) Other genera within the *Firmicutes* phyla include *Staphylococcus* and *Enterococcus*.

Bacteroidetes, or *Bacteroidota*, is a gram-negative bacteria phylum that plays a role in the upkeep of healthy gut microbiota. In addition to the metabolic functions in the gut, *Bacteroidetes* are important for immune function as they are known to affect cytokine production and activate lymphocytes. (Gibiino et al., 2018) *Bacteroidetes* phylum includes genera such as *Sphingobacteria* and *Bacteroides*. *Bacteroides* are known to cause serious

disease when entering the circulation. Especially lipopolysaccharide produced by *Bacteroides fragilis*, has meaningful role in *Bacteroides* pathophysiology (Mancuso et al., 2005).

1.3.4 The source of cbDNA

Since blood is an enclosed niche, the source of cbDNA present in blood, in absence of disease or injury, is up to question. While the source of cbDNA in blood is not confirmed, it is commonly proposed to originate predominately from the gut (Ho et al., 2021). The translocation of bacteria from the gut to circulation seems to be the most popular theory for the source of cbDNA in HBM. While gut microbiome is separated from the circulation by epithelial cells and an immunological barrier, in multiple disease settings, including SLE, leakage of bacteria from the gut to circulation has been reported. This leaky gut and the causational presence of bacteria and their metabolites in the circulation is known to affect the disease state of SLE. (Christovich and Luo, 2022)

Diffusion of bacteria from the gut to circulation in absence of disease or epithelial barrier disruption is a more novel line of thought. It has been shown that bacteria can enter circulation through intestinal mucus-secreting goblet cells, dendritic cells, and mucosal lymph cells without affecting epithelial barriers of tight junctions (Niess et al., 2005; Wiest et al., 2014; Castillo et al., 2019). In healthy state however, immune barriers and the reticuloendothelial system clear most of the crossing bacteria and their metabolites, keeping most of the bacterial load out of the circulation. It is suggested that the immune cells and intestinal cells filter specific bacteria preventing their entrance to the circulation, while letting other bacteria through (Païssé et al., 2016). This could explain why the reported composition of HBM is not identical to the gut microbiota but shows more resemblance to the microbiota of oral cavities and skin. On the other hand, the composition of cbDNA is also seen as evidence to the notion that gut microbiome is not the main source of cbDNA in HBM. Some studies suggest that translocation from skin and oral sources are the more prevalent routes for cbDNA acquisition. (Khan et al., 2022)

In addition to gut, other bacterial niches such as oral cavity, skin, lungs and birth canal are also proposed to contribute to the HBM (Païssé et al., 2016; Khan et al., 2022). Oral microbiota can enter the circulation when tight junctions of oral mucosa are compromised. Similarly, skin microbiota can enter when skin barrier is damaged. Escape of lung microbiota into the circulation is also proposed. In addition, there is evidence of bacterial translocation between mother and foetus during pregnancy, while the placenta has been proven to be

without microbiome in absence of infection. (Iwai, 2009; Cogen et al., 2008; Goraya et al., 2022; De Goffau et al., 2019) The full extent to which these routes of bacterial translocation might contribute to HBM under normal physiological setting is still under debate. Outside normal function, bacteria can enter the circulation when it is exposed to external sources of microbes, for example, in cases of surgery, physical trauma or insect and animal bites (Goraya et al., 2022).

While bacterial translocation is the most popular theory for cbDNA presence in the blood, bacteria derived extracellular vesicles (EVs) are suggested to be another possible route for cbDNA (Ricci et al., 2020). Bacteria are known to release EVs containing multiple compounds including cytoplasmic components, toxins and bacterial DNA (Domingues and Nielsen, 2017). These EVs have been located in the circulation in disease conditions such as cancers and mental disorders (Kim et al., 2020; Lee et al., 2020; Rhee et al., 2020). In the end, whether the cbDNA present in the circulation is caused by EVs, transient bacteraemia events or live possibly dormant bacteria is up to debate (Castillo et al., 2019; Wolf et al., 2005; Ricci et al., 2020). More extensive research is needed to determine the sources of cbDNA.

1.4 Significance of cbDNA in disease settings

After the discovery of cbDNA in healthy blood, studies of the relationships between altering composition of cbDNA and different diseases compared to the cbDNA composition of healthy volunteers have been conducted. Substantial differences were found between the cbDNA composition in healthy and diseased subjects in many disease settings, with equal sample processing (Goraya et al., 2022).

The profile of cbDNA shows significant variation between healthy and diabetic subjects. In diabetic patients, bacteria such as *Faecalibacterium, Akkermansia*, and *Bifidobacterium* are less prevalent in leucocytes than in HBM. On the other hand, bacteria including *Escherichia coli, Bacteroides fragilis*, and *Lactobacillus* are over represented in both diabetic and pre-diabetic populations. (Ghaemi et al., 2022) In addition to bacterial profile variation, overall load of bacteria in blood seems to indicate type 2 diabetes, with higher bacterial loads being connected to increased blood glucose, insulin resistance and systemic inflammation (Amar et al., 2011).

Association between cbDNA composition and liver cirrhosis and pancreatis has been discovered. Liver cirrhosis patient showed increased microbial diversity in blood compared to healthy subjects. *Enterobacteriaceae* were significantly more abundant in cirrhosis patients,

while *Akkermansia, Rikenellaceae*, and *Erysipelotrichales* were more abundant in controls (Kajihara et al., 2019). In liver fibrosis, the overall amount of cbDNA wan increased (Khan et al., 2022). In pancreatitis on the other hand, an increase in *Firmicutes* and *Bacteroidetes* and a reduction in *Actinobacteria* was reported (Li et al., 2018).

Dysbiosis in the HBM can be an underlying cause of cardiovascular diseases (Goraya et al., 2022). Total reduction in blood bacterial load with increased amount of *Proteobacteria* was discovered in patients with acute cardiovascular event (Amar et al., 2013). Especially, increased *Actinobacteria* to *Proteobacteria* ratio seems to characterise cardiovascular disease (Dinakaran et al., 2014).

In addition to diabetes, hepatic and pancreatic conditions and cardiovascular diseases, many other diseases and conditions have been shown to have circulatory bacterial component. Common variable immunodeficiency patients with inflammatory complications had greater bacterial load in the blood compared to healthy subjects (Ho et al., 2021). Blood bacteria composition in rheumatoid arthritis, ankylosing spondylitis and psoriatic arthritis patients differed from HBM composition (Hammad et al., 2019). Bacterial lipopolysaccharide secretions were found in abundance from the blood of Parkinson's and Alzheimer's disease patients indicating bacterial presence in blood (Pretorius et al., 2016). cbDNA composition has been studied also in dermatological diseases, such as psoriasis, hidradenitis suppurativa and rosacea (Khan et al., 2022). In colorectal cancer, 28 bacterial species were identified to have potential to distinguish between healthy subjects and cancer patients. Similar correlations have been discovered in cases of lung cancer and melanoma (Xiao et al., 2021; Ricci et al., 2020). Bacterial translocation has also been connected to human immunodeficiency virus infection and inflammatory bowel disease (Ho et al., 2021). The role of cbDNA still remains largely unexplored in many diseases including many neurocognitive disorders, SLE and psychiatric disorders (Goraya et al., 2022).

The notion of cbDNA being used as a biomarker for disease onset or progression has been made in multiple diseases including diabetes, cardiovascular disease and heart failure, liver cirrhosis and fibrosis, colorectal cancer, atherosclerosis and pancreatic carcinoma (Goraya et al., 2022). Composition of cbDNA is known to have function on several major systems in the body including immune system surveillance and development, chronic inflammation, blood coagulation and blood–brain barrier permeability (Kell and Pretorius, 2018; Nikkari et al., 2001). In the future cbDNA could prove to be invaluable tool for diagnostics and prognostics.

Further research on the topic might also shed light on the pathophysiological changes caused by HMB dysbiosis and create new therapeutic targets for drug development.

1.5 16S ribosomal RNA PCR

Traditionally the identification of bacteria from blood in cases of infection is done by blood cultures. However, in absence of infection the amounts of bacteria precent in blood are miniscule and the bacteria precent can be in dormant state making it impossible to detect using the traditional culture methods. Also, due to limitations on the culturing techniques some bacteria strains cannot be cultivated from blood even if precent in meaningful amounts. To combat these issues, methos independent from cultivation have been developed including PCR methods and targeted metagenomics. (Nikkari et al., 2001; Païssé et al., 2016; Panaiotov et al., 2018; Whittle et al., 2018) In addition to the amplification and sequencing of the 16S ribosomal RNA (rRNA) gene discussed here, other methods for bacteria identification from blood samples include shotgun metagenome sequencing, fluorescent in situ hybridisation and quantitative PCR.

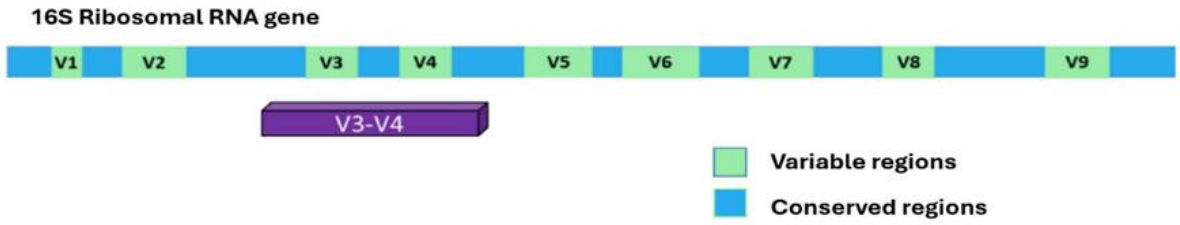


Figure 1. The structure of the bacterial 16S rRNA gene with the nine variable regions and surrounding conserved regions. Adapted from (Ricci et al., 2020)

Amplification of the 16S rRNA gene from the genome of bacteria with PCR, followed by next generation sequencing (NGS), is one of the main methods for identifying bacterial taxonomy from tissue samples. The operon expressing the 5S, 16S and 23S genes coding rRNA is highly conserved amongst different species of bacteria. Each bacterium has 1 to 15 copies of the 16S rRNA gene in their genome (Figure 1), while the gene itself is approximately 1500 bp long and contains nine bacteria-specific hypervariable regions (V1-V9) (Johnson et al., 2019). These regions are surrounded by highly conserved and well characterised areas of the genome. By amplifying and sequencing these hypervariable regions with the surrounding conserved areas, in this case V3-V4, and comparing them to known bacterial genomes, the bacteria, from which the DNA originated form, can be taxonomically identified. (Ricci et al., 2020)

The primers used in this study (Figure 2) target the V3-V4 region of the 16S rRNA gene. The primers were chosen based on assessment of 16S primers by Klindworth et al. in 2013, where the used primes were noted as the most promising (Klindworth et al., 2013). The primers consist of the locus specific sequence and overhang adapter sequences to which sequencing indices and sequence adapters are attached to. These primers yield an amplicon of approximately 550 bp including the amplified region (~460 bp), the locus specific primers and overhang adapters. In addition to V3-V4, other regions of the 16S rRNA gene, including V4 and V4-V5, can be targeted and, consequently, the best region to target is up to debate (Kozich et al., 2013).

16S Amplicon PCR Forward Primer = 5'
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACG
GGNGGCWGCAG

16S Amplicon PCR Reverse Primer = 5'
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTA
CHVGGGTATCTAATCC

Figure 2. The Amplicon forward and reverse primers with the overhang adapter sequences (red) and locus-specific sequence (black).

When analysing the cbDNA composition of blood with metagenomic tools such as 16S rRNA PCR and NGS, some limitations do exist. The amount of cbDNA in the blood is low and thus the extraction process of the bacterial DNA from samples can be difficult due to abundance of eukaryotic genomic products. Blood also hosts a plenty of PCR inhibitors further complicating the amplification process. (Païssé et al., 2016) Contamination of the samples is one of the most important limitations of the method. Due to the low amount of cbDNA present in blood, the samples are extremely prone to contamination from laboratory equipment, reagents, skin, and environmental bacteria. Most common laboratory contaminants include DNA from bacteria of the *Bacillus, Flavobacteria, Fusobacteria, Propionibacterium*, and *Serratia* genera (Glassing et al., 2016). Extensive care is needed during sample collection and processing to always minimize the possibility of contamination. In addition to contamination from live bacteria, bacterial DNA from external sources can also contaminate the samples. To combat the issue of contamination, extensive experimental controls are required although not routinely used (Castillo et al., 2019; Goraya et al., 2022; Whittle et al., 2018). Even with controls it is important to remain cautious and realistic about the possibility of contamination while interpretating 16S sequencing results.

In addition to sample preparation, the accuracy of assigning the taxonomy for the sequencing data can also be an issue. In species level, and in cases of some genera, 16S rRNA sequencing has low power to reliably differentiate between operational taxonomic units (OTUs) (Janda and Abbott, 2007). This is partly due to the shortness (only few hundred base pairs) of the

sequences produced by NGS methods such as Illumina MiSeq. The cbDNA found by sequencing could also be derived from free bacterial DNA from immune degradation, not just DNA precent within the studied blood fractions. (Païssé et al., 2016) In addition, separation of live bacteria from bacterial DNA is not possible with this method (Khan et al., 2022). Other aspects that could alternate results include background contamination and temporary transient bacteraemia events (Goraya et al., 2022). While NGS of the amplified 16S rRNA gene is an extremely useful method that is used as the gold standard for detection of bacterial taxonomy, it is not without fault and understanding the limitations of the method is essential for reliable data collection and analysis.

1.6 Aims and significance

The role cbDNA in health and disease has received increasing attention since its discovery. Multiple disease settings have been proven to be associated with the dysbiosis of blood microbiome and the potential role cbDNA in disease pathology and progression offers and interesting point of view for biomarker and drug development. While the role of cbDNA in health and disease remain elusive, new discoveries in the field are made with increasing pace. New and more reliable methods for detecting bacterial presence in different niches around the body are constantly developed. The ability to detect the presence of these organisms in decreasing abundances from blood is critical for developing understanding of the role of HBM and cbDNA in health and in the pathology of various diseases. With the current evidence, blood can no longer be considered a sterile environment, but a bacterial niche hosting diverse taxa of bacteria.

This study is a precursory study for the role of cbDNA and HBM dysbiosis in SLE, where possible connections between cbDNA composition and disease prognosis and onset remain unexplored. While SLE itself is not yet investigated, other rheumatic and autoimmune mediated conditions, including rheumatic arthritis, have shown association between cbDNA composition and disease development (Hammad et al., 2019). This strengthens our hypothesis of cbDNA composition alternation in SLE patients when compared to HBM.

The aim of this study is to establish protocol for cbDNA extraction from blood samples followed by 16S rRNA PCR and NGS using the Illumina MiSeq. With these methods we aim to taxonomically characterise the cbDNA composition of healthy volunteers. In the future these results could be used to compare the cbDNA composition of SLE patients to the HBM. In addition to characterising the profile of HBM, we aim to discover whether there is

difference to the localisation of bacteria from different genera between blood fractions. As a secondary objective we are interested in finding out how stabile HBM is when compared between different individuals and different time points. To accomplish these goals, we used 16S rRNA PCR and Illumina MiSeq next generation sequencing to characterise the taxonomic profile of HBM in WB, BC and PBMC samples.

2 Results

2.1 Sample preparation

We conducted a series of experiments with the aim of characterising the cbDNA composition in healthy blood. Blood samples from five different subjects were collected and the DNA form the three blood fractions (BC, WB and PBMCs) was isolated. In addition, samples from one of the donors were collected at two additional time points. Some of the samples were spiked with known amounts of bacteria to study the localisation of the different bacteria between the fractions. After the isolated DNA concentrations were measured, we performed V3-V4 16S rRNA PCR followed by Illumina MiSeq sequencing and data analysis.

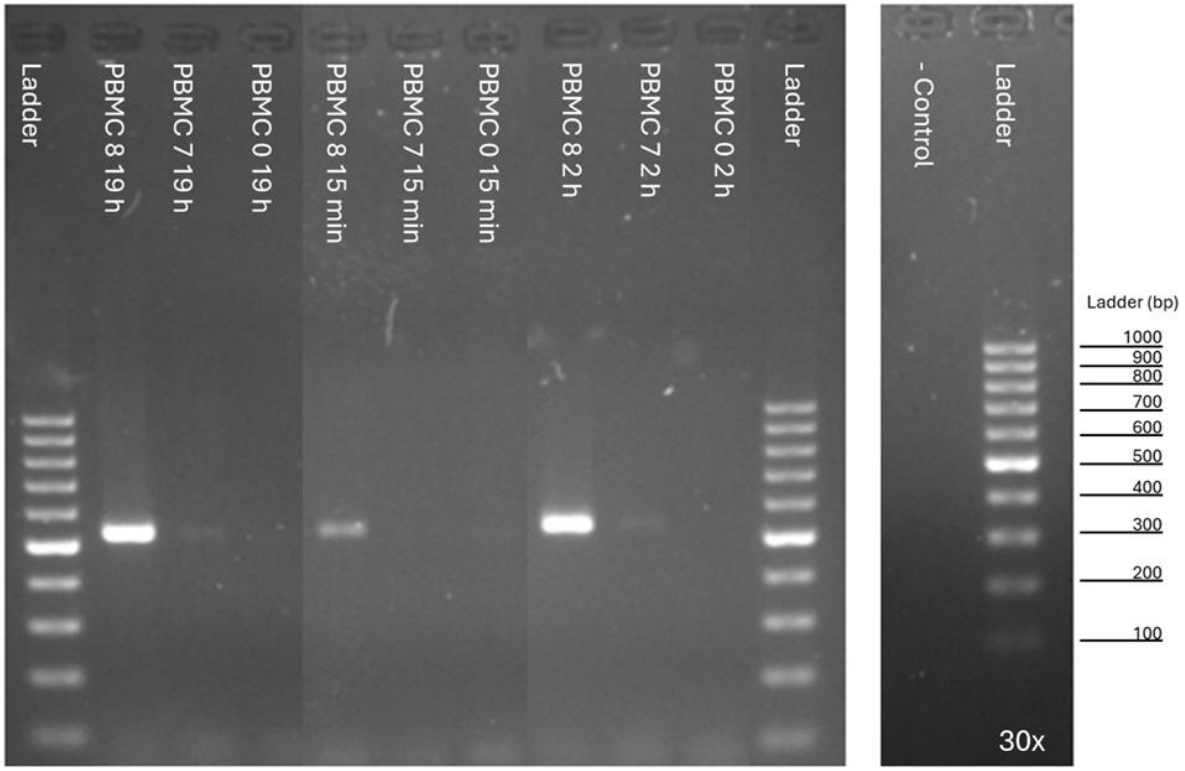


Figure 3. PBMC samples spiked with *E. coli* incubated for 19 h, 15 min and 2 h. Spiked samples (PBMC 8 and PBMC 7, c = 1.5 x 10^8/ml and 1.5 x 10^7/ml) and unspiked samples (PBMC 0) incubated for 19 h, 15 minutes or 2 h. Samples incubated for 15 minutes showed less signal at the 550 bp range, with only a weak band in the higher concentration, than samples incubated for 2 or 19 h, both of which showed strong bands in the higher concentrations and weak bands in the lower one. The unspiked samples and negative control showed no signal.

Before the actual experiment, we conducted trials determining the optimal incubation time and bacteria cocktail concentrations for the spiked samples, as well as the optimal the number of cycles for the amplicon PCR step. The result image form the incubation time test (Figure 3) showed the PBMC samples spiked with *E. coli* (PBMC 8 and PBMC 7, c = 1.5 x 10^8/ml and 1.5 x 10^7/ml) and unspiked samples (PBMC 0) incubated for 19 h, 15 minutes or 2 h. Samples

incubated for 15 minutes showed less signal, with only a weak band at approximately 550 bp in the higher concentration, than samples incubated for 2 or 19 h, both of which showed strong bands in the higher concentrations and weak bands in the lower one. The unspiked samples and negative control showed no signal. Due to high amount of erythrocyte degradation during the 19 h incubation, 2 h was chosen as the incubation time for the final spiked samples.

We also evaluated different concentrations of bacterial cocktail (*E. coli* and *E. faecalis*) with the results shown in figure 4. The PBMC samples were spiked with the bacterial cocktail from concentration 1 x 10^7/ml to 1 x 10^2/ml (PBMC 7 to PBMC 2). The results showed strong signal in PBMC 7 and 6 samples and dim line in PBMC 5 at the 550 bp mark. Positive control showed strong signal and negative control stayed empty. Since PBMC 6 was the lowest concentration to show strong signal, we chose that as the strongest concentration of the bacterial cocktail for the final spiked samples.

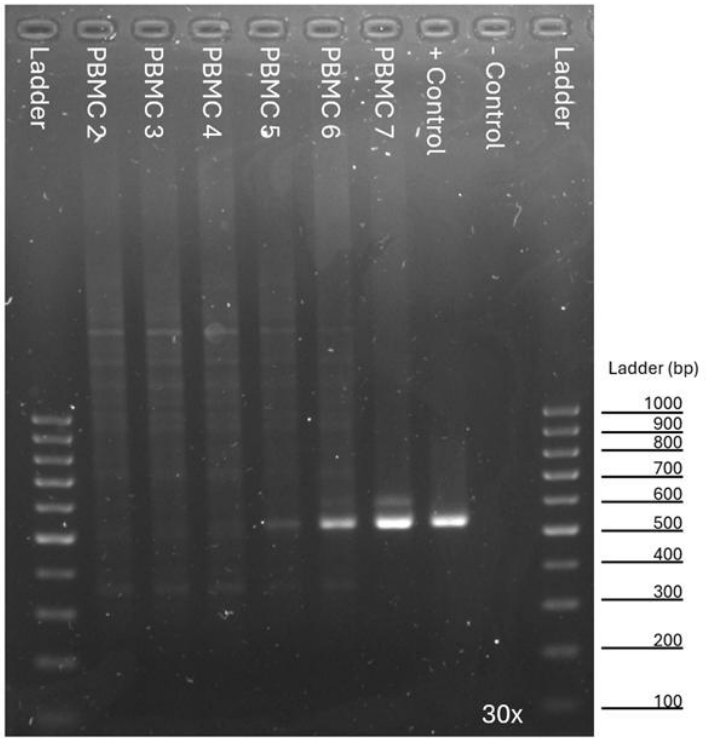


Figure 4. PBMC samples spiked with bacterial cocktail from concentration 1 x 10^7/ml to 1 x 10^2/ml (PBMC 7 to PBMC 2). The results showed strong lines in PBMC 7 and 6 samples and dim line in PBMC 5 at the 550 bp mark. Positive control showed strong signal and negative control stayed empty.

In addition to bacteria concentration and incubation time tests, we also experimented on the amount of PCR cycles. The same samples from the previous bacterial concentration test (with 30 cycles) were run on the same PCR program for 35 and 40 cycles (Figure 5). Although the intensity of the bands remained similar in all three cycle amounts, the negative control also stayed clear indicating that no unspecific amplification occurred even at the higher cycle counts. Due to the low expected amount of the cbDNA present in healthy blood, we chose to continue with 40 cycles for further experiments.

The final samples were prepared, spiked and the DNA was isolated from the extracted blood fractions. Before 16S rRNA PCR, the concentrations of the isolated DNA were measured

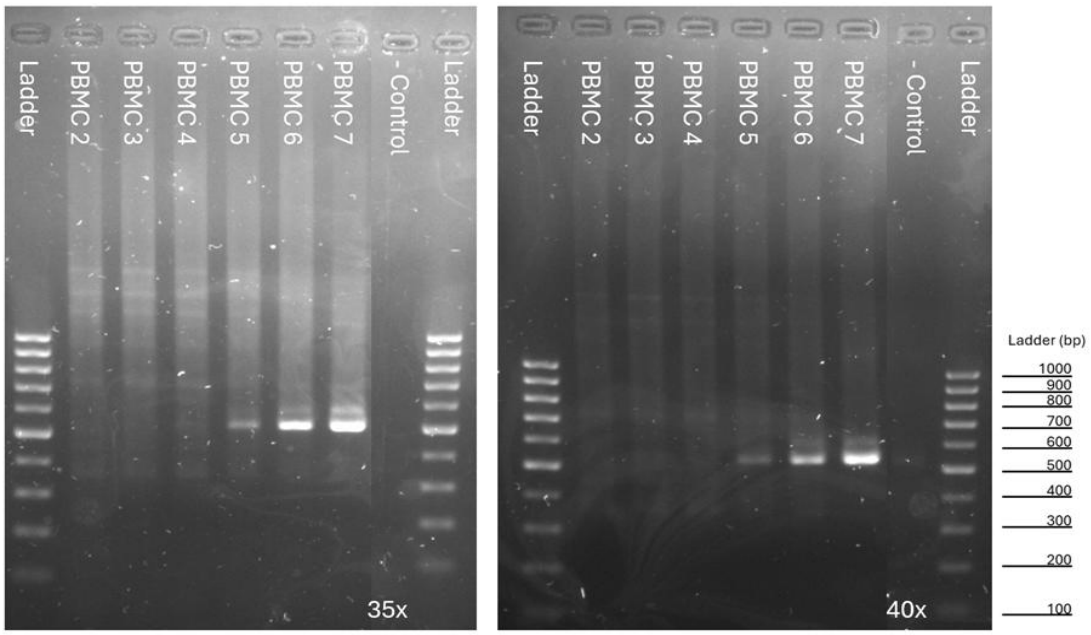


Figure 5. PBMC samples run for 35 and 40 PCR cycles. The samples from the previous bacterial concentration experiment were run on the PCR program for 35 and 40 cycles. Although the intensity of the bands remained similar in all cycle amounts, the negative control stayed clear indicating that no unspecific amplification occurred even at the higher cycle counts.

Table 1. Measured DNA concentrations of the final samples before and after 16S rRNA PCR. Before PCR, all samples showed measurable DNA concentrations, however, six samples (BC 0 E2, PBMC 0 E1 and PBMC 0 A/B/C/D) had concentrations below 6.25 ng/µl. After PCR, all samples had concentrations well over 1.7 ng/µl (4 nM).

	Concentration (ng/µl)	
Sample	Pre-PCR	Post-PCR
PBMC 0 E2	11,2	135
PBMC 2 E	11	35,15
PBMC 3 E	11,4	66,55
PBMC 4 E	11,2	104,6
PBMC 5 E	11,6	58,45
PBMC 6 E	10,6	69
BC 0 E3	12	148
BC 2 E	12	115
BC 3 E	11,8	43,45
BC 4 E	11,6	87,05
BC 5 E	11	74,6
BC 6 E	11,4	83,45
WB 0 E2	24,2	30,85
WB 2 E	20,6	45,05
WB 3 E	12	107
WB 4 E	27,8	81,9
WB 5 E	11,8	34,2
WB 6 E	11,2	134
BC 0 E1	42	91,85
BC 0 E2	0,85	58,7
PBMC 0 E1	0,313	80
WB 0 E1	7,64	69
BC 0 A	10,8	71,4
PBMC 0 A	0,167	44,65
BC 0 B	27,5	48,95
PBMC 0 B	1,24	62,25
BC 0 C	39,3	62,65
PBMC 0 C	0,34	55,35
BC 0 D	16,7	122
PBMC 0 D	0,145	43,35

(Table 1). All samples were diluted to 6.25 ng/µl for the PCR, therefore ideally the DNA concentrations should be above the 6.25 ng/µl limit. All samples showed measurable DNA concentrations, however, six samples (BC 0 E2, PBMC 0 E1 and PBMC 0 A/B/C/D) had concentrations below 6.25 ng/µl. Because the concentrations of these samples were still measurable, we continued to the 16S rRNA PCR with all 30 samples. We also prepared a control sample of possible contamination during DNA extraction by running the extraction protocol with PBS. This sample showed no presence of DNA during concentration measurements with fluorometer, nor any signal on agarose gel electrophoresis (results not shown).

The amplicon PCR (40 cycles) was performed, and the results were run to 1.5 % TAE gel (Figure 6). The samples are named according to the blood fraction (PBMC/BC/WB), concentration of bacteria cocktail (0-6), the donor (A-E) and finally the replicate (1-3). In the spiked samples

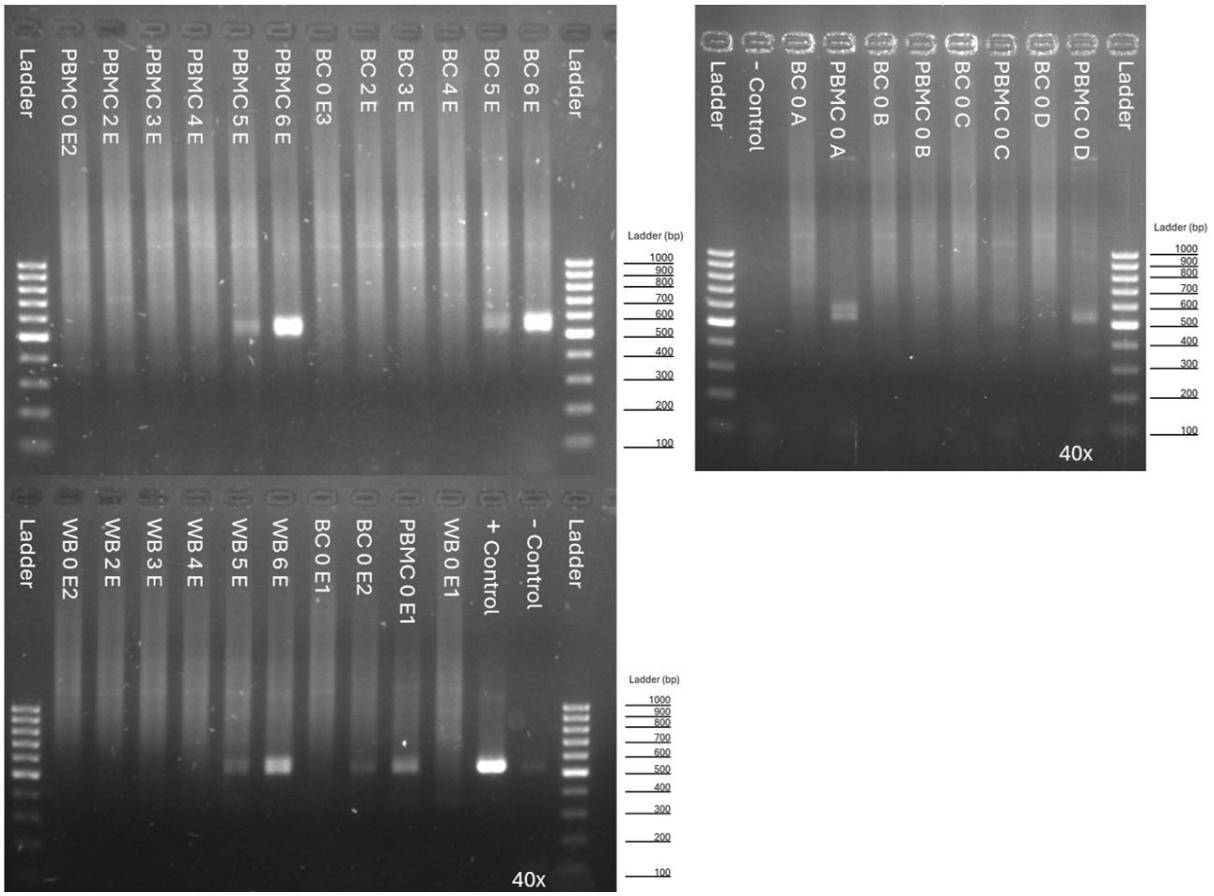


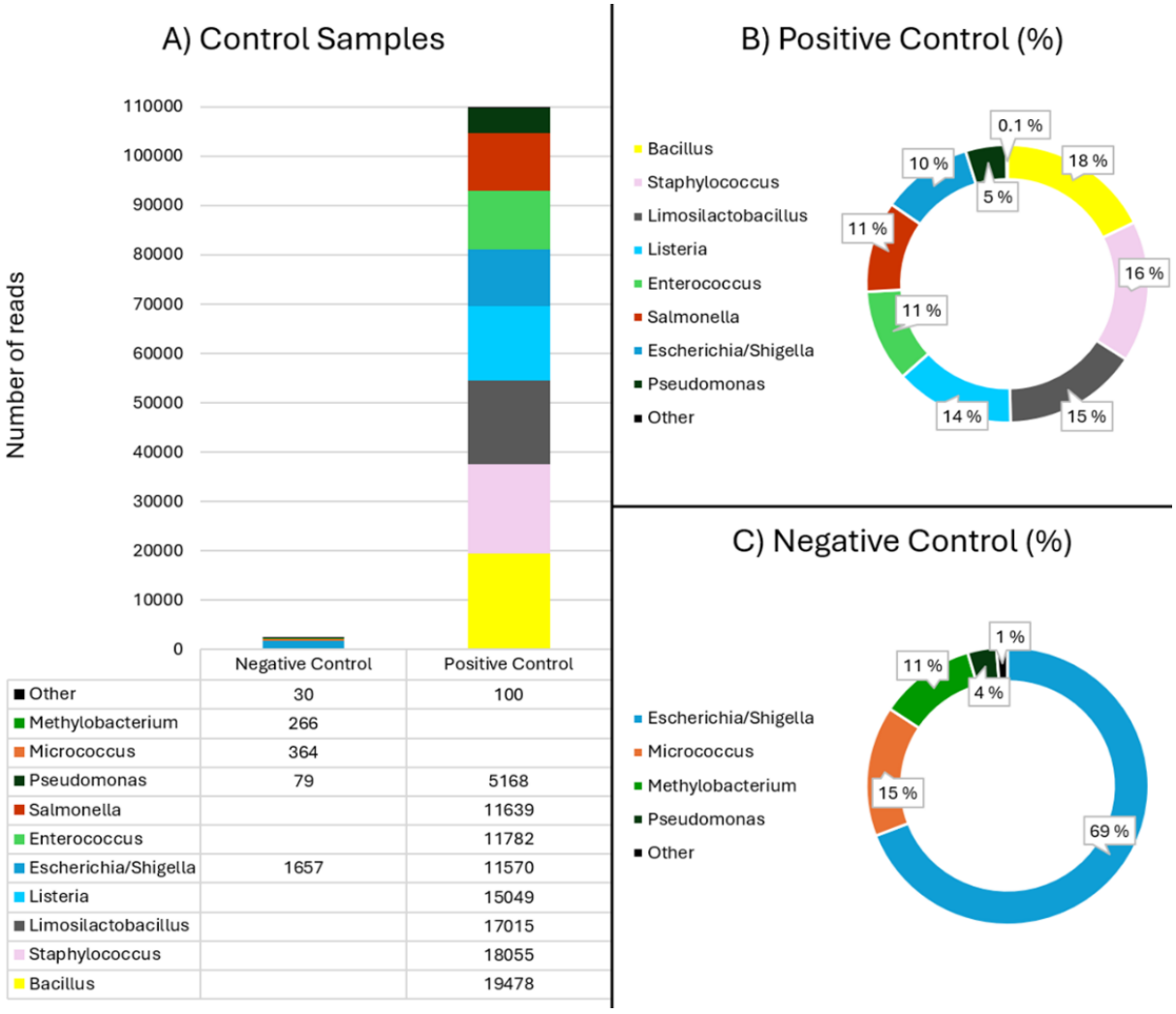
Figure 6. The amplicon PCR (40 cycles) results. The samples are named according to the blood fraction (PBMC/BC/WB), concentration of bacteria cocktail (0-6), the donor (A-E) and finally the replicate (1/2). In the spiked samples PBMC/BC/WB 2-6 E (*S. aureus, E. faecalis, E. coli, Ps. aeruginosa, B. fragilis* and *Cl. perfringens*, c = 2 x 10^6/ml to 2 x 10^2/ml) all samples with the highest bacteria cocktail concentrations showed strong signal at the 550 bp marker, while the second highest concentrations showed weaker signal. From the unspiked samples BC 0 E2, PBMC 0 E1, PBMC 0 A and PBMC 0 D showed signal at 550 bp. Negative controls remained empty while the positive control showed strong signal.

PBMC/BC/WB 2-6 E (*S. aureus, E. faecalis, E. coli, Ps. aeruginosa, B. fragilis* and *Cl. perfringens*, c = 2 x 10^6/ml to 2 x 10^2/ml) all samples with the highest bacteria cocktail concentrations showed strong signal at the 550 bp marker, while the second highest concentrations showed weaker signal. From the unspiked samples BC 0 E2, PBMC 0 E1, PBMC 0 A and PBMC 0 D showed signal at 550 bp. Negative controls remained empty while the positive control showed strong signal.

Index PCR was performed after the amplicon PCR, followed by post-PCR DNA concentration measurements (Table 1). All samples showed the amplification of bacterial DNA, with concentrations well over the required 1.7 ng/µl (or 4 nM) for NGS. The samples were diluted to 4 nM and pooled with PhiX control before sequencing with Illumina MiSeq.

2.2 Control samples

Overall, the sequencing of the samples produced 870606 reads including the positive and negative controls. After data cleanup and processing we were left with the total of 403947 reads. The negative control sample had a total of 2369 reads most of which (69 %) belonging to the genus *Escherichia/Shigella* (Figure 7 A and C). The second common genus was *Micrococcus* (15 %) followed by *Methylobacterium* (11 %) and *Pseudomonas* (4 %). The commercial positive control (Figure 7 A and B) consisting of ten different bacteria genera (*Listeria* 12%, *Pseudomonas* 12%, *Bacillus* 12%, *Escherichia* 12%, *Salmonella* 12%, *Limosilactobacillus* 12%, *Enterococcus* 12%, *Staphylococcus* 12%, *Saccharomyces* 2%, and *Cryptococcus* 2%) was used to ensure the appropriate function of the sequencing protocol. It yielded 109856 reads total and, as is shown in figure 7 B, the composition of the sequenced positive control was 18 % of *Bacillus*, 16 % of *Staphylococcus*, 15 % of *Limosilactobacillus*, 14 % of *Listeria*, 11 % of *Enterococcus* and *Salmonella*, 10 % of *Escherichia/Shigella* and 5% of *Pseudomonas*.

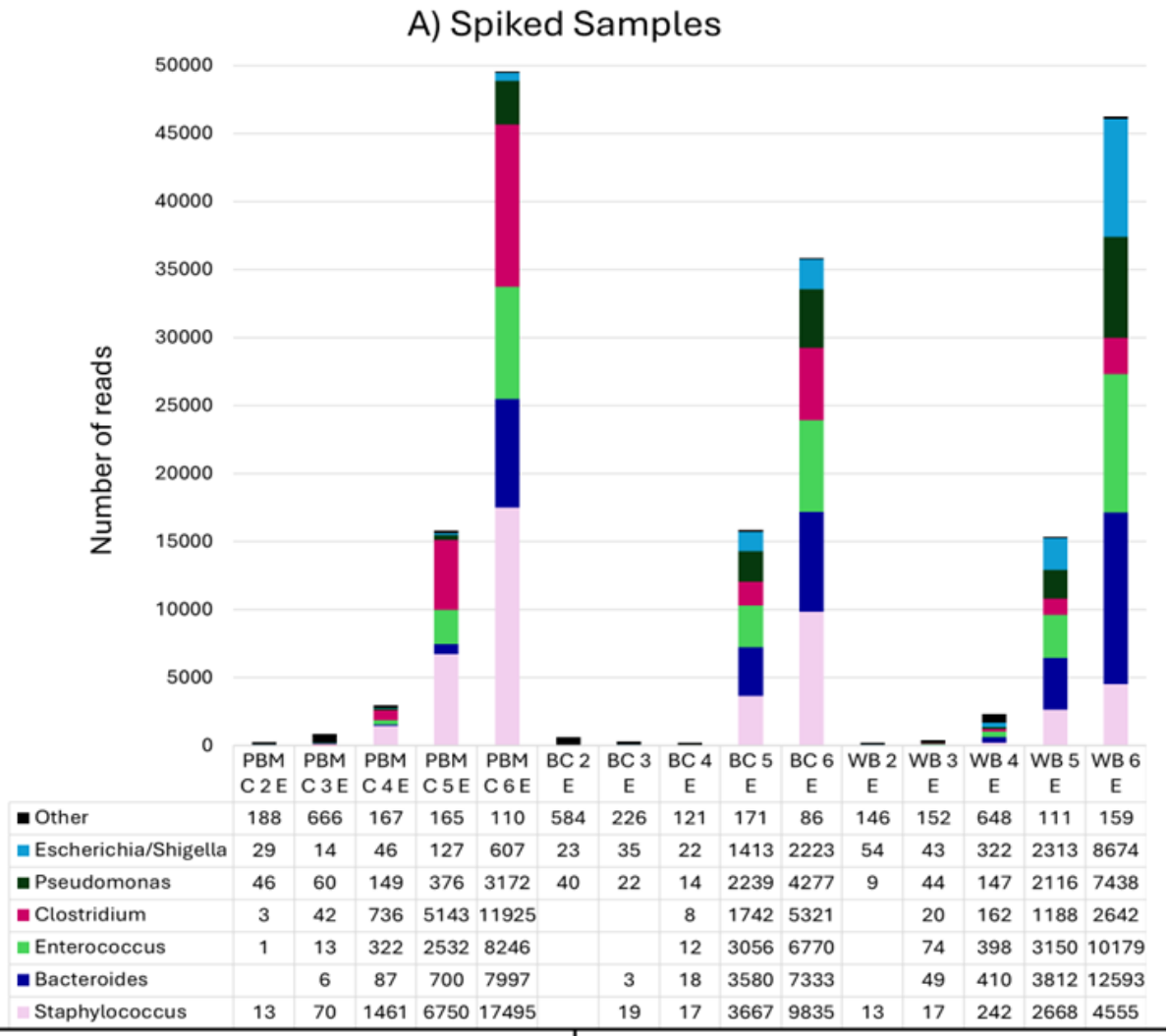


	Negative Control	Positive Control
Other	30	100
Methylobacterium	266	
Micrococcus	364	
Pseudomonas	79	5168
Salmonella		11639
Enterococcus		11782
Escherichia/Shigella	1657	11570
Listeria		15049
Limosilactobacillus		17015
Staphylococcus		18055
Bacillus		19478

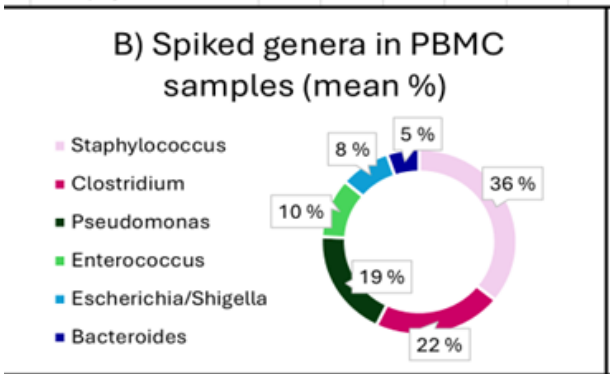
Figure 7. The composition of negative and positive control. A) The read counts of the controls. Negative control had a total of 2369 reads and positive control had 109856 reads. B) The genera composition of the positive control as a percentage of the total read count. *Bacillus* was the most common genus with 18 % of total reads followed by *Staphylococcus* (16 %), *Limosilactobacillus* (15 %), *Listeria* (14 %), *Enterococcus* and *Salmonella* (11 %), *Escherichia/Shigella* (10 %) and *Pseudomonas* (5 %). The remaining 0.1 % consisted of other genera with a presence of less than 1 % each. C) The genera composition of the negative control as a percentage of the total read count. *Escherichia/Shigella* was the most common genus with 69 % of total reads followed by *Micrococcus* (15 %), *Methylobacterium* (11 %) and *Pseudomonas* (4 %). The remaining 1 % consisted of other genera with a presence of less than 1 % each.

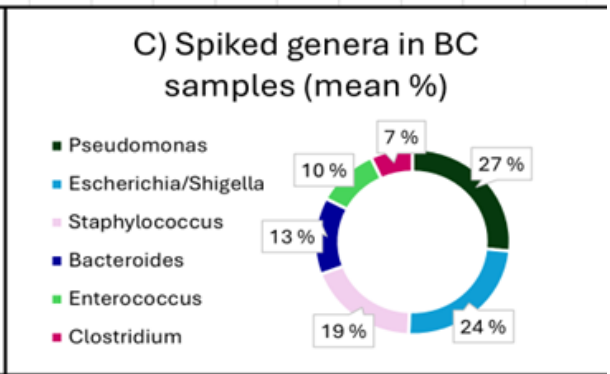
2.3 Spiked samples

The number of reads for spiked samples (Figure 8 A) varied from 212 reads (BC 4 E) to 49552 (PBMC 6 E) with the mean of 12459 reads per sample. The samples were spiked with a cocktail of six genera (*Escherichia/Shigella, Pseudomonas, Clostridium, Enterococcus, Bacteroides* and *Staphylococcus*) in a dilution series (c = 2 x 10^6, 2 x 10^5, 2 x 10^4, 2 x 10^3 and 2 x 10^2). In figure 8 B, C and D, the presence of each spiked genera in spiked PBMC, BC and WB samples are presented as the mean of the percentages in each sample group. Other genera, excluding the six spiked genera, were not included in figures 8 B, C and D, but the reads are shown in figure 8 A. Out of the six spiked genera *Staphylococcus* was the most common in PBMC samples (36 %) followed by *Clostridium* (22 %), *Pseudomonas* (19 %), *Enterococcus* (10 %), *Escherichia/Shigella* (8%) and *Bacteroides* (5%). *Pseudomonas*, on the other hand, *Pseudomonas* was the most common genus in BC samples (27 %) followed by *Escherichia/Shigella* (24 %), *Staphylococcus* (19 %), *Bacteroides* (13 %), *Enterococcus* (10%) and *Clostridium* (7%). In WB samples, *Escherichia/Shigella* was the most common genus (29 %) followed by *Bacteroides* and *Enterococcus* (19 %), *Pseudomonas* (14 %), *Staphylococcus* (13%) and *Clostridium* (6%).



	PBM C 2 E	PBM C 3 E	PBM C 4 E	PBM C 5 E	PBM C 6 E	BC 2 E	BC 3 E	BC 4 E	BC 5 E	BC 6 E	WB 2 E	WB 3 E	WB 4 E	WB 5 E	WB 6 E
Other	188	666	167	165	110	584	226	121	171	86	146	152	648	111	159
Escherichia/Shigella	29	14	46	127	607	23	35	22	1413	2223	54	43	322	2313	8674
Pseudomonas	46	60	149	376	3172	40	22	14	2239	4277	9	44	147	2116	7438
Clostridium	3	42	736	5143	11925			8	1742	5321		20	162	1188	2642
Enterococcus	1	13	322	2532	8246			12	3056	6770		74	398	3150	10179
Bacteroides		6	87	700	7997		3	18	3580	7333		49	410	3812	12593
Staphylococcus	13	70	1461	6750	17495		19	17	3667	9835	13	17	242	2668	4555





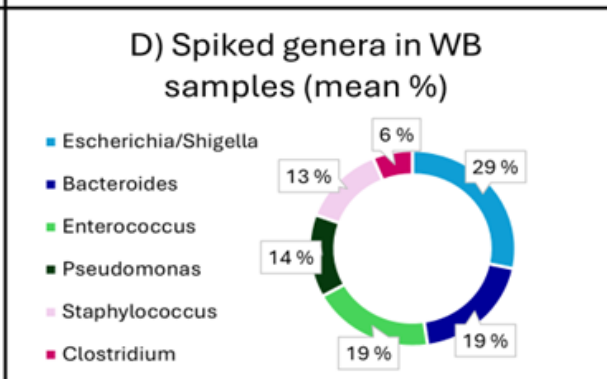


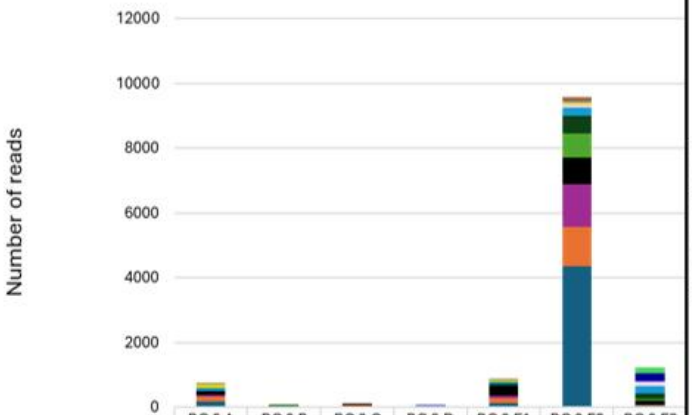
Figure 8. The composition of the spiked PBMC, BC and WB samples. The samples were spiked with a cocktail of six bacteria in a dilution series (c = 2 x 10^6, 2 x 10^5, 2 x 10^4, 2 x 10^3 and 2 x 10^2). Other genera, excluding the six spiked genera, were not included in graphs B, C or D. A) the read counts of the spiked samples displaying the reads of the spiked genera (*Escherichia/Shigella*, *Pseudomonas, Clostridium, Enterococcus, Bacteroides* and *Staphylococcus*) and others (Other). The number of reads varied from 212 reads (BC 4 E) to 49552 (PBMC 6 E) with the mean of 12459 reads per sample. B) The presence of each spiked genera in spiked PBMC samples as the mean of the percentages in each sample. Out of the six spiked genera *Staphylococcus* was the most common in PBMC samples (36 %) followed by *Clostridium* (22 %), *Pseudomonas* (19 %), *Enterococcus* (10 %), *Escherichia/Shigella* (8%) and *Bacteroides* (5%). C) The presence of each spiked genera in spiked BC samples as the mean of the percentages in each sample. Out of the six spiked genera *Pseudomonas* was the most common in BC samples (27 %) followed by *Escherichia/Shigella* (24 %), *Staphylococcus* (19 %), *Bacteroides* (13 %), *Enterococcus* (10%) and *Clostridium* (7%). D) The presence of each spiked genera in spiked WB samples as the mean of the percentages in each sample. Out of the six spiked genera *Escherichia/Shigella* was the most common in WB samples (29 %) followed by *Bacteroides* and *Enterococcus* (19 %), *Pseudomonas* (14 %), *Staphylococcus* (13%) and *Clostridium* (6%).

2.4 Unspiked samples

The unspiked samples were prepared to study the cbDNA composition of HBM. Figure 9 shows the read counts of the unspiked samples according to the blood fraction. In PBMC samples the counts varied from 295 (PBMC 0 E2) to 25013 (PBMC 0 E1) with a mean of 12637 reads per sample. The mean of the read count in BC samples was smaller, varying from 98 (BC 0 C) to 9566 (BC 0 E2) with a mean of 1820 reads per sample. Finally, the WB samples had split number of reads ranging from 102 (WB 0 E2) to 908 (WB 0 E1) with a mean of 505 reads per sample. Sample PBMC 0 A had an unusually large amount of *Pseudomonas* reads (15232 reads) that were removed from the results as they are most likely due to a contamination of the sample.

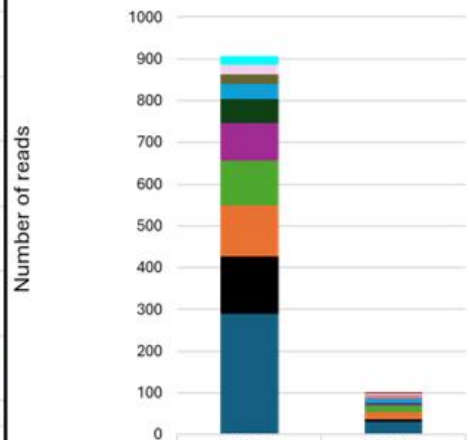
The genera and phyla composition of the unspiked samples altered between blood fractions. *Corynebacterium* and *Micrococcus* were the two most common genera in each fraction (Figure 10 A, C and E). The six most common genera in the PBMC samples were *Corynebacterium* (29 %), *Micrococcus* (17 %), other genera with less than 1 % presence in the sample group (13 %), *Escherichia/Shigella* (10 %), *Dietzia* (8 %) and unclassified *Enterobacteriaceae* (4 %). In BC samples, the six most common genera were *Corynebacterium* (23 %), *Micrococcus* (15 %), other genera (14 %), *Escherichia/Shigella* (13 %), *Dietzia* (8 %) and *Caldibacillus* (4 %). Finally, in the unspiked WB samples, the six most common genera were *Corynebacterium* (31 %), *Micrococcus* (15 %), unclassified *Enterobacteriaceae* (14 %), other (11 %), *Escherichia/Shigella* (7 %) and *Dietzia* (6 %).

A) Unspiked BC samples



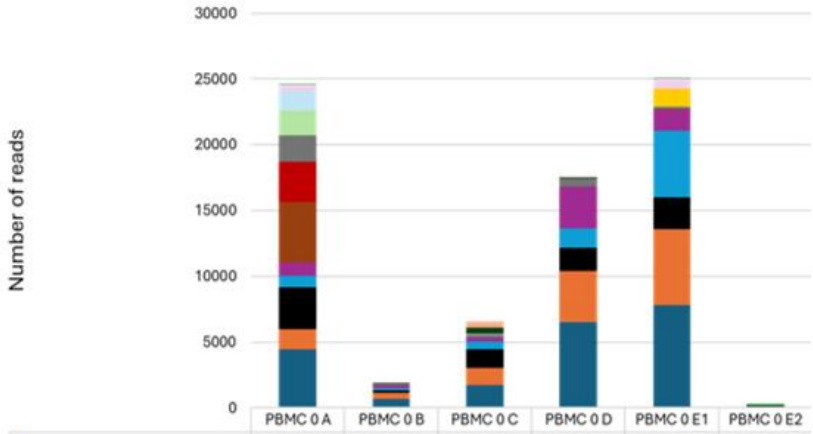
	BC 0 A	BC 0 B	BC 0 C	BC 0 D	BC 0 E1	BC 0 E2	BC 0 E3
Stenotrophomonas				11	2		3
Caldibacillus		37					
Moraxella	22		3		12	60	
Cutibacterium	11	6	4	10	9	82	4
Enterococcus	7						170
Paracoccus	107				45	79	1
Bacteroides							248
Staphylococcus				5	11	106	159
Escherichia/Shigella	105	14	18	22	70	254	224
Pseudomonas			3		61	527	166
Enterobacteriaceae unclassified					12	758	40
Other	123	22	3	8	296	816	132
Dietzia	43		16	10	78	1311	7
Micrococcus	165	9	32	9	146	1224	27
Corynebacterium	172	31	19	28	130	4349	46

B) Unspiked WB samples



	WB 0 E1	WB 0 E2
Moraxella		3
Erwiniaceae unclassified		5
Streptococcus		7
Lawsonella	22	
Staphylococcus	22	
Klebsiella	24	
Escherichia/Shigella	36	11
Pseudomonas	58	3
Dietzia	90	3
Enterobacteriaceae unclassified	107	16
Micrococcus	121	17
Other	139	6
Corynebacterium	289	31

C) Unspiked PBMC samples



	PBMC 0 A	PBMC 0 B	PBMC 0 C	PBMC 0 D	PBMC 0 E1	PBMC 0 E2
Enterobacteriaceae unclassified	3				6	69
Blautia			439			
Pseudomonas	0	5	469	81		33
Staphylococcus	450	49	51		721	
Paracoccus			30	39	1379	18
Gemella	1479					
Xanthobacteraceae unclassified	1926					
Cutibacterium	1974	74	215	556	174	6
Comamonadaceae unclassified	3065					
Moraxella	4585	2			87	
Dietzia	1022	257	282	3212	1560	9
Escherichia/Shigella	885	147	626	1457	5076	30
Other	3204	205	1384	1779	2434	42
Micrococcus	1521	475	1322	3873	5746	16
Corynebacterium	4444	670	1734	6525	7830	72

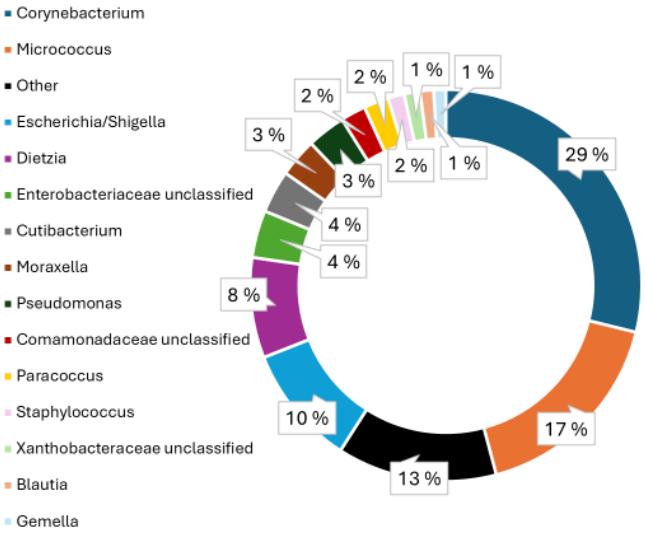
Figure 9. The read counts of the unspiked samples according to the blood fraction. A) The read counts of unspiked BC samples. The counts varied from 98 (BC 0 C) to 9566 (BC 0 E2) with the mean of 1820 reads per sample. B) The read counts of unspiked WB samples. The counts varied from 102 (WB 0 E2) to 908 (WB 0 E1) with the mean of 505 reads per sample. C) The read counts of unspiked PBMC samples. The counts varied from 295 (PBMC 0 E2) to 25013 (PBMC 0 E1) with the mean of 12637 reads per sample. From sample PBMC 0 A, 15232 reads in the *Pseudomonas* genera caused by contamination were removed from the graphs (red rectangle).

The order of abundance in phyla level was the same in all sample groups (Figure 10 B, D and F). *Actinobacteria* was the most common phylum amounting to the mean of 62 %, 53 % and 58% of reads in PBMC, BC and WB samples respectively, followed by *Proteobacteria* (28 %, 30 % and 34 %) as the second most common. *Firmicutes* (8 %, 12 % and 6 %) and *Bacteroidetes* (2 %, 4 % and 1 %) were also present in all sample groups.

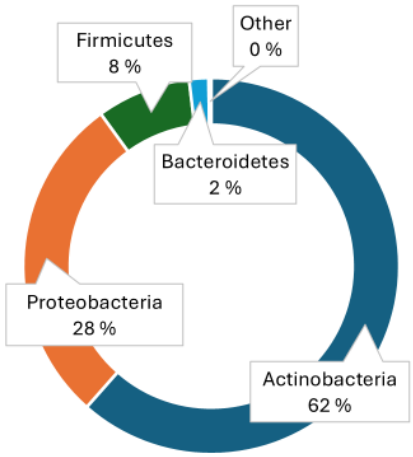
Samples from subject E were taken at different time points to show how the cbDNA composition changes between samplings. The taken BC samples and their genera composition can be seen from figure 11. As can be seen, the cbDNA composition in blood greatly changes between time points.

As cbDNA composition is known to be substantially unique to each person, one aim of this experiment was to characterise differences of cbDNA composition between subjects. A total of 5 subjects donated blood samples for the experiment (labelled A-E). The individual composition of bacterial genera found in the blood samples from the five subjects is shown in figure 12, presented as the mean of the percentages of each genus from all samples taken from the same subject. Unspiked samples from subject A (total of 2 samples, n = 2) had the composition of *Corynebacterium* (20 %), *Micrococcus* (14 %), *Moraxella* (11 %), *Escherichia/Shigella* (9 %), other genera with less than 1 % presence each in the sample group (8 %) and *Paracoccus* (7 %) as the six most common genera. Unspiked samples from subject B (n = 2) on the other hand showed the six most common genera as *Corynebacterium (31 %), Micrococcus (16 %), Caldibacillus (16 %), Escherichia/Shigella* (10 %), *Dietzia* (7 %) and *Cutibacterium* (4 %). Subject C (n = 2) had *Micrococcus* (26 %), *Corynebacterium* (23 %), *Escherichia/Shigella* (14 % %), *Dietzia* (10 %), other (7 %) and *Pseudomonas* (5 %). Subject D's samples (n = 2) had *Corynebacterium* (32 %), *Micrococcus* (15 %), *Escherichia/Shigella* (15 %), *Dietzia* (14 %), *Cutibacterium* (7 %) and other (5 %). Lastly, unspiked samples from subject E (n = 7) showed the six most common genera as *Corynebacterium* (26 %), other (15 %), *Micrococcus* (13 %), *Escherichia/Shigella* (10 %), unclassified *Enterobacteriaceae* (8 %) and *Pseudomonas* (7 %).

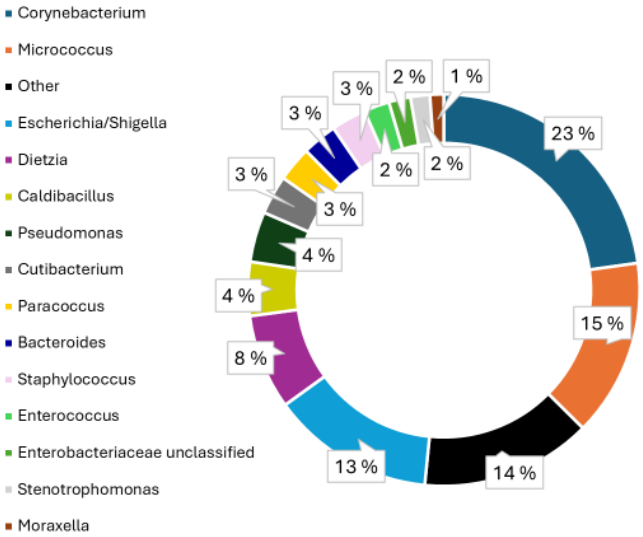
A) Unspiked PBMC samples per Genera (mean %)



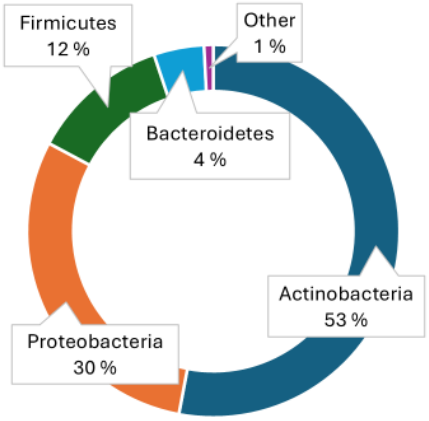
B) Unspiked PBMC samples per Phyla (mean %)



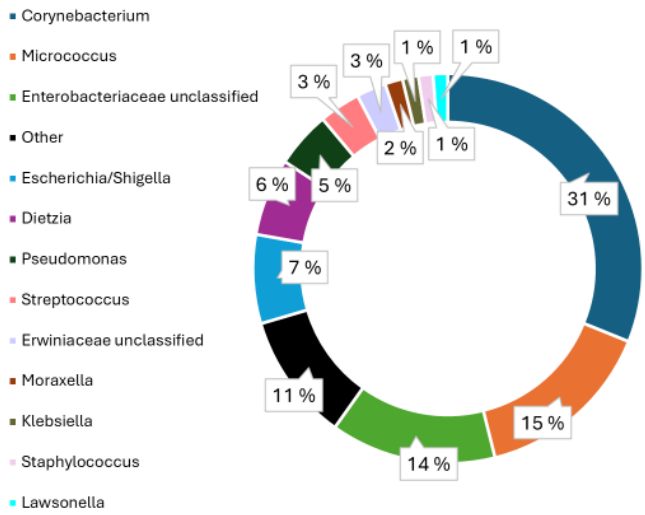
C) Unspiked BC samples per Genera (mean %)



D) Unspiked BC samples per Phyla (mean %)



E) Unspiked WB samples per Genera (mean %)



F) Unspiked WB samples per Phyla (mean %)

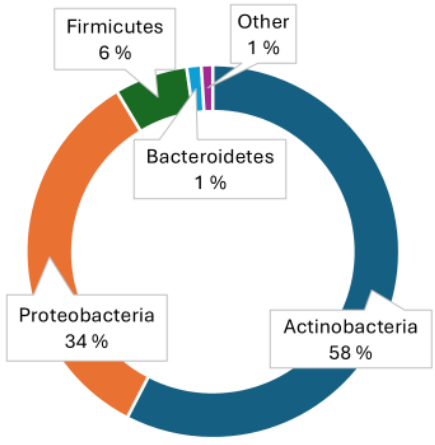


Figure 10. The genera and phyla composition of unspiked samples presented as the mean of the percentages in each sample group. Genera and phyla with less than 1 % presence in their sample group were clustered together under “Other”. A) The composition of genera in unspiked PBMC samples. The six most common genera were *Corynebacterium* (29 %), *Micrococcus* (17 %), other (13 %), *Escherichia/Shigella* (10 %), *Dietzia* (8 %) and unclassified *Enterobacteriaceae* (4 %). B) The composition of phyla in unspiked PBMC samples. Most common phylum was *Actinobacteria* (62 %) followed by *Proteobacteria* (28 %), *Firmicutes* (8 %) and *Bacteroidetes* (2 %). C) The composition of genera in unspiked BC samples. The six most common genera were *Corynebacterium* (23 %), *Micrococcus* (15 %), other (14 %), *Escherichia/Shigella* (13 %), *Dietzia* (8 %) and *Caldibacillus* (4 %). D) The composition of phyla in unspiked BC samples. Most common phylum was *Actinobacteria* (53 %) followed by *Proteobacteria* (30 %), *Firmicutes* (12 %) and *Bacteroidetes* (4 %). Mean 1% of the reads are from other phyla. E) The composition of genera in unspiked WB samples. The six most common genera were *Corynebacterium* (31 %), *Micrococcus* (15 %), unclassified *Enterobacteriaceae* (14 %), other (11 %), *Escherichia/Shigella* (7 %) and *Dietzia* (6 %). F) The composition of phyla in unspiked WB samples. Most common phylum was *Actinobacteria* (58 %) followed by *Proteobacteria* (34 %), *Firmicutes* (6 %) and *Bacteroidetes* (1 %). Mean 1% of the reads are from other phyla.

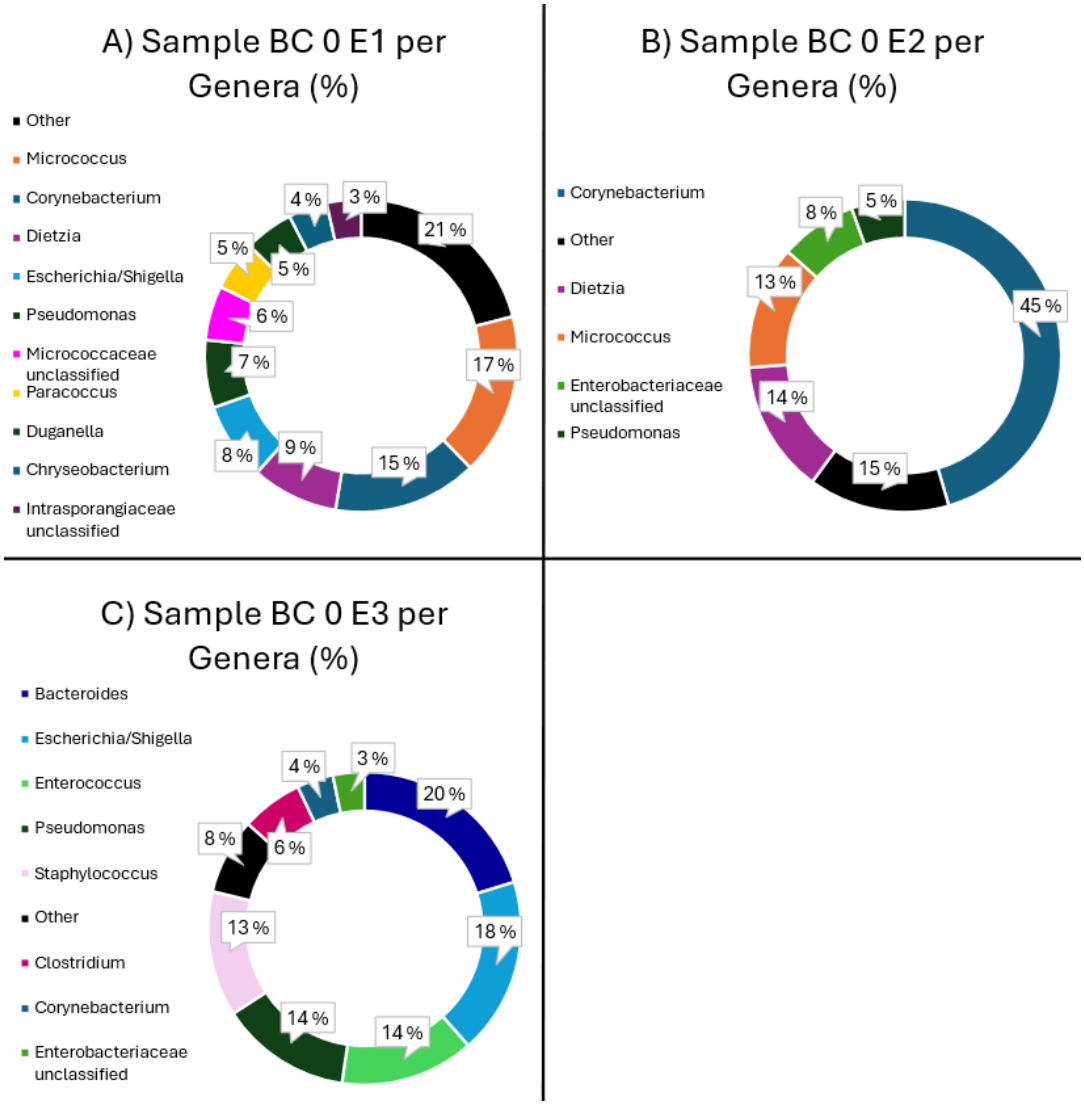
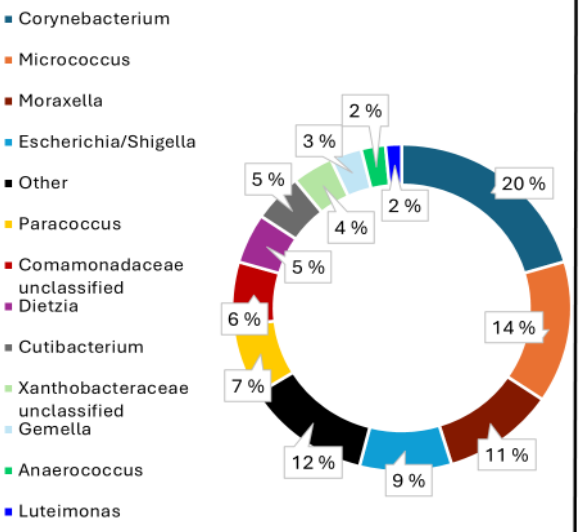
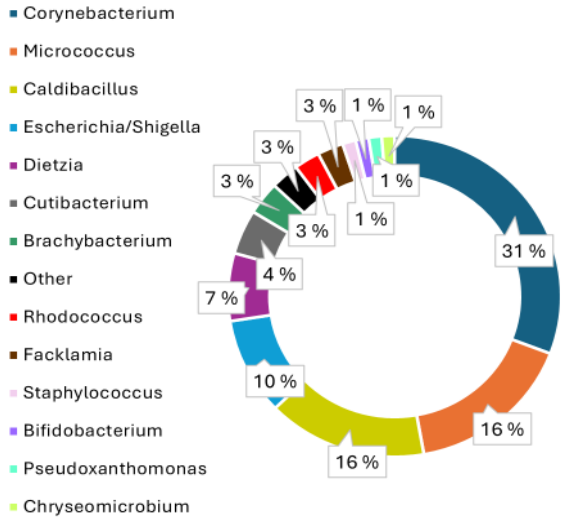


Figure 11. The BC 0 samples from subject E taken at different time points (A, B and C). The composition of cbDNA radically changes between the different timepoints.

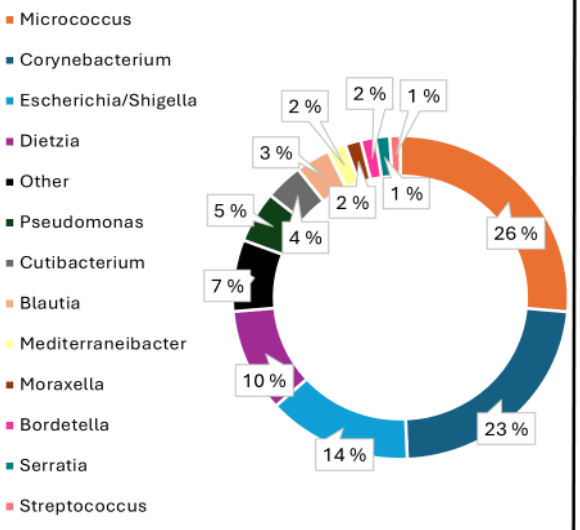
A) Subject A samples per Genera (mean %)



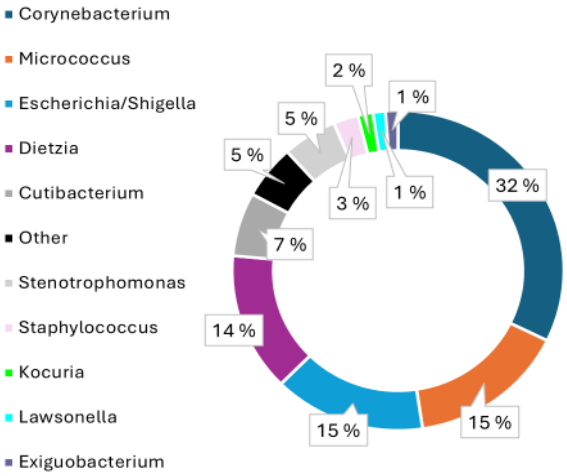
B) Subject B samples per Genera (mean %)



C) Subject C samples per Genera (mean %)



D) Subject D samples per Genera (mean %)



E) Subject E samples per Genera (mean %)

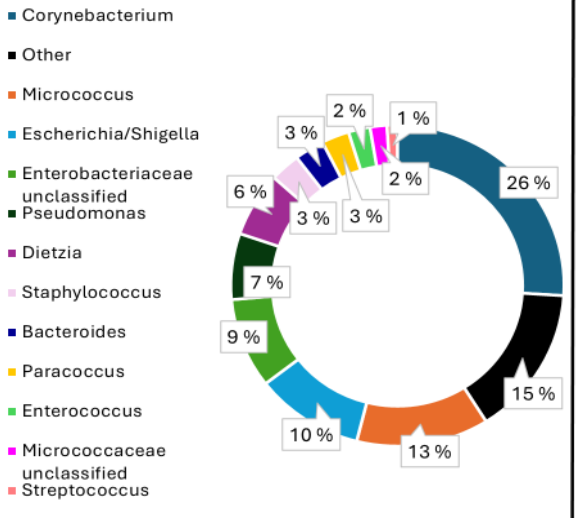


Figure 12. The genera composition of unspiked samples according to subject (A, B, C, D and E) presented as the mean of the percentages in each sample group. Genera with less than 1 % presence in their sample group were clustered together under “Other”. A) Unspiked samples from subject A (n = 2). The six most common genera were *Corynebacterium* (20 %), *Micrococcus* (14 %), *Moraxella* (11 %), *Escherichia/Shigella* (9 %), other (8 %) and *Paracoccus* (7 %). B) Unspiked samples from subject B (n = 2). The six most common genera were *Corynebacterium* (31 %), *Micrococcus* (16 %), *Caldibacillus* (16 %), *Escherichia/Shigella* (10 %), *Dietzia* (7 %) and *Cutibacterium* (4 %). C) Unspiked samples from subject C (n = 2). The six most common genera were *Micrococcus* (26 %), *Corynebacterium* (23 %), *Escherichia/Shigella* (14 % %), *Dietzia* (10 %), other (7 %) and *Pseudomonas* (5 %). D) Unspiked samples from subject D (n = 2). The six most common genera were *Corynebacterium* (32 %), *Micrococcus* (15 %), *Escherichia/Shigella* (15 %), *Dietzia* (14 %), *Cutibacterium* (7 %) and other (5 %). E) Unspiked samples from subject E (n = 7). The six most common genera were *Corynebacterium* (26 %), other (15 %), *Micrococcus* (13 %), *Escherichia/Shigella* (10 %), unclassified *Enterobacteriaceae* (8 %) and *Pseudomonas* (7 %).

3 Discussion

3.1 Methods

After experimenting with the study conditions, we chose to continue to the actual study with 2-hour incubation of spiked samples with concentrations of 2 x 10^6, 2 x 10^5, 2 x 10^4, 2 x 10^3 and 2 x 10^2 bacteria/ml. Incubation of 15 minutes yielded vaguer bands in the agarose gel electrophoresis depicting reduced bacterial uptake of the PBMCs compared to samples incubated for 2 h (Figure 3). On the other hand, samples incubated for 19 hours had similarly prevailing signal on the gel, but during PBMC isolation steps considerable amount of RBC derived debris was present both atop and within the PBMC layer of the density gradient centrifugation. Presence of debris in the sample could have affected the results as DNA released from the RBCs was intermingled with PBMCs. In the experiment regarding the concentration of the bacterial cocktail (Figure 4), we chose to continue with 10^6 as the highest concentration as it was the lowest concentration to show strong signal in the preliminary experiments. Concentrations higher that that would have caused increased risk of contaminating the other samples after PCR especially at the higher cycle amounts. PCR with high number of cycles, 35 to 40, is known to sometimes induce unspecific amplification in the samples producing visible bands in absence of template DNA (Moriyama et al., 2008). As the amount of cbDNA in blood samples is slight, higher number of cycles is necessary for proper amplification. We evaluated cycle counts of 30, 35 and 40 and did not observe any unspecific amplification in any group, thus we chose to continue the experiments with 40 cycles (Figure 5).

During the sample preparation all used reagents and equipment were sterile. Sterility, however, does not account for microbial DNA contaminants, but only living micro-organisms. During DNA extraction and PCR, we used negative controls to ensure no contamination from reagents or equipment. However, some reagents and equipment used only during sample collection and preparation, such as the heparin vacutainer tubes, Ficoll-Paque Plus and beads used for mechanical lysis, were not evaluated for bacterial DNA contaminants. In addition of possible contamination from laboratory equipment and reagents, contamination from skin during sample collection and handling is possible despite careful avoidance of all contamination.

Before and during the DNA extraction, steps were taken to maximise the yield of bacterial DNA. Mechanical lysis with 0.1 mm glass bead was one of those steps as described by Païssé et al. (Païssé et al., 2016). During DNA extraction protocol, the incubation time of the lysate in 70 ̊C was increased from 15 min to 20 min for the same reason. The purity of the DNA extracted with the DNA extraction kit should be 1.60 to 1.90 A_{260}/A_{280}. (Macherey-Nagel, 2022)

With the use of 16S PCR and NGS, the cbDNA located could be from live or dead bacteria or free bacterial DNA from immunodegradation. Also, whether the cbDNA is from bacteria endemic to blood or from transient bacteraemia events, cannot be determines using this method. (Païssé et al., 2016) To assess the effect of background contamination, the negative control was used during sequencing. From figure 6 depicting the agarose gel electrophoresis results after amplicon PCR, we can see that only spiked samples with the two highest concentrations, in addition to some 0-samples, show signal on the gel. However, after index PCR and NGS (Table 1 and Figure 9) we can see that even the samples with no signal have bacterial DNA in them. With such low concentrations of cbDNA in the blood samples and PCR inhibitors in the blood to further complicate things, gel electrophoresis is not particularly sensitive method of detecting presence of cbDNA, but used more as a control method to ensure the PCR has functioned as intended.

During computational analysis of the sequencing data, we merged nearly identical sequences together. It is estimated that 1 difference in 100 bases is most likely due to sequencing error and not biological differences (Hiltemann et al., 2017). As the length of our sequence of interest was about 460 bp (without primers or overhang adapters), the threshold of combining similar sequences during pre-clustering was set to 4 mismatches between the reads. Moreover, during computational analysis, we determined the taxonomy of each assigned OTU to the genera level. This was because 16S sequencing is not often reliable method to differentiate between OTUs at species level (Bosshard et al., 2006; Mignard and Flandrois, 2006). During computational analysis of the data, sample PBMC 0 A was noticed to have a disproportionally large amount of *Pseudomonas* reads (15232 reads). The source of these reads was assumed to be contamination, and the reads were taken out of the following data analysis (Figure 9).

The sample size of five individuals and 15 unspiked samples is relatively small, and definitive conclusions based on this data alone cannot be made. However, when considered with data from previous studies of the same nature, some deductions can be made.

3.2 Results

The purpose of the negative and positive control in the sequencing pool was to detect background contamination of the samples as well as the proper function of the protocol (Figure 7). Zymo Biomics Microbial Community DNA Standard consisting of *Listeria* 12%, *Pseudomonas* 12%, *Bacillus* 12%, *Escherichia* 12%, *Salmonella* 12%, *Limosilactobacillus* 12%, *Enterococcus* 12%, *Staphylococcus* 12%, *Saccharomyces* 2%, and *Cryptococcus* 2% was used as the positive control. When comparing the amounts of each genus after sequencing we can see that the composition shows acceptable variation to the original (*Listeria* 14 %, *Pseudomonas* 5%, *Bacillus* 18 %, *Escherichia/Shigella* 10 %, *Salmonella* 11 %, Limosilactobacillus 15 %, *Enterococcus* 11 %, *Staphylococcus* 16 %) proving our sequencing protocol to be mostly effective in mirroring the genera present in our samples. Negative control shows us surprisingly large amount of reads with *Escherichia/Shigella* genus overrepresented likely due to contamination. Other genera present in the negative control include *Micrococcus*, a possible contaminant from skin, and *Methylobacterium*, a common contaminant of DNA extraction kits (Salter et al., 2014). The abundant presence of *Escherichia/Shigella* in the negative control sample, if due to background contamination, could mean that the genus is overrepresented in other samples as well.

From the spiked samples we can see the read count increasing in correlation to the increasing concentration of the bacteria cocktail (Figure 8). Based on these results *Clostridium* and *Staphylococcus* genera seem to localise to the PBMCs more than to BC of WB. *Pseudomonas* on the other hand seem to prefer the BC, where *Staphylococcus* also has notable presence. *Escherichia/Shigella*, *Bacteroidetes* and *Enterococcus* localise predominantly to the WB instead of leucocytes, with *Escherichia/Shigella* showing some affinity to BC as well. The reason why the spiked genera are not equally represented in WB is most likely due to more efficient immunodegradation of some genera of bacteria. The affinity of certain genera to BC or PBMCs could be explained by the different leucocyte compositions of the fractions. Different leucocytes might have more affinity certain bacterial genera leading to increased endocytosis (Mohr et al., 2006; Taha et al., 2016; Wolf et al., 2005). Especially the lack of granulocytes in PBMCs could cause alteration in internalisation and immunodegradation between the blood fractions.

In the unspiked samples, the read counts vary a lot between the blood fractions (mean of 12637 reads per sample in PBMCs, mean of 1820 reads per sample in BC and mean of 505

reads per sample in WB) (Figure 9). This cannot be entirely explained by the dynamic changes of HMB composition that naturally occur as time goes on, nor individual differences between donors of blood factions, as the difference in read counts is observed between samples taken at the same time from the same donor (PBMC and BC samples from subjects A, B, C and D) and the differences between the blood fractions are over 10-fold in the unspiked samples. In comparison, spiked samples, which are all taken from the same person at the same time, show similar read counts between blood fractions, only the composition of the reads vary significantly (Figure 8). As mentioned before, PBMCs and BC might have varied immunodegradation properties due to their different leucocyte composition. The increased read counts in PBMCs compared to BC and WB could be due to impaired immunodegradation within the fraction due to the lack of granulocytes. When compared to the spiked samples, the amount of cbDNA in the unspiked samples is miniscule, and even a low amount of immunodegradation can cause significant effects when the remaining bacterial DNA is amplified. Another explanation for the varying read counts between BC and PBMC samples could be that the cells which are present in both BC and PBMC fractions (lymphocytes and monocytes) carry most of the bacterial load of the BC. The elimination of platelets and granulocytes could cause more concentrated cbDNA presence in PBMCs compared to BC. Between samples taken at different time points, the changes in the composition and abundance of cbDNA is explained by the dynamic changes to the HMB that can be affected by factors including diet, time of day, stress, and other environmental factors. It is important to note that all sample groups have a small sample size, especially WB from which we only had two samples and thus these results cannot be generalised.

Corynebacterium, Micrococcus, Escherichia/Shigella and *Dietzia* are the most common genera in all unspiked fractions (Figure 10). WB samples have a substantial portion of unclassified *Enterobacteriaceae,* the family that includes *Escherichia/Shigella* genera. As mentioned before, 16S sequencing sometimes struggles with differentiation between genera (Bosshard et al., 2006; Mignard and Flandrois, 2006). *Escherichia* and *Shigella* genera are closely related and their reliable differentiation is difficult at best, hence they are of the grouped together (Devanga Ragupathi et al., 2018). The composition of cbDNA between the fractions at phylum level is similar in all groups. *Actinobacteria* dominates, followed by *Proteobacteria, Firmicutes* and some *Bacteroidetes*. This differs from the previously reported findings, where *Proteobacteria* was the most dominant phylum, followed by *Actinobacteria, Firmicutes* and *Bacteroidetes* (Païssé et al., 2016). This difference may be due to the small

samples size of the experiment, causing results that do not represent the full population to be emphasised. Another reason might be local differences caused by genetic of environmental factors as all subjects were in this experiment were Finnish. Possible contamination of samples, especially from skin sources during sampling, might also cause possible overrepresentation of *Actinobacteria*.

As mentioned before, *Bacteroidetes* and *Firmicutes* are the most prominent phyla in gut in addition to less prevalent presence of *Actinobacteria* and *Proteobacteria*. Genera such as *Bacteroides, Prevotella, Clostridium, Bifidobacteria, Eubacterium* and *Ruminococcus* are present in abundance in the gut microbiome (Adak and Khan, 2019). Out of these genera only *Bacteroides* were discovered in low amounts from the BC samples (Figure 10). Escherichia is notably present in the gut microbiota, but can also be commonly isolated from the skin (Petkovšek et al., 2009). Similarly, Pseudomonas can be found on both skin and gut where they act as opportunistic pathogens (Qin et al., 2022). Skin on the other hand is dominated by *Actinobacteria,* including *Corynebacterium*, *Micrococcus*, *Dietzia* and *Cutibacterium,* all of which are notably present in our samples. Oral microbiota is known to inhabit *Streptococcus* species as well as *Moraxella.* Based on our results the composition of our samples is dominated by genera associated with skin microbiome, with some genera from oral and gut microbiome also present.

The composition of found cbDNA from this study resemble more that of the skin than gut. While similar notions about the cuticular source of cbDNA has been made (Khan et al., 2022), skin is also a major source of contamination for biological samples. *Dietzia* and *Corynebacterium* are both inhabitants of the skin microbiome. Out of the two *Corynebacterium* is known to be a common contaminant (Tauch et al., 2016). Whether the source of *Corynebacterium* in our samples is from actual biological cbDNA of from contaminant sources, cannot be reliably determined based on this experiment alone and requires more research. *Micrococcus* species have been previously found in HBM and our results mimics these results (Damgaard et al., 2015). In case of *Escherichia/Shigella*, the abundant presence of the genera in the negative controls sample might insinuate more copious occurrence of the genera in the data compared to the actual biological presence in the blood. While some *Micrococcus* was found in the positive control. other genera present in the unspiked samples were not precent in the negative control and in these cases the effect of background contamination can be ruled insignificant. The most common laboratory

contaminants, *Bacillus, Flavobacteria, Fusobacteria, Propionibacterium* and *Serratia* genera, nevertheless, were not present in meaningful amounts in any of the samples.

As seen in the figure 12, the genera composition varies significantly between individuals. When genera composition is compared, some genera are present only in some individuals. For example, *Caldibacillus* is present in blood of subject B, but not significantly in others. Subject A on the other hand has the most diverse genera composition out of the subjects, displaying genera such as *Moraxella* and *Paracoccus*. The genera composition also radically changes between time points as seen from figure 11. In this experiment, environmental factors affecting the composition of cbDNA, such as diet, the time of day the samples were taken and physical activity, were not followed, leading to dramatically varying results from samples taken at different time points. This highlights the need for standardised sampling schedule, maybe including fasting before sampling, in the future experiments to gain more comparable results between time points. Also, if cbDNA is to be used as a biomarker for disease progression or onset, a standard profile of cbDNA composition should be determined to each person to reliably quantify changes in the profile. Conclusions about disease state could hardly be made form single sample. Our results show that the HBM profile is unique to each individual and deeply dynamic. However, our sample size of 2 to 7 samples per subject is limited and no definitive conclusions can be made on this data alone.

3.3 Conclusions

This experiment was conducted as a pilot study for the role of HBM in SLE. We characterised the composition of cbDNA form five healthy individuals and set up a protocol for cbDNA extraction, amplification, sequencing, and analysis. As a secondary objective we characterised the localisation of six common bacterial genera to BC, PBMCs and WB as the localisation could be crucial factor in determining the usefulness of cbDNA as a biomarker in disease settings. In addition to cbDNA composition between the blood factions, we also analysed the data comparing the HBM of different individuals to each other. Finally, we checked to see whether the samples from same individuals alter between timepoints.

Our results show that HBM is highly dynamic, diverse, and individual, with more resemblance to skin microbiome than to the gut. We confirmed that the localisation of bacterial genera alters between blood fractions and that the composition of cbDNA varies between individuals as well as time points. While this experiment provided valuable insight into the composition of HBM, the results also highlighted the need of rigorous controls

throughout the experiments when analysing cbDNA composition. In the future research, especially when determining differences between HBM and the cbDNA composition in disease, a much larger sample size is required to make conclusions as individual samples do not give full picture of the dynamic HBM consistency. As a note to future research the conditions of sampling should be standardised to produce comparable results between time points. In the future studies the methods used in these experimental methods can be used to determine HBM's relationship to SLE.

Research on HBM dysbiosis has important implications on human health and disease. This study was the first step in understanding the dynamic function of the HBM and in the future after more research possible biomarkers indicated by cbDNA composition could be discovered not only to SLE, but to other chronic diseases as well. The possibility of novel therapeutic targets affecting HBM dysbiosis is another exiting future prospect of cbDNA research.

4 Materials and methods

4.1 Method validation

We experimented on different incubation times and bacteria cocktail concentrations for the spiked samples as well as the cycle amounts for the amplicon PCR before performing the final protocol. To discover the optimal incubation time for spiked sampled, nine PBMC samples spiked with *E. coli* (c = 1.5 x 10^8 or 1.5 x 10^7 bacterial/ml) and unspiked PBMC samples were incubated for 19 h, 15 minutes or 2 h RT with one sample of each concentration in each group. Sample collection, the isolation of PBMCs, the spiking with *E. coli,* mechanical lysis, DNA extraction and amplicon PCR were done with the same methods as detailed in section 4.2, but with different concentrations and bacteria.

To test the best concentration of bacteria cocktail by incubating PBMC samples spiked with bacterial cocktail (*E. coli* and *E. faecalis*) in concentrations of 1 x 10^7, 1 x 10^6, 1 x 10^5, 1 x 10^4, 1 x 10^3 and 1 x 10^2 (six samples) for 2 h RT. Sample collection, the isolation of PBMCs, the spiking with *E. coli* and *E. faecalis* cocktail, mechanical lysis, DNA extraction and amplicon PCR were done with the same methods as detailed in section 4.2, but with different concentrations and bacteria. The PCR cycle detailed in section 4.4.1 was replicated three times with varying cycle amounts (30 x, 35x and 40x) with the DNA isolated from the previous PBMC samples spiked with *E. coli* and *E. faecalis* to evaluate for optimal cycle settings.

4.2 Sample collection and preparation

4.2.1 Sample collection

The blood samples were collected from the healthy volunteers (n = 5) after cleaning and disinfecting the area. Approximately 6 ml of blood was drawn into 10 ml vacutainer tubes with heparin and processed during the same day. All steps until after Amplicon PCR were done in laminar flow cabinets, except for DNA concentration measurements with Qubit. Filter pipette tips were used in all steps. Contamination of the samples was avoided at all times.

4.2.2 Spiking

A bacteria cocktail consisting of 1 ml of *S. aureus, E. faecalis, E. coli, Ps. aeruginosa, B. fragilis* and *Cl. perfringens* (c = 2 x 10^8 bacteria/ml) was mixed. Subsequently, a 10-fold

dilution series from 2 x 10^8 bacteria/ml to 2 x 10^2 bacteria/ml was prepared by diluting 1 ml of the bacteria cocktail with 9 ml PBS (gibco sterile filtered PBS, pH 7.4, Life Technologies Limited, UK) and repeating the process six times. In sterile 15 ml falcon tubes, 3 ml of each bacteria cocktail dilution (c = 2 x 10^6, 2 x 10^5, 2 x 10^4, 2 x 10^3 and 2 x 10^2) was mixed with two 6 ml tubes of blood separately, in addition, two 6 ml tubes of blood were mixed with 3 ml of PBS. The falcon tubes were mixed and incubated for 2 h in room temperature (RT). After incubation, 1 ml of blood incubated with bacteria was taken aside from each set of the samples for lysis (WB samples). The rest of the samples were further diluted with PBS to 1:1 ratio with the original volume of blood, followed by BC/PBMC isolation.

4.2.3 Isolation of BC

The fresh samples of blood (6 ml each) were diluted 1:1 with PBS and mixed (unspiked samples). The diluted samples, both spiked and unspiked, were centrifuged at RT and 800 x *g* for 10 min with deacceleration set to 0. The plasma layer was carefully discarded and the BC layer collected with sterile transfer pipettes. The samples were washed twice with 1:1 ration of sample to PBS and resuspended to 1ml of PBS. The BC samples were transferred to bead tubes for lysis.

4.2.4 Isolation of PBMCs

The fresh samples of blood (6 ml each) were diluted 1:1 with PBS and mixed. 9 ml of the samples were carefully layered on top of 5 ml of RT Ficoll-Paque Plus (Ficoll-Paque™ PLUS, cytiva) in sterile 15 ml falcon tubes. The tubes were centrifuged at RT and 400 x *g* for 30 min with deacceleration set to 0. After centrifugation, the PBMC layer was carefully collected and washed twice with 1:1 ratio of sample to PBS. After washes the cells were resuspended in 400 µl of PBS and transferred to bead tubes for lysis.

4.2.5 Mechanical lysis

The samples were mechanically lysed before DNA extraction to improve the yield of bacterial DNA (Païssé et al., 2016). The samples were lysed in sterile 5 ml tubes with 0.1 mm glass beads (PowerBead Tubes, Glass 0.1 mm, Qiagen, Germany) in 30 Hz for 1.5 min (TissueLyzer II, Qiagen, USA) followed by centrifugation at RT and 14 000 RPM for 20 min. The supernatant was collected into sterile 1.5 ml microcentrifuge tubes and stored at -20 °C until DNA extraction.

4.3 DNA extraction

The DNA from the lysis samples was extracted with NucleoSpin Blood Kit (NucleoSpin® Blood, Macherey-Nagel, Germany) following the protocol in the user manual (Macherey-Nagel, 2022). 200 µl of the lysis samples, or 200 µl or PBS for control sample, and 25 µl of proteinase K were added to 1.5 ml microcentrifuge tubes followed by 200 µl of B3 buffer and vigorous vortexing (~15 s). The samples were incubated for 20 min in 70 °C (vortex during incubation at 10 min). After incubation, 210µl of absolute ethanol (ETAX, min. 99.5 p-%, Altia Oyj, Finland) was added and to adjust DNA binding conditions followed by vortexing. The samples were loaded into NucleoSpin Blood columns and centrifuged at RT and 11 000 x *g* for 1 min. After the first centrifugation, the flow-through was discarded and 500 µl of BW buffer was added to the column followed by centrifugation with the same parameters as the first one. After the second centrifugation the flow-through was again discarded and 600 µl of B5 buffer was added before the third centrifugation. The flow-through was discarded as before and the columns were centrifuged again with the same parameters to dry the membrane. Before the final centrifugation, 100 µl of 70 °C BE buffer was added to the columns and incubated 1 min at RT. The final flow-through was collected to sterile 1.5 ml microcentrifuge tubes and stored at -20 °C.

The DNA concentrations were measured from the extracted samples using Qubit 2.0 (Qubit 2.0 Fluorometer and reagent kit, Life Technologies, USA). Qubit High Sensitivity (HS) working solution was mixed (1 µl Qubit HS reagent per sample and 199 µl Qubit HS buffer per sample). Standard curve was created by measuring Qubit HS standards: 190 µl Qubit HS working solution mixed with 10 µl Qubit HS Standard 1 or 2, followed by 2 min incubation at RT. The samples were measured with same amounts of Qubit HS working solution (190 µl) and sample (10 µl) mixed, before 2 min incubation at RT. Control sample with water instead of sample was also measured. For any samples that had over 600 ng/ml DNA concentration, Broad Range (BR) Qubit was performed using the same protocol while replacing Qubit HS reagent, buffer and standards with Qubit BR reagent, buffer, and standards.

4.4 16S ribosomal RNA PCR

4.4.1 Amplicon PCR

Before the amplicon PCR, the isolated DNA templates were diluted to 6.25 ng/µl with PCR pure aqua. Samples that had concentration lower than 6.25 ng/µl were not diluted.

For the 16S amplicon PCR, master mix was prepared on cold blocks by combining 16.5 µl of 2 x KAPA Hifi hotstart ready mix (Roche, USA), 1 µl of both Amplicon forward and reverse primers (6.6 µM, Figure 2) and 2.5 µl of PCR pure aqua per sample. In sterile PCR strips (8-well 0.2 ml PCR tube strips, Nippon Genetics) on cold blocks, 21 µl of the master mix and 12 µl of the template, or PCR aqua in case of negative controls, were mixed. The PCR was run on BioRad T100 Thermal Cycler for 40 cycles, with on cycle consisting of initial denaturation (95 °C, 3 min), denaturation (95 °C, 30 s), annealing (55 °C, 30 s), extension (72 °C, 30 s) and final extension (72 °C, 5 min).

1.5 % TAE gel was prepared for the gel electrophoresis of the PCR products. For one gel, 1.8 g of standard agarose was microwaved with 120 ml of 1 x TAE until dissolved. 6 µl of Midori green DNA dye (Midori green Advance DNA Stain, Nippon Genetics) was added to the solution before pouring it to the mould and allowing it to set for an hour. After setting, the gel was submerged in 1 x TAE buffer. 8 µl of the PCR products were mixed with 2 µl of loading dye (6X DNA Loading Dye, Thermo Scientific, Lithuania) to create the samples for electrophoresis. 8 µl of the samples were loaded into the wells alongside some wells with 3.5 µl of 100 bp ladder (GeneRuler 100 bp DNA Ladder, Thermo Scientific, Lithuania). The electrophoresis was run for 60 min at 120 V. The gel was imaged using GelDoc XR+ Imaging system (BioRad).

The rest of the PCR products were purified using AMpure XP beads (AMpure XP, Beckman Coulter, USA). 20 µl of RT and well mixed beads were added to the PCR products and mixed well before incubating them at RT for 5 min while shaking, followed by 2 min incubation on a magnetic plate. The supernatant was removed and 200 µl of freshly prepared 80 % ethanol was added before the subsequent 30 s incubation. The ethanol wash was repeated once. After the second wash, the supernatant was all carefully removed, and the beads were allowed to dry for about 6 min until they became visibly matte. Once dry, the beads were resuspended in 52.5 µl of PCR pure aqua and incubated for 2 min off magnet and 2 min on the magnetic plate. The supernatant was collected and stored ad -20 °C until index PCR.

4.4.2 Index PCR

To add index tags, index PCR was performed to the purified amplicon PCR products. The master mix was prepared on cold blocks by combining 25 µl of 2 x KAPA Hifi hotstart ready mix and 10 µl of PCR pure aqua per sample. 35 µl of the master mix, 5 µl of the purified amplicon PCR product, including negative control sample, and 10 µl of Illumina UD indexes

(IDT for Illumina, DNA/RNA UD indexes, Illumina, USA) were combined carefully for each sample ensuring no contamination from other indexes. The PCR was run on BioRad for 8 cycles, with the same programme as the amplicon PCR.

The index PCR products were purified using the AMpure XP beads. 56 µl of RT well suspended beads were added to the index PCR products, mixed well, and incubated for 5 min in RT off the magnet while shaking, followed by 2 min incubation on the magnetic plate. The supernatant was discarded, and the beads were washed twice with 200 µl of fresh 80 % ethanol similarly as in the purification of the amplicon PCR products. After the washes, the beads were allowed to dry for about 6 min and afterwards resuspended into 27.5 µl of PCR pure aqua. The suspension was incubated 2 min off the magnet and 2 min on the magnet plate. 25 µl of the supernatant was collected for DNA concentration measurements and sequencing. The purified samples were stored in -20 °C.

The DNA concentrations of the purified index PCR products were measured with the same protocol as detailed before in section 4.3. All samples were measured twice, and the mean was used in the next steps.

4.5 Library preparation and sequencing

The purified index PCR products were diluted to 4 nM with PCR pure aqua (negative control samples were not diluted), the run cassette (MiSeq Reagent Kit v3, Illumina, USA) was defrosted and 0.2 M NaOH was prepared (800 µl of PCR pure aqua and 200 µl 1 M NaOH (Sodium hydroxide solution, 1.0N, Sigma Life science, UK)). From each of the diluted PCR products and controls (PCR product of the negative control, and positive control (ZymoBiomics microbial community DNA standard, Zymo Research, USA)), 5 µl was combined to create the sample pool while the pool was kept on ice. The pool was denatured by combining 5 µl of 0.2 M NaOH to 5 µl of the 4 nM pool. The denatured pool vas incubated for 5 min at RT and diluted to 20 pM by adding 990 µl of cold HT1 buffer (Illumina MiSeq Reagent Kit). This sample library was further diluted to 4 pM by mixing 114 µl of the 20 pM sample library with 456 µl of cold HT1.

PhiX control (10 nm PhiX library, PhiX Control v3, Illumina, USA) was prepared by first diluting the PhiX to 4 nM with PCR pure aqua and then denaturing the library with 5 µl of 0.2 M NaOH. The denatured library was incubated for 5 min at RT before 990 µl of cold HT1 was added (c = 20 pM). For the MiSeq run the PhiX library was further diluted with cold HT1

to the concentration of 8 pM. The PhiX library, sample library and HT1 buffer were kept on ice. Finally, 60 µl of the 8 pM PhiX library was combined with 570 µl of the 4 pM sample library (~20 % PhiX concentration). The final combined library was incubated at 96 °C for 2 min, mixed and incubated for 5 min on ice just before loading the sample (600 µl) to the cassette. The Illumina MiSeq sequencer was prepared according to the machine's instructions and the run was set for paired end reads for read lengths of 301 with adapter trimming on.

4.6 Computational methods

The sequencing data was cleaned, aligned and classified modifying the Mothur MiSeq standard operating procedure (Schloss et al., 2009) in Galaxy open-source platform version 24.2.4.dev0 (Batut et al., 2018; Hiltemann et al., 2017, 2023).The forward and reverse reads were paired and the overlapping sections combined with the "Make.contigs" command. Reads under 250 bp or over 600 bp were removed as well as all with ambiguous base calls using the "Screen.seqs" tool (minlenght: 250, maxlenght: 600 and maxambig: 0). To identify unique reads and record the number of identical sequences we ran the "Uniqie.seqs" tool for the trimmed data followed by "Count.seqs" tool. The sequences were aligned with Mothur formatted SILVA-based bacterial reference alignment using "Align.seqs" command (Mothur formatted SILVA release v102 (Quast et al., 2012; Yilmaz et al., 2014; Glöckner et al., 2017; Yarza et al., 2014; Schloss et al., 2009)), and any reads that did not overlap with the V3-V4 region (positions 6388 - 25316) were removed as well as any overhang using "Screen.seqs" (start: 6388, end: 25316, maxhomop: 8) and "Filter.seqs" (vertical: yes, trump: .) tools. The "Unique.seqs" command was rerun to remove any duplicate sequences. Next the alignment was cleaned and redundancy removed. From the remaining sequences, all near identical sequences (with less than 4 mismatches) were merged in pre-clustering with "Pre.cluster" tool (diffs: 4), followed by chimera removal with "Chimera.vsearch" (dereplicate: yes) and "Remove.seqs" tools. The next step was to assign taxonomy to the sequences. This was done with the Mothur-formatted version of the Ribosomal Database Project 16S rRNA reference (RDP version 19 (Schloss et al., 2009; Cole et al., 2014)) and the "Classify.seqs" tool. After assigning taxonomy, all non-bacterial sequences were removed with the "Remove.lineage" tool. The sequences were then clustered into OTUs based on 97 % sequence identity and the OTUs were classified at genus level with "Cluster.split" (taxlevel: 4, cutoff: 0.03) and "Classify.otu" tools (Westcott and Schloss, 2017).

The subsequent read counts were further analysed in Microsoft Excel. Percentage of each genus of phyla precent in each sample was calculated and the mean of these percentages within the sample group in question was reported in addition to the read counts. All genera or phyla with less than 1 % mean presence in the according sample group was pooled under the label ‘Other’. In spiked samples all but the genera used for spiking were labelled ‘Other’.

5 Acknowledgements

I would like to thank my supervisors Docent Arno Hänninen and PhD Teemu Kallonen for their support and guidance throughout the thesis project. Thank you to members of Hänninen group, Mikrobisto group and all personnel of Medisiina D 7th floor for the continuous help and support during this project. Special thanks to Päivi Haaranen for helping and guiding me in the laboratory and to Juha Grönroos for providing the bacteria to this experiment. Finally, thanks to my family and friends who pushed, supported, and encouraged me in the process.

6 Abbreviations list

SLE	Systemic lupus erythematosus
cbDNA	Circulating bacterial DNA
HBM	Healthy blood microbiome
WB	Whole blood
BC	Buffy coat
PBMC	Peripheral blood mononuclear cell
EV	Extracellular vesicle
rRNA	Ribosomal RNA
NGS	Next generation sequencing
OTU	Operational taxonomic unit
HS	High sensitivity
BR	Broad range

References

Adak, A., and M.R. Khan. 2019. An insight into gut microbiota and its functionalities. *Cell Mol Life Sci*. 76:473–493. doi:10.1007/s00018-018-2943-4.

Akhil, A., R. Bansal, K. Anupam, A. Tandon, and A. Bhatnagar. 2023. Systemic lupus erythematosus: latest insight into etiopathogenesis. *Rheumatol Int*. 43:1381–1393. doi:10.1007/s00296-023-05346-x.

Amar, J., C. Lange, G. Payros, C. Garret, C. Chabo, O. Lantieri, M. Courtney, M. Marre, M.A. Charles, B. Balkau, R. Burcelin, and D.E.S.I.R. Study Group. 2013. Blood microbiota dysbiosis is associated with the onset of cardiovascular events in a large general population: the D.E.S.I.R. study. *PLoS One*. 8:e54461. doi:10.1371/journal.pone.0054461.

Amar, J., M. Serino, C. Lange, C. Chabo, J. Iacovoni, S. Mondot, P. Lepage, C. Klopp, J. Mariette, O. Bouchez, L. Perez, M. Courtney, M. Marre, P. Klopp, O. Lantieri, J. Doré, M.A. Charles, B. Balkau, and R. Burcelin. 2011. Involvement of tissue bacteria in the onset of diabetes in humans: evidence for a concept. *Diabetologia*. 54:3055–3061. doi:10.1007/s00125-011-2329-8.

Ameer, M.A., H. Chaudhry, J. Mushtaq, O.S. Khan, M. Babar, T. Hashim, S. Zeb, M.A. Tariq, S.R. Patlolla, J. Ali, S.N. Hashim, and S. Hashim. 2022. An Overview of Systemic Lupus Erythematosus (SLE) Pathogenesis, Classification, and Management. *Cureus*. 14:e30330. doi:10.7759/cureus.30330.

Aringer, M., K. Costenbader, D. Daikh, R. Brinks, M. Mosca, R. Ramsey-Goldman, J.S. Smolen, D. Wofsy, D.T. Boumpas, D.L. Kamen, D. Jayne, R. Cervera, N. Costedoat-Chalumeau, B. Diamond, D.D. Gladman, B. Hahn, F. Hiepe, S. Jacobsen, D. Khanna, K. Lerstrøm, E. Massarotti, J. McCune, G. Ruiz-Irastorza, J. Sanchez-Guerrero, M. Schneider, M. Urowitz, G. Bertsias, B.F. Hoyer, N. Leuchten, C. Tani, S.K. Tedeschi, Z. Touma, G. Schmajuk, B. Anic, F. Assan, T.M. Chan, A.E. Clarke, M.K. Crow, L. Czirják, A. Doria, W. Graninger, B. Halda-Kiss, S. Hasni, P.M. Izmirly, M. Jung, G. Kumánovics, X. Mariette, I. Padjen, J.M. Pego-Reigosa, J. Romero-Diaz, Í. Rúa-Figueroa Fernández, R. Seror, G.H. Stummvoll, Y. Tanaka, M.G. Tektonidou, C. Vasconcelos, E.M. Vital, D.J. Wallace, S. Yavuz, P.L. Meroni, M.J. Fritzler, R. Naden, T. Dörner, and S.R. Johnson. 2019. 2019 European League Against Rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus. *Ann Rheum Dis*. 78:1151–1159. doi:10.1136/annrheumdis-2018-214819.

Baker, J.L., J.L. Mark Welch, K.M. Kauffman, J.S. McLean, and X. He. 2024. The oral microbiome: diversity, biogeography and human health. *Nat Rev Microbiol*. 22:89–104. doi:10.1038/s41579-023-00963-6.

Batut, B., S. Hiltemann, A. Bagnacani, D. Baker, V. Bhardwaj, C. Blank, A. Bretaudeau, L. Brillet-Guéguen, M. Čech, J. Chilton, D. Clements, O. Doppelt-Azeroual, A. Erxleben, M.A. Freeberg, S. Gladman, Y. Hoogstrate, H.-R. Hotz, T. Houwaart, P. Jagtap, D. Larivière, G. Le Corguillé, T. Manke, F. Mareuil, F. Ramírez, D. Ryan, F.C. Sigloch, N. Soranzo, J. Wolff, P. Videm, M. Wolfien, A. Wubuli, D. Yusuf, J. Taylor, R. Backofen, A. Nekrutenko, and B. Grüning. 2018. Community-Driven Data

Analysis Training for Biology. *Cell Systems*. 6:752-758.e1. doi:10.1016/j.cels.2018.05.012.

Bosshard, P.P., R. Zbinden, S. Abels, B. Böddinghaus, M. Altwegg, and E.C. Böttger. 2006. 16S rRNA gene sequencing versus the API 20 NE system and the VITEK 2 ID-GNB card for identification of nonfermenting Gram-negative bacteria in the clinical laboratory. *J Clin Microbiol*. 44:1359–1366. doi:10.1128/JCM.44.4.1359-1366.2006.

Cao, S.-Y., C.-N. Zhao, X.-Y. Xu, G.-Y. Tang, H. Corke, R.-Y. Gan, and H.-B. Li. 2019. Dietary plants, gut microbiota, and obesity: Effects and mechanisms. *Trends in Food Science & Technology*. 92:194–204. doi:10.1016/j.tifs.2019.08.004.

Castillo, D.J., R.F. Rifkin, D.A. Cowan, and M. Potgieter. 2019. The Healthy Human Blood Microbiome: Fact or Fiction? *Front Cell Infect Microbiol*. 9:148. doi:10.3389/fcimb.2019.00148.

Chee, W.J.Y., S.Y. Chew, and L.T.L. Than. 2020. Vaginal microbiota and the potential of Lactobacillus derivatives in maintaining vaginal health. *Microb Cell Fact*. 19:203. doi:10.1186/s12934-020-01464-4.

Christovich, A., and X.M. Luo. 2022. Gut Microbiota, Leaky Gut, and Autoimmune Diseases. *Front Immunol*. 13:946248. doi:10.3389/fimmu.2022.946248.

Cogen, A.L., V. Nizet, and R.L. Gallo. 2008. Skin microbiota: a source of disease or defence? *Br J Dermatol*. 158:442–455. doi:10.1111/j.1365-2133.2008.08437.x.

Cole, J.R., Q. Wang, J.A. Fish, B. Chai, D.M. McGarrell, Y. Sun, C.T. Brown, A. Porras-Alfaro, C.R. Kuske, and J.M. Tiedje. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucl. Acids Res.* 42:D633–D642. doi:10.1093/nar/gkt1244.

Damgaard, C., K. Magnussen, C. Enevold, M. Nilsson, T. Tolker-Nielsen, P. Holmstrup, and C.H. Nielsen. 2015. Viable Bacteria Associated with Red Blood Cells and Plasma in Freshly Drawn Blood Donations. *PLoS ONE*. 10:e0120826. doi:10.1371/journal.pone.0120826.

De Goffau, M.C., S. Lager, U. Sovio, F. Gaccioli, E. Cook, S.J. Peacock, J. Parkhill, D.S. Charnock-Jones, and G.C.S. Smith. 2019. Human placenta has no microbiome but can contain potential pathogens. *Nature*. 572:329–334. doi:10.1038/s41586-019-1451-5.

Devanga Ragupathi, N.K., D.P. Muthuirulandi Sethuvel, F.Y. Inbanathan, and B. Veeraraghavan. 2018. Accurate differentiation of Escherichia coli and Shigella serogroups: challenges and strategies. *New Microbes New Infect*. 21:58–62. doi:10.1016/j.nmni.2017.09.003.

Dinakaran, V., A. Rathinavel, M. Pushpanathan, R. Sivakumar, P. Gunasekaran, and J. Rajendhran. 2014. Elevated levels of circulating DNA in cardiovascular disease patients: metagenomic profiling of microbiome in the circulation. *PLoS One*. 9:e105221. doi:10.1371/journal.pone.0105221.

Domingues, S., and K.M. Nielsen. 2017. Membrane vesicles and horizontal gene transfer in prokaryotes. *Current Opinion in Microbiology*. 38:16–21. doi:10.1016/j.mib.2017.03.012.

Ghaemi, F., A. Fateh, A.A. Sepahy, M. Zangeneh, M. Ghanei, and S.D. Siadat. 2022. Blood microbiota composition in Iranian pre-diabetic and type 2 diabetic patients. *HAB*. 29:243–248. doi:10.3233/HAB-210450.

Gibiino, G., L.R. Lopetuso, F. Scaldaferri, G. Rizzatti, C. Binda, and A. Gasbarrini. 2018. Exploring Bacteroidetes: Metabolic key points and immunological tricks of our gut commensals. *Digestive and Liver Disease*. 50:635–639. doi:10.1016/j.dld.2018.03.016.

Glassing, A., S.E. Dowd, S. Galandiuk, B. Davis, and R.J. Chiodini. 2016. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog*. 8:24. doi:10.1186/s13099-016-0103-7.

Glöckner, F.O., P. Yilmaz, C. Quast, J. Gerken, A. Beccati, A. Ciuprina, G. Bruns, P. Yarza, J. Peplies, R. Westram, and W. Ludwig. 2017. 25 years of serving the community with ribosomal RNA gene reference databases and tools. *Journal of Biotechnology*. 261:169–176. doi:10.1016/j.jbiotec.2017.06.1198.

Goraya, M.U., R. Li, A. Mannan, L. Gu, H. Deng, and G. Wang. 2022. Human circulating bacteria and dysbiosis in non-infectious diseases. *Front Cell Infect Microbiol*. 12:932702. doi:10.3389/fcimb.2022.932702.

Grice, E.A., and J.A. Segre. 2011. The skin microbiome. *Nat Rev Microbiol*. 9:244–253. doi:10.1038/nrmicro2537.

Hammad, D.B.M., S.L. Hider, V.C. Liyanapathirana, and D.P. Tonge. 2019. Molecular Characterization of Circulating Microbiome Signatures in Rheumatoid Arthritis. *Front Cell Infect Microbiol*. 9:440. doi:10.3389/fcimb.2019.00440.

Hiltemann, S., B. Batut, and D. Clements. 2017. 16S Microbial Analysis with mothur (extended) (Galaxy Training Materials).

Hiltemann, S., H. Rasche, S. Gladman, H.-R. Hotz, D. Larivière, D. Blankenberg, P.D. Jagtap, T. Wollmann, A. Bretaudeau, N. Goué, T.J. Griffin, C. Royaux, Y. Le Bras, S. Mehta, A. Syme, F. Coppens, B. Droesbeke, N. Soranzo, W. Bacon, F. Psomopoulos, C. Gallardo-Alba, J. Davis, M.C. Föll, M. Fahrner, M.A. Doyle, B. Serrano-Solano, A.C. Fouilloux, P. Van Heusden, W. Maier, D. Clements, F. Heyl, Galaxy Training Network, B. Grüning, and B. Batut. 2023. Galaxy Training: A powerful framework for teaching! *PLoS Comput Biol*. 19:e1010752. doi:10.1371/journal.pcbi.1010752.

Ho, H., L. Radigan, G. Bongers, A. El-Shamy, and C. Cunningham-Rundles. 2021. Circulating bioactive bacterial DNA is associated with immune activation and complications in common variable immunodeficiency. *JCI Insight*. 6:e144777. doi:10.1172/jci.insight.144777.

Iwai, T. 2009. Periodontal bacteremia and various vascular diseases. *J of Periodontal Research*. 44:689–694. doi:10.1111/j.1600-0765.2008.01165.x.

Janda, J.M., and S.L. Abbott. 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol*. 45:2761–2764. doi:10.1128/JCM.01228-07.

Jin, S., D. Wetzel, and M. Schirmer. 2022. Deciphering mechanisms and implications of bacterial translocation in human health and disease. *Current Opinion in Microbiology*. 67:102147. doi:10.1016/j.mib.2022.102147.

Johnson, J.S., D.J. Spakowicz, B.-Y. Hong, L.M. Petersen, P. Demkowicz, L. Chen, S.R. Leopold, B.M. Hanson, H.O. Agresta, M. Gerstein, E. Sodergren, and G.M. Weinstock. 2019. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun*. 10:5029. doi:10.1038/s41467-019-13036-1.

Justiz Vaillant, A.A., A. Goyal, and M. Varacallo. 2023. Systemic Lupus Erythematosus. *In* StatPearls. StatPearls Publishing, Treasure Island (FL).

Kajihara, M., S. Koido, T. Kanai, Z. Ito, Y. Matsumoto, K. Takakura, M. Saruta, K. Kato, T. Odamaki, J.-Z. Xiao, N. Sato, and T. Ohkusa. 2019. Characterisation of blood microbiota in patients with liver cirrhosis. *Eur J Gastroenterol Hepatol*. 31:1577–1583. doi:10.1097/MEG.0000000000001494.

Kell, D.B., and E. Pretorius. 2018. No effects without causes: the Iron Dysregulation and Dormant Microbes hypothesis for chronic, inflammatory diseases. *Biol Rev Camb Philos Soc*. 93:1518–1557. doi:10.1111/brv.12407.

Khan, I., I. Khan, Z. Jianye, Z. Xiaohua, M. Khan, M.G. Hilal, M.A. Kakakhel, A. Mehmood, A. Lizhe, and L. Zhiqiang. 2022. Exploring blood microbial communities and their influence on human cardiovascular disease. *J Clin Lab Anal*. 36:e24354. doi:10.1002/jcla.24354.

Kim, S.I., N. Kang, S. Leem, J. Yang, H. Jo, M. Lee, H.S. Kim, D.N. Dhanasekaran, Y.-K. Kim, T. Park, and Y.S. Song. 2020. Metagenomic Analysis of Serum Microbe-Derived Extracellular Vesicles and Diagnostic Models to Differentiate Ovarian Cancer and Benign Ovarian Tumor. *Cancers*. 12:1309. doi:10.3390/cancers12051309.

Klindworth, A., E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, and F.O. Glöckner. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*. 41:e1–e1. doi:10.1093/nar/gks808.

Kozich, J.J., S.L. Westcott, N.T. Baxter, S.K. Highlander, and P.D. Schloss. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol*. 79:5112–5120. doi:10.1128/AEM.01043-13.

Kumar, S., and A. Kumar. 2022. Microbial pathogenesis in inflammatory bowel diseases. *Microbial Pathogenesis*. 163:105383. doi:10.1016/j.micpath.2021.105383.

Kwan, B.C.-H., K.-M. Chow, C.-B. Leung, M.-C. Law, P.M.-S. Cheng, V. Yu, P.K.-T. Li, and C.-C. Szeto. 2013. Circulating bacterial-derived DNA fragments as a marker * of systemic inflammation in peritoneal dialysis. *Nephrology Dialysis Transplantation*. 28:2139–2145. doi:10.1093/ndt/gft100.

Lee, H., H.K. Lee, S.K. Min, and W.H. Lee. 2020. 16S rDNA microbiome composition pattern analysis as a diagnostic biomarker for biliary tract cancer. *World J Surg Onc*. 18:19. doi:10.1186/s12957-020-1793-3.

Li, Q., C. Wang, C. Tang, X. Zhao, Q. He, and J. Li. 2018. Identification and Characterization of Blood and Neutrophil-Associated Microbiomes in Patients with Severe Acute Pancreatitis Using Next-Generation Sequencing. *Front. Cell. Infect. Microbiol.* 8:5. doi:10.3389/fcimb.2018.00005.

Macherey-Nagel. 2022. User manual: genomic DNA from blood. *Bioanalysis*. Rev. 17.

Mancuso, G., A. Midiri, C. Biondo, C. Beninati, M. Gambuzza, D. Macrì, A. Bellantoni, A. Weintraub, T. Espevik, and G. Teti. 2005. Bacteroides fragilis-derived lipopolysaccharide produces cell activation and lethal toxicity via toll-like receptor 4. *Infect Immun*. 73:5620–5627. doi:10.1128/IAI.73.9.5620-5627.2005.

Martel, J., C.-Y. Wu, P.-R. Huang, W.-Y. Cheng, and J.D. Young. 2017. Pleomorphic bacteria-like structures in human blood represent non-living membrane vesicles and protein particles. *Sci Rep*. 7:10650. doi:10.1038/s41598-017-10479-8.

Mignard, S., and J.P. Flandrois. 2006. 16S rRNA sequencing in routine bacterial identification: A 30-month experiment. *Journal of Microbiological Methods*. 67:574–581. doi:10.1016/j.mimet.2006.05.009.

Mitchell, A.J., W.D. Gray, M. Schroeder, H. Yi, J.V. Taylor, R.S. Dillard, Z. Ke, E.R. Wright, D. Stephens, J.D. Roback, and C.D. Searles. 2016. Pleomorphic Structures in Human Blood Are Red Blood Cell-Derived Microparticles, Not Bacteria. *PLoS ONE*. 11:e0163582. doi:10.1371/journal.pone.0163582.

Mohr, H., A. Bayer, U. Gravemann, and T.H. Müller. 2006. Elimination and multiplication of bacteria during preparation and storage of buffy coat–derived platelet concentrates. *Transfusion*. 46:949–955. doi:10.1111/j.1537-2995.2006.00827.x.

Moriyama, K., C. Ando, K. Tashiro, S. Kuhara, S. Okamura, S. Nakano, Y. Takagi, T. Miki, Y. Nakashima, and H. Hirakawa. 2008. Polymerase chain reaction detection of bacterial 16S rRNA gene in human blood. *Microbiology and Immunology*. 52:375–382. doi:10.1111/j.1348-0421.2008.00048.x.

Mu, Q., H. Zhang, and X.M. Luo. 2015. SLE: Another Autoimmune Disorder Influenced by Microbes and Diet? *Front Immunol*. 6:608. doi:10.3389/fimmu.2015.00608.

Niess, J.H., S. Brand, X. Gu, L. Landsman, S. Jung, B.A. McCormick, J.M. Vyas, M. Boes, H.L. Ploegh, J.G. Fox, D.R. Littman, and H.-C. Reinecker. 2005. CX_3 CR1-Mediated Dendritic Cell Access to the Intestinal Lumen and Bacterial Clearance. *Science*. 307:254–258. doi:10.1126/science.1102901.

Nikkari, S., I.J. McLaughlin, W. Bi, D.E. Dodge, and D.A. Relman. 2001. Does Blood of Healthy Subjects Contain Bacterial Ribosomal DNA? *J Clin Microbiol*. 39:1956–1959. doi:10.1128/JCM.39.5.1956-1959.2001.

Païssé, S., C. Valle, F. Servant, M. Courtney, R. Burcelin, J. Amar, and B. Lelouvier. 2016. Comprehensive description of blood microbiome from healthy donors assessed by 16S targeted metagenomic sequencing. *Transfusion*. 56:1138–1147. doi:10.1111/trf.13477.

Pan, Q., F. Guo, Y. Huang, A. Li, S. Chen, J. Chen, H.-F. Liu, and Q. Pan. 2021. Gut Microbiota Dysbiosis in Systemic Lupus Erythematosus: Novel Insights into Mechanisms and Promising Therapeutic Strategies. *Front Immunol*. 12:799788. doi:10.3389/fimmu.2021.799788.

Panaiotov, S., G. Filevski, M. Equestre, E. Nikolova, and R. Kalfin. 2018. Cultural Isolation and Characteristics of the Blood Microbiome of Healthy Individuals. *AiM*. 08:406–421. doi:10.4236/aim.2018.85027.

Petkovšek, Ž., K. Eleršič, M. Gubina, D. Žgur-Bertok, and M. Starčič Erjavec. 2009. Virulence Potential of Escherichia coli Isolates from Skin and Soft Tissue Infections. *J Clin Microbiol*. 47:1811–1817. doi:10.1128/JCM.01421-08.

Pongen, Y.L., D. Thirumurugan, R. Ramasubburayan, and S. Prakash. 2023. Harnessing actinobacteria potential for cancer prevention and treatment. *Microbial Pathogenesis*. 183:106324. doi:10.1016/j.micpath.2023.106324.

Potgieter, M., J. Bester, D.B. Kell, and E. Pretorius. 2015. The dormant blood microbiome in chronic, inflammatory diseases. *FEMS Microbiol Rev*. 39:567–591. doi:10.1093/femsre/fuv013.

Pretorius, E., S. Mbotwe, J. Bester, C.J. Robinson, and D.B. Kell. 2016. Acute induction of anomalous and amyloidogenic blood clotting by molecular amplification of highly substoichiometric levels of bacterial lipopolysaccharide. *J R Soc Interface*. 13:20160539. doi:10.1098/rsif.2016.0539.

Qin, S., W. Xiao, C. Zhou, Q. Pu, X. Deng, L. Lan, H. Liang, X. Song, and M. Wu. 2022. Pseudomonas aeruginosa: pathogenesis, virulence factors, antibiotic resistance, interaction with host, technology advances and emerging therapeutics. *Sig Transduct Target Ther*. 7:199. doi:10.1038/s41392-022-01056-1.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F.O. Glöckner. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*. 41:D590–D596. doi:10.1093/nar/gks1219.

Rhee, S.J., H. Kim, Y. Lee, H.J. Lee, C.H.K. Park, J. Yang, Y.-K. Kim, S. Kym, and Y.M. Ahn. 2020. Comparison of serum microbiome composition in bipolar and major depressive disorders. *Journal of Psychiatric Research*. 123:31–38. doi:10.1016/j.jpsychires.2020.01.004.

Ricci, V., D. Carcione, S. Messina, G.I. Colombo, and Y. D'Alessandra. 2020. Circulating 16S RNA in Biofluids: Extracellular Vesicles as Mirrors of Human Microbiome? *IJMS*. 21:8959. doi:10.3390/ijms21238959.

Salter, S.J., M.J. Cox, E.M. Turek, S.T. Calus, W.O. Cookson, M.F. Moffatt, P. Turner, J. Parkhill, N.J. Loman, and A.W. Walker. 2014. Reagent and laboratory contamination

can critically impact sequence-based microbiome analyses. *BMC Biol*. 12:87. doi:10.1186/s12915-014-0087-z.

Schloss, P.D., S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A. Lesniewski, B.B. Oakley, D.H. Parks, C.J. Robinson, J.W. Sahl, B. Stres, G.G. Thallinger, D.J. Van Horn, and C.F. Weber. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol*. 75:7537–7541. doi:10.1128/AEM.01541-09.

Sender, R., S. Fuchs, and R. Milo. 2016a. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell*. 164:337–340. doi:10.1016/j.cell.2016.01.013.

Sender, R., S. Fuchs, and R. Milo. 2016b. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol*. 14:e1002533. doi:10.1371/journal.pbio.1002533.

Taha, M., M. Kalab, Q. -L. Yi, E. Maurer, C. Jenkins, P. Schubert, and S. Ramirez-Arcos. 2016. Bacterial survival and distribution during buffy coat platelet production. *Vox Sanguinis*. 111:333–340. doi:10.1111/vox.12427.

Tauch, A., I. Fernández-Natal, and F. Soriano. 2016. A microbiological and clinical review on Corynebacterium kroppenstedtii. *International Journal of Infectious Diseases*. 48:33–39. doi:10.1016/j.ijid.2016.04.023.

Tedeschi, G.G., D. Amici, and M. Paparelli. 1969. Incorporation of Nucleosides and Amino-acids in Human Erythrocyte Suspensions: Possible Relation with a Diffuse Infection of Mycoplasms or Bacteria in the L Form. *Nature*. 222:1285–1286. doi:10.1038/2221285a0.

Tian, J., D. Zhang, X. Yao, Y. Huang, and Q. Lu. 2023. Global epidemiology of systemic lupus erythematosus: a comprehensive systematic analysis and modelling study. *Ann Rheum Dis*. 82:351–356. doi:10.1136/ard-2022-223035.

Velmurugan, G., V. Dinakaran, J. Rajendhran, and K. Swaminathan. 2020. Blood Microbiota and Circulating Microbial Metabolites in Diabetes and Cardiovascular Disease. *Trends in Endocrinology & Metabolism*. 31:835–847. doi:10.1016/j.tem.2020.01.013.

Weckerle, C.E., and T.B. Niewold. 2011. The unexplained female predominance of systemic lupus erythematosus: clues from genetic and cytokine studies. *Clin Rev Allergy Immunol*. 40:42–49. doi:10.1007/s12016-009-8192-4.

Westcott, S.L., and P.D. Schloss. 2017. OptiClust, an Improved Method for Assigning Amplicon-Based Sequence Data to Operational Taxonomic Units. *mSphere*. 2:e00073-17. doi:10.1128/mSphereDirect.00073-17.

Whittle, E., M.O. Leonard, R. Harrison, T.W. Gant, and D.P. Tonge. 2018. Multi-Method Characterization of the Human Circulating Microbiome. *Front Microbiol*. 9:3266. doi:10.3389/fmicb.2018.03266.

Wiest, R., M. Lawson, and M. Geuking. 2014. Pathological bacterial translocation in liver cirrhosis. *Journal of Hepatology*. 60:197–209. doi:10.1016/j.jhep.2013.07.044.

Wolf, K., E. Fischer, and T. Hackstadt. 2005. Degradation of Chlamydia pneumoniae by peripheral blood monocytic cells. *Infect Immun*. 73:4560–4570. doi:10.1128/IAI.73.8.4560-4570.2005.

Xiao, Q., W. Lu, X. Kong, Y.W. Shao, Y. Hu, A. Wang, H. Bao, R. Cao, K. Liu, X. Wang, X. Wu, S. Zheng, Y. Yuan, and K. Ding. 2021. Alterations of circulating bacterial DNA in colorectal cancer and adenoma: A proof-of-concept study. *Cancer Letters*. 499:201–208. doi:10.1016/j.canlet.2020.11.030.

Yamaguchi, M., Y. Terao, Y. Mori-Yamaguchi, H. Domon, Y. Sakaue, T. Yagi, K. Nishino, A. Yamaguchi, V. Nizet, and S. Kawabata. 2013. Streptococcus pneumoniae invades erythrocytes and utilizes them to evade human innate immunity. *PLoS One*. 8:e77282. doi:10.1371/journal.pone.0077282.

Yarza, P., P. Yilmaz, E. Pruesse, F.O. Glöckner, W. Ludwig, K.-H. Schleifer, W.B. Whitman, J. Euzéby, R. Amann, and R. Rosselló-Móra. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol*. 12:635–645. doi:10.1038/nrmicro3330.

Yilmaz, P., L.W. Parfrey, P. Yarza, J. Gerken, E. Pruesse, C. Quast, T. Schweer, J. Peplies, W. Ludwig, and F.O. Glöckner. 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucl. Acids Res.* 42:D643–D648. doi:10.1093/nar/gkt1209.