TURUN YLIOPISTO

Optimization of DNA extraction from prostate tissue for metagenomic sequencing

Molecular Biosciences, Cell Biology, Master of Science
Master's thesis

Author:
Petteri Hirva

22.05.2025
Turku

Master's thesis

Major: Cell biology
Author: Petteri Hirva
Title: Optimization of DNA extraction from prostate tissue for metagenomic sequencing
Supervisors: LT Peter Boström, FM Sanja Vanhatalo
Number of pages: 56
Date: 23.05.2025

The gut microbiome composition of prostate cancer patients has been observed to differ from that of healthy patients. A limited number of studies have shown that the prostate may harbor a unique microbiome, which could contribute to the development of prostate cancer. However, a reliable method for determining tissue microbiota is still lacking. Next generation sequencing of the 16S rRNA gene and the whole DNA in Shotgun Metagenomic Sequencing are most often used for microbiome composition analysis, but both methods have known experimental and computational challenges. In this study, we compared whether removing host background (host depletion) affected the microbial profiles of prostate tissue samples analyzed with 16S sequencing and Shotgun metagenomics.

Prostate tissue samples were obtained from Turku University Hospital. Tissue samples were homogenized and divided in two subsample sets, of which one was treated with MolYsis™ Basic5 kit and the other went straight to microbial DNA extraction, which was performed to both sets. The extraction was done using Chemagic ™ DNA Stool 200 mg Kit H96 with Magnetic Separation Module I extraction robot. The analysis also included negative controls (OMNIgene fluid, DNA/RNA Shield fluid), extraction controls (Chemagic Lysis Buffer 1) and ZymoBIOMICS Gut microbiome standards. The microbial composition was determined by using 16S rRNA gene amplicon sequencing targeting the V3–V4 hypervariable regions as well as Shotgun Metagenomics. Both pooled libraries were sequenced with the Illumina platform.

The number of reads of the samples received from 16S rRNA gene amplicon sequencing was on average 21×10^5 while the average for Shotgun Metagenomics was 44×10^6. Majority of the reads were either host reads or unclassified by the database. All the species shared by negative controls and samples that might be attributable to contamination were excluded, which left a small number of bacteria species. Moreover, the microbial profiles were different between host depletion samples and untreated samples which refers that the host depletion changed the microbial profile of a prostate tissue sample. In the future, the V1–V2 region of the 16S rRNA gene could be tested instead of V3–V4 region for tissue microbiome analysis to check if it has better specificity compared to V3–V4. Shotgun metagenomics on the other hand would need more sequencing depth.

Keywords: Microbiome, Prostate microbiome, Next generation sequencing

Contents

1 Introduction

1.1 Introduction to prostate cancer

The prostate gland (PG) is a small pear-shaped organ that is a part of the male reproductive system. It weighs around 20 g and is located below the bladder and surrounds a part of the urethra. Anatomically, PG can be divided into different compartments to recognize the development of pathological changes in the PG. The compartments are central, transitional, periurethral and peripheral zone. An epithelium consisting of basal cells and secretory ciliary cells are covering the tubulo alveolar glands of the prostate. The glandular structures are separated from the supportive connective tissue by a basement membrane. The growth and development of the prostate gland is regulated by the male hormones secreted by testes. PG's primary function is to produce the fluid component of semen. Moreover, PG secretes prostate-specific antigen (PSA) which aids sperm mobility. So far, elevated PSA levels (>3) are the most accurate indicator of prostate cancer (PCa), which is the 4th most common cancer type in men after lung, colorectum, and liver cancer. (Patologia, 2012.)

PCa develops typically in men aged over 45 (median 68). In Europe, the survival rate has been between 50-90 percent after 5 years from the diagnosis, in the absence of other causes of death except cancer. (Ferlay et al. 2024, ECIS 2024.) In addition to elevated PSA levels, the common understanding has been that higher levels of free testosterone (T) on the serum are linked to increased risk for prostate cancer (Asanad et al., 2024; Pierorazio et al., 2010; Watts et al., 2022). However, some studies have shown this may not always be the case (Mearini et al., 2013).

1.1.1 Gleason Grading

Gleason grade grouping was invented by Dr. Donald Gleason when he noticed that cancerous cells follow 5 distinct patterns while changing from normal cells to tumor cells. The scale is graded from one to five, with five being the highest grade, where the cells typically do not have resemblance to normal prostate cells at all. While making diagnosis, pathologists grade the most and second most predominant pattern, for instance 3+4. The patterns sum up to Gleason score, which can theoretically range from 2-10, although the total is rarely under 6 making the range from 6-10. The Gleason score relates closely to ISUP grade group, which was proposed by the

International Society of Urologic Pathologists (ISUP) in 2012. ISUP doesn't change the histopathological diagnosis but describes more accurately the diagnostics as the rarity of Gleason 1 and 2. Based on large patient studies, the prognostic value of Gleason scoring has been demonstrated to be very precise. (Patologia, 2012.)

Whether a tissue sample taken from a diagnosed patient is benign or malign is determined by the nature of the tissue. Typically, the benign core doesn't have any carcinoma in the sample, and malign has at least to some extent. However, tissue cores with lower than 30% of carcinoma are not optimal samples in terms of their representativeness of the hole prostate as will be further discussed. Samples that have carcinoma in them will be given Gleason grade depending on the appearance of the carcinoma. (Freedland et al., 2003; Phipps et al., 2005.)

1.2 The microbial etiology of cancer

1.2.1 Gut microbiota

The human gut microbiome is a sum of all microbes, bacteria, fungi and viruses and their genes. Bacteria are classified taxonomically according to phyla, classes, orders and families. Everyone develops a unique gut microbiome profile which consists of more than 100 trillion micro-organisms. The most dominant gut bacterial phyla are Firmicutes and Bacteroidetes which form 90% of microbiota composition. According to Helander and Fändriks et al (2014), the human gastrointestinal (GI) tract covers the total surface area of 32 m^2 and the microbiome composition vary in each part of the GI tract. Majority of microbes reside in the large intestine as the environment is less acidic and hostile than in the small intestine. The total weight of GM is approximately around 2 kilograms, and the density of bacterial cells has been estimated to be at 10^{10} to 10^{12} per milliliter. According to (Rinninella et al., 2019) this would make the GI one of the most densely populated microbial habitats on earth. The development of GM begins immediately after birth and is shaped throughout life. However, the changes in early life, such as type of delivery, methods of milk feeding, and their composition and use of antibiotics, are the most significant factors contributing to GM composition. Later in life, the use of antibiotics and diet is the major factor affecting the health of the GM. (Shalon et al., 2023.)

1.2.2 Tumorigenesis and bacteria

Recent studies have shown a possible relationship between PCa and gut microbiota (GM) (Fujita et al., 2022; Liss et al., 2018). The GM composition of PCa patients has been observed to differ from that of healthy patients, which has raised a question about the possibility of making a PCa risk assessment based on GM signature (Kalinen et al 2024, Liss et al 2024). Moreover, the RNA sequences of gut microbes found in prostate tissue are similar to those found in stool samples of PCa patients. Thus, it has been suggested that the GM would have a direct or an indirect pathway through which it affects the progression of PCa (Kalinen et al 2024; Lachance 2024). Additionally, the relative amount of *Firmicutes* was associated with higher serum testosterone levels independent of host factors. The amount of Firmicutes and Lachnospira were also in higher abundances in patients with PCa (Kalinen et al., 2024; Matsushita et al., 2022).

McCulloch and Trinchieri (2021) found a reduction in alpha diversity of GM both in human patients and mice models. Furthermore, they found an alteration in certain commensal bacteria, which can synthesize androgens and testosterone from precursor pregnenolone, and this change is believed to be associated with PCa aggressiveness. Recent findings suggest that gut microbiota (GM) composition differs significantly between stages of prostate cancer progression. Notably, castration-resistant prostate cancer (CRPC) mice exhibited a more enriched and diverse GM profile compared to hormone-sensitive prostate cancer (HSPC) mice, implying a potential role for gut bacteria in driving disease advancement. Specifically, *Ruminococcus* (*Mediterraneibacter*) *gnavus* and *Bacteroides acidifaciens* were found in higher abundance in CRPC mice. Consistent with these findings, CRPC patients receiving androgen deprivation therapy (ADT) showed elevated levels of *Ruminococcus* species, including DSM_100440 and OM05_10BH, alongside a marked reduction in *Prevotellaceae* species. These observations support the hypothesis that specific microbial shifts may contribute to prostate cancer progression and resistance to hormone therapy. However, it is challenging to prove whether the alterations in the GM occurred before the PCa or whether the PCa changed the GM composition afterward. (McCulloch & Trinchieri, 2021.)

Numerous studies have shown that diet plays a major role in shaping gut microbiota (GM) composition. A high-fat, highly processed Western diet has been strongly associated with gut dysbiosis—an imbalance in the gut microbiome that can disrupt normal bodily functions

through multiple pathways. This dysregulation has been linked to a range of chronic conditions, including low-grade inflammation, increased intestinal permeability ("leaky gut"), colorectal cancer (CRC), and inflammatory bowel diseases (IBD) (Adolph and Tilg, 2024). Given these connections, research exploring the relationship between prostate cancer (PCa) and GM is essential for understanding the broader role of GM in human health.

Sepich-Poore et al (2021) reported that there is experimental evidence that some microbes such as *Fusobacterium nucleatum* and *Salmonella* strains amplify tumorigenesis by promoting inflammation and activating proto-oncogenes (Fig. 1). The mechanisms rely behind the activation of E-cadherin-Wnt- β-catenin signaling pathway via virulence factors like FadA and AvrA. By activating β-catenin, FadA can promote inflammatory and oncogenic responses which can enhance tumorigenesis in colorectal cancer. AvrA is a product secreted by certain *Salmonella* strains which can affect eukaryotic signaling pathways by upregulating β-catenin. Eventually this could promote the production of proto-oncogene c-Myc and thus, enhance tumorigenesis. In addition, Wang et al (2024) found *F. nucleatum* to be related to genetic and epigenetic lesions, such as microsatellite instability (MSI) and CpG island methylator phenotype (CIMP), in CRC. The abundance of *F. nucleatum*, a gram negative, non-spore forming anaerobe, has been reported to have significantly increased in stool samples of patients with CRC (Wang et al. 2024). Guo et al. (2020) found that *F. nucleatum* infection could potentially cause tumor cells to secrete certain kind of exosomes that can promote metastatic behavior. However, no connection has been reported between PCa and *F. nucleatum*. (Gao et al. 2022; Guo et al.; Sepich-Poore et al. 2021)

According to Fujita et al (2022), overall diversity is not significantly different between patients with PCa and healthy controls. However, certain species have been reported to be more abundant in PCa. Men with castration resistant prostate cancer (CRPC) have an increased abundance of bacteria with androgenic functions. *Akkermansia muciniphila* is related to maintenance of healthy gut wall and *Ruminococcus* to DHEA and testosterone production which are downstream metabolites of pregnenolone and hydroxy pregnenolone. In addition, in patients with CRPC, *Ruminococcus* was associated with poor prognosis and *Prevotella* with favorable. According to (Liss et al., 2018) a case control study found a higher abundance of *Bacteroides massiliensis* in PCa patients and a higher *Faecalibacterium prausnitzii* and *Eubacterium rectale* in controls group which suggests a possible link to micronutrient metabolism. With a follow-up study, involving 133 rectal swabs samples collected at least two

weeks prior to transrectal prostate biopsy, the same group found that bacterial taxa associated with carbohydrate metabolism pathways were enriched in PCa patients. In contrast, bacteria involved in natural B-vitamin production were underrepresented. Despite these functional differences, the overall bacterial community structure was largely similar between patients with and without prostate cancer.

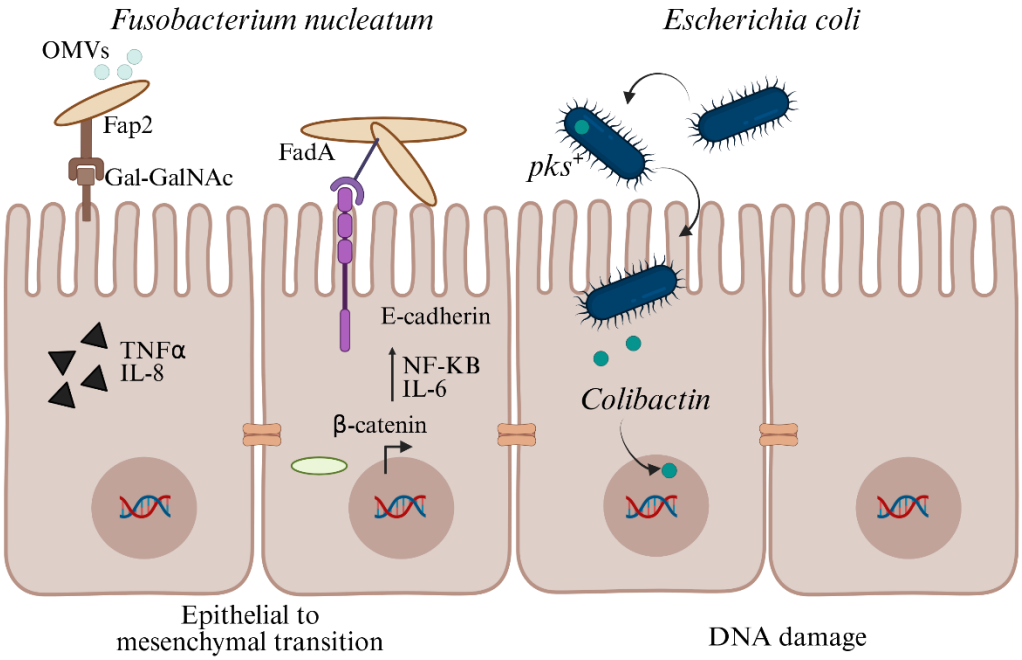

Figure 1. Impact of *Fusobacterium nucleatum* and *Escherichia coli* on neoplastic processes in epithelial cells. *F. nucleatum* might trigger cancer via virulence factors like FadA adhesin which starts a proinflammatory cascade mediated by NF-κB and IL-6. Fap2, another adhesin that interacts with D-galactose-b (1–3)-N-acetyl-D-galactosamine (Gal-GalNAc) at the tumor surface enhances cellular proliferation with Wnt/β-catenin pathway increasing proinflammatory cytokine production. *E. coli* on the other hand has an arsenal of virulence factors and toxins capable of pathogenic functions. Cullin et al. (2021), borrowed and edited.

1.2.3 Intratumoral microbiome

The current research emphasizes often the well-known connection between GM and health. However, emerging studies have reported an association between intratumoral microbiome and various cancer types, such as lung cancer, colorectal cancer and pancreatic cancer. Suggestions state that microbes found from solid tumor environments could be derived from the gut. Moreover, only around 25% of the microbes found from lung, colorectal and pancreatic tumors were originally from the gut. (Gao et al., 2022; Guo et al., 2021; Riquelme et al., 2019.)

So far, the number of studies about intratumoral PCa microbiome is limited. The current view states, as prostate can get exposed to numerous micro-organisms through the urethra, that this would eventually cause chronic inflammation and stress through epigenetic alterations, mutagenesis and reactive oxygen species. In addition, recent research has found an association between carcinogenesis risk and the infection history of any sexually transmitted infection. (Miyake et al. 2022.) Intratumoral microbiome has also been suggested to recruit inflammatory mediators as well as immune cells such as T, B and NK cells from periphery. The species such as *Klebsiella pneumoniae*, *Fusobacterium* and *Enterobacter asburiae* have been reported being associated with pancreatic and breast cancer. (Cavarretta et al., 2017; Cullin et al., 2021; Nejman et al., 2020.).

1.2.4 Microbial mechanisms promoting cancer

The mechanism by which intratumoral microbiome and gut microbiome affect tumor behavior is not yet fully understood. One of the mechanisms suggested by recent studies is through the secretion of exosomes. Exosomes are 40-100 nm vesicles including various proteins lipids and RNAs and they regulate intercellular communication between different cell types. Exosomes derived from tumors can transfer miRNA and proteins promoting tumor metastasis in normal tissues through various mechanisms. According to Cullin et al. (2021) cancer-associated microbes could potentially reach the tumor environment by peri-intestinal translocation through the pancreatic duct. Factors such as dysbiosis and epithelial damage on the gut wall could also help microbes to travel.

In colorectal cancer, Bertocchi et al. (2021), reported that *E. coli* migrated to the liver after damage in gut barrier, promoting inflammatory response and metastasis. The contact-dependent interactions suggested by Cullin et al included direct bacteria-host cell interactions as well as contact independent interactions. Species such as H. pylori via CagA- E-cadherin/β-catenin association, *F. nucleatum* with the E-cadherin-FadA interaction and *S. enterica* by upregulation of β-catenin pathways a few of the examples by which pathogens have been reported to be able to promote tumor progression. *Escherichia*, *Helicobacter* and *Salmonella* have been suggested to be able to bind to a certain kind of receptor which would ultimately lead to inhibition of cellular proliferation and DNA damage. Contact independent mechanism affects through the production of bioactive molecules that can travel to the target tumor via the systemic circulation (Fig. 2). According to Rossi et al. (2020), the most common toxins are lipopolysaccharides

(LPS) and acetate that can promote carcinogenesis by upregulating EMT and angiogenesis. Host derived metabolites such as the secondary bile acids (sBAs), deoxycholic acid (DCA) and litholic acid (LCA) were reported to be associated with CRC and hepatocellular carcinoma. Moreover, hydrogen sulfide, a toxic compound produced by microbes, was found to be elevated in CRC biopsies compared to healthy tissues. (Bertocchi et al. 2021; Blachier et al. 2021; Cullin et al. 2021; Rossi et al. 2020.)

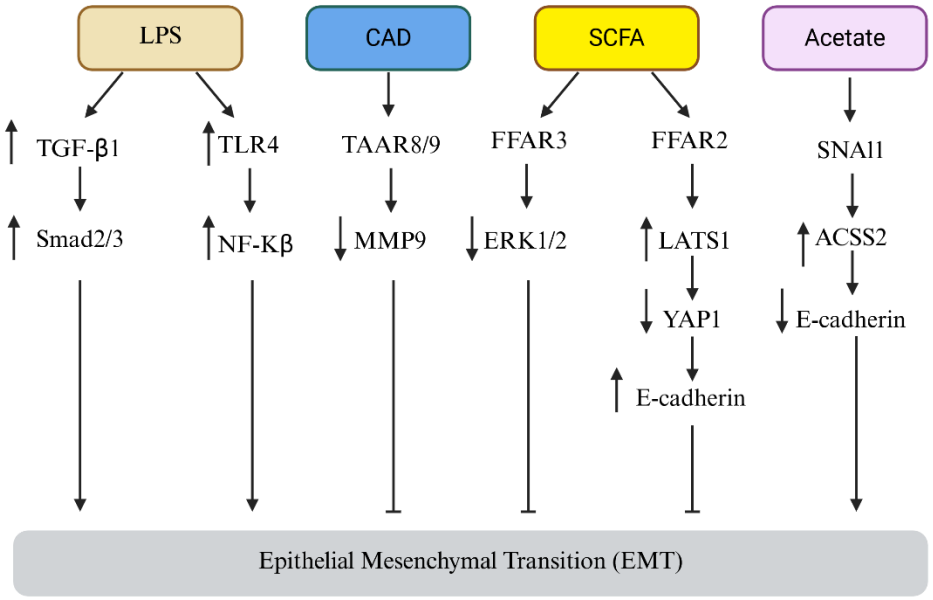

Figure 2. Metabolic pathways involved in Epithelial-Mesenchymal transition (EMT) in cancer. Lipopolysaccharide (LPS) has EMT promoting effect through upregulation of transforming growth factor beta-1 (TGFβ-1) and Mothers Against Decapentaplegic Homolog 2/3 (Smad2/3) or through increasing expression of Toll Like Receptor 4 (TLR4) and NF-κB. Cadaverine (CAD) might inhibit EMT in breast cancer cells by activating trace amino acid receptors 8 and 9 (TAAR8/9) that modulate the expression of metalloproteinase9 (MMP9). Short chain fatty acids (SCFAs) can also inhibit EMT by activating Free Fatty Acid Receptor 2 (FFAR2), which leads to inhibition of Hippo-Yap pathway and increases expression of E-cadherin and FFAR2. Eventually this results in mitogen-activated protein kinase (MAPK) signaling inhibition. Acetate can promote EMT by increasing Snail Family Transcriptional Repressor 1 (SNA1) and Acyl-CoA Synthetase Short chain Family Member 2 (ACSS2) in renal carcinoma cells. Rossi et al (2020), borrowed and edited.

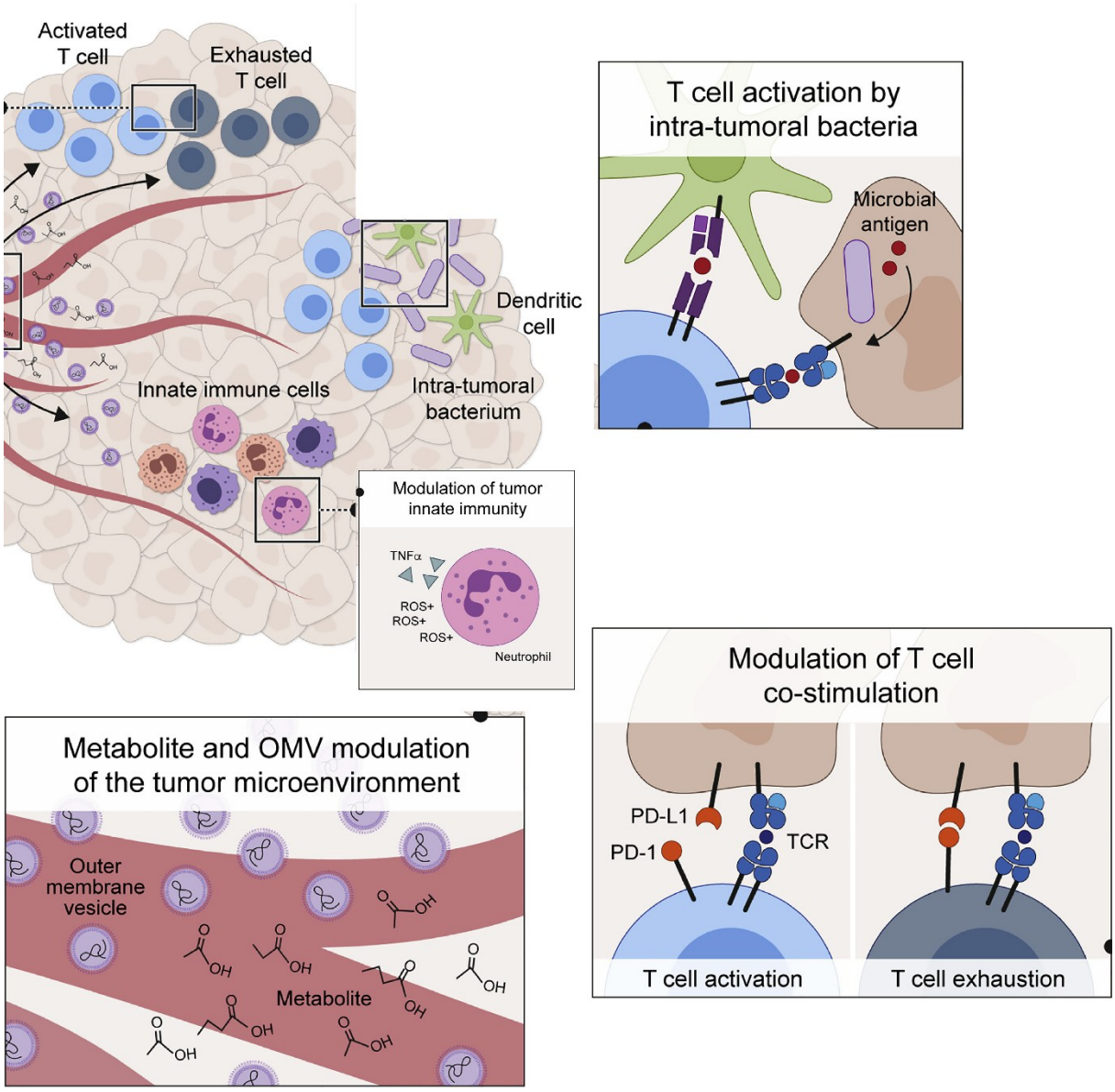

Figure 3. Interaction between tumor immune microenvironment and microbiome. Bacteria release small molecules and outer membrane vesicles (OMVs) that can attract immune cells like neutrophils, stimulating them to release chemicals like tumor necrosis factor alpha (TNF-α) and reactive oxygen species (ROS) which help to fight against cancer cells. Bacterial metabolites and OMVs might also modulate T cell co-stimulation impacting TME T cell activation and exhaustion. Lastly, bacterial antigens might be presented to T cells which helps to activate them. Cullin et al., (2021), borrowed and edited.

1.3 Microbiota in cancer diagnostics

One of common targets for microbial diagnostics is based on gut microbiome. Typically, the pathotype level of specific taxa, species or strain is investigated and applied to diagnostics. Almost for a century, culture-based methods, have been popular in clinical microbiology.

However, it is time consuming, and anaerobes are difficult to culture with basic techniques. A common way to surveillance antibiotic resistant pathogens is to culture the microbiome of a hospitalized patient and screen for resistant bacteria like *Enterococcus spp*. (Damhorst et al., 2021.) According to Sepich-Poore et al (2020), determining microbial DNA and RNA signature can be detected in blood samples. They reported that stringent filtering criteria in over 10,000 screened patients were able to identify microbial plasma signature, used for cancer type prediction, could be used to specify healthy tissue profiles (Sepich-Poore et al. 2024). Moreover, strongly cancer associated species such as *F. nucleatum* that have been linked with various cancer types like CRC, PCa and breast cancer (BC), are strong indicators for tumor metastases. According to Parhi et al (2020) *F. nucleatum* was noticed to upregulate Gal-GalNAc, a cell binding tumor upregulator in breast cancer. (Cullin et al., 2021; Parhi et al., 2020; Sepich-Poore et al., 2024; Xu et al., 2021.) Another diagnostic practice applied on microbiome is identifying taxa or metabolic patterns as biomarkers with metagenomic sequencing. According to Wirbel et al (2019), a meta-analysis of four geographically different cohorts identified 4 clusters of 29 bacteria species associated with colorectal cancer and tumor location. Additionally, three virulence factor coding genes were associated with colorectal cancer. Other potential targets for diagnostics include transcriptome, proteomics and metabolomics analyses, all of which have the capability to determine diagnostic biomarkers. For instance, transcriptomics can reveal gene expression of active organisms and their pathways. Metabolomics on the other hand identify metabolites produced by known pathogens like *Clostridioides difficile* causing diarrhea. (Damhorst et al., 2021; Wirbel et al., 2019.)

1.4 Analyzing microbiome

1.4.1 Next generation sequencing

Next generation sequencing (NGS) of the 16S rRNA gene and the whole DNA in Shotgun Metagenomic Sequencing are most often used for determining microbiome composition, however, both methods have known experimental and computational challenges. In both, a critical measure is the acquired sequencing depth which is strongly affected by the amount of host reads (Bharti and Grimm 2021).

A common way to determine microbiome composition of the target environment is to extract the bacterial DNA and RNA through biopsies or fecal, swipe or environmental samples.

However, studies have reported variation between extraction methods. Typically, low-abundance bacteria are left undetected because of the high concentrations of human DNA in tissue samples. For certain sample types like stool samples, which naturally have a small amount of host DNA (<0,5%) regular NGS works fluently. Samples like skin swabs or respiratory and tissue samples on the other hand hold a greater amount of host DNA, which makes it harder for NGS methods to separate host DNA from microbial. Moreover, the cost of Shotgun Metagenomic sequencing rises as the depth increases, and even then, it is not guaranteed that Shotgun Metagenomic sequencing would detect low-abundance specimens from sample types that hold >99% host DNA. (Kim et al. 2024.)

Removing human background DNA (host depletion) before the high-throughput sequencing phase, shifts the microbial composition (Kim et al., 2024a). Furthermore, metagenomic sequencing without host depletion underestimates the diversity of microbes in the sample. Nevertheless, in their study, Kim et al showed also that host depletion decreased not only human DNA concentration but also bacterial DNA, especially with gram-negative bacteria, that are potentially more vulnerable to host depletion lysis after freezing.

Microbial-enrichment methods (MEM) described by Wu-Woods et al. (2023), have been reported to be promising, giving more than 1000-fold removal of host DNA from solid mammalian tissue, while preserving most of the microbial composition. MEM is similar to commercial host-depletion kits like MolYsis, QIAamp and lyPMA, all of which include two main steps: selective lysis followed by nucleic-acid removal. However, MEM uses bead-beating procedure with larger beads (1,4 mm), for targeted host cell lysis. To ease the pre-treatment phase, studies usually apply bead-beating to the protocol to achieve well homogenized lysate before removing host DNA from low-abundance microbial samples. Enzymatic protocol is performed with Benzonase and Proteinase K to degrade host DNA. According to Wu-Woods et al, MEM enabled detection of low-abundance microbial taxa, pathways and genes at relative abundances low as 10^{-10}, respectively. (Bharti & Grimm, 2021; Kim et al., 2024a; Wu-Woods et al., 2023)

1.4.2 16S RNA sequencing

Next generation sequencing (NGS) of the 16S rRNA gene is one of the most used methods for determining microbial composition. 16S rRNA sequencing is based on the amplification of 16S rRNA gene region which codes for the bacterial 16S subunit. The 16S rRNA gene is

approximately 1,500 base pairs long and consists of highly conserved regions interspersed with nine hypervariable regions (V1–V9). The conserved regions can be targeted by specific primers, enabling the detection of a broad range of bacterial taxa that share these sequences. However, differences between the conserved regions and differences in the annealing capability of different primers can result in an unequal amplification of bacteria. (Abellan-Schneyder et al. 2021.) Typically, the sequencing of the V3–V4 region of the rRNA gene has been reported to be most accurate and effective. However, 16S rRNA sequencing come with distinct flaws such as amplification bias, chimera generation and variation in amplified regions of the 16S rRNA gene. 16S sequencing typically identifies bacteria on all taxonomic levels but the accuracy fells when it comes to species or strain level. Moreover, reference data interpretation, software parameters and data curation cause challenges for repeatability across studies. (Austin & Korem, 2024; Cullin et al., 2021; Schloss, 2018.) Library preparation steps for 16S V3–V4 sequencing include usually (1) genomic DNA extraction (gDNA), (2) normalization of the gDNA, (3) 16S rRNA gene amplification PCR, (4) PCR-product clean-up and quantification and (5) normalization of the PCR products and pooling" (Fig. 4). The steps are not conserved, and the changes can affect reproducibility and sequencing outcomes. However, the diversity and microbiome compositions are generally unaffected by protocol streamlining in gut commensal analysis. (Celis et al. 2022.)

Sequencing can be performed a long read sequencing or short-amplicon sequencing. The advantage of sequencing the whole gene, with third-generation sequencers such as Oxford Nanopore MinION is the long read length and improved identification of bacterial taxa. Long read sequencing comes with higher error rate (up to 15% per sequence), and limited capability in high-throughput studies as well as higher costs. On the other hand, short-amplicon sequencing, targeting certain regions like V1-V2 (Fig. 5), V3, or V3–V4 /V5 and V4 which are the most common ones, has the potential to identify a larger number of bacteria. However, taxonomic classification may differ significantly between targeted regions and cross-study comparison will not necessarily be possible if the protocols differ too much. Nevertheless, a large portion of studies use short-amplicon sequencing that produces short sequences (≤ 300 bp). Overall, 16S rRNA gene amplicon sequencing comes with less effort compared to Shotgun metagenomic sequencing as well as with lower costs. (Abellan-Schneyder et al., 2021; Johnson et al., 2019.)

Typically, after sequencing the library, data is clustered with different reference databases such as Silva and GreenGenes2. Abellan-Schneyder et al. (2021) found that not only trimming of amplicons is crucial and different truncated-length combinations should be tested for each study, but they also noticed that certain specific but important taxa are not picked up by certain databases like GreenGenes. (Johnson et al., 2019.)

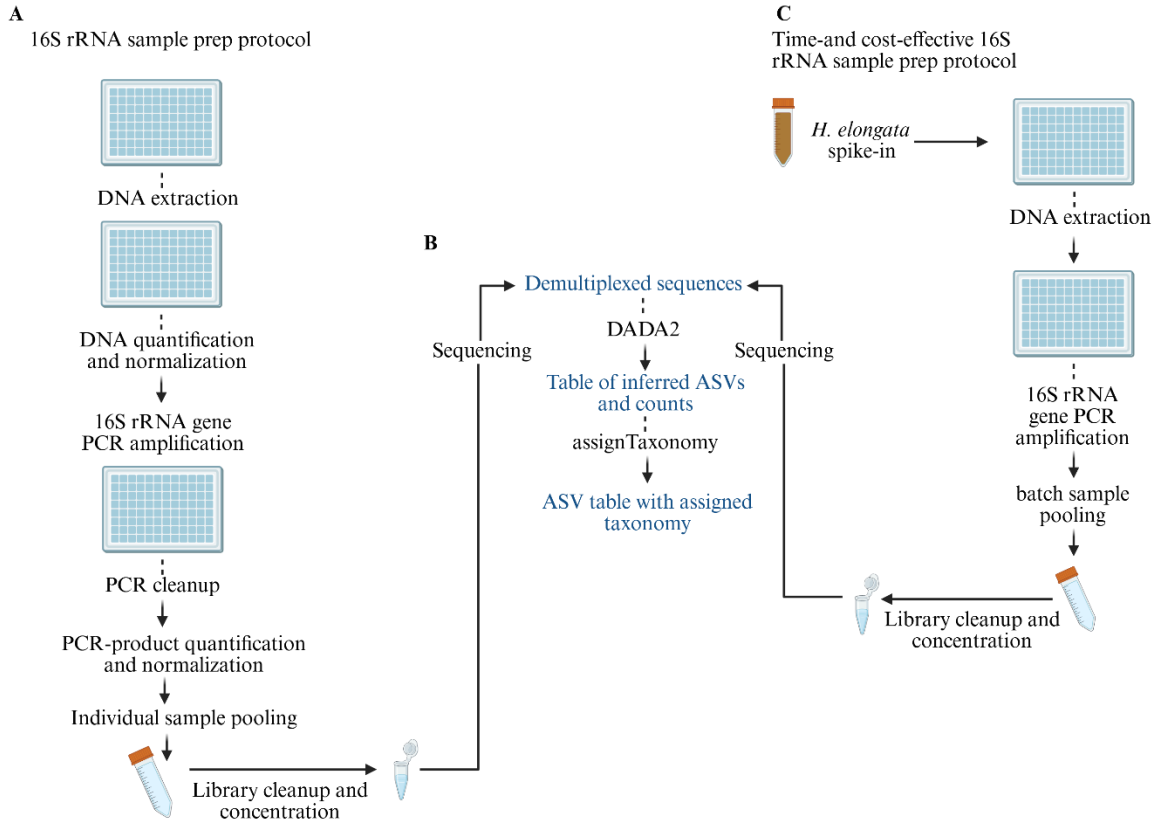

Figure. 4. 16S rRNA sequencing protocol. **A)** Most common protocol for preparation of 16S rRNA gene library. **B)** Pipeline using DADA2 (Callahan et al., 2016). **C)** Optimized protocol reducing costs/time, without losing accuracy and reproducibility. Celis et al., (2022), borrowed and edited (Biorender).

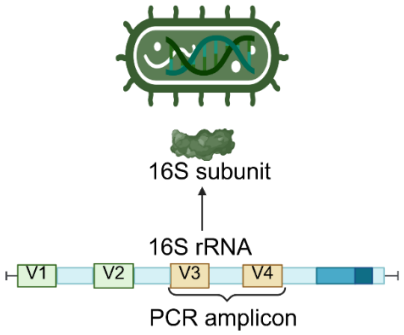

Figure 5. The basic principle of 16S V3–V4 rRNA sequencing.

1.4.3 Shotgun metagenomic sequencing

Shotgun metagenome sequencing allows both mapping of taxonomic composition of a sample as well as evaluation of genes that can potentially encode for metabolic pathways. SMS is based on the random DNA fragmentation of the sample which are sequenced using high-throughput platform, like Illumina. Overlapping sequences are reconstructed and identified by an appropriate computational tool, such as CLC Metagenomic Workbench or Kraken pipeline. SMS has the capability to provide more accurate information about the species- and strain compositions compared to 16S rRNA sequencing. SMS gives information not only about taxonomic compositions but also about functional bacterial genes and part of genomes which cannot be achieved by 16S rRNA sequencing. However, SMS has also several disadvantages as it typically requires more sequencing depth to cover metagenomics and this requires more data space, raising the costs of each sequencing cassette. Generally, in microbiome research, SMS has been reported to have significantly higher resolution, especially detecting low abundant species (>500,000 reads). However, several studies have shown that relative species abundance distribution has been similar between SMS and 16S sequencing samples with low microbial load (Durazzi et al., 2021). Thus, in cases in which microbial load is very low, SMS could perform even worse than 16S sequencing. Furthermore, the bacterial taxonomic accuracy at the genus level would be the same compared to 16S V4 sequencing (Usyk et al., 2023).

1.4.4 Contamination and low bacterial biomass analysis

Low biomass microbiome sample typically covers biopsies, tissue swabs and lavages. So far, 16S sequencing has been the most popular for bacterial DNA analysis from these samples. However, till today, no current method has been standardized for the analysis of low biomass samples. Thus, studies must have relied on less standardized methods containing mainly variations of the 16S sequencing of rRNA gene. According to current knowledge, the sample biomass is the most influential factor for anticipating the representativeness of microbiome composition. (Eisenhofer et al., 2019.) Alpha diversity and species richness tend to increase with mechanical lysing as well as with an increase in biomass. Furthermore, samples of which bacterial densities were below 10^6, resulted in a loss of sample identity, based on cluster analysis tested for appropriate protocols. (Kennedy et al., 2023; Villette et al., 2021)

Due to high sensitivity, NGS methods can detect contaminant DNA and cross contamination, which is a common problem in microbiome research. Contaminant DNA can be derived from

various sources from obtaining the sample, preparation and laboratory environment as well as from researchers, tools, kits and reagents. On the other hand, cross-contamination can also take place in various stages of the sample processing and treatment. (Austin & Korem, 2024.) Most common problem being the transfer of sample DNA from one sample to another (well-to-well leakage) but also a phenomenon called tag-switching, in which adjacent barcodes hop into sample wells or tubes. Lastly, cross-contamination can take place in through index-hopping, where certain sequencing platforms mismatch indexing reads to sequencing reads. (Eisenhofer et al. 2019.) Moreover, the effect of contamination is emphasized especially in low bacterial biomass analysis, in which the number of microorganisms is naturally very low. Low-biomass samples are very susceptible to overamplification bias, as PCR cycle numbers increase, contaminating microbes can become over-represented. This not only makes the determination of the microbial composition from the background and contamination harder but also drawing the line between background and sample compositions even more biased. Thus, the role of contamination is often downplayed by microbiome studies, as well as the usage of appropriate negative controls and reporting negative results. Furthermore, the most difficult thing in microbiome research is to prove something to be non-existent, especially in the case of organs which are supposed to be sterile. (Knight et al., 2018.)

1.4.5 Bioinformatics

Advances in DNA sequencing technologies have made it possible to generate large amounts of data, which must be analyzed accurately and efficiently. Nowadays, many computational tools, methods, and standards are available for this purpose. The two most common sequencing approaches—16S rRNA sequencing and shotgun metagenomic sequencing (SMS)—differ in the amount and type of data they produce. 16S V3–V4 sequencing focuses on a specific region of a single gene, resulting in a smaller and more manageable dataset. In contrast, SMS sequences entire genomes, producing much larger and more complex datasets. Because of these differences, the data analysis workflows for each method vary, but both can be used to study the taxonomy, function, and evolution of microbial communities.

Marker gene analysis is common for 16S rRNA and shotgun metagenomic sequencing (Fig. 6). The sequenced data is obtained as reads, which are either unassembled DNA or mRNA sequences. The first step involves removing sequencing errors and poor-quality reads by trimming a certain amount from the sequence and removing poor quality reads. Typically, this is followed by clustering similar sequences into operational taxonomic units (OTUs), usually

with 97% similarity. OTU merges sequence variants including ones with sequence error and this can be used for determining abundance profile. The downside of OTU clustering is that it fails to recognize real biological sequence variation like single nucleotide polymorphisms (SNPs) that would otherwise be determined into single OTUs. According to Knight et al. (2018), a method called oligotyping improves traditional OTU picking by looking for small and specific differences in bacterial DNA. Thus, it can distinguish bacteria that are almost identical. After identification, taxonomic names are assigned to microbial sequences in the data. Databases like Greengenes, Silva or RDP are commonly used as their specificity should be fairly accurate. However, the databases lack sensitivity as the number of unknown taxa is often large (Knight et al., 2018). According to Knight et al, exact sequence variants (ESV) would offer higher resolution and more reproducibility across datasets compared to OTUs and amplicon sequence variants (ASV). ASV is based on exact sequence recognition by denoising sequencing data. On the downside, ASV is computationally a bit more demanding, and it requires high-quality reads to operate (commonly Illumina platform). Databases for ASV analysis include for instance Greengenes, DADA2 and QIIME 2.

Typically, after OTU clustering, microbial alpha and beta diversities are investigated. Alpha diversity determines the species amount and diversity within the individual samples, beta diversity on the other hand compares dissimilarities between each pair of samples. There are several methods to determine biodiversity. For determining alpha diversity, Shannon and Simpson's diversity index are common ways to further analyze the species' richness and evenness as well as the probability that two individuals randomly belong to different species. Chao1 estimates true species diversity and can give more accurate information species richness if sequencing wasn't done very deep. Different measures like Bray-Curtis, weighted UniFrac or Jaccard can be tested for beta diversity analysis of the microbial composition. Bray-Curtis compares the community structure of the species and gives information about dominance patterns, which species is over-represented. Weighted UniFrac on the other hand grants information about the evolutionary differences between closely related species. Lastly, Jaccard investigates presence/absence of species. (Knight et al., 2018.)

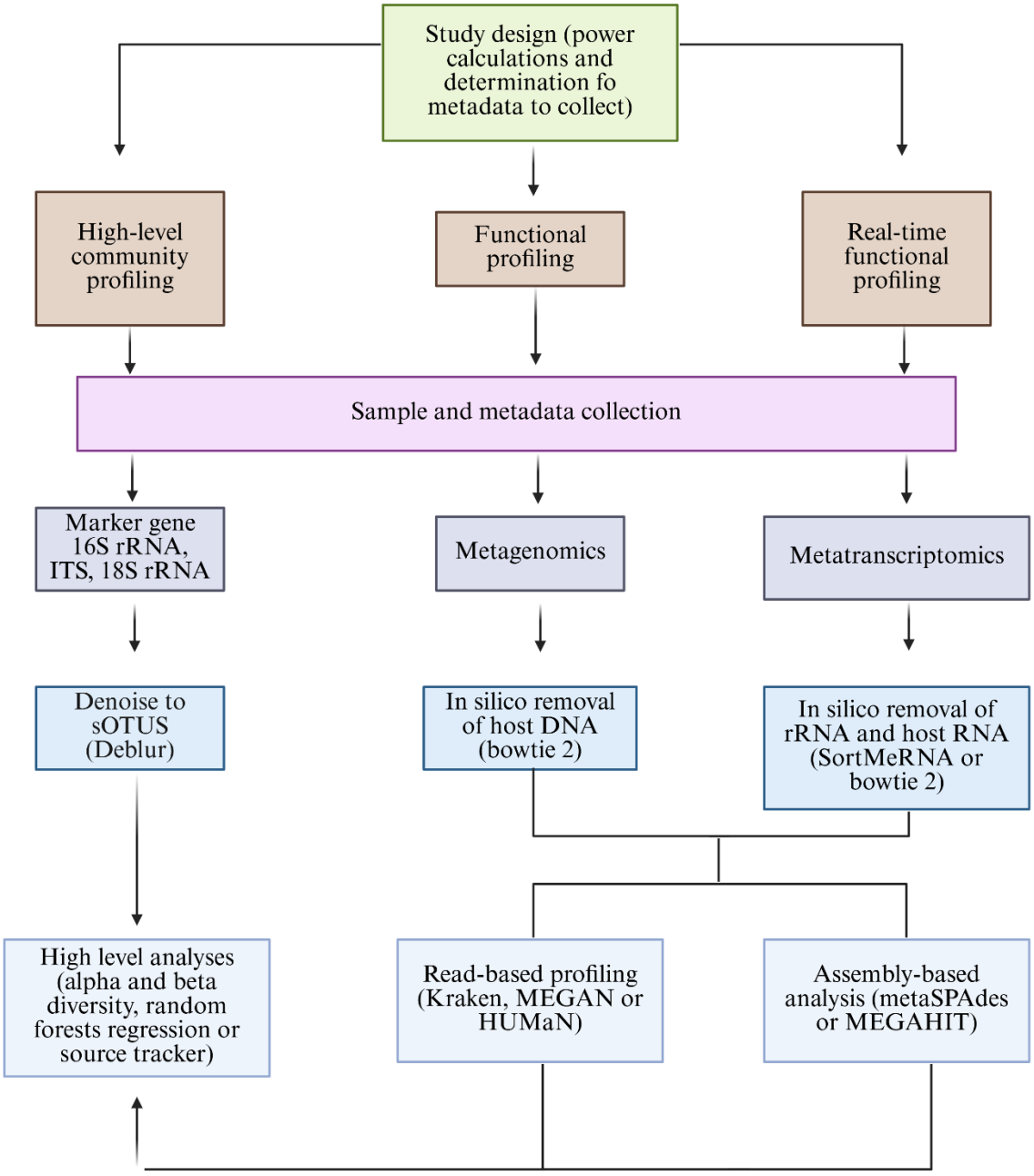

Figure 6. Optimal workflow for metagenomic and metatranscriptomic 16S ribosomal RNA sequencing. Deblur is recommended for 16s RNA sequencing to identify specific DNA sequences (sOTUs). This is considered faster and more consistent compared to similar tools, like DADA2. After cleaning metagenomic and meta transcriptomic data from host DNA and RNA, analysis tools like Kraken, MEGAN or HUMAnN or through genome assembly tools like metaSPAdes and MEGAHIT are recommended. After data processing, more specific analyses like diversity analyses, taxonomy profiling or machine learning can be used to investigate patterns. Random forest regression can be used to estimate time since death or microbiome development. SourceTracke which is a Bayesian tool, helps to determine where microbial communities come from based on their environment. Knight et al., (2018), borrowed and edited (Biorender).

1.5 Aims

The study aims to optimize the DNA isolation method for prostate biopsies to determine the prostate microbiome. Several studies have shown that the prostate would harness a unique microbiome, however, the number of reads tends to be always low as well as the use of negative controls is limited. We aimed to investigate if removing host background DNA before 16S V3–V4 and Shotgun Metagenomic sequencing altered the microbial profiles of the prostate tissue samples. Furthermore, we looked for parameters that could affect and improve the prostate tissue extraction protocol, as there is no well-established methodology yet. Lastly, we aimed to look for similarities between intratumoral prostate tissue and gut microbiome.

2 Materials and methods

2.1 Optimization of DNA extraction from prostate tissue.

In this study, we investigated parameters that might affect DNA extraction from prostate tissue. We evaluated the effect of host depletion, tissue weight, DNA concentration on microbial profiles of prostate as well as bioinformatics parameters related to 16S V3–V4 sequencing (Fig. 7). We investigated the optimal time for bead-beating as well as the optimal amount of reagents for pre-treatment.

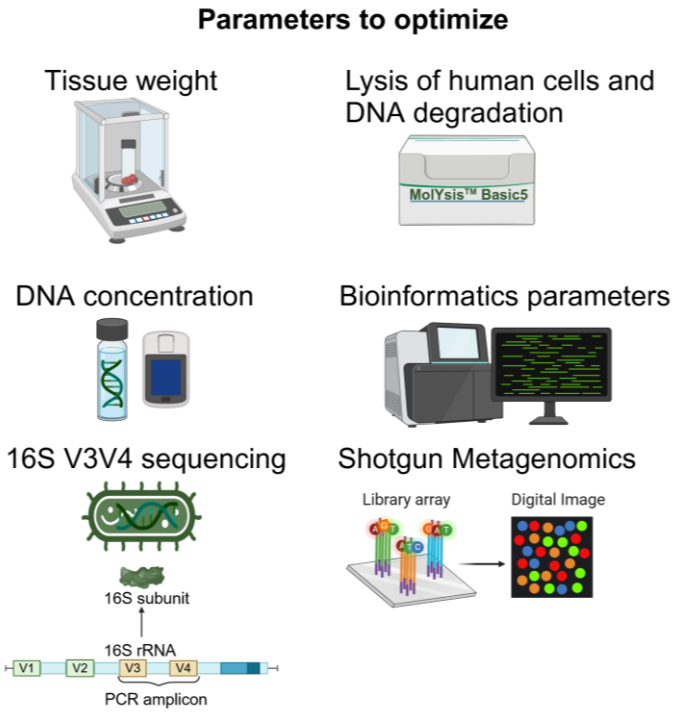

Figure 7. Optimization of DNA extraction from prostate tissue. (Biorender)

2.2 Study design and patients

The optimization of DNA extraction is part of a larger study, prostate cancer, and gut microbiota (Promic), in which the prostate tissue samples, called cores, are collected by Turku University Hospital's Pathology unit, after surgery. Histological Core Services (Histocore), imaged the cores and handed them to Turku Prostate Cancer Consortium (TPCC), which weighed and stored them at -80 °C. In this study, we had 17 prostate tissue samples from 15 different patients (Fig. 8.). The samples were obtained from three distinct types of surgeries, transurethral resection of the prostate, robotic assisted laparoscopic prostatectomy, and cystectomy. Tissue

cores have been annotated based on their representativeness of the whole cancer of the removed prostate. Moreover, piloted cores were used to optimize the DNA extraction from prostate tissue for metagenomic sequencing.

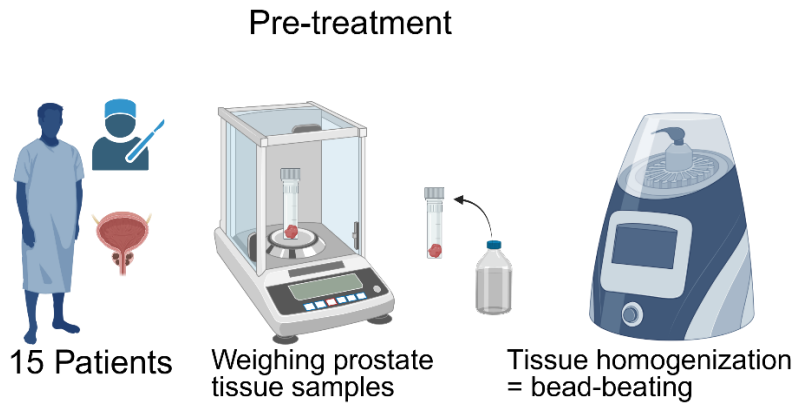

Figure 8. Sample preparation for DNA extraction. (Biorender**)**

2.3 Prostate tissue annotation.

The cancer profile of the prostate tissue cores was determined using Philips digital pathology image management system. Prostate tissue cores were annotated by analyzing three different levels of tissue sectioning from each patient's prostate tissue core (Fig 9.). Cases in which carcinomas were above 30% of the core surface were chosen to represent malign prostate tissue. For each malign core, a benign equivalent was determined. The annotations included the following information: malign/benign status, ISUP Grade group, tertiary Gleason, whole prostate cancer (%), Gleason (4&5), margin status, healthy lymph nodes/metastasized, pT-class, weight and inflammation status ranging from 1 to 3.

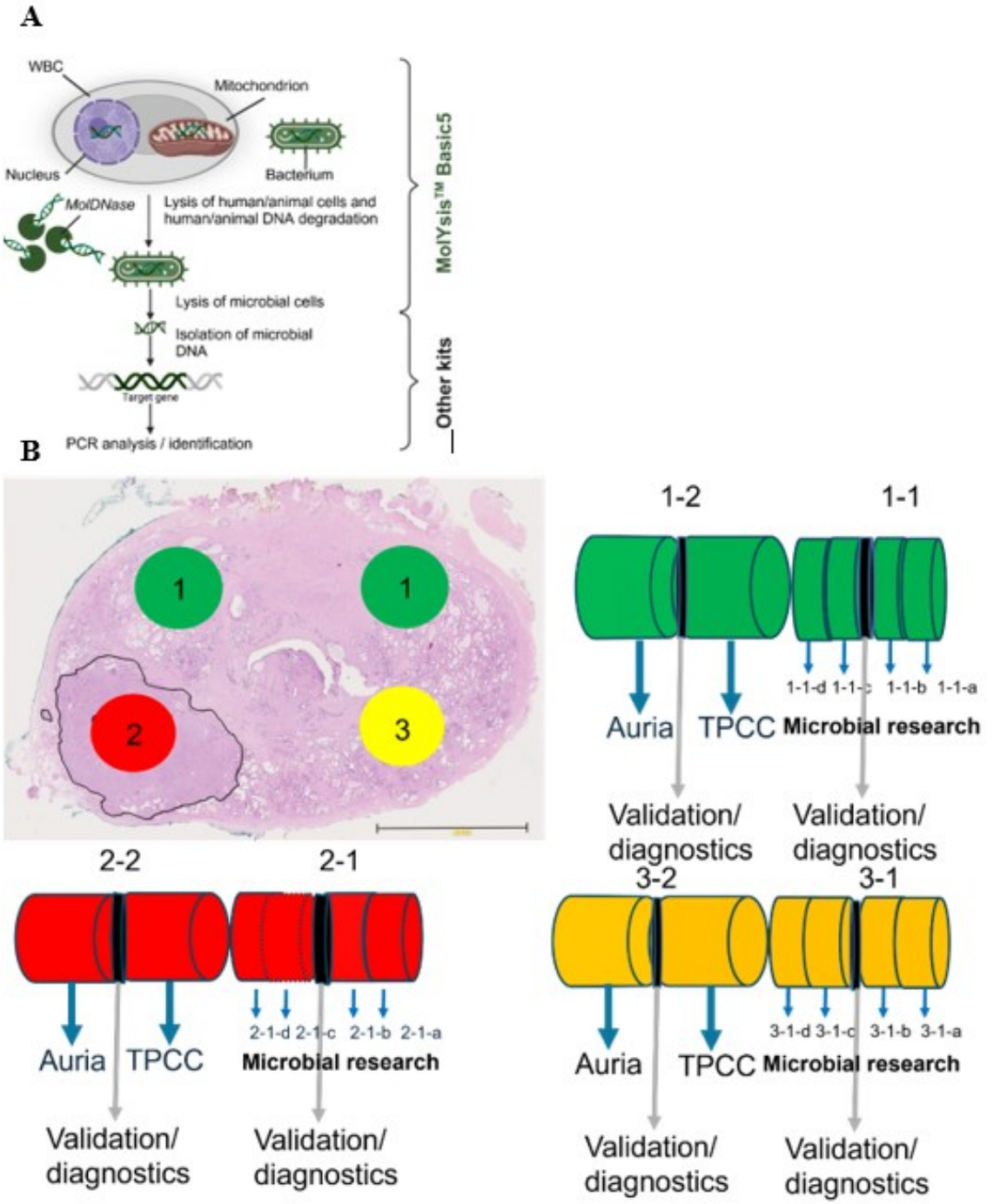

Figure 9. **A)** MolYsis host depletion technology (Biorender). **B)** Prostate tissue core mapping.

2.4 Sample collection and processing.

As a first step, all tissue samples—including those intended for total DNA extraction and those undergoing host DNA depletion—were weighed and mechanically homogenized (bead-beating) using 2mm ZR BashingBead™ Tubes (Zymo Research Corp., Irvine, CA, USA) to ensure uniform sample processing. 800 ul of SU buffer was added up to 1ml before samples were homogenized for 180s, 2500 rpm. Samples were vortexed lightly shortly after and all the

liquid was moved to two tubes with the total volume divided 1:1. Samples not meant for host depletion were frozen -20 °C. For host DNA depletion, the MolYsis™ Basic5 kit (Molzym GmbH & Co. KG, Bremen, Germany) was subsequently used with minor adjustments to manufacturer's instructions. This kit enriches microbial DNA by selectively lysing human/animal cells and degrading host DNA, while preserving bacterial and fungal DNA. This pretreatment enhances the accuracy and reliability of downstream molecular analyses by reducing host background. Bacteria were then centrifuged and treated by BugLysis reagent for the degradation of cell walls. A spike-in control using *Enterococcus faecalis* (ATCC 29212) at a known bacterial cell concentration (11.7×10^8), preserved in glycerol, was added to one of the samples (RALP-1433) prior to processing. Following completion of the protocol, all remaining samples were stored at -20 °C.

2.5 DNA extraction

DNA extraction was done using Chemagic ™ DNA Stool 200 mg Kit H96 (Revvity, Inc., Waltham, MA, USA) with Magnetic Separation Module I (MSM I) extraction robot (PerkinElmer) and performed according to the manufacturer's instructions. In addition to positive spike-in control (*E. Faecalis*) extraction also included negative controls (OMNIgene fluid, DNA/RNA Shield fluid) and PCR grade water as well as extraction controls (Chemagic Lysis Buffer 1) and ZymoBIOMICS Gut (Zymo Research, USA).

2.6 16S metagenomic sequencing library

The microbial composition of the prostate tissue samples was determined by targeting bacterial V3–V4 hypervariable region of the 16S rRNA gene. The sequencing libraries were prepared according to Illumina library preparation protocol (https://support.illumina.com /content/dam/illumina support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf) using the KAPA HiFi HotStart ReadyMix kit (KAPA biosystems, Roche, Basel, Switzerland). Protocol was modified by increasing DNA template amount in the amplicon PCR reaction from standard 12.5 ng to 75 ng. In addition, to achieve an adequate amount for library quality monitoring, the total volume of the PCR reaction was increased from 25 ul to 33 ul. Sequences of the V3–V4 gene specific full length forward and reverse primers were 5' TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CCT ACG GGN GGC WGC AG-3' and-5'GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGAC TAC HVG GGT ATC TAA TCC-3' (Klindworth et al., 2013). Control

samples covering negative DNA extraction control, negative PCR control and positive mock community (ZymoBiomics microbial community DNA standard, Zymo Research) were included in every run. Sample specific Nextera XT indices were added in the second PCR using Nextera XT Kit (Illumina, San Diego, California, USA) which resulted in approximately 630 bp PCR product. The products were purified with AMPure XP Magnetic beads (Beckman Coulter, Inc., Brea, CA, USA) along with DynaMagTM-96 magnetic plate (Life Technologies). Moreover, the DNA concentrations of the final libraries were evaluated with Qubit fluorometer (Life Technologies) and dsDNA HS Assay Kit (Thermo Fisher Scientific). This was followed by equimolar pooling (4nM), denaturation and dilution to 4 pM concentration. For sequencing control and to increase the nucleotide diversity 1 % PhiX (Illumina) was added to the pool. The sequencing was performed using MiSeq Reagent Kit v3 and 2 x 300 bp paired end run on Miseq System (Illumina). Amplicon PCR was performed under following conditions; initial denaturation at 95 C for 3 min, followed by 30 cycles consisting of denaturation at 95 C for 30s, annealing at 55 C for 30s and extension at 72 C for 30s, and with a final extension at 72 C for 5 min. After PCR, 8 µl of the product was analyzed with 1,5% TAE agarose gel (100V, 60 min). PCR products were purified with in-house purification protocol (appendix x) with AMPure XP magnetic beads (Becman Coulter, USA).

2.7 Shotgun Metagenomics sequencing library

Shotgun Metagenomics sequencing library was prepared using the Nextera XT DNA Library Prep Kit according to manufacturer's instructions. The prostate tissue samples were amplified in 22 ul reactions containing Nextera PCR Master Mix, 1 uM of each primer, 5 µl template gDNA, 5 µl Amplicon Tagment mix and 10 µl Tagment DNA buffer. Thermal profile included an initial 72 °C × 180 s, 95 °C × 30 s denaturation, followed by 12 cycles of denaturation of 95 °C × 10 s, annealing at 55 °C × 30 and extension at 72 °C × 30 s. Plus final extension at 72 °C × 5 min. Amplification was confirmed by running 5 µl of PCR product on a 1,5% TAE agarose gel, by visualization of ≈ 550 bp band. Sample specific Nextera XT indices were added in the second PCR using Nextera XT Kit (Illumina, San Diego, California, USA) which resulted in approximately 630 bp PCR product. The DNA concentrations of libraries were measured using Qubit fluorometer (Invitrogen) applying High sensitivity assay and sample pooling was done at standardized concentration. The pooled library was sequenced on the Illumina MiSeq platform (Illumina, California, USA) utilizing 2 × 300 bp chemistry. Samples were then analyzed with Bioanalyzer and sequenced with Nextseq.

2.8 Statistical analysis and data visualization

Sequencing reads obtained from 16S rRNA gene V3–V4 region amplicons were analyzed using CLC Genomics Workbench and CLC Microbial Genomics Module (QIAGEN, Aarhus, Denmark) as well as the Kraken2/Bracken pipeline. Initial quality control and adapter trimming were performed using the workflow “Data QC and OTU Clustering” in CLC, with default settings unless otherwise stated. Trimming parameters included a minimum Phred quality score of 20 and minimum read length of 150 bp. OTU clustering was carried out at 97% similarity using both the SILVA 138 and Greengenes2 reference databases. Taxonomic classification and relative abundance profiles were generated at the genus and species levels. Decontamination was performed by excluding all bacterial taxa detected in negative control samples. In addition, ASV analysis was conducted using the “ASV Abundance Table with Greengenes2” workflow, following default settings. Alpha diversity was assessed using total diversity tool. Beta diversity was not evaluated and is therefore not reported. Statistical analysis and further visualization of taxonomic and diversity data were performed using RStudio (R version 2024.12.1) and built-in tools within the CLC Microbial Genomics Module.

3 Results

In this study, we had prostate tissue samples from 15 patients. The samples were obtained from three distinct types of surgery techniques, transurethral resection of the prostate, robotic assisted laparoscopic prostatectomy, and cystectomy. The frozen tissue samples were weighed, homogenized, and divided into two sub-sample sets. Host depletion (HD) samples were treated with the the MolYsis™ Basic5 kit based on selective lysis, and the DNA extraction (DNA Ex) sample set went straight to further processing. These two sample groups were compared using 16S V3–V4 or Shotgun metagenomics sequencing for microbiome analysis.

3.1 DNA yield

The average tissue sample weight was 97.4 mg, the host depletion (HD) samples had an average concentration of <0.05 ng/µl, and the DNA extraction (DNA Ex) samples had an average concentration of 30.1 ng/µl. We found no significant correlation between tissue weights and DNA concentrations. After applying host depletion kit, in most samples the amount of DNA was below the Qubit detection range (<0,05). Due to low concentrations of DNA in HD samples, we speculate that the prostate would likely harness very little number of bacteria or the host depletion kit had just removed most of the bacterial DNA as well. For DNA extraction samples the amounts were significantly higher, due to human DNA. In two (2/17) of our samples, the tissue weight was not weighed and in four (4/17) of our samples the concentration was not recorded due to technical difficulties (Table 1.).

Table 1. Prostate tissue weight did not correlate with DNA concentration. Green = high DNA, Red = low DNA. HD = Host depletion, DNA Ex = DNA extraction, C = Cystectomy, R = Robotic assisted radical prostatectomy, T = Transurethral resection of the prostate, BE = Benign, MA = Malign

ID	Tissue weight (mg)	HD (ng/µl)	DNA Ex (ng/µl)
1C	43	<0,05	7,52
1T	46	0,11	42,6
1R MA	78	<0,05	22,8
2T	105	<0,05	36,7
2R MA	35	<0,05	40,7
2R BE	27	<0,05	14,6
3R BE	80	<0,05	19,9
4R	264	2,08	41,7
5R	69	<0,05	N/A
6R	72	0,163	33,7
7R	N/A	<0,05	25,1
8R	119	1,97	N/A
9R	51	<0,05	40,4
10R	220	0,068	N/A
11R	83	<0,05	52,7
12R	169	<0,05	N/A
13R	N/A	<0,05	22,8

3.2 16S V3–V4 sequencing

3.2.1 Amplicon PCR

Amplicon PCR for 16S V3–V4 sequencing a few visible bands. PCR was performed to validate samples going for 16S V3–V4 analysis. The size of the amplicon was 550 kb, and an appropriate positive spike-in *Enterococcus faecalis* control and negative controls were applied for host depletion and DNA extraction samples.

Besides positive *E. Faecalis* control, 1R MA HD and 1R MA DNA Ex gave visible bands (Fig. 11A). The rest of the samples were scarce in DNA and no bands were recognized. However, our negative HD and DNA Ex controls as well as positive *E. faecalis* control confirmed that removal of human DNA with the MolYsis kit had worked accordingly (Fig. 11B).

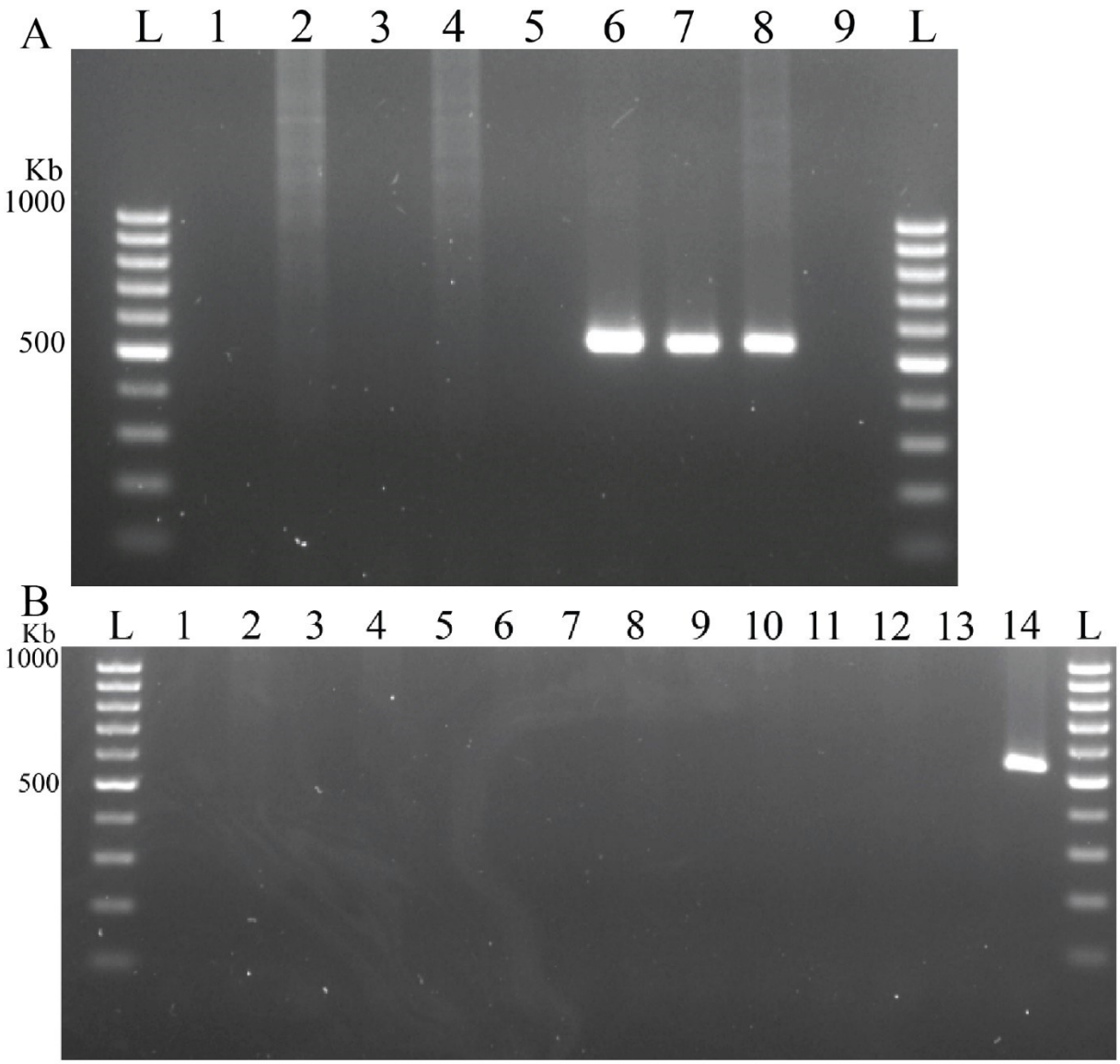

Fig. 11. Amplicon PCR for 16S V3–V4 sequencing. The size of the amplicon is 550 kb. **A)** 1,3,7 treated with MolYsis™ Basic5 kit for human DNA removal. 2,4,8 with normal DNA extraction. Visible bands from samples 1R MA HD and 1R MA DNA Ex. Sample 6 contains *E. Faecalis* control of which the visible band is from. Negative PCR controls in wells 5 and 9. **B)** 1,3,5,7,9,11 treated with MolYsis™ Basic5 kit for human DNA removal. 2,4,6, 8, 10, 12 with normal DNA extraction. Negative PCR control in 13. Positive *E. faecalis* control in 14.

3.2.2 Diversity analysis

CLC Metagenomic workbench was used in 16S V3–V4 and Shotgun Metagenomic analyses. For OTU clustering, Silva138 and greengenes2 databases were applied to determine the microbial compositions. One of the samples, 3R BE DNA extraction, the host DNA removal samples held higher number of observed species compared to DNA extraction samples. After

performing diversity analysis to sequences from which low-quality sections and adapter sequences (reads) were removed, the amount of reads for DNA Ex samples was greatly reduced. Overall, the richness were low as well as the number of bacterial reads (Fig 12).

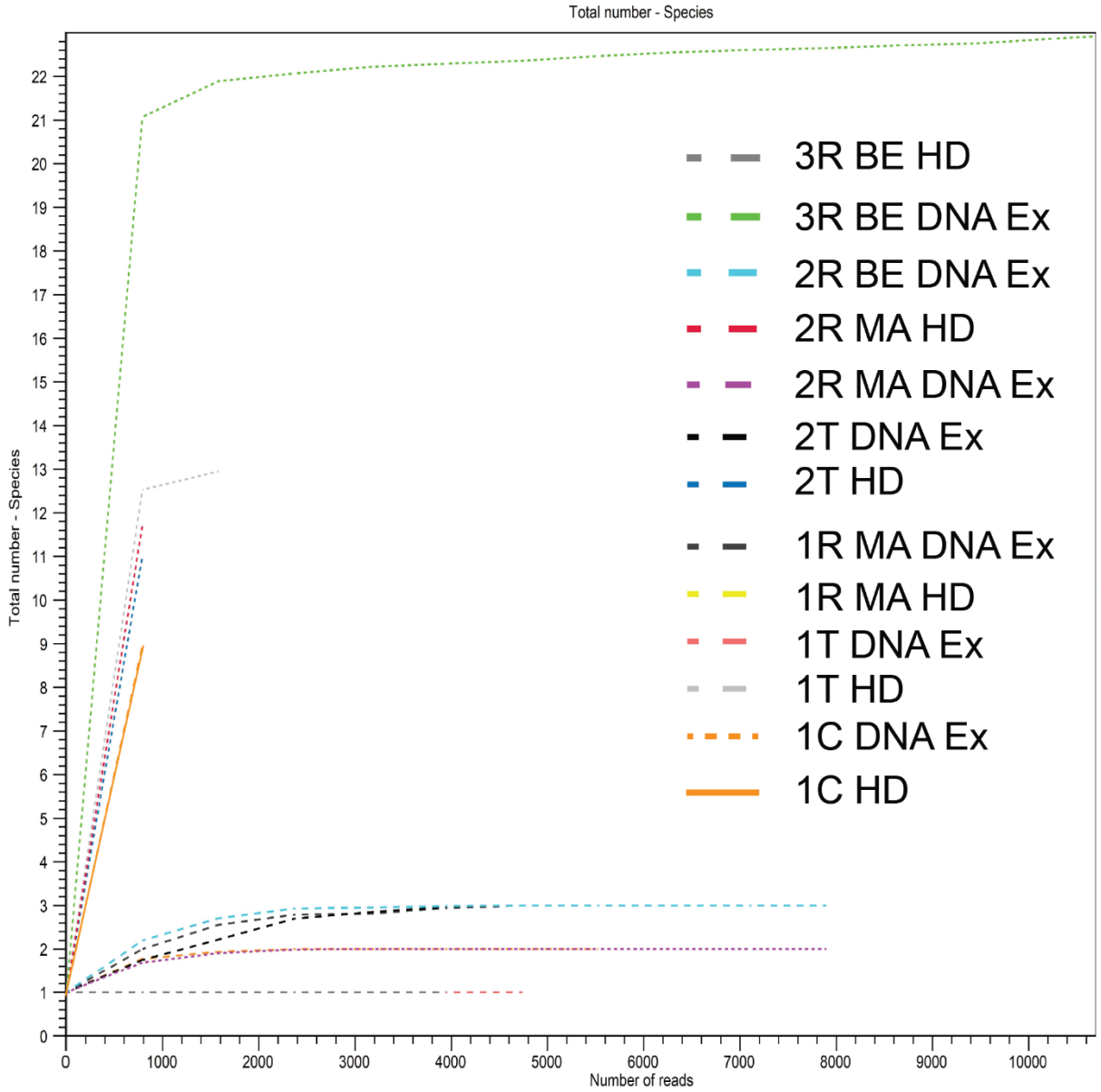

Figure 12. 16S V3–V4 sequencing rarefaction curves showing observed species richness in individual samples. Highest number of species in 3R BE DNA Ex, value. HD = Host depletion, NC = negative control, DNA Ex = DNA extraction, C = Cystectomy, R = Robotic assisted radical prostatectomy, T = Transurethral resection of the prostate, BE = Benign, MA = Malign

3.2.3 16S V3–V4 sequencing controls

The average amount of bacterial reads was 3500 for negative DNA extraction controls and 2200 for negative host depletion controls. The read amount for PCR control was around 3800. For prostate tissue samples from which human DNA was removed, the average read amount was

6200 and DNA extraction samples 28300. The average proportion of total sequencing reads usable for analysis was 67% of total reads for host depletion and 26% of total reads for DNA extraction samples.

3.2.4 16S V3–V4 sequencing samples

The average reads for prostate tissue samples were $2x10^4$ (67% of total reads) in host depletion and $28x10^5$ in DNA extraction (26% of total reads). Majority of reads was unclassified or was blasted as host reads. These reads were removed. Removing host DNA backgrounds altered the microbial profiles both in negative host depletion and DNA extraction controls as well as in tissue samples (Fig. 13-14.) The most abundant genera in negative controls were *Enterobacteriaceae*, *Escherichia-Shigella*, *Corynebacterium* and *Micrococcaceae*, *Micrococcus*.

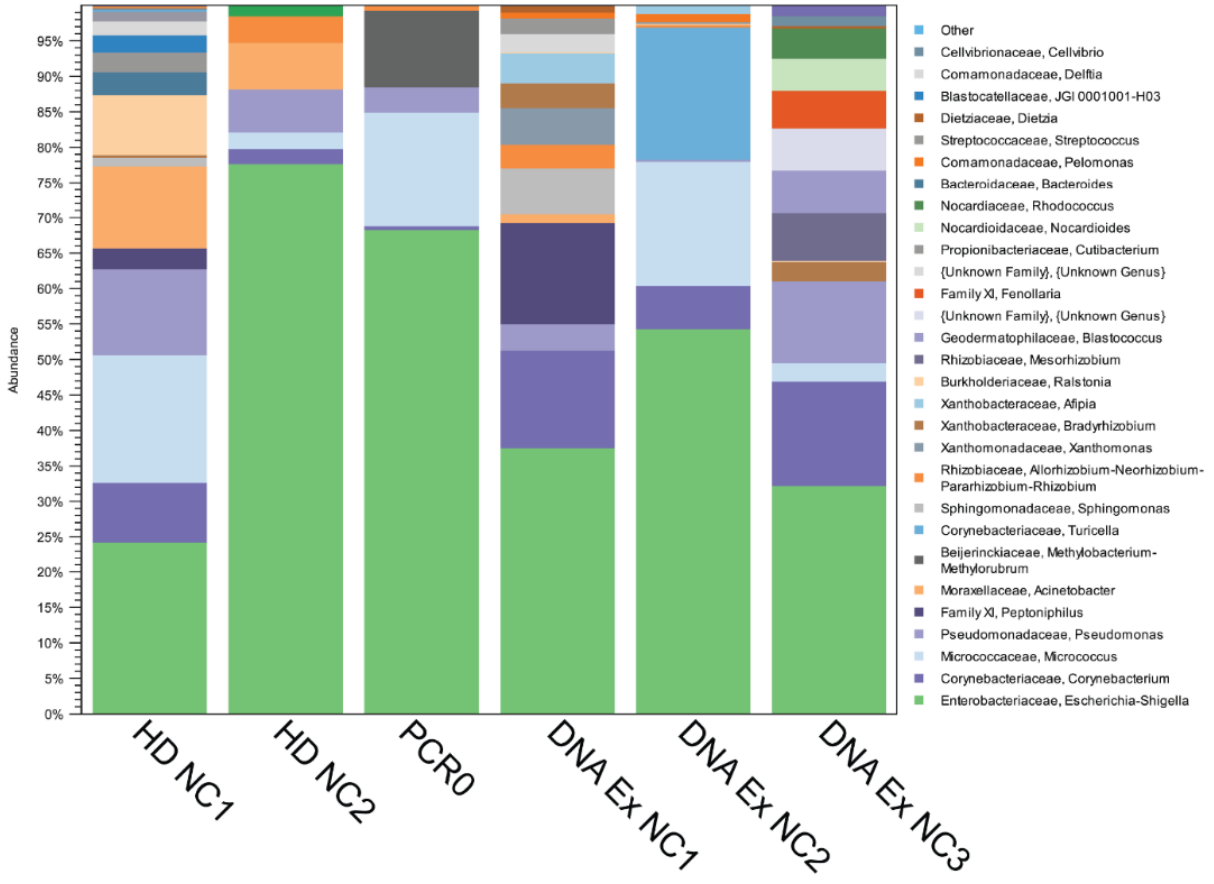

Figure. 13. Relative abundances of bacteria at the genus level in negative control samples. The most abundant genera were *Escherichia-Shigella*, *Corynebacterium* and *Micrococcus*. 3R BE DNA Ex left no microbial profile after decontamination with CLC Metagenomic Workbench, Silva138 (97% threshold). HD = Host depletion, NC = negative control, DNA Ex = DNA extraction, C = Cystectomy, R = Robotic assisted radical prostatectomy, T = Transurethral resection of the prostate, BE = Benign, MA = Malign. Genus-level abundances were determined using OTU clustering at 97% identity.

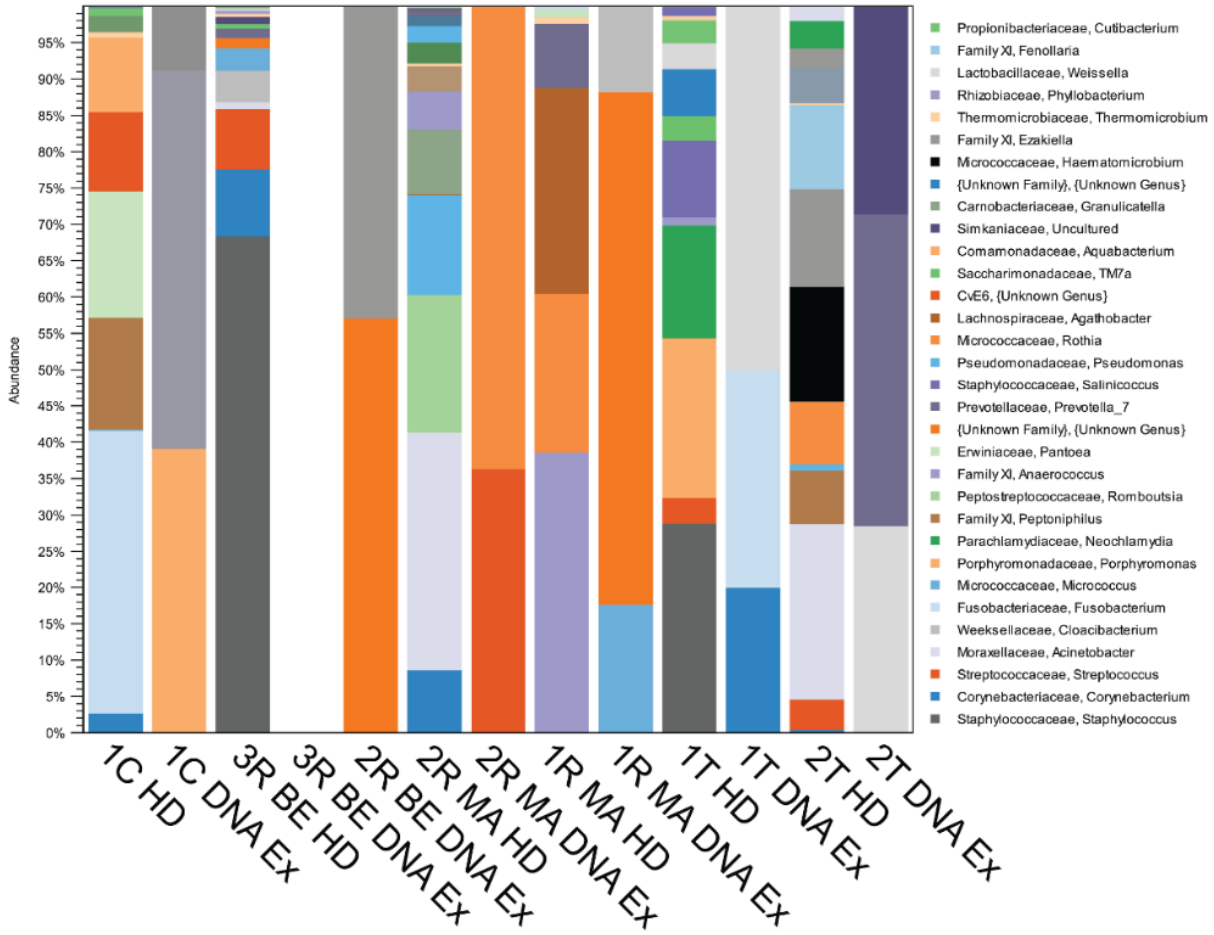

Figure. 14. Relative abundances of bacteria at the genus level in PCa samples. Most abundant genus were *Staphylococcus*, *Corynebacterium* and *Streptococcus*. Analyzed with CLC Metagenomic Workbench, Silva138 (97% threshold). HD = Host depletion, NC = negative control, DNA Ex = DNA extraction, C = Cystectomy, R = Robotic assisted radical prostatectomy, T = Transurethral resection of the prostate, BE = Benign, MA = Malign. Genus-level abundances were determined using OTU clustering at 97% identity.

Removing species included in the control samples, left several potential species deriving from prostate tissue environment. To avoid contamination, we excluded all the species included in the control samples despite the risk of losing original species from tissue samples that could have been got into the control samples. After this decontamination, sample 3R BE Host depletion left no microbial profile (Fig. 17). The most abundant species in control samples were *Escherichia coli str. K-12, Escherichia coli BL21, Micrococcus sp. X-48 (*Fig.16*)*. In prostate tissue samples, *Staphylococcus spZQ3, Staphylococcus pasteuri, Corynebacterium tuberculosteriacum* and *Cloacibacterium normanense* were the most abundant ones. *F. nucleatum* was still found from the sample 1C HD. Microbial species profiles were significantly

different between HD and DNA Ex sub-samples of the same tissue piece. Otherwise, the variation between samples was high, and no consistent microbial signature was associated with the two DNA extraction methods (Fig. 17).

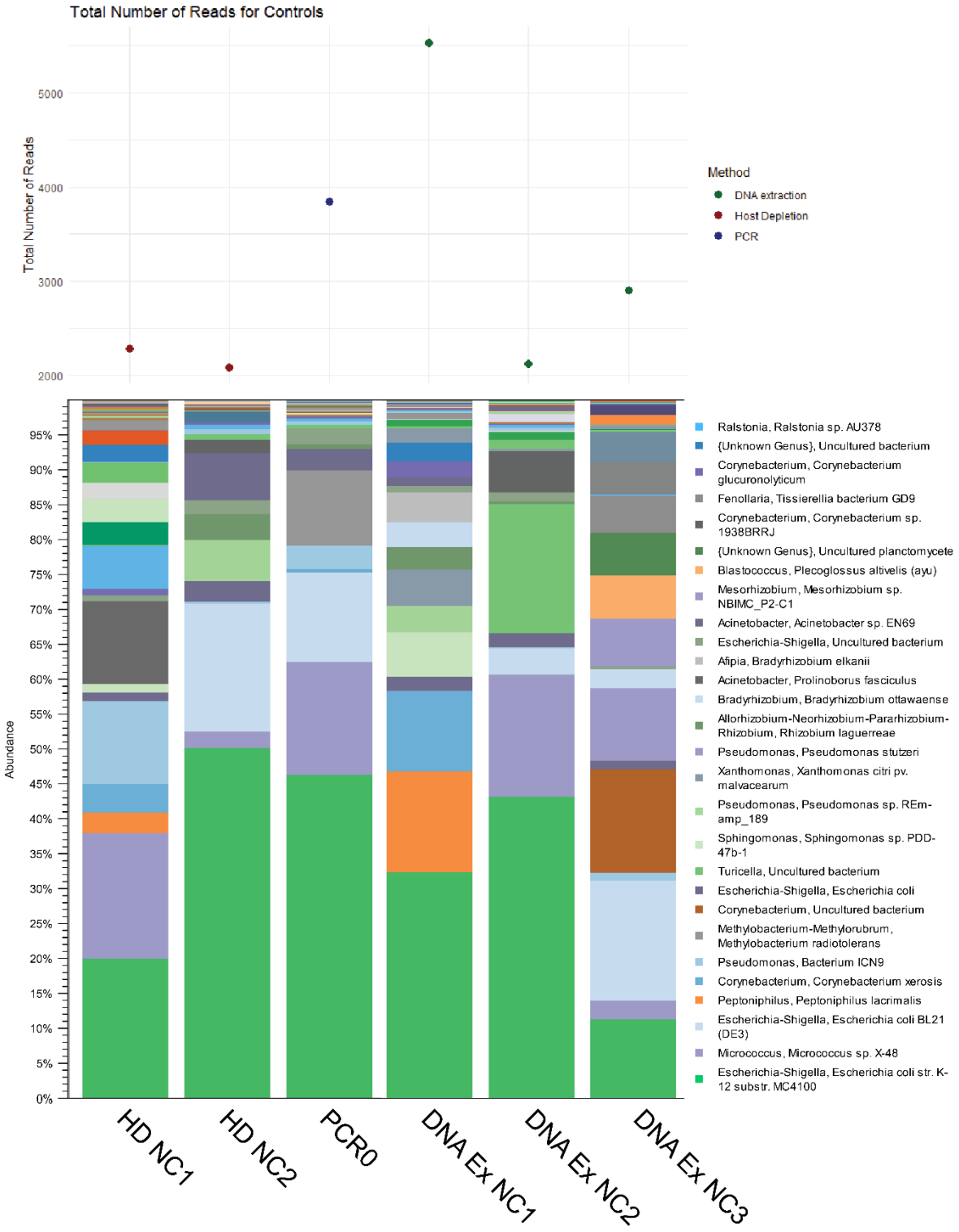

Figure. 16. Relative abundances of bacteria at the subspecies level in negative control samples. Most abundant subspecies were *Escherichia coli* str. K-12, *Micrococcus* sp. X-48 and *Escherichia coli* BL21. Analyzed with CLC Metagenomic Workbench, Silva138 (97% threshold). HD = Host depletion,

NC = negative control, DNA Ex = DNA extraction, C = Cystotectomy, R = Robotic assisted radical prostatectomy, T = Transurethral resection of the prostate, BE = Benign, MA = Malign. Species-level abundances were determined using OTU clustering at 97% identity.

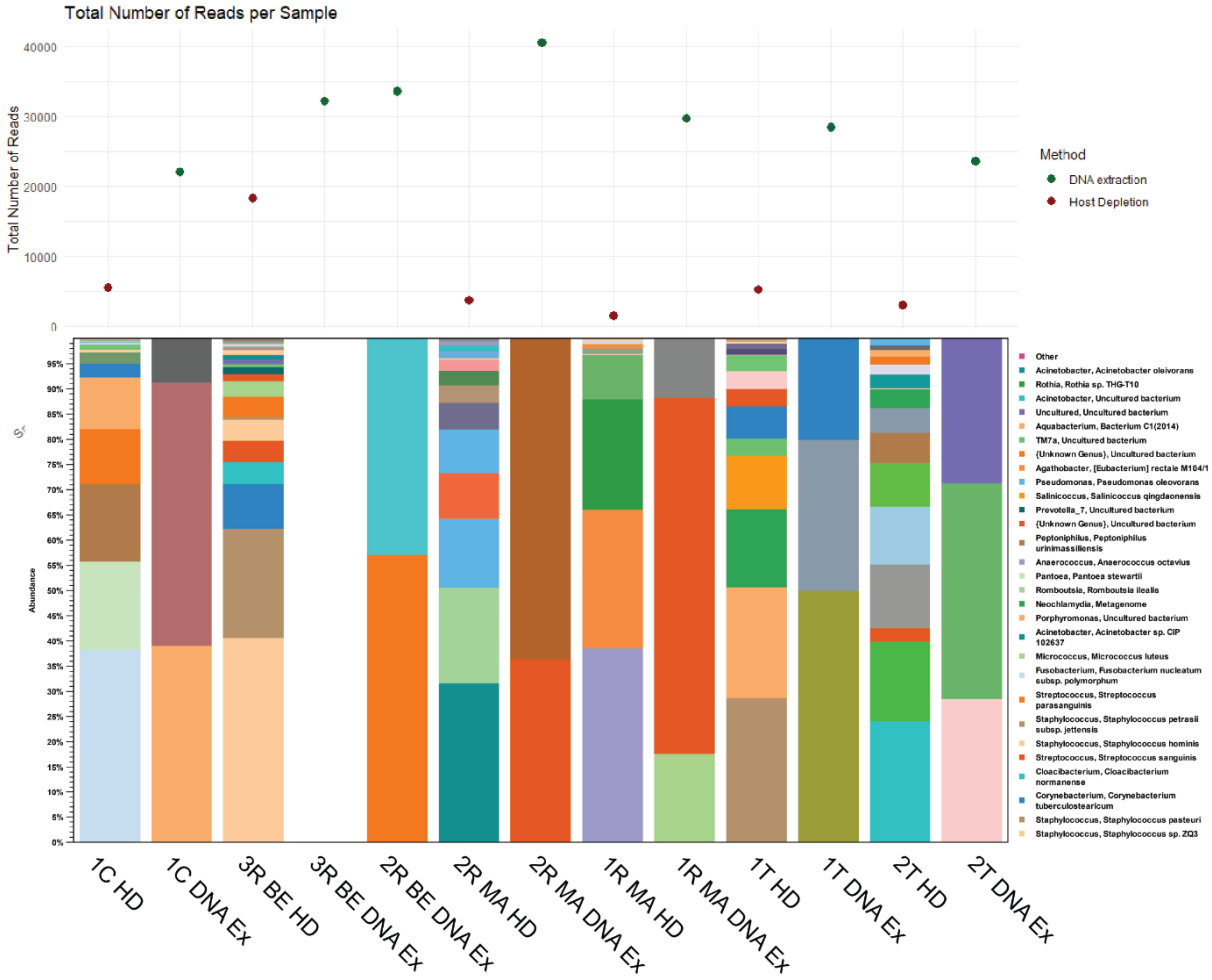

Figure. 17. Relative abundances of bacteria at the species level in PCa samples. Most abundant species were Staphylococcus spZQ3, *Staphylococcus pasteuri, Corynebacterium tuberculosteriacum* and *Cloacibacterium normanense*. 3R BE DNA Ex left no microbial profile after decontamination with CLC Metagenomic Workbench, Silva138 (97% threshold). HD = Host depletion, NC = negative control, DNA Ex = DNA extraction, C = Cystotectomy, R = Robotic assisted radical prostatectomy, T = Transurethral resection of the prostate, BE = Benign, MA = Malign. Species-level abundances were determined using OTU clustering at 97% identity.

ASV analysis was performed to the 16S V3–V4 sequencing samples using greengenes2 database. The same phenomenon was observed as in OTU clustering analysis and relative abundance table. Removing human DNA background from the negative control samples altered microbial profiles compared to DNA extraction samples. ASV analysis found smaller amount

of bacterial genera compared to OTU clustering with silva138 database. *Staphylococcus* and *Peptoniphilus* were the only common genera. The most common genera in negative control samples were *Micrococcus*, *Peptoniphilus A* and *Bradyrhizobium* and for prostate tissue samples, an unknown genus belonging to *Peptostreptococcaceae*, *Anaerococcus* and *C7867-001* after decontamination. Samples 1C HD, 2R MA DNA Ex, 1T DNA Ex and 2T DNA Ex left no microbial profile after decontamination (Fig. 19).

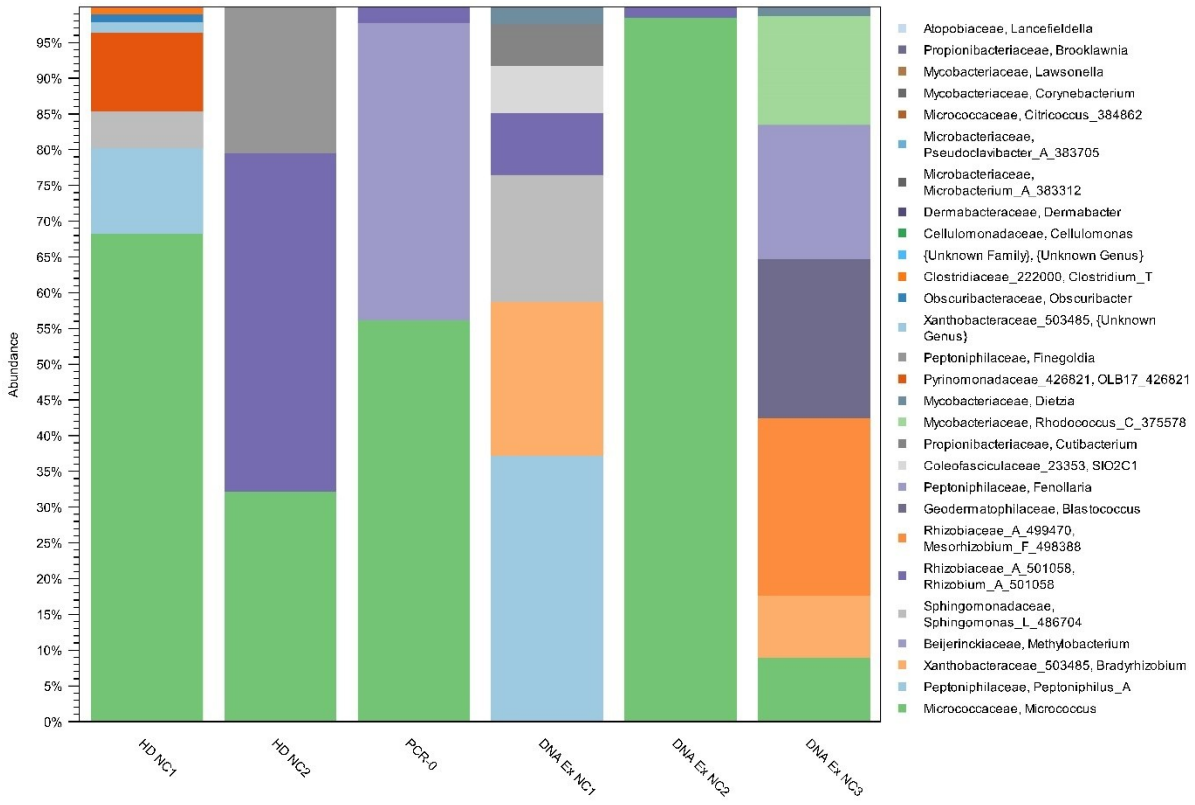

Figure. 18. Relative abundances of bacteria at the genus level in negative control samples. The most common genera in negative control samples were *Micrococcus*, *Peptoniphilus A* and *Bradyrhizobium*. Analyzed with CLC Metagenomic Workbench. HD = Host depletion, NC = negative control, DNA Ex = DNA extraction, C = Cystotectomy, R = Robotic assisted radical prostatectomy, T = Transurethral resection of the prostate, BE = Benign, MA = Malign. Genus-level abundances were derived using ASV analysis with Silva138 (97% threshold).

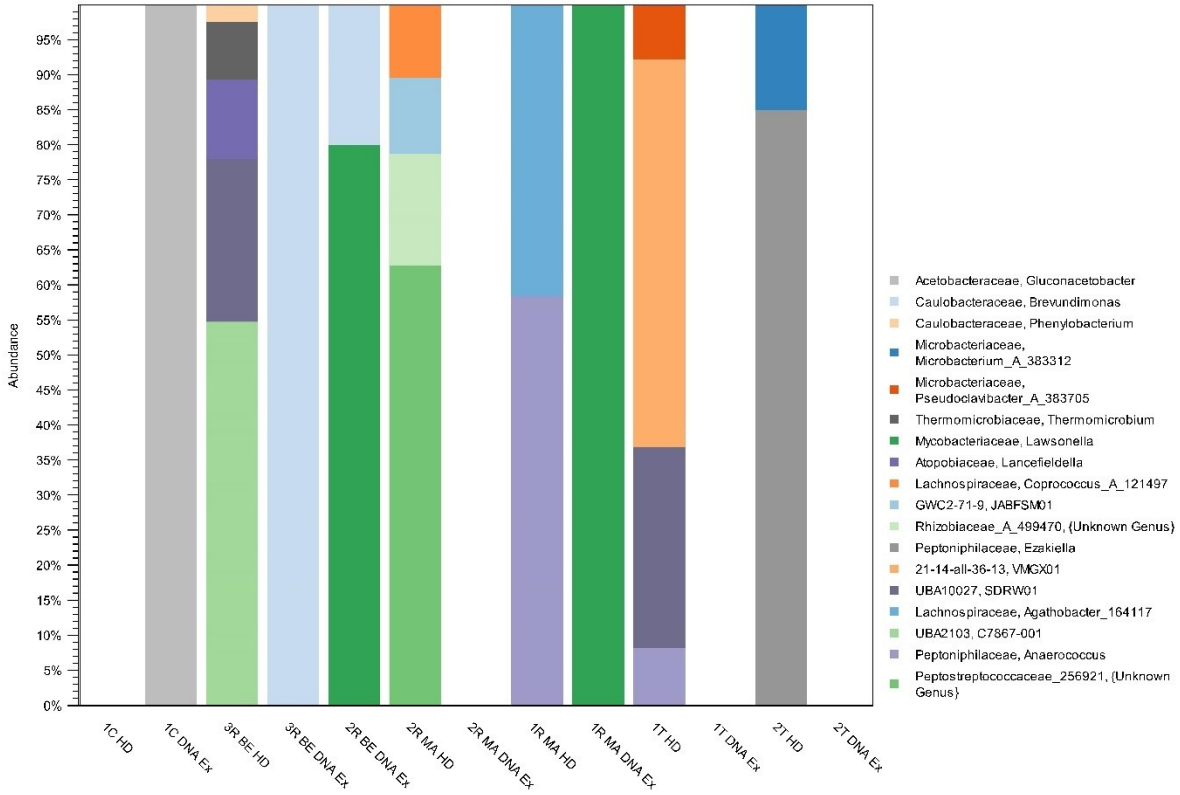

Figure. 19. 16S V3–V4 sequencing with sequence numbers of V3–V4 gene (reads) and relative abundances of bacteria at the species level with ASV analysis in prostate tissue samples. Most common genera were an unknown genus belonging to *Peptostreptococcaceae*, *Anaerococcus* and *C7867-001* after decontamination. Samples 1C HD, 2R MA DNA Ex, 1T DNA Ex and 2T DNA Ex left no microbial profile after decontamination. Analyzed with CLC Megagenomic Workbench. HD = Host depletion, NC = negative control, DNA Ex = DNA extraction, C = Cystotectomy, R = Robotic assisted radical prostatectomy, T = Transurethral resection of the prostate, BE = Benign, MA = Malign. Species-level abundances were derived using ASV analysis with Silva138 (97% threshold).

3.3 Shotgun Metagenomics

3.3.1 Shotgun metagenomics sequencing controls

Prostate tissue samples used for Shotgun Metagenomic analyses were annotated having an inflammation status of 2 or greater referring to possible prostatitis. This was due to hypothesis that the presence of inflammation would increase the likelihood of present microbes. Amplicon PCR of the V3-V4 region was performed to select samples going for Shotgun Metagenomic analysis. Amplicon PCR for Shotgun Metagenomics gave only faint bands. The size of the amplicon was 550 kb, and an appropriate positive spike-in *E.* Faecalis control and positive PCR water control as well as negative controls were applied for host depletion and DNA extraction

samples. Amplicon PCR confirmed that host depletion had worked as spiked sample left a visible band meaning that spiked cells were not lysed when treated with host depletion kit. (Fig. 20., samples A1-A10) as there was practically no band from any of the samples excluding spike control (A4) and positive control (A10). On the other hand, DNA extraction samples (B1-B10) were hypothesized to have even a small visible band, however, there was only a small faint band on top of the 500 kb ladder size on samples B7/B8. Based on the results, subsample sets 5 and 8 were chosen for Shotgun Metagenomics.

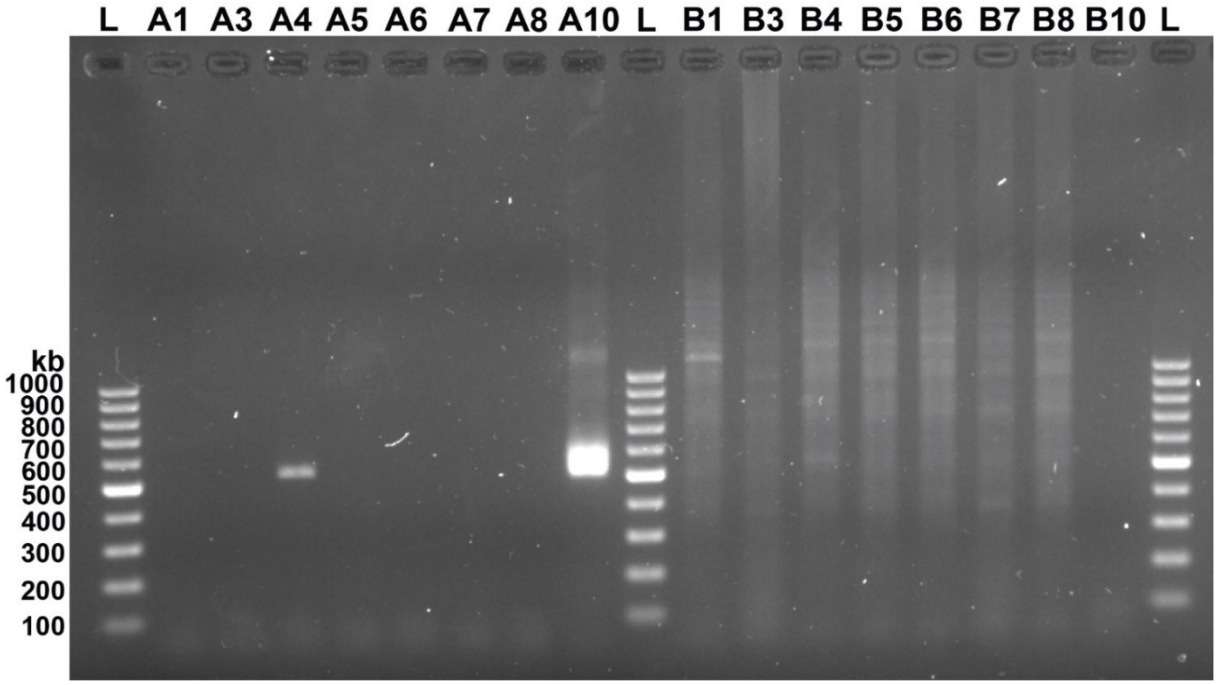

Figure 20. Amplicon PCR. A1-A8 treated with MolYsis™ Basic5 kit for host DNA removal. B1-B8 are equal to samples A1-A8 but without host depletion before DNA extraction. The size of the amplicon is 550 kb. Sample A4 contains *E. faecalis* control of which the visible band is from. A10 is a positive PCR control, B10 a negative PCR control.

3.3.2 Shotgun metagenomics sequencing tissue samples

Shotgun metagenomics samples were analyzed first with the CLC Metagenomic Workbench and then with Kraken/Braken pipeline. The average number of reads per sample obtained from the analysis was $39x10^6$ of which <0,1% was identified as microbial. The number of unidentified reads was on average $12x10^6$. Human reads were excluded with CLC Metagenomic Workbench, but CLC had difficulties in microbial identification, which is why Kraken/Braken was performed. Most common genera were *Cutibacterium*, *Homo sapiens*, *Ralstonia* and

Corynebacterium. Despite using both the CLC Metagenomic Workbench and Kraken databases to remove human DNA sequences, Bracken still identified a large number of reads as being of human origin. This is likely because not all host (human) reads are successfully mapped to the CLC human reference, and Kraken classifies both microbial and human reads. As a result, some host-derived reads remain even after host removal steps. With Kraken, these reads can't be specifically removed from the dataset, but they can be excluded from the microbial profile, as they do not truly represent microbial content. The genera common between 16S V3–V4 sequencing and Shotgun Metagenomics were *Staphylococcus* and *Corynebacterium*, both common for skin microbiome. Overall, shotgun Metagenomics species abundances were low.

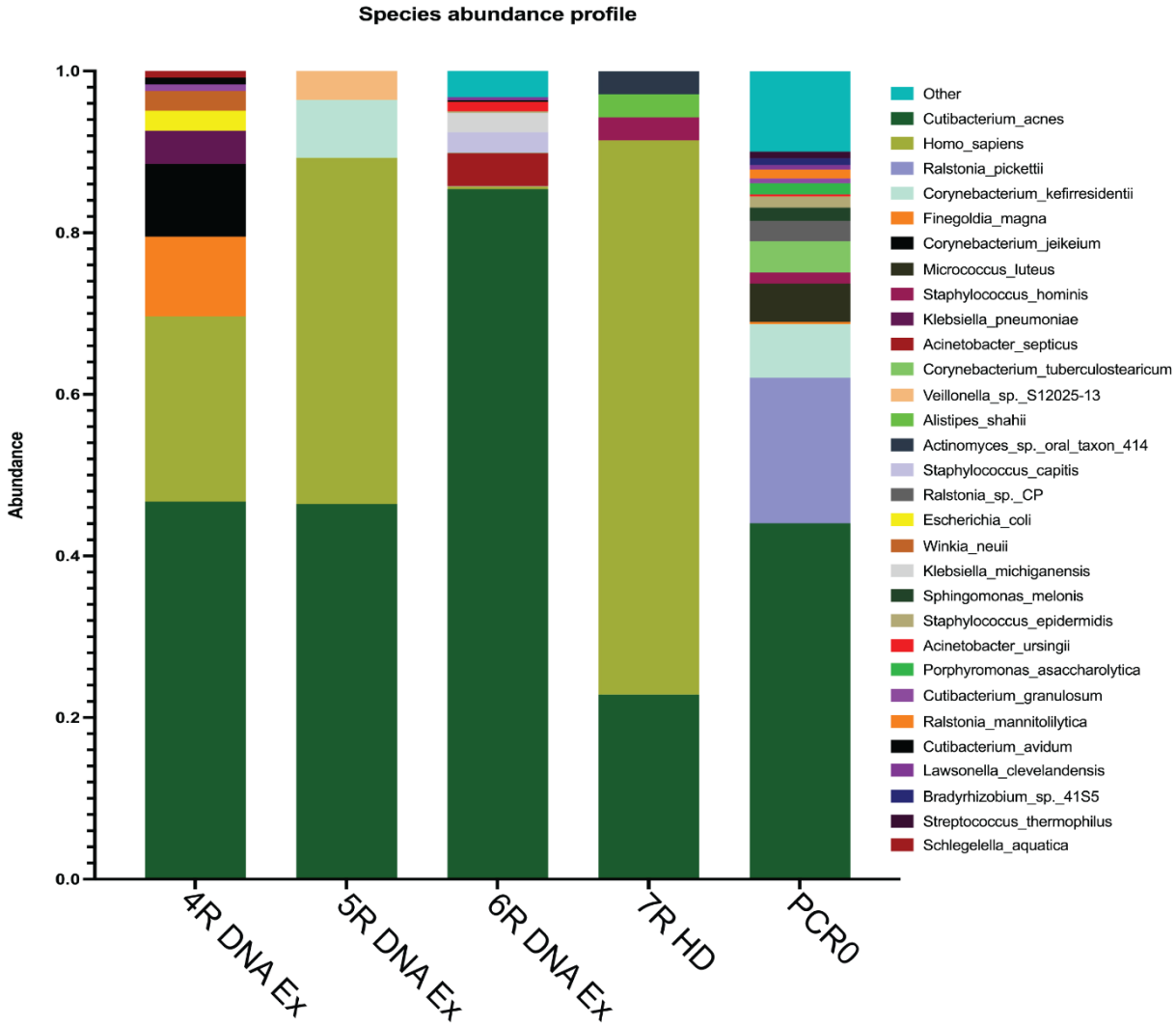

Figure 21. Shotgun metagenomics. Relative abundances of classified microbial species in prostate cancer tissue samples. Most common genera were *Cutibacterium*, *Homo sapiens*, *Ralstonia* and *Corynebacterium*. Analysis was performed with Kraken/Bracken pipeline. R = Robotic assisted radical prostatectomy, DNA Ex = DNA extraction, HD = Host depletion

4 Discussion

4.1.1 16S V3–V4 sequencing

Removing host DNA background altered the microbial profiles of the prostate tissue samples. In addition, the abundance of profiles received from ASV analysis were different compared to OTU clustering. This could be potentially due to ASVs ability to distinguish only real sequence differences and not arbitrary 3% thresholds or overestimation of species composition by OTU clustering. Generally, human DNA removal is considered to improve bacterial signal despite reducing the total amount of reads significantly (Lazarevic et al., 2022; Marquet et al., 2022). As a result, this is often seen as greater riches, diversity and different dominant taxa, which is also seen in our samples. However, removing human DNA can also lead to unspecified degradation of bacterial DNA during pre-treatment as well as unspecific amplification of human DNA in amplicon PCR step in library preparation. Moreover, the lower the amount of DNA in the samples, the more prone they are to contamination (6.1). Reads amount we received from 16S V3–V4 sequencing were extremely low. Compared to similar studies (Barot et al., 2024; Heravi et al., 2020; Kim et al., 2024.) with DNA extraction from tissues, our read amounts were below average (<33 400 reads per sample). However, studies do not often report the read amount in control samples or there is only little talk about the significance of the reads or the effect of low-abundance microbial environment on reads and contamination (Barot et al., 2024; Cavarretta et al., 2017; Kim et al., 2024). Thus, this comprises in reproducibility as there is no common guideline for control reads threshold nor a method to minimize the effect of contamination.

4.1.2 DNA yield

The prostate tissue samples were weighed, and DNA yield was determined for both host depletion and DNA extraction sample sets (4.1). We noted that DNA concentrations didn't correlate with tissue weight, but samples from which human DNA was removed had significantly lower amounts of DNA, as expected. Thus, based on our results, an optimal tissue size is between 70-90 mg to make bead-beating (3.4) more efficient and less time consuming, as the time required for lysis increased significantly with tissues weighing more than 90 mg.

4.1.3 Controls

As microbial contamination is a major problem, especially in low abundance microbial extraction protocols that are extremely prone to contamination (5.4), the usage of appropriate amount of different types of controls is important. A common method is to sequence extraction controls along with the samples and decontaminate the samples by removing the bacteria included in the controls (Dyrhovden et al., 2021). In our study, we used PCR grade water with negative PCR controls as well as positive spike-in *E. faecalis* control to validate the functionality of host depletion and PCR. The most abundant genus was *Escherichia-Shigella*, present in all control samples. Commonly this is caused by contamination from water, kit reagents, laboratory and laboratory surfaces (Liu et al., 2022; Salter et al., 2014; Weiss et al., 2014). In addition, several other genera, such as *Pseudomonas*, *Micrococcus* and *Staphylococcus* identified in this study are also potentially associated with contamination. After excluding all the species present in negative control samples from the prostate tissue samples, the majority of the genera and species were still common for skin microbiome. This can be due to the contamination which could have occurred in another phase of the protocol, such as sample collection phase, amplicon PCR or library preparation

4.1.4 Diversity analysis

We estimated alpha and beta-diversities using CLC Metagenomic Workbench diversity tools. The samples from which human DNA was removed had higher bacterial alpha diversity compared to DNA extraction samples. However, total diversities were still underwhelming despite 3R BE DNA Ex sample, which was for an unknown reason much more abundant in richness/number of species. The average read amount for host depletion samples was 6200 and 28300 for DNA extraction samples. After performing diversity analysis to our trimmed sequences, from which low-quality sections and adapter sequences (reads) were removed the amount of reads for DNA Ex samples was greatly reduced. Considering the low number of reads in both control and prostate tissue samples, it was challenging to determine a baseline threshold for contamination.

4.1.5 Essential species

Regarding cancer, our samples included several species that have been associated with different cancers. *Fusobacterium nucleatum* has been associated with colorectal cancer and prostate cancer due to its pro-inflammatory capabilities. Many species have been related to colorectal

cancer. *Peptostreptococcaceae* family has also been related to being more prevalent in CRC patients causing inflammation and dysbiosis. Especially *Peptostreptococcus anerobius* has been observed to be enriched in colorectal cancer patients. Studies suggest that *P. anerobius* attaches to the CRC mucosa with specific integrins activating PI3K-Akt pathway which promotes inflammation accelerating tumor development (Liu et al., 2024; Long et al., 2019; Mao et al., 2023). On the other hand, species of *Lachnospiraceae* family might give protection against CRC by enhancing the immune system by promoting the activation of tumor infiltrating $CD8^+$ T cells (Zhang et al., 2023).

Escherichia-Shigella is commonly found in the human GI tract and is a potential causer of shigellosis which can result in diarrhea and fever. *Corynebacterium* and *Micrococcus* on the other hand are typically part of skin microbiome. *Corynebacterium* and *Micrococcus* refer to contamination, either from the environment or from the prostate tissue samples. The most abundant genus in prostate tissue samples were *Staphylococcus*, *Corynebacterium and Streptococcus*, both of which are typical for skin microbiome. The only genus linked strongly to prostate cancer was *Fusobacterium* found from sample 1C HD. *Fusobacterium nucleatum* is part of the normal pharyngeal microbiota, however it has been suggested that *F. nucleatum* has the capability to translocate to the colon and contribute to the progression of colorectal carcinoma by increasing cell proliferation in animal models. The mechanism by which *F. nucleatum* operates is via the interaction of adhesins Fap2 and FadA, which interact with the immune system and cancer cells. FadA is a surface adhesin, capable of regulating proinflammatory cytokine production (Boyanova et al. 2023; Minarovits 2021). Species of the genus such as *Porphyromonas*, *Streptococcus* and *Prevotella* have been associated with oral, gastrointestinal and cervical cancers. (Bai et al. 2024; Díaz-Basabe et al. 2024; Rai et al. 2021.)

4.2 Shotgun metagenomics sequencing

In the V3-V4 amplicon PCR, we received only faint bands from all samples except positive *Enterococcus Faecalis* spike-in control indicating that the amount of bacterial DNA was low. We chose the most promising samples for Shotgun Metagenomic analysis (Fig. 20) and determined the relative abundances with CLC Metagenomic Workbench as well with Kraken/Braken pipeline. The amount of bacterial/microbial reference database matches was under 0,1% from the total read amount with both methods. Most common genera were *Cutibacterium*, *Ralstonia* and *Corynebacterium*. All thes genera are potential contaminants

from skin or reagent environment. Even after removing all the human DNA sequences with the CLC Metagenomic Workbench database, Braken identified a large number of reads to be from human origin. The reason could be potentially that some bacteria share common DNA sequences with humans which would explain the overlap identification by Kraken/Braken or if CLC Metagenomic Workbench failed to remove all the human reads during host read decontamination. The species common between 16S V3–V4 sequencing and Shotgun Metagenomics were *Staphylococcus* and *Corynebacterium*, both common for skin microbiome.

4.2.1 Controls

In Shotgun Metagenomic analysis, we used only negative PCR control as the amplicon PCR confirmed both negative PCR controls for HD and DNA Ex samples to be working along with spike-in control. (Fig. 20). However, for an unknown reason, the species' richness was the greatest in this negative PCR control. This raises the question whether the prostate had microbiome at all as negative control baseline is set relatively high and the microbial profiles of prostate tissue samples are hard to differentiate from this background.

4.2.2 Essential species

Genera and species blasted from shotgun metagenomic sequencing were much lower compared to ones analyzed with 16S V3–V4 sequencing. However, in our study, we found several species that have been associated with prostate cancer. *Cutibacterium acnes* (formerly known as *Propinibacterium acnes*) is a common skin contaminant but it has been also linked to chronic inflammation promoting prostate cancer via secretion of IL-6 and IL-8 in infected cells. Response was mediated by transcription factors NF-κB and STAT3. In addition, according to Fassi et al (2011), *P. acnes* induced cell response activated COX2 prostaglandin as well as the plasminogen-matrix metalloproteinase pathways eventually leading to initiation of cellular transformation. (Fassi Fehri et al., 2011.)

Klebsiella pneumoniae, a pathogen known for causing pneumonia, has been also linked to a few cancers such as hepatocellular carcinoma, bladder cancer and colorectal cancer. In hepatocellular carcinoma, surface protein PBP1B of *K. pneumoniae* activates TLR4 on HCC cells which causes cell proliferation and activates oncogenic signaling (Wang et al., 2025). In bladder cancer, *K. pneumoniae* was one of the 5 most abundant species found in bladder-cancer tissue samples, compared to the urine. *K. pneumoniae* is known to cause DNA double-strand

breaks which leads to instability and cell cycle arrest. Together with colibactin toxin and secretion of proinflammatory cytokines in the colon with *Klebsiella* present, this can result in chronic inflammation and epithelial cell proliferation. (Mansour et al., 2020.) Lastly, stated by (Aschtgen et al., 2022), in addition to secreting genotoxin colibactin, LPS from *K. pneumoniae* has been noticed also to inhibit p53 at the mRNA level, through activation of TLR4- NF-κB pathway.

Veillonella parvula is commonly associated with lung cancer. According to (Zeng et al., 2023), it turned out in in vivo studies that *V. parvula* promoted the growth of lung adenocarcinoma in mice via suppression of tumor-associated T-lymphocytes and peripheral T-lymphocytes. Further analysis clarified that *V. parvula* induced CCN4 expression and activated NOD-like receptor as well as NF-κB pathway's Nod2/CCN4 signaling resulting in cell proliferation in the tumor microenvironment. According to (Zhou et al., 2023) both *Veillonella* and *Streptococcus* were more abundant in plasma of lung cancer patients compared to healthy controls. *Streptococcus* has been shown to promote lung tumorigenesis by activating of NF-κB pathway, through binding PspC to PAFR.

4.3 Challenges and limitations

Both 16S V3–V4 and shotgun metagenomic sequencing are prone to bias due to high risk of contamination within and between samples. Contamination can occur in various stages beginning from the surgery all the way through sample preparation, pre-treatment and final analysis. As an organ prostate is supposed to be nearly sterile. However, immediately after removal, the prostate gets exposed to the environment and its microbes. No matter how sterile the hospital environment is designed to be, there is always going to be contamination to a certain extent. After the prostates were removed, they were delivered to TYKS Pathology unit for the making of the prostate core samples (**Fig 9B**). Cores were then frozen at -80 degrees Celsius.

Cores were melted/thawed and pre-treated with MolYsis™ Basic5 kit. During the pre-treatment, samples were processed in a laminar flow hood. Pre-treatment included several steps which exposed samples for contamination. Even though the samples were dealt with extreme caution, contamination couldn't be fully prevented. Thus, it is hard to tell if the signal received is from actual prostate microbiome, cross-contamination or reagent contamination This significantly affects downstream analyses, especially in studies involving low microbial

abundance samples, where even small amounts of contamination can compromise the accuracy of metagenomic analyses. (Chrisman et al., 2022; Merchant et al., 2014; Salter et al., 2014.)

Various tools have been developed to reduce contamination from laboratory protocols. These tools are based on either sequencing a reagent-only or blank sample to establish a baseline for contamination level or measure the total amount of target DNA in a sample assuming that the lower the amount of target DNA, the higher the level of contamination. Although these tools have improved the accuracy of several microbiome studies, their underlying assumptions might fail in certain experimental designs. This is particularly a problem for studies with low microbial abundance, or when factors are not accounted for in the controls or when the amount of contaminant DNA is resembles the actual sample.

Another significant issue is the presence of contamination in reference databases. According to Chrisman et al (2022), a number of studies have reported presence of human DNA contamination in non-primate reference genomes as well as in GenBank. (Chrisman et al., 2022; Longo et al., 2011; Steinegger & Salzberg, 2020.) A rough way to perform decontamination is to exclude all the species included in the control samples. However, during this process, a number of species naturally derived from the prostate might be lost (Dyrhovden et al., 2021; Eisenhofer et al., 2019). Additionally, the threshold for low microbial abundance samples is naturally lower. As there is no common consensus where this line should be drawn, studies might vary in thresh hold making reproducibility and reliability worse.

So far, the optimal extraction protocol to extract microbial components from prostate tissue is absent. To achieve reliable results, optimizing DNA extraction from prostate tissue for metagenomic studies is crucial as extracting bacterial DNA tends to contaminate very easily. Moreover, negative results are often not mentioned or there is only a brief mention of the issue. Therefore, to provide reliable research data on the isolation of the prostate tissue microbiome, it is crucial to use proper controls and to give enough comprehensive information on negative outcomes. Furthermore, future research is needed to understand the lying mechanism and evaluate whether gut microbiota signature could be used to assess the risk of prostate cancer and to help to understand the relationship between gut and health.

4.4 Prospects

Reliable and reproducible DNA extraction methods from low-microbial abundance samples, such as tissue samples are still absent. Extracting microbial DNA from tissue samples has a challenge of non-specific amplification of human DNA. According to López-Aladid et al. (2023) V1–V2 region is more sensitive and specific, exhibiting the highest resolving power for taxonomy identification from respiratory samples. In this study, they compared the resolving power of regions V1–V2, V3–V4 and V5–V7 and V7–V9 with 33 human sputum samples. The alpha diversity was found to be significantly higher for V1–V2, V3–V4 and V5–V7 regions compared to V7–V9. Walker et al. (2020) reported that a few human biopsy sample types had a significant off-target amplification percentage when using 16S V3–V4 sequencing. Breast tumor samples were most affected sample type and when re-analyzed with V1–V2 primer set, a considerable reduction in off target amplification was reached. As prostate and mammary glands share many similarities, both comprised of luminal and basal/myoepithelial cells, V1–V2 could be potentially more accurate compared to V3–V4 sequencing (Fig. 22). (López-Aladid et al. 2023; Walker et al. 2020).

As discussed, due to high risk of contamination, it is hard to prove microbiome to be non-existent in such studies. We noted that various studies have found a prostate microbiome from PCa patients as well as several essential species that might alter the tumor microenvironment. Gut microbiome is an increasingly popular topic, and its significance has been noted on various aspects of human health along with PCa. However, tissue microbiome should be evaluated to validate and understand interactions between GM and PCa properly. Thus, optimizing not only the protocol for DNA extraction but also the protocol for reporting results and drawing baselines for contamination background is crucial to achieve reliable and reproducible results.

5 Conclusions

This study aimed to optimize a reliable and reproducible DNA extraction method from prostate tissue. We also pursued to investigate if prostate had a unique microbiome, while observing the effect of host DNA removal to microbial profiles of PCa tissue samples. The optimization included PCa samples obtained by three different surgery techniques, robotic-assisted laparoscopic prostatectomy, transurethral resection of the prostate as well as prostate tissue samples from cystectomy. All samples were weighed, bead-beaten, and split into two sets—one treated with the MolYsis™ Basic5 kit to remove human DNA, the other untreated. Negative and PCR controls were included for both. Microbial composition was analyzed using 16S V3–V4 and shotgun metagenomic sequencing. Removing host DNA background altered the microbial profiles and enhanced the richness of species in PCa samples. Thus, host depletion can lessen the costs of Shotgun metagenomic sequencing, as more shallow depth is required after majority of the human DNA has been removed. However, the samples were very low in bacteria making determination of contamination difficult and the host depletion kit can also degrade microbial DNA as well. Further studies are required to reduce the burden of unspecified DNA in low-abundance tissue samples as well as to prove if prostate tissue has microbiome at all.

Acknowledgements

I want to thank my supervisors Sanja Vanhatalo and Peter Boström for their invaluable support and advice during my thesis. I'm in gratitude to Katri Kylä-Mattila and Päivi Haaranen for introducing me to basics of the microbial work. I'd also like to thank Teemu Kallonen for his experienced guidance in difficult situations. A special thanks to Pekka Taimen and Peter Boström for giving me the opportunity to work with so many amazing people and for supporting me in every way possible during this process.

References

Abellan-Schneyder, I., Matchado, M. S., Reitmeier, S., Sommer, A., Sewald, Z., Baumbach, J., List, M., & Neuhaus, K. (2021). Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. *mSphere*, *6*(1), e01202-20. https://doi.org/10.1128/mSphere.01202-20

Aschtgen, M.-S., Fragkoulis, K., Sanz, G., Normark, S., Selivanova, G., Henriques-Normark, B., & Peuget, S. (2022). Enterobacteria impair host p53 tumor suppressor activity through mRNA destabilization. *Oncogene*, *41*(15), 2173–2186. https://doi.org/10.1038/s41388-022-02238-5

Austin, G. I., & Korem, T. (2024). Planning and Analyzing a Low-Biomass Microbiome Study: A Data Analysis Perspective. *The Journal of Infectious Diseases*, jiae378. https://doi.org/10.1093/infdis/jiae378

Barot, S. V., Sangwan, N., Nair, K. G., Schmit, S. L., Xiang, S., Kamath, S., Liska, D., & Khorana, A. A. (2024). Distinct intratumoral microbiome of young-onset and average-onset colorectal cancer. *eBioMedicine*, *100*, 104980. https://doi.org/10.1016/j.ebiom.2024.104980

Bharti, R., & Grimm, D. G. (2021). Current challenges and best-practice protocols for microbiome analysis. *Briefings in Bioinformatics*, *22*(1), 178–193. https://doi.org/10.1093/bib/bbz155

Cavarretta, I., Ferrarese, R., Cazzaniga, W., Saita, D., Lucianò, R., Ceresola, E. R., Locatelli, I., Visconti, L., Lavorgna, G., Briganti, A., Nebuloni, M., Doglioni, C., Clementi, M., Montorsi, F., Canducci, F., & Salonia, A. (2017). The Microbiome of the Prostate Tumor Microenvironment. *European Urology*, *72*(4), 625–631. https://doi.org/10.1016/j.eururo.2017.03.029

Celis, A. I., Aranda-Díaz, A., Culver, R., Xue, K., Relman, D., Shi, H., & Huang, K. C. (2022). Optimization of the 16S rRNA sequencing analysis pipeline for studying in vitro communities of gut commensals. *iScience*, *25*(4), 103907. https://doi.org/10.1016/j.isci.2022.103907

Chrisman, B., He, C., Jung, J.-Y., Stockham, N., Paskov, K., Washington, P., & Wall, D. P. (2022). The human "contaminome": Bacterial, viral, and computational contamination in whole

genome sequences from 1000 families. *Scientific Reports*, *12*(1), 9863. https://doi.org/10.1038/s41598-022-13269-z

Cullin, N., Azevedo Antunes, C., Straussman, R., Stein-Thoeringer, C. K., & Elinav, E. (2021). Microbiome and cancer. *Cancer Cell*, *39*(10), 1317–1341. https://doi.org/10.1016/j.ccell.2021.08.006

Damhorst, G. L., Adelman, M. W., Woodworth, M. H., & Kraft, C. S. (2021). Current Capabilities of Gut Microbiome–Based Diagnostics and the Promise of Clinical Application. *The Journal of Infectious Diseases*, *223*(Supplement_3), S270–S275. https://doi.org/10.1093/infdis/jiaa689

Durazzi, F., Sala, C., Castellani, G., Manfreda, G., Remondini, D., & De Cesare, A. (2021). Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Scientific Reports*, *11*(1), 3030. https://doi.org/10.1038/s41598-021-82726-y

Dyrhovden, R., Rippin, M., Øvrebø, K. K., Nygaard, R. M., Ulvestad, E., & Kommedal, Ø. (2021). Managing Contamination and Diverse Bacterial Loads in 16S rRNA Deep Sequencing of Clinical Samples: Implications of the Law of Small Numbers. *mBio*, *12*(3), e00598-21. https://doi.org/10.1128/mBio.00598-21

Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., & Weyrich, L. S. (2019). Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends in Microbiology*, *27*(2), 105–117. https://doi.org/10.1016/j.tim.2018.11.003

Fassi Fehri, L., Mak, T. N., Laube, B., Brinkmann, V., Ogilvie, L. A., Mollenkopf, H., Lein, M., Schmidt, T., Meyer, T. F., & Brüggemann, H. (2011). Prevalence of Propionibacterium acnes in diseased prostates and its inflammatory and transforming activity on prostate epithelial cells. *International Journal of Medical Microbiology*, *301*(1), 69–78. https://doi.org/10.1016/j.ijmm.2010.08.014

Freedland, S. J., Aronson, W. J., Terris, M. K., Kane, C. J., Amling, C. L., Dorey, F., & Presti, J. C. (2003). The percentage of prostate needle biopsy cores with carcinoma from the more

involved side of the biopsy as a predictor of prostate specific antigen recurrence after radical prostatectomy: Results from the Shared Equal Access Regional Cancer Hospital (SEARCH) database. *Cancer*, *98*(11), 2344–2350. https://doi.org/10.1002/cncr.11809

Gao, F., Yu, B., Rao, B., Sun, Y., Yu, J., Wang, D., Cui, G., & Ren, Z. (2022). The effect of the intratumoral microbiome on tumor occurrence, progression, prognosis and treatment. *Frontiers in Immunology*, *13*, 1051987. https://doi.org/10.3389/fimmu.2022.1051987

Guo, W., Zhang, Y., Guo, S., Mei, Z., Liao, H., Dong, H., Wu, K., Ye, H., Zhang, Y., Zhu, Y., Lang, J., Hu, L., Jin, G., & Kong, X. (2021). Tumor microbiome contributes to an aggressive phenotype in the basal-like subtype of pancreatic cancer. *Communications Biology*, *4*(1), 1019. https://doi.org/10.1038/s42003-021-02557-5

Heravi, F. S., Zakrzewski, M., Vickery, K., & Hu, H. (2020). Host DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples. *Journal of Microbiological Methods*, *170*, 105856. https://doi.org/10.1016/j.mimet.2020.105856

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

Kennedy, K. M., De Goffau, M. C., Perez-Muñoz, M. E., Arrieta, M.-C., Bäckhed, F., Bork, P., Braun, T., Bushman, F. D., Dore, J., De Vos, W. M., Earl, A. M., Eisen, J. A., Elovitz, M. A., Ganal-Vonarburg, S. C., Gänzle, M. G., Garrett, W. S., Hall, L. J., Hornef, M. W., Huttenhower, C., … Walter, J. (2023). Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. *Nature*, *613*(7945), 639–649. https://doi.org/10.1038/s41586-022-05546-8

Kim, J. H., Seo, H., Kim, S., Rahim, M. A., Jo, S., Barman, I., Tajdozian, H., Sarafraz, F., Song, H.-Y., & Song, Y. S. (2024a). Different Prostatic Tissue Microbiomes between High- and Low-Grade Prostate Cancer Pathogenesis. *International Journal of Molecular Sciences*, *25*(16), 8943. https://doi.org/10.3390/ijms25168943

Kim, J. H., Seo, H., Kim, S., Rahim, M. A., Jo, S., Barman, I., Tajdozian, H., Sarafraz, F., Song, H.-Y., & Song, Y. S. (2024b). Different Prostatic Tissue Microbiomes between High- and Low-Grade Prostate Cancer Pathogenesis. *International Journal of Molecular Sciences*, *25*(16), 8943. https://doi.org/10.3390/ijms25168943

Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L.-I., McDonald, D., Melnik, A. V., Morton, J. T., Navas, J., Quinn, R. A., Sanders, J. G., Swafford, A. D., Thompson, L. R., Tripathi, A., Xu, Z. Z., … Dorrestein, P. C. (2018). Best practices for analysing microbiomes. *Nature Reviews Microbiology*, *16*(7), 410–422. https://doi.org/10.1038/s41579-018-0029-9

Lazarevic, V., Gaïa, N., Girard, M., Mauffrey, F., Ruppé, E., & Schrenzel, J. (2022). Effect of bacterial DNA enrichment on detection and quantification of bacteria in an infected tissue model by metagenomic next-generation sequencing. *ISME Communications*, *2*(1), 122. https://doi.org/10.1038/s43705-022-00208-2

Liss, M. A., White, J. R., Goros, M., Gelfond, J., Leach, R., Johnson-Pais, T., Lai, Z., Rourke, E., Basler, J., Ankerst, D., & Shah, D. P. (2018). Metabolic Biosynthesis Pathways Identified from Fecal Microbiome Associated with Prostate Cancer. *European Urology*, *74*(5), 575–582. https://doi.org/10.1016/j.eururo.2018.06.033

Liu, Y., Elworth, R. A. L., Jochum, M. D., Aagaard, K. M., & Treangen, T. J. (2022). De novo identification of microbial contaminants in low microbial biomass microbiomes with Squeegee. *Nature Communications*, *13*(1), 6799. https://doi.org/10.1038/s41467-022-34409-z

Liu, Y., Wong, C. C., Ding, Y., Gao, M., Wen, J., Lau, H. C.-H., Cheung, A. H.-K., Huang, D., Huang, H., & Yu, J. (2024). Peptostreptococcus anaerobius mediates anti-PD1 therapy resistance and exacerbates colorectal cancer via myeloid-derived suppressor cells in mice. *Nature Microbiology*, *9*(6), 1467–1482. https://doi.org/10.1038/s41564-024-01695-w

Long, X., Wong, C. C., Tong, L., Chu, E. S. H., Ho Szeto, C., Go, M. Y. Y., Coker, O. O., Chan, A. W. H., Chan, F. K. L., Sung, J. J. Y., & Yu, J. (2019). Peptostreptococcus anaerobius promotes colorectal carcinogenesis and modulates tumour immunity. *Nature Microbiology*, *4*(12), 2319–2330. https://doi.org/10.1038/s41564-019-0541-3

Longo, M. S., O'Neill, M. J., & O'Neill, R. J. (2011). Abundant Human DNA Contamination Identified in Non-Primate Genome Databases. *PLoS ONE*, *6*(2), e16410. https://doi.org/10.1371/journal.pone.0016410

Mansour, B., Monyók, Á., Makra, N., Gajdács, M., Vadnay, I., Ligeti, B., Juhász, J., Szabó, D., & Ostorházi, E. (2020). Bladder cancer-related microbiota: Examining differences in urine and tissue samples. *Scientific Reports*, *10*(1), 11042. https://doi.org/10.1038/s41598-020-67443-2

Mao, Y., Xiao, X., Zhang, J., Mou, X., & Zhao, W. (2023). Designing a multi-epitope vaccine against Peptostreptococcus anaerobius based on an immunoinformatics approach. *Synthetic and Systems Biotechnology*, *8*(4), 757–770. https://doi.org/10.1016/j.synbio.2023.11.004

Marquet, M., Zöllkau, J., Pastuschek, J., Viehweger, A., Schleußner, E., Makarewicz, O., Pletz, M. W., Ehricht, R., & Brandt, C. (2022). Evaluation of microbiome enrichment and host DNA depletion in human vaginal samples using Oxford Nanopore's adaptive sequencing. *Scientific Reports*, *12*(1), 4000. https://doi.org/10.1038/s41598-022-08003-8

McCulloch, J. A., & Trinchieri, G. (2021). Gut bacteria enable prostate cancer growth. *Science*, *374*(6564), 154–155. https://doi.org/10.1126/science.abl7070

Merchant, S., Wood, D. E., & Salzberg, S. L. (2014). Unexpected cross-species contamination in genome sequencing projects. *PeerJ*, *2*, e675. https://doi.org/10.7717/peerj.675

Nejman, D., Livyatan, I., Fuks, G., Gavert, N., Zwang, Y., Geller, L. T., Rotter-Maskowitz, A., Weiser, R., Mallel, G., Gigi, E., Meltser, A., Douglas, G. M., Kamer, I., Gopalakrishnan, V., Dadosh, T., Levin-Zaidman, S., Avnet, S., Atlan, T., Cooper, Z. A., … Straussman, R. (2020). The human tumor microbiome is composed of tumor type–specific intracellular bacteria. *Science*, *368*(6494), 973–980. https://doi.org/10.1126/science.aay9189

Phipps, S., Yang, T. H. J., Habib, F. K., Reuben, R. L., & McNeill, S. A. (2005). Measurement of tissue mechanical characteristics to distinguish between benign and malignant prostatic disease. *Urology*, *66*(2), 447–450. https://doi.org/10.1016/j.urology.2005.03.017

Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G. A. D., Gasbarrini, A., & Mele, M. C. (2019). What is the Healthy Gut Microbiota Composition? A Changing Ecosystem

across Age, Environment, Diet, and Diseases. *Microorganisms*, *7*(1), 14. https://doi.org/10.3390/microorganisms7010014

Riquelme, E., Zhang, Y., Zhang, L., Montiel, M., Zoltan, M., Dong, W., Quesada, P., Sahin, I., Chandra, V., San Lucas, A., Scheet, P., Xu, H., Hanash, S. M., Feng, L., Burks, J. K., Do, K.-A., Peterson, C. B., Nejman, D., Tzeng, C.-W. D., … McAllister, F. (2019). Tumor Microbiome Diversity and Composition Influence Pancreatic Cancer Outcomes. *Cell*, *178*(4), 795-806.e12. https://doi.org/10.1016/j.cell.2019.07.008

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., & Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, *12*(1), 87. https://doi.org/10.1186/s12915-014-0087-z

Schloss, P. D. (2018). Identifying and Overcoming Threats to Reproducibility, Replicability, Robustness, and Generalizability in Microbiome Research. *mBio*, *9*(3), e00525-18. https://doi.org/10.1128/mBio.00525-18

Shalon, D., Culver, R. N., Grembi, J. A., Folz, J., Treit, P. V., Shi, H., Rosenberger, F. A., Dethlefsen, L., Meng, X., Yaffe, E., Aranda-Díaz, A., Geyer, P. E., Mueller-Reif, J. B., Spencer, S., Patterson, A. D., Triadafilopoulos, G., Holmes, S. P., Mann, M., Fiehn, O., … Huang, K. C. (2023). Profiling the human intestinal environment under physiological conditions. *Nature*, *617*(7961), 581–591. https://doi.org/10.1038/s41586-023-05989-7

Steinegger, M., & Salzberg, S. L. (2020). Terminating contamination: Large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biology*, *21*(1), 115. https://doi.org/10.1186/s13059-020-02023-1

Usyk, M., Peters, B. A., Karthikeyan, S., McDonald, D., Sollecito, C. C., Vazquez-Baeza, Y., Shaffer, J. P., Gellman, M. D., Talavera, G. A., Daviglus, M. L., Thyagarajan, B., Knight, R., Qi, Q., Kaplan, R., & Burk, R. D. (2023). Comprehensive evaluation of shotgun metagenomics, amplicon sequencing, and harmonization of these platforms for epidemiological studies. *Cell Reports Methods*, *3*(1), 100391. https://doi.org/10.1016/j.crmeth.2022.100391

Villette, R., Autaa, G., Hind, S., Holm, J. B., Moreno-Sabater, A., & Larsen, M. (2021). Refinement of 16S rRNA gene analysis for low biomass biospecimens. *Scientific Reports*, *11*(1), 10741. https://doi.org/10.1038/s41598-021-90226-2

Wang, X., Fang, Y., Liang, W., Cai, Y., Wong, C. C., Wang, J., Wang, N., Lau, H. C.-H., Jiao, Y., Zhou, X., Ye, L., Mo, M., Yang, T., Fan, M., Song, L., Zhou, H., Zhao, Q., Chu, E. S.-H., Liang, M., … Yu, J. (2025). Gut–liver translocation of pathogen Klebsiella pneumoniae promotes hepatocellular carcinoma in mice. *Nature Microbiology*, *10*(1), 169–184. https://doi.org/10.1038/s41564-024-01890-9

Weiss, S., Amir, A., Hyde, E. R., Metcalf, J. L., Song, S. J., & Knight, R. (2014). Tracking down the sources of experimental contamination in microbiome studies. *Genome Biology*, *15*(12), 564. https://doi.org/10.1186/s13059-014-0564-2

Wirbel, J., Pyl, P. T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J. S., Voigt, A. Y., Palleja, A., Ponnudurai, R., Sunagawa, S., Coelho, L. P., Schrotz-King, P., Vogtmann, E., Habermann, N., Niméus, E., Thomas, A. M., Manghi, P., Gandini, S., … Zeller, G. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nature Medicine*, *25*(4), 679–689. https://doi.org/10.1038/s41591-019-0406-6

Wu-Woods, N. J., Barlow, J. T., Trigodet, F., Shaw, D. G., Romano, A. E., Jabri, B., Eren, A. M., & Ismagilov, R. F. (2023). Microbial-enrichment method enables high-throughput metagenomic characterization from host-rich samples. *Nature Methods*, *20*(11), 1672–1682. https://doi.org/10.1038/s41592-023-02025-4

Zeng, W., Wang, Y., Wang, Z., Yu, M., Liu, K., Zhao, C., Pan, Y., & Ma, S. (2023). Veillonella parvula promotes the proliferation of lung adenocarcinoma through the nucleotide oligomerization domain 2/cellular communication network factor 4/nuclear factor kappa B pathway. *Discover Oncology*, *14*(1), 129. https://doi.org/10.1007/s12672-023-00748-6

Zhang, X., Yu, D., Wu, D., Gao, X., Shao, F., Zhao, M., Wang, J., Ma, J., Wang, W., Qin, X., Chen, Y., Xia, P., & Wang, S. (2023). Tissue-resident Lachnospiraceae family bacteria protect

against colorectal carcinogenesis by promoting tumor immune surveillance. *Cell Host & Microbe*, *31*(3), 418-432.e8. https://doi.org/10.1016/j.chom.2023.01.013

Zhou, H., Liao, J., Leng, Q., Chinthalapally, M., Dhilipkannah, P., & Jiang, F. (2023). Circulating Bacterial DNA as Plasma Biomarkers for Lung Cancer Early Detection. *Microorganisms*, *11*(3), 582. https://doi.org/10.3390/microorganisms11030582