„BioTechNan – Program Interdyscyplinarnych Środowiskowych Studiów Doktoranckich KNOW z obszaru Biotechnologii i Nanotechnologii”

Hirszfeld Institute of Immunology and Experimental Therapy
Polish Academy of Sciences

Viral and bacterial components of the human microbiome: profiling in selected gastrointestinal disease states

Aleksander Szymczak

Dissertation completed in Laboratory of Phage Molecular Biology
Hirszfeld Institute of Immunology and Experimental Therapy
Polish Academy of Sciences
Wrocław, Poland
Supervisors: Prof dr hab. Krystyna Dąbrowska,
Prof. dr hab. Wojciech Witkiewicz

Wrocław, 2022

Projekt współfinansowany przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Wirusowe i bakteryjne składniki mikrobiomu ludzkiego: profilowanie w wybranych stanach chorobowych przewodu pokarmowego.

The results presented in this dissertation have been partially published in the publication:

- Szymczak A, Ferenc S, Majewska J, Miernikiewicz P, Gnus J, Witkiewicz W, Dąbrowska K. Application of 16S rRNA gene sequencing in *Helicobacter pylori* detection. PeerJ. 2020 May 13;8:e9099. doi: 10.7717/peerj.9099. PMID: 32440373; PMCID: PMC7229771.

The results presented in this dissertation were presented at the following scientific conferences:

- Transfer between ecological niches within human body: how gut phageome expands to serum phageome Szymczak A, Gnus J, Ferenc S, Miernikiewicz P, Majewska J, Gembara K, Strapagiel D, Stączek P, Dąbrowska K, Viruses of Microbes 2022, Abstract Book, p 167
- "Correlation between SNPs in human genomes and bacterial and viral complexity of human microbiome" Szymczak A, Gembara K, Ferenc S, Strapagiel D, Majewska J, Miernikiewicz P, Lach J, Gnus J, Stączek P, Witkiewicz W, Dąbrowska K, EMBL Symposium: New Approaches and Concepts in Microbiology 7-9.06.2021, Online Abstract Book
- „Comparison of human stomach microbiome and virome fractions" Szymczak A, Ferenc S, Miernikiewicz P, Majewska J, Gnus J, Witkiewicz W, Dąbrowska K Oxford Bacteriophage Conference – Phages 2019, 11-12.09.2019, Oxford UK, Book of Web p. 36
- "Microbiome comparison of Helicobacter pylori infected and non-infected patients" Szymczak A, Majewska J, Ferenc S, Miernikiewicz P, Witkiewicz W, Dąbrowska K. Microbiome Microbiome comparison of Helicobacter pylori infected and non-infected patients. 8th International Weigl Conference, 26-28.06.2019, Łódź, Abstract Book p. 130

The research presented in this dissertation was carried out in part as part of the project:

- Doctoral thesis carried out under the program of „BioTechNan – Program Interdyscyplinarnych Środowiskowych Studiów Doktoranckich KNOW z obszaru Biotechnologii i Nanotechnologii”
- National Science Centre in Poland grant OPUS 15 UMO-2018/29/B/NZ6/01659.

Podziękowania

Na powstanie i finalny kształt niniejszej pracy doktorskiej miało wpływ wiele osób, którym chciałbym szczególnie podziękować w tym miejscu.

Chciałbym podziękować prof. dr hab. Krystynie Dąbrowskiej za pomoc, motywację do prowadzenia przeze mnie badań i wsparcie merytoryczne.

Dziękuję Prof. dr hab. Wojciechowi Witkiewiczowi za cenne wskazówki, oraz całą pomoc udzieloną mi w trakcie trwania studiów doktoranckich.

Pragnę również podziękować wszystkim Koleżankom i Kolegom z którymi pracowałem w Instytucie Immunologii i Terapii Doświadczalnej PAN oraz Ośrodku Badawczo-Rozwojowym Wojewódzkiego Szpitala Specjalistycznego.

Szczególne podziękowania kieruję do całego zespołu Laboratorium Biologii Molekularnej Bakteriofagów za pomoc w pracy laboratoryjnej oraz wspólnie spędzony czas.

Na koniec chciałbym również podziękować moim Najbliższym: Dziadkom, Rodzicom za motywację do dążenia do swoich celów.

Nade wszystko chciałbym podziękować mojej Żonie Justynie oraz Czarkowi. Dziękuję za cierpliwość, dobrą radę oraz nieocenione wsparcie w najtrudniejszych momentach powstawania tej pracy. Bez Was ta praca nie zostałaby nigdy ukończona.

Table of contents

1. Abbreviations:

1. ADRB2 - Adrenoceptor Beta 2
2. AMMS – Asseco Medical Management Solutions
3. APS - Adenosine-5'-phosphosulfate
4. ATP - Adenosine triphosphate
5. AURKA – Aurora Kinase A gene
6. BAM – Binary Alignment Map
7. BCF - BIM Collaboration Format
8. BLAST – Basic Local Alignment Search Tool
9. Bp – Base pairs
10. BWA – Burrows – Wheeler Aligner
11. CACNA1A - calcium voltage-gated channel subunit alpha1 A
12. CCDC33 - Coiled-Coil Domain Containing 33
13. CDKAL1 - CDK5 Regulatory Subunit Associated Protein 1 Like 1
14. CEO – Chief Executive Officer
15. CFTR - Cystic Fibrosis Transmembrane Regulator
16. CI – Confidence Interval
17. DAMP – Damage associated molecular patterns
18. ddNTP - dideoxy-Nucleoside triphosphates

19. DNA – Deoxyribonucleic Acid
20. ELANE - neutrophil elastase
21. FUT2 - alpha-1, 2-L- fucosyltransferase
22. GHRL - Ghrelin And Obestatin Prepropeptide
23. GI – Gastrointestinal Tract
24. GRCh38 - Genome Reference Consortium Human Build 38
25. GWAS – Genome-wide association studies
26. HIIET – Hirszfeld Institute of Immunology and Experimental Therapy
27. HMP – Human Microbiome Project
28. IBD – Inflammatory Bowel Disease
29. IL – Interleukina
30. ISP – Ion Sphere Particles
31. ICD - International Classification of Diseases
32. Kb – kilobases
33. LCT – functional lactase gene
34. LINC01192 - Long Intergenic Non-Protein Coding RNA 1192
35. LPS – lipopolysaccharide
36. LTA – Lymphotoxin Alpha
37. MADCAM1 - Mucosal Vascular Addressin Cell Adhesion Molecule 1
38. MTHFR – Methylenetetrahydrofolate
39. MTTP - Microsomal Triglyceride Transfer Protein
40. MUC1 - Mucin 1, Cell Surface Associated
41. NCBI - National Center for Biotechnology Information
42. NF-kB - Nuclear factor- Kappa Beta
43. ng – nanograms
44. NGS – Next Generation Sequencing
45. NOD1 - Nucleotide Binding Oligomerization Domain Containing 1
46. OT2 – One Touch 2
47. OTU – operational taxonomy units
48. PBS - Phosphate buffered saline
49. PC – Principal Component

50. PCA – Principal Component Analysis
51. PCR – Polymerase Chain Reaction
52. PGLYRP4 – Peptidoglyvan recognition protein 4
53. PGM – Personal Genome Machine
54. pH – potential of hydrogen
55. RDP – Ribosomal Data Project
56. RNA – Ribonucleic Acid
57. RPM – Revolutions per minute
58. rRNA – Ribosomal Ribonucleic Acid
59. SAM – Sequence Alignment Map
60. SLC2A9 - Solute Carrier Family 2 Member 9
61. SLIT3 - Slit Guidance Ligand 3
62. SNP - Single Nucleotide Polymorphism
63. TLR – Toll like receptor
64. TMEM91 - Transmembrane Protein 91
65. TNF – Tumor Necrosis Factor
66. TSPAN8 – Tetraspanin 8
67. ul – microliters
68. US – United States
69. UV – Ultraviolet
70. VCF – Variant Call Format
71. VDR – Vitamin D Receptor
72. VLP – Virus-like proteins
73. ZNF – Zinc Finger Protein

2. Used Materials:

1. 0,22 µm syringe filters (Millipore)
2. 0.2 ml Tubes, flat cap (Thermofisher Scientific)
3. 1.5 ml Eppendorf Safe-Lock Tubes (Eppendorf)
4. 16S Metagenomics Kit (Thermofisher Scientific)
5. 2 ml Eppendorf Safe-Lock Tubes (Eppendorf)

6. AMPure XP Reagent for PCR purification (Beckman Coulter)
7. $CaCl_2$ (Sigma-Aldrich)
8. Ep. T.I.P.S. Pipette Tips 10,20-200,100-1000 (Eppendorf)
9. Ethanol 96% (Sigma-Aldrich)
10. Genomic Micro AX Blood Gravity (A&A Biotechnology)
11. GenomiPhi V2 DNA Amplification Kit (Cytiva Life Sciences)
12. Illumina DNA Prep [Nextera DNA Flex before 2021] (Illumina)
13. Ion 316 Chip Kit v2 BC (Thermofisher Scientific)
14. Ion 318 Chip Kit v2 BC (Thermofisher Scientific)
15. Ion Ampliseq Library Kit Plus (Thermofisher Scientific)
16. Ion PGM Hi-Q View OT2 Kit (Thermofisher Scientific)
17. Ion PGM Hi-Q View Sequencing Kit (Thermofisher Scientific)
18. Ion PGM Wash 2 Bottle Kit (Thermofisher Scientific)
19. Ion Plus Fragment Library Kit (Thermofisher Scientific)
20. Ion Universal Library Quantification Kit (Thermofisher)
21. Ion Xpress Barcode Adapters 1-16 Kit (Thermofisher Scientific)
22. Microseal PCR plate Seal (Bio-rad)
23. Miseq Reagent Kit V3 (Illumina)
24. Micro Beat Bead Gravity AX kit (A&A Biotechnology)
25. NaOH (Sigma-Aldrich)
26. Nextera DNA CD Indexes - 96 indexes, 96 samples (Illumina)
27. NextSeq 500/550 Mid Output kit (Illumina)
28. PCR-plates, 96 well, non-skirted (Thermofisher Scientific)
29. PhiX sequencing Control V3 (Illumina)
30. Phosphate buffered saline (Sigma-Aldrich)
31. QuantiFluor ds DNA system 1 ml (Promega)
32. Qubit dsDNA High Sensitivity Kit (Thermofisher Scientific)
33. Sherlock AX (A&A Biotechnology)
34. Sterile Water for Injection (USP)
35. Tris-HCl buffer, pH 8.0 (Sigma-Aldrich)
36. TUBE PA Easy Seal ultracentrifugation tubes PK/50 (Thermofisher Scientific)

3. Used Devices:

1. CFX 96 Real-Time PCR system (Bio-Rad)
2. Ion Torrent PGM sequencing machine (Thermofisher)
3. Ion OneTouch 2 System (Thermofisher)
4. Ion PGM Torrent Server (Thermofisher)
5. Denovix DS-11 Spectrophotometer (Denovix)
6. ABI Applied Biosystems 9902 Veriti PCR Thermal Cycler (Applied Biosystems)
7. Eppendorf Mastercycler Nexus 2XE (Eppendorf)
8. Thermomixer C (Eppendorf)
9. NextSeq 550 sequencing system (Illumina)
10. MiSeq system (Illumina)
11. Centrifuge 5424R (Eppendorf)
12. Centrifuge 5810 (Eppendorf)
13. Vortex-Genie 1 (Scientific Industries)
14. Roller Digital shaker (IKA)
15. Quantus Fluorometer (Promega)
16. Qubit 2.0 (Thermofisher Scientific)

4. Bioethical statement

Biological material used in descripted experiments was taken by physicians working in Regional Samples were collected in accordance with the principles of good clinical practice and the Declaration of Helsinki. Consent no. KB/nr 8/rok 2017 has been given by Bioethical commission in Research and Development Center, Regional Hospital in Wroclaw.

5. Index of Tables

- Table 2. Script in Python 3.6 used for data cleaning and Kraken2 screening analysis for Paired End Reads from Illumina sequencing.
- Table 3. Bash script used for converting FASTQ files derived from IonTorrent PGM sequencer to Annotated Variant Call Format (VCF) files for further statistical analysis. VCF format is required as an input for further bioinformatic operations. It contains information about identified SNPs and their positions in the investigated genome.
- Table 4. Script used for merging 16S rRNA data from kraken2 reports. Jupyter lab software was used to perform all actions in cited scripts. For better readability for the purposes of this document, the individual cell results have been omitted.
- Table 5. Script used for merging VCF format files. Jupyter lab software was used to perform all actions in cited scripts. For better readability for the purposes of this document, the individual cell results have been omitted.
- Table 6. An individual script created for identification of organisms' taxonomy based on taxomizr package. The script automates annotation based on taxids presented in output BLAST files. Finally, data derived from FASTQ files were saved as csv and proceeded for statistical data analysis. All data were normalized against total number of reads in each sample
- Table 7. Shannon Wiener Index equation transformed into Python function to calculate values for specified samples
- Table 8. A script used to implement function for PCA analysis on csv generated from 16S rRNA, phageome, and SNPs identification. R in 3.1 version was used.
- Table 9. Individual function to implement Cramer's V test for nominal variable calculation in Python 3.6
- Table 10. ICD-10 classified diseases diagnosed by physicians during endoscopy examination of patients enrolled to this study
- Table 11. Correlations between bacterial species found in the microbiome composition in patients and ICD classified illnesses diagnosed among them. P-value was calculated using z-test function from statsmodels Python 3.6 library
- Table 12. Correlations between bacteriophage genera found in the microbiome composition in patients and ICD classified illnesses diagnosed in these patients. P-value was calculated using z-test function from statsmodels Python 3.6 library.

- Table 13. SNPs that correlate with significant change in bacterial microbiome composition measured by Shannon-Wiener Index. T-test was performed using CI = 9% and $p<0.05$.
- Table 14. Correlations between presence of selected bacterial species associated with gastritis and diversity in patients' microbiomes. P-value calculated using t-test.
- Table 15. Concomitance of variants and composition of bacteriophages' part of microbiome. Estimation was obtained using collections library in Python 3.6.
- Table 16. Specific alleles related to increased frequency of *Helicobacter pylori* presence in stomach. Table created using *counter* Python 3.6 library on Ampliseq sequencing data and bacterial microbiome composition features. Cramer's V test was used to confirm that presence of *H. pylori* is correlated with above variants with $p<0.01$.
- Table 17. Alleles related to bacteria related to increased frequency of *Rothia mucilaginosa*. Table created using *counter* Python 3.6 library on Ampliseq sequencing data and bacterial microbiome composition features. Cramer's V test was used to confirm that presence of *R. mucilaginosa* is correlated with above variants with $p<0.005$.
- Table 18. Alleles related to bacteria related to increased frequency of *Prevotella melanonigica.* Table created using *counter* Python 3.6 library on Ampliseq sequencing data and bacterial microbiome composition features. Cramer's V test was used to confirm that presence of *P. melanogenica* is correlated with above variants with $p<0.01$.
- Table 19. Alleles related to bacteria related to increased frequency of *Neisseria subflava.* Table created using *counter* Python 3.6 library on Ampliseq sequencing data and bacterial microbiome composition features. Cramer's V test was used to confirm that presence of *N. subflava,* is correlated with above variants with $p<0.01$.
- Table 20. Alleles related to bacteria related to increased frequency of *Prevotella jejuni.* Table created using *counter* Python 3.6 library on Ampliseq sequencing data and bacterial microbiome composition features. Cramer's V test was used to confirm that presence of *P. jejuni* correlated with above variants with $p<0.01$.

6. Index of Figures

- Figure 6. An example reports for Illumina sequencing generated using FASTQC software. Panel (A) – distribution of quality scores among generated reads. Panel (B) – distribution of Phred quality scores among single base pairs in generated reads.
- Figure 7. SNPs which correlate with significant change in Shannon-Wiener diversity index. Significance was calculated with Confidence Interval = 95% for $p<0.05$ using t-test. Results for samples with an identified single nucleotide allele are marked as "variant". Results for samples without an identified single nucleotide allele are marked as "Reference".
- Figure 8. Bacteriophages presence which correlates with significant change in Shannon-Wiener diversity index. Significance was calculated with Confidence Interval = 95% for $p<0.05$ using t-test. Results for samples with specified bacteriophage are marked as "present". Results for samples without specified bacteriophage are marked as "absent".
- Figure 9. Two-dimensional biplot showing the relationship between the different genera of bacteria, which affects their clustering in PCA analysis. Patients with complete phage and bacterial microbiome composition data as well as sequencing SNPs were used for analysis. Analysis was done using bacterial microbiome composition data.
- Figure 10. Biplot visualizing dimensions that represent bacteriophage genera derived from metagenomics sequencing and that have the highest influence on PCA plots. Patients with complete phage and bacterial microbiome composition data as well as sequencing SNPs were used for analysis. The presence of some bacterial groups in this Figure is due to their annotation in the NCBI database: some of the bacteriophages are prophages for instance: *Haemophilus, Streptomyces,* thus annotated as a part of bacterial genomes.
- Figure 11. Principal Component Analysis applied to different Operational Taxonomy Levels in bacterial microbiome. Plot created using ggplot2 library in R. Panel (A) shows samples clustered by bacterial family's microbiome composition, panel (B) shows samples clustered by bacterial genus microbiome composition, panel (C) shows samples clustered by bacterial species.

- Figure 12. Principal Component Analysis applied to different Operational Taxonomy Levels in phageome. Plot created using ggplot2 library in R. Panel (A) shows samples clustered by bacteriophages' species microbiome composition, panel (B) shows samples clustered by phages' genus microbiome composition.
- Figure 13. Scree plots generated to evaluate explained Variance among calculated Principal Components. Data generated using ggplot2 R' package. Panel (A) present scree plot for bacteriophages' microbiome composition while Panel (B) presents scree plot for bacterial microbiome composition.
- Figure 14. Principal Component Analysis applied to bacterial microbiome composition data. Plot created using ggplot2 library in R. Panel (A) present data generated basing on bacterial composition of microbiome while panel (B) shows data for Bacteriophages' part.
- Figure 15. Principal Component Analysis applied to bacterial microbiome composition data and grouped by presence of pathogens related to gastritis. Figure created using ggplot2 library in R. Panel (A) present visualization of presence of H. pylori, Panel (B) is dedicated to *N. subflava* presence, Panel (C) is for *P. melaninogica*, Panel (D) shows data related to *R. mucialignosa*.
- Figure 16. 3D Principal Component Analysis plotted based on phageome components of human microbiome. individual items have been labeled according to whether specific chronic illness was diagnosed. Figure was created using pca3d package in R. Panel (A) shows bacteriophage composition data with labeled patients with claimed hypertension. Panel (B) shows bacteriophage composition data with labeled patients with claimed diabetes.
- Figure 17. 3D Principal Component Analysis plotted based on bacterial components of human microbiome. individual items have been labeled according to whether specific chronic illness was diagnosed. Figure was created using pca3d package in R. Panel (A) shows bacterial composition data with labeled patients with claimed hypertension. Panel (B) shows bacterial composition data with labeled patients with claimed diabetes.
- Figure 18. Bacterial components microbiome profiling based on 16S rRNA sequencing belonging to sample "E35". Figure created using Krona Tools.

7. Index of equations

8. Introduction

8.1. Human microbiome concept

Once upon a time researchers decided to understand better bacteria that inhabit human’s body. They decided to run Human Microbiome Project (HMP) (Peterson et al., 2009), to try to comprehend composition of microorganisms that can be found in different parts of body. Initiative was taken by four research centers in the US: University of Maryland, Berkeley Lab, Joint Genome Institute, and University of Colorado Boulder that have obtained funding from Common Fund support by National Institute of Health. Two phases of the program were planned. From 2008 till 2013, they collected samples from 300 individuals, including 5 parts of their bodies: nasal cavity, oral cavity, skin, gastrointestinal tract (GI) and urogenital tract. Initially, researchers had only 5 objectives to complete. They wanted to create a reference set of minimum 3000 microbiota genome sequences, to reveal the complexity of the bacterial community among body sites where sample were taken. They wanted to find whether changes

in human microbiome composition may have an effect on occurrence of diseases, to develop new workflows for computational data analysis, and setting up procedures to examine human microbiota in accordance with ethical principles. An interview was conducted with all sample donors to evaluate their health status. As a result, researchers expected to find correlations between characteristic microorganisms' communities and the declared well-being. Two different technological approaches were used to obtain such information. 16S rRNA gene sequencing was used to characterize bacterial part of metagenome isolated from selected niches (Prodan et al., 2020). This gene encodes a small subunit of bacterial ribosomes. All bacteria possess a characteristic variant of this gene. 16S rRNA coding regions contain short conserved fragments common to all groups and hypervariable regions (Kang et al., 2021) Sequencing of these hypervariable regions allows to assign bacteria up to their species. The second technological approach applied by the consortium was whole genome shotgun sequencing which provided extensive information about metabolism pathways, antibiotic resistance genes, presence of fungi, and both prokaryotic and eukaryotic viruses (Castelino et al., 2017; Ranjan et al., 2016).

This type of study was possible due to the recent advance of DNA sequencing methods. Recently worked out milestone called "Next-Generation Sequencing" (NGS) revolutionized the approach for identification of genetic information preserved in both prokaryotic and eucaryotic cells as well. Before developing high throughoutput sequencing like NGS, it was difficult to study microorganisms which were unable to grow in laboratory conditions. With the ability to study entire bacterial fractions using DNA sequencing, completely new species and metabolic pathways that were previously unknown were discovered.

Phase 1 of the Human Microbiome Project ended in 2013, yielding with terabytes of data which are being a subject of analyses to present days. Due to outstanding results, the consortium proceeded to the next phase. More and more questions arose during the project. It became obvious that data required deeper analysis. Scientists involved in HMP decided to create integrative HMP (iHMP). The new approach included creation of a database, which contained metagenome data that can be linked to current host health status described for each individual. Easily accessible datasets were published and they became a unique source of information for scientists all over the world. As a consequence of phase 2 HMP, three

clinical studies were performed, which included host phenotype – microbiome interactions. Investigators analyzed association of preterm births to vaginal microbiome of mothers (Vinturache et al., 2016), presence of inflammatory bowel disease (IBD) to composition of gastrointestinal tract microbiota (Zuo & Ng, 2018), and diabetes 2 was correlated to nasal and GI tract microbiomes (Kumpitsch et al., 2019). As it has been stated on official HMP website, second phase was ended in 2016. From the beginning in the year 2008, to the year 2017, over 50 papers resulting from HMP were published in top ranked scientific journals. It was a huge booster for microbiome research not only in the US, but all over the world. During almost 10 years, NIH supported microbiome research with more than $1 billion of which HMP accounted for 215 million. Sixty-five Principal Investigators have been selected to manage tasks under this project. Estimates suggest that the second amount was spent outside the United States on similar studies aimed at defining microbiome relationships. What is more, only in 2020 over 1$ Billion grants were awarded to microbiome related research and 75 000 papers were published in that scientific area. Despite crossing the billion-dollar barrier, this is not a record amount. The largest support spent so far on microbiome was in 2019, when funding for such projects reached $1.9 billion (Peterson et al., 2009; Ursell et al., 2012).

8.2. Dark ages of microbiota research

Basing on DNA sequencing development, it could be concluded that microbiome research started in early 2000s, but the fact is, it was born same time as microbiology. Antonie van Leewenhoek, known for constructing the first microscope, was the first unintentional microbiome researcher. In his notes in 1680 he reported observations of samples taken from mouth and those from stools that were completely different. It is the first documented investigation aimed to understand microorganisms living in human organism. Choice of sampling locations was dedicated to hypothesis that had formerly been posed: diseases were associated with bad smells coming from human body, so the researcher chose those emitting the worst. Unfortunately, van Leewenhoek did not develop any theory about specific pathogens causing the odors and health problems (Bardell, 1982; Leewenhoeck, 1684).

In the years 1888 – 1891 Sergei Winogradsky was working in Swiss Polytechnic Institute in Zurich. This Russian biologist spend time of his early research in botanist laboratory. He

studied *Mycoderma vini* which was responsible for a plague on sugar beets and nowadays it is known because of its ability to spoil wine. Windogradsky should be considered as the first microbiologist focused on environmental microbiome composition. He stated that there was a necessity to observe bacteria in their natural settings. In his publication in the year 1949, he claimed that microbes acted differently in natural habitats than in laboratory prepared growth media. He developed a tool named "Winogradsky column". It was a simple device which for the first time allowed to culture wide range of microorganisms in conditions similar to natural environment (Ackert, 2007; Dworkin & Gutnick, 2012).

Breakthrough for microbiology and human microbiome research was the work of Louis Pasteur and Robert Koch. These chemist and physician from France and Germany (respectively), have brought the world new tools to discover the world of microbes. Staining techniques optimized by Koch are used to present days in microscopic observations. One of his assistants called Julius Petri designed shallow dish named after him. This dish is the major symbol of microbiology widely recognizable and used by researchers on daily basis for culturing bacteria. The other one, Walther Hesse, developed the first solid medium. That time, scientific world was aware of the existence of microorganisms that inhabitant the human body. However, specific identification of these species was too difficult (Spichler et al., 2015; Surana & Kasper, 2014).

In 1901, Eli Metchnikoff who was a Russian biologist focused on zoology and immunology presented a concept that use of lactic acid bacteria can improve condition of human gastrointestinal tract and it is the key to a long and healthy life. That was the moment when the concept of "probiotics" was presented and when a market of probiotics, currently estimated for more than $34 billion, was born. His hypothesis was published in his "The Prolongation of Life: Optimistic Studies" (Metchnikoff, 1907); to reinforce the idea he was supporting, he himself drank a glass of sour milk every day. He stated that longevity of Bulgarian peasants resulted from their increased yogurt consumption. Unfortunately, the concept of probiotics was discontinued until the end of XX century, when modern microbiological research confirmed some of his statements (Mackowiak, 2013; Podolsky, 2012).

Soon, in 1909, Arthur I Kendal, a bacteriologist living in the US, demonstrated effects of diet on the communities of microorganisms present in intestinal tract in monkeys and further

implications for the animals' health (Kendall, 1909). Development of germ-free animal models in 1950s can be considered as the first milestone in modern microbiome research. It opened the new way to examine host-microbe interactions. Rene Dubos from Rockefeller Institute for Medical Research was the first who observed that pathogenicity of bacteria was dependent on external conditions. He confirmed the theory presented before by Arthur I Kendall. Dubose's team provided information about correlation between diet, stress, and microbes living in human body determining human health status(Aziz, 2009).

Because of the technical limitations in microbiological research that could only investigate culturable bacteria, sequencing techniques designed in 1980s opened a completely new possibilities in microbiome research. Carl Woese and George E. Fox were microbiologist who set foundation for further big-scale metagenomic analysis. Their concept of "tree of life" based on molecular markers is used to present day for characterizing newly discovered organisms (Woese & Fox, 1977). Nevertheless, the true revolution of microbiome research was about to come.

8.3. Microbiome that surrounds human

Ribosomal Data Project (RDP) aims to collect ribosomal sequences that have been identified around the world. Bacterial and archeal 16S rRNA sequences, as well as fungal 28S rRNA sequences collected within RDP are crucial for taxonomy identification of sequenced microbiomes. These genes consist of hypervariable regions flanked by conserved fragments. It makes them perfect candidates for highly efficient targeted sequencing. In 2012, RDP contained around 1 million of documented oligonucleotides. Unfortunately, only 5% of them were not derived directly from human microbiomes (Vasileiadis et al., 2012) Researchers however started identification of microbial communities from sites where according to previous assumptions living organisms including microbes could not exist, for instance from thermal geysers, volcanoes, or ocean depths. New technological solutions for detection of whole microbial communities have made it possible to describe the variability of bacteria found across continents (Prodan et al., 2020).

An exemplary experiment was conducted by Lauber et al.(Lauber et al., 2009). They collected soil samples from 88 sites in the Americas. Using a pyrosequencing method, they identified 16S rRNA gene fragments belonging to bacteria existing there, which allowed for detection of

25 phyla across all samples and for demonstration that pH can be a good predictor for microbial soil composition (Lauber et al., 2009). Another approach defined bacteria that were living in Hot Springs placed in Yellowstone. For a long time, this environment was considered as uninhabited because of the extreme conditions there. Characteristic communities of bacteria living in this place have been identified and new relationships between different groups of microorganisms have been demonstrated (Miller et al., 2009). NGS identification of 16S rRNA enabled H Song group and coworkers to examinate biodiversity in fresh volcanic ash. Diverse microbiota was observed, in spite of extreme environmental conditions (Song et al., 2020).

The most abundant communities of microorganisms can be found in water that covers 71% of surface on planet Earth. Studies of bacterial samples from the deepest parts of the ocean were impaired, since it was impossible to create adequate laboratory conditions to imitate natural growth conditions of microorganisms. NGS become a solution allowing studies of the unculturable bacterial communities, including those in the deep ocean.

Samples from various water layers were collected in a place with the deepest water on earth – Marina Trench. Surprising ecological diversity was discovered, increased in deeper water layers (>4000 m). While surface waters were occupied mostly by Cyanobacteria, lower parts turned out to be rich in microorganisms that produce proteins involved in degrading nitrate to ammonia and urea transport (Xue et al., 2020).

Recently, it has become a standard to upload raw sequences for every published study, including microbiome research. Big amounts of data generated by NGS are publicly available in open databases portals. Accessible data from defined metagenomes are a great source of sequences for further analyses (Zaparucha et al., 2018). Environmental microbiomes are not only containing prokaryotes and eukaryotes. The most abundant part are viruses, for instance up to 10^{10} of viruses in 1 gram of soil. Vast majority of bacteria are infected by bacteriophages that strongly contribute to microbial diversity in the environment (Kuzyakov & Mason-Jones, 2018).

8.4. Sanger revolution

First DNA sequences were identified in 1970s with the use of two-dimensional chromatography. First DNA sequencing method was optimized by Ray Wu in 1968 when he

reported the first partial sequence of lambda DNA bacteriophage. In 1973 Walter Gilbert and Allan Maxam annotated 24 base pairs using their original method to know *lac* operator in *E. coli*. Nevertheless, breakthrough came with Frederick Sanger. His method known as "chain termination" turned out to be a gold standard of DNA sequencing for a long time. It was known that DNA polymerase uses dNTPs to build complementary strand during amplification process. In 1977 his group optimized a method based on 4 manual Polymerase Chain Reactions (PCR). The process was fast and it was impossible to control with systems available those days. Sanger overcame that problem by developing a new way of termination of amplification process in the desired short time. According to four bases that build DNA they prepared 4 different PCR reactions. Each of them contained a small amount of single type of dideoxy-Nucleoside triphosphates (ddNTP) representing Adenine (ddATP), Thymine (ddTTP), Guanine (ddGTP) and Cytosine (ddCTP). Once ddNTP was fused into the DNA strand, no other base could be fused and reaction was terminated. This step was followed by gel electrophoresis. Samples from all four reactions were placed in separate lines in the gel. DNA fragments migrate in the gel in a electric field with a speed dependent on their mass. This phenomenon makes it possible to relate length of DNA fragments to specific type of ddNTP that terminated PCR reaction (Heather & Chain, 2016a; Shendure et al., 2017). Frederick Sanger is the most famous scientist in the field of DNA sequencing who got Nobel Prize in 1980 for his "dideoxy" method.

Human Genome Project was an idea to reveal the sequence of human genome and it employed the Sanger method. Officially, in 1990 the program has been launched. News headers were announcing solving the most complex mystery that humanity ever had, but in the end, researchers knew that it was only a praeludium for a multitude of further projects. Even though only euchromatic regions of human genome were sequenced, completion of the program was announced in 2003. Approximately 92% of the genome was assembled and eventually more data were acquired than ever before. Soon it became obvious that more robust, faster, and what is the most important, cheaper techniques will be needed. Approximately 3 billion base pairs cost about $2.7 billion in the time of target completion. Today, cost of sequencing is much lower than 1000$ per genome and it takes no more than 24h. However, we are approaching the fifty fifth anniversary of the first report on DNA sequencing (Heather & Chain, 2016a).

The main disadvantage of Sanger sequencing is the use of electrophoretic separation for identification of molecular mass of DNA fragments. This method is efficient for particles with relatively low masses. It is however much more complicated in longer DNA fragments, where a single base does not represent a significant part of the whole mass of the whole molecule. For this reason, sequencing of DNA particles longer than 700 base pairs with this method is difficult(Heather & Chain, 2016b). This technique has however been used in laboratories for almost 25 years, until 2000s when the second-generation sequencing technologies (called also Next-Generation Sequencing) replaced it. It does not mean that this technique has been abandoned. Sanger pattern of DNA sequencing is still used in novel tools designed by engineers. Modern techniques parallelized PCR reactions that were previously prepared manually. The development of variations derived from Sanger sequencing has resulted in a whole branch of techniques named as "Sequencing by synthesis". All of them are based on PCR amplification termination and reading base-dependent signals (Heather & Chain, 2016a).

8.5. Pyrosequencing

In 2004, 454 sequencers developed by Life Sciences from Branford in Connecticut presented a revolutionary instrument that enabled parallel sequencing for thousands of DNA molecules at the same time. They based their technology on pyrosequencing, which is an extension of the "Sequencing by synthesis" optimized by Sanger. It was developed as early as in 1996, but had not enter commercial use before 2004 (Shendure et al., 2017). In this method, the signal from a specific nucleotide is detected through chemiluminescence. It allowed for the first time to sequence many isolates derived from different individuals in one sequencing sample by the use of the barcoding method. That required DNA preparation by ligation of strands to short oligonucleotides(Heather & Chain, 2016b). Sequences of these short chains were defined, and then during bioinformatic analysis it was possible to distinguish oligonucleotides in one isolate from those in the others. Pyrosequencing begins with addition of one type of nucleotides. Polymerase incorporates them into single chains attached to the beads according to the principle of complementarity. Each nucleotide attachment is accompanied by the release of pyrophosphate to form a phosphodiester bond. This leads to a cascade reaction in the solution where ATP sulfurylase (Sulfate adenylyltransferase) converts pyrophosphate to ATP in the presence of APS (Adenosine-5'-phosphosulfate). The luciferin is also present in the reaction

mixture, and with ATP it is transformed by the enzyme luciferase with emitting the light, which is detected by the device. The intensity of the light is directly proportional to the amount of ATP produced. So called pyrogram shows the amount of light produced in time, which represents the number of nucleotides attached. The reaction mixture is then regenerated by the enzyme apyrase, which degrades ATP and unbound nucleotides. After this event, the cycle repeats with the addition of another type of nucleotide. Although the sequencers have been improved and now allow sequences of up to 400 bp in length to be read, the principle of the reaction and the operation of the sequencers remain unchanged. (Ahmadian et al., 2006; Nyrén, 2007; Ronaghi et al., 1998). .

8.6. Ion Semiconductor Sequencing

Almost 13 years ago Chris Toumazou, a founder and CEO of DNA Electronics developed a technology that made possible to sequence DNA on silicon-based chips. The company addressed the need of real-time sequencing. In 2010, this technology was licensed to other sequencing company called Ion Torrent Systems Inc. This tool was intended for fast, reliable, and portable gene sequencing in the real time. Physicochemical phenomena are standing behind semiconductor sequencing: during amplification, integration of dNTPs is accompanied by secretion of pyrophosphate and hydrogen ion that is charged positively. Semiconductor chip is filled with microwells that contain previously prepared single-stranded template DNA molecules. During sequencing, solutions containing (separately) A, T, C, or G bases, flow through microwells along with polymerase activation. When dNTP is incorporated and positively charged hydrogen is released, then the bottom layer of the chip detects this signal, since underneath the microwell surface, an ion sensitive layer is placed. Hydrogen ion release change the pH of the solution in microwells. This is detected by ion-sensitive transistors, and information is related to the type of nucleotide that has been used in the flow. Sequencer than washes out unincorporated dNTPs and the next flow is applied (Goswami & Sanan-Mishra, 2022; Gupta & Gupta, 2014).

At the beginning of this technology development, Ion Torrent sequencers allowed for 50 base pairs long reads with 100 Megabases detectable in one run. Recent machines that use this technology allow for 600 bp-long reads and about 50 Gigabases yield (10^9 base pairs).

Currently, the time consumed by sequencing depends on the size of the used chip (Cheng et al., 2013; Katsnelson, 2010; Y. Wang, Wen, et al., 2014).

What makes Ion Torrent technology unique is not only the way of pH measurement, but the process that precedes sequencing. Preparing of microwells requires a specific type of PCR called emulsion PCR. In this process, oil and microbeads are added to the reagent mixture in order to create billions of microbeads-containing water droplets separated by oil fraction (emulsion). These microbeads anchorage single strain DNA fragments. These fragments are amplified in emulsion PCR reaction to produce many thousands of amplified DNA molecules of the same type (sequence) on one bead. This large number of resulting readable strands will help to avoid reading errors using bioinformatics techniques. Prior to placing chips into the instrument, beads are applied on the chip, where only one bead covered by a single DNA strand of the same origin is supposed to be placed in one microwell. This is required for proper signal detection (Bertolini et al., 2015).

Reliability of the instrument was confirmed in 2014 when for the first time Ion Torrent Personal Genome Machine was approved by the U.S. Food and Drug Administration for clinical use. It enabled medical laboratories to implement Next-Generation Sequencing into their routine work. In the beginning of new era of sequencing, Ion Torrent was known for its advantages like rapidity and relatively low operating cost. Despite the fact that speed of sequencing is limited by the time of solution flow, sequencing can be monitored in real time. Unfortunately, ion conductor sequencing has significant problems when it has to deal with homopolymers. Then, in one cycle the change of the pH is greater than usual and it generates significantly stronger signal. The longer a homorepeat is, the more difficult to estimate its length. In some application the limited length of reads could be a problem as well, since even with the most advanced chemistry and chips it can only be approximately 600 bp. For this reason, calculating sequence coverage during experiment design is more challenging (Bertolini et al., 2015; Y. Wang, Wen, et al., 2014).

8.7. Sequencing by synthesis – Illumina variation

Illumina technology has its roots is in mid 1990s in the United Kingdom. That time scientists in Cambridge for the first time observed fluorescently dyed nucleotides. First, they wanted to

observe motion of the polymerase. Eventually they came up with an idea that their technology could be used in DNA sequencing, so they decided to form Solexa company in 1998. Soon, they brought their technology to the point where they were able to sequence a full genome of bacteriophage phiX-174 using only their instruments; this phage DNA is still used to calibrate sequencers based on this technology. Their first commercial instrument called "Solexa Genome Analyzer" was presented in 2006. It was a milestone in human genome sequencing. For the first time, in one experiment, researchers could obtain up to 1 Gigabase of data. Rapid success aroused the interest of other companies and in 2007 Illumina acquired Solexa. Since then, companies have advanced their technology to create the most popular DNA sequencing platform. In 2019, they presented their recent invention. Nova Seq, which is the most advanced sequencer when one takes into consideration data that could be generated by a single run. According to the manufacturer, it is able to generate up to 20 Terabases (10^{12}). It is 1666x more than the first pyrosequencer could generate ever. This resulted in a significant price reduction for the whole human genome sequencing. Now it can be lower that 1000$ (Farrer et al., 2009; J. Wu et al., 2012).

The technology of sequencing by synthesis based on fluorescent dyes remains same, but the range of parallelization has changed. The entire process of nucleotide detection in the Illumina sequencer takes place on a flow plate, to which short nucleotide sequences are pre-attached. In the first step, oligonucleotide adapters are attached by ligation to investigated DNA fragments, at both ends. DNA is denatured to single-stranded molecules with ligated adapters and then attached to short complementary sequences on the plate. Fragments that are not bound at this stage are washed off. Next, isothermal amplification of bound DNA is performed. Process consists of cyclic denaturation, attachment of 3'-strand ends to complementary oligonucleotide sequences on the plate, and strand extension. The 3' ends of the nucleic acids hybridize with adjacent short chains immobilized on the plate forming a characteristic bridge structure. The final step is to wash away the double-stranded DNA fragments and attach dideoxynucleotides to the 3' ends of the strand. This is to prevent the formation of non-specific bonds during the actual sequencing process (Heather & Chain, 2016b; Slatko et al., 2018).

Sequence by synthesis developed by Illumina is based on optical detection of identified nucleotides. It is different than Ion Torrent method which was based on pH changes

measurement. Each of the four types of nucleotides has its own unique fluorescent tag. Once attached, the fluorophore is excited by a laser and the specific signal is detected. These attached fluorescent tags also act as reaction terminators. As long as the fluorophore is not excited and dissociated from the nucleotide, it blocks the attachment of subsequent base pairs by its presence at the 3' hydroxyl position. This prevents mass attachment of nucleotides before the signal is read, and also ensures the synchronous sequencing of DNA fragment simultaneously on the flow plate. This reaction is performed repeatedly and simultaneously on all individual nucleic acid strands coated on the plate (Besser et al., 2018; Guo et al., 2010).

8.8. Mobility of sequencing by nanopore

Nowadays, next-generation sequencers have become popular, and they are used on daily routine in laboratory work. Nevertheless, they are still physically too big to be used in mobile laboratories where samples could not be transported to a regular lab. Concept of a portable sequencer was created early in 1989, but the technology at that time was not sufficient to make this idea a reality.

The first working nanopore sensor was described in Nature Nanotechnology in 2001 (J. Li et al., 2001). The first functional sequencing with nanopores was completed in March 2011. Soon, the first sequencer that fits into the pocket was presented by Nanopore – MinION (Wang, Wen, et al., 2014)

The technology is based on sequencing using nanopores. This solution does not use optical system for nucleotide identification. The use of such a solution made it possible to extend the length of readouts even to several thousand base pairs and to miniaturize the device to a weight of 90 grams. The device contains 2048 sequencing channels inside, each channel can sequence one continuous DNA strand at a time. Channels are formed with α-HL which is a membrane channel protein. Diameter of approximately 1.4 nm allows only single stranded DNA or RNA to be squeezed through the hole. DNA is translocated through pores and characteristic for each nucleotide shifts of ionic current is detected by sensors. The lengths, amplitudes, and variances of such changes are detected and parametrized. The software recognizes characteristic parameters for individual nucleotide sequences and it produces a two-dimensional diagram based on the parameters. This is currently the handiest sequencer that also allows to observe

the process in the real time able to read strands up to 50 kilobases (Kb) long and with 100 bp/s rate. Sample preparation comprise adding adapters that connect to the 3' and 5' ends of DNA fragments. They facilitate DNA capture and loading of the process helicase enzyme (Caldwell & Spies, 2017; Y. Wang, Yang, et al., 2014). This enzyme executes the unidirectional passage of the DNA molecule through the nanopores. This innovative method allows for recognition of nucleotide structures; thus, it is also possible to recognize methylated forms of nucleotides. With the use of this sequencer, read lengths of up to 50,000 base pairs can be obtained (Lu, Giordano, et al., 2016; Madoui et al., 2015; Mikheyev & Tin, 2014).

Today, space technology has also made significant advances. Since 1998, an international space station has been orbiting the Earth. Astronauts are constantly conducting research there under supervision of experts on the globe surface. One of the major areas of the research is monitoring conditions of microgravity. Microorganisms in the microgravity environment are also an important part of space research. Unfortunately, conducting biomonitoring of bacteria in space is difficult due to the limited space of the launch modules. A program called Microbial Tracking is conducted to assess the microbiological cleanliness of the space station modules, as well as the vehicles used to transport people there. DNA sequencing is used to monitor the emergence of drug resistance genes in bacteria (Avila-Herrera et al., 2020). What is more, DNA sequencing is one of a few experiments which can be performed by the astronauts in the space. In 2016, the first sequencer based on nanopore sequencing was placed in the vehicle targeted to International Space Station. Soon, astronaut Kelly Rubins for the first time performed DNA sequencing outside the planet Earth (Rainey, 2016).

Milestones in sequencing and microbiome research are presented on timeline in the Figure 1.

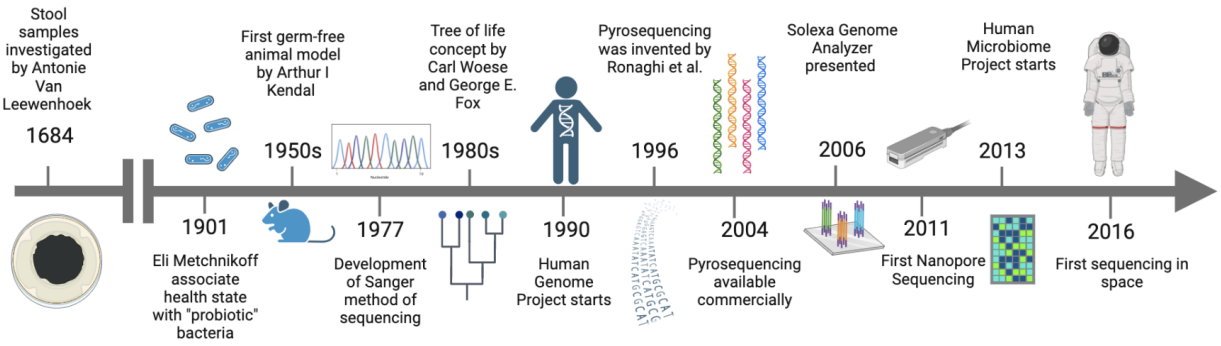

Figure 1. Timeline presents highlights of the sequencing and microbiome research. Figure was created using Biorender software.

8.9. Human microbiota diversity

Microbial Tracking program launched a few years ago by NASA revealed that most of microbiota which can be found in space vehicles and modules are most likely derived from people who have been working there (Avila-Herrera et al., 2020; Checinska Sielaff et al., 2019). The boom of discoveries on diversity and abundance of microbes living in human bodies have caused problems in nomenclature. There is also a need to discriminate between microbiome, as the whole community which consist of microbes and their genes, and microbiota, that is taxa that belong to microorganisms. However, in scientific literature, these terms are often used interchangeably (Peterson et al., 2009; Ursell et al., 2012).

It has been estimated that each human is a unique host for 10-100 trillion microorganisms. What is more interesting, shotgun sequencing revealed almost 3.3 million non-redundant genes only in human gut. Despite of coding regions that could be detected in other parts of body, this is over 150x more than entire human genome consist of (approximately 22000) (Abdellah et al., 2004; Qin et al., 2010). The vastness and complexity of the human microbiome has inspired the idea of considering the microbiota as just another human organ. Notably, microbial community in human body has secretory functions, protective functions against infections, immunomodulatory effects, and it may have impact even on daily mood(Baquero & Nombela, 2012; Clarke et al., 2014). One of the most significant discoveries regarding the human microbiome was an observation that it is so variable that a specific composition can be

attributed to a single individual. Small differences in diet, daily routines, cosmetics, and antibiotic usage can significantly change composition of microbes that reside in human body (Hasan & Yang, 2019).

A group of students from Harvard University created a code that was able to recognize around 80% of individuals only using microbial sequences identified in the individuals. Data used in this study was collected within HMP and the group used machine learning algorithms to create "microbiota fingerprints" and to identify them in a blank trial (Franzosa et al., 2015).

The bacterial kingdom is the first that comes to mind when thinking about the human microbiome. Other groups of microorganisms, like archaea, fungi and, most importantly, viruses are often underestimated. Population-level metagenomics studies identified that human gut microbiome consists of approximately 40% of archaea (Franzosa et al., 2015). Some researchers propose that properties of Archeones in the gut environment can be used in a similar way as we use probiotic properties of certain bacteria(Brugère et al., 2014; Horz, 2015). Moreover, human fungal microbiota, sometimes called "mycobiome", is much less abundant than the bacterial one (Shah et al., 2021; Tiew et al., 2020). Common colonization of human urogenital tract, oral cavity, and gastrointestinal tract with fungi is widely known. Most of them are *Candida* spp, that often escape immune control and cause infections. These strains mostly inhabit human gut or other parts of the body with mucosal surfaces. Usually, *Candida* spp. remain undetected and does not cause any pathological symptoms. This may change when the sites it inhabits become inflamed, or the body struggles with immunosuppression for an extended period (Fidel, 2002; Huffnagle & Noverr, 2013; Sogin et al., 2006).

Identification of microbiotic fungi is different to that of bacteria. They do not code for 16S rRNA genes, so identification can be done by amplification of Internal Transcribed Spacer region (ITS2) which is the genetic marker of fungi, or by shotgun metagenome sequencing with whole genome analysis (Donovan et al., 2018; Huffnagle & Noverr, 2013; van Tilburg Bernardes et al., 2020).

The least identified though the most abundant part of microbiota are viruses called virome. Importantly, viruses infecting human cells are not a major part of the virome. Human virome contains viruses that infect other microbial components such as bacteria (infected by bacteriophages) or archea, and viruses consumed with food (only temporarily in human body).

These viruses can be either RNA or DNA (double- or single-stranded). Their diversity includes both morphology, coding genes, and genome size (Liang & Bushman, 2021). Most viruses that can be found in human body belongs to bacteriophages (phages). Each bacteriophage infects only its specific bacterial host, which makes them an excellent weapon against bacterial pathogens that could break homeostasis – microbiota balance. About 2500 different virus-like particles were identified in the gut, but possibilities of robust identification are still very limited (Federici et al., 2020) Unfortunately, most of phage sequences are not annotated in databases, and thus they cannot be identified in metagenome screening studies. The major difficulty in identifying viral communities is the fact that they do not have a universal genetic marker such as 16S rRNA gene in bacteria, or ITS in fungi. It is necessary to perform whole genome sequencing and assembly to be able to define which virus has been detected. The other challenge results from phage ability to multiply in two separate cycles – lytic and lysogenic. Lysogenic phages can be present in microbiota, but 'hidden' in the form of prophages incorporated into bacterial genomes (Federici et al., 2020; Liang & Bushman, 2021).

8.10. Human microbiome diversity

Hypocrates claimed that "All diseases begin in the gut", and this statement can today be linked to microbiome. One of the major microbial communities in human body resides in the human gut. Microbiome communities inhabiting our body are not only diverse among human individuals, but human body provides various environments (niches) where conditions are different. It results in completely different composition of microbes that inhabit different body sites. Deep 16S rRNA gene sequencing revealed that bacteria play an important role in the proper functioning of organs where the presence of bacteria is not intuitive. For instance, such a distinctive organ as the human eye also has its own characteristic set of microorganisms. Conjunctival swabs and NGS sequencing delivered proofs that eye was inhabited with a vast number of microbes that had not been described before. Bacterial community is represented in that part of eye by core genera such as *Pseudomonas*, *Propionibacterium*, *Bradyrhizobium*, and others(Q. Dong et al., 2011). The presence of bacteria in the human urinary tract is an important parameter determined during the examination. Although the presence of bacteria in the bladder has been linked only to infections, it has been proven that their presence in the urogenital tract is normal. This part of the microbiome evolves with a person's age and lifestyle.

Bacterial communities exist there in homeostasis. Moreover, most of the composition of urinary tract microbiome consists of bacteriophages (Perez-Carrasco et al., 2021). Upper reproductive tract of woman is colonized by microorganisms as well. This site considered for a long time as completely sterile is significantly more diverse than skin and UTI surface (Franasiak & Scott, 2015).

Two major body sites that are niches for the human microbiome are skin and gastrointestinal tract. Skin is colonized by microbiome from the beginning of human life. First microbiota is transmitted from mother during birth and further developed during breastfeeding. Microorganisms can colonize whole skin and they have crucial impact on shaping tolerance of T-cells to further commensal microbe presence. Interruption of that process may have serious health consequences manifested during growing up and ageing (Nagao & Segre, 2015; Scharschmidt, 2017; Scharschmidt et al., 2015). Investigation of human skin microbiome allowed for distinguishing 20 different sites of the body where configurations of bacterial fractions were significantly different (Dréno et al., 2016; Grice & Segre, 2011). Major condition that affects diversity and abundance of human skin microbiota is pH (ranging from 4.2 on pale skin to 6.25 on soles), temperature, and humidity. What is interesting, although the skin is exposed to many external factors that can be potentially bacteriostatic, these are of no significance for the bacterial fractions that make up the microbiome (Grice & Segre, 2011; Oh et al., 2016).

One of the most common examples of a skin disease directly related to a commonly occurring microorganism is acne. *Propionibacterium acnes* was for a long time considered as a major factor that cause skin disease. Nevertheless, recent studies show that the lack of bacteriophages specific to *P. acnes* may be involved too. Because of ability to control the abundance of that anaerobic bacilli bacteriophages specific to *P. acnes* may have influence on host's skin condition. Their presence maintain skin in better condition and prohibit expansion of *P. acnes* colonies. (Barnard et al., 2016; Ellis et al., 2019; J. Liu et al., 2015). Other bacteria that may have serious effect on skin condition are *Staphylococcus epidermidis* and *Malessezia* sp. that were considered commensal for a long time. Nevertheless, recent studies demonstrated that they are able to cause difficult skin infections (Akaza et al., 2016; Nishijima et al., 2000). *Staphylococcus aureus* which is an important element of skin microbiome in many individuals

can be linked to atopic dermatitis. Its presence on skin correlates to increased inflammation and severity of that disease(Ellis et al., 2019).

A few studies demonstrate that improper condition and composition of skin microbiome in early years of human life may result in eczema, rosacea, and psoriasis occurrence (Assarsson et al., 2018; Ellis et al., 2019; Grice & Segre, 2011; Whitfeld et al., 2011). While Eli Metchnikoff proposed drinking sour milk to keep condition of the whole body, using probiotics on human skin can prevent or reduce the probability of developing skin disease. It is known that even oral treatment with *Lactobacillus johnsonii* can improve immune functions of skin. Probiotics may affect toll-like receptors which can prevent atopic dermatitis (Assarsson et al., 2018; Ellis et al., 2019; Roudsari et al., 2015). From the perspective of microbime research, the most interesting is the GI tract. Depending on the part of this system, we can identify different communities of bacteria living there. It was first discovered in the 18th century by the aforementioned Leeuwenhoek, who examined material taken from the stools and mouths of patients suffering from various diseases (Bardell, 1982). In the oral cavity, where the GI tract has its beginning, most identified microbes are usually transient biota. These microbes come from food and inhaled air. Due to the conditions in the oral cavity, it is inhabited by predominantly aerobic microorganisms. Within the oral microbiome, several niches can be distinguished that are characterized by different living conditions for microbes: buccal epithelium, maxillary anterior, tongue dorsum, tonsils or tooth surface. Most of them are inhabited by microbes that belong to genera like Gemella, Granulicatella, Streptococcus and Veilonnella. Around 700 species of bacteria were identified and detected there (Deo & Deshmukh, 2019). One should not forget about communities that are not members of the Prokaryote clade. These include *Candida, Aspergillus Saccharomycetales* and *Fusarium* (Sharma et al., 2018). Pre-digested and chewed food is moved to the stomach. Bacterial communities often exist in the oral cavity in the form of biofilm to prevent the transition to more unfavorable conditions of the human stomach (Kilian et al., 2016).

Stomach for long was considered sterile because of its internal conditions. Gastric juice makes it extremally hostile to any life forms. This solution consisting of hydrochloric acid and pepsin (enzyme endopeptidase) with pH reaching 1.7 (Dressman et al., 1990). However, some microbes including viruses are able to create a core community that inhabits that organ.

Researchers revealed presence of *Firmicutes*, *Bacteroidetes*, and *Actinobacteria* which dominate gastric mucosal samples. One of the most recognizable species is *Helicobacter pylori*. (Nardone & Compare, 2015b; Z. Wang et al., 2020; Y. Yang et al., 2021). Some studies suggest that microbiota found in the stomach is transient only. Its presence is related to a diet and dependent on food that comes through esophagus. Bacteria take advantage of the fact that the pH of gastric juice varies. It is different before and during a meal. When food enters the stomach, the pH temporarily rises, and this also provides an opportunity for the bacteria to have a limited hold on the stomach mucosa (Nardone & Compare, 2015b; I. Yang et al., 2013). Bacteria also take advantage of any decrease in immunity and irritation within the stomach.

Inflammation of the gastric mucosa enforce the use of gastric juice neutralizers or blockers. This creates very favorable conditions for colonization by microorganisms. Studies have shown that an increase in pH is clearly correlated with the diversity of microbiota that we can identify in the stomach. These studies have confirmed this phenomenon in both adults and children. Proton pump inhibitors, which are routinely used to treat reflux and inflammatory gastric disorders, significantly increase the diversity of the microbiome that can be identified in stomach (del Piano et al., 2012, 2014) (description of stomach microbiome is further developed in the section "Specific microbiome of stomach").

Behind the pyloric region of the stomach there is the duodenum – the first part of small intestine. Studying the small intestine microbiome is difficult due to the fact that it is difficult to obtain samples directly from this organ. The largely digested food particles coming from the stomach are mixed in the duodenum with other digestive solutions from the pancreas and liver. Also, partial absorption of water and nutrients is started there. Approximately 60% of the genera's Operational Taxonomy Units (OTUs) identified in that part of GI consist of *Streptococcus*, *Actinomyces*, *Propionibacterium*, and *Granulicatella*. In duodenum, microorganisms have to contend with peristaltic movements, active enzymes, and limited access to oxygen(Angelakis et al., 2012, 2015).

Bacterial density is significantly higher than in stomach, but it represents only a fraction of what can be found in further sections of the intestine. As with the stomach, transitional microbiota predominates duodenum. Peristaltic movements strongly contribute to that, as shown in an experiment where bacterial colony-forming units sampled from different parts of

the small intestine differing in the dynamics of these movements were compared. In the ileum the number of detectable bacteria was several orders higher (Hayashi et al., 2005; Kastl et al., 2020).

Distal sections of the GI tract include the large intestine, which is the most often investigated in terms of microbiome composition. Peristaltic movements slow down there and the final absorption of water from the consumed food takes place. The diversity of bacteria identified in large intestine is the highest within the human body. Bacteria found in this section of the digestive tract have been found responsible for a myriad of health conditions in our bodies. For instance, the gut-brain axis has been demonstrated, that is, human microbiota can affect brain functions via secretion of several different compounds. Interactions can be mediated through the immune system, tryptophan metabolism, or the secretion of bacterial metabolites such as short-chain fatty acids, amino acids, or peptidoglycan, and others. The composition of microbiota in the gut and the resulting profile of microbial products that are secreted may play a role in conditions such as autism, overweight, schizophrenia, Parkinson's disease, and Alzheimer's disease. This part of the digestive system has even been a subject of its own variant of the Human Microbiome Project - the American Gut project and the British Gut project, that are directed solely at studying diversity in large intestine (McDonald et al., 2018; Savio et al., 2017).

Colonization of the gut begins shortly after birth, where the mother's microbiota is passed on. This is influenced by the way a person comes into the world. The consequences of cesarean delivery for microbiome composition can be detected even in adults. It results in lower representation of probiotic bacteria and lower abundance of *Bacteroides* species eventually causing gut homeostatis disorders. Increased population of *Enterobacteriaceae* lead to enhanced lipopolysaccharide (LPS) production. Inappropriate exposure for microbial factor is correlated with development of asthma during growing up in children (Cryan et al., 2019; Hemarajata & Versalovic, 2012; G. Kim et al., 2020; Lyon, 2018). Gut is also a home for the largest portion of microorganisms known as probiotics. They are responsible for maintaining homeostasis in the intestine. A balance of microbial fractions should be achieved, where no fraction overgrowth others. Overrepresentation of one particular microbial fraction can lead for instance to irritable bowel syndrome or inflammatory bowel disease. Application of

probiotics has a positive effect on controlling symptoms of the aforementioned conditions. A chronic dysbiosis in the gut leads to a chronic inflammation. This may further induce formation of colorectal cancer (Hemarajata & Versalovic, 2012).

8.11. Specific microbiome of stomach

In 1982, Robin Warren and Barren Marshall raised doubts on the sterility of human stomach. *Campylobacter pyloridis*, renamed two years later to *Helicobacter pylori*, is able to survive and colonize such challenging environment as stomach. This environment conditions are determined by presence of gastric juice with one of the lowest pH levels found in animal bodies, thickness of the mucus layer, and presence of gastric peristalsis. Moreover, microbiota in the mouth convert the nitrogen in food and saliva into nitrite; the process is mediated by bacteria from *Lactobacilli* genus. When nitrites come into interaction with gastric juice, they are converted into a powerful antibacterial compound, nitric oxide. *H. pylori* developed several capabilities which allow for colonizing of stomach. One of the best-known adaptations is mediated by the enzyme urease, much more efficient than development cell walls tough enough to resist the low pH in the stomach. Around 10% of all produced proteins in *H. pylori* cells is intracellular urease. The bacterium has the unique ability to change pH of the environment, since the enzyme catalyzes the reaction of hydrolytic degradation of urea to ammonia and carbon dioxide. Molecules of the urease can increase periplasmatic pH and membrane potential. Urea which is a direct substrate for the enzyme is acquired through the bacterial inner membrane. One of the functions attributed to this protein is also interaction with human cells that may be crucial in colonizing the gastric mucosal layer. It has been shown that the urease enzyme also has a pro-inflammatory activity and exerts inactivating effects on neutrophil cells (Baj et al., 2020; Hathroubi et al., 2018).

Coding of urease is not just assigned to *Helicobacter pylori*. Other transients that are a part of the gastric microbiome also benefit from the utility of this enzyme. *Morganella morgani* and *Yersinia enterocolitica* are pathogens that are able to survive for a longer period of time in an acidic environment. They also use urease to change the environment to a more alkaline (Uberti et al., 2013; Young et al., 1996).

The hydrolysis of urea by urease changes the consistency of the gastric mucosal layer. It converts to a more mucous-like consistency, which makes it easier for bacteria to shuffle through. The consistency of the mucus changes when pH rises above its normal pH 1-2. *Helicobacter pylori* has a spiral shape that allows it to move faster in the conditions it creates. In addition, this bacterium also has a flagellum made up of two subunits, being long in relation to the cell length. Usually, the cell has 2 to 6 flagella, and they are activated by contact with an acidic environment. Thanks to these adaptations, *H. pylori* is able to migrate faster into the gastric mucosal layer and to start its colonization. Environmental stress conditions affecting bacteria residing in the stomach can lead to damage of genetic material and thus prevent colonization. *H. pylori* through homologous recombination is however able to repair DNA damaged by low pH and oxidative stress. Its genome also encodes a number of proteins such as RecA, which attach to the DNA strand and mediate its repair (Clyne et al., 1995).

Lactic acid bacilli that have also been identified as a part of the gastric microbiome have also developed ability to survive in the unfavorable environment of gastric pH. The pH tolerance of these microorganisms depends on the range of the pH gradient between the extracellular and intracellular environments. When this difference exceeds a certain acceptable value, the bacterial cells die. Gram-positive bacteria use a mechanism based on the F_0F_1-ATPase that increases the intracellular pH under stressful conditions for the cell. This mechanism is activated at the transcriptional level (Deckers-Hebestreit & Altendorf, 1996). Enzyme takes part in changing transmembrane ion gradient due to high concentration of ATP in prokaryotic cell. It makes inside of the bacteria more resistant for outside changing ion potential. Increased pH tolerance among these microorganisms in the presence of simple sugars has also been documented. This causes these bacteria to have the best chance of surviving in gastric juice conditions when the pH is higher due to food intake or inflammation occurrence (Charalampopoulos et al., 2003).

Escherichia coli has also evolved its unique ability to avoid the negative effects of low pH in the stomach. This bacterium is able to rearrange lipids in the outer membrane, leading to increased tolerance to this stress factor(Foster, 2004; Lund et al., 2014; Y. Xu et al., 2020). *Salmonella enterica* serovar Typhimurium is a pathogen which infection begins with the ingestion of contaminated water or food. It is identified as a transient microbiota in the

stomach. Among the many virulence factors present in this bacterium are those that allow it to survive neutralization by gastric juice in the stomach during the initial stages of digestion. *S. enterica* has a system of defense against acidotic stress based on decarboxylation of arginine, lysins, and ornithine. This system is however less efficient in the absence of atmospheric oxygen. *Brucella spp* belongs to gram-negative intracellular pathogens that can be present in stomach microbiome. These bacteria are linked to food production and processing, often transmitted from animals to humans through consumption of unpasteurized milk and contaminated dairy products. *Brucallae* is equipped with the low pH defense systems similar to those in other pathogens. It has the ability to produce urease, like *H. pylori*, however there is no information about its intracellular activity. It is also protected by a decarboxylation-based system similar to *Salmonella* (Xu et al., 2020).

Streptococcus spp. is most abundant in microbiota in oral cavity, which makes these bacteria natural inhabitants and permanent colonizers of the stomach environment. They produce active F-ATPase that enable to alkalize the surrounding environment. *Streptococcus spp.* have specific sites in their genome that directly protect DNA strands from degradation due to exposure of the cell to low pH. These bacteria have a system for transporting potassium ions to cells, which allows them to maintain ion balance when the pH gradient between the inside and outside of the cell becomes too high (Quivey et al., 2000; Sheng & Marquis, 2006).

In the analysis of surviving elements of the stomach microbiome, the most numerous group - viruses - must be considered as well. Viruses do not demonstrate similar (to bacterial) adaptations to persist in the stomach environment. However, in their genomes, they code factors that make it easier to survive those difficult conditions. Most of Caudovirales, the most abundant group of phages, cannot stay active when pH in the environment drops below 4. Examples of exceptions are M13 and T1 phages specific to *Escherichia coli*. They withstand pH as low as 2 and 3 respectively. Bacteriophages have also the possibility to survive harsh environmental conditions inside bacterial cells as an incorporated part of the genome (prophage), or a plasmid, when entering other life cycles than a lytic one. In terms of phage therapy bacteriophages are expected to persist active in various environments, including these with low pH. Currently researchers' efforts are directed to investigate molecular mechanism

of phage resistance to difficult conditions in stomach (Jończyk et al., 2011; Nobrega et al., 2016).

8.12. Human genome effects on microbiome

Human microbiome consists of an enormous number of genes coded by all of microorganisms inhabiting all body sites. Nevertheless, genes encoded by the host may interplay with the microbiome. The complexity of the environment in which microorganisms live is also shaped by genetic determinants. Genome-wide association studies (GWAS) have shown that the human genome may contain 4.1 to 5 million different variants, more than 99% of which are Single Nucleotide Polymorphisms (SNPs) and indels. NGS sequencing techniques have enabled multiomic studies of potential influence of human genome on profiles of human microbiome. To design an experiment that examines possible relations between bacterial fraction occurrence and hosts' SNPs profiles, one must consider the complexity of the data to be analyzed. Demographics can also shape the microbiome. Gender, age, and geographic origin of studied individuals must be considered.

Symbiosis of humans with microorganisms is beneficial in many ways, among others, by bacterial competition for receptors on human cell surfaces. Due to colonization by commensal microorganisms, pathogens have no room to expand their colonies. Considering the importance of the microbiome in human life, there is a need to know the mechanisms that determine the inhabitation of ecological niches in the human body. Not only from the perspective of bacterial adaptation, but also from the position of the host, which creates the conditions of this environment. One example of correlation between host genotype and bacteria involves the genus *Bifidobacterium* in fecal microbiota. SNP marked as rs4988235 was identified as associated with a lack of lactase and increased abundance of *Bifidobacterium* in examined samples. In this case, it is supposed to have a direct influence on the presence of Bifidobacteria (Bonder et al., 2016).

Another identified human genetic factor that affects microbiota composition is the SLIT3 gene. This is a gene that encodes a secretory protein that is responsible for cellular transport. Unclassified Clostridiaceae had association with the SLIT3 variant rs10055309 detected in UK twins. Upregulation of SLIT3 expression was found associated to colorectal adenomas

(Goodrich et al., 2016). Study by Igartua et al. (2017) performed 16S rRNA amplicon sequencing of 144 adult Europeans' microbiomes. They identified 148 653 genetic variants among all annotated variants. The showed association between some of them with diagnosed diseases. *Dermacoccus* presence within nasal microbiome was identified in all 144 adult individuals of a European population collected in summer and winter months. Identification of *Dermacoccus* was negatively correlated with atopic dermatitis. An association of Micrococcacae with a variant of the PGLYRP4 gene was identified. This is the gene associated with the immune response to peptidoglycan contained in the cell walls of gram-positive bacteria (Igartua et al., 2017a).

Another study was performed by Dayama et al. (2020) with patients that were diagnosed with cysitic fibrosis. That disease is caused by cystic fibrosis transmembrane conductor regulatory (CFTR) gene mutations. Reduced expression leads to lack of hydratation in mucus, difficulties in respiratory, digestive, and reproductive functions (Dayama et al., 2020). Variants of CFTR gene were correlated with decreased abundance of Ruminococcaceace and *Butyricimonas*. These groups of bacteria are considered as butyrate producing microbes. While, on the other hand Actinobacteria and *Clostridium* were more abundant than in healthy individuals. Butyrate promotes bacterial growth, and it also is an anti-inflammatory agent. In the intestine, butyrate plays an important role in maintaining homeostasis (Dayama et al., 2020).

One clue that the human genome influences the microbiome that resides in the body was a study involving related family members. In that study, Lee et al. (2011) compared the fecal microbiota from twins living in Korea and the United States. The results showed that there were significant geographic differences. Nevertheless, authors revealed that differences between continent fractions were same as between monozygotic and dizygotic tweens living on different continents (S. Lee et al., 2011).

Data left behind by the Human Microbiome Project were analyzed for possible correlations of the microbiome with the genome of the carriers studied. Bioinformatics analysis revealed an association between the FUT2, functional lactase gene (LCT) and the presence of *Bifidobacterium longum*. Publicly available databases were searched for genetic profiles where the host was fucosylotrasferase secretor genotype. Linear regression revealed that *B. longum* had the strongest correlation. Individuals without premature stop codon in FUT2 gene were

characterized by significantly increased abundance of that bacteria. Very similar effect was observed according to LCT. This gene is responsible for lactase coding in human gut. Lactase production is regulated by allele in rs4988235 close to LCT. Latest findings have shown that a diary rich diet is correlated positively with hypolactasia gene variant and Bifidobacteria abundance in human gut (Kato et al., 2018; Kolde et al., 2018a).

Genome wide studies included individuals from norther Germany was performed by Wang et al (J. Wang et al., 2016). Authors examinated taxa derived from metagenomic analysis with multiple genetic loci. They described association between gene with Porphyromonadaceae. This gene is responsible for one of the proteins that belongs to glucose transporter family. It plays a crucial role in maintaining glucose homeostasis. What is more, they identified that (rs62295801) is correlated with *Lactobacillales* which can potentially affect response to vitamin A and increased liver cholesterol level. These rod-shaped bacteria are mostly known as probiotic. They are permanent inhabitants of human gut. Beta-diversity between 1812 individuals from Northern German was analyzed and did not show any significant results above satisfactory threshold. This allows for a hypothesis that presence of genetic variants encoding Vitamin D receptor (VDR) are not related to the abundance of detectable bacteria. Nevertheless, VDR gene locus was associated with increased abundance of *Parabacteroides*. These are gram negative, anaerobic that colonize GI tract. Species from that type are related to inflammation, autoimmune disorders and autism spectrum disorders (Ezeji et al., 2021; J. Wang et al., 2016).

9. Materials

9.1. Samples collection from human participants of the study

Patients underwent gastroscopy by gastroenterologists working at the Regional Specialized Hospital in Wroclaw. Two qualified surgeons participated in the project: Jan Gnus, MD PhD and Stanisław Ferenc, MD PhD. Sample collection started on 30.01.2018. They were collected during endoscopy examination by the physicians in the Endoscopy Department of Regional Specialist Hospital in Wroclaw. Prior to the examinations, patients consented to the use of biological samples collected from them, and gave interview on chronic diseases, medtications they were taking, and on antibiotics use during recent 6 months. Specimen was taken from the

pyloric part of the stomach. That specific place was chosen according to available literature that indicates it as the most inhabited by microorganisms part of human stomach and this area is recommended for diagnostics of *Helicobacter pylori* infection (Bayerdorffer et al., 1989). Two separate specimens were taken from each patient, blood samples were collected for further SNPs analysis. One was proceeded to histological examination to evaluate parameters according to Sydney scale. The project was completed under approval of the bioethical committee (Komisja bioetyczna przy Wojewódzkim Szpitalu Specjalistycznym we Wrocławiu Ośrodek Badawczo-Rozwojowy). The project approval number is KB/nr 8/rok 2017.

Patients' data were derived from the interview and from medical database Asseco Medical Management Solutions (AMMS). The data that were obtained in this way are: age, gender, chronic diseases affecting the patient, medications taken permanently, antibiotics, drugs from the group of proton pump inhibitors, Pandas version 1.4.2 was used on sheets to construct conditional and pivot tables. Numpy was used for array analysis in numerical data (Harris et al., 2020). Scikit-bio – python package with algorithms, and resources for bioinformatics was used for index calculations.

9.2. DNA isolation for NGS sequencing

Each sample collected from a patient was immediately placed in PBS and forwarded for separate viral and bacterial DNA isolation. Specimens in PBS were floated on 3D shaker for 3h in 4 Celsius degrees and centrifuged for 10 000 RPMs for 10 minutes. Supernatant was used for further bacteriophage separation and pellet was proceeded to bacterial DNA isolation with Micro Beat Bead Gravity AX kit (A&A Biotechnology). Bacteriophage separation was performed with cesium chloride (CsCl) ultracentrifugation as follow. Supernatant was filtered through 0.22μm pore membrane and then loaded on CsCl gradient. CsCl density gradient ranged from the highest density 1.7g/ml, followed by 1.5 g/ml, 1.35 g/ml and 1.15 g/ml (scheme is presented in the Figure 1.). Centrifugation was performed overnight at 62 000 g at 4 Celsius degrees in a swinging bucket rotor. Half a milliliter of content between 1.35 and 1.5 of CsCL solution was collected with a syringe for further phage DNA isolation. DNA isolation was performed with Sherlock AX (A&A Biotechnology) kit. Samples were proceeded to Genomiphi V2 DNA Amplification (Cytiva). Human genomic DNA was isolated from collected blood samples using Genomic Micro AX Blood Gravity Kit (A&A Biotechnology).

DNA isolated from biopsy samples using the Micro Beat Bead Gravity AX Kit (A&A Biotechnology) was proceeded for bacterial DNA sequencing. DNA samples isolated from human blood using Sherlock AX (A&A Biotechnology) were forwarded to Ampliseq – SNP sequencing. DNA obtained from biopsy samples and ultracentrifuged were proceeded to Shotgun sequencing for Phageome investigation.

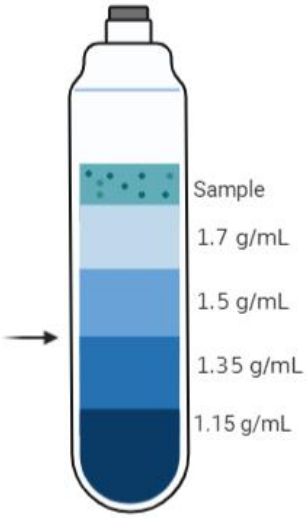

Figure 2. Scheme of the ultracentrifugation tube with CsCl gradient layers loaded on; arrow indicates bacteriophage fraction location

9.3. Library preparation for 16S rRNA sequencing.

Initially, DNA isolated in the previous step was quantified using Qubit 2.0 with dsHigh Sensitivity Assay Kit (Thermofisher). Minimum concentration of DNA that was necessary for further steps (no less than 2 ng/µl) was achieved in 144 samples (out of 145 samples processed), these samples were proceeded for further analysis. Amplification of 16S rRNA regions were performed with Ion 16S Metagenomics Kit (Thermofisher). It contains 2 sets of primers that cover V2, V3, V4, V6, V7, V8 hypervariable regions. Specific Environmental Master Mix which is a part of the assay provides conditions to perform PCR even with a high concentration of PCR inhibitors. Amplified DNA was purified with AMPure XP Reagents (Beckman) – silicone magnetic beads. DNA fragmentation was done with Ion Plus Fragment Library Kit (Thermofisher). Purified DNA was measured again with dsHigh Sensitivity Assay kit on Qubit 2.0 fluorometer. Amount of DNA recommended by the manufacturer was used for library preparation according to manufacturer's instructions. For the purpose of sample pooling

oligonucleotide barcodes were included into the library preparation; Ion Xpress Barcode Adapters kit (Thermofisher) was used as recommended for enzymatic fragmentation with Ion Plus Fragment Library Kit. The final concentration of libraries was measured with Ion Universal Library Quantification Kit (Thermofisher) using BioRad CFX386. Prepared library was stored in -20 Celsius degrees.

9.4. Library preparation for phageome sequencing

Eluted DNA was quantified using Quantus Fluorometer with QuantiFluor ds DNA system kit. Minimum concentration of DNA that was considered as sufficient was 1 ng/µl Samples with concentration lower than mentioned were discarded. Purity of isolated DNA was estimated using DeNovix spectrophotometer. Absorbance 260 / Absorbance 280 ratio was calculated to identify contaminations which absorb UV light e.g., proteins, organic solvents. As the quality requirement, DNA concentration in eluted solution should be higher than 1 ng/µl. Out of 140 isolated viromes, 99 met that requirement and they were proceeded to amplification and library preparation. DNA amplified by GenomiPhi V2 DNA Amplification Kit (Cytiva Life Sciences) was finally quantified using Quantus Fluorometer and QuantiFluor dsDNA Kit (Promega). Illumina DNA Prep (Illumina) was used to prepare sequencing libraries. It was chosen because of its specification that includes broad DNA input range and applicability for both short and long sequences. All samples were barcoded using Nextera DNA CD indexes (Illumina). Final concentration of libraries was calculated using Quantus Fluorometer results. Prepared libraries were stored in -20 Celsius degrees.

9.5. Library preparation for Single Nucleotide Polymorphisms sequencing

Quantitation of DNA isolated from blood cells was performed using Qubit 2.0 with ds High Sensitivity Assay Kit (Thermofisher). As a minimum requirement, concentration of DNA as 1 ng/µl was set Samples with human DNA concentration lower than required were discarded. Samples which absorbance ratio 260/280 did not amount to ~1.8 were discarded due to residue protein contamination. 111 (out of 141) samples remained after quality control using DeNovix spectrophotometer. A custom Ampliseq panel was designed using Ion AmpliSeq Designer and ordered as Ampliseq Custom Panel (Thermofisher). It contained 218 SNP targets within 181 different amplicons (Supplementary material 1). Targets were designed using Ion Ampliseq

Designer. For the amplification of dedicated regions Ion Ampliseq Library Kit Plus (Thermofisher) was used according to manufacturer's instructions. Amplified DNA was purified with AMPure XP Reagents (Beckman) – silicone magnetic beads. DNA fragmentation was performed using Ion Ampliseq Library Kit Plus. For the purpose of sample pooling oligonucleotide barcodes were included into the library preparation: Ion Xpress Barcode Adapters kit (Thermofisher) was used. The final concentration of libraries was calculated using Ion Universal Library Quantification Kit (Thermofisher) with BioRad CFX386. Prepared library was stored in -20 Celsius degrees.

9.6. Ion Torrent Sequencing

Quantified libraries were proceeded to emulsion PCR using Ion One Touch 2. Amount of 5 µl was taken from each sample and pooled into one 0.2 ml tube (Thermofisher Scientific). Reagent mixture for emulsion PCR contained Ion PGM Hi-Q regent Mix, Hi-Q Enzyme Mix, and Ion Hi-Q ISPs. Whole volume of the freshly prepared mix was poured on Ion One Touch Reaction filter and placed into the dedicated device. Emulsion PCR was performed with *assisted* option and with use of Ion PGM Hi-Q OT2 Kit -400. Enrichment part of the process was run using automated Ion OneTouch ES. This part of the process was done same way for 16S rRNA and for Ampliseq samples.

The enriched spheres coated with DNA fragments (prepared as described above) were mixed by pipetting and then centrifuged for 4 minutes at 14500 RPM. Supernatant was discarded from the pellet so that approximately 27 µl remained in the tube. Two microliters of primers were added by pipetting. Sample was placed in thermocycler on a 2-step program – 95 Celsius degrees for 2 minutes, followed by 37 Celsius degrees for another 2 minutes. Next, samples were set aside to achieve room temperature. In the meantime, "sequencing chip check" was performed on Ion Torrent Personal Genome Machine. For 16S rRNA sequencing Ion 316 Chips were used, while for SNPs detection Ion 318 Chips were used. All the operation with the sequencing chip was performed without protective gloves and with proper grounding (as recommended). A pre-programmed sequencing plan – "16S rRNA Sequencing" was selected on the instrument and manufacturer's instructions were followed until the chip calibration was completed. Sequencing plan involved 750 flows of each ddNTPs through inserted chip. It was set to generate reads as long as 400 bps. After sequencing process, basecalling was

automatically run. This process was based on translating conducting data collected during sequencing into nucleotide-based reads. Once the samples reached room temperature, 1 µl of Ion PGM Hi-Q Sequencing polymerase was added. The samples were mixed by pipetting and incubated for 5 minutes on the bench. Filling solution was removed from the sequencing chip that passed calibration and then the sample was applied on the chip. The chip along with the sample was centrifuged 3 times for one minute on a dedicated centrifuge at maximum speed. After each centrifugation, the liquid on the chip was pipetted five times. This step was to evenly distribute the DNA coated beads in the wells on the chip. After the 3rd centrifugation, excess beads and fluid were removed with a pipette. The chip from Ion 318 Chip Kit v2 BC (Thermofisher Scientific) was placed inside the Ion Torren Personal Genome Machine. It was correctly recognized by the machine, and then after final plan approval, sequencing was initiated.

9.7. Illumina sequencing

Previously quantified libraries were denatured with 1M NaOH (Sigma-Aldrich). Fresh dilution of NaOH was used. For the purpose of NextSeq 500/550 Mid Output kit (Illumina) was used. Equal volumes of sample and 1M NaOH was incubated for 5 minutes and then followed by adding 200 mM Tris-HCL (Sigma-Aldrich) with pH 7. HT1 solution was thawed at room temperature. According to manufacturer's instructions prechilled Hybridization Buffer was mixed with the library pool to get total volume of 1 ml. Simultaneously, control library of PhiX sequencing Control V3 (Illumina) was prepared and finally mixed with the sample. Final loading concentration contained denatured library, denatured control library (PhiX), and hybridization buffer. Initialization of NextSeq 550 was fully automated and required pre-run cleaning with predefined solutions. Sample was loaded on the cartridge and then putted directly into the device. Following instructions were available on touch screen, all of the reagents were placed and sequencing was initiated.

9.8. Files extraction

FASTQ files were downloaded from sequencing devices using software dedicated to Illumina and Ion Torrent devices, that is BaseSpace (Illumina) and TorrentExporter (Thermofisher Scientific) respectively. Files from IonTorrent PGM were exported as Single End Sequences,

thus reads pairing was not applicable. Single End reads are specific for Ion Torrent sequencing technology. Illumina sequencing is based on sequencing by synthesis replicating small clusters of DNA; during that process “bridges” consisting of double stranded DNA are created. This specificity allows for creating Paired End reads. Data were transferred to internal computer cluster in Hirszfeld Institute of Immunology and Experimental Therapy using WeTransfer. Further Analysis was done according to protocols presented below.

9.9. 16S rRNA data analysis

To create FASTQ files from 16S rRNA sequencing FileExplorer plugin was used on IonTorrent Server. Data was transferred to local computer cluster in the Hirszfeld Institute of Immunology and Experimental Therapy. Firstly, FastQC – a quality control tool for High Throughput Sequence was used to determine quality of performed runs. This tool has been created for screening potential errors in datasets created using NGS technologies. To remove low quality sequences from the analysis, Trimmomatic V0.32 was used with minimal length of 50 base pairs (bp). Sequence mapping was performed with Kraken2. Kraken2 is a taxonomic classification tool which makes use of k-mer matching. This is more efficient and faster than classical approach by BLAST algorithm. Database that was used was Kraken Microbial database (September, 2018) which involved all of 16S rRNA genes publicly available in the public repositories. Report files were subjected to further statistical analysis while output files were used for Krona chart creation. This tool creates interactive taxonomic visualization of each sequenced sample. To perform all above steps Python 3.6 was used and an individual script was created, this script is presented in the Table 1.

Table 1. An individual script for Python 3.6 used for 16S rRNA data analysis with sequences derived from Ion Torrent PGM instrument – Single End Reads saved in FASTQ files.

```
import glob,os,subprocess, sys
from venv import create
#take first argument in loop
folder = (sys.argv[1])
print(folder)
#change destination
destination = "cd {}".format(folder)
print(destination)
destinationcommand = subprocess.run(destination,shell=True)
```

```
def importhomeland():
    stream = os.popen('pwd')
    out = stream.read().strip()
    return out
#Change workspace of script
#execetute commands
def executecommand(command):
    subprocess.run(command, shell = True)
def createfolder(name):
    if not os.path.isdir(name):
        os.mkdir(name)
workspace = importhomeland()+"/"+folder
dir = os.chdir(workspace)
print(workspace)
#make needed folders
createfolder("trimmed")
createfolder("outputs")
createfolder("reportskraken")
createfolder("Kronacharts")
#Run Trimmomatic
for i in glob.glob("*fastq"):
    samplename = i.split(".fastq")[0]
    print("Processing_"+samplename)
    #create command
    trimmomatic= "java -Xms4g -Xmx4g -jar /home/fagi/trimmomatic/Trimmomatic-0.39/trimmomatic-0.39.jar SE -
threads 32 -phred33 {} trimmed/trimmed_{}.fastq TRAILING:20 MINLEN:50".format(i, samplename)
    #save trimmomatic output in folder trimmed
    executecommand(trimmomatic)
    #commandtrimmomatic = subprocess.run(trimmomatic, shell =True)
#Run Kraken2
for i in glob.glob("trimmed/*fastq"):
    #extract filename
    trimmedname = i.split(".fastq")
    #create kraken2 command
    kraken2 = "/home/fagi/kraken2/kraken2 --threads 16 --db /home/fagi/kraken_microbial  --fastq-input {}  --report
reportskraken/report_{} --output outputs/output_{}.krona".format(i,trimmedname,trimmedname)
    #save reports to reportskraken and output for krona in outputs
    executecommand(kraken2)
    #commandkraken2 = subprocess.run(kraken2, shell =True)
#Run Krona Chart Creation
for i in glob.glob("outputs/*krona"):
    #extract name for krona
    krakenname = i.split(".krona")[0].split("_")[1]
    krona = "/home/fagi/krona/bin/ktImportTaxonomy -q 2 -t 3 {} -o Kronacharts/{}.html".format(i,krakenname)
    executecommand(krona)
```

```
commandkrona = subprocess.run(krona, shell = True)
```

9.10. Viral Data analysis

FASTQ files were obtained by demultiplexing. Data was transferred to local computer cluster in the Institute of Immunology and Experimental Therapy. FastQC was applied to detect possible low-quality files. Low quality sequences were removed from the analysis by Trimmomatic V0.32 (minimal length of 50 bps). Due to significant contribution of human DNA, it was removed from analysis by mapping sequences to Human Genome GRCh38 (GenBank Accession: GCA_000001405.15) and excluding positive hits from further analysis. That part of analysis was done using bowtie2. Initially, unaligned sequences were processed using Kraken2 aligner, with Kraken Microbial Database (September, 2018). Due to dynamic changes in taxonomy annotations related to bacteriophages it turned out to be outdated. There was a need to prepare custom database and use another, more customizable aligner. Therefore, sequences unaligned to the reference human genome were proceeded to BLAST analysis. BLAST algorithm was used along with parameters that allowed 2 mismatches in aligned sequence and minimum length of 50 bps. Blast was used locally on the computer at HIEET however, the new script was not needed because the cleaned readings from the kraken2 screening were used for it. Reference for this part of analysis was built from all DNA Virus sequences available in public National Center for Biotechnology Information Genomes repository. BLAST report files were subjected to further statistical analysis. Python 3.6 was used and an individual script was created, this script is presented in the Table 2.

Table 2. Script in Python 3.6 used for data cleaning and Kraken2 screening analysis for Paired End Reads from Illumina sequencing.

```
import glob,os,subprocess, sys
#take first argument in loop
folder = (sys.argv[1])
print(folder)
#change destination
destination = "cd {}".format(folder)
print(destination)
destinationcommand = subprocess.run(destination,shell=True)
   #Change workspace of script
```

```
def importhomeland():
    stream = os.popen('pwd')
    out = stream.read().strip()
    return out
#execute commands
def executecommand(command):
    subprocess.run(command, shell = True)
#create folders
def createfolder(name):
    if not os.path.isdir(name):
        os.mkdir(name)
workspace = importhomeland()+"/"+folder
dir = os.chdir(workspace)
print(workspace)
pathtofagi = "/bighdd/metagenomics/IlluminaSeq/workspace/"
trimmedpath = pathtofagi+"trimmed_"+folder
outputspath = pathtofagi+"outputs_"+folder
reportspath = pathtofagi+"reportskraken_"+folder
#kronapath = pathtofagi+"Kronacharts_"+folder
cleanseqpath = pathtofagi+"cleanseq_"+folder
normalreportpath= pathtofagi+"normalreport_"+folder


createfolder(pathtofagi+"trimmed_"+folder)
createfolder(pathtofagi+"outputs_"+folder)
createfolder(pathtofagi+"reportskraken_"+folder)
#createfolder(pathtofagi+"Kronacharts_"+folder)
createfolder(pathtofagi+"cleanseq_"+folder)
createfolder(pathtofagi+"unclassified_"+folder)
createfolder(pathtofagi+"normalreport_"+folder)
#Run Trimmomatic
for i in glob.glob("*R1*fastq"):
   print(i)
   samplename = i.split(".fastq")[0]
   readname = i.split("_")[0]
   print("Processing_"+samplename)
   #create command
   trimmomatic= "java -Xms4g -Xmx4g -jar /home/fagi/trimmomatic/Trimmomatic-0.39/trimmomatic-0.39.jar PE -
threads 32 -phred33 {}*R1* {}*R2* {}/clean_{}_R1.fastq /dev/null {}/clean_{}_R2.fastq /dev/null TRAILING:20
MINLEN:50".format(readname,readname,trimmedpath,readname,trimmedpath, readname)
   print(trimmomatic)
#save trimmomatic output in folder trimmed
  executecommand(trimmomatic)
#discard human sequences
for i in glob.glob(trimmedpath+"/*R1*fastq"):
```

```
    print(i)
    trimmedname = i.split(".fastq")[0]
    print(trimmedname)
    readname = i.split("_")[2]
    print(trimmedpath+readname)
    print(readname)
    print(cleanseqpath)
    bowtiediscard = "/home/fagi/miniconda3/bin/bowtie2 --threads 16 --very-sensitive-local -x
/home/fagi/bowtie2grch38/GCA_000001405.15_GRCh38_full_analysis_set.fna.bowtie_index -1 {}*R1.fastq -2
{}*R2.fastq --un-conc {}/clear_{}.fastq -S
/dev/null".format(trimmedpath+"/*"+readname,trimmedpath+"/*"+readname,cleanseqpath, readname)
    print(bowtiediscard)
   executecommand(bowtiediscard)
#Run kraken2
for i in glob.glob(cleanseqpath+"/*.1.fastq"):
    #extract filename
    trimmedname = i.split(".fastq")[0].split(".")[0]
    outname = trimmedname.split("/")[-1]
    print(trimmedname)
    #create kraken2 command
    kraken2 = "/home/fagi/kraken2/kraken2 --threads 32 --db /home/fagi/kraken_microbial --paired {}*.1.fastq {}*.2.fastq
--report {}/report_{}".format(trimmedname,trimmedname,normalreportpath,outname)
    print(kraken2)
    #save reports to reportskraken and output for krona in outputs
    executecommand(kraken2)
```

9.11. Single Nucleotide Polymorphism Ampliseq Data analysis

FASTQ files were generated on IonTorrent server and then exported using FileExplorer plugin to computer cluster in the Institute of Immunology and Experimental Therapy. FastQC tool was applied to evaluate quality of the results. Low quality sequences were removed from the analysis using V0.32 Trimmomatic with minimal length of 50 bp. BWA (Burrows – Wheeler Aligner) was used to align DNA sequences to a reference human genome: Human Genome GRCh38 (GenBank Accession: GCA_000001405.15) that was indexed with the “bwa index” command. Generated Sequence Alignment Map (SAM) files were converted to BAM (Binary Alignment Map) with SamTools version 1.14. BAM files were sorted and duplicates were marked using Picard. Sorting is a mandatory process that prepares BAMs for next steps of analysis. It is a predefined set of CLI (command line) programs that were designed for processing of files in formats such as SAM/BAM/VCF BIM Collaboration Format (BCF) tools

were designed to perform manipulations on Variant Call Format (VCF). This was the tool that was used to transform sorted BAM files to VCF as applicable for further analysis. This step is called variant calling which refers to the identification of variants within sequenced data. VCF files were further processed with SNPeff (Cingolani et al., 2012) tool which identifies SNPs present in sequenced samples. SNPeff includes a very large database of publicly accessible, described SNPs, and providing additional information about these SNPs such as functionality, assigned gene, or possible association with human diseases. Bash shell was used and an individual script was created, this script is presented in the Table 3.

Table 3. Bash script used for converting FASTQ files derived from IonTorrent PGM sequencer to Annotated Variant Call Format (VCF) files for further statistical analysis. VCF format is required as an input for further bioinformatic operations. It contains information about identified SNPs and their positions in the investigated genome.

```
mkdir trimmed
echo "\e[36m Trimming Started \e[0m"
for i in *.fastq;
do
  sam=$(echo "$i" | rev |cut -d"." -f2- | rev )
    echo "$sam"
    java -Xms4g -Xmx4g -jar /home/fagi/trimmomatic/Trimmomatic-0.39/trimmomatic-0.39.jar SE -
threads 64 -phred33 "$sam".fastq trimmed/"$sam".fastq TRAILING:20 MINLEN:50
done #cut sequences with trimmomatic
echo "\e[36m Trimming Done \e[0m "
mkdir samfiles
mkdir bamfiles
echo "\e[36m Human GRCh38 mapping \e[0m "
cd trimmed/
#Check if FASTQ contains more than 100k reads to analyse
for i in *.fastq;
do
    sam=$(echo "$i" | rev |cut -d"." -f2- | rev )
    echo "$sam"
    count=$(echo $(cat $sam.fastq|wc -l)/4|bc)
    echo "$count"
    if [ "100000" -gt $count ];then
        rm "$sam.fastq"
        echo "removed $sam"
    fi
done #cut sequences with trimmomatic
```

```
for x in *.fastq #Map human genome for SNPs
do
   sam=$(echo "$x" | rev |cut -d"." -f2- | rev )
   echo "\e[32m Mapping $sam to GRCh \e[0m"
   bwa mem -t 16 /home/fagi/grch38BWA/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna
$sam.fastq > ../samfiles/$sam.sam
done
cd ../samfiles

#Fixing FLAG
#for z in *.bam
#do
#   sam=$(echo "$z" | rev |cut -d"." -f2- | rev )
#   echo "$sam"
#   samtools fixmate -O bam "$sam".bam fixed"$sam".bam
#done
#Sam to bam
echo "\e[36m SAM -> BAM \e[0m"
for y in *.sam
do
   sam=$(echo "$y" | rev |cut -d"." -f2- | rev )
   echo "\e[32m Translating $sam to BAM \e[0m"
   samtools view -S -b $sam.sam > ../bamfiles/$sam.bam # Translate SAM --> BAM
done
#Sorting reads according to chromosome position
cd ../bamfiles
echo "\e[36m Sorting \e[0m"
for u in *.bam
do
   sam=$(echo "$u" | rev |cut -d"." -f2- | rev )
   echo "\e[32m Sorting $sam \e[0m"
   #samtools view -hbS - | samtools sort -m 1000000000 - $sam.bam
   samtools sort -O bam -o sorted"$sam".bam $u
done
#Marking duplicates and removing them
echo "\e[36m Mark Duplicates \e[0m"
for u in sorted*.bam
do
   sam=$(echo "$u" | rev |cut -d"." -f2- | rev )
   echo "\e[32m Checking for duplicates in $sam \e[0m"
   java -jar /home/fagi/picard/picard/build/libs/picard.jar MarkDuplicates I=$u O=Duplicate$sam.bam
M=Duplicates$sam_metrics.txt REMOVE_DUPLICATES=true
done
# samtools faidx /
```

```
#java -jar /home/fagi/picard.jar CreateSequenceDictionary
R=GCA_000001405.15_GRCh38_no_alt_analysis_set.fna O=reference.dict
#Indel Realignment ------- Not needed after last software update
#echo "\e[36m INDEL Realignment \e[0m
mkdir ../VCFfiles
# Basecalling na vcf bam ->> VC
echo "\e[36m Basecall BCF -> VCF \e[0m"
for z in Duplicatesorted*.bam
do
   sam=$(echo "$z" | rev |cut -d"." -f2- | rev )
   echo "\e[32m Basecalling $sam \e[0m"
   /home/fagi/miniconda3/bin/bcftools mpileup -Ou -f
/home/fagi/GRCh38SNP/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna $z |
/home/fagi/miniconda3/bin/bcftools call -vmO z -o $sam.bcf #Variant calling z bcftools
   echo "\e[32m Filtering $sam and translating to VCF \e[0m"
   /home/fagi/miniconda3/bin/bcftools filter -i 'QUAL>100' $sam.bcf > ../VCFfiles/$sam.vcf #PHRED
score filtering
done
cd ../VCFfiles
for z in *.vcf
do
   sam=$(echo "$z" | rev |cut -d"." -f2- | rev )
   echo "\e[32m Annotating SNPs with snpEff $sam \e[0m"
   java -Xmx4g -jar /home/fagi/snpEff/snpEff.jar -c /home/fagi/snpEff/snpEff.config -v GRCh38.86 $z
> Annotated$sam.vcf
done
```

9.12. Data cleaning in Pandas

Data derived from all 3 types of NGS data analysis were transformed into common formats for further statistical data analysis. For this purpose, each folder with respective data was merged into one comma separated values (csv) file. 16SrRNA data before transformation was saved in kraken2 output files containing data on classified microorganisms, accuracy, and counts per taxonomy level. Bash shell was used and an individual script was created, this script is presented in the Table 4.

Table 4. Script used for merging 16S rRNA data from kraken2 reports. Jupyter lab software was used to perform all actions in cited scripts. For better readability for the purposes of this document, the individual cell results have been omitted.

```
#Initial settings creation
```

```
import pandas as pd
import glob,os
import re
import seaborn
#import workspace
workspace = "/home/fagi/16SrRNA"
#set workspace
dir = os.chdir(workspace)
pd.set_option('display.max_rows', 500)
pd.set_option('display.max_columns', 500)
pd.set_option('display.width', 1000
!ls reportold/
listofviromesrna = []
for z in glob.glob("reportold/*"):
    #print(z.split("_")[1])
    y = z.split("_")[1]
    listofviromesrna.append(y)
#Create merged table

merged = pd.DataFrame(columns=['SAMPLE'])
for z in glob.glob("reportold/*"):
    name = z.split("_")[1]
    df = pd.read_csv(z,sep = "\t", header = None)
    idx = 0
    df.insert(loc = idx, column = "SAMPLE", value = name)
    merged = merged.append(df)
    #print(name)
merged.columns = ["SAMPLE","percentage","fragments", "directly", "tax","NCBI_number","name"]
selected = merged[merged["tax"]=="G"]
len(selected)
# Function to read each csv transformed file
def readreport(csv,taxname,threshold):
    samplename = csv.split("_")[1]
    dffile = pd.read_csv(csv, sep = "\t",header = None)
    dffile["SAMPLE"] = samplename
    #print("check")
    dffile.columns = ["percentage","fragments", "directly", "tax","NCBI_number","name","SAMPLE"]
    num = dffile[dffile["name"].str.contains("Viruses")].index
    dffile = dffile.iloc[num[0]:]
    dffile = dffile[dffile["fragments"]>= threshold]
    #print("check")
    selected = dffile[dffile["tax"] == taxname]
    #print("check")
    finalselected = pd.DataFrame()
    finalselected[samplename] = selected["name"]
```

```
    return finalselected.reset_index(drop=True), selected
#Checkpoint
len(df["SAMPLE"].unique())
#Pivoting table
z = pd.pivot_table(df, values= "fragments", index = "name", columns = "SAMPLE")
```

AmpliSeq sequencing yielded with VCF files as a result. Data such as SNP's reference, position, quality, gene were derived from this data and merged into one csv file to make it readable for python libraries used for statistical data analysis. Bash shell was used and an individual script was created, this script is presented in the Table 5.

Table 5. Script used for merging VCF format files. Jupyter lab software was used to perform all actions in cited scripts. For better readability for the purposes of this document, the individual cell results have been omitted.

```
#Import presets
import pandas as pd
import glob,os
import io
import os
import pandas as pd
import re
workspace = "/home/fagi/SNP/"
dir = os.chdir(workspace)
pd.set_option('display.max_rows', 1500)
pd.set_option('display.max_columns', 1500)
pd.set_option('display.width', 1500)
#Read VCF files based on github.com/dceoy/ repository
#Create dataframe with sample columns
merged = pd.DataFrame(columns=['SAMPLE'])
#look for files in specific folder
for z in glob.glob("R_2018_12_06_06_15_41_user_SN2-127-43-49_SNP_Ludzie/VCF_files/*vcf"):
    #read datafile
    df = read_vcf(z)
    idx = 0
    #insert columns with sample
    df.insert(loc = idx, column = "SAMPLE", value = z)
    #append next dataframe to it
    merged = merged.append(df)
    #read only specific columns
    merged = merged[['SAMPLE', "CHROM", 'POS', "REF", "ALT", "INFO"]]
    merged["SNP"] = merged["CHROM"].astype(str)+":"+merged["POS"].astype(str)
```

```
#Define annotated file
def readannotated(datafile):
    #Read file
    df=read_vcf(datafile)
    #Set index to 0
    idx=0
    #insert new column
    df.insert(loc = idx, column = "SAMPLE", value = datafile)
    #Pick only few columns from file
    df = df[['SAMPLE', "CHROM", 'POS', "REF", "ALT", "INFO"]]
    #Create columns SNP with merging Chrom and Pos
    df["SNP"] = df["CHROM"].astype(str)+":"+df["POS"].astype(str)
    #separate column Info
    info = df["INFO"].str.split("ANN=", expand = True)
    #separate info by |
    gene = info[1].str.split("|", expand = True)
    #create gene column
    df["GENE"] = gene[3]
    #create function column
    df["FUNCTION"] = gene[7]
    #create type column
    df["TYPE"] = gene[1]
    df["mutation"] = df["REF"]+df["ALT"]
    df["SNPmutation"] = df["SNP"]+" | "+df["mutation"]+" | "+df["GENE"] + " | "+df["FUNCTION"]
    return df
#import and merge annotated files
z = 0
merged = pd.DataFrame()
testing = pd.DataFrame()
for file in glob.glob("*/VCF_files/Annotated*.vcf"):
    #print(file)
    try:
        df = readannotated(file)
    except KeyError:
        print(file)
        z = z+1
        pass
    merged = merged.append(df)
print(z)
```

Data exported from BaseSpace Illumina sequencing and analyze yielded with many output files from BLAST aligner. Those files were imported to Jupyter Lab environment using Pandas (Python Data Analysis Library). The files contained information about aligned sequences and

NCBI taxid numbers. To retrieve full taxonomy data, such as classification and nomenclature for all of the organisms annotated in public reference databases, taxonomizr (0.9.3) package was used in RStudio 4.1 environment. Its function is to allocate taxonomy to an NCBI accession number. This package allowed to download data dumps (last updated on 26.02.2022) and to create a local repository for custom taxonomic assignment. R 4.1 was used to create individual script. It is presented in the Table 6.

Table 6. An individual script created for identification of organisms' taxonomy based on taxomizr package. The script automates annotation based on taxids presented in output BLAST files.

```
#read commands from bash input
args <- commandArgs()
#load needed libraries
library("dplyr")
library("taxonomizr")
#read blastn result
sampledf <- read.csv(args[1], sep="\t",header = FALSE) %>% select(V2,V10, V11)
#transform to list unique accession numbers
array <- as.character(unique(dplyr::pull(sampledf, V2)))
#identify taxids based on accession numbers
taxidarray <- accessionToTaxa(array, "accessionTaxa.sql", version = c("version","base"))
#create table with accession numebers and taxids
sampletaxdf <- data.frame(array, taxidarray)
#Assign taxonomy to taxids
taxonomylevels <- getTaxonomy(taxidarray, sqlFile = "nameNode.sqlite",desiredTaxa = c("family",
"genus",
                                                    "species"))
# make final table and save file as csv
finalidentification <- bind_cols(sampletaxdf,taxonomylevels)
write.csv(finalidentification, args[1])
```

Finally, data derived from FASTQ files were saved as csv and proceeded for statistical data analysis. All data were normalized against total number of reads in each sample.

9.13. Statistical Data Analysis

9.13.1 Comparison of detected SNP distribution to microbiome diversity (Shannon)

Correlation between detected SNPs and the diversity index in 16S rRNA and bacteriophage elements of microbiome was estimated using Shannon-Wiener equation that is the most widely used indicator of biodiversity (Spellerberg & Fedor, 2003). Its value determines the probability that two random elements from a sample belong to different species. Taxonomy level of species was used to calculate the index value for bacteria. Function based on provided equation was used to calculate index values and showed in the Table 7 and Equation 1. respectively.

$$H=-\sum_{i=1}^{s}(p_i \ln p_i)$$

Equation.1 Shannon Wiener Index equation. *s* - the number of Operational Taxonomy Units (OTUs). *pi* - the ratio of the number of individuals of a given OTUs by *i*.

Table 7. Shannon Wiener Index equation transformed into Python function to calculate values for specified samples.

```
def calculateshannon(row):
    table["taxcol"] = row
    table["log"] = np.log(table["taxcol"])
    table["log*pi"] = table["taxcol"]*table["log"]
    shannon = -table["log*pi"].sum()
    return(shannon)
```

9.13.2 Principal Component Analysis (PCA) of microbiome diversity data

PCA was performed using Rstudio environment in 3.1 version. Following libraries were imported and applied: “dplyr”, “tibble”, “ggplot2”, “ggfortify”,tibble”,”stringr”,”data.table”; function (code) created for this analysis is presented in the Table 8.. PCA is a method of dimension reduction in analysis of large datasets with a high number of dimensions (variables). In my data, each variable like bacterial genus or bacteriophage family can be presented as a different dimension. PCA is primarily used to reduce the variables describing a phenomenon and to discover possible regularities between features. It allows for identification of variables that have a strong effect on the appearance of individual principal components, that is, those

that form a homogeneous group. Finally, PCA plots are projections of multidimensional data on a 2-D plane. R 4.1 was used to create individual script. It is presented in the Table 8. Biplot Figures was created using “factoextra” library in R. Creating PCA plot is only possible if variance can be calculated, so AmpliSeq sequencing cannot be visualized that way. Dimensions used in these Figures were chosen based on their variance coverage from scree plots generated using the same library in R. Scree plot data is presented in the Figure 3.

Table 8. A script used to implement function for PCA analysis on csv generated from 16S rRNA, phageome, and SNPs identification. R in 3.1 version was used.

```
createpca <- function(csv, patientfile, settitle, column) {
  #import patient file
  patientscsv <- read.csv(patientfile, sep =";")
  #import data file
  print(csv)
  my_data <- read.csv(csv,sep = "\t")  %>%
    select(-X) %>% select(patients) %>% column_to_rownames("tax")
  my_data[is.na(my_data)] <- 0

  #Count matrix
  pca <- prcomp(my_data, scale = T)

  #Manipulate tables
  pca_df <- as.data.frame(pca$rotation)
  #extract from index
  pca_df <- tibble::rownames_to_column(pca_df, "patientnr")
  #join the Table with patients
  pca_df <- dplyr::left_join(pca_df,patientscsv, by = "patientnr")

  #PCAloadings <- data.frame(Variables = rownames(pca$rotation), pca$rotation)
  #Plot PCA
  pcaplot0 <- ggplot(pca_df,aes(PC1,PC2)) +
    geom_point(aes(color=get(column)),size = 10) +
    geom_text(label = pca_df$patientnr, size = 1) +
    labs(title=settitle,
         x ="PC1", y = "PC2")
  ggsave(paste0("PCAplots/",column,".jpg"))
  return(pcaplot0)
}
```

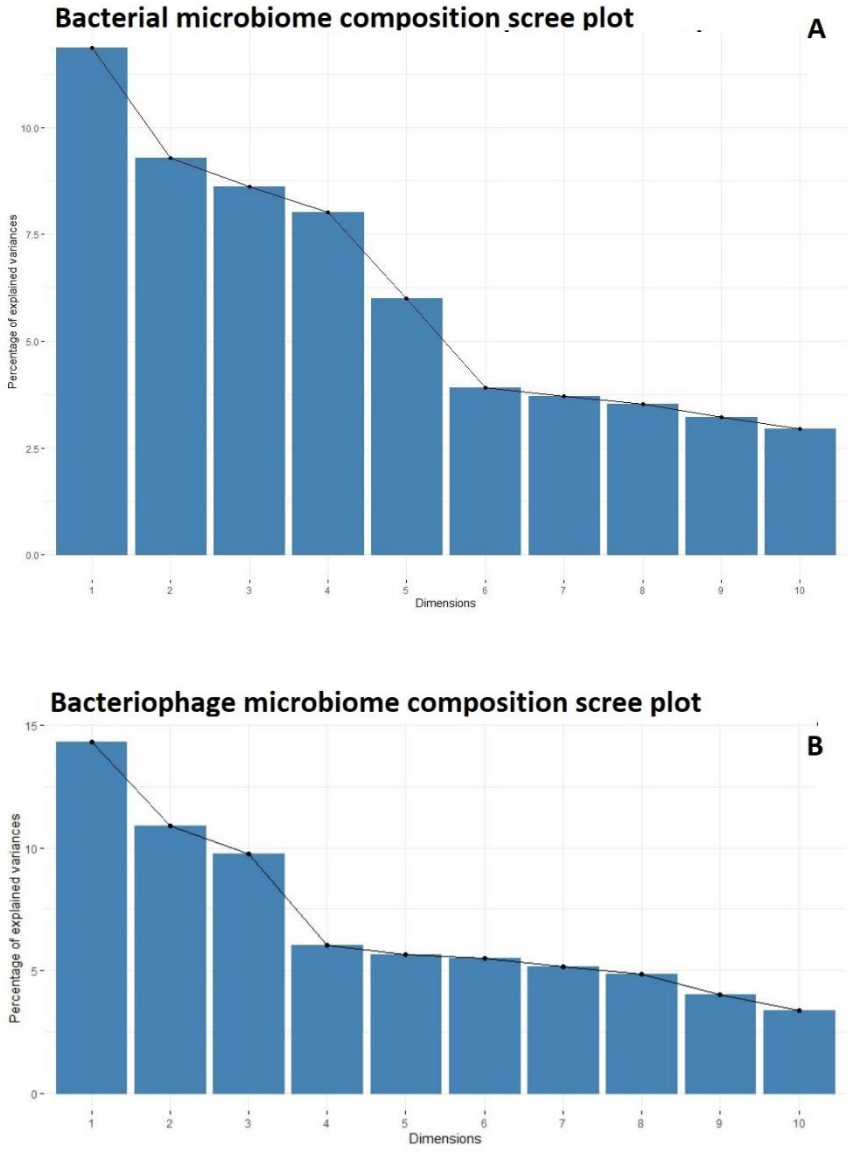

Figure 3. Scree plots presenting eigenvalues for each individual Dimension. Variances were calculated using *get_eigenvalue()* function in R's factoextra library. Two dimensions that cover most of the percentage of explained variances were chosen for biplot creating. Panel (A) is related to bacterial microbiome composition data, while panel (B) to bacteriophage microbiome composition data.

PCA plots in 3 dimensions were created using "pca3d" library in R. using a individual created function. Variables calculated previously were used in pca3d function to visualize 3 dimensions at one time and to mark outliers.

9.13.3 Analysis of correlation between SNPs distribution and occurrence of bacteria presence using Cramer's V test

Cramer's V test was used as a measure of association between two nominal variables, based on Pearson's chi-squared statistic. This assumption is satisfied by the factors that we can specify in the binary system. Where their occurrence is the number 1, and their absence is the number 0. For this purpose, contingency tables were created using Pandas Dataframes. Cramer's V equation was transformed into Python 3.6 function and presented in the Table 9.

Table 9. Individual function to implement Cramer's V test for nominal variable calculation in Python 3.6

```
def cramers_corrected_stat(cramermatrix):
    chi2 = scipy.chi2_contingency(cramermatrix)[0]
    number1 = cramermatrix.sum()
    phi2 = chi2/number1
    z,v = cramermatrix.shape
    phi2corr = max(0, phi2 - ((k-1)*(r-1))/(number1-1))
    rcorr = z - ((z-1)**2)/(number1-1)
    kcorr = v - ((v-1)**2)/(number1-1)
    return np.sqrt(phi2corr / min( (kcorr-1), (rcorr-1)))
```

9.14. Test of proportion

Correlation between bacterial and bacteriophage presence was calculated using test for proportions based on normal z-test. Function from statsmodel library coded in Python 3.6 was used to compute results. Formula used for calculation is presented in Equation 2. Quantitative data from microbiome composition sequencing was transformed into categorical where variables were defined as "present" and "absent" in relation to bacteriophages or bacteria.

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Equation 2. Normal z-test equation used in statsmodel library in function *statsmodels.stats.proportion.proportions_ztest*¶ written in Python 3.6. $\hat{p}$ – sample proportion, p – population proportion, n – sample size.

10. Results

10.1. Samples collected from patients

148 gastric samples were collected from patients over 2 years of collection in the Endoscopy Department, Regional Specialist Hospital in Wroclaw. Among the patients included in the study, there were 91 women and 56 men. 141 Blood samples was collected from these patients. The average age of women was 54, and men was 59. During the assessment by the physicians, ICD (International Classification of Diseases) codes corresponding to the syndromes labeled were assigned to each patient (Drösler et al., 2021). All the patients' ICDs are shown in the Table 10.

Table 10. ICD-10 classified diseases diagnosed by physicians during endoscopy examination of patients enrolled to this study

ICD Code	ICD Meaning	Number of diagnosed patients
D37.1	Neoplasm of uncertain or unknown behavior: Stomach	1
D48	Neoplasm of uncertain or unknown behavior of other and unspecified sites.	1
D50	Iron deficiency anemia	1
D50.9	Iron deficiency anaemia, unspecified	1
K21	Gastro-esophageal reflux disease	17
K25	Gastric ulcer	1
K26	Duodenal ulcer	1
K26.7	Duodenal ulcer Chronic without haemorrhage or perforation	2
K29	Gastritis and duodenitis	64
K29.5	Unspecified chronic gastritis	5
K29.6	Other gastritis	2
K29.9	Gastroduodenitis, unspecified	1
K30	Functional dyspepsia	2
K31	Other diseases of stomach and duodenum	1

	Neurotic	
K31.8	Other specified diseases of stomach and duodenum	3
K44	Diaphragmatic hernia without obstruction or gangrenę	1
K52.9	Noninfective gastroenteritis and colitis, unspecified.	1
K59	Other functional intestinal disorders	1
K63	Other diseases of intestine	2
K86	Other diseases of pancreas	1
K92	Other diseases of digestive system	1
Q40.1	Congenital hiatus hernia.	1
R05	Cough	1
R10	Abdominal and pelvic pain	5
R10.1	Pain localized to upper abdomen	3
R10.4	Other and unspecified abdominal pain.	9
R11	Nausea and vomiting	1
R29.8	Other and unspecified symptoms and signs involving the nervous and musculoskeletal systems	1
R53	Malaise and fatigue	1
Z03.8	Encounter for observation for other suspected diseases and conditions ruled ou	1
Z03.9	Observation for suspected disease or condition, unspecified	1

10.2. Bacterial DNA sequencing

In total, 145 tissue bioptates were proceeded for bacterial DNA isolation using Micro Beat Bead Gravity AX kit (A&A Biotechnology). Process was successful in 144 cases which were forwarded for 16S rRNA library preparation. Final quantification of library using Ion Universal Library Quantification Kit (Thermofisher) showed that in 128 samples the library was correctly prepared and barcoded. Seven sequencing sessions were performed for bacterial microbiome composition investigation. Reports from Ion Torren PGM system are presented in the Figure 4.

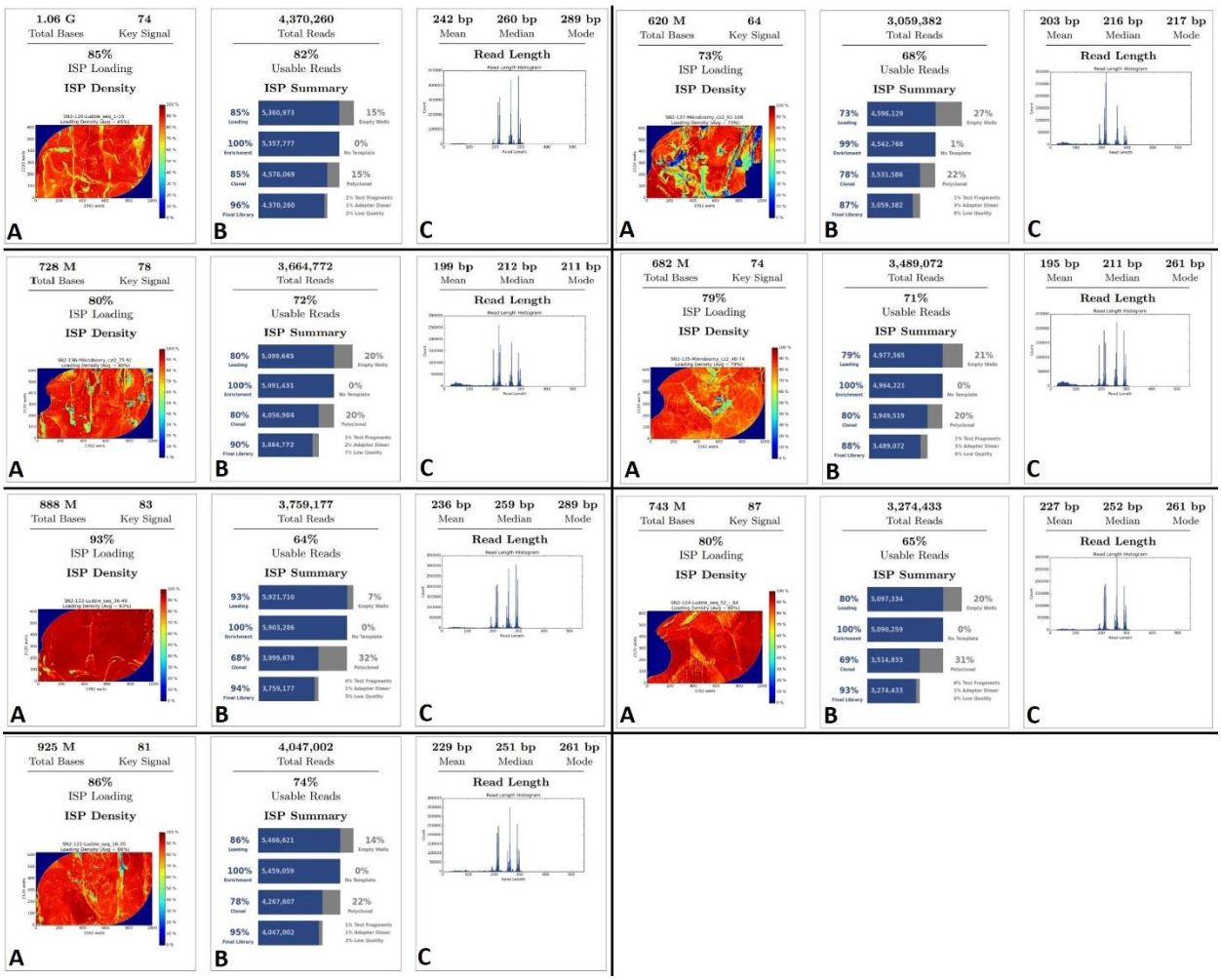

Figure 4. Reports for 16S rRNA amplicon sequencing generated with Ion Torrent PGM on IonTorrent Server. The panels for each sample present Quality control data regarding 16S rRNA amplicons. Panel (A) - Distribution of ISPs on the chip; redness increases with the density of beads at a given location on the chip. Panel (B) - Statistics on sequenced ISPs. The individual percentages show whether the process of preparing the library was carried out correctly. Panel (C) - Statistics on the single reads that were successfully generated during sequencing.

10.3. Ampliseq human genome sequencing

Blood samples (141) were collected from patients of Endoscopy Department, Regional Specialist Hospital in Wroclaw. DNA was isolated using Sherlock AX (A&A Biotechnology). Process was successful in 111 samples, and these were forwarded to library preparation using

Ion Ampliseq Library Kit Plus (Thermofisher Scientific). Final quantification of library using Ion Universal Library Quantification Kit (Thermofisher Scientific) showed that in 72 samples the library was correctly prepared and barcoded. Twelve sequencing sessions were performed for Ampliseq SNPs investigation. Ten example reports from Ion Torren PGM system are presented in the Figure 5.

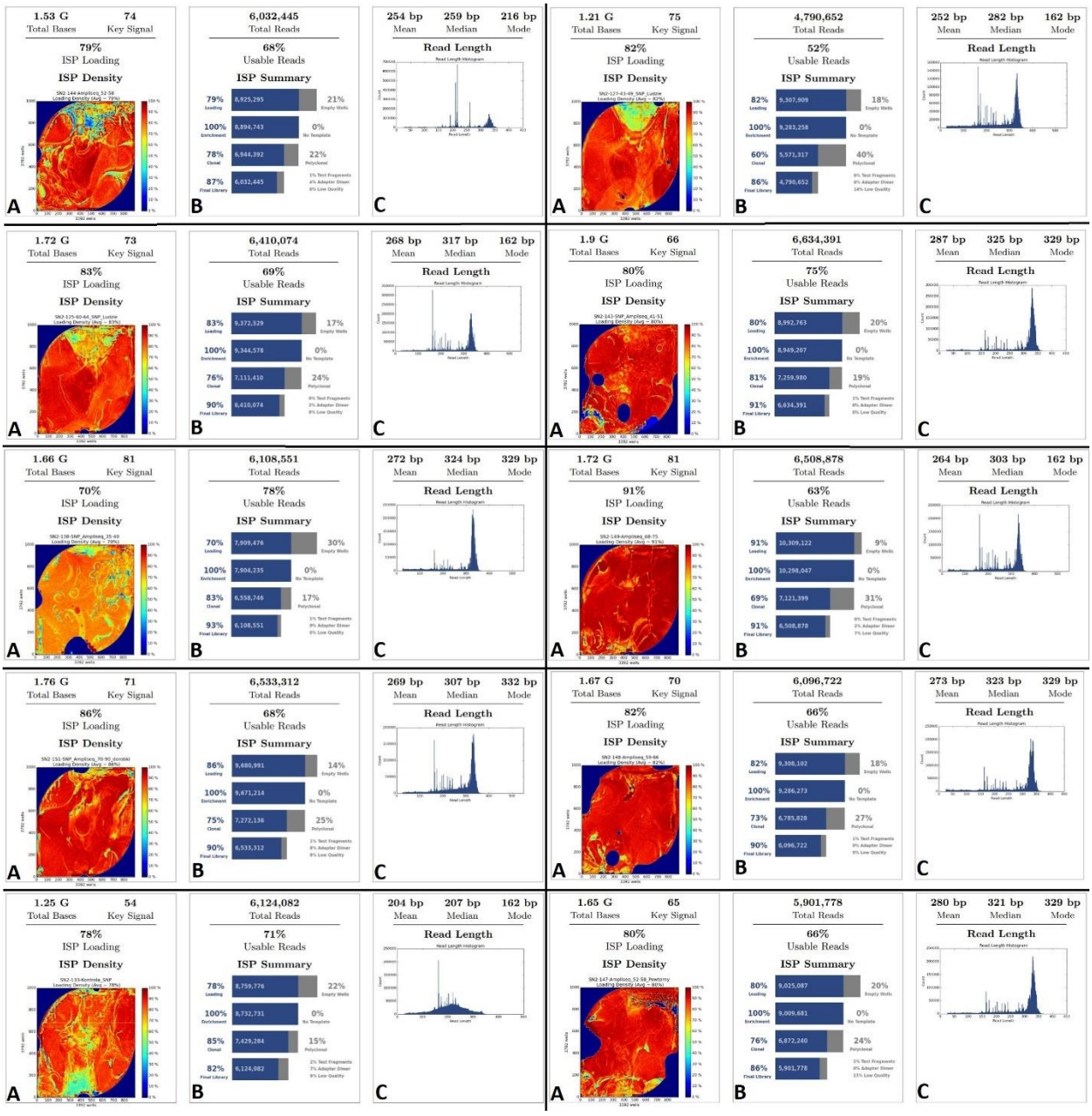

Figure 5. Selected reports for Ampliseq amplicon sequencing generated with Ion Torrent PGM on IonTorrent Server. 3 Panels to each sample present Quality control data regarding Ampliseq

amplicons. Panel (A) - Distribution of ISPs on the chip. The redder the color, the higher the density of beads at a given location on the chip. Panel (B) - Statistics on sequenced ISPs. The individual percentages show whether the process of preparing the library was carried out correctly. Panel (C) - Statistics on the single reads that were successfully generated during sequencing.

10.4. Phageome Sequencing

In total, 145 solutions with viral part of stomach microbiome were proceeded for ultracentrifugation and further DNA isolation using Sherlock AX (A&A Biotechnology). Final DNA quantification was performed using QuantiFluor ds DNA system 1 ml (Promega) and Quantus Fluorometer (Promega). 101 samples were forwarded for sequencing library preparation using Illumina DNA prep kit (Illumina). In two cases process of creating the library was failed. Finally, 99 samples were sequenced during 4 sequencing sessions using NextSeq 550 device and NextSeq 500/550 Mid Output kit (Illumina). Viral part of the microbiome sequencing was performed offline because of the local software incompatibility with online Illumina software, therefore automatic reports were not generated as IonTorrent Server did in previous sequencing sessions. Nevertheless, each file generated by the sequencer was manually checked with FASTQC software. An example sequencing report is shown in the Figure 6.

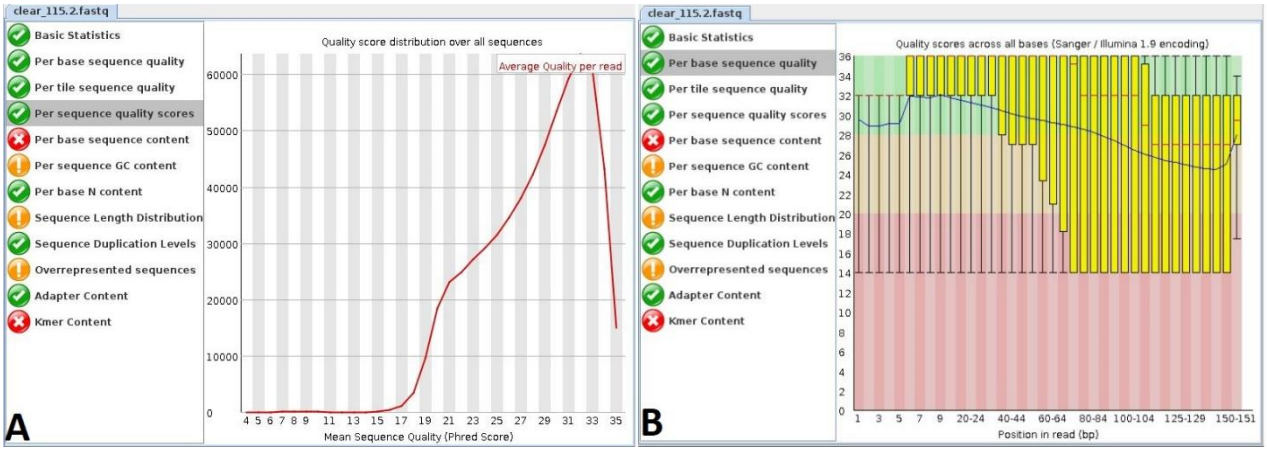

Figure 6. An example reports for Illumina sequencing generated using FASTQC software. Panel (A) – distribution of quality scores among generated reads. Panel (B) – distribution of Phred quality scores among single base pairs in generated reads.

10.5. Microbiome components correlate with disease states

To evaluate possible correlation between observed diseases (comorbidities) and microbiome composition 143 samples were sequenced to identify bacterial 16S rRNA coding regions. Using z-test function implemented in Python library, p values were calculated. Ninety-nine phageome samples were analyzed using the same test to evaluate specific elements of phageome presence in samples from patients with diagnosed diseases in accordance with the International Classification of Diseases.

10.5.1. Bacterial components in specific disease states

Fourteen correlations were found by statistical analysis of bacterial components and ICD-classified diseases. It turned out, that associations were related to diseases classified as: K29 – "Gastritis and duodenitis", R.10.4 – "Other and unspecified abdominal pain", K21 – "Gastro-oesophageal reflux disease". One proportion z-test was used to compare proportion in patients where ICD-classified disease was diagnosed (observed group) and patients without any disease diagnosed or claimed in the interview (control group). All correlations were defined in CI at level of 0.95 and with p-value lower than 0.05. Bacterial species listed in the Table 11. were enriched in patients with specific disease states.

Table 11. Correlations between bacterial species found in the microbiome composition and ICD classified illnesses diagnosed in patients. P-value was calculated using z-test function from statsmodels Python 3.6 library.

Bacterial species	ICD classification	p-value
Escherichia albertii	K29	0.02
Variovorax sp. PAMC 28711	K29	0.01
Fusobacterium nucleatum	K29	0.02
Escherichia albertii	R10.4	0.01
Klebsiella pneumoniae	R10.4	0.02
Klebsiella variicola	R10.4	0.03
Raoultella ornithinolytica	R10.4	0.03
Serratia marcescens	R10.4	0.01
Variovorax sp. PAMC 28711	R10.4	0.03
Phreatobacter cathodiphilus	R10.4	0.03

Escherichia albertii	K21	0.03
Hydrogenophaga sp. PBC	K21	0.03
Variovorax sp. PAMC 28711	K21	0.03
Haemophilus sp. oral taxon 036	K21	0.02

10.5.2. Phageome components in specific disease states

Three correlations were found by statistical analysis of bacteriophage components with ICD-classified diseases. Associations are present among diseases classified as: K29 – "Gastritis and duodenitis", K63 – "Other diseases of intestine", K29.6 – "Other gastritis". Those correlations were defined in CI at level of 0.95 and with p-value lower than minimum 0.05. One proportion z-test was used to compare proportion in patients where ICD-classified disease was diagnosed (observed group) and patients without any disease diagnosed or claimed in the interview (control group). Bacteriophage family listed in the Table 12. were enriched in specific disease states.

Table 12. Correlations between bacteriophage genera found in the microbiome composition in patients and ICD classified illnesses diagnosed in these patients. P-value was calculated using z-test function from statsmodels Python 3.6 library.

Bacteriophage family	ICD classification	p-value
Clostridioides prophages	K29	0.005
Tequatrovirus	K63	0.02
Inovirus	K29.6	0.02

10.6. SNPs correlated to microbiome diversity inex

Among the sequenced samples, 44 samples were applicable for analysis of possible correlations between SNPs and bacterial microbiome diversity. Analysis revealed 13 SNPs which correlated to significant microbiome diversity change, as demonstrated by changes in the value of the Shannon index (t-test was performed using CI = 95% with $p < 0.05$). Alignments at the level of bacterial species were analyzed, that is at the lowest taxonomy level that was annotated in FASTQ analysis. Changes in Shannon index values correlated to the identified SNPs and the reference allele is shown in the Table 13. and in the Figure 7.

Table 13. SNPs that correlate with significant change in bacterial microbiome composition measured by Shannon-Wiener Index. T-test was performed using CI = 9% and $p<0.05$

Variant	Gene	Variant type	Change in Shannon index value
G>A	TLR5	Stop gained	Increase
C>T	IL-1B	Upstream gene	Decrease
T>C	TLR2	Synonymus	Increase
G>A	TLR10	Upstream gene	Decrease
T>A	TLR1	Missense	Decrease
dupCT	IL6	Upstream gene	Increase
TCT>CCC	IL6	Upstream gene	Increase
A>G	IL6	Upstream gene	Increase
C>T	IL6	Upstream gene	Increase
C>A	IL6	Upstream gene	Increase
C>G	IL6	Upstream gene	Increase
T>G	IL6	Upstream gene	Increase

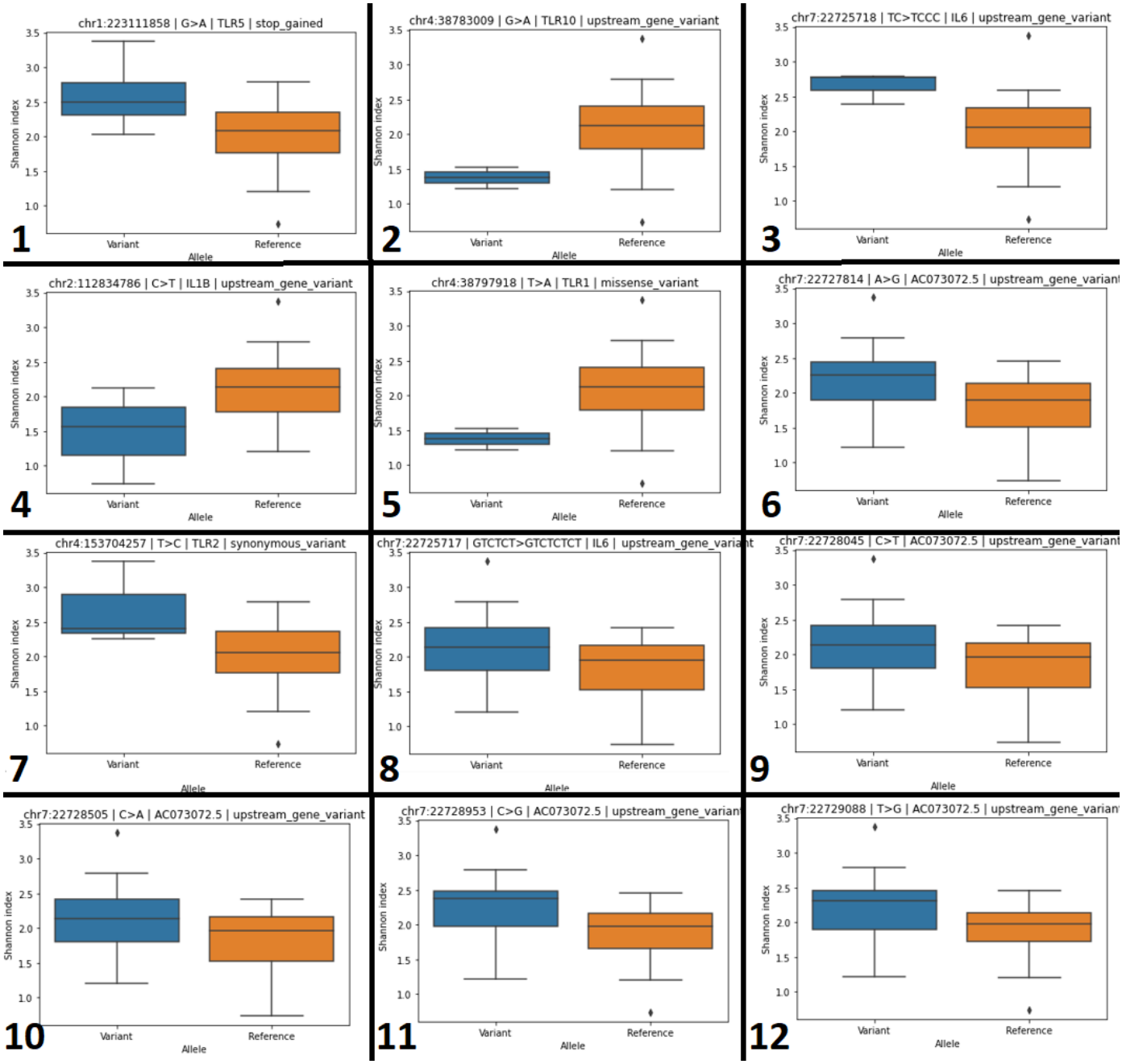

Figure 7. SNPs which correlate with significant change in Shannon-Wiener diversity index. Significance was calculated with Confidence Interval = 95% for $p<0.05$ using t-test. Results for samples with an identified single nucleotide allele are marked as "variant". Results for samples without an identified single nucleotide allele are marked as "Reference".

10.7. Characteristic microbiome elements that correlate to decreased diversity in bacterial microbiome

10.7.1 Bacteriophage groups correlated to lower diversity among bacterial population

Similar to analysis of SNPs, correlation between bacteriophage occurrence and Shannon diversity of bacterial microbiome composition based on 16S rRNA genes sequencing is presented in the Figure 8. Two genera of bacteriophages turned out to be significantly associated to bacterial composition of patients' microbiome. Presence of Brussowvirus and Triavirus genus in the human microbiome correlated to significantly decreased bacterial diversity measured by Shannon index with $p<0.05$ (estimated using t-test). Brussowvirus are viruses that infect bacteria from *Streptococcus sp.* while Triavirus is specific to *Staphylococcus sp.* Both groups represent long non-contractile tails morphology belonging to Siphoviridae.

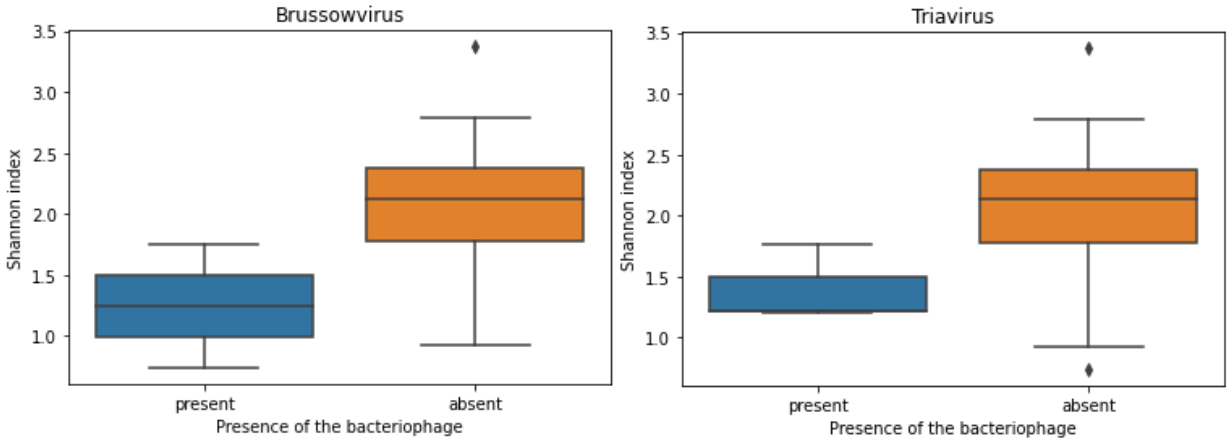

Figure 8. Bacteriophages presence which correlates with significant change in Shannon-Wiener diversity index. Significance was calculated with Confidence Interval = 95% for $p<0.05$ using t-test. Results for samples with specified bacteriophage are marked as "present". Results for samples without specified bacteriophage are marked as "absent".

10.7.2 Bacterial composition may correlate to bacterial population diversity in microbiomes

Shannon diversity was related to the occurrence of specific bacterial pathogens. They were selected on the basis of the literature analysis as related to gastritis. The analyzed species were: *Helicobacter Pylori* (Watari et al., 2014), *Rothia mucilanigosa* (J. Liu et al., 2018) *Prevotella melaninogenica* (T. Dong et al., 2017) *Neisseria subflava* (Nakamura et al., 2006)

Prevotella jejuni (Könönen & Gursoy, 2021a), *Salmonella enterica* (Zha et al., 2019). The analysis showed that microbiomes containing *S. enterica* and *P. melaninogenica* had significantly lower diversity than others. Presence of *R. mucilaginosa* correlated to significantly increased diversity expressed by Shannon index. Interestingly, no significant differences for *H. pylori* were observed. These results are presented in the Table 14.

Pathogen	Positives Alfa-Diversity	Negatives Alfa – Diversity	Significance
Helicobacter pylori	2.12	2.09	The *t*-value is -1.38324. The *p*-value is 0.08451. The result is *not* significant at $p < .05$.
Rothia mucilanigosa	1.78	2.09	The *t*-value is -2.46475. The *p*-value is 0.007524. The result is significant at $p < .05$
Prevotella melaninogenica	2.08	1.82	The *t*-value is 1.78231. The *p*-value is 0.038544. The result is significant at $p < .05$.
Neisseria subflava	2.06	2.11	Not enough data
Prevotella jejuni	2.07	1.72	The *t*-value is 1.33291. The *p*-value is 0.092474. The result is *not* significant at $p < .05$.
Salmonella enterica	2.08	1.63	The *t*-value is -2.65711. The *p*-value is 0.004447. The result is significant at $p < .05$.

Table 14. Correlations between presence of selected bacterial species associated with gastritis and diversity in patients' microbiomes. P-value calculated using t-test.

10.8. Principal Component Analysis (PCA)

To perform PCA analysis data table created with pandas Python 3.6 library was used. Reads annotated to specific Operational Taxonomy Units belonging bacteria and bacteriophages were normalized to number of total paired-end reads in each sample.

10.8.1 Biplots

Based on variance, a biplot was created to visualize which elements of the investigated microbiome components in represented dimensions. Figures were created using "factoextra" library in R. Creating PCA plot is only possible if variance can be calculated, so SNPs presence data cannot be visualized that way. A clustering with variance in individual components of the bacterial and viral microbiome was performed. The samples were also colored according to the single nucleotide variant they carried. Acute angle between two arrows in biplots means strong association between a corresponding row and column in the calculation. The direction and lengths of the vectors in relation to each other, expressed on the axes of the graph, indicate the influence in each dimension. In the Figure 9. it can be noticed that for instance *Synechococcus sp and Kocuria sp.* have a strong effect on sample clustering in a two-dimensional graph. Presence of *Corynebacterium sp.* is negatively associated with *Lactobacillus.*

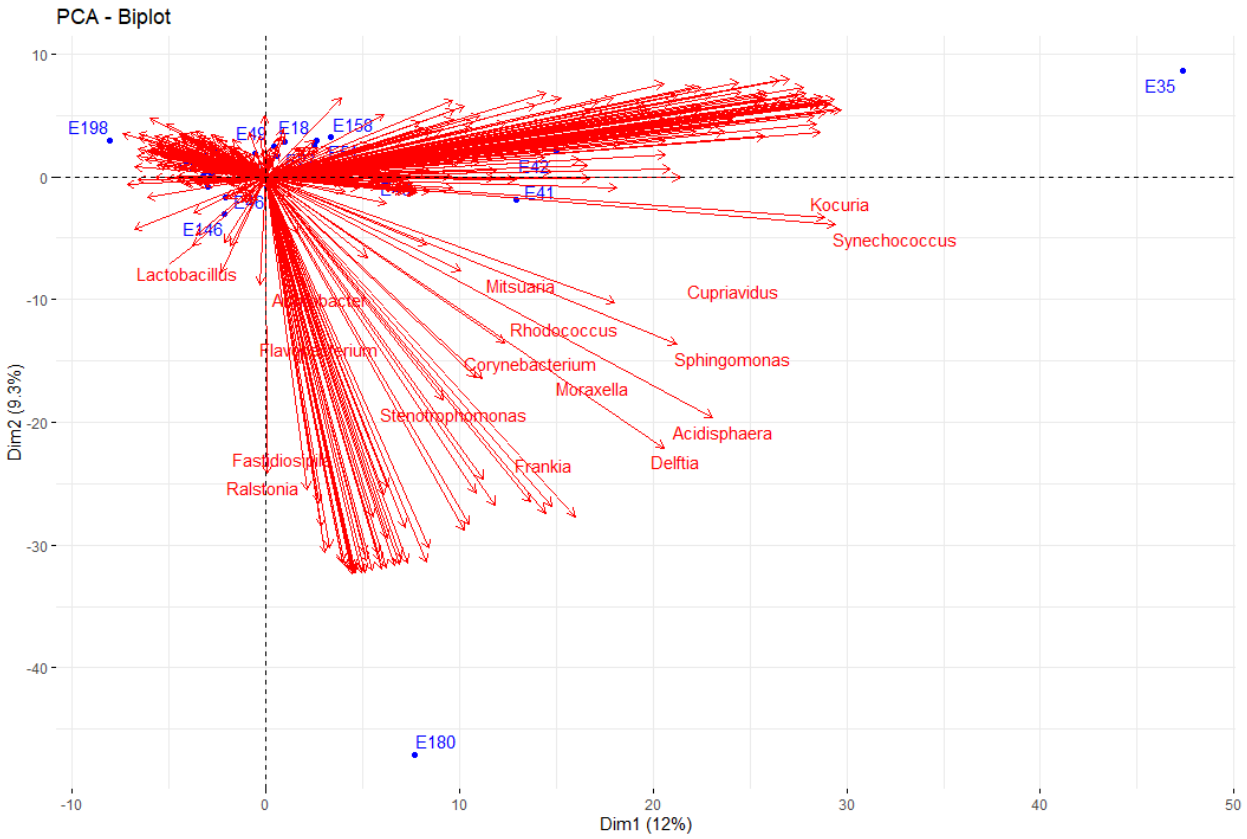

Figure 9. Two-dimensional biplot showing the relationship between the different genera of bacteria, which affects their clustering in PCA analysis. Patients with complete phage and bacterial microbiome composition data as well as sequencing SNPs were used for analysis. Analysis was done using bacterial microbiome composition data.

Phageome composition sequencing data was used to create biplot that show associations between different genus of bacteriophages. It can be observed that Tequatrovirus and Lambdavirus are negatively associated to Triavirus. Traversvirus, Triavirus, Bcepfunavirus, Lambdavirus and Tequatrovirus have the strongest impact on sample clustering. Results are present in the Figure 10.

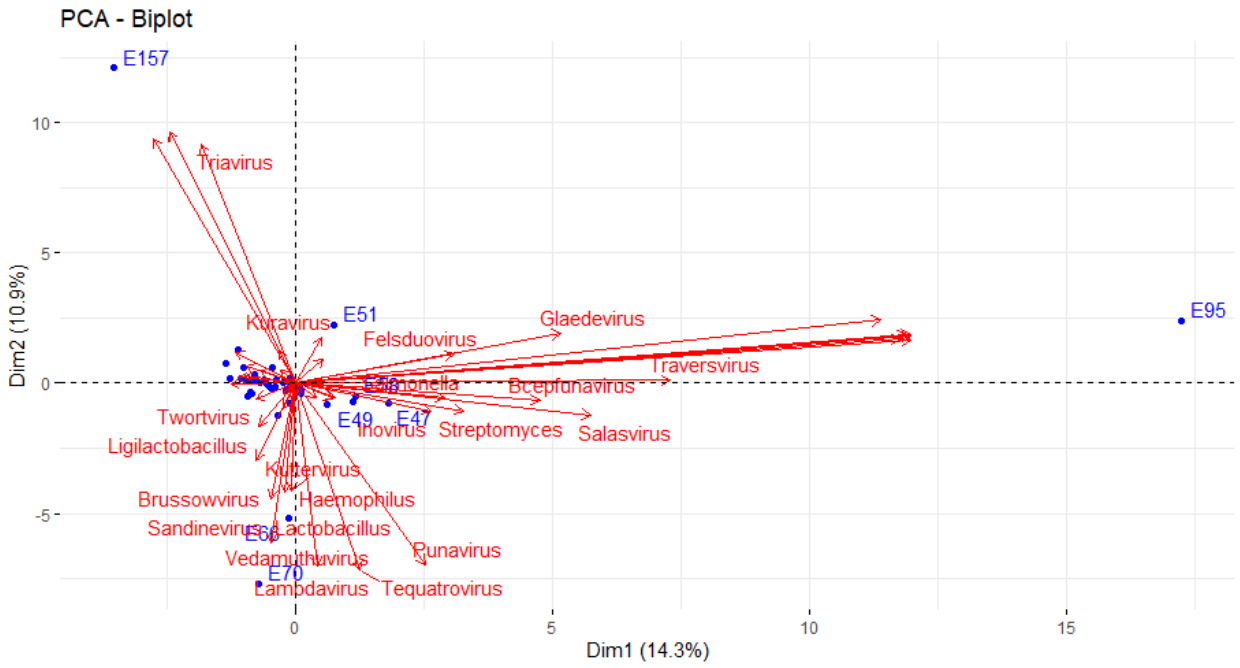

Figure 10. Biplot visualizing dimensions that represent **bacteriophage genera** derived from metagenomics sequencing and that have the highest influence on PCA plots. Patients with complete phage and bacterial microbiome composition data as well as sequencing SNPs were used for analysis. The presence of some bacterial groups in this Figure is due to their annotation in the NCBI database: some of the bacteriophages are prophages for instance: *Haemophilus*, *Streptomyces*, thus annotated as a part of bacterial genomes.

10.8.2 Clustering by disease state using Principal Component Analysis

PCA analysis did not reveal any clustering among data sorted by ICD diagnosed state – Figure 11. Panels were generated using *ggplot2* R package and *prcomp()* function. Samples were

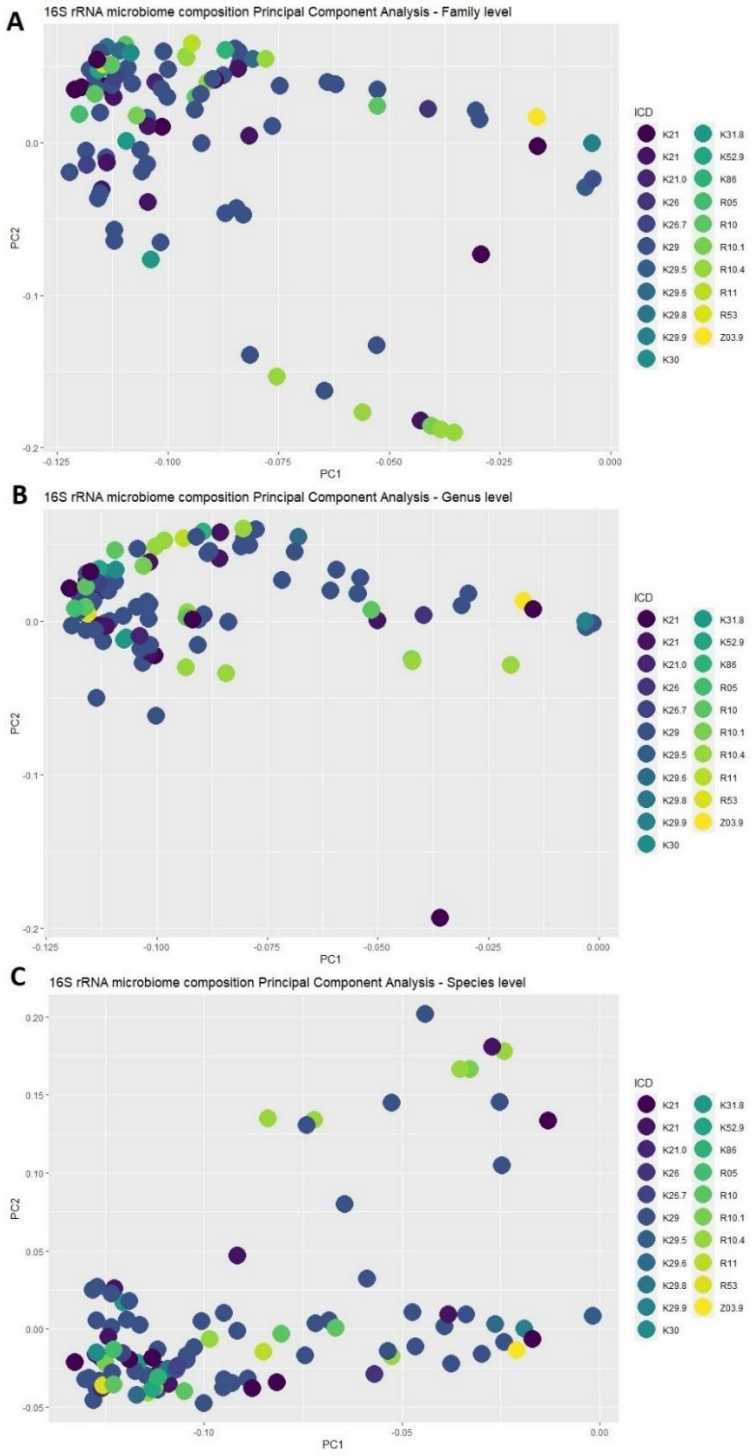

Figure 11. Principal Component Analysis applied to different Operational Taxonomy Levels in bacterial microbiome. Plot created using ggplot2 library in R. Panel (A) shows samples clustered by bacterial families microbiome composition, panel (B) shows samples clustered by bacterial genus microbiome composition, panel (C) shows samples clustered by bacterial species.

clustered according to variance calculated from bacterial microbiome quantitation data.

Differences between samples are insignificant to distinguish clusters among 21 different disease states according to ICD.

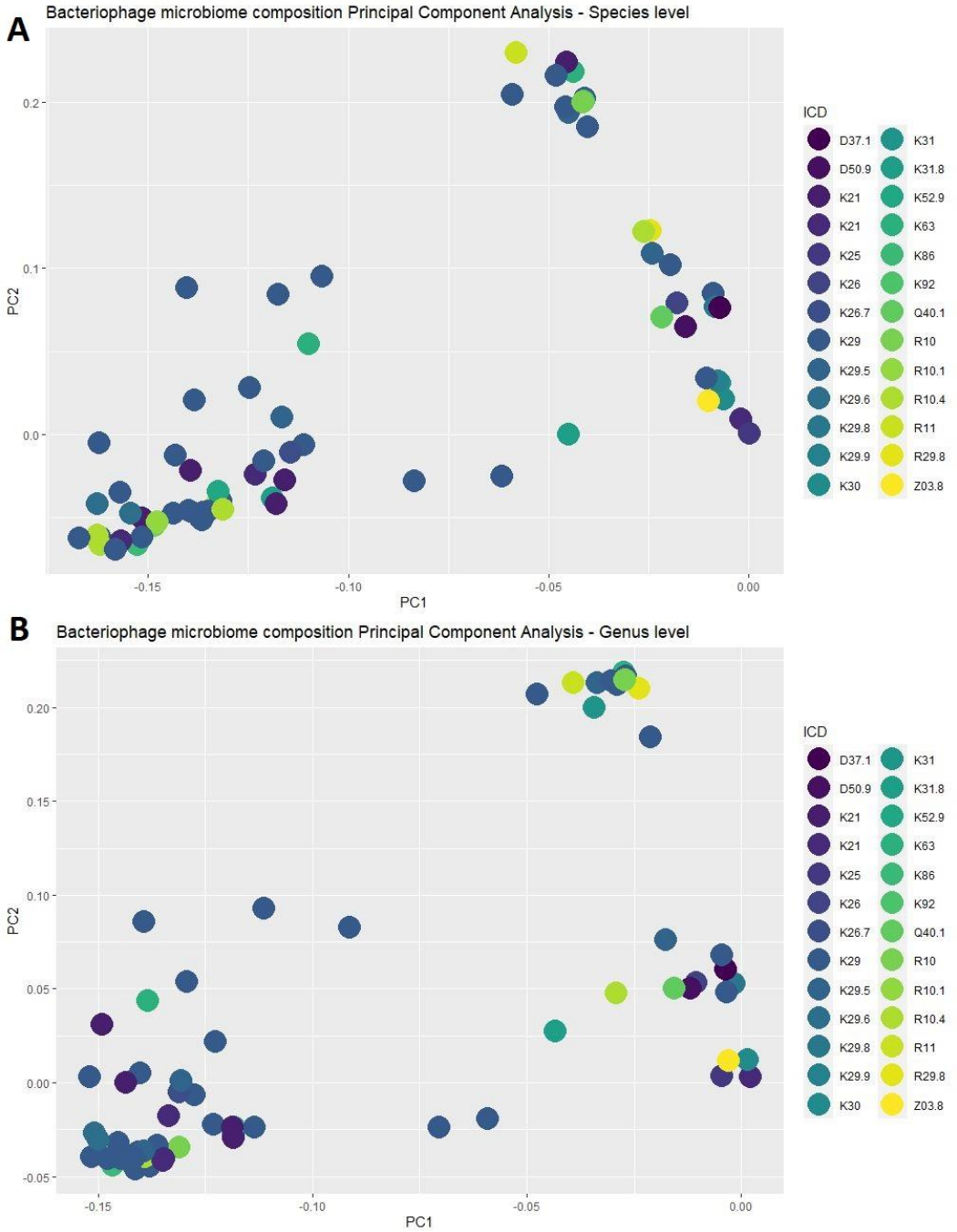

Figure 12. Principal Component Analysis applied to different Operational Taxonomy Units in phageome. Plot created using ggplot2 library in R. Panel (A) shows samples clustered by bacteriophages' species microbiome composition, panel (B) shows samples clustered by phages' genus microbiome composition.

Clustering by ICD diagnosed illness was done to phageome composition data. Differences between clustered samples were small and insignificant. Twenty-two diseases according to ICD were diagnosed in group of the study. Data is presented in the Figure 12. PC1 and PC2 components were chosen because they covered over 75% of Variance Explained on Scree plots presented in the Figure 13.

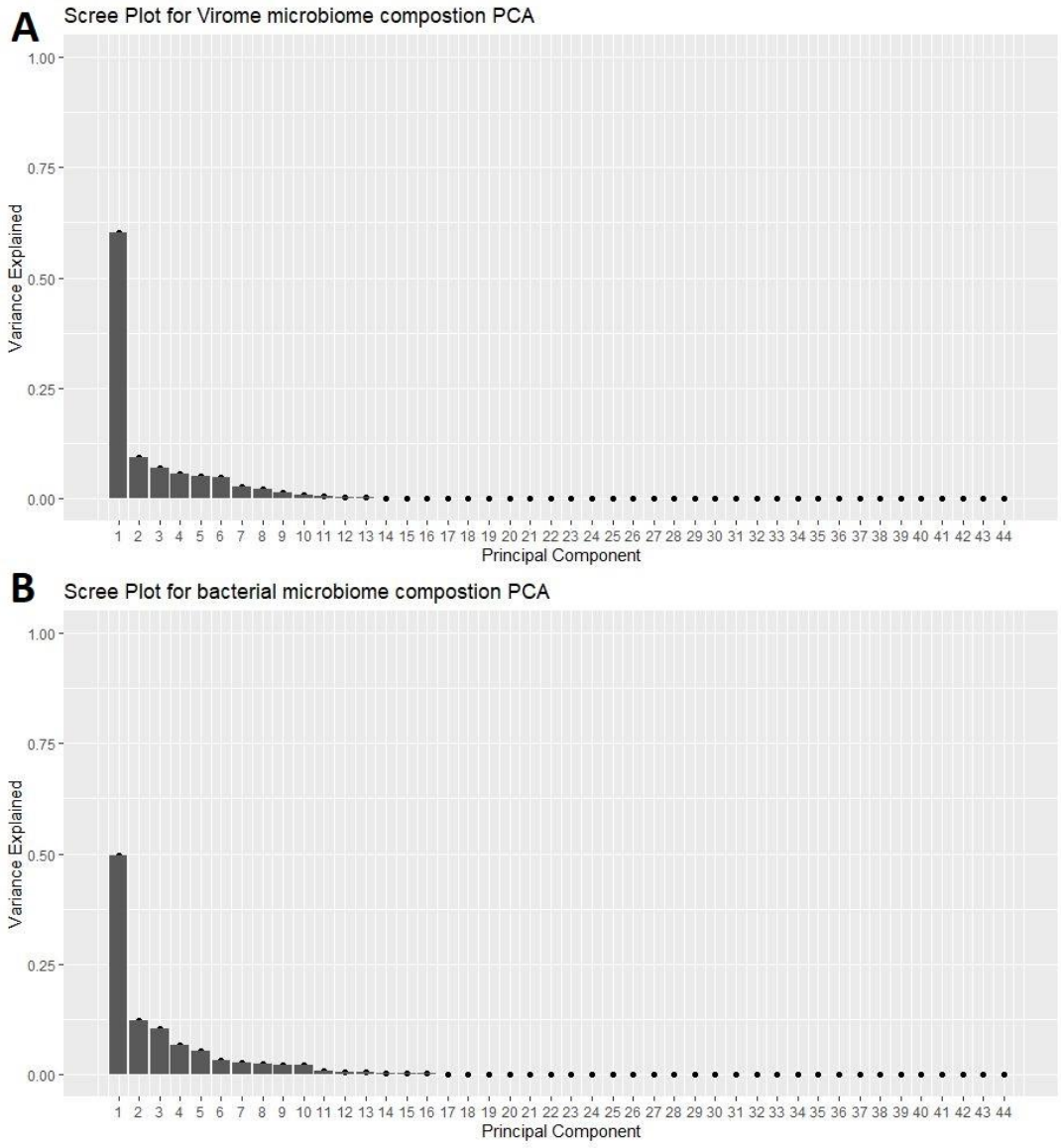

Figure 13. Scree plots generated to evaluate explained Variance among calculated Principal Components. Data generated using ggplot2 R' package. Panel (A) present scree plot for bacteriophages' microbiome composition while Panel (B) presents scree plot for bacterial microbiome composition.

10.9. Microbiome composition in relation to Single Nucleotide Polymorphisms

The clustering of samples based on the composition of the microbiome developed from 16S rRNA genes sequencing was evaluated. The samples were colored according to specific bacterial species that were identified. No significant correlations were found. Results are present in the Figure 14.

16S rRNA gene sequencing data was tested for 1190 possible variants clustering. It was done by PCA plotting visual rating. The number of variants in the analysis is equal to OTUs number annotated in all samples. SNPs couldn't be visualized with PCA, due to their binary character. Nevertheless, cross analysis including polymorphism occurrence was applied. SNPs were tested in 940 different configurations in relation to phageome and bacterial microbiome composition. No significant clustering was confirmed by visual observation of generated plots. Exemplary plots are presented in the Figure 15. where samples with complete 16S rRNA and SNPs sequencing data were grouped by the variance generated basing on the bacterial or bacteriophage microbiome composition.

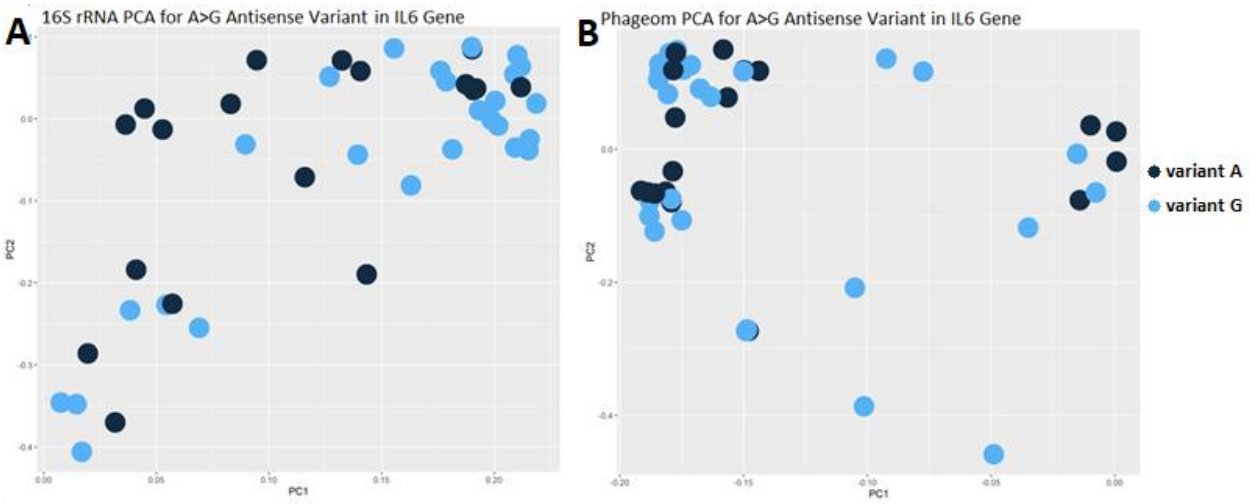

Figure 14. Principal Component Analysis applied to bacterial microbiome composition data. Plot created using ggplot2 library in R. Panel (A) presents data generated basing on bacterial composition of microbiome while panel (B) shows data for the bacteriophages' part of microbiome.

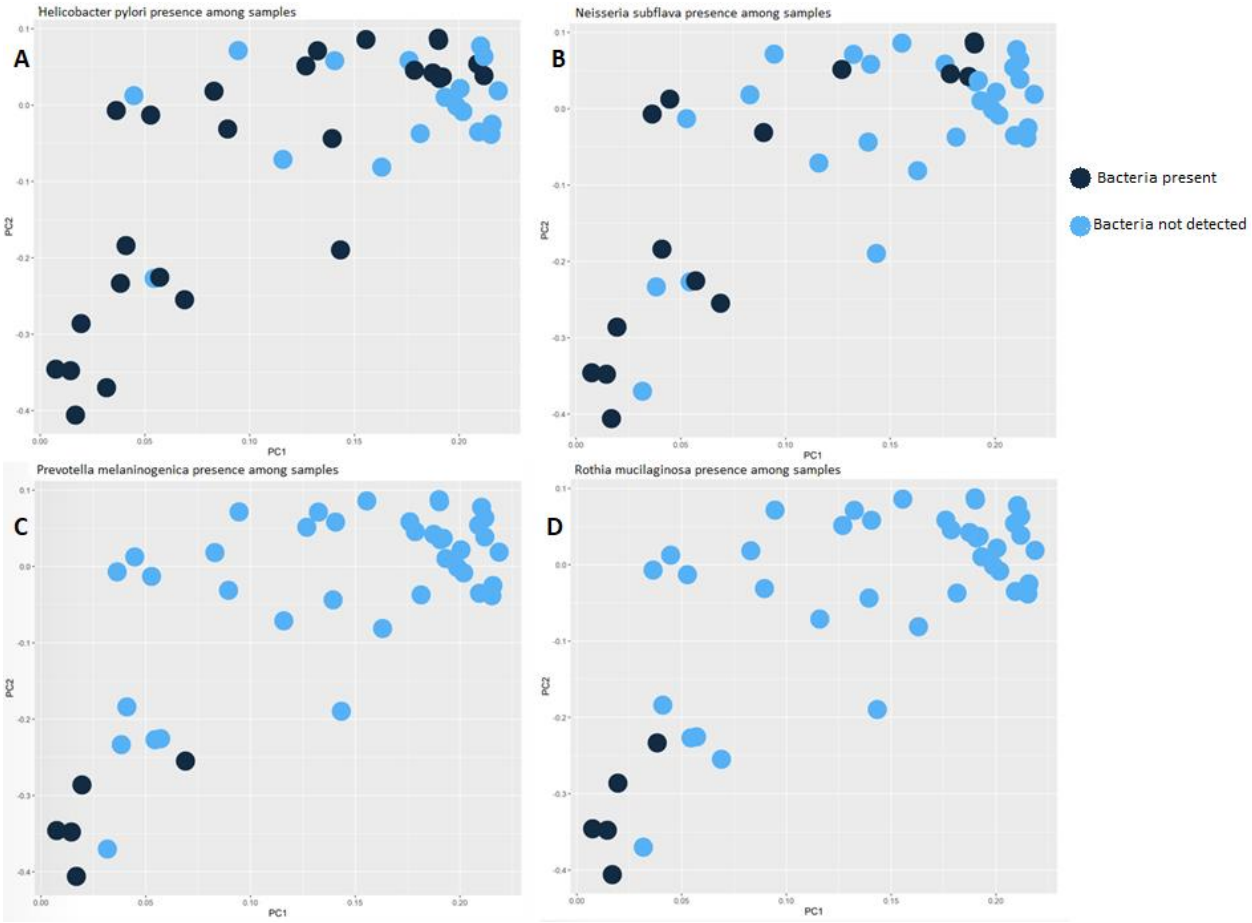

Figure 15. Principal Component Analysis applied to bacterial microbiome composition data and grouped by presence of pathogens related to gastritis. Figure created using ggplot2 library in R. Panel (A) present visualization of presence of *H. pylori*, Panel (B) is dedicated to *N. subflava* presence, Panel (C) is for *P. melaninogica*, Panel (D) shows data related to *R. mucialignosa*.

10.10. 3D Principal Component Analysis of microbiomes in the context of comorbidities

3D PCA analysis was performed using normalized results derived from 16S rRNA genes sequencing, phageome sequencing, and medical data from patients. Patients were sorted according to their health disorders.

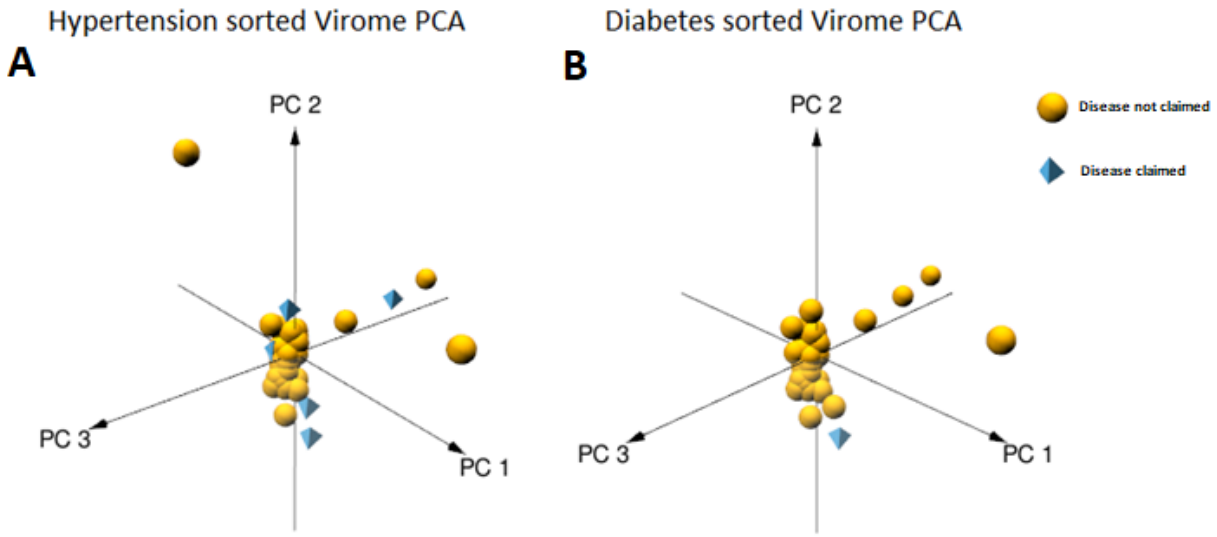

Figure 16. 3D Principal Component Analysis plotted based on phageome components of human microbiome; individual objects have been labeled according to whether specific chronic illness was diagnosed. Figure was created using *pca3d* package in R. Panel (A) shows bacteriophage composition data with labeled patients with claimed hypertension. Panel (B) shows bacteriophage composition data with labeled patients with claimed diabetes.

Analysis presented in the Figure 16 shows that phageomes of patients with chronic diseases tend to have significantly different composition to those in healthy ones. Similar analysis (Figure 16) that involved bacterial microbiome composition showed that these differences ware related only to bacteriophage composition while there were no significant differences among bacterial composition.

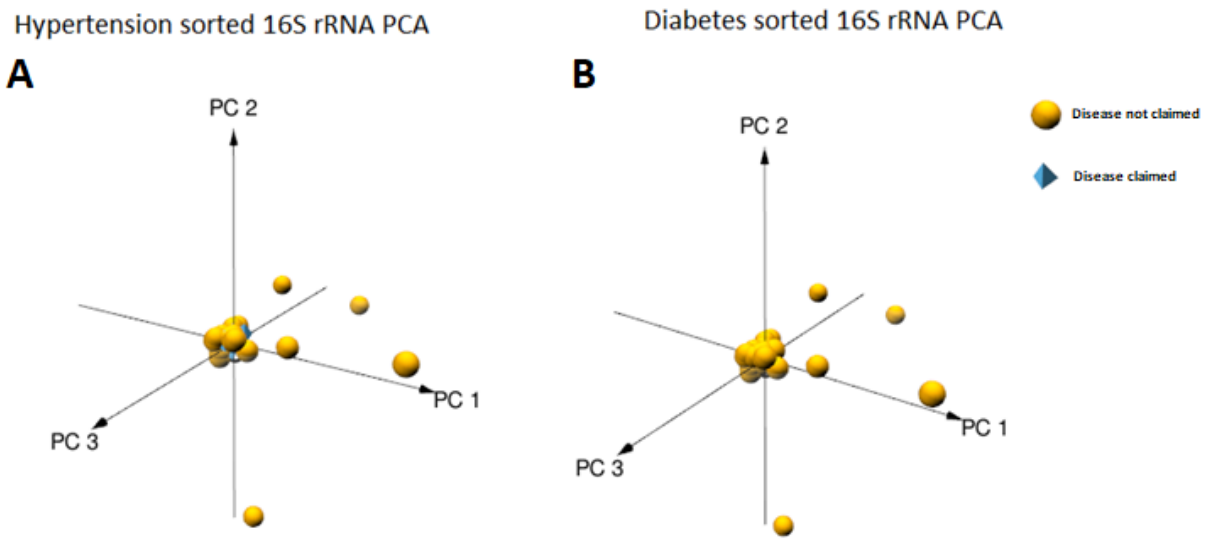

Figure 17. 3D Principal Component Analysis plotted based on bacterial components of human microbiome; individual objects have been labeled according to whether specific chronic illness was diagnosed. Figure was created using *pca3d* package in R. Panel (A) shows bacterial composition data with labeled patients with claimed hypertension. Panel (B) shows bacterial composition data with labeled patients with claimed diabetes.

10.11. Concomitance of SNPs and specific groups of bacteriophages

Cramer's V test was used to asses correlation between occurrence of particular SNPs in a patient genome and particular bacteriophages (or their groups) in phageomes, since these represent nominal variables. Quantitative data was transformed into binary format by the use of threshold = 10, that is presence of an OTU was considered positive in a sample when more than 10 reads were annotated to this taxonomy id. This type of calculation is based on Pearson's chi-squared statistic. The analysis revealed concomitance of some investigated allels and the presence of bacteriophage genera, listed in the Table 15.

Table 15. Concomitance of SNPs variants and composition of bacteriophages' part of microbiome. Estimation was obtained using *collections* library in Python 3.6.

Chromosome coordinates (rsID)	Type of Variant	Gene	Description - The samples in which the listed variant was detected were associated with the identification of specific groups of bacteriophages.
Regulation of immune response genes			
chr2:112837577-112837585 (rs3917345)	delCCAA (Insertion and Deletion)-	*Il-1 beta*	In 37 cases the Salmonella prophages were detected while in 5 they were absent. A similar association occurred with other groups where the Staphylococcus prophages were identified in 36 samples and absent in 6, Lambdavirus present in 30 and absent in 12, Kayvirus present in 28 and absent in 14, Punavirus present in 28 cases and absent in 14, Teseptimavirus present in 25 and absent in 17.
chr2:112830955 (rs2464906)	**A>G**	*Il-1 beta*	Kayvirus was identified in 27 samples while not identified in 12 with this variant. Lambdavirus was identified in 28 samples while not identified in 11 with this variant. Teseptimavirus was identified in 23 samples while not identified in 16 with this variant.
Chr2:112834078 (rs1143630)	**T>G**	*Il-1 beta*	Lambdavirus was identified in 27 samples while not identified in 11 with this variant. Teseptimavirus

			was identified in 26 samples while not identified in 12 with this variant.
chr2:112830976 (rs1143642)	**A>G**	*Il-1 beta*	Lambdavirus was identified in 27 samples while not identified in 11 with this variant. Kayvirus was identified in 26 samples while not identified in 12 with specified variant. Punavirus was found in 26 and not identified in 12 samples with this variant
chr2:112835941 (rs1143629)	**G>A**	*Il-1 beta*	Lambdavirus was identified in 25 samples while not identified in 9 with this variant. Punavirus was identified in 25 samples while not identified in 9 with this variant. Kayvirus was identified in 23 samples while not identified in 11 with this variant.
chr2:112829544 (rs2853550)	**A>G**	*Il-1 beta*	Staphylococcus prophages were identified in 35 samples and absent in 5, Staphylococcus prophages present in 34 and absent in 6, Lambdavirus present in 28 and absent in 12, Punavirus present in 27 cases and not identified in 13, Kayvirus present in 27 and not identified in 17.
chr1:67189524 (no rsID)	**A>C**	*IL23R*	Lambdavirus was identified in 26 samples while not identified in 11 with this variant. Punavirus was

			identified in 25 samples while not identified in 12 with this variant. Kayvirus was identified in 24 samples while not identified in 13 with this variant.
chr12:68249814 (rs976748)	**G>A**	*IL22*	Kayvirus was identified in 24 samples while not identified in 9 with this variant. Lambdavirus was identified in 25 samples while not identified in 10 with this variant. Punavirus was identified in 24 samples while not identified in 11 with this variant.
rchr7:22728045 (rs2069833)	**C>T**	*IL- 6*	Kayvirus was identified in 24 samples while not identified in 9 with this variant. Lambdavirus was identified in 24 samples while not identified in 9 with this variant. Punavirus was identified in 22 samples while not identified in 11 with this variant.
chr4:25229876 (no rsID)	**T>C**	*PI4K2B*	Lambdavirus was identified in 28 samples while not identified in 11 with this variant. Kayvirus was identified in 27 samples while not identified in 12 with this variant. Punavirus was identified in 27 samples while not identified in 10 with this variant. Teseptimavirus was identified in 23 samples while not identified in 16 with this variant.

chr1:11794400 (rs4846051)	**G>A**	*MTHFR*	Lambdavirus was identified in 26 samples while not identified in 9 with this variant. Punavirus was identified in 25 samples while not identified in 10 with this variant. Kayvirus was found in 24 and not identified in 11 samples with this variant. Teseptimavirus was found in 21 and not identified in 14 samples with this variant
chr1:153341436 (rs3014859)	A>G	*PGLYRP 4*	Lambdavirus was identified in 24 samples while not identified in 11 with this variant. Kayvirus was identified in 23 samples while not identified in 12 with this variant. Punavirus was identified in 23 samples while not identified in 12 with this variant. Teseptimavirus was identified in 14 samples while not identified in 21 with this variant.
chr1:153341677(no rsID)	**C>A**	*PGLYRP 4*	Lambdavirus was identified in 25 samples while not identified in 11 with this variant. Kayvirus was identified in 24 samples while not identified in 12 with this variant. Punavirus was identified in 24 samples while not identified in 12 with this variant.
chr19:41355432 - rs11466313	**dupCTC / insCTCATGTC**	*TGFB1*	Lambdavirus was identified in 24 samples while not identified in 11 with this variant. Punavirus was

	CCTG $(CCCT)_2$CCTC		identified in 24 samples while not identified in 11 with this variant. Kayvirus was identified in 23 samples while not identified in 12 with this variant.
Cell metabolism			
chr20:56386407 (rs1047972)	**T>C**	*AURKA*	Lambdavirus was identified in 24 samples while not identified in 11 with this variant. Kayvirus was identified in 23 samples while not identified in 12 with this variant. Punavirus was identified in 23 samples while not identified in 12 with this variant. Teseptimavirus was identified in 14 samples while not identified in 21 with this variant.
chr2:160357664 (rs6718526)	**T>C**	*RBMS1*	Lambdavirus was identified in 26 samples while not identified in 10 with this variant. Kayvirus was identified in 25 samples while not identified in 11 with this variant. Punavirus was identified in 25 samples while not identified in 11 with this variant

Most common association was related to IL1 -beta gene. Six different SNPs in this gene could be identified as associated to presence of bacteriophages that belong to Kayvirus, Punavirus, Lambdavirus, Teseptimavirus. Two SNPs in the range of the gene PGLYRP4 were identified as associated to presence of bacteriophage families like Lambdavirus, Kayvirus and Punavirus. Single variants were identified regarding other genes: Ili 22, Il-6, MTHFR, Pl42KB, AURKA

and RBMS1 and their coexistence with Kayvirus, Punavirus, Lambdavirus, and Teseptimavirus bacteriophages.

10.12. Gastritis – related bacteria association to Single Nucleotide Polymorphisms

Correlation between identified SNPs and pathogens related to stomach diseases was checked. According to literature, following pathogens were chosen for the analysis: *Helicobacter pylori* (Watari et al., 2014) *Prevotella jejuni* (Könönen & Gursoy, 2021a) and *Salmonella enterica* (Zha et al., 2019). Cramer's V test was used to compare categorical datasets for their potential correlation. This analysis revealed that specific alleles in 11 genes were related to *H. pylori* identification in bioptates from the stomach. Variants in IL22, TSPAN8, CCDC33, CACNA1A, IL23R, IL-1B, CDKAL1, LTA, IL6, NOD1, TLR4 genes were identified as those with increased prevalence in individuals with *H. pylori* detected in stomach bioptates ($p<0.01$). Investigated correlations are presented in the Table 16.

Table 16. Specific alleles related to increased frequency of *Helicobacter pylori* presence in stomach. Table created using *counter* Python 3.6 library on Ampliseq sequencing data and bacterial microbiome composition features. Cramer's V test was used to confirm that presence of *H. pylori* is correlated with above variants with $p<0.01$.

Position	Allele	Gene
chr12:68251064	A>G	*IL22*
chr12:68255353	T>C	*IL22*
chr12:71183321	G>C	*TSPAN8*
chr15:74312311	C>A	*CCDC33*
chr19:13529284	TAATACTAATACAATAC	*CACNA1A*
chr19:13529487	T>C	*CACNA1A*
chr1:67165579	C>T	*IL23R*
chr1:67222666	T>C	*IL23R*
chr1:67238627	A>T	*IL23R*
chr1:67240275	G>A	*IL23R*

chr2:112832890	C>T	*IL-1B*
chr2:112837574	AATACCAAA	*IL-1B*
chr2:112837576	TACCAACCAATACCAA	*IL-1B*
chr6:20657114	T>C	*CDKAL1*
chr6:31575981	G>A	*LTA*
chr6:31579012	T>C	*LTA*
chr7:22726866	A>C	*IL6*
chr7:30446094	A>C	*NOD1*
chr9:117702840	T>C	*TLR4*

Table 17. Alleles related to increased frequency of *Rothia mucilaginosa*. Table created using *counter* Python 3.6 library on Ampliseq sequencing data and bacterial microbiome composition features. Cramer's V test was used to confirm that presence of *R. mucilaginosa* is correlated with above variants with $p<0.005$.

position	Variant	Gene
chr12:68249918	T>C	*IL22*
chr12:68255258	C>T	*IL22*
chr19:41363851	C>T	*TMEM91*
chr19:498524	A>G	*MADCAM1*
chr19:855574	CGGCG	*ELANE*
chr1:67209833	C>T	*IL23R*
chr3:10290784	C>T	*GHRL*
chr3:159991864	T>C	*IL12A*
chr4:38783012	G>T	*TLR10*
chr6:20716864	A>G	*CDKAL1*
chr7:30454406	T>A	*NOD1*
chr8:117172544	C>T	*SLC30A8*

The analysis revealed that specific alleles in 11 genes were related to increased *Rothia mucilaginosa* detection in bioptates from human stomach. Alleles in IL22, TMEM91, MADCAM1, ELANE, IL23R, GHRL, IL12A, TLR10, CDKAL1, NOD1, SLC30A8 genes were identified as those with increased prevalence in individuals with *Rothia mucilaginosa* detected in stomach bioptates ($p<0.005$). Data is presented in the Table 17.

Table 18. Alleles related to increased frequency of *Prevotella melanonigica.* Table created using *counter* Python 3.6 library on Ampliseq sequencing data and bacterial microbiome composition features. Cramer's V test was used to confirm that presence of *P. melanogenica* is correlated with above variants with $p<0.01$.

position	Variant	Gene
chr19:41332301	G>A	*TGFB1*
chr19:41337556	T>C	*TGFB1*
chr1:186681619	T>C	*PTGS2*
chr1:67204530	G>A	*IL23R*
chr1:67215986	T>G	*IL23R*
chr3:159992214	A>T	*IL12A*
chr5:148826812	C>T	*ADRB2*
chr5:148826910	G>C	*ADRB2*
chr6:20660912	T>A	*CDKAL1*
chr7:128723233	G>A	*FAM71F1*
chr7:30446125	A>G	*NOD1*
chr7:30454406	T>A	*NOD1*

Cramer's V analysis revealed that specific alleles in 8 genes were related to increased *Rothia mucilaginosa* detection in biopsies from human stomach. Alleles in *TGFB1, PTGS2, IL23R, IL12A, ADRB2, CDKAL1, FAM71F1, NOD1* genes were identified as those with increased prevalence in individuals with *Prevotella melaninogenica* detected in stomach's bacterial microbiome ($p<0.01$). Data is presented in the Table 18.

Table 19. Alleles related to increased frequency of *Neisseria subflava.* Table created using *counter* Python 3.6 library on Ampliseq sequencing data and bacterial microbiome composition features. Cramer's V test was used to confirm that presence of *N. subflava,* is correlated with above variants with $p<0.01$.

position	Variant	Gene
chr10:112996282	A>T	*TCF7L2*
chr10:112998590	C>T	*TCF7L2*
chr16:10948382	G>A	*CLEC16A*
chr19:41354391	A>G	*TGFB1*
chr1:11794400	G>A	*MTHFR*
chr1:155192276	C>T	*MUC1*
chr1:67189464	G>A	*IL23R*
chr1:67204530	G>A	*IL23R*
chr1:67222666	T>C	*IL23R*
chr1:67240272	TCCTC	*IL23R*
chr3:10290784	C>T	*GHRL*
chr4:102267552	C>T	*SLC39A8*
chr4:99583507	T>C	*MTTP*
chr6:20660790	G>A	*CDKAL1*

Cramer's V analysis revealed that specific alleles in 10 genes were related to increased *Neisseria subflava* detection in bioptates from human stomach. Alleles in *TCF7L2, CLEC16A, TGFB1, MTHFR, MUC1, IL23R, GHRL, SCL39A8, MTTP, CDKAL* genes were identified as those with increased prevalence in individuals with *Neisseria subflava.* detected in stomach bioptates ($p<0.01$). As many as 4 variants were related only to IL23R gene. Data is presented in the Table 19.

Table 20. Alleles related to increased frequency of *Prevotella jejuni*. Table created using *counter* Python 3.6 library on Ampliseq sequencing data and bacterial microbiome composition features. Cramer's V test was used to confirm that presence of *P. jejuni* correlated with above variants with $p<0.01$.

position	Variant	Gene
chr10:43573110	A>C	*ZNF239*
chr12:68252741	T>C	*IL22*
chr12:68253933	G>A	*IL22*
chr16:53767042	T>C	*FTO*
chr16:53776774	T>C	*FTO*
chr16:53786591	G>A	*FTO*
chr16:53786615	T>A	*FTO*
chr1:67219760	G>A	*IL23R*
chr1:67222666	T>C	*IL23R*
chr6:20657114	T>C	*CDKAL1*

The analysis revealed specific alleles in 5 genes as related to increased *Prevotella jejuni* detection in bioptates from human stomach. Variants in ZNF239, IL22, FTO, IL23R, CDKAL1 genes were identified as those with increased prevalence in individuals with *P. jejuni* detected in stomach bioptates ($p<0.01$). Four among these variants are located within FTO gene. Data is presented in the Table 20. No significant variants were identified as related to *Salmonella enterica* presence among patients' bacterial microbiome.

10.13. Single Nucleotide Polymorphisms, bacteriophages, and their hosts tripartite association

Correlation has been tested between bacteriophage presence and host's variants detected in the genome. Cramer's V test was used and bacteriophage' families taxonomic level was investigated. Increased presence of two phage families were associated with variants in the

confidence interval of 0.95 and $p<0.05$: Kayvirus and Punavirus. Genes in which variants have been detected are involved in cytokine release (Il-1 beta, IL-22, Il-6, cell signaling (TGFP1, PI4K2B) and amino acids processing (AURKA). Results of the analysis of associations between bacteriophages and host's genome variants are presented in the Table 21.

Table 21. Variants related to increased presence of bacteriophages in patients' phageomes; significantly more frequently represented phage families and their hosts are listed. Correlations were derived using Cramer's V test and ViralHost Database.

Position	Gene	rsID	Allele	Phage and its hosts
chr1:11794400	*MTHFR*	rs4846051	G>A	**Kayvirus**; Staphylococcus simulans, cohti, codimenti, lugdunensis **Punavirus**; Salmonella enterica, Aeromonas veronii
chr12:68249814	*IL22*	rs976748	G>A	**Kayvirus**; Staphylococcus simulans, cohti, codimenti, nepalensis
chr19:41355432	*TGFB1*	rs11466313	dupCTC / insCTCATGTCCCTG	**Kayvirus**; Staphylococcus simulans, cohti, codimenti, lugdunensis Punavirus; Salmonella enterica, Aeromonas salmonicida
chr2:112835941	*Il-1 beta*	rs1143629	G>A	**Kayvirus**, Staphylococcus aureus, epidermidis, pasteuri, pettenkoferi
chr2:112837577-112837585	*Il-1 beta*	rs3917345	delCCAA	**Kayvirus**, Staphylococcus haemolyticus
chr20:56386407	*AURKA*	rs1047972	T>C	**Kayvirus**, Staphylococcus haemolyticus

chr4:25229876	*PI4K2B*	No rsID	T>C	**Kayvirus**, Staphylococcus captis
chr7:22725717	*IL6-AS1*	rs35081782	dupCT	**Kayvirus**, Staphylococcus captis

11. Discussion

In this study correlation between bacterial species and bacteriophages with ICD classification diseases was observed. SNPs present in genes involved in innate immune response are associated with changes in the microbiome diversity. Presence of specific bacteriophage or bacterial taxons associates to bacterial diversity in microbiomes. Also, gastritis-related bacteria presence has its links to some human SNPs. Thus, tripartite interdependencies between human host, bacterial, and bacteriophage components of human microbiomes can be observed, as well as their links to human health status.

11.1. Bacterial taxons in ICD-classified diseases

According to Table 11. some bacterial species have correlation with diseases diagnosed in the investigated patients. Environmental bacteria *Variovorax* in this study was found associated to 3 health disorders according to ICD classification, these disorders were: K29 – "Gastritis and duodenitis", R.10.4 – "Other and unspecified abdominal pain", K21 – "Gastro-oesophageal reflux disease". *Variovorax* is a Gram-negative motile bacterium. It was isolated from many extremophile environments, like Antarctic or hypersaline lakes (Shrestha et al., 2022). This bacterium could be observed in soil and freshwater, and its ability to adapt to challenging conditions may be the reason why it was found so frequently in stomach microbiome(Hwang et al., 2022). Some reports indicated other bacteria (from *Variovorax sp*.) that may be involved in symptomatic gastritis development, however Han *et al.* (H. S. Han et al., 2019) reported that *Variovorax paradoxus* was significantly correlated with histological gastritis, which is in line with the data presented herein (Table 11). Unclassified *Variovorax* was confirmed as one of the *H. pylori* infections-associated pathogens (D. Wang et al., 2022). Study by (Ndegwa et al.,

2020) showed that *Variovorax* sp. was significantly enriched in samples, where *H. pylori* infection was diagnosed. Nevertheless, here this association was not observed (Table 11).

Escherichia albertii was found also linked to K29 – "Gastritis and duodenitis", R.10.4 – "Other and unspecified abdominal pain", K21 – "Gastro-oesophageal reflux disease" ICD disease states (Table 11). Because of the possession of *eae* gene, some of *E. albertii* strains are often falsely identified as enteropathogenic or enterohemorrhagic *E. coli*. Pathogenicity of *E. albertii* is related to its ability to attach to intestinal epithelial cell surfaces and to produce Shiga toxin, a causative agent of the bloody diarrhea in infected people. (Ooka et al., 2012). Study by Palmas, et al (2021) shows that this bacterium is significantly correlated with microbiome composition which may be found more frequently in people who suffer from obesity (Palmas et al., 2021). Gallardo, et al (2017) demonstrated that *E. albertii* was confirmed as indicative among other species for diarrheagenic *Escherichia coli* group compared to the virus infected and healthy groups (Gallardo et al., 2017). Results presented herein show that *E. albertii* presence is correlated with the occurrence of gastrointestinal diseases (Table 11).

Oral bacterium as *Fusobacterium nucleatum* was associated with Gastritis and dudenitis diseases classified in ICD as "K29" (Table 11). This bacterium is one the most intensively studied because of its correlation with many cancer cases. Its increased incidence was shown in esophageal cancer, pancreatic cancer, hepatocellular carcinoma, and gastric cancer (Y. Liu et al., 2019a, 2019b). Study performed by Surlin et al, (2020) show *F. nucleatum* as an aggravating factor for gastric cancer (Șurlin et al., 2020). Results presented herein show that *F. nucleatum* is related to gastritis and duodenitis (Table 11) which is one of the factors that increase probability of stomach cancer development (Waldum & Fossmark, 2021).

Klebsiella pneumoniae is one of the most widespread pathogens in the world. It leads to serious oral-gastrointestinal disorders. Presence of oral *Klebsiella* spp. is indicatory for Inflammatory Bowel Disease. In mice, oral application of this bacterium leads to liver abscess formation (Atarashi et al., 2017; Chen et al., 2014; Sung et al., 2018). Data retrieved from Human Microbiome Project shows that *K. pneumoniae* is present among 3.9% of all human gastrointestinal samples (Gorrie et al., 2017). In the research conducted for this manuscript, this number is significantly higher, because *K. pneumoniae* was found in 62,9% patients where at least one ICD-classified disease state was diagnosed. In contrast, in patients without

diagnosed diseases, it was 40%. Other studies show that *Klebsiella* prevalence may vary according to geographical region and hospital or non-hospital environments (Chang et al., 2021; Ssekatawa et al., 2021). Research presented herein shows that *K. pneumoaniae* is significantly associated with disease state causing serious implication in the gastric part of the gastrointestinal tract. This is consistent with studies by Chang et al (2021) and Ssekatawa et al (2021), where an increased prevalence of this pathogen was found among patients with gastrointestinal diseases

The location of the stomach in the gastrointestinal tract means that the microbiota found inside is often transient and derived from the environment. Bacteria get there together with food. Many conditions may change abilities of the stomach that protect from infections (for example, a low pH changed to a more alkaline one decreases the level of protection). This provides a good opportunity for the growth of microorganisms from the environment, which are able to survive there longer due to their adaptability (Beasley et al., 2015; Lopetuso et al., 2014; Noto & Peek, 2017). Increased gastric pH can be associated with *H. pylori* infection (Sung et al., 2018). *Klebsiella variicola, Raoultella ornithinolytica, Phreatobacter cathodiphilus, Hydrogenophaga sp. PBC, Haemophilus sp*. oral taxon 036, *Serratia marcescens* are opportunistic pathogens that may be found in soil, water, and unprocessed food. Under favorable conditions such as gastritis, they are able to survive in stomach (Diricks et al., 2022; Hajjar et al., 2020; Khanna et al., 2013; S. J. Kim et al., 2018; Rodríguez-Medina et al., 2019). All these bacterial taxons have been found in this work correlated to gastrointestinal diseases (Table 11). Gastritis can cause diversity changes in microbiome, and make it more available for environmental organisms. Yang et al, (2019) confirmed that for instance *H. pylori* infection was correlated with fecal microbiota switch (L. Yang et al., 2019).

11.2. Bacteriophages in ICD-classified disease states

Increased presence of Clostridioides prophages was identified among patients with “K29 - Gastritis and duodenitis” disease state (Table 12). It was calculated using z-test and showed significance with $p<0.005$. Clostridioides genus groups a few anaerobic species, able to form endospores species, including *Clostridium difficile.* This is a dangerous pathogen which in rare cases can even lead to death. Its pathogenicity is regulated by prophage activity that is involved

in toxin production (Fortier, 2018). Occurrence of gastritis and duodenitis in patients was correlated with IBD occurrence among patients (Sonnenberg et al., 2011). *C. difficile* is one of the major factors in pathogenesis of IBD (Nitzan et al., 2013). In the study presented herein none of the patients' intestines were simultaneously investigated, so data that could potentially confirm presence of *Clostridum*-related IBD and gastritis were not available.

Significantly higher prevalence of Tequatrovirus phages was identified among patients with "K63 - Other diseases of intestine". It was calculated using z-test and showed significance with $p<0.05$ (Table 12). Tequatrovirus is a genus of bacteriophages that are specific, among others, to a wide spectrum of pathogens present in food and causing contaminations, like pathogenic *E. coli* or *Shigella* (Pham-Khanh et al., 2019; Zhou et al., 2022). These bacteriophages could not maintain their activity in typical, low pH conditions (lower than 3 (Inbaraj et al., 2022)) identified in stomach, so their presence indicates that at the moment of sample collection stomach microbiome was disturbed. ICD code K63 that was found correlated to this phage group indicates that other diseases were ongoing in the same time and they potentially could cause dysbiosis in stomach. Presence of Tequatrovirus in stomach may also be determined by an overgrow of their host in other parts of human gastrointestinal tract. Overgrow of particular type of bacteria may be followed by overgrowth of bacteriophages specific to these bacteria (de Paepe et al., 2014).My results suggest that in some patients in the studied group, "K63 - Other diseases of intestine" disease state was caused by *Escherichia sp.* overgrowth, with a subsequent increase in Tequatroviruses detected in their phageomes (Table 12.). This observation correlates with increased incidence of "R10.4 Other and unspecified abdominal pain" diagnosis.

Increased presence of Inovirus bacteriophages was identified among patients with "K29.6 - Other Gastritis" disease state (Table 12). Inovirus is a group of filamentus bacteriophages specific to several Gram-negative pathogens that may determine serious human intestine diseases such as *Bacteroides fragilis* and *Fusobacterium nucleatum.* Studies performed by other groups indicate that bacteriophages belonging to Inoviridae are very widespread in both plant and human microbiomes (Handley & Devkota, 2019; Hannigan et al., 2018; Nakatsu et al., 2022). Inoviruses are not able to maintain activity during exposition to pH lower than 2 (Day, 2008), so their presence in samples collected in the study presented in this dissertation

can be associated with disturbances in the giant gastric folds, diminished acid secretion, or excessive mucus secretion (Jia et al., 2022; Sánchez-Alcoholado et al., 2020). Association with K29.6 – "Other gastritis" presented in the conducted research may suggest that in the other parts of the patients' intestine another infection was developing, resulting in identification of bacteriophage specific to those bacteria in the stomach.

11.3. Single Nucleotide Polymorphisms impacts microbiome diversity

Microbiome diversity is one of the parameters that allow us to take a holistic view on microorganisms' communities. Its variations are significant for pathogenesis in diseases not only in human gut but in all organs where microbiomes could be detected such as lungs, skin, etc (Hufnagl et al., 2020; Mirsepasi-Lauridsen et al., 2018; Swarte et al., 2020, 2022). The diversity of the microbiome as expressed by the Shannon index can be an indicator of the presence of a disease state (Joossens et al., 2011; Manor et al., 2020; Morgan et al., 2012; Org et al., 2017). In this study, I discovered that the microbiome diversity index was linked to specific genetic variants (patient genotypes) in immune-related genes.

In my study, I observed variants present in the human genome that were associated with changes in the diversity of the bacterial microbiome as expressed by the Shannon index. Variant in chr1:223111858 G>A in *TLR5* was previously described as potential polymorphism associated with systemic lupus erythematosus – which is an autoimmunological disease of unknown pathogenesis and potential association with *Burkholderia pseudomallei* infection (Y. H. Lee et al., 2016; West et al., 2013). *TLR5* gene has a significant role in recognition of potentially harmful pathogens in human organisms. It is used by the immune system to build mucosal barrier to prevent illnesses associate with inflammation in the intestines (J. Yang & Yan, 2017). Lee et al, (2016) showed that variant changes when observed in the middle of the *TLR5* gene do not have to bring negative effects for a patient. It suggests that variant chr1:223111858 G>A may have enhanced effect on *TLR5* receptor and increase its efficiency. Therefore, the higher diversity of bacterial microorganisms has been observed herein (Table 13), thus suggesting a lower risk of overgrow by pathogens.

Upstream gene variant was observed in the *IL-1β* gene (Table 13) and it was significantly related to decreased microbiome diversity. *IL-1β* is related to production of a crucial cytokine:

interleukin 1 (IL-1β), involved in immunological response to viruses, bacteria, and fungi. IL-1β induces production of interferons, IL-6 or TNF. Expression of *IL-1β*affects lymphocytes T and development of lymphocytes B (Lopez-Castejon & Brough, 2011). Variant in chr2:112834786 is located upstream the gene. In this dissertation, a significant decrease of diversity in stomach bacterial microbiome was observed in association with this variant. Similar fluctuations in microbiome diversity have already been reported in other mammals. For instance, in mice, Il-1 receptors deficiency resulted in microbiome diversity drop (Rogier et al., 2017). Variant may have impact on IL-1β expression, so host-response to pathogens may work with decreased efficiency. It creates an opportunity for some pathogens to overgrow which may result in decreased diversity.

Upstream gene variant was observed in the *TLR2* gene (Table 13) and it was significantly related to increased microbiome diversity. Toll-like receptor 2 coded by *TLR2* gene plays a crucial role in the innate immune response to a wide variety of bacteria and other microorganisms. Some studies showed in mouse model that this gene's knockout render animals more susceptible to sepsis development by *Staphylococcus aureus,* meningitis by *Streptococcus pneumoniae,* and to *Mycobacterium tuberculosis* infection. TLR2 receptor is activated by bacterial cell wall elements (Fujiwara et al., 2018; Texereau et al., 2005). Although the detected variant is synonymous, studies show that it can potentially affect messenger RNA splicing, stability, and structure as well as protein folding (Hunt et al., 2009). It suggests that variant in chr4:153704257 T>C by changing its gene expression, influences the environment of the gastric microbiome increasing its diversity, as observed in the study presented herein (Table 13).

Variant in *TLR10* gene is significantly related to decreased microbiome diversity. G>A variant in chr4:387830009 in the coding region of *TLR10* gene is an upstream gene variant. This gene expression occurs in B cells and plasmacytoid dendritic cells. TLR10 ligands and their origin are not determined. Modifications of microbiota in different cancers are results of activation of specific TLRs, but the exact mechanism is still unknown (le Noci et al., 2021; Yinhang et al., 2022). Literature shows that some polymorphisms increase susceptibility to various infections either bacterial or viral (Henrick et al., 2019; Y. Wang et al., 2018). Among patients

investigated in this study decrease of microbiome diversity can be observed in association with this variant (Table 13.) which indicates that variant can impair the gene function.

Variant T>A in the TLR1 gene was observed and significantly related to decreased microbiome diversity (Table 13). Missense variant in the chr4:38797918 in the middle of *TLR1* gene indicates a change of amino acid that is substituted. This change resulted in decreased diversity of bacterial microbiome composition in patients' stomach. *TLR1* gene works together with *TLR2* gene since their products create a heterodimer. It can recognize lipoproteins that are characteristic for bacterial cells (Burgueño & Abreu, 2020; Mcdermott & Huffnagle, 2014). Study by Kamdar et al, (2018) shows that sensing of the gut microbiota by TLR1 provides crucial signals for regulation of colonic epithelium functions and inflammation which results in prevention of pathogen attachment to the mucosa (Kamdar et al., 2018; Takeuchi et al., 2002). Missense mutation might deregulate function of the TLR1 gene, so it is not working efficiently. This promotes overgrowth of pathogenic bacteria and leads to a state of dysbiosis and at the same time reduced diversity, as observed herein (Table 13.).

In the *IL-6* gene, 7 upstream gene variants were observed (Table 13). All of them increased diversity of bacterial microbiome composition in patients' stomach. Interleukin 6 is involved in the pathogenesis of IBD. Its role is to differentiate T helper 17 cells and mediate destructive inflammatory response. Overexpression of that gene is observed during gram-negative bacteria overgrowth inside gut (Higuchi et al., 2018; Shahini & Shahini, 2022; S. Wu et al., 2022). In mice, no significant changes in microbiome diversity were observed after Il-6 knockout(S. Wu et al., 2022). Nevertheless, in this study I have observed that the specific variants identified in the investigated patients were associated to increased diversity in patients' microbiomes. For instance, by altering the structure of the expressed RNA, these variants may affect the efficiency of IL-6 expression, as well as the formation of structures and interaction with other fragments of the genome.

11.4. Brussowirus and Triavirus correlate to decreased microbiome diversity

Results presented in the Figure 8. reveal two bacteriophages in stomach microbiomes that were associated with decreased bacterial diversity expressed in Shannon diversity index. Brussowvirus is a dsDNA virus with genome of 40 kb. It has a non-contractile tail and an icosahedral head. This phage is specific to *Streptococcus spp.* including *S. agalactiae*, and its presence was confirmed in environmental biomes (Abril et al., 2020; Hanemaaijer et al., 2021). Triavirus is a dsDNA bacteriophage specific to *Staphylococcus spp.* including *S. aureus* MRSA. This virus contains a prolate head and a non-contractile tail (Feng et al., 2021; Klimka et al., 2021). Presence of these two bacteriophages can be potentially linked to presence of their hosts that may potentially be pathogenic and cause inflammatory states or dysbiosis in the stomach or intestines. They presence can indicate overgrowth of some bacteria and decreasing the diversity among bacterial composition of microbiome.

11.5. Bacterial components influence microbial diversity

Six bacterial species were selected basing on literature data and investigated whether they are linked to changes in microbiome diversity. *Helicobacter pylori* is the most investigated pathogen related to gastric cancer. Thanks to urease secretion and pathogenicity islands, this bacterium can survive in the challenging environment of gastric juice and peristaltic movements (Alexander et al., 2021; Kinoshita-Daitoku et al., 2021). Nevertheless, results of this work do not confirm a significant effect of the presence of *H. pylori* on the diversity of the bacterial microbiome. T-test applied to the data analysed herein revealed that differences in infected and infection-free individuals were insignificant (Table 14.). It is highly probable that this was due to the fact that *H. pylori* can exist in coccoid forms - remaining enzymatically inactive. 16S sequencing is more sensitive in detecting this form of *H. pylori* infection than other commonly used methods for *H.pylori* detection (Gantuya et al., 2021; Szymczak et al., 2020), so possibly forms that did not affect gut conditions prevailed in this study.

Rothia mucilaginosa was confirmed as associated to gastritis in studies on oral microbiota. Due to the structure of the digestive tract, a part of the microbiome from the oral cavity can also be temporarily observed in the stomach. *R. mucilaginosa* is a facultative anaerobe and opportunistic bacterium that may cause, among others, bacteremia, endocarditis, and pneumonia. It typically affects immunocompromised patients (F. teng Li et al., 2012; Z. M. Sun et al., 2013). By its presence it significantly impacts diversity of bacterial microbiome increasing it. Recently, Rigauts et al, (2022) reported that *R. mucilaginosa* expressed anti-inflammatory activity in lungs (Rigauts et al., 2022). This suggests that the documented negative correlation of this bacterium presence with pro-inflammatory markers may prevent dysbiosis and overgrowth which increase alpha-diversity, as observed also herein (Table 14).

Prevotella melaninogenica as an anaerobic commensal microorganism, one of the first species that inhabits human oral cavity immediately after birth. Its presence was increased in tumor tissues in stomach where *H. pylori* was eradicated (Könönen & Gursoy, 2021b; S. Wang et al., 2022; Waskito et al., 2022). *P melninogenica* significantly decrease diversity in stomach microbiome. Study shows that it may be responsible for triggering expression of antimicrobial peptides, cytokines(P. Xu et al., 2022). Increased inflammatory responses may lead to dysbiosis and decreased microbiome diversity index (J. Wang et al., 2020) which was also observed in data from this work, as presented in the Table 14.

Neisseria subflava is one of pathogens that strongly correlates with *H. pylori* presence. It can induce interleukin 8 production which cause the reduction of gastric acid secretion (Cui et al., 2022; Duan et al., 2022; Miyata et al., 2019). *N. subflava* has also been reported as a pathogen causing meningitidis and endocarditis (Baraldès et al., 2000). However, studies conducted on this patient group did not yield enough data to statistically evaluate its relation to microbiome diversity (Table 14).

The presence of *Prevotella jejuni* was confirmed in the group of patients investigated in this study. *P. jejuni* is a pathogen that was related to symptomatic gastritis. Its anaerobic properties make it capable of surviving in the stomach environment (Könönen & Gursoy, 2021a). However, my observations did not confirm that the presence of *P. jejuni* does not significantly correlated to any changes in the diversity of the microbiome (Table 14).

Salmonella enterica is one of the most spread foodborne pathogens that may have serious health implications. In this study *S. enterica* was included to evaluate whether gut pathogens may cause complications in upper sections of the gastrointestinal tract (Knodler & Elfenbein, 2019; Zha et al., 2019). *S. enterica* presence significantly decreases bacterial microbiome diversity which was presented in the Table 14. Previous studies report that its presence is strongly associated with inflammatory states and lowered bacterial Shannon index. Nevertheless, there is no literature data that suggest such effect in the stomach. As presented in the Table 14, *S. enterica* may correlate to decreased diversity in microbiome and thus potentially cause problems in in the upper sections of the gastrointestinal tract, particularly in stomach.

11.6. Principal component analysis - biplots

In two dimensional biplot (vectors angles analysis) in the Figure8. one can observe that probiotic bacterial genera negatively correlate with potential pathogenic groups like *Acinetobacter, Moraxella, Flavobacterium,* or *Corynebacterium.* Samples "E35" and "E180" are clearly outstanding from other samples. Sample "E35" was collected from a 44 year old woman suffering from lower abdominal pains. Histological results revealed focal inflammation and erosion in the stomach. I have found that her bacterial microbiome was completely dominated by Proteobacteria and Enterobacterieaceae that are normal components only for fecal microbiome (Litvak et al., 2017; Shin et al., 2015). Here one can observe a dysbiosis developed by overgrowth of Enterobacteriaceae family. Typically, human stomach microbiome should be dominated by *Firmicutes, Bacteroidetes,* and *Actinobacteria* (Mukhopadhya et al., 2012; Nardone & Compare, 2015a). Microbiome profiling for "E35" patient is visualized in the Figure 18.

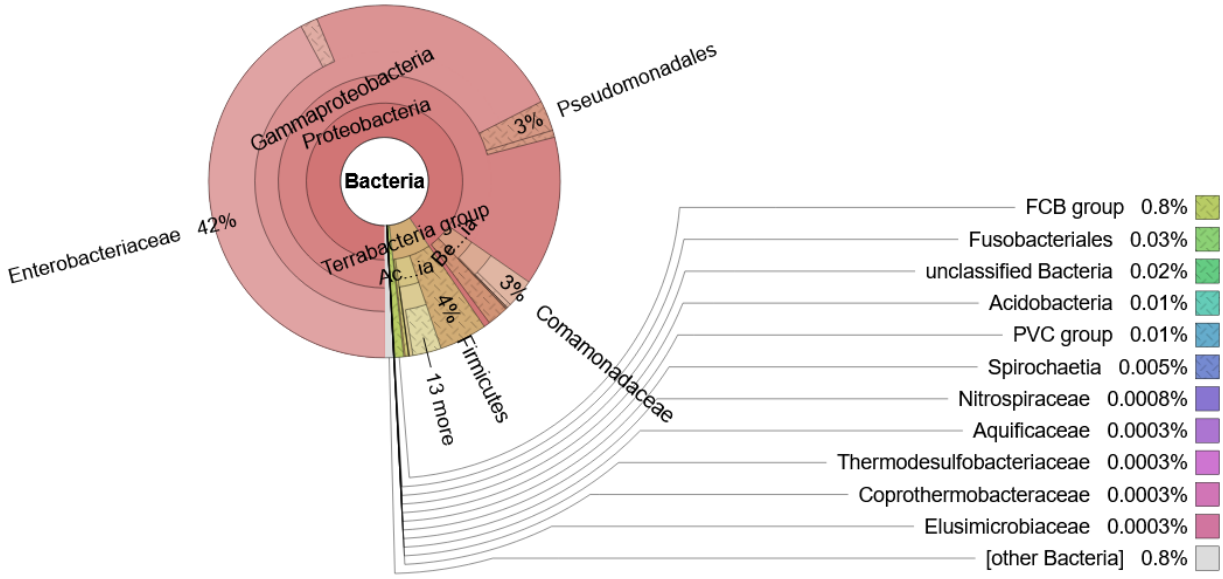

Figure 18. Bacterial microbiome profiling based on 16S rRNA genes sequencing in sample "E35". Figure created using Krona Tools.

Sample "E180" was collected from a 43-year-old woman after a successful *H. pylori* eradication. She was however diagnosed with "K29" – Acute hemorrhagic gastritis. Based on taxonomic analysis of her stomach microbiome presented in the Figure 18, it is possible to confirm the presence of bacterial groups that significantly affected outlying location of this sample on the biplot chart, like Moraxellaceae (*Acinetobacter*), Burkholderaceaceace (*Ralsonia*), Clostridiales (*Fastidiosipila*).

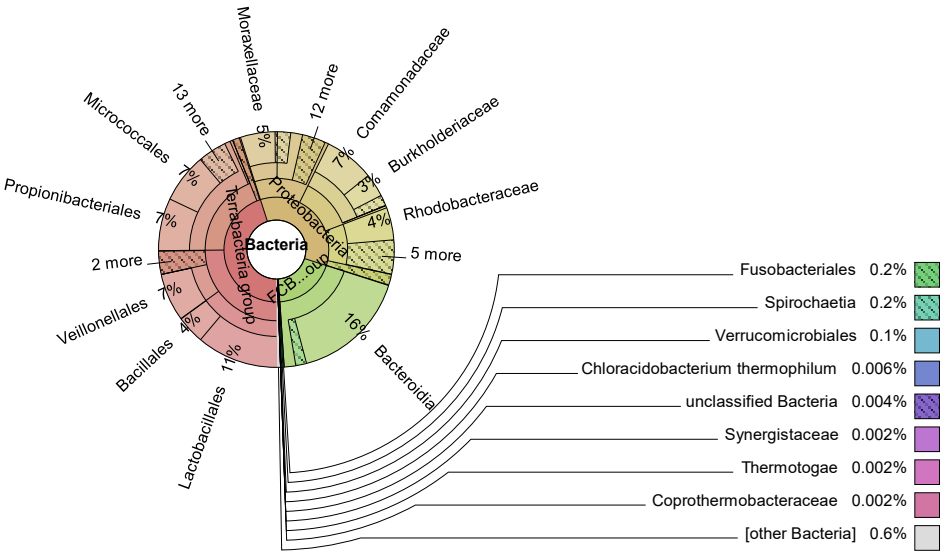

Figure 19. Bacterial microbiome profiling based on 16S rRNA genes sequencing in sample "E180". Figure created using Krona Tools.

Biplot based on phageome data presented in the Figure9. revealed that Triavirus had a significant contribution of the variables to the principal component analysis. In this study, Triavirus was also observed as correlated with significant decrease of bacterial microbiome diversity. Further analysis demonstrated that its presence was completely opposite contributing to PCs than Brussowvirus (Figure 9.) that was also associated with decreased Shannon index. Sample "E157" was collected from a 58-year-old woman with asthma and diabetes, and after gallbladder excision. Her gastric disorder was classified as "K29.9-Gastroduodenitis". Her phageome was completely dominated by Staphyloccoccus prophages, as demonstrated in the Figure 19 – 91.51% of identified sequences was derived from Staphylococcus prophages. Such dominance of one factor resulted in an outlier position of E157 on the biplot (Figure 9). This strongly suggests that this patient suffered from a staphylococcus-linked dysbiosis and the dominant phage type played a role of a biomarker.

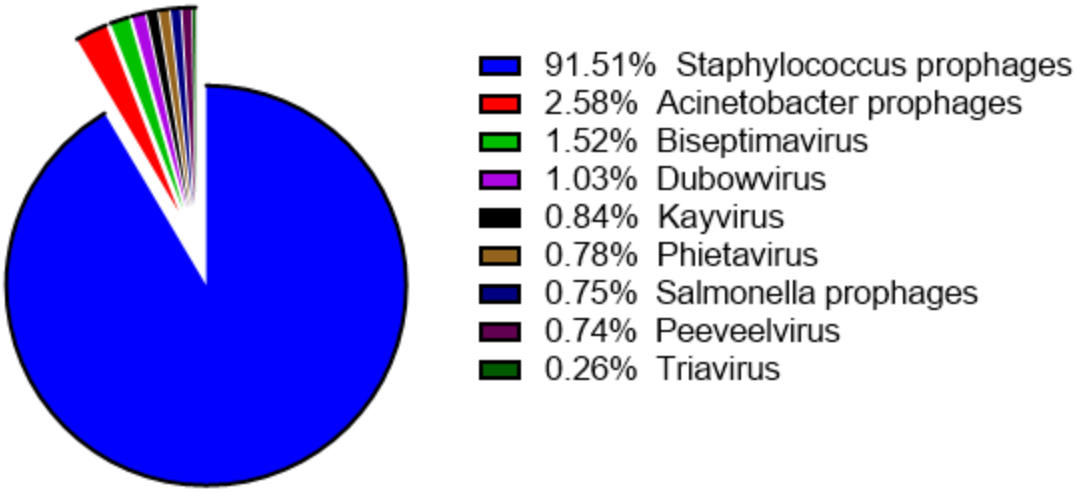

Figure 20. Phageome profiling based on shotgun sequencing in sample “E157”. Figure created using Krona Tools.

Sample “E95” was collected from a 49 years old woman with no chronic illnesses declared neither this patient declared any medications taken permanently. The patient was diagnosed with “K29 - Gastritis and duodenitis”. Her phageome was dominated by Salmonella prophages and Tequatroviruses (Figure 20). According to the biplot (Figure 9), both these groups had significant contribution to the principal component. The overrepresentation of one taxon may be the cause of the apparent position on the biplot chart, which can be described as an outlier. Phageome profile of the patient is presented in the Figure 21. This composition suggests that the occurrence of gastric complaints classified as K29 may be caused by dysbiosis through the proliferation of bacteria of the genus *Salmonella* and *Escherichia* where phages specific to them accounted for 54.88% and 19.67% of the total phageome, respectively.

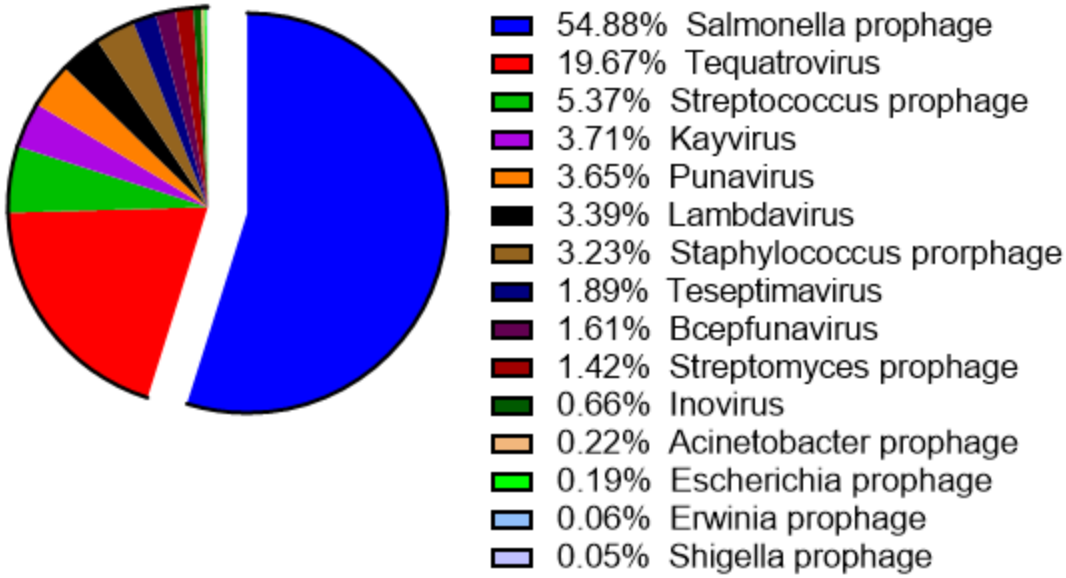

Figure 21. Phageome microbiome profiling based on shotgun sequencing in sample "E95". Figure created using Krona Tools.

11.7. Lack of clear clusters in bacterial and phageome composition by diseases states and Single Nucleotide Polymorphisms

Principal Component Analysis on microbiome composition data allows to distinguish clusters of factors that may significantly impact the microbiome shape at different taxonomy levels. Exemplary plots from the data in this study are presented in the Figure 13. Clustering helps to detect associations that may be difficult to find when analyzing only similarities between sequences (Sankaran & Holmes, 2019; P. Zhang et al., 2019). PCA dimension reduction automatically leads to K-means clustering using an objective function (Ding & He, 2004). Data from the literature suggest that the deeper is the sequencing, the more subtle differences between clusters can be detected (Ansari et al., 2015; Konishi et al., 2019; Lu, Ren, et al., 2016; Ramírez et al., 2016). The differences between the samples did not carry over to entire groups of bacteria either. They concerned the overrepresentation of single groups. Small differences in the composition of the microbiome can however affect an organism and not necessarily be apparent when analyzing large amounts of data (Björklund & Björklund, 2019; Jollife & Cadima, 2016).

In this study, no significant clustering results were observed in PCA plotting between investigated SNPs and microbiome composition. Exemplary plots are presented in the Figure 14. The differences relating to the x and y axes between the different samples in the graphs are not enough distinct to be statistically significant. Other studies revealed that variants in FUT2 and LCT genes could be responsible for increased colonization with probiotic strains such as *Bifidobacterium longum.* Genome Wide Associated Studies also suggested that alleles could only affect gene expression in individual organs and, consequently, affect the microbiomes locally, that is, only the microbiome present in these organs (Awany et al., 2019; Kolde et al., 2018b; Rothschild et al., 2018). Of note, typical cohort studies for genotyping include hundreds or thousands of samples (Awany et al., 2019; Rothschild et al., 2018), while this study was limited to 44 complete samples allowing for analysis of 16S rRNA – SNPs interactions. This may result in problems with identifying of correlations. On the other hand, populations may differ in their characteristics. Here a Polish population was investigated and it is not clear how far results from other populations can be extrapolated to this one.

3D PCA scores were previously used in microbiome research to compare composition of microbiota between samples treated with different factors (Shang et al., 2016; L. Yang et al., 2020). In this study, 3D PCA analysis is presented in the Figure 15 and 16. During the interview prior to sample collecting patients informed on their chronic diseases/comorbidities. Samples were clustered by the microbiome composition separately based on phageome and bacterial data. Figure 15. demonstrates that samples from patients with hypertension stand out further on the chart because of their different bacteriophage composition. Bacterial composition does not allow for the same conclusions. Samples from patients with hypertension did not stand out from the main cluster of bacterial patients' microbiomes. Previous studies have shown that some viruses can be hypertension biomarkers. Han et al, (2018) even reported correlated systems between viruses and bacteria in human gut virome. They found ubiquitous viral-bacterial associations in healthy individuals, from pre-hypertensives to hypertensive patients (M. Han et al., 2018). Another investigation by de Jonge et al, (2022) showed bacteriophages closely related to metabolic syndrome. *Candidatus Heliusviridae,* which is widely spread gut phage lineage, was determined as a biomarker to their study group (de Jonge et al., 2022). Previously described results showing associations between the presence of bacteriophages and the ailments with which metabolic syndrome is associated suggest that stomach phageome may

be related to hypertension occurrence among patients. Nevertheless, study on a bigger group is needed to evaluate this finding.

11.8. Gene variants related to bacteriophages' hosts presence

Genetic variants in immune responses-related genes that are associated to specific bacteriophage presence in microbiomes are also known to be a part of defense against eukaryotic viral infections. Those identified in this study are presented in the Table 15. IL-1β seems to be crucial in inducing the inflammation state (K. S. Kim et al., 2015). IL-1β is expressed in a wide range of tissues and a wide variety of cells. Its presence was confirmed in gastrointestinal tract either IL-1β a promotes differentiation of monocytes into conventional dendritic cells and supports the proliferation of activated B-lymphocytes. IL-1 family cytokine members trigger innate inflammation and works as damage-associated molecular patterns (DAMPs) (Kaneko et al., 2019; K. S. Kim et al., 2015). Van Belleghem et al. showed that *Staphylococcus* specific phage, as well as *Pseudomonas* specific phage triggered immune response based on IL-1β. Both these phages belong to Kayviruses (van Belleghem et al., 2017). Lambdavirus, Kayvirus, and Salmonella prophages are viruses specific to potential harmful bacteria that in terms of decreased activity of IL-1β may have an opportunity to develop infection. Variants within the IL-1β gene can impair its function. Impaired IL-1β function can result in the proliferation of particular pathogens, and thus increase the presence of bacteriophages specific to them. Such a phenomenon may result in concomitance along with individual alleles. It is possible that immunological disfunctions related to variants in IL-1β coding gene result in higher frequency of pathogenic bacteria in microbiomes and this is linked to increased representation of bacteriophages that infect these bacteria.

Alleles in the area of *IL23R* gene were previously reported as related to gastritis (Zandi et al., 2014). This gene encodes a pro-inflammatory cytokine and it can be activated by macrophages and dendritic cells. Studies in mice showed that variants of *IL23R* gene might play a role in autoimmune diseases, also, specific mutations could lead to lower expression of this gene in tissues related to gastrointestinal tract (Peng et al., 2017; Tang et al., 2012). Deficiency of an important pro-inflammatory cytokine might result in conditions promoting survival of the

hosts of Punavirus (*Escherichia sp.*), Lambdavirus (*Esherichia sp.*), and Kayvirus (*Staphylococcus sp.*) bacteriophages in the stomach, as observed in this study (Table 15).

Interleukin 22 (IL-22) plays an important role in inducing antimicrobial immunity and maintaining the integrity of the mucosal barrier in the intestine. IL-22 exhibits a variety of metabolic benefits as it improves insulin sensitivity, preserves intestinal mucosal barrier and endocrine function, reduces endotoxemia and chronic inflammation, and maintains lipid metabolism in the liver and adipose tissues (X. Wang et al., 2014). Other studies showed that proper functions of IL-22 have a positive influence on microbiome diversity which leads to preventive properties (Hammer et al., 2017). Nevertheless, IL-22 disruption can cause dysbiosis and also impair mucosal immunity (X. Wang, Ota, et al., 2014). Data obtained in my study show that the variant rs976748 is associated with presence of Kayvirus bacteriophages. This SNP is upstream gene variant – sequence located 5' of the IL-22 gene. A variant that occurs upstream of a transcript but within the gene's coding region due to alternately transcribed isoforms and can alter the expression of this gene. The impairment of its function can cause favorable conditions for the overgrowth of potentially pathogenic *E. coli* that is a host for bacteriophages Lambdavirus and Punavirus- the groups detected as more frequent in this study (Table 15).

Interleukin 6 belongs to the family of cytokines that takes part in B-cell and T-cell stimulation. Neutralizing IL-6 family cytokines decreases risk of autoimmune diseases, but significantly increases the risk of developing a bacterial infection (Rose-John, 2018). Associations between microbiome diversity and interleukin 6 secretion was previously described by Smith et al, (2019). Increased abundance of Bacteroidetes and Firmicutes was associated with Interleukin-6 increased concentration (Smith et al., 2019). Rs2069833 allele identified in this study as associated to phageome composition is located in a non-coding region that previously was reported as one of the factors strongly associated with prostate cancer risk increase (Pierce et al., 2009). This variant may have influence on final concentration of Il-6 circulating in blood. Concomitance with phages specific to *Streptococcus sp.*, *Staphylococcus sp.*, and *Escherichia sp.* may indicate the creation of conditions with decreased immunological control that allows for the development of pathogens from these bacterial families.

MTHFR variant rs4846051 was previously described in the literature as related to drug metabolism pathways and birth defects (Ahn & Park, 2017; Aneji et al., 2012; Borobia et al., 2018; D'cruz et al., 2020; Rose-John, 2018). Encoding methylenetetrahydrofolate reductase is important for homocysteine recycling process. Mutations in this gene lead to decreased efficiency of the enzyme activity. This further leads to increased level of homocysteine in blood. Increased susceptibility to pathogen infection such as *H. pylori* was documented where homocysteine level was elevated (Park et al., 2017). High concomitance with Lambdavirus specific to *Escherichia coli* opportunistic pathogen may indicate that with these variant favorable conditions are created for this bacterium in the stomach environment.

Peptidoglycan recognition proteins coded by *PGLYRP4* gene have a crucial role in antimicrobial and anticancer immunological processes. Spatial structure of the main protein's domain is similar to bacteriophage type 2 amidases that are involved in cell wall binding and exolytic activity among phages (Igartua et al., 2017b; Son et al., 2018; Z. Zhang et al., 2012). PGLYRP can affect microbiome composition in several way such as: regulation of reactive oxygen species, presence in mammalian milk and expression in mucous membranes throughout all segments of the gastrointestinal tract. Deficiency of PGLYRP protein leads to dysbiosis and significant drop of microbiome diversity which may lead to health complications (Royet et al., 2011; Z. Zhang et al., 2012). Changes in the diversity of the microbiome by regulating the recognition of factors on the surface of bacterial cell walls can also lead to regulation of the composition of local microbiome. Disruption of the immune response in the gastric mucosal layer can lead to the creation of favorable conditions for infection by opportunistic microorganisms belonging to groups against which Lambdavirus (*E. coli*), Kavirus (*S. aureus*) and Punavirus (*S. enterica*) bacteriophages are specific; these phages were found corelated to *PGLYRP4* gene in this study (Table 15).

Variants among involved in cell metabolism genes such AURKA and RBMS1 were identified with concomitance in Lambdavirus, Teseptimavirus and Punavirus (Table 15). Deficiency of Aurora-A protein encoded by Aurka gene lead to dysbiosis and obesity promotion in human while mutations in RBMS1 gene is associated with increased risk of type 2 diabetes(Sánchez-

Maldonado et al., 2022; N. Sun et al., 2021). In both cases dysbiosis in bacterial microbiomes can be linked to changes in phageomes, which is in line with my observations presented herein.

11.9. Genes related to gastritis-associated pathogens

Helicobacter pylori has been demonstrated in many studies as associated with gastritis, development of stomach cancer, and reflux. It is a Gram-negative bacterium that is able to colonize gastric epithelial cells and gastric mucosa. Interleukin 22 expression level is significantly higher among patients with *H. pylori* infection. Data regarding *H. pylori* in this study is presented in the Table 16. *H. pylori* triggers immunological response that can damage mucosal cell layer in stomach creating opportunity for *H. pylori* to develop persistent infection (Sanaii et al., 2019). Increased TSPAN8 expression promotes gastric cancer (Wei et al., 2015). Patients infected with *H. pylori* and carriers of specific -31CC genotype had significantly lower expression of IL-1β. Additionally, compared to uninfected controls, patients with duodenal ulcers who had *H. pylori* infection had a significantly higher frequency of the T/C haplotype of the IL-1β 511 and IL-1β 31 loci (Chakravorty et al., 2006). The IL-1β -511C and -31T alleles, as well as the -511C/-31T and -511T/-31T haplotypes, are linked to an elevated risk of chronic gastritis and stomach ulcer, as demonstrated in the southern Mexican population (Ren et al., 2019). Gastric cancer is more likely to occur in a subgroup of people who have *Helicobacter pylori* infection and carry the LTA +252G allele. There may also occur an interaction between IL-10 and LTA that further increases the risk of gastric cancer (He et al., 2012). Nod1 detection of *H. pylori* was dependent on the bacterial type IV secretion system, encoded by the *H. pylori* cag pathogenicity island, delivering peptidoglycan to host cells (Viala et al., 2004). TLR expression facilitates interaction between *H. pylori* and gastric cancer cells. TLR expression by gastric carcinoma cells may be harmful since *H. pylori* might provide substances that promote stomach cancer, like IL-8, through epithelial TLR expression (Schmaußer et al., 2005). The correlation of the presence of *H. pylori* with variants in genes responsible for the immune response (IL22, CCDC33, IL23R, IL1B, IL6, NOD1, TLR4) and the cell cycle (TSPAN8, CACNA1A, CDKAL1, LTA) suggests that they create favorable conditions for the growth of this pathogen by disrupting their function. Correlations presented in the Table 16. along with reports from other studies indicate that variants among reported

genes may increase risk of gastrointestinal disorders together with the biological agent that might be the bacterium *H. pylori*.

Data regarding *R. mucilaginosa* in this study is presented in the Table 17. The stomach biopsy sample yielded a total of 18 non-*H. pylori* bacterial genera (43 species), the majority of which were gram-positive bacteria, including *R. mucilaginosa* (Hu et al., 2012). Alleles correlated with the occurrence of *R. mucilaginosa* may suggest that by weakening the immune response due to a reduction in gene efficiency (like the genes L22, TMEM91, MADCAM1, ELANE, IL23R, GHRL, IL12A, TLR10, CDKAL1, NOD1, SLC30A8), they create conditions for the growth of this bacterium. Due to its proven co-infection properties in gastrointestinal disorders, mutations directly affecting the growth promotion of *R. mucilaginosa* may contribute to an increased risk of gastrointestinal disease.

Prevotella melaninogenica bacterium creates acid in the medium and can survive in acidic conditions when growth on blood-containing media. Data regarding *P. melaninogenica* in this study is presented in the Table 18. Enrichment with *Prevotella spp*., namely *P. melaninogenica*, may cause rise of stomach pH above levels characteristic for non-atrophic gastritis. Additionally, iron inhibits proliferation of *Prevotella* species. In the atrophic gastritis, the gastric mucosa is destroyed, causing blood and tissue fluids to flow into the gastric fluids, which may lead to a higher iron level than in non-atrophic gastritis (T. Dong et al., 2017b). Prostaglandins are a group of hormones, derivatives of arachidonic acid, which are formed in the tissues of the human body and act at the site where they are formed. Prostaglandins are responsible for such physiological processes in the body as the response to pathogens. *Prevotella* spp. growth is therefore inhibited in the group of people with atrophic gastritis. Mutations among MUC1 gene may promote bacterial growth by lowering mucosal immunity. Activity of MUC-C was associated with colorectal and gastric carcinomas (Kufe, 2009). The presence of *P. melaninogenica* significantly affects the diversity of the bacterial microbiota in the stomach (Table 14.) and is significantly related to variant in MUC1 gene (Table 19.). Through reduced mucosal immunity, favorable conditions are created for the proliferation of potential pathogens like *P. melaninogenica.* This may be the reason for the reduced diversity of the microbiome.

During the analysis of the composition of the gastric microbiome in this group of patients, *Neisseria sp*. was identified. Detailed analysis of sequences revealed that these *Neisseria* strains were mostly *Neisseria subflava*. In terms of pathological interactions, *Neisseria sp.* and *H. pylori* coinfection was strongly linked to lymph follicle formation (Nakamura et al., 2006b). Data regarding *N. subflava* in this study is presented in the table 19. In European populations, a variant in the fat mass and obesity-associated gene (FTO) is linked to body mass index (C. Zhang et al., 2009). CLEC16A – C-type lectin domain containing 16A. Expression of this gene is highly associated with B-lymphocytes, NK, and dendritic cells activation. Cell surface-associated mucin 1, also known as polymorphic epithelial mucin or epithelial membrane antigen or EMA, is a mucin in humans encoded by the MUC1 gene. MUC1 is a glycoprotein with extensive O-glycosylation of its extracellular domain. Mutations in genes related to immunological response processes can lead to its impairment. In light of reports where *N. subflava* may be a concomitant factor in *H. pylori* infections, this may suggest that identified variants (Table 19) could potentially increase the risk of gastritis and related conditions.

Data presented in the Table 20 is related to *Prevotella jejuni*. Its cells are Gram-negative bacilli that are obligately anaerobic and non-motile, it can bind to human intestinal tissues and to agglutinate human AB and O erythrocytes (Hedberg et al., 2013). Analysis of data from patients included in this study showed that a variant within the CDKAL1 gene (chr6:20657114 T>C) is strongly associated with the presence of *P jejuni* in gastric sample (Table 20). CDKAL1 alleles were identified as a susceptibility gene for type 2 diabetes in five subsequent GWAS in several population cohorts of European and Asian ancestry. It has also been reported that CDKAL1 genetic variations are linked to -cell function (Schroner et al., 2012). detected variants in genes that may be responsible for increasing the risk of developing type two diabetes or obesity may negatively affect the diversity of the microbiome. At the same time, this creates conditions for overgrowth of pathogenic bacteria such as *P jejuni.*

11.10. Host and phage associations with Single Nucleotide Polymorphisms.

Data presented in the Table 21. shows association between bacteriophages and their potential hosts and SNPs detected in hosts' genomes. These mutations include genes from both cells

signaling, immune responses, and metabolism. Disorders in these strategic processes can affect the development of many diseases, as well as the propagation of bacterial survival in body niches such as the human stomach. These alleles, through potential impairment of the function of proteins encoded by the modified genes, may lead to changes in the composition of the phageome of the human stomach by affecting the bacterial part of the microbiota (Cadwell, 2015; Ma et al., 2018; Manrique et al., 2016; Virgin, 2014; Zárate et al., 2017). The mutations shown may predispose certain microorganisms to grow within a particular microbiota. However, the tripartite relationship described here is also susceptible to external factors. Both the bacterial and viral composition of humans is not predetermined by their genetic factors. External factors such as diet, developmental environment, or the geographic zone where a person leads his life must also be taken into account in the formation of such (Cadwell, 2015; Zárate et al., 2017).

12. Conclusions

- Specific health disorders according to the ICD10 classification correlate with bacterial components of the gastric microbiota
- Specific health disorders according to the ICD10 classification correlate with bacteriophage elements of the gastric microbiome such as Clostridioides prophages, Tequatrovirus, Inovirus
- Differences between bacterial microbiome compositions clustered ICD-10 classified samples are insignificant in the investigated group
- Change in phageome composition in chronic hypertensive disease has been observed, but changes in the composition of bacterial components of the microbiome have not been observed, thus suggesting a direct link between phage community and the disease (not mediated by bacterial host contribution)
- There is a link between mutations in genes responsible for cell signaling and immune response and presence of bacteriophages Kayvirus, Punavirus, Lambdavirus or Teseptimavirus in the investigated phageomes

- Single nucleotide variants identified and correlated with gastritis-associated pathogens can affect the entire microbiome composition and predispose specific species to grow thereby leading to dysbiosis
- In most cases, phageome composition seem to depend on bacterial composition of microbiome and enriched components of phageome probably indicate that relevant bacterial hosts were enriched in the bacterial part of microbiome
- The presence of specific alleles in genes can promote the growth of certain bacteria by impairing the function of the genes in which they occur

13. Streszczenie

Postęp w dziedzinie sekwencjonowania następnej generacji zapoczątkował bardziej szczegółowe badania ludzkiego mikrobiomu. Metoda ta pozwoliła na jednoczesne sekwencjonowanie wielu próbek, co obniżyło koszty, a także zbadanie frakcji mikroorganizmów, które wcześniej nie mogły być badane ze względu na trudności w hodowli.

Podczas trwania tego projektu, od pacjentów Pracowni Endoskopii Wojewódzkiego Szpitala Specjalistycznego, podczas badania gastroskopii pobierany był wycinek części odźwiernikowej żołądka. Od pacjentów pobierana była także krew do dalszej analizy pojedynczych mutacji nukleotydowych. Informacje o pacjentach zostały zebrane podczas wywiadu poprzedzającego badanie, a także za pomocą bazy danych medycznych Asseco Medical Management Solutions (AMMS). Po wyizolowaniu DNA z próbek, zostały one zsekwencjonowane za pomocą aparatury Illumina NextSeq550 oraz Ion Torrent Personal Genome Machine. Dane ze wszystkich rodzajów analizy zostałe ustandaryzowane i przeznaczone do analizy statystycznej.

Analiza statystyczna składników bakteryjnych i chorób sklasyfikowanych w Międzynarodowej Statystycznej Klasyfikacji Chorób ICD-10 ujawniła czternaście korelacji. Analiza statystyczna komponentów bakteriofagowych ujawniła trzy korelacje z chorobami sklasyfikowanymi w ICD-10. Trzynaście pojedynczych mutacji nukleotydowych, jest powiązane ze statystycznie istotnymi zmianami w różnorodności mikrobiomu. Sześć polimorfizmów nukleotydowych zostało powiązanych z obecnością bakteriofagów z rodzin Kayvirus, Punavirus, Lambdavirus oraz Teseptimavirus. Warianty pojedynczych nukleotydów występujące w patogenach związanych z zapaleniem błony śluzowej żołądka mogą zmieniać cały skład mikrobiomu i predysponować określone gatunki do wzrostu, co skutkuje dysbiozą. Obecność specyficznych alleli w genach może promować wzrost bakterii poprzez upośledzenie funkcji genów, w których występują.

SNP znalezione w patogenach związanych z zapaleniem błony śluzowej żołądka mogą wpływać na skład mikrobiomu i predysponować określone gatunki do wzrostu, co skutkuje dysbiozą i rozwojem stanów chorobowych.

14. Abstract

The advancement of Next-Generation Sequencing has supported the study of the human microbiome. This method allowed for the simultaneous sequencing of multiple samples, which reduced the cost. It enabled researchers to learn about microorganism fractions that could not previously be studied due to cultivation difficulties.

Patients in this study were subjected to gastroscopy by gastroenterologists at the Regional Specialized Hospital in Wroclaw. Over the course of two years, 148 gastric samples were collected from patients at the Endoscopy Department of the Regional Specialist Hospital in Wroclaw. Each patient had two separate specimens taken, and blood samples were collected for further Single Nucleotide Polymorphisms (SNP) analysis. Patients' information was gathered through interviews and the medical database Asseco Medical Management Solutions (AMMS). Following the isolation of DNA from samples, Illumina NextSeq550 and Ion Torrent Personal Genome Machine were used for DNA Next Generation Sequencing. Data from all three types of NGS data analysis were converted into standard formats for further statistical data analysis.

Statistical analysis of bacterial components and ICD-classified diseases revealed fourteen correlations. Statistical analysis of bacteriophage components revealed three correlations with ICD-classified diseases. The analysis revealed 13 SNPs that were associated with significant changes in microbiome diversity. Six different Single Nucleotide Polymorphisms in this gene have been linked to the presence of bacteriophages from the Kayvirus, Punavirus, Lambdavirus, and Teseptimavirus families. Single nucleotide variants found in gastritis-associated pathogens can alter the entire microbiome composition and predispose specific species to grow, resulting in dysbiosis. The presence of specific alleles in genes can promote bacterial growth by impairing the function of the genes in which they occur.

SNPs found in gastritis-associated pathogens can alter the entire microbiome composition and predispose specific species to grow, resulting in dysbiosis and developing disease states.

15. Bibliography

Abdellah, Z., Ahmadi, A., Ahmed, S., Aimable, M., Ainscough, R., Almeida, J., Almond, C., Ambler, A., Ambrose, K., Ambrose, K., Andrew, R., Andrews, D., Andrews, N., Andrews, D., Apweiler, E., Arbery, H., Archer, B., Ash, G., Ashcroft, K., … Kamholz, S. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–945. https://doi.org/10.1038/NATURE03001

Abril, A. G., Carrera, M., Böhme, K., Barros-Velázquez, J., Cañas, B., Rama, J. L. R., Villa, T. G., & Calo-Mata, P. (2020). Characterization of Bacteriophage Peptides of Pathogenic Streptococcus by LC-ESI-MS/MS: Bacteriophage Phylogenomics and Their Relationship to Their Host. *Frontiers in Microbiology*, *11*, 1241. https://doi.org/10.3389/FMICB.2020.01241/BIBTEX

Ackert, L. T. (2007). The "cycle of life" in ecology: Sergei Vinogradskii's Soil Microbiology, 1885-1940. *Journal of the History of Biology*, *40*(1), 109–145. https://doi.org/10.1007/S10739-006-9104-6

Ahmadian, A., Ehn, M., & Hober, S. (2006). Pyrosequencing: History, biochemistry and future. *Clinica Chimica Acta*, *363*(1–2), 83–94. https://doi.org/10.1016/J.CCCN.2005.04.038

Ahn, E., & Park, T. (2017). Analysis of population-specific pharmacogenomic variants using next-generation sequencing data. *Scientific Reports*, *7*(1). https://doi.org/10.1038/S41598-017-08468-Y

Akaza, N., Akamatsu, H., Numata, S., Yamada, S., Yagami, A., Nakata, S., & Matsunaga, K. (2016). Microorganisms inhabiting follicular contents of facial acne are not only Propionibacterium but also Malassezia spp. *The Journal of Dermatology*, *43*(8), 906–911. https://doi.org/10.1111/1346-8138.13245

Alexander, S. M., Retnakumar, R. J., Chouhan, D., Devi, T. N. B., Dharmaseelan, S., Devadas, K., Thapa, N., Tamang, J. P., Lamtha, S. C., & Chattopadhyay, S. (2021). Helicobacter pylori in Human Stomach: The Inconsistencies in Clinical Outcomes and the Probable Causes. *Frontiers in Microbiology*, *12*, 2277. https://doi.org/10.3389/FMICB.2021.713955/BIBTEX

Aneji, C. N., Northrup, H., & Sing Au, K. (2012). Deep sequencing study of the MTHFR gene to identify variants associated with myelomeningocele. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, *94*(2), 84–90. https://doi.org/10.1002/BDRA.22884

Angelakis, E., Armougom, F., Carrière, F., Bachar, D., Laugier, R., Lagier, J. C., Robert, C., Michelle, C., Henrissat, B., & Raoult, D. (2015). A Metagenomic Investigation of the Duodenal Microbiota Reveals Links with Obesity. *PLoS ONE*, *10*(9). https://doi.org/10.1371/JOURNAL.PONE.0137784

Angelakis, E., Armougom, F., Million, M., & Raoult, D. (2012). The relationship between gut microbiota and weight gain in humans. *Future Microbiology*, *7*(1), 91–109. https://doi.org/10.2217/FMB.11.142

Ansari, D., Andersson, R., Bauden, M. P., Andersson, B., Connolly, J. B., Welinder, C., Sasor, A., & Marko-Varga, G. (2015). Protein deep sequencing applied to biobank samples from patients with pancreatic cancer. *Journal of Cancer Research and Clinical Oncology*, *141*(2), 369–380. https://doi.org/10.1007/S00432-014-1817-X/FIGURES/6

Assarsson, M., Duvetorp, A., Dienus, O., Söderman, J., & Seifert, O. (2018). Significant Changes in the Skin Microbiome in Patients with Chronic Plaque Psoriasis after Treatment with Narrowband Ultraviolet B. *Acta Dermato-Venereologica*, *98*(4), 428–436. https://doi.org/10.2340/00015555-2859

Atarashi, K., Suda, W., Luo, C., Kawaguchi, T., Motoo, I., Narushima, S., Kiguchi, Y., Yasuma, K., Watanabe, E., Tanoue, T., Thaiss, C. A., Sato, M., Toyooka, K., Said, H. S., Yamagami, H., Rice, S. A., Gevers, D., Johnson, R. C., Segre, J. A., … Honda, K. (2017). Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation. *Science (New York, N.Y.)*, *358*(6361), 359–365. https://doi.org/10.1126/SCIENCE.AAN4526

Avila-Herrera, A., Thissen, J., Urbaniak, C., Be, N. A., Smith, D. J., Karouia, F., Mehta, S., Venkateswaran, K., & Jaing, C. (2020). Crewmember microbiome may influence microbial composition of ISS habitable surfaces. *PLOS ONE*, *15*(4), e0231838. https://doi.org/10.1371/JOURNAL.PONE.0231838

Awany, D., Allali, I., Dalvie, S., Hemmings, S., Mwaikono, K. S., Thomford, N. E., Gomez, A., Mulder, N., & Chimusa, E. R. (2019). Host and microbiome genome-wide association studies: Current state and challenges. *Frontiers in Genetics*, *10*(JAN), 637. https://doi.org/10.3389/FGENE.2018.00637/BIBTEX

Aziz, R. K. (2009). A hundred-year-old insight into the gut microbiome! *Gut Pathogens*, *1*(1), 21. https://doi.org/10.1186/1757-4749-1-21

Baj, J., Forma, A., Sitarz, M., Portincasa, P., Garruti, G., Krasowska, D., & Maciejewski, R. (2020). Helicobacter pylori Virulence Factors-Mechanisms of Bacterial Pathogenicity in the Gastric Microenvironment. *Cells*, *10*(1), 1–37. https://doi.org/10.3390/CELLS10010027

Baquero, F., & Nombela, C. (2012). The microbiome as a human organ. *Clinical Microbiology and Infection*, *18*(SUPPL. 4), 2–4. https://doi.org/10.1111/J.1469-0691.2012.03916.X

Baraldès, M. A., Domingo, P., Barrio, J. L., Pericas, R., Gurguí, M., & Vazquez, G. (2000). Meningitis due to Neisseria subflava: Case report and review. *Clinical Infectious Diseases*, *30*(3), 615–617. https://doi.org/10.1086/313700/2/30-3-615-TBL001.GIF

Bardell, D. (1982). The Roles of the Sense of Taste and Clean Teeth in the Discovery of Bacteria by Antoni van Leeuwenhoek. *MICROBIOLOGICAL REVIEWS*, *47*(1), 121–126.

Barnard, E., Shi, B., Kang, D., Craft, N., & Li, H. (2016). The balance of metagenomic elements shapes the skin microbiome in acne and health. *Scientific Reports*, *6*. https://doi.org/10.1038/SREP39491

Bayerdorffer, E., Oertel, H., Lehn, N., Kasper, G., Mannes, G. A., Sauerbruch, T., & Stoltet, M. (1989). Topographic association between active gastritis and Campylobacter pylori colonisation. In *J Clin Pathol* (Vol. 42).

Beasley, D. E., Koltz, A. M., Lambert, J. E., Fierer, N., & Dunn, R. R. (2015). The Evolution of Stomach Acidity and Its Relevance to the Human Microbiome. *PLOS ONE*, *10*(7), e0134116. https://doi.org/10.1371/JOURNAL.PONE.0134116

Bertolini, F., Ghionda, M. C., D'Alessandro, E., Geraci, C., Chiofalo, V., & Fontanesi, L. (2015). A Next Generation Semiconductor Based Sequencing Approach for the Identification of Meat Species in DNA Mixtures. *PLoS ONE*, *10*(4). https://doi.org/10.1371/JOURNAL.PONE.0121701

Besser, J., Carleton, H. A., Gerner-Smidt, P., Lindsey, R. L., & Trees, E. (2018). Next-Generation Sequencing Technologies and their Application to the Study and Control of Bacterial Infections. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *24*(4), 335. https://doi.org/10.1016/J.CMI.2017.10.013

Björklund, M., & Björklund, B. (2019). Be careful with your principal components. *Evolution*, *73*(10), 2151–2158. https://doi.org/10.1111/EVO.13835

Bonder, M. J., Kurilshikov, A., Tigchelaar, E. F., Mujagic, Z., Imhann, F., Vila, A. V., Deelen, P., Vatanen, T., Schirmer, M., Smeekens, S. P., Zhernakova, D. v., Jankipersadsing, S. A., Jaeger, M., Oosting, M., Cenit, M. C., Masclee, A. A. M., Swertz, M. A., Li, Y., Kumar, V., … Zhernakova, A. (2016). The effect of host genetics on the gut microbiome. *Nature Genetics*, *48*(11), 1407–1412. https://doi.org/10.1038/NG.3663

Borobia, A. M., Dapia, I., Tong, H. Y., Arias, P., Muñoz, M., Tenorio, J., Hernández, R., García García, I., Gordo, G., Ramírez, E., Frías, J., Lapunzina, P., & Carcas, A. J. (2018). Clinical Implementation of Pharmacogenetic Testing in a Hospital of the Spanish National Health System: Strategy and Experience Over 3 Years. *Clinical and Translational Science*, *11*(2), 189–199. https://doi.org/10.1111/CTS.12526

Brugère, J. F., Borrel, G., Gaci, N., Tottey, W., O'Toole, P. W., & Malpuech-Brugère, C. (2014). Archaebiotics: Proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease. *Gut Microbes*, *5*(1), 5. https://doi.org/10.4161/GMIC.26749

Burgueño, J. F., & Abreu, M. T. (2020). Epithelial Toll-like receptors and their role in gut homeostasis and disease. *Nature Reviews Gastroenterology & Hepatology 2020 17:5*, *17*(5), 263–278. https://doi.org/10.1038/s41575-019-0261-4

Cadwell, K. (2015). The Virome in Host Health and Disease. *Immunity*, *42*(5), 805–813. https://doi.org/10.1016/J.IMMUNI.2015.05.003

Caldwell, C. C., & Spies, M. (2017). Helicase SPRNTing through the nanopore. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(45), 11809–11811. https://doi.org/10.1073/PNAS.1716866114/ASSET/383B4895-C537-4739-AB8A-9479DD47732E/ASSETS/GRAPHIC/PNAS.1716866114FIG01.JPEG

Castelino, M., Eyre, S., Moat, J., Fox, G., Martin, P., Ho, P., Upton, M., & Barton, A. (2017). Optimisation of methods for bacterial skin microbiome investigation: primer selection and comparison of the 454 versus MiSeq platform. *BMC Microbiology*, *17*(1). https://doi.org/10.1186/S12866-017-0927-4

Chakravorty, M., Ghosh, A., Choudhury, A., Santra, A., Hembrum, J., & Roychoudhury, S. (2006). Interaction between IL1B gene promoter polymorphisms in determining susceptibility to Helicobacter pylori associated duodenal ulcer. *Human Mutation*, *27*(5), 411–419. https://doi.org/10.1002/HUMU.20299

Chang, D., Sharma, L., dela Cruz, C. S., & Zhang, D. (2021). Clinical Epidemiology, Risk Factors, and Control Strategies of Klebsiella pneumoniae Infection. *Frontiers in Microbiology*, *12*, 3955. https://doi.org/10.3389/FMICB.2021.750662/BIBTEX

Charalampopoulos, D., Pandiella, S. S., & Webb, C. (2003). Evaluation of the effect of malt, wheat and barley extracts on the viability of potentially probiotic lactic acid bacteria under acidic conditions. *International Journal of Food Microbiology*, *82*(2), 133–141. https://doi.org/10.1016/S0168-1605(02)00248-9

Checinska Sielaff, A., Urbaniak, C., Mohan, G. B. M., Stepanov, V. G., Tran, Q., Wood, J. M., Minich, J., McDonald, D., Mayer, T., Knight, R., Karouia, F., Fox, G. E., & Venkateswaran, K. (2019). Characterization of the total and viable bacterial and fungal communities associated with the International Space Station surfaces. *Microbiome*, *7*(1), 1–21. https://doi.org/10.1186/S40168-019-0666-X/FIGURES/7

Chen, N., Wang, L. lin, Xue, J., Ma, X. bo, Zhao, S., Rong, R. xue, Li, H. quan, Ding, L., Zheng, M. zhi, Chen, Y. ying, Duan, F., & Shen, Y. liang. (2014). Different metabolic profiles of K1 serotype and non-serotype K1 and K2 Klebsiella pneumoniae isolates in oral infection mice model. *Microbial Pathogenesis*, *75*, 41–48. https://doi.org/10.1016/J.MICPATH.2014.08.006

Cheng, C. S., Rai, K., Garber, M., Hollinger, A., Robbins, D., Anderson, S., Macbeth, A., Tzou, A., Carneiro, M. O., Raychowdhury, R., Russ, C., Hacohen, N., Gershenwald, J. E., Lennon, N., Nusbaum, C., Chin, L., Regev, A., & Amit, I. (2013). Semiconductor-based DNA sequencing of histone modification states. *Nature Communications*, *4*. https://doi.org/10.1038/NCOMMS3672

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Clarke, G., Stilling, R. M., Kennedy, P. J., Stanton, C., Cryan, J. F., & Dinan, T. G. (2014). Minireview: Gut microbiota: the neglected endocrine organ. *Molecular Endocrinology (Baltimore, Md.)*, *28*(8), 1221–1238. https://doi.org/10.1210/ME.2014-1108

Clyne, M., Labigne, A., & Drumm, B. (1995). Helicobacter pylori requires an acidic environment to survive in the presence of urea. *Infection and Immunity*, *63*(5), 1669–1673. https://doi.org/10.1128/IAI.63.5.1669-1673.1995

Cryan, J. F., O'riordan, K. J., Cowan, C. S. M., Sandhu, K. v., Bastiaanssen, T. F. S., Boehme, M., Codagnone, M. G., Cussotto, S., Fulling, C., Golubeva, A. v., Guzzetta, K. E., Jaggar, M., Long-Smith, C. M., Lyte, J. M., Martin, J. A., Molinero-Perez, A., Moloney, G., Morelli, E., Morillas, E., … Dinan, T. G. (2019). The Microbiota-Gut-Brain Axis. *Physiological Reviews*, *99*(4), 1877–2013. https://doi.org/10.1152/PHYSREV.00018.2018

Cui, J., Hou, S., Liu, B., Yang, M., Wei, L., Du, S., & Li, S. (2022). Species composition and overall diversity are significantly correlated between the tongue coating and gastric fluid microbiomes in gastritis patients. *BMC Medical Genomics*, *15*(1). https://doi.org/10.1186/S12920-022-01209-9

Day, L. A. (2008). Inoviruses. *Encyclopedia of Virology*, 117–124. https://doi.org/10.1016/B978-012374410-4.00398-8

Dayama, G., Priya, S., Niccum, D. E., Khoruts, A., & Blekhman, R. (2020). Interactions between the gut microbiome and host gene regulation in cystic fibrosis. *Genome Medicine*, *12*(1), 1–15. https://doi.org/10.1186/S13073-020-0710-2/FIGURES/3

D'cruz, L. G., McEleney, K. G., Tan, K. B. C., Shukla, P., Gardiner, P. v., Connolly, P., Conway, C., Cobice, D., & Gibson, D. S. (2020). Clinical and Laboratory Associations with Methotrexate Metabolism Gene Polymorphisms in Rheumatoid Arthritis. *Journal of Personalized Medicine*, *10*(4), 1–14. https://doi.org/10.3390/JPM10040149

de Jonge, P. A., Wortelboer, K., Scheithauer, T. P. M., van den Born, B. J. H., Zwinderman, A. H., Nobrega, F. L., Dutilh, B. E., Nieuwdorp, M., & Herrema, H. (2022). Gut virome profiling identifies a widespread bacteriophage family associated with metabolic syndrome. *Nature Communications 2022 13:1*, *13*(1), 1–15. https://doi.org/10.1038/s41467-022-31390-5

de Paepe, M., Leclerc, M., Tinsley, C. R., & Petit, M. A. (2014). Bacteriophages: An underestimated role in human and animal health? *Frontiers in Cellular and Infection Microbiology*, *5*(MAR), 39. https://doi.org/10.3389/FCIMB.2014.00039/BIBTEX

Deckers-Hebestreit, G., & Altendorf, K. (1996). The F0F1-type ATP synthases of bacteria: structure and function of the F0 complex. *Annual Review of Microbiology*, *50*, 791–824. https://doi.org/10.1146/ANNUREV.MICRO.50.1.791

Deo, P. N., & Deshmukh, R. (2019). Oral microbiome: Unveiling the fundamentals. *Journal of Oral and Maxillofacial Pathology : JOMFP*, *23*(1), 122. https://doi.org/10.4103/JOMFP.JOMFP_304_18

Ding, C., & He, X. (2004). K-means clustering via principal component analysis. *Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004*, 225–232. https://doi.org/10.1145/1015330.1015408

Diricks, M., Kohl, T. A., Käding, N., Leshchinskiy, V., Hauswaldt, S., Jiménez Vázquez, O., Utpatel, C., Niemann, S., Rupp, J., & Merker, M. (2022). Whole genome sequencing-based classification of human-related Haemophilus species and detection of antimicrobial resistance genes. *Genome Medicine*, *14*(1), 1–15. https://doi.org/10.1186/S13073-022-01017-X/FIGURES/3

Dong, Q., Brulc, J. M., Iovieno, A., Bates, B., Garoutte, A., Miller, D., Revanna, K. v., Gao, X., Antonopoulos, D. A., Slepak, V. Z., & Shestopalov, V. I. (2011). Diversity of bacteria at healthy human conjunctiva. *Investigative Ophthalmology and Visual Science*, *52*(8), 5408–5413. https://doi.org/10.1167/IOVS.10-6939/-/DCSUPPLEMENTAL

Dong, T., Feng, Q., Liu, F., Chang, L. K., Zhou, X., Han, M., Tian, X., Zhong, N., & Liu, S. (2017a). Alteration of stomach microbiota compositions in the progression of gastritis induces nitric oxide in gastric cell. *Experimental and Therapeutic Medicine*, *13*(6), 2793. https://doi.org/10.3892/ETM.2017.4373

Dong, T., Feng, Q., Liu, F., Chang, L. K., Zhou, X., Han, M., Tian, X., Zhong, N., & Liu, S. (2017b). Alteration of stomach microbiota compositions in the progression of gastritis induces nitric oxide in gastric cell. *Experimental and Therapeutic Medicine*, *13*(6), 2793–2800. https://doi.org/10.3892/ETM.2017.4373/HTML

Donovan, P. D., Gonzalez, G., Higgins, D. G., Butler, G., & Ito, K. (2018). Identification of fungi in shotgun metagenomics datasets. *PloS One*, *13*(2). https://doi.org/10.1371/JOURNAL.PONE.0192898

Dréno, B., Araviiskaia, E., Berardesca, E., Gontijo, G., Sanchez Viera, M., Xiang, L. F., Martin, R., & Bieber, T. (2016). Microbiome in healthy skin, update for dermatologists. *Journal of the European Academy of Dermatology and Venereology : JEADV*, *30*(12), 2038–2047. https://doi.org/10.1111/JDV.13965

Dressman, J. B., Berardi, R. R., Dermentzoglou, L. C., Russell, T. L., Schmaltz, S. P., Barnett, J. L., & Jarvenpaa, K. M. (1990). Upper gastrointestinal (GI) pH in young, healthy men and women. *Pharmaceutical Research*, *7*(7), 756–761. https://doi.org/10.1023/A:1015827908309

Drösler, S. E., Weber, S., & Chute, C. G. (2021). ICD-11 extension codes support detailed clinical abstraction and comprehensive classification. *BMC Medical Informatics and Decision Making*, *21*. https://doi.org/10.1186/S12911-021-01635-2

Duan, X., Chen, P., Xu, X., Han, M., & Li, J. (2022). Role of Gastric Microorganisms Other than Helicobacter pylori in the Development and Treatment of Gastric Diseases. *BioMed Research International*, *2022*. https://doi.org/10.1155/2022/6263423

Dworkin, M., & Gutnick, D. (2012). Sergei Winogradsky: a founder of modern microbiology and the first microbial ecologist. *FEMS Microbiology Reviews*, *36*(2), 364–379. https://doi.org/10.1111/J.1574-6976.2011.00299.X

Ellis, S. R., Nguyen, M., Vaughn, A. R., Notay, M., Burney, W. A., Sandhu, S., & Sivamani, R. K. (2019). The Skin and Gut Microbiome and Its Role in Common Dermatologic Conditions. *Microorganisms*, *7*(11), 550. https://doi.org/10.3390/MICROORGANISMS7110550

Ezeji, J. C., Sarikonda, D. K., Hopperton, A., Erkkila, H. L., Cohen, D. E., Martinez, S. P., Cominelli, F., Kuwahara, T., Dichosa, A. E. K., Good, C. E., Jacobs, M. R., Khoretonenko, M., Veloo, A., & Rodriguez-Palacios, A. (2021). Parabacteroides distasonis: intriguing aerotolerant gut anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in human health. *Gut Microbes*, *13*(1). https://doi.org/10.1080/19490976.2021.1922241

Farrer, R. A., Kemen, E., Jones, J. D. G., & Studholme, D. J. (2009). De novo assembly of the Pseudomonas syringae pv. syringae B728a genome using Illumina/Solexa short sequence reads. *FEMS Microbiology Letters*, *291*(1), 103–111. https://doi.org/10.1111/J.1574-6968.2008.01441.X

Federici, S., Nobs, S. P., & Elinav, E. (2020). Phages and their potential to modulate the microbiome and immunity. *Cellular & Molecular Immunology*. https://doi.org/10.1038/s41423-020-00532-4

Feng, T., Leptihn, S., Dong, K., Loh, B., Zhang, Y., Stefan, M. I., Li, M., Guo, X., & Cui, Z. (2021). JD419, a Staphylococcus aureus Phage With a Unique Morphology and Broad Host Range. *Frontiers in Microbiology*, *12*, 840. https://doi.org/10.3389/FMICB.2021.602902/BIBTEX

Fidel, P. L. (2002). Immunity to Candida. *Oral Diseases*, *8*(SUPPL. 2), 69–75. https://doi.org/10.1034/j.1601-0825.2002.00015.x

Fortier, L. C. (2018). Bacteriophages contribute to shaping clostridioides (Clostridium) difficile species. *Frontiers in Microbiology*, *9*(AUG), 2033. https://doi.org/10.3389/FMICB.2018.02033/BIBTEX

Foster, J. W. (2004). Escherichia coli acid resistance: tales of an amateur acidophile. *Nat. Rev. Microbiol.*, *2*(11), 898–907. https://doi.org/10.1038/nrmicro1021

Franasiak, J. M., & Scott, R. T. (2015). Reproductive tract microbiome in assisted reproductive technologies. *Fertility and Sterility*, *104*(6), 1364–1371. https://doi.org/10.1016/J.FERTNSTERT.2015.10.012

Franzosa, E. A., Huang, K., Meadow, J. F., Gevers, D., Lemon, K. P., Bohannan, B. J. M., & Huttenhower, C. (2015). Identifying personal microbiomes using metagenomic codes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(22), E2930–E2938. https://doi.org/10.1073/PNAS.1423854112/-/DCSUPPLEMENTAL

Fujiwara, M., Anstadt, E. J., Flynn, B., Morse, K., Ng, C., Paczkowski, P., Zhou, J., Mackay, S., Wasko, N., Nichols, F., & Clark, R. B. (2018). Enhanced TLR2 responses in multiple sclerosis. *Clinical and Experimental Immunology*, *193*(3), 313–326. https://doi.org/10.1111/CEI.13150

Gallardo, P., Izquierdo, M., Vidal, R. M., Chamorro-Veloso, N., Rosselló-Móra, R., O'Ryan, M., & Farfán, M. J. (2017). Distinctive gut microbiota is associated with Diarrheagenic Escherichia coli infections in chilean children. *Frontiers in Cellular and Infection Microbiology*, *7*(OCT), 424. https://doi.org/10.3389/FCIMB.2017.00424/FULL

Gantuya, B., el Serag, H. B., Saruuljavkhlan, B., Azzaya, D., Matsumoto, T., Uchida, T., Oyuntsetseg, K., Oyunbileg, N., Davaadorj, D., & Yamaoka, Y. (2021). Advantage of 16S rRNA amplicon sequencing in Helicobacter pylori diagnosis. *Helicobacter*, *26*(3). https://doi.org/10.1111/hel.12790

Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R., Ober, C., Spector, T. D., Bell, J. T., Clark, A. G., & Ley, R. E. (2016). Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host & Microbe*, *19*(5), 731–743. https://doi.org/10.1016/J.CHOM.2016.04.017

Gorrie, C. L., Mirc Eta, M., Wick, R. R., Edwards, D. J., Thomson, N. R., Strugnell, R. A., Pratt, N. F., Garlick, J. S., Watson, K. M., Pilcher, D. v., McGloughlin, S. A., Spelman, D. W., Jenney, A. W. J., & Holt, K. E. (2017). Gastrointestinal Carriage Is a Major Reservoir of Klebsiella pneumoniae Infection in Intensive Care Patients. *Clinical Infectious Diseases*, *65*(2), 208–215. https://doi.org/10.1093/CID/CIX270

Goswami, K., & Sanan-Mishra, N. (2022). RNA-seq for revealing the function of the transcriptome. *Bioinformatics*, 105–129. https://doi.org/10.1016/B978-0-323-89775-4.00002-X

Grice, E. A., & Segre, J. A. (2011). The skin microbiome. *Nature Reviews. Microbiology*, *9*(4), 244–253. https://doi.org/10.1038/NRMICRO2537

Guo, J., Yu, L., Turro, N. J., & Ju, J. (2010). An Integrated System for DNA Sequencing by Synthesis Using Novel Nucleotide Analogues. *Accounts of Chemical Research*, *43*(4), 551. https://doi.org/10.1021/AR900255C

Gupta, A. K., & Gupta, U. D. (2014). Next Generation Sequencing and Its Applications. *Animal Biotechnology: Models in Discovery and Translation*, 345–367. https://doi.org/10.1016/B978-0-12-416002-6.00019-5

Hajjar, R., Ambaraghassi, G., Sebajang, H., Schwenter, F., & Su, S. H. (2020). Raoultella ornithinolytica: Emergence and Resistance. *Infection and Drug Resistance*, *13*, 1091. https://doi.org/10.2147/IDR.S191387

Hammer, A. M., Morris, N. L., Cannon, A. R., Khan, O. M., Gagnon, R. C., Movtchan, N. v., van Langeveld, I., Li, X., Gao, B., & Choudhry, M. A. (2017). Interleukin-22 Prevents Microbial Dysbiosis and Promotes Intestinal Barrier Regeneration Following Acute Injury. *Shock (Augusta, Ga.)*, *48*(6), 657. https://doi.org/10.1097/SHK.0000000000000900

Han, H. S., Lee, S. Y., Oh, S. Y., Moon, H. W., Cho, H., & Kim, J. H. (2019). Correlations of the Gastric and Duodenal Microbiota with Histological, Endoscopic, and Symptomatic Gastritis. *Journal of Clinical Medicine*, *8*(3). https://doi.org/10.3390/JCM8030312

Han, M., Yang, P., Zhong, C., & Ning, K. (2018). The Human Gut Virome in Hypertension. *Frontiers in Microbiology*, *9*, 3150. https://doi.org/10.3389/FMICB.2018.03150/BIBTEX

Handley, S. A., & Devkota, S. (2019). Going viral: A novel role for bacteriophage in colorectal cancer. *MBio*, *10*(1). https://doi.org/10.1128/MBIO.02626-18

Hanemaaijer, L., Kelleher, P., Neve, H., Franz, C. M. A. P., de Waal, P. P., van Peij, N. N. M. E., van Sinderen, D., & Mahony, J. (2021). Biodiversity of phages infecting the dairy bacterium streptococcus thermophilus. *Microorganisms*, *9*(9), 1822. https://doi.org/10.3390/MICROORGANISMS9091822/S1

Hannigan, G. D., Duhaime, M. B., Ruffin, M. T., Koumpouras, C. C., & Schloss, P. D. (2018). Diagnostic potential and interactive dynamics of the colorectal cancer virome. *MBio*, *9*(6). https://doi.org/10.1128/MBIO.02248-18/ASSET/A78B8AEB-CCA5-45A5-B25C-DC266B2BBCB1/ASSETS/GRAPHIC/MBO0061841830006.JPEG

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/S41586-020-2649-2

Hasan, N., & Yang, H. (2019). Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ*, *7*(8). https://doi.org/10.7717/PEERJ.7502

Hathroubi, S., Servetas, S. L., Windham, I., Merrell, D. S., & Ottemann, K. M. (2018). Helicobacter pylori Biofilm Formation and Its Potential Role in Pathogenesis. *Microbiology and Molecular Biology Reviews : MMBR*, *82*(2). https://doi.org/10.1128/MMBR.00001-18

Hayashi, H., Takahashi, R., Nishi, T., Sakamoto, M., & Benno, Y. (2005). Molecular analysis of jejunal, ileal, caecal and rectosigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. *Journal of Medical*

Microbiology, *54*(11), 1093–1101. https://doi.org/10.1099/JMM.0.45935-0/CITE/REFWORKS

He, B., Pan, Y., Xu, Y., Nie, Z., Chen, L., Gu, L., & Wang, S. (2012). Increased Risk for Gastric Cancer in Carriers of the Lymphotoxin-α+252G Variant Infected by Helicobacter pylori. *Https://Home.Liebertpub.Com/Gtmb*, *16*(1), 9–14. https://doi.org/10.1089/GTMB.2011.0078

Heather, J. M., & Chain, B. (2016a). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1. https://doi.org/10.1016/J.YGENO.2015.11.003

Heather, J. M., & Chain, B. (2016b). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1. https://doi.org/10.1016/J.YGENO.2015.11.003

Hedberg, M. E., Israelsson, A., Moore, E. R. B., Svensson-Stadler, L., Wai, S. N., Pietz, G., Sandstrom, O., Hernell, O., Hammarstrom, M. L., & Hammarstrom, S. (2013). Prevotella jejuni sp. nov., isolated from the small intestine of a child with coeliac disease. *International Journal of Systematic and Evolutionary Microbiology*, *63*(Pt 11), 4218. https://doi.org/10.1099/IJS.0.052647-0

Hemarajata, P., & Versalovic, J. (2012). Effects of probiotics on gut microbiota: mechanisms of intestinal immunomodulation and neuromodulation: *Http://Dx.Doi.Org/10.1177/1756283X12459294*, *6*(1), 39–51. https://doi.org/10.1177/1756283X12459294

Henrick, B. M., Yao, X. D., Zahoor, M. A., Abimiku, A., Osawe, S., & Rosenthal, K. L. (2019). TLR10 senses HIV-1 proteins and significantly enhances HIV-1 infection. *Frontiers in Immunology*, *10*(MAR), 482. https://doi.org/10.3389/FIMMU.2019.00482/FULL

Higuchi, B. S., Rodrigues, N., Gonzaga, M. I., Paiolo, J. C. C., Stefanutto, N., Omori, W. P., Pinheiro, D. G., Brisotti, J. L., Matheucci, E., Mariano, V. S., & de Oliveira, G. L. V. (2018). Intestinal dysbiosis in autoimmune diabetes is correlated with poor glycemic control and increased interleukin-6: A pilot study. *Frontiers in Immunology*, *9*(JUL), 1689. https://doi.org/10.3389/FIMMU.2018.01689/BIBTEX

Horz, H. P. (2015). Archaeal Lineages within the Human Microbiome: Absent, Rare or Elusive? *Life*, *5*(2), 1333. https://doi.org/10.3390/LIFE5021333

Hu, Y., He, L. H., Xiao, D., Liu, G. D., Gu, Y. X., Tao, X. X., & Zhang, J. Z. (2012). Bacterial flora concurrent with Helicobacter pylori in the stomach of patients with upper gastrointestinal diseases. *World Journal of Gastroenterology : WJG*, *18*(11), 1257. https://doi.org/10.3748/WJG.V18.I11.1257

Huffnagle, G. B., & Noverr, M. C. (2013). The emerging world of the fungal microbiome. *Trends in Microbiology*, *21*(7), 334. https://doi.org/10.1016/J.TIM.2013.04.002

Hufnagl, K., Pali-Schöll, I., Roth-Walter, F., & Jensen-Jarolim, E. (2020). Dysbiosis of the gut and lung microbiome has a role in asthma. *Seminars in Immunopathology 2020 42:1*, *42*(1), 75–93. https://doi.org/10.1007/S00281-019-00775-Y

Hunt, R., Sauna, Z. E., Ambudkar, S. v., Gottesman, M. M., & Kimchi-Sarfaty, C. (2009). Silent (synonymous) SNPs: should we care about them? *Methods in Molecular Biology (Clifton, N.J.)*, *578*, 23–39. https://doi.org/10.1007/978-1-60327-411-1_2

Hwang, J., Kim, B., Lee, M. J., Nam, Y., Youn, U. J., Lee, C. S., Oh, T. J., Park, H. H., Do, H., & Lee, J. H. (2022). Structural basis for the substrate specificity of an S-formylglutathione hydrolase derived from Variovorax sp. PAMC 28711. *Biochemical and Biophysical Research Communications*, *629*, 159–164. https://doi.org/10.1016/J.BBRC.2022.09.008

Igartua, C., Davenport, E. R., Gilad, Y., Nicolae, D. L., Pinto, J., & Ober, C. (2017a). Host genetic variation in mucosal immunity pathways influences the upper airway microbiome. *Microbiome*, *5*(1). https://doi.org/10.1186/S40168-016-0227-5

Igartua, C., Davenport, E. R., Gilad, Y., Nicolae, D. L., Pinto, J., & Ober, C. (2017b). Host genetic variation in mucosal immunity pathways influences the upper airway microbiome. *Microbiome*, *5*(1), 1–17. https://doi.org/10.1186/S40168-016-0227-5/FIGURES/7

Inbaraj, S., Angappan, | M, Thomas, | Prasad, Kumar, M., Irungbam, K., Med, |, Verma, R., Nagaleekar Viswas, | K, Abhishek, |, Rawat, M., & Chaudhuri, P. (2022). Isolation and characterization of bacteriophage Ib_pec2 against shigatoxigenic Escherichia coli. *Journal of Basic Microbiology*. https://doi.org/10.1002/JOBM.202200398

Jia, K., Peng, Y., Chen, X., Jian, H., Jin, M., Yi, Z., Su, M., Dong, X., & Yi, M. (2022). A Novel Inovirus Reprograms Metabolism and Motility of Marine Alteromonas. *Microbiology Spectrum*. https://doi.org/10.1128/SPECTRUM.03388-22

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). https://doi.org/10.1098/RSTA.2015.0202

Jończyk, E., Kłak, M., Międzybrodzki, R., & Górski, A. (2011). The influence of external factors on bacteriophages—review. *Folia Microbiologica*, *56*(3), 191. https://doi.org/10.1007/S12223-011-0039-8

Joossens, M., Huys, G., Cnockaert, M., de Preter, V., Verbeke, K., Rutgeerts, P., Vandamme, P., & Vermeire, S. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut*, *60*(5), 631–637. https://doi.org/10.1136/gut.2010.223263

Kamdar, K., Johnson, A. M. F., Chac, D., Myers, K., Kulur, V., Truevillian, K., & DePaolo, R. W. (2018). Innate Recognition of the Microbiota by TLR1 Promotes Epithelial

Homeostasis and Prevents Chronic Inflammation. *The Journal of Immunology*, *201*(1), 230–242. https://doi.org/10.4049/JIMMUNOL.1701216

Kaneko, N., Kurata, M., Yamamoto, T., Morikawa, S., & Masumoto, J. (2019). The role of interleukin-1 in general pathology. *Inflammation and Regeneration 2019 39:1*, *39*(1), 1–16. https://doi.org/10.1186/S41232-019-0101-5

Kang, X., Deng, D. M., Crielaard, W., & Brandt, B. W. (2021). Reprocessing 16S rRNA Gene Amplicon Sequencing Studies: (Meta)Data Issues, Robustness, and Reproducibility. *Frontiers in Cellular and Infection Microbiology*, *11*. https://doi.org/10.3389/FCIMB.2021.720637

Kastl, A. J., Terry, N. A., Wu, G. D., & Albenberg, L. G. (2020). The Structure and Function of the Human Small Intestinal Microbiota: Current Understanding and Future Directions. *Cellular and Molecular Gastroenterology and Hepatology*, *9*(1), 33–45. https://doi.org/10.1016/J.JCMGH.2019.07.006

Kato, K., Ishida, S., Tanaka, M., Mitsuyama, E., Xiao, J. Z., & Odamaki, T. (2018). Association between functional lactase variants and a high abundance of Bifidobacterium in the gut of healthy Japanese people. *PLoS ONE*, *13*(10). https://doi.org/10.1371/JOURNAL.PONE.0206189

Katsnelson, A. (2010). DNA sequencing for the masses. *Nature*. https://doi.org/10.1038/NEWS.2010.674

Kendall, A. I. (1909). Some Observations on the Study of the Intestinal Bacteria. *Journal of Biological Chemistry*, *6*(6), 499–507. https://doi.org/10.1016/S0021-9258(18)91596-6

Khanna, A., Khanna, M., & Aggarwal, A. (2013). Serratia Marcescens- A Rare Opportunistic Nosocomial Pathogen and Measures to Limit its Spread in Hospitalized Patients. *Journal of Clinical and Diagnostic Research : JCDR*, *7*(2), 243. https://doi.org/10.7860/JCDR/2013/5010.2737

Kilian, M., Chapple, I. L. C., Hannig, M., Marsh, P. D., Meuric, V., Pedersen, A. M. L., Tonetti, M. S., Wade, W. G., & Zaura, E. (2016). The oral microbiome - an update for oral healthcare professionals. *British Dental Journal*, *221*(10), 657–666. https://doi.org/10.1038/SJ.BDJ.2016.865

Kim, G., Bae, J., Kim, M. J., Kwon, H., Park, G., Kim, S. J., Choe, Y. H., Kim, J., Park, S. H., Choe, B. H., Shin, H., & Kang, B. (2020). Delayed Establishment of Gut Microbiota in Infants Delivered by Cesarean Section. *Frontiers in Microbiology*, *11*, 2099. https://doi.org/10.3389/FMICB.2020.02099/BIBTEX

Kim, K. S., Jung, H., Shin, I. K., Choi, B. R., & Kim, D. H. (2015). Induction of interleukin-1 beta (IL-1β) is a critical component of lung inflammation during influenza A (H1N1) virus infection. *Journal of Medical Virology*, *87*(7), 1104–1112. https://doi.org/10.1002/JMV.24138

Kim, S. J., Ahn, J. H., Heo, J., Cho, H., Weon, H. Y., Hong, S. B., Kim, J. S., & Kwon, S. W. (2018). Phreatobacter cathodiphilus sp. Nov., isolated from a cathode of a microbial fuel cell. *International Journal of Systematic and Evolutionary Microbiology*, *68*(9), 2855–2859. https://doi.org/10.1099/IJSEM.0.002904/CITE/REFWORKS

Kinoshita-Daitoku, R., Kiga, K., Miyakoshi, M., Otsubo, R., Ogura, Y., Sanada, T., Bo, Z., Phuoc, T. V., Okano, T., Iida, T., Yokomori, R., Kuroda, E., Hirukawa, S., Tanaka, M., Sood, A., Subsomwong, P., Ashida, H., Binh, T. T., Nguyen, L. T., … Mimuro, H. (2021). A bacterial small RNA regulates the adaptation of Helicobacter pylori to the host environment. *Nature Communications 2021 12:1*, *12*(1), 1–12. https://doi.org/10.1038/s41467-021-22317-7

Klimka, A., Mertins, S., Nicolai, A. K., Rummler, L. M., Higgins, P. G., Günther, S. D., Tosetti, B., Krut, O., & Krönke, M. (2021). Epitope-specific immunity against Staphylococcus aureus coproporphyrinogen III oxidase. *Vaccines*, *6*(11). https://doi.org/10.1038/s41541-020-00268-2

Knodler, L. A., & Elfenbein, J. R. (2019). Salmonella enterica. *Trends in Microbiology*, *27*(11), 964–965. https://doi.org/10.1016/J.TIM.2019.05.002

Kolde, R., Franzosa, E. A., Rahnavard, G., Hall, A. B., Vlamakis, H., Stevens, C., Daly, M. J., Xavier, R. J., & Huttenhower, C. (2018a). Host genetic variation and its microbiome interactions within the Human Microbiome Project. *Genome Medicine*, *10*(1), 1–13. https://doi.org/10.1186/S13073-018-0515-8/FIGURES/4

Kolde, R., Franzosa, E. A., Rahnavard, G., Hall, A. B., Vlamakis, H., Stevens, C., Daly, M. J., Xavier, R. J., & Huttenhower, C. (2018b). Host genetic variation and its microbiome interactions within the Human Microbiome Project. *Genome Medicine*, *10*(1), 1–13. https://doi.org/10.1186/S13073-018-0515-8/FIGURES/4

Konishi, T., Matsukuma, S., Fuji, H., Nakamura, D., Satou, N., & Okano, K. (2019). Principal Component Analysis applied directly to Sequence Matrix. *Scientific Reports 2019 9:1*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-55253-0

Könönen, E., & Gursoy, U. K. (2021a). Oral Prevotella Species and Their Connection to Events of Clinical Relevance in Gastrointestinal and Respiratory Tracts. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/FMICB.2021.798763

Könönen, E., & Gursoy, U. K. (2021b). Oral Prevotella Species and Their Connection to Events of Clinical Relevance in Gastrointestinal and Respiratory Tracts. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/FMICB.2021.798763

Kufe, D. W. (2009). Functional targeting of the MUC1 oncogene in human cancers. *Cancer Biology & Therapy*, *8*(13), 1197–1203. https://doi.org/10.4161/cbt.8.13.8844

Kumpitsch, C., Koskinen, K., Schöpf, V., & Moissl-Eichinger, C. (2019). The microbiome of the upper respiratory tract in health and disease. *BMC Biology 2019 17:1*, *17*(1), 1–20. https://doi.org/10.1186/S12915-019-0703-Z

Kuzyakov, Y., & Mason-Jones, K. (2018). Viruses in soil: Nano-scale undead drivers of microbial life, biogeochemical turnover and ecosystem functions. *Soil Biology and Biochemistry*, *127*, 305–317. https://doi.org/10.1016/J.SOILBIO.2018.09.032

Lauber, C. L., Hamady, M., Knight, R., & Fierer, N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology*, *75*(15), 5111–5120. https://doi.org/10.1128/AEM.00335-09/SUPPL_FILE/SUPPLEMENTARY_FIGURE_AND_TABLES.PDF

le Noci, V., Bernardo, G., Bianchi, F., Tagliabue, E., Sommariva, M., & Sfondrini, L. (2021). Toll Like Receptors as Sensors of the Tumor Microbial Dysbiosis: Implications in Cancer Progression. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/FCELL.2021.732192

Lee, S., Sung, J., Lee, J. E., & Ko, G. P. (2011). Comparison of the Gut Microbiotas of Healthy Adult Twins Living in South Korea and the United States. *Applied and Environmental Microbiology*, *77*(20), 7433. https://doi.org/10.1128/AEM.05490-11

Lee, Y. H., Choi, S. J., Ji, J. D., & Song, G. G. (2016). Association between toll-like receptor polymorphisms and systemic lupus erythematosus: A meta-analysis update. *Lupus*, *25*(6), 593–601. https://doi.org/10.1177/0961203315622823

Leewenhoeck, A. (1684). An abstract of a letter from Mr. Anthony Leevvenhoeck at Delft, dated Sep. 17. 1683. Containing some microscopical observations, about animals in the scurf of the teeth, the substance call'd worms in the nose, the cuticula consisting of scales. *Philosophical Transactions of the Royal Society of London*, *14*(159), 568–574. https://doi.org/10.1098/RSTL.1684.0030

Li, F. teng, Zhao, J., & Pang, X. yan. (2012). [The oral microbial fingerprint on the greasy tongue coating of patients with chronic gastritis]. *Zhongguo Zhong Xi Yi Jie He Za Zhi Zhongguo Zhongxiyi Jiehe Zazhi = Chinese Journal of Integrated Traditional and Western Medicine*, *32*(10), 1331–1335. https://europepmc.org/article/med/23163140

Li, J., Stein, D., McMullan, C., Branton, D., Aziz, M. J., & Golovchenko, J. A. (2001). Ion-beam sculpting at nanometre length scales. *Nature*, *412*(6843), 166–169. https://doi.org/10.1038/35084037

Liang, G., & Bushman, F. D. (2021). The human virome: assembly, composition and host interactions. *Nature Reviews Microbiology 2021 19:8*, *19*(8), 514–527. https://doi.org/10.1038/s41579-021-00536-5

Litvak, Y., Byndloss, M. X., Tsolis, R. M., & Bäumler, A. J. (2017). Dysbiotic Proteobacteria expansion: a microbial signature of epithelial dysfunction. *Current Opinion in Microbiology*, *39*, 1–6. https://doi.org/10.1016/j.mib.2017.07.003

Liu, J., Xue, Y., & Zhou, L. (2018). Detection of gastritis-associated pathogens by culturing of gastric juice and mucosa. *International Journal of Clinical and Experimental Pathology*, *11*(4), 2214. /pmc/articles/PMC6958204/

Liu, J., Yan, R., Zhong, Q., Ngo, S., Bangayan, N. J., Nguyen, L., Lui, T., Liu, M., Erfe, M. C., Craft, N., Tomida, S., & Li, H. (2015). The diversity and host interactions of Propionibacterium acnes bacteriophages on human skin. *The ISME Journal*, *9*(9), 2078–2093. https://doi.org/10.1038/ISMEJ.2015.47

Liu, Y., Baba, Y., Ishimoto, T., Iwatsuki, M., Hiyoshi, Y., Miyamoto, Y., Yoshida, N., Wu, R., & Baba, H. (2019a). Progress in characterizing the linkage between Fusobacterium nucleatum and gastrointestinal cancer. *Journal of Gastroenterology*, *54*(1), 33–41. https://doi.org/10.1007/S00535-018-1512-9/TABLES/2

Liu, Y., Baba, Y., Ishimoto, T., Iwatsuki, M., Hiyoshi, Y., Miyamoto, Y., Yoshida, N., Wu, R., & Baba, H. (2019b). Progress in characterizing the linkage between Fusobacterium nucleatum and gastrointestinal cancer. *Journal of Gastroenterology*, *54*(1), 33–41. https://doi.org/10.1007/S00535-018-1512-9/TABLES/2

Lopetuso, L. R., Scaldaferri, F., Franceschi, F., & Gasbarrini, A. (2014). The gastrointestinal microbiome – Functional interference between stomach and intestine. *Best Practice & Research Clinical Gastroenterology*, *28*(6), 995–1002. https://doi.org/10.1016/J.BPG.2014.10.004

Lopez-Castejon, G., & Brough, D. (2011). Understanding the mechanism of IL-1β secretion. *Cytokine & Growth Factor Reviews*, *22*(4), 189. https://doi.org/10.1016/J.CYTOGFR.2011.10.001

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics and Bioinformatics*, *14*(5), 265–279. https://doi.org/10.1016/j.gpb.2016.05.004

Lu, H., Ren, Z., Li, A., Zhang, H., Jiang, J., Xu, S., Luo, Q., Zhou, K., Sun, X., Zheng, S., & Li, L. (2016). Deep sequencing reveals microbiota dysbiosis of tongue coat in patients with liver carcinoma. *Scientific Reports 2016 6:1*, *6*(1), 1–11. https://doi.org/10.1038/srep33142

Lund, P., Tramonti, A., de Biase, D., Pasteur-Fondazione, I., & Bolognetti, C. (2014). Coping with low pH: molecular strategies in neutralophilic bacteria. *FEMS Microbiology Reviews*, *38*(6), 1091–1125. https://doi.org/10.1111/1574-6976.12076

Lyon, L. (2018). 'All disease begins in the gut': was Hippocrates right? *Brain*, *141*(3), e20–e20. https://doi.org/10.1093/BRAIN/AWY017

Ma, Y., You, X., Mai, G., Tokuyasu, T., & Liu, C. (2018). A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome*, *6*(1), 1–12. https://doi.org/10.1186/S40168-018-0410-Y/FIGURES/6

Mackowiak, P. A. (2013). Recycling Metchnikoff: Probiotics, the Intestinal Microbiome and the Quest for Long Life. *Frontiers in Public Health*, *1*(NOV). https://doi.org/10.3389/FPUBH.2013.00052

Madoui, M. A., Engelen, S., Cruaud, C., Belser, C., Bertrand, L., Alberti, A., Lemainque, A., Wincker, P., & Aury, J. M. (2015). Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics*, *16*(1). https://doi.org/10.1186/s12864-015-1519-z

Manor, O., Dai, C. L., Kornilov, S. A., Smith, B., Price, N. D., Lovejoy, J. C., Gibbons, S. M., & Magis, A. T. (2020). Health and disease markers correlate with gut microbiome composition across thousands of people. *Nature Communications 2020 11:1*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-18871-1

Manrique, P., Bolduc, B., Walk, S. T., van Oost, J. der, de Vos, W. M., & Young, M. J. (2016). Healthy human gut phageome. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(37), 10400–10405. https://doi.org/10.1073/PNAS.1601060113/SUPPL_FILE/PNAS.201601060SI.PDF

Mcdermott, A. J., & Huffnagle, G. B. (2014). The microbiome and regulation of mucosal immunity. *Immunology*, *142*(1), 24–31. https://doi.org/10.1111/IMM.12231

McDonald, D., Hyde, E., Debelius, J. W., Morton, J. T., Gonzalez, A., Ackermann, G., Aksenov, A. A., Behsaz, B., Brennan, C., Chen, Y., DeRight Goldasich, L., Dorrestein, P. C., Dunn, R. R., Fahimipour, A. K., Gaffney, J., Gilbert, J. A., Gogul, G., Green, J. L., Hugenholtz, P., … Gunderson, B. (2018). American Gut: an Open Platform for Citizen Science Microbiome Research. *MSystems*, *3*(3). https://doi.org/10.1128/MSYSTEMS.00031-18/SUPPL_FILE/SYS003182229ST5.XLSX

Metchnikoff, E. (1907). *The prolongation of life; optimistic studies*,. Heinemann.

Mikheyev, A. S., & Tin, M. M. Y. (2014). A first look at the Oxford Nanopore MinION sequencer. *Molecular Ecology Resources*, *14*(6), 1097–1102. https://doi.org/10.1111/1755-0998.12324

Miller, S. R., Strong, A. L., Jones, K. L., & Ungerer, M. C. (2009). Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Applied and Environmental Microbiology*, *75*(13), 4565–4572. https://doi.org/10.1128/AEM.02792-08/SUPPL_FILE/TABLE_S3.XLS

Mirsepasi-Lauridsen, H. C., Vrankx, K., Engberg, J., Friis-Møller, A., Brynskov, J., Nordgaard-Lassen, I., Petersen, A. M., & Krogfelt, K. A. (2018). Disease-specific enteric microbiome dysbiosis in Inflammatory Bowel disease. *Frontiers in Medicine*, *5*(NOV), 304. https://doi.org/10.3389/FMED.2018.00304/BIBTEX

Miyata, N., Hayashi, Y., Hayashi, S., Sato, K., Hirai, Y., Yamamoto, H., & Sugano, K. (2019). Lipopolysaccharides From Non-Helicobacter pylori Gastric Bacteria Potently Stimulate Interleukin-8 Production in Gastric Epithelial Cells. *Clinical and Translational Gastroenterology*, *10*(3), 1–8. https://doi.org/10.14309/CTG.0000000000000024

Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. v., Reyes, J. A., Shah, S. A., LeLeiko, N., Snapper, S. B., Bousvaros, A., Korzenik, J., Sands, B. E., Xavier, R. J., & Huttenhower, C. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.*, *13*(9). https://doi.org/10.1186/gb-2012-13-9-r79

Mukhopadhya, I., Hansen, R., El-Omar, E. M., & Hold, G. L. (2012). IBD—what role do Proteobacteria play? *Nature Reviews Gastroenterology & Hepatology 2012 9:4*, *9*(4), 219–230. https://doi.org/10.1038/nrgastro.2012.14

Nagao, K., & Segre, J. A. (2015). "Bringing Up Baby" to Tolerate Germs. *Immunity*, *43*(5), 842–844. https://doi.org/10.1016/j.immuni.2015.10.020

Nakamura, M., Matsui, H., Serizawa, H., Murayama, S. Y., Yamaguchi, T., Takahashi, T., Matsumoto, T., Yamada, H., Hibi, T., & Tsuchimoto, K. (2006a). Coinfection of Helicobacter pylori and Neisseria subflava is closely associated with lymph follicle formation in human stomach. *Alimentary Pharmacology & Therapeutics*, *24*(SUPPL.4), 207–213. https://doi.org/10.1111/J.1365-2036.2006.00047.X

Nakamura, M., Matsui, H., Serizawa, H., Murayama, S. Y., Yamaguchi, T., Takahashi, T., Matsumoto, T., Yamada, H., Hibi, T., & Tsuchimoto, K. (2006b). Coinfection of Helicobacter pylori and Neisseria subflava is closely associated with lymph follicle formation in human stomach. *Alimentary Pharmacology & Therapeutics Symposium Series*, *2*(1), 207–213. https://doi.org/10.1111/J.1746-6342.2006.00047.X

Nardone, G., & Compare, D. (2015a). The human gastric microbiota: Is it time to rethink the pathogenesis of stomach diseases? *Http://Dx.Doi.Org/10.1177/2050640614566846*, *3*(3), 255–260. https://doi.org/10.1177/2050640614566846

Nardone, G., & Compare, D. (2015b). The human gastric microbiota: Is it time to rethink the pathogenesis of stomach diseases? *United European Gastroenterology Journal*, *3*(3), 255. https://doi.org/10.1177/2050640614566846

Ndegwa, N., Ploner, A., Andersson, A. F., Zagai, U., Andreasson, A., Vieth, M., Talley, N. J., Agreus, L., & Ye, W. (2020). Gastric microbiota in a low- helicobacter pylori prevalence general population and their associations with gastric lesions. *Clinical and Translational Gastroenterology*, *11*(7). https://doi.org/10.14309/CTG.0000000000000191

Nishijima, S., Kurokawa, I., Katoh, N., & Watanabe, K. (2000). The bacteriology of acne vulgaris and antimicrobial susceptibility of Propionibacterium acnes and Staphylococcus epidermidis isolated from acne lesions. *The Journal of Dermatology*, *27*(5), 318–323. https://doi.org/10.1111/J.1346-8138.2000.TB02174.X

Nitzan, O., Elias, M., Chazan, B., Raz, R., & Saliba, W. (2013). Clostridium difficile and inflammatory bowel disease: Role in pathogenesis and implications in treatment. *World Journal of Gastroenterology : WJG*, *19*(43), 7577. https://doi.org/10.3748/WJG.V19.I43.7577

Nobrega, F. L., Costa, A. R., Santos, J. F., Siliakus, M. F., van Lent, J. W. M., Kengen, S. W. M., Azeredo, J., & Kluskens, L. D. (2016). Genetically manipulated phages with improved pH resistance for oral administration in veterinary medicine. *Scientific Reports*, *6*. https://doi.org/10.1038/SREP39235

Noto, J. M., & Peek, R. M. (2017). The gastric microbiome, its interaction with Helicobacter pylori, and its potential role in the progression to stomach cancer. *PLOS Pathogens*, *13*(10), e1006573. https://doi.org/10.1371/JOURNAL.PPAT.1006573

Nyrén, P. (2007). The History of Pyrosequencing®. *Pyrosequencing Protocols*, 1–13. https://doi.org/10.1385/1-59745-377-3:1

Oh, J., Byrd, A. L., Park, M., Kong, H. H., & Segre, J. A. (2016). Temporal Stability of the Human Skin Microbiome. *Cell*, *165*(4), 854–866. https://doi.org/10.1016/J.CELL.2016.04.008

Org, E., Blum, Y., Kasela, S., Mehrabian, M., Kuusisto, J., Kangas, A. J., Soininen, P., Wang, Z., Ala-Korpela, M., Hazen, S. L., Laakso, M., & Lusis, A. J. (2017). Relationships between gut microbiota, plasma metabolites, and metabolic syndrome traits in the METSIM cohort. *Genome Biol.*, *18*(1), 1–14. https://doi.org/10.1186/s13059-017-1194-2

Palmas, V., Pisanu, S., Madau, V., Casula, E., Deledda, A., Cusano, R., Uva, P., Vascellari, S., Loviselli, A., Manzin, A., & Velluzzi, F. (2021). Gut microbiota markers associated with obesity and overweight in Italian adults. *Scientific Reports 2021 11:1*, *11*(1), 1–14. https://doi.org/10.1038/s41598-021-84928-w

Park, A. M., Omura, S., Fujita, M., Sato, F., & Tsunoda, I. (2017). Helicobacter pylori and gut microbiota in multiple sclerosis versus Alzheimer's disease: 10 pitfalls of microbiome studies. *Clinical and Experimental Neuroimmunology*, *8*(3), 215–232. https://doi.org/10.1111/CEN3.12401

Peng, L. L., Wang, Y., Zhu, F. L., Xu, W. D., Ji, X. L., & Ni, J. (2017). IL-23R mutation is associated with ulcerative colitis: A systemic review and meta-analysis. *Oncotarget*, *8*(3), 4849. https://doi.org/10.18632/ONCOTARGET.13607

Perez-Carrasco, V., Soriano-Lerma, A., Soriano, M., Gutiérrez-Fernández, J., & Garcia-Salcedo, J. A. (2021). Urinary Microbiome: Yin and Yang of the Urinary Tract. *Frontiers in Cellular and Infection Microbiology*, *11*, 421. https://doi.org/10.3389/FCIMB.2021.617002/BIBTEX

Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J. A., Bonazzi, V., McEwen, J. E., Wetterstrand, K. A., Deal, C., Baker, C. C., di Francesco, V., Howcroft,

T. K., Karp, R. W., Lunsford, R. D., Wellington, C. R., Belachew, T., Wright, M., Giblin, C., … Guyer, M. (2009). The NIH Human Microbiome Project. *Genome Research*, *19*(12), 2317–2323. https://doi.org/10.1101/GR.096651.109

Pham-Khanh, N. H., Sunahara, H., Yamadeya, H., Sakai, M., Nakayama, T., Yamamoto, H., Truong Thi Bich, V., Miyanaga, K., & Kamei, K. (2019). Isolation, Characterisation and Complete Genome Sequence of a Tequatrovirus Phage, Escherichia phage KIT03, Which Simultaneously Infects Escherichia coli O157:H7 and Salmonella enterica. *Current Microbiology*, *76*(10), 1130–1137. https://doi.org/10.1007/S00284-019-01738-0/FIGURES/3

Pierce, B. L., Biggs, M. L., Decambre, M., Reiner, A. P., Li, C., Fitzpatrick, A., Carlson, C. S., Stanford, J. L., & Austin, M. A. (2009). C-reactive protein, interleukin-6, and prostate cancer risk in men aged 65 years and older. *Cancer Causes & Control : CCC*, *20*(7), 1193. https://doi.org/10.1007/S10552-009-9320-4

Podolsky, S. H. (2012). Metchnikoff and the microbiome. *The Lancet*, *380*(9856), 1810–1811. https://doi.org/10.1016/S0140-6736(12)62018-2

Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A. H., Nieuwdorp, M., & Levin, E. (2020). Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLoS ONE*, *15*(1). https://doi.org/10.1371/journal.pone.0227434

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., … Zoetendal, E. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59–65. https://doi.org/10.1038/NATURE08821

Quivey, R. G., Kuhnert, W. L., & Hahn, K. (2000). Adaptation of oral streptococci to low pH. *Advances in Microbial Physiology*, *42*, 239–274. https://doi.org/10.1016/S0065-2911(00)42004-7

Rainey, K. (2016). *First DNA Sequencing in Space a Game Changer*. http://www.nasa.gov/mission_pages/station/research/news/dna_sequencing

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165. https://doi.org/10.1093/NAR/GKW257

Ranjan, R., Rani, A., Metwally, A., McGee, H. S., & Perkins, D. L. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical and Biophysical Research Communications*, *469*(4), 967. https://doi.org/10.1016/J.BBRC.2015.12.083

Ren, H. Y., Wen, L. S., Geng, Y. H., Huang, J. bin, Liu, J. F., Shen, D. Y., & Meng, J. R. (2019). Association between IL-1B gene polymorphisms and Helicobacter pylori

infection: A meta-analysis. *Microbial Pathogenesis*, *137*, 103769. https://doi.org/10.1016/J.MICPATH.2019.103769

Rigauts, C., Aizawa, J., Taylor, S. L., Rogers, G. B., Govaerts, M., Cos, P., Ostyn, L., Sims, S., Vandeplassche, E., Sze, M., Dondelinger, Y., Vereecke, L., van Acker, H., Simpson, J. L., Burr, L., Willems, A., Tunney, M. M., Cigana, C., Bragonzi, A., … Crabbé, A. (2022). Rothia mucilaginosa is an anti-inflammatory bacterium in the respiratory tract of patients with chronic lung disease. *European Respiratory Journal*, *59*(5), 11. https://doi.org/10.1183/13993003.01293-2021

Rodríguez-Medina, N., Barrios-Camacho, H., Duran-Bedolla, J., & Garza-Ramos, U. (2019). Klebsiella variicola: an emerging pathogen in humans. *Emerging Microbes & Infections*, *8*(1), 973. https://doi.org/10.1080/22221751.2019.1634981

Rogier, R., Ederveen, T. H. A., Boekhorst, J., Wopereis, H., Scher, J. U., Manasson, J., Frambach, S. J. C. M., Knol, J., Garssen, J., van der Kraan, P. M., Koenders, M. I., van den Berg, W. B., van Hijum, S. A. F. T., & Abdollahi-Roodsaz, S. (2017). Aberrant intestinal microbiota due to IL-1 receptor antagonist deficiency promotes IL-17- and TLR4-dependent arthritis. *Microbiome*, *5*(1), 63. https://doi.org/10.1186/S40168-017-0278-2/FIGURES/7

Ronaghi, M., Uhlén, M., & Nyrén, P. (1998). A sequencing method based on real-time pyrophosphate. *Science*, *281*(5375), 363–365. https://doi.org/10.1126/SCIENCE.281.5375.363

Rose-John, S. (2018). Interleukin-6 Family Cytokines. *Cold Spring Harbor Perspectives in Biology*, *10*(2). https://doi.org/10.1101/CSHPERSPECT.A028415

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P. I., Godneva, A., Kalka, I. N., Bar, N., Shilo, S., Lador, D., Vila, A. V., Zmora, N., Pevsner-Fischer, M., Israeli, D., Kosower, N., Malka, G., Wolf, B. C., … Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature 2018 555:7695*, *555*(7695), 210–215. https://doi.org/10.1038/nature25973

Roudsari, M. R., Karimi, R., Sohrabvandi, S., & Mortazavian, A. M. (2015). Health effects of probiotics on the skin. *Critical Reviews in Food Science and Nutrition*, *55*(9), 1219–1240. https://doi.org/10.1080/10408398.2012.680078

Royet, J., Gupta, D., & Dziarski, R. (2011). Peptidoglycan recognition proteins: modulators of the microbiome and inflammation. *Nature Reviews Immunology 2011 11:12*, *11*(12), 837–851. https://doi.org/10.1038/nri3089

Sanaii, A., Shirzad, H., Haghighian, M., Rahimian, G., Soltani, A., Shafigh, M., Tahmasbi, K., & Bagheri, N. (2019). Role of Th22 cells in Helicobacter pylori-related gastritis and peptic ulcer diseases. *Molecular Biology Reports*, *46*(6), 5703–5712. https://doi.org/10.1007/S11033-019-05004-1/FIGURES/9

Sánchez-Alcoholado, L., Ramos-Molina, B., Otero, A., Laborda-Illanes, A., Ordóñez, R., Medina, J. A., Gómez-Millán, J., & Queipo-Ortuño, M. I. (2020). The Role of the Gut Microbiome in Colorectal Cancer Development and Therapy Response. *Cancers 2020, Vol. 12, Page 1406*, *12*(6), 1406. https://doi.org/10.3390/CANCERS12061406

Sánchez-Maldonado, J. M., Collado, R., Cabrera-Serrano, A. J., Horst, R. ter, Gálvez-Montosa, F., Robles-Fernández, I., Arenas-Rodríguez, V., Cano-Gutiérrez, B., Bakker, O., Bravo-Fernández, M. I., García-Verdejo, F. J., López, J. A. L., Olivares-Ruiz, J., López-Nevot, M. Á., Fernández-Puerta, L., Cózar-Olmo, J. M., Li, Y., Netea, M. G., Jurado, M., … Sainz, J. (2022). Type 2 Diabetes-Related Variants Influence the Risk of Developing Prostate Cancer: A Population-Based Case-Control Study and Meta-Analysis. *Cancers*, *14*(10). https://doi.org/10.3390/CANCERS14102376/S1

Sankaran, K., & Holmes, S. (2019). Multitable methods for microbiome data integration. *Frontiers in Genetics*, *10*(JUN), 627. https://doi.org/10.3389/FGENE.2019.00627/BIBTEX

Savio, L. del, Prainsack, B., & Buyx, A. (2017). Motivations of participants in the citizen science of microbiomics: data from the British Gut Project. *Genetics in Medicine*, *19*(8), 959–961. https://doi.org/10.1038/GIM.2016.208

Scharschmidt, T. C. (2017). Establishing Tolerance to Commensal Skin Bacteria: Timing Is Everything. *Dermatologic Clinics*, *35*(1), 1–9. https://doi.org/10.1016/j.det.2016.07.007

Scharschmidt, T. C., Vasquez, K. S., Truong, H. A., Gearty, S. v., Pauli, M. L., Nosbaum, A., Gratz, I. K., Otto, M., Moon, J. J., Liese, J., Abbas, A. K., Fischbach, M. A., & Rosenblum, M. D. (2015). A Wave of Regulatory T Cells into Neonatal Skin Mediates Tolerance to Commensal Microbes. *Immunity*, *43*(5), 1011–1021. https://doi.org/10.1016/J.IMMUNI.2015.10.016

Schmaußer, B., Andrulis, M., Endrich, S., Müller-Hermelink, H. K., & Eck, M. (2005). Toll-like receptors TLR4, TLR5 and TLR9 on gastric carcinoma cells: An implication for interaction with Helicobacter pylori. *International Journal of Medical Microbiology*, *295*(3), 179–185. https://doi.org/10.1016/J.IJMM.2005.02.009

Schroner, Z., Javorský, M., Halušková, J., Klimčáková, L., Babjaková, E., Fabianová, M., Slabá, E., Kozárová, M., & Tkáč, I. (2012). Variation in CDKAL1 Gene Is Associated With Therapeutic Response to Sulphonylureas. *Physiol. Res*, *61*, 177–183. https://doi.org/10.33549/physiolres.932228

Shah, S., Locca, A., Dorsett, Y., Cantoni, C., Ghezzi, L., Lin, Q., Bokoliya, S., Panier, H., Suther, C., Gormley, M., Liu, Y., Evans, E., Mikesell, R., Obert, K., Salter, A., Cross, A. H., Tarr, P. I., Lovett-Racke, A., Piccio, L., & Zhou, Y. (2021). Alterations of the gut mycobiome in patients with MS. *EBioMedicine*, *71*, 103557. https://doi.org/10.1016/J.EBIOM.2021.103557

Shahini, A., & Shahini, A. (2022). Role of interleukin-6-mediated inflammation in the pathogenesis of inflammatory bowel disease: focus on the available therapeutic approaches and gut microbiome. *Journal of Cell Communication and Signaling 2022*, 1–20. https://doi.org/10.1007/S12079-022-00695-X

Shang, Q., Shi, J., Song, G., Zhang, M., Cai, C., Hao, J., Li, G., & Yu, G. (2016). Structural modulation of gut microbiota by chondroitin sulfate and its oligosaccharide. *International Journal of Biological Macromolecules*, *89*, 489–498. https://doi.org/10.1016/J.IJBIOMAC.2016.04.091

Sharma, N., Bhatia, S., Sodhi, A. S., & Batra, N. (2018). Oral microbiome and health. *AIMS Microbiology*, *4*(1), 42. https://doi.org/10.3934/MICROBIOL.2018.1.42

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: past, present and future. *Nature 2017 550:7676*, *550*(7676), 345–353. https://doi.org/10.1038/nature24286

Sheng, J., & Marquis, R. E. (2006). Enhanced acid resistance of oral streptococci at lethal pH values associated with acid-tolerant catabolism and with ATP synthase activity. *FEMS Microbiology Letters*, *262*(1), 93–98. https://doi.org/10.1111/J.1574-6968.2006.00374.X

Shin, N. R., Whon, T. W., & Bae, J. W. (2015). Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends in Biotechnology*, *33*(9), 496–503. https://doi.org/10.1016/J.TIBTECH.2015.06.011

Shrestha, P., Karmacharya, J., Han, S.-R., Lee, J. H., Park, H., & Oh, T.-J. (2022). Complete Genome Sequence and Comparative Genome Analysis of Variovorax sp. Strains PAMC28711, PAMC26660, and PAMC28562 and Trehalose Metabolic Pathways in Antarctica Isolates. *International Journal of Microbiology*, *2022*, 1–13. https://doi.org/10.1155/2022/5067074

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, *122*(1), e59. https://doi.org/10.1002/CPMB.59

Smith, R. P., Easson, C., Lyle, S. M., Kapoor, R., Donnelly, C. P., Davidson, E. J., Parikh, E., Lopez, J. v., & Tartar, J. L. (2019). Gut microbiome diversity is associated with sleep physiology in humans. *PloS One*, *14*(10). https://doi.org/10.1371/JOURNAL.PONE.0222394

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., & Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere." *Proceedings of the National Academy of Sciences of the United States of America*, *103*(32), 12115–12120. https://doi.org/10.1073/PNAS.0605127103

Son, B., Kong, M., & Ryu, S. (2018). The Auxiliary Role of the Amidase Domain in Cell Wall Binding and Exolytic Activity of Staphylococcal Phage Endolysins. *Viruses*, *10*(6). https://doi.org/10.3390/V10060284

Song, H., Kerfahi, D., Takahashi, K., Nixon, S. L., Tripathi, B. M., Kim, H., Tateno, R., & Adams, J. (2020). Metagenomic analysis reveals rapid development of soil biota on fresh volcanic ash. *Scientific Reports 2020 10:1*, *10*(1), 1–15. https://doi.org/10.1038/s41598-020-78413-z

Sonnenberg, A., Melton, S. D., & Genta, R. M. (2011). Frequent occurrence of gastritis and duodenitis in patients with inflammatory bowel disease. *Inflammatory Bowel Diseases*, *17*(1), 39–44. https://doi.org/10.1002/IBD.21356

Spellerberg, I. F., & Fedor, P. J. (2003). A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener' Index. *Global Ecology and Biogeography*, *12*(3), 177–179. https://doi.org/10.1046/J.1466-822X.2003.00015.X

Spichler, A., Hurwitz, B. L., Armstrong, D. G., & Lipsky, B. A. (2015). Microbiology of diabetic foot infections: From Louis Pasteur to "crime scene investigation." *BMC Medicine*, *13*(1), 1–13. https://doi.org/10.1186/S12916-014-0232-0/FIGURES/4

Ssekatawa, K., Byarugaba, D. K., Nakavuma, J. L., Kato, C. D., Ejobi, F., Tweyongyere, R., & Eddie, W. M. (2021). Prevalence of pathogenic Klebsiella pneumoniae based on PCR capsular typing harbouring carbapenemases encoding genes in Uganda tertiary hospitals. *Antimicrobial Resistance and Infection Control*, *10*(1), 1–10. https://doi.org/10.1186/S13756-021-00923-W/TABLES/6

Sun, N., Meng, F., Zhao, J., Li, X., Li, R., Han, J., Chen, X., Cheng, W., Yang, X., Kou, Y., Zheng, K., Yang, J., & Ikezoe, T. (2021). Aurka deficiency in the intestinal epithelium promotes age-induced obesity via propionate-mediated AKT activation. *International Journal of Biological Sciences*, *17*(5), 1302. https://doi.org/10.7150/IJBS.56477

Sun, Z. M., Zhao, J., Qian, P., Wang, Y. Q., Zhang, W. F., Guo, C. R., Pang, X. Y., Wang, S. C., Li, F. F., & Li, Q. (2013). Metabolic markers and microecological characteristics of tongue coating in patients with chronic gastritis. *BMC Complementary and Alternative Medicine*, *13*(1), 1–10. https://doi.org/10.1186/1472-6882-13-227/FIGURES/7

Sung, J., Kim, N., Lee, J., Hwang, Y. J., Kim, H. W., Chung, J. W., Kim, J. W., & Lee, D. H. (2018). Associations among Gastric Juice pH, Atrophic Gastritis, Intestinal Metaplasia and Helicobacter pylori Infection. *Gut and Liver*, *12*(2), 158. https://doi.org/10.5009/GNL17063

Surana, N. K., & Kasper, D. L. (2014). Deciphering the tête-à-tête between the microbiota and the immune system. *The Journal of Clinical Investigation*, *124*(10), 4197. https://doi.org/10.1172/JCI72332

Șurlin, P., Nicolae, F. M., Șurlin, V. M., Pătrașcu, Ștefan, Ungureanu, B. S., Didilescu, A. C., & Gheonea, D. I. (2020). Could Periodontal Disease through Periopathogen Fusobacterium nucleatum Be an Aggravating Factor for Gastric Cancer? *Journal of*

Clinical Medicine 2020, Vol. 9, Page 3885, *9*(12), 3885. https://doi.org/10.3390/JCM9123885

Swarte, J. C., Douwes, R. M., Hu, S., Vila, A. V., Eisenga, M. F., van Londen, M., Gomes-Neto, A. W., Weersma, R. K., Harmsen, H. J. M., & Bakker, S. J. L. (2020). Characteristics and Dysbiosis of the Gut Microbiome in Renal Transplant Recipients. *Journal of Clinical Medicine 2020, Vol. 9, Page 386*, *9*(2), 386. https://doi.org/10.3390/JCM9020386

Swarte, J. C., Li, Y., Hu, S., Björk, J. R., Gacesa, R., Vich Vila, A., Douwes, R. M., Collij, V., Kurilshikov, A., Post, A., Klaassen, M. A. Y., Eisenga, M. F., Gomes-Neto, A. W., Kremer, D., Jansen, B. H., Knobbe, T. J., Berger, S. P., Sanders, J. S. F., Heiner-Fokkema, M. R., … Weersma, R. K. (2022). Gut microbiome dysbiosis is associated with increased mortality after solid organ transplantation. *Science Translational Medicine*, *14*(660), eabn7566. https://doi.org/10.1126/SCITRANSLMED.ABN7566/SUPPL_FILE/SCITRANSLMED.ABN7566_MDAR_REPRODUCIBILITY_CHECKLIST.PDF

Szymczak, A., Ferenc, S., Majewska, J., Miernikiewicz, P., Gnus, J., Witkiewicz, W., & Dąbrowska, K. (2020). Application of 16S rRNA gene sequencing in Helicobacter pylori detection. *PeerJ*, *8*(3). https://doi.org/10.7717/PEERJ.9099

Takeuchi, O., Sato, S., Horiuchi, T., Hoshino, K., Takeda, K., Dong, Z., Modlin, R. L., & Akira, S. (2002). Cutting edge: role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. *Journal of Immunology (Baltimore, Md. : 1950)*, *169*(1), 10–14. https://doi.org/10.4049/JIMMUNOL.169.1.10

Tang, C., Chen, S., Qian, H., & Huang, W. (2012). Interleukin-23: as a drug target for autoimmune inflammatory diseases. *Immunology*, *135*(2), 112. https://doi.org/10.1111/J.1365-2567.2011.03522.X

Texereau, J., Chiche, J.-D., Taylor, W., Choukroun, G., Comba, B., & Mira, J.-P. (2005). The Importance of Toll-Like Receptor 2 Polymorphisms in Severe Infections. *Clinical Infectious Diseases*, *41*, 408–423.

Tiew, P. Y., mac Aogain, M., Ali, N. A. B. M., Thng, K. X., Goh, K., Lau, K. J. X., & Chotirmall, S. H. (2020). The Mycobiome in Health and Disease: Emerging Concepts, Methodologies and Challenges. *Mycopathologia*, *185*(2), 207. https://doi.org/10.1007/S11046-019-00413-Z

Uberti, A. F., Olivera-Severo, D., Wassermann, G. E., Scopel-Guerra, A., Moraes, J. A., Barcellos-de-Souza, P., Barja-Fidalgo, C., & Carlini, C. R. (2013). Pro-inflammatory properties and neutrophil activation by Helicobacter pylori urease. *Toxicon : Official Journal of the International Society on Toxinology*, *69*, 240–249. https://doi.org/10.1016/J.TOXICON.2013.02.009

Ursell, L. K., Metcalf, J. L., Parfrey, L. W., & Knight, R. (2012). Defining the Human Microbiome. *Nutrition Reviews*, *70*(Suppl 1), S38. https://doi.org/10.1111/J.1753-4887.2012.00493.X

van Belleghem, J. D., Clement, F., Merabishvili, M., Lavigne, R., & Vaneechoutte, M. (2017). Pro- and anti-inflammatory responses of peripheral blood mononuclear cells induced by Staphylococcus aureus and Pseudomonas aeruginosa phages. *Scientific Reports*, *7*(1). https://doi.org/10.1038/S41598-017-08336-9

van Tilburg Bernardes, E., Pettersen, V. K., Gutierrez, M. W., Laforest-Lapointe, I., Jendzjowsky, N. G., Cavin, J. B., Vicentini, F. A., Keenan, C. M., Ramay, H. R., Samara, J., MacNaughton, W. K., Wilson, R. J. A., Kelly, M. M., McCoy, K. D., Sharkey, K. A., & Arrieta, M. C. (2020). Intestinal fungi are causally implicated in microbiome assembly and immune development in mice. *Nature Communications 2020 11:1*, *11*(1), 1–16. https://doi.org/10.1038/s41467-020-16431-1

Vasileiadis, S., Puglisi, E., Arena, M., Cappa, F., Cocconcelli, P. S., & Trevisan, M. (2012). Correction: Soil Bacterial Diversity Screening Using Single 16S rRNA Gene V Regions Coupled with Multi-Million Read Generating Sequencing Technologies. *PLoS ONE*, *7*(8), 42671. https://doi.org/10.1371/ANNOTATION/BF36B938-523B-4C9F-AE86-16DB7E8D39C1

Viala, J., Chaput, C., Boneca, I. G., Cardona, A., Girardin, S. E., Moran, A. P., Athman, R., Mémet, S., Huerre, M. R., Coyle, A. J., DiStefano, P. S., Sansonetti, P. J., Labigne, A., Bertin, J., Philpott, D. J., & Ferrero, R. L. (2004). Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island. *Nature Immunology 2004 5:11*, *5*(11), 1166–1174. https://doi.org/10.1038/ni1131

Vinturache, A. E., Gyamfi-Bannerman, C., Hwang, J., Mysorekar, I. U., & Jacobsson, B. (2016). *Maternal microbiome e A pathway to preterm birth*. https://doi.org/10.1016/j.siny.2016.02.004

Virgin, H. W. (2014). The Virome in Mammalian Physiology and Disease. *Cell*, *157*(1), 142–150. https://doi.org/10.1016/J.CELL.2014.02.032

Waldum, H., & Fossmark, R. (2021). Gastritis, Gastric Polyps and Gastric Cancer. *International Journal of Molecular Sciences*, *22*(12). https://doi.org/10.3390/IJMS22126548

Wang, D., Zhang, T., Lu, Y., Wang, C., Wu, Y., Li, J., Tao, Y., Deng, L., Zhang, X., & Ma, J. (2022). Helicobacter pylori infection affects the human gastric microbiome, as revealed by metagenomic sequencing. *FEBS Open Bio*, *12*(6), 1188–1196. https://doi.org/10.1002/2211-5463.13390

Wang, J., Chen, W. D., & Wang, Y. D. (2020). The Relationship Between Gut Microbiota and Inflammatory Diseases: The Role of Macrophages. *Frontiers in Microbiology*, *11*, 1065. https://doi.org/10.3389/FMICB.2020.01065/BIBTEX

Wang, J., Thingholm, L. B., Skiecevičie, J., Rausch, P., Kummen, M., Hov, J. R., Degenhardt, F., Heinsen, F. A., Rühlemann, M. C., Szymczak, S., Holm, K., Esko, T., Sun, J., Pricop-Jeckstadt, M., Al-Dury, S., Bohov, P., Bethune, J., Sommer, F., Ellinghaus, D., … Franke, A. (2016). Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nature Genetics 2016 48:11*, *48*(11), 1396–1406. https://doi.org/10.1038/ng.3695

Wang, S., Kuang, J., Zhang, H., Chen, W., Zheng, X., Wang, J., Huang, F., Ge, K., Li, M., Zhao, M., Rajani, C., Zhu, J., Zhao, A., & Jia, W. (2022). Bile Acid–Microbiome Interaction Promotes Gastric Carcinogenesis. *Advanced Science*, *9*(16). https://doi.org/10.1002/ADVS.202200263

Wang, X., Ota, N., Manzanillo, P., Kates, L., Zavala-Solorio, J., Eidenschenk, C., Zhang, J., Lesch, J., Lee, W. P., Ross, J., Diehl, L., van Bruggen, N., Kolumam, G., & Ouyang, W. (2014). Interleukin-22 alleviates metabolic disorders and restores mucosal immunity in diabetes. *Nature 2014 514:7521*, *514*(7521), 237–241. https://doi.org/10.1038/nature13564

Wang, Y., Wen, Z., Shen, J., Cheng, W., Li, J., Qin, X., Ma, D., & Shi, Y. (2014). Comparison of the performance of Ion Torrent chips in noninvasive prenatal trisomy detection. *Journal of Human Genetics*, *59*(7), 393–396. https://doi.org/10.1038/JHG.2014.40

Wang, Y., Yang, Q., & Wang, Z. (2014). The evolution of nanopore sequencing. *Frontiers in Genetics*, *5*(DEC). https://doi.org/10.3389/FGENE.2014.00449

Wang, Y., Zhang, M. M., Huang, W. W., Wu, S. Q., Wang, M. G., Tang, X. Y., Sandford, A. J., & He, J. Q. (2018). Polymorphisms in Toll-Like Receptor 10 and Tuberculosis Susceptibility: Evidence from Three Independent Series. *Frontiers in Immunology*, *9*(FEB). https://doi.org/10.3389/FIMMU.2018.00309

Wang, Z., Gao, X., Zeng, R., Wu, Q., Sun, H., Wu, W., Zhang, X., Sun, G., Yan, B., Wu, L., Ren, R., Guo, M., Peng, L., & Yang, Y. (2020). Changes of the Gastric Mucosal Microbiome Associated With Histological Stages of Gastric Carcinogenesis. *Frontiers in Microbiology*, *11*, 997. https://doi.org/10.3389/FMICB.2020.00997/BIBTEX

Waskito, L. A., Rezkitha, Y. A. A., Vilaichone, R. korn, Sugihartono, T., Mustika, S., Dewa Nyoman Wibawa, I., Yamaoka, Y., & Miftahussurur, M. (2022). The role of non-Helicobacter pylori bacteria in the pathogenesis of gastroduodenal diseases. *Gut Pathogens*, *14*(1). https://doi.org/10.1186/S13099-022-00494-0

Watari, J., Chen, N., Amenta, P. S., Fukui, H., Oshima, T., Tomita, T., Miwa, H., Lim, K. J., & Das, K. M. (2014). Helicobacter pylori associated chronic gastritis, clinical syndromes, precancerous lesions, and pathogenesis of gastric cancer development. *World Journal of Gastroenterology : WJG*, *20*(18), 5461. https://doi.org/10.3748/WJG.V20.I18.5461

Wei, L., Li, Y., & Suo, Z. (2015). TSPAN8 promotes gastric cancer growth and metastasis via ERK MAPK pathway. *International Journal of Clinical and Experimental Medicine*, *8*(6), 8599. /pmc/articles/PMC4537951/

West, T. E., Chantratita, N., Chierakul, W., Limmathurotsakul, D., Wuthiekanun, V., Myers, N. D., Emond, M. J., Wurfel, M. M., Hawn, T. R., Peacock, S. J., & Skerrett, S. J. (2013). Impaired TLR5 functionality is associated with survival in melioidosis. *Journal of Immunology (Baltimore, Md. : 1950)*, *190*(7), 3373–3379. https://doi.org/10.4049/JIMMUNOL.1202974

Whitfeld, M., Gunasingam, N., Leow, L. J., Shirato, K., & Preda, V. (2011). Staphylococcus epidermidis: a possible role in the pustules of rosacea. *Journal of the American Academy of Dermatology*, *64*(1), 49–52. https://doi.org/10.1016/J.JAAD.2009.12.036

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(11), 5088–5090. https://doi.org/10.1073/PNAS.74.11.5088

Wu, J., Liu, B., Cheng, F., Ramchiary, N., Choi, S. R., Lim, Y. P., & Wang, X. W. (2012). Sequencing of chloroplast genome using whole cellular DNA and solexa sequencing technology. *Frontiers in Plant Science*, *3*(NOV). https://doi.org/10.3389/FPLS.2012.00243

Wu, S., Zhang, Y., Ma, J., Liu, Y., Li, W., Wang, T., Xu, X., Wang, Y., Cheng, K., & Zhuang, R. (2022). Interleukin-6 absence triggers intestinal microbiota dysbiosis and mucosal immunity in mice. *Cytokine*, *153*, 155841. https://doi.org/10.1016/J.CYTO.2022.155841

Xu, P., Shao, R. R., Zhang, S., Tan, Z. W., Guo, Y. T., & He, Y. (2022). The mechanism on Prevotella melaninogenica promoting the inflammatory progression of oral lichen planus. *Clinical and Experimental Immunology*, *209*(2), 215–224. https://doi.org/10.1093/CEI/UXAC054

Xu, Y., Zhao, Z., Tong, W., Ding, Y., Liu, B., Shi, Y., Wang, J., Sun, S., Liu, M., Wang, Y., Qi, Q., Xian, M., & Zhao, G. (2020). An acid-tolerance response system protecting exponentially growing Escherichia coli. *Nature Communications 2020 11:1*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-15350-5

Xue, C. X., Liu, J., Lea-Smith, D. J., Rowley, G., Lin, H., Zheng, Y., Zhu, X. Y., Liang, J., Ahmad, W., Todd, J. D., & Zhang, X. H. (2020). Insights into the Vertical Stratification of Microbial Ecological Roles across the Deepest Seawater Column on Earth. *Microorganisms*, *8*(9), 1–21. https://doi.org/10.3390/MICROORGANISMS8091309

Yang, I., Nell, S., & Suerbaum, S. (2013). Survival in hostile territory: the microbiota of the stomach. *FEMS Microbiology Reviews*, *37*(5), 736–761. https://doi.org/10.1111/1574-6976.12027

Yang, J., & Yan, H. (2017). TLR5: beyond the recognition of flagellin. *Cellular & Molecular Immunology 2017 14:12*, *14*(12), 1017–1019. https://doi.org/10.1038/cmi.2017.122

Yang, L., Zhang, J., Xu, J., Wei, X., Yang, J., Liu, Y., Li, H., Zhao, C., Wang, Y., Zhang, L., & Gai, Z. (2019). Helicobacter pylori Infection Aggravates Dysbiosis of Gut Microbiome in Children With Gastritis. *Frontiers in Cellular and Infection Microbiology*, *9*, 375. https://doi.org/10.3389/FCIMB.2019.00375/BIBTEX

Yang, L., Zhao, Y., Huang, J., Zhang, H., Lin, Q., Han, L., Liu, J., Wang, J., & Liu, H. (2020). Insoluble dietary fiber from soy hulls regulates the gut microbiota in vitro and increases the abundance of bifidobacteriales and lactobacillales. *Journal of Food Science and Technology*, *57*(1), 152–162. https://doi.org/10.1007/S13197-019-04041-9/FIGURES/3

Yang, Y., Ji, R., Zhao, X., Cao, X., Wang, Q., Jiang, Q., Zhang, Y., Zheng, W., Wu, X., & Yang, A. (2021). Alterations in Gastric Mucosal Microbiota in Gastric Carcinogenesis: A Systematic Review and Meta-Analysis. *Frontiers in Medicine*, *8*, 2487. https://doi.org/10.3389/FMED.2021.754959/BIBTEX

Yinhang, W., Wei, W., Jing, Z., Qing, Z., Yani, Z., Yangyanqiu, W., & Shuwen, H. (2022). Biological roles of toll-like receptors and gut microbiota in colorectal cancer. *Https://Doi.Org/10.2217/Fmb-2021-0072*, *17*(13), 1071–1089. https://doi.org/10.2217/FMB-2021-0072

Young, G. M., Amid, D., & Miller, V. L. (1996). A bifunctional urease enhances survival of pathogenic Yersinia enterocolitica and Morganella morganii at low pH. *Journal of Bacteriology*, *178*(22), 6487–6495. https://doi.org/10.1128/JB.178.22.6487-6495.1996

Zandi, F., Shirzad, H., Bagheri, N., Rahimian, G., Salimzadeh, L., Azadegan, F., Yousefzadeh Eshkevari, K., Fatahi, F., Ahmadi, A., Gharib, A., Gholami, S., & Zamanzad, B. (2014). Evaluation of H.pylori infection and IL23R gene polymorphism in dyspeptic subjects. *Life Science Journal*, *11*(SPEC.ISS.2), 40–46.

Zaparucha, A., de Berardinis, V., & Vaxelaire-Vergne, C. (2018). Chapter 1 Genome Mining for Enzyme Discovery. *RSC Catalysis Series*, *2018-January*(32), 1–27. https://doi.org/10.1039/9781788010450-00001

Zárate, S., Taboada, B., Yocupicio-Monroy, M., & Arias, C. F. (2017). Human Virome. *Archives of Medical Research*, *48*(8), 701–716. https://doi.org/10.1016/J.ARCMED.2018.01.005

Zha, L., Garrett, S., & Sun, J. (2019). Salmonella Infection in Chronic Inflammation and Gastrointestinal Cancer. *Diseases*, *7*(1), 28. https://doi.org/10.3390/DISEASES7010028

Zhang, C., Zhang, M., Wang, S., Han, R., Cao, Y., Hua, W., Mao, Y., Zhang, X., Pang, X., Wei, C., Zhao, G., Chen, Y., & Zhao, L. (2009). Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. *The ISME Journal 2010 4:2*, *4*(2), 232–241. https://doi.org/10.1038/ismej.2009.112

Zhang, P., West, N. P., Chen, P. Y., Thang, M. W. C., Price, G., Cripps, A. W., & Cox, A. J. (2019). Selection of microbial biomarkers with genetic algorithm and principal

component analysis. *BMC Bioinformatics*, *20*(Suppl 6). https://doi.org/10.1186/S12859-019-3001-4

Zhang, Z., Long, Q., & Xie, J. (2012). Roles of Peptidoglycan Recognition Protein (PGRP) in Immunity and Implications for Novel Anti-infective Measures. *Critical Reviews™ in Eukaryotic Gene Expression*, *22*(3), 259–268. https://doi.org/10.1615/CRITREVEUKARGENEEXPR.V22.I3.90

Zhou, Y., Li, L., Han, K., Wang, L., Cao, Y., Ma, D., & Wang, X. (2022). A Polyvalent Broad-Spectrum Escherichia Phage Tequatrovirus EP01 Capable of Controlling Salmonella and Escherichia coli Contamination in Foods. *Viruses 2022, Vol. 14, Page 286*, *14*(2), 286. https://doi.org/10.3390/V14020286

Zuo, T., & Ng, S. C. (2018). The Gut Microbiota in the Pathogenesis and Therapeutics of Inflammatory bowel disease. *Frontiers in Microbiology*, *9*(SEP), 2247. https://doi.org/10.3389/FMICB.2018.02247/BIBTEX