ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
INSTITUTE OF HEALTH SCIENCES

ESTABLISHING A HUMAN GUT MICROBIOME BIOINFORMATICS ANALYSIS PIPELINE TO STUDY A DIETARY TREATMENT RESPONSE FOR METHYLMALONIC ACIDEMIA PATIENTS

BERKAY YEKTA EKREN

M.Sc. THESIS

DEPARTMENT OF BIOSTATISTICS AND BIOINFORMATICS

SUPERVISOR
Prof. Dr. O. Uğur Sezerman

ISTANBUL-2023

ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
INSTITUTE OF HEALTH SCIENCES

ESTABLISHING A HUMAN GUT MICROBIOME BIOINFORMATICS ANALYSIS PIPELINE TO STUDY A DIETARY TREATMENT RESPONSE FOR METHYLMALONIC ACIDEMIA PATIENTS

BERKAY YEKTA EKREN

M.Sc. THESIS

DEPARTMENT OF BIOSTATISTICS AND BIOINFORMATICS

SUPERVISOR
Prof. Dr. O. Uğur Sezerman

ISTANBUL-2023

DECLARATION

I declare that this thesis work is my own work, I had no unethical behavior at any stages from the planning to the writing of the thesis, I obtained all the information in this thesis in accordance with academic and ethical rules, I cited all the information and comments that were not obtained with this thesis work, and I provided resources in the list of references. I also declare that there was no violation of any patents and copyrights during the study and writing of this thesis.

02/06/2023

Berkay Yekta Ekren

PREFACE AND ACKNOWLEDGEMENT

I would like to express my first and foremost thanks to my supervisor, Prof. Dr. O. Uğur Sezerman, who helped me in every step of this thesis. He introduced me to the world of bioinformatics and biostatistics and inspired me to do better. He has become my idol, showing me his guidance, support and utmost patience as my teacher. Furthermore, I would like to thank my jury members and substitute jury members, Prof. Dr. Emel Timuçin, Dr. Engin Köse, Prof. Dr. Eda Turanlı and Prof. Dr. Emirhan Nemutlu for their numerous contributions.

Moreover, I would like to thank Orhan Özcan, PhD., who helped me at every step of the bioinformatics pipeline creation. Every single time I asked a question, without mistake, he would answer and explain the process in detail.

It has been my honor working with my colleagues at Sezerman Lab and Epigenetiks Inc. and I thank them deeply for their help and moral support along the way.

Last but not least, my family and friends supported me with all their heart during my thesis study. Only with their incredible moral support that I was able to commit to my thesis.

TABLE OF CONTENTS

DECLARATION iii

PREFACE AND ACKNOWLEDGEMENT iv

TABLE OF CONTENTS v

LIST OF FIGURES vii

LIST OF TABLES ix

ÖZET 1

ABSTRACT 2

1 INTRODUCTION AND AIM 3

1.1 Organic Acidemias 3
1.1.1 Propionic acidemia 3
1.1.2 Methylmalonic acidemia 4
1.2 Human Gut Metagenome Sequencing And *in silico* Analyses 7

2 BACKGROUND 10

2.1 Bacteria 10
2.2 Propionate 11
2.3 Next Generation Sequencing (NGS) - Generations and Techniques 13
2.3.1 Generations 13
2.3.2 Techniques 14
2.3.2.1 Whole genome shotgun sequencing (WGS) 14
2.3.2.2 Whole exome sequencing (WES) 14
2.3.2.3 Native RNA sequencing (RNA-Seq) 15
2.3.2.4 Methylation sequencing 15
2.3.2.5 Chip sequencing (ChIP-Seq) 15
2.3.2.6 Targeted DNA sequencing 16
2.3.2.6.1 Identification of enteric microbial community 17
2.3.2.6.2 16S rRNA gene databases 19
2.4 Bioinformatics Analyses For 16S Microbiome Data 19
2.4.1 Long reads 20
2.4.1.1 Basecallers 20
2.4.1.2 FastQC 20
2.4.1.3 BBTools 20
2.4.1.4 Magic-BLAST 21
2.4.1.5 SAMtools 21
2.4.1.6 BLAST+ 21
2.4.2 Short reads 21

3 MATERIALS AND METHODS 23

3.1 Sample Collection 23
3.2 Microbiome Sequencing 24
3.2.1 gDNA isolation 24
3.2.2 Polymerase Chain Reaction (PCR) 24
3.2.3 16S targeted amplicon sequencing 25
3.3 Bioinformatics And Statistical Analyses 26
3.3.1 Bioinformatics analysis 26
3.3.2 Biostatistics analysis 27
3.3.2.1 Data imputation 27
3.3.2.2 Hypothesis Tests 28
3.3.2.2.1 Alpha value 28
3.3.2.2.2 Shapiro-Wilk normality test 28
3.3.2.2.3 T hypothesis test 28
3.3.2.2.4 Wilcoxon hypothesis test 29
3.3.2.2.5 ANOVA hypothesis test 29
3.3.2.2.6 Kruskal-Wallis hypothesis test 30
3.3.2.3 P-value adjustment 30
3.3.2.4 Regression 30
3.3.2.4.1 Linear regression 31
3.3.2.4.2 Nonlinear regression 31
3.3.3 Diversity analysis 32
3.3.3.1 Alpha diversity 32
3.3.3.2 Beta diversity 33

4 RESULTS 34

4.1 Phyla Level OTU Calculation Results 34
4.2 Species Level OTU Calculation Results 36
4.3 OTUs Phyla & Species Level Alpha - Beta Diversity Calculations 37
4.4 Species Level Propionate Producing OTUs 38
4.5 Linear Regression Results 39

5 DISCUSSION 45

6 CONCLUSION 48

7 REFERENCES 50

8 APPENDIX 61

9 CURRICULUM VITAE 69

LIST OF FIGURES

Figure 1. Propanoate or Succinyl-CoA Production pathway. One of the metabolic pathways, the succinyl-CoA production pathway is involved in the main energy production metabolism. Also, energy production from some amino acids and fatty acids can only be carried out if gut bacterial community derived propionate is converted into succinyl-CoA for its participation in TCA cycle (also known as Krebs cycle or citric acid cycle) 6
Figure 2. Taxonomic list starting from domain (superkingdom) and following with more detail, kingdom, phylum, class, order, family, genus and species. A group of organisms belonging to the same species will belong to the same genus while organisms which are under the same genus will be under the same family as the taxonomic classification..... 11
Figure 3. The created single-end long-read Oxford Nanopore Microbiome bioinformatics pipeline for the methylmalonic acidemia patients. The pipeline starts with the fast5 Nanopore signal files and results with OTU files 27
Figure 4. (a) Stacked bar plot of all phyla and their relative abundances. Phyla with relative abundance more than 1% are depicted in the plot above. The rest of the phyla relative abundances were added together and labelled as “Others” and marked with “grey”. Enteric microbiota composition of each sample and the averages of each group was shown in the plot. (b) The distribution of phyla in the cohort in the separate treatment steps. In total, 30 different phyla were found, and while some are shared between the treatment steps, some are observed to be belonging to a specific treatment stage. The treatment step 1 is depicted with S1 (orange) while S2 represents Step 2 (green), Step3 is shown as S3 (purple) and, lastly, S4 is the treatment step 4 (pink). 35
Figure 5. (a) OTUs at species level with relative abundance (%) values. The OTUs which had relative abundance values lower than 1% were summed and depicted under the name “Others”. (b) The distribution of species in the cohort in the separate treatment steps. In total, 2744 different species were found, and while some are shared between the treatment steps, some are observed to be belonging to a specific treatment stage. The treatment step 1 is depicted with S1 (orange) while S2

represents Step 2 (green), Step3 is shown as S3 (purple) and, lastly, S4 is the treatment step 4 (pink). 36
Figure 6. The figure above shows the alpha and beta diversity visualizations with box and PCoA plots comparing the first step of the treatment and the latter steps. While S1 represents the first step of the treatment (orange), S2, the second step (green), S3, the third (purple), and, finally, S4 indicates the fourth treatment step (pink). 38
Figure 7. Detected propionate producing bacteria with relative abundance values in a decreasing pattern are given. The S1 (orange) stands for step 1 of the treatment, S2 (green) depicts the treatment step 2 while S3 (purple) is the step 3 and S4 (pink) represents the last step of the treatment duration 39
Figure 8. Correlation plot for best significant OTUs at genus (a) and species (b) level. Significant OTUs for the first 3 treatment steps were incorporated while calculating the correlations. Dark colors (dark blue or dark red) represent high correlation values along with the circles with a large size 40

LIST OF TABLES

Table 1: Bacteria found in the human gut microbiota and known to produce propionate as metabolites……..…….……..……...12
Table 2: Sample information. 8 participants with MMA were included in the study with varying childhood ages, between 1.5 and 13.5……...23
Table 3: Treatment steps with protein containing dietary supplements ……24
Table 4: 16S targeted PCR conditions. After an initial denaturation at 95°C for 1 minute, with 25 PCR cycles, denaturation at 95°C for 20 seconds, annealing at 55°C for 30 seconds, and extension at 65°C for 2 minutes was carried out. Lastly, the final extension was at 65°C for 5 minutes……..……25
Table 5: Linear regression results for the methylmalonic acid contents of patients' urine and relative abundance of OTUs at the genus level. The coefficients show the contribution of the changes in the relative abundance of the respective bacteria. Positive values indicate that methylmalonic acid levels in urea increase as the respective bacteria is increasing while negative values show that appetite levels and the respective bacteria are negatively correlated. The p-value shows the importance of the respective bacteria while the significance column indicates the importance level of the bacteria in the created model. .: <0.1 not significant *: <0.05/significant; **: <0.01/very significant; ***: <0.001/very very significant…..……42
Table 6: Linear regression results for the methylmalonic acid contents of patients' urine and relative abundance of OTUs at the species level. The coefficients show the contribution of the changes in the relative abundance of the respective bacteria. Positive values indicate that methylmalonic acid in urea increases as the respective bacteria is increasing while negative values show that methylmalonic acid and the respective bacteria are negatively correlated. The p-value shows the importance of the respective bacteria while the significance column indicates the importance level of the bacteria in the created model. *: <0.05/significant; **: <0.01/very significant; ***: <0.001/very very significant……..……43
Table 7: Linear regression results for the appetites of patients and relative abundance of OTUs at the genus level. The coefficients show the contribution of the changes in the relative abundance of the respective bacteria. Positive values indicate that

appetite levels increase as the respective bacteria is increasing while negative values show that appetite levels and the respective bacteria are negatively correlated. The p-value shows the importance of the respective bacteria while the significance column indicates the importance level of the bacteria in the created model. *: <0.05/significant; **: <0.01/very significant; ***: <0.001/very very significant…….....44

ÖZET

Metilmalonik Asidemi Hastalarında Diyet Tedavisi Yanıtını İncelemek için İnsan Bağırsak Mikrobiyom Biyoinformatik Analiz Akışının Oluşturulması

İnsan bağırsağı, burada barınan mikroorganizmalar aracılığıyla vücut fonksiyonlarında çeşitli temel roller üstlenerek insan vücudundaki en önemli yerlerden biri olarak bilinmektedir. 6'dan az karbon atomuna sahip kısa zincirli yağ asitlerinin (SCFA'lar), enerji üretim yolundaki temel bileşenler olduğu bilinmektedir. Belirli amino asitlerin ve tek zincirli yağ asitlerinin katabolizması ve propiyonil-CoA'nın sentezlenmesi için propionat temel bir bileşendir. Metilmalonil-CoA mutaz, metilmalonil-CoA epimeraz ve kobalamin yolağı genlerindeki belirli patojenik mutasyonlar nedeniyle insan vücudu, biriktikten sonra metilmalonik asidemiye yol açan toksik ara maddeler üretebilir. Yalnızca insan bağırsağı bakterileri tarafından üretilen SCFA'lar, özellikle propiyonat, insan kalın bağırsağında üretilen mikrobiyal metabolitlerdir ve bunların üretimini teşvik etmenin sayısız sağlık yararı vardır. Sağlıklı insanların aksine, organik asidemi hastalarında SCFA'ların artması, zararlı semptomlara yol açabilen toksik agregatlara neden olur. 16S hedefli NGS tekniği kullanılarak, enterik mikrobiyota bileşimi belirlenebilir ve tam uzunlukta 16S bölge dizilimi ile Oxford Nanopore Technologies kullanılarak bakterilerin tür düzeyinde tanımlanması başarıyla gerçekleştirilebilir. Bu tezin amacı, diyet tedavisi ile insan bağırsaklarında propiyonat üreten bakterilerin azalmasını gözlemlemektir. Bu çalışmaya dahil edilen 4 farklı diyet takviyesi ile tedavi edilen 8 birey bulunmaktadır. Bakteriyel bolluklar, kurulan biyoinformatik analiz akışı aracılığıyla elde edilmiştir. Bakteriyel bolluk gözlemlerinin yanı sıra propiyonat ürettiği bilinen bakterilerin bolluk verilerindeki değişimler de araştırılmştır.

Anahtar Sözcükler: İnsan Bağırsak Mikrobiyomu, Organik asidemi, Metilmalonik asidemi, 16S rRNA gen bölgesi, Yeni Jenerasyon Dizileme

ABSTRACT

Establishing a Human Gut Microbiome Bioinformatics Analysis Pipeline to Study a Dietary Treatment Response for Methylmalonic Acidemia Patients

Human gut is known to be one of the most important locations in the human body via playing various essential roles in bodily functions through the microorganisms habituating here. Short chain fatty acids (SCFAs), with less than 6 carbon atoms, are known to be essential components in the energy production pathway. For the catabolism of certain amino acids and odd-chain fatty acids and synthesizing of propionyl-CoA, propionate is known to be an essential component. Due to certain pathogenic mutations in methylmalonyl-CoA mutase, methylmalonyl-CoA epimerase, and the cobalamin pathway genes, the human body can produce toxic intermediates that lead to methylmalonic acidemia after their accumulation. Only produced by the human gut bacteria, SCFAs, especially propionate, are microbial metabolites produced in the human large intestine and promoting their production has numerous health benefits. Contrary to healthy humans, increase of SCFAs in organic acidemia patients causes toxic aggregates, which may lead to detrimental symptoms. Using 16S targeted NGS technique, the enteric microbiota composition could be determined, and using Oxford Nanopore Technologies with full length 16S region sequenced, species level identification of bacteria could be successfully carried out. Aim of this thesis is to observe a decrease of the propionate producing bacteria in the human intestines with a dietary treatment. There are 8 individuals who are treated with 4 different dietary supplements included in this study. The bacterial abundances were obtained through the established bioinformatics pipeline. Along with the bacterial abundance observations, the changes in the abundance data of the bacteria known to produce propionate were investigated.

Keywords: Human Gut Microbiome, Organic acidemias, Methylmalonic acidemia, 16S rRNA gene region, New Generation Sequencing

1 INTRODUCTION AND AIM

1.1 Organic Acidemias

Being the primary method of energy production in human cells, the TCA cycle is one of the most essential pathways of the human body. Hence, a disruption in this pathway may have many adverse effects. There are a few conditions that cause the disruption of this pathway and aggregation of certain intermediates in the human body. These are generally called organic acidemias [1-3]. Being inherited conditions, organic acidemias in which the incorporation of succinyl CoA in the metabolism is prevented cause various intermediates, like propionic acid and methylmalonic acid, to build up in body, which results in various symptoms ranging from occasional lethargy to developmental retardation and, in some cases, coma and death [1-3].

1.1.1 Propionic acidemia

One of the genetic defects impairing the metabolic pathways and related to the propanoate (succinyl-CoA) pathway is known as propionic acidemia. Caused by mutations in the genes *PCCA* and *PCCB* which code for propionyl-CoA carboxylase heterodimer enzyme, propionic acidemia is known to be seen [4]. The enzyme has two subunits, alpha and beta coded by the aforementioned genes, respectively [4-5]. Due to the disruption of the enzyme's function, the normally toxic materials, such as propionyl-CoA, buildup in human blood, urine and tissues, therefore, and cause severe damage to the brain and peripheral nervous system [6-8]. Being an autosomal recessive condition, propionic acidemia can show various symptoms including loss of appetite, vomiting, lethargy, and in more serious cases heart abnormalities, seizures, coma and even death [8].

1.1.2 Methylmalonic acidemia

Another metabolic defect, the other commonly seen organic acidemia type is methylmalonic acidemia (isolated). It can be detected in humans via the high concentration of organic acids (methylmalonic acid) in the urine. It is determined to be a three-sided metabolic defect: vitamin B12-responsive, vitamin B12-unresponsive and methylmalonyl-CoA epimerase deficiency phenotypes of methylmalonic acidemia [8-11,16]. This metabolic defect may be caused by mutations in any of the various genes. Mutations of *MCEE, MMAA*, *MMAB*, *MMADHC* and *MMUT* genes are known to be impactful for the methylmalonic acidemia through methylmalonic acid (MMA) aggregation [10-21]. Of these, the *MCEE* gene, known as the Methylmalonyl-CoA Epimerase coding gene is responsible for catalyzing the conformational change between D-Methylmalonyl-CoA and L-Methylmalonyl-CoA [12-14,16]. Any mutation in the *MMAA* and *MMAB* genes, which code for enzymes that transport vitamin B12 (cobalamin) to mitochondria then form adenosylcobalamin, respectively, blocks the pathway, causing the further aggregation of methylmalonic acid [14-17]. Another gene that is crucial for the fatty acid and cholesterol degradation is *MMADHC*. This gene facilitates the vitamin B12 (cobalamin) to be transformed into adenosylcobalamin, which is required for the correct functionality of the methylmalonyl-CoA mutase preventing the methylmalonic acid buildup [15-19]. Lastly, *MMUT* gene codes for methylmalonyl-CoA mutase (MCM). The enzyme is responsible for converting methylmalonyl-CoA to succinyl-CoA [16,19-22]. There are various mutations reported regarding this gene causing inefficient to mediocre activity in MCM. While the mut^0 genotype is observed in missense alleles, the mut^- genotype shows a low or inefficient activity of MCM enzyme in mitochondria [21]. A few examples the mut^- genotype showing mutations are Y100C (tyrosine to cysteine at 100th amino acid-aa), R108H (arginine to histidine at 108th aa), N366S (asparagine to serine at 366th aa), V633G (valine to glycine at 633rd aa), R694W (arginine to tryptophan at 694th aa), R694L (arginine to leucine at 694th aa) and M700K (methionine to lysine at 700th aa) [21]. The most commonly seen missense mutations for mut^0 genotype are N219Y

(asparagine to tyrosine at 219th aa), G637E (glycine to glutamic acid at 637th aa), and G602R (glycine to arginine at 602nd aa). [22, 23].

Methylmalonic acidemia is characterized by the abnormally high levels of MMA in the blood and body tissues. Hence, the excess amount of organic acid in the body is excreted through urine and is considered to be the marker of the methylmalonic acidemia [16]. Like all other organic acidemias, methylmalonic acidemia is inherited through recessive autosomal pattern, too, and its occurrence in populations shows high variance between different regions. Generally, its prevalence varies between 1 in every 11000 people to 1 in every 100000 people. This value changes to 1/45000 to 1/20000 in North America while approximately to 1/26000 in China [15-19].

There are various types of treatments that exist to combat the symptoms of this metabolic dysfunction. In case of a vitamin B12-responsive type, primarily, vitamin B12 supplementation is provided to the patient. Furthermore, a specialized restricted diet in which limited natural protein intake with a high-calorie is administered. Recurring vomiting, eating difficulties and growth failure should be addressed during treatments. Additionally, carnitine as a food supplement, was seen to be effective in patients with carnitine deficiency. Lastly, albeit interestingly, reducing the propionate producing agents in gut flora is proposed for treatment purposes [16].

Early detection of this condition in individuals can prevent or reduce the chance of occurrence of early mortality, neurodevelopmental retardation and intellectual disability, some movement disorders, irreversible cerebral damage, pancreatitis, growth failure, functional immune system impairment, bone marrow failure, optic nerve atrophy, arrhythmias, cardiomyopathy, hepatic fibrosis and cancer, and renal cancer [16-18].

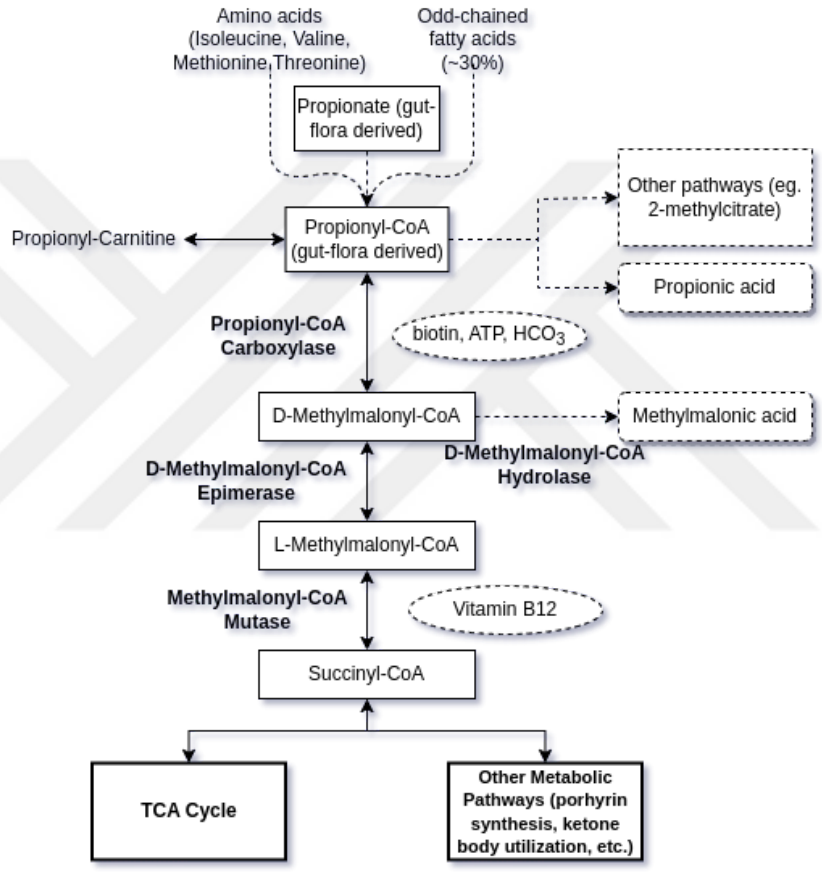

Figure 1. Propanoate or Succinyl-CoA Production pathway. One of the metabolic pathways, the succinyl-CoA production pathway is involved in the main energy production metabolism. Also, energy production from some amino acids and fatty acids can only be carried out if gut bacterial community derived propionate is converted into succinyl-CoA for its participation in TCA cycle (also known as Krebs cycle or citric acid cycle).

1.2 Human Gut Metagenome Sequencing And *in silico* Analyses

With the development of NGS technology, sequencing assays have become more common approach to untangle the human health. And via the multi-omics approach, different aspects of human health have been tried to be integrated with each other using various *in silico* methods for a deeper understanding of human biology.

Being one of the most commonly seen inherited organic acidemias, methylmalonic acidemia and its relation to the human gut microbiome wasn't researched before [135]. Nevertheless, propionic acidemia, other known to be a common organic acidemia, and its impact on the gut microbiota have been started to be highlighted with a 2021 study by Bordugo *et al*. [136]. The study employed an antibiotic (metronidazole) and a specific diet with a probiotic supplementation on only one patient through five months. The sequencing was carried out with Illumina MiSeq platform encompassing only the V3-V4 hypervariable regions of 16S rRNA gene, and low-coverage whole metagenome shotgun sequencing. The study revealed that while *Bacteroides fragilis*, belonging to the Bacteroidetes phylum, was a propionate producer, some *Bifidobacterium* and *Ruminococcus* species were keys characterizing the patient's gut microbiota [136]. Another, a more recent study which included two care centers (form the UK and Austria) having 29 total individuals (15 with propionic acidemia and 14 healthy) with varying ages between 3 and 35, focused on human gut microbiota diversity after dietary intakes [137]. The gut microbiome composition was determined through 16S V3-V4 gene regions with the Illumina MiSeq platform while metagenomic shotgun sequencing was applied to a subset of the patients, through which the mutations in the host-specific genes were observed [137]. Acidemia patients and controls from each health center were compared in terms on alpha diversity, and relative abundance. Consequently, *Dorea*, *Faecalibacterium*, *Roseburia*, *Streptococcus* and *Subdoligranulum* genera were found to be more (relative abundance) with both 16S and whole metagenome shotgun sequencing in propionic acidemia patients [137].

Other than organic acidemias, in the last 10 years, human gut microbiome studies demonstrated the enteric bacterial community's impact on the human diseases, such as cardiovascular diseases, type 1 diabetes, type 2 diabetes, irritable bowel syndrome (IBS), inflammatory bowel diseases (IBDs), colorectal cancers (CLCs), gouty arthritis, etc., through various *in silico* methods [107, 108, 109]. These methods are generally known as artificial-intelligence-based machine learning algorithms. Regression modelling, decision trees, random forest, support vector machines, elastic net, neural networks, etc. are a few examples for the machine learning (ML) and deep learning (DL) algorithms that are used for these *in silico* analyses.

Being one of the leading conditions that are fatal to humans is cardiovascular diseases. In 2020, Aryal et al. showed a relation between the individuals with a cardiovascular disease and the dysbiosis of the gut microbiota in these patients [107]. In the process, 951 (478 cardiovascular disease and 473 healthy) individuals' 16S rRNA gene sequencing data (fecal) belonging to the American Gut Project were used. In total, 5 different ML methods were implemented and compared to each other, and the best model was selected among them via using the area under the receiver operating characteristic (AUROC) curve. The best achieved AUROC curve value was ~0.70 with random forest via using only the 25 highly contributing OTU features while developing the model, where 0.00 represents the perfect anti-discrimination, 0.50 is random guessing and 1.00 shows a perfect discrimination between the groups [107].

Beura et al. employed metagenome-scale community metabolic modelling in the detection of a relation between the human gut microbiome and type 1 and type 2 diabetes, IBDs, CLCs, etc. [108]. MICOM and the microbiome modelling toolbox (MM toolbox), which incorporate the steady-state modelling strategy, are the most commonly used tools for this purpose. Using the partial-length 16S rRNA gene amplicon sequencing data of 83 human fecal samples, where 42 are healthy and 41 are gout patients, *Faecalibacterium* genus was found to be a determining feature for gout arthritis [108]. For IBD, using 108 (83 patient, 25 healthy) individuals' partial

16S rRNA gene sequences, *Roseburia spp., Faecalibacterium prausnitzii* and *Eubacterium rectale* were found to be decreased in dysbiotic IBD microbiomes. Furthermore, *Fusobacterium spp.* was found strongly associated with the CRC microbiomes after incorporating 586 individuals (365 cancer cases and 251 healthy individuals) in MM toolbox [108].

The aim of this study is to understand the role of the gut microbiota on the methylmalonic acidemia development and treatment response in terms of metagenome data.

2 BACKGROUND

2.1 Bacteria

Though the exact number is still unknown, the average estimated number of cells in the human body is estimated to be between $3x10^{13}$ and $4x10^{13}$. With this knowledge, considering that the human body is estimated to contain microorganism cells approximately the same number of cells, it is a necessity to assume that the human body functions are deeply dependent on the functions of the microorganisms habituating our bodies, which is called human body microbiota [86, 87]. Forming the biggest portion of the microbiota, the domain (or superkingdom), bacteria have a very diverse community [90]. The bacteria living in the human gastrointestinal system, specifically human intestines, are thought to contain the most diverse bacterial flora in the human body after the bacterial community habituating skin, which is thought to be similar to the Earth in its capability of nurturing the organisms habituating it [91].

Bacterial communities in human intestines are very diverse and create a balance among each other and lead a symbiotic life with each other and humans, their hosts. This eventuality bolsters the homeostasis in the human body, allowing it to continue its necessary biological functions. Thus, most of the bacteria habituating the human intestines are known to be harmless and even some are beneficial, in contrast to some of the microorganisms being pathogenic to humans. As opposed to common opinion, however, even some of the pathogenic bacteria are known to alleviate the symptoms of some of the human gastrointestinal diseases [92, 93].

As aforementioned, these levels under domain includes kingdom, phyla, class, order, family, genus, and species, as the main taxonomic levels whereas there are other organisms that fall under unclassified regions between these taxonomic levels, and generally have sub- prefix in their taxonomic classifications (Figure 2) [94].

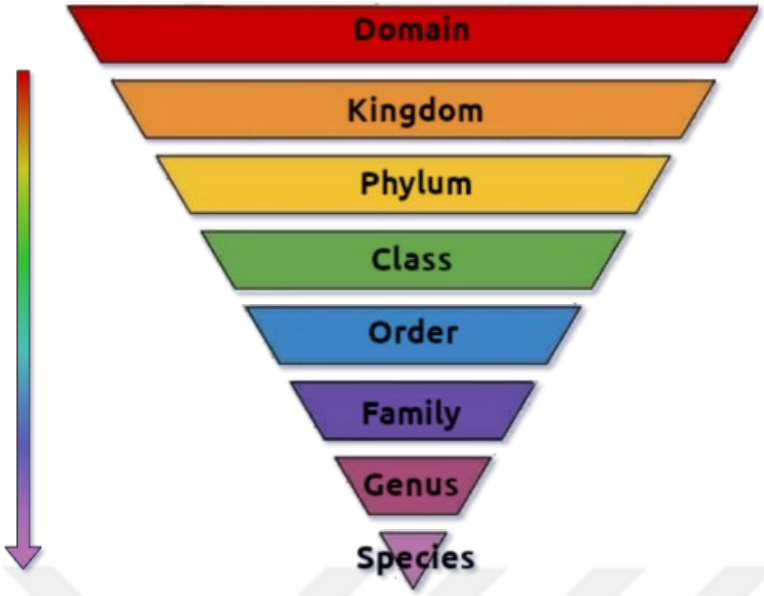

Figure 2. Taxonomic list starting from domain (superkingdom) and following with more detail, kingdom, phylum, class, order, family, genus and species. A group of organisms belonging to the same species will belong to the same genus while organisms which are under the same genus will be under the same family as the taxonomic classification.

2.2 Propionate

The enteric bacterial community has many varying contributions in a balanced and healthy intestine. One of them is producing metabolites which cannot be produced by human cells themselves, metabolites like propionate [93].

Being one of the most commonly found SCFAs in the human body, propionate is known to be involved in certain amino acid and odd-chain fatty acid degradation in tricarboxylic acid cycle, thus, primary energy production in human cells [92]. Propionate (also known as propanoate) is a 3-carbon conjugate base of propionic acid which has the molecular formula of $C_3H_6O_2$ or CH_3CH_2COOH.

Propionate cannot be produced by the human cells, but the enteric microbiota is responsible for the creation of propionate (gut flora derived). There are a number of bacteria that are known to produce propionate as metabolites (Table 1), such as *Akkermansia muciniphila*, *Roseburia inulinivorans, Bacteroides vulgatus*, *Bifidobacterium longum* [93-103].

Table 1: Bacteria found in the human gut microbiota and known to produce propionate as metabolites [93-103].

Roseburia inulinivorans	*Bacteroides thetaiotaomicron*	*Lactobacillus agilis*
Coprococcus catus	*Dalister succinatiphilus*	*Lactobacillus acidophilus*
Ruminococcus obeum	*Dialister invisus*	*Bifidobacterium longum*
Eubacterium hallii	*Megasphaera elsdenii*	*Bifidobacterium bifidum*
Akkermansia muciniphila	*Phascolarctobacterium succinatutens*	*Prevotella copri*
Veillonella parvula	*Lactobacillus rhamnosus*	*Alistipes putredinis*
Bacteroides vulgatus	*Lactobacillus gasseri*	*Blautia obeum*
Bacteroides uniformis	*Lactobacillus salivarius*	*Clostridium methylpentosum*

In the propionate pathway, also known as propanoate or succinyl-CoA production pathway, the addition of coenzyme A and later enzymatic reactions result in succinyl-CoA production, which has an essential task in the connecting TCA cycle (Figure 2). The pathway is used in catabolism of some amino acids, like isoleucine, methionine, threonine and valine, all of which are in L (laevus, left in Latin) conformation, and a portion of odd-chain fatty acids [104]. Being one of the most crucial energy production pathways, this pathway occurs in the mitochondria [105]. With propionate, a bacterial metabolism product, propionyl-CoA is produced. After carboxylase reaction, adding one more Carbon atom to propionyl-CoA, a 3-carbon molecule, d-methylmalonyl-CoA is produced. However, due to, most of the time, L enantiomers are naturally processed in human bodies, its conformation is changed to L-methylmalonyl-CoA. After vitamin 12, also known as cobalamin, is transformed to adenosylcobalamin, succinyl-CoA production takes place, and energy production through the TCA cycle and other metabolic pathways carry on (Figure 2) [106].

2.3 Next Generation Sequencing (NGS) - Generations and Techniques

2.3.1 Generations

There are various techniques to analyze the genome of a microorganism, such as sequencing technologies. Currently, the most well-known and trusted sequencing technology is Sanger Sequencing (first generation sequencing); while this is considered to be the gold-standard by the research community, it has several handicaps, compared to the more recent new generation sequencing (NGS) technologies [25-28]. With the developing technology, the better processing power and parallel sequencing allowed NGS to flourish and became the preferred sequencing option compared to Sanger sequencing.

Next generation sequencing represents three in-use and one under-development generations of sequencing technologies. Of these, while the second-generation sequencing focuses on short sequence reads and is represented by Illumina Sequencing, Roche/454 Sequencing technique compared to WGS for identification and classification of the microorganisms of the human gut microbiome.

Nowadays, Illumina sequencing is more widely used than any other next generation sequencing technology. However, compared to Illumina sequencing technologies, having longer reads is shown to be more advantageous when trying to classify and annotate the microorganisms [25]. Having a capacity for longer reads, hence making species level identification possible through only a gene, than its 2nd generation counterparts and necessitating relatively lower funds compared to other sequencing technologies, Oxford Nanopore Technologies (ONT) was preferred to be used in this study.

2.3.2 Techniques

2.3.2.1 Whole genome shotgun sequencing (WGS)

Whole genome shotgun sequencing is the most comprehensive of the DNA sequencing types and its requirements vary between different sequencing technologies. For instance, WGS preparation for the short read sequencer, Illumina, requires fragmentation primers for each fragment of DNA to be read. Whereas, for the long read sequencer ONT, fragmentation reaction isn't required for the sequencing preparation (ONT, Cat No SQK-LSK109) (v. ACDE_9064_v109_revP_14Aug2019). Recently, a shotgun metagenomic NGS what is called mNGS has started to be used for microbial community identification. mNGS is very similar to the WGS sequencing in terms of sequencing technique and library preparation. Although it has been suggested in the mid-2000s and gives high resolution results, it's still considered too costly compared to its counterparts in classification and identification of microbial communities [28]. Pyrosequencing (Roche 454) and Ion Torrent Sequencing, the third and, currently in mainstream use, last generation of sequencing technologies represent long sequence reads which include Oxford Nanopore (ONT) and Pacific Biosciences Sequencing (PacBio) Technologies [25-27]. Currently, the last, fourth generation of sequencing was proposed to perform the assay *in situ* with both targeted and non-targeted options [28].

2.3.2.2 Whole exome sequencing (WES)

A eukaryotic genome is separated into two main parts: intron (intragenic region) and exon (expressed region). The intronic regions of a genome are made up of the non-expressed and non-operator regions of a genome while exons include the rest of the genome [29, 30]. The whole exome sequencing technique pertains to the nucleotide sequences of all of the expressed regions that the sequenced organism has. Thus, this technique is generally used in human genetic testing to identify the

mutations and variants that possibly lead to a genetic defect and possible treatment plan for various cancer types [31].

2.3.2.3 Native RNA sequencing (RNA-Seq)

RNA, being similar yet different to DNA, is made up of nucleotides with alternating phosphate groups and ribose sugar backbones, unlike DNA with the deoxyribose sugar-based backbone, which causes RNA to be less stable. The RNA-Seq techniques include total RNA, mRNA, SmallRNA, Single-cell RNA and targeted RNA sequencing, etc. This technique is employed to observe the expression levels under different conditions. The study is also known as transcriptomics.

2.3.2.4 Methylation sequencing

Genome wide methylation sequencing mostly depends on the bisulfite conversion via PCR of the unmethylated cytosines, hence, the methylation ratio in the genome could be calculated from the percentage of the GC content. This way, the CpG (5'-cytosine-phosphate-guanine-3'), CHH (5'-cytosine-phosphate-nonguanine-3') and CHG (5'-cytosine-phosphate-noncytosine-3') islands are able to be identified [32-34].

2.3.2.5 Chip sequencing (ChIP-Seq)

ChIP (chromatin immunoprecipitation) assays are also used for identification of DNA binding transcription factors in genome-wide assays, which is also known as the ChIP-seq application of NGS [35]. This technology allows in-depth examination of gene expression regulations in various pathways and protein-nucleic acid interactions across the organism genome.

2.3.2.6 Targeted DNA sequencing

Unlike the WGS assays, targeted sequencing experiments require the isolated gDNA to be limited to a specific genome. First of all, to limit the gDNA to the wanted gene region, a polymerase chain reaction (PCR) with the appropriately specific forward and reverse primers would be carried out. Amongst the sequencing technologies, the most basic library preparation for Illumina sequencing belongs to targeted DNA sequencing. Whereas, whole genome sequencing library preparation for ONT is much shorter (Cat No. SQK-LSK109).

The conventional PCR process is composed of five steps, in which three of them repeat. Initial denaturation is responsible for activating hot-start polymerase enzymes and untangling the DNA chains for replication. The denaturation step, being the first one to be repeated through PCR cycling, helps longer DNA sequences to unfold and it increases the specificity of the amplification and is usually carried out at the same temperature as that of the initial denaturation step. The second cycling step is the primer annealing step. This step's temperature is usually calculated according to the melting temperature (T_m) of the desired primer used in the reaction. The approximate T_m of the primers are calculated using the 4*(G+C)+2*(A+T) formula. If a more accurate estimation of the T_m is desired, including the salt concentration ($[Na^+]$) in the calculation is required. Being the final repeated PCR step, the primer extension step extends the primers from the 3' end, complementary to the template DNA strand (5' to 3' end). The extension time is determined by the length of the target DNA amplicon and activity efficiency of the DNA polymerase enzyme used. The number of PCR cycles determines the amplification amount of the target DNA as 2^n of copies where "n" is the number of cycles. For instance, more than 1 billion copies can be achieved in the 30th PCR cycle. However, due to the amplification errors during the polymerase activity, PCR bias is considered to be a limiting factor for PCR cycles in microbiota studies. In the first days of the PCR, the main limitation was the fluctuation of the temperature and its effects on the DNA polymerase [36]. Using a high-quality DNA polymerase, limiting the temperature fluctuation and PCR cycles,

and choosing appropriately optimized primers may minimize the bias caused by the PCR process [37].

2.3.2.6.1 Identification of enteric microbial community

There are various methods used to differentiate the prokaryote and eukaryote microorganisms, such as the length of their genome, introns that aren't transcribed for future protein production, specific genes, etc. One of the most evolutionarily conserved and hypervariable genes are the ribosomal genes in both eukaryote and prokaryote microorganisms [38-39]. Along with the ITS (1 and 2) regions of eukaryotes, the rRNA genes are the mainstream genes to identify most of the microorganisms and distinguish one from another [39].

In the early days of microbiology, bacteria were classified according to their phenotypic properties, shape, spore production, structure of cell wall like properties [40]. However, nowadays, especially due to the impressive development of technology in the last decades, classifying microorganisms through their phenotypic properties has been used more scarcely, and their genotype is becoming more important when assigning in which taxonomic genus a newly discovered microorganism belongs [41].

In prokaryotes, Bacteria and Archaea, there are 3 ribosomal subunit genes present, which are 5S, 16S and 23S, and are found together in a rRNA operon [41]. Of these, 16S regions each represent smaller subunits while 5S and 23S are the genomic components of the larger subunit of the ribosomal RNA structures [43-45]. Especially, the 16S rRNA gene region, due to its conservation through the evolutionary stages and hypervariability in the genome from organism to organism, is considered to be the major differentiating DNA region in prokaryotic microorganisms [38].

The human gut possesses one of the richest microbial communities in the human body [68]. With various methods, it has been made possible to identify the microbes

habituating in the human intestines. For targeted sequencing, due to the high complexity of the human gut microbiome, the hypervariable region, 16S rRNA gene region is the preferred targeted region for the DNA based sequencing [46]. The 2nd generation sequencing, prefers choosing specific regions in the 16S rDNA, which are V3 or V3-V4 hypervariable regions of the complete nine regions of the 16S rRNA gene. On the other hand, with 3rd generation long read sequencing, the full-length targeted 16S rRNA gene DNA sequencing is targeted with PCR using the universal primers (27F forward – 5'- AGAGTTTGATCMTGGCTCAG-3' and 1492R reverse – 5' - CGGTTACCTTGTTACGACTT - 3') for the better identification of the found bacteria [47]. Hence, Illumina MiSeq 16S short-read technique remains behind in its resolution of diversity at genus and species levels in comparison to Oxford Nanopore (ONT) long-read sequencing. The single molecule long read 3rd generation sequencing technologies, such as PacBio and Oxford Nanopore, are known for their higher error rates than their 2nd generation counterparts [48]. Being approximately 1500 bp long in each prokaryotic organism, the full-length sequencing of the 16S rRNA gene region helps identify the microorganisms even at the species level [48-52].

While determining the quality of the reads for the ONT long read sequences, the value N50 is one of the most important metrics to determine the capability of the obtained reads. The N50 value represents the length of a sequence that when the sum of the read sequences longer than the N50 length reaches the 50% of the total of the complete read lengths. It is mostly similar to the median of lengths with higher weight to longer reads. Similar to N50, there are other values like N90, L50 and L90, where L50 is the number of the sequences until the length of N50, that are used to evaluate the quality of the obtained long read sequences. Albeit these values represent different metrics, N50 remains the most widely used metric among these [53].

With the decrease in the sequencing costs in later years, other methods for microbial identification started to be employed. Shotgun metagenome sequencing has become an option that some studies started to employ in their sequencing assays.

Better resolution is expected to be obtained through this method due to the whole genomes of the organisms being captured. However, consequently, since it captures whole genomes both the microbiota and host sequences, the necessary number of reads to be obtained in the assay increases drastically. Furthermore, in terms of bioinformatics approach, both the mapping and assembly-based metagenomic profiling methods introduce their own challenges, especially in highly complex samples, like enteric microbiome. Albeit, it has many disadvantages, a successful investigation offers a better functional analysis of the organisms and being able to characterize these organisms to sub-species and strain levels, which contribute to the development of science greatly [132, 133].

2.3.2.6.2 16S rRNA gene databases

There are a few bacterial and archaeal 16S sequence databases, such as RDP (Ribosomal Database Project), GreenGenes, SILVA and NCBI 16S databases [54-58]. Having approximately 1.5kb in length, the 16S rRNA gene region of prokaryotes is constantly being updated due to recent findings in taxonomy classification depending on the organism genomes. However, some of the above listed databases, not having been maintained and updated, are considered to be less useful, and remain obsolete. For instance, using the NCBI 16S database having been updated at the beginning of 2023 (as of 03/Jan/2023) results with the most up-to-date taxonomic information with regards to correct classification of 16S sequence data.

2.4 Bioinformatics Analyses For 16S Microbiome Data

Depending on the sequencing procedure, the bioinformatics analyses vary from the beginning. The use of short and long reads determines the quality of bases read, and this impacts the approach to the analysis of the 16S microbiome data.

2.4.1 Long reads

Long NGS sequencers like PacBio (Pacific Biosciences) and ONT (Oxford Nanopore Technologies are capable of sequencing reads longer than 500 base pairs as massively parallel sequencing, while the short reads can only be up to 400 bp in length. Hence, longer reads require a different method and tools for *in silico* analysis.

2.4.1.1 Basecallers

Basecalling is the transformation of the signal files that are created with the sequencing assay to the readable "fastq" files. Among various basecallers for ONT sequencing, such as Bonito, Albacore, Guppy, Fast-Bonito and etc., older versions of Guppy (ver. 3.4.1) remain inefficient in comparison to Bonito and Fast-Bonito [59-60]. Guppy version 6.0.6 is the last version as of 20/03/2022.

2.4.1.2 FastQC

FastQC is a read quality control visualization tool with both graphical and command line user interfaces [61]. It is used to visualize and determine the overall quality and length of each sequence.

2.4.1.3 BBTools

BBTools is a collection of tools for manipulating most of the NGS high throughput data based on Java. Among these, bbmap, bbduk and bbmerge are used to manipulate the sequence data [60]. bbduk is used for trimming and error correction while bbmerge merges paired end reads by aligning the forward and reverse reads together. Additionally, bbmap is a global aligner for the NGS sequences.

2.4.1.4 Magic-BLAST

Magic-BLAST is an RNA and DNA mapping tool for NGS assays. It incorporates the BLAST algorithm for alignments and it uses custom prepared and generated databases for its function [61].

2.4.1.5 SAMtools

SAMtools is an NGS focused, high throughput data manipulator tool. It is used to create consensus sequences, SNVs analysis, data format conversion, sorting, merging and indexing by using the sequence alignment results [62].

2.4.1.6 BLAST+

By utilizing only, the sequence data, Basic Local Alignment Search Tool (BLAST) uses its own algorithm, which is known to be more sensitive in comparison to the greedy algorithm employed by megablast, a faster version of BLAST, to find the similarity of a sequence to a database [63]. With the command line interface (CLI), different functions allow one to align either amino acid or nucleotide sequences to a protein or nucleotide-based database.

2.4.2 Short reads

The technologies outputting short reads, like MGI and Illumina sequencing technologies can sequence partial 16S regions for each bacterium, especially the hypervariable regions of V3 or V3-V4, among the 9 regions of 16S. The data obtained from this sequencing has relatively lower resolution, as mentioned above, however, has very high read qualities compared to its long read counterparts with generally 85-95% of the reads being higher than expected quality thresholds (Q30) [66-67].

Starting with read quality control, error correction, merging the paired end reads with both forward and reverse reads, and clustering the reads for operational taxonomic units (OTUs) and amplicon sequence variants (ASVs) table creation before assigning the appropriate taxonomic nomenclature to each OTU is the standard Illumina 16S analysis. To accomplish this, various tools, like BBMap aligner [62], DADA2 [68] and Qiime2 pipeline [69], which offers Deblur [70] and DADA2 in its analysis pipeline, can be used. Preceding Qiime2, Qiime (v. 1.x) is, still being used by some studies [137]. However, due to its development having been stopped on 01/01/2018, it is considered outdated and not officially supported by its creators. Therefore, the use of this pipeline is not recommended.

3 MATERIALS AND METHODS

3.1 Sample Collection

After obtaining the consents from 8 individuals with methylmalonic acidemia disorder, their fecal samples of the patients at four different time points were collected. The obtained samples were sequenced with 16S targeted sequencing using Oxford Nanopore Technologies (ONT, Oxford, UK) – MinION (Mk1B) device and SQK-16S024 ONT Kit (ONT, v. 16S_9086_v1_revM_14Aug2019). The sample sets collected from patients were determined as the 1st step, 2nd step, 3rd step and 4th step of treatment with a dietary nutrient mixture (Table 2).

Table 2. Sample information. 8 participants with MMA were included in the study with varying childhood ages, between 1.5 and 13.5.

SAMPLE ID	TREATMENT STEP(S)	SEX	AGE
1	Steps 1-2-3-4	Female	2.5
2	Steps 1-2-3-4	Male	8.5
3	Steps 1-2-3-4	Female	5
4	Steps 1-2-3-4	Female	1.5
5	Steps 1-2-3-4	Female	13.5
6	Steps 1-2-3-4	Male	1.5
7	Steps 1-2-3-4	Female	6
8	Steps 1-2-3-4	Male	7.5

In each treatment step, a different treatment was applied for at least 2-month duration to observe the patients' responses. Step 1 included a treatment with a high protein content dietary supplement, having 50% natural and 50% synthetic protein content. While the protein content was lowered in the latter steps, the 2nd step had a ratio of 50%-50% distribution between natural and synthetic protein content, 3rd step had 80% natural and 20% synthetic proteins, lastly, the 4th step only had the addition of metronidazole application differing from the 3rd step (Table 3).

Table 3. Treatment steps with protein containing dietary supplements.

STEPS	CONTENT
Step 1	High protein content - 50% natural protein 50% synthetic protein
Step 2	Low protein content - 50% natural protein 50% synthetic protein
Step 3	Low protein content - 80% natural protein 20% synthetic protein
Step 4	Low protein content - 80% natural protein 20% synthetic protein + metronidazole

3.2 Microbiome Sequencing

3.2.1 gDNA isolation

The samples were collected to a tube with a protective solution, then stored at +4◦C for gDNA isolation. For each sample, up to 200mg was used for gDNA extraction with ZymoBIOMICS DNA Miniprep Kit (Zymo Research, Cat. No. D4300). The real DNA concentrations were measured with Qubit 2.0 (Thermo Fisher, Cat. No. Q32866) Qubit dsDNA HS Assay Kit (Invitrogen, Cat. No. Q32854), and the DNA purity was determined with NanodropOne (Thermo Fisher, Cat. No. ND-ONE-W). Since the sequencing was be targeted to 16S gene region, with PCR, the purity of samples carries an important role. The reference values to confirm the purity of the gDNA samples were obtained from ONT (260/280 ~1.80 and 260/230 ~2.0-2.2). The samples that couldn't justify the purity requirements were purified using Agencourt AMPure XP beads (Beckman Coulter, Cat. No. A63881).

3.2.2 Polymerase Chain Reaction (PCR)

For the 16S rRNA gene region targeted sequencing assay, the gene should be excised and multiplied before the assay is carried out for a better resolution. For this purpose, polymerase chain reaction (PCR) was carried out. In the reaction, the forward and reverse universal primers for the complete 16S gene region were used [71]. The ONT 16S sequencing kit has its own barcodes, which are attached to the universal primers for the PCR. The PCR conditions were optimized for NGS reads

longer than 1 kb by ONT and detailed in the 16S sequencing kit guidelines (v. 16S_9086_v1_revM_14Aug2019, last updated on 21/04/2021) (Table 4).

Table 4: 16S targeted PCR conditions. After an initial denaturation at 95°C for 1 minute, with 25 PCR cycles, denaturation at 95°C for 20 seconds, annealing at 55°C for 30 seconds, and extension at 65°C for 2 minutes was carried out. Lastly, the final extension was at 65°C for 5 minutes.

PCR STEP	TEMPERATURE (°C)	TIME	CYCLES
Initial Denaturation	95°C	1 min	x1
Denaturation	95°C	20 secs	x25
Annealing	55°C	30 secs	
Extension	65°C	2 mins	
Final Extension	65°C	5 mins	x1
Hold	4°C	∞	x1

3.2.3 16S targeted amplicon sequencing

During the multiplex library preparation, the obtained gDNA were amplified with PCR process using the ONT-16S barcoding kit (ONT, Cat. No. SQK-16S024) and the kit's guidelines (v. 16S_9086_v1_revM_14Aug2019, last updated on 21/04/2021) were followed. The prepared library was loaded onto the ONT MinION Flowcell (v. 9.4.1, Cat. No. FLO-MIN106D) and the 16S sequencing was performed.

3.3 Bioinformatics And Statistical Analyses

3.3.1 Bioinformatics analysis

During the sequencing, the data were obtained as “fast5” files using MinKNOW (v. 22.03.5) GUI program, N50 of sequencing was found to be approximately 1.48 kb (the average V1-V9 length is 1.45 kb). The adapter and barcode removal and first quality filtering were performed with ONT-guppy (v. 6.0.6) CLI program, and the “fastq” formatted sequencing files were made ready for downstream analyses after this process for each sample. The quality score of the completed NGS assay was determined with FastQC (v. 0.11.9) and the average score was found to be Q19 (Phred score). Using BBTools (v. 39.01), seqtk (v. 1.3), magicblast (v. 1.6.0) and samtools (v. 1.13), the consensus sequences were created, and these sequences were annotated with NCBI blastn (v. 2.0.14) (Figure 3). The genus and species level taxonomic annotations were carried out with a 95% identity threshold. After the OTU table creation for each sample, the relative abundance values were calculated using phylum and species levels.

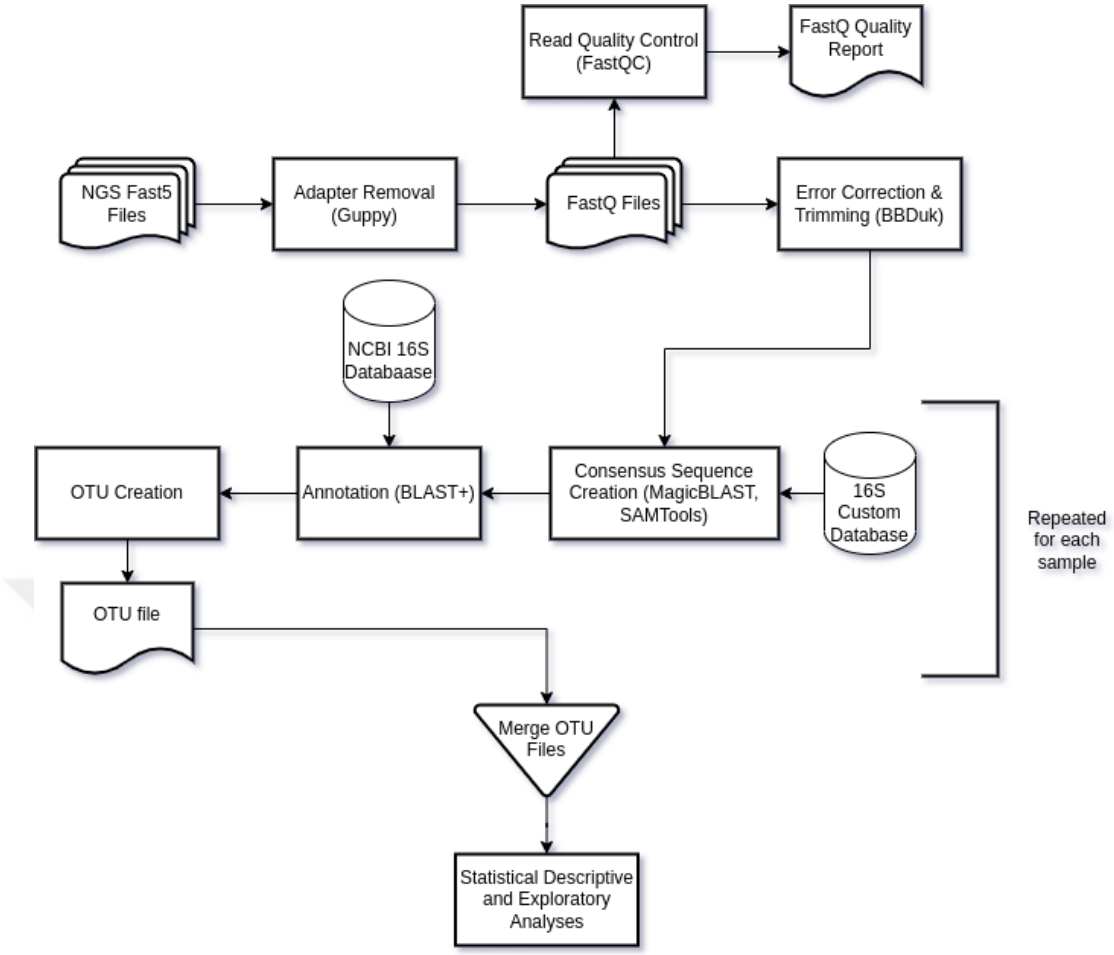

Figure 3. The created single-end long-read Oxford Nanopore Microbiome bioinformatics pipeline for the methylmalonic acidemia patients. The pipeline starts with the fast5 Nanopore signal files and results with OTU files.

3.3.2 Biostatistics analysis

3.3.2.1 Data imputation

Data imputation is a collection of statistical techniques for filling the missing data for further statistical analyses when some of the data couldn't be collected. There are various reasons for this action to be carried out. For instance, when data collected is limited and all of the data are crucial for any of the analyses that are to be carried out, or the missing data may skew the data distribution preventing the statistician to apply the correct hypothesis tests.

In our study, not to interfere with the distributional properties of the dataset, median values were incorporated for the missing data.

3.3.2.2 Hypothesis Tests

3.3.2.2.1 Alpha value

Alpha value, which is usually depicted as α, as in Greek character, is the probability of a type 1 error in any statistical test result. In any hypothesis testing, such as Student's t-test, Wilcoxon Signed Rank test, etc., the result shows a p-value which determines the significance of the test result. Because hypothesis tests reveal the non-randomness with a level of confidence, the level of confidence is chosen by the statistician performing the tests, which is determined by the value alpha. Commonly, though there are many claiming it incorrect, this value is chosen to be 5%, or 0.05. This could be selected as 0.1 or 0.01, but the confidence interval changes as a result. In our study, the alpha value was set to 0.05.

3.3.2.2.2 Shapiro-Wilk normality test

Prior to any other test, the distribution of the data should be checked. Shapiro Normality test is a hypothesis test where the null hypothesis is considered to be that the data is normally distributed. Therefore, in the case that the resulting p-value is less than the preselected alpha value, the null hypothesis is rejected [72]. This process can be done using QQ-plots, as well [73]. In our study, Shapiro-Wilk's Normality test was applied for each data.

3.3.2.2.3 T hypothesis test

T-tests are applied to normally distributed datasets with only two groups. However, the type of the hypothesis test applied changes when the data obtained contains the data belonging to the same individuals with time lapse. In this case, the test should be carried out as a paired, otherwise, unpaired test should be chosen [74].

Another parameter that separates T-tests into two is the variance. The data variance differences of the groups are checked with the F-test. It demonstrates whether the variances of the two groups compared have similar variances or not. Should the F-test resulted p-value be lower than the chosen alpha, the null hypothesis, which proposes that the variances are the same, is rejected. Under this circumstance, Welch's T-test should be applied, otherwise, Student's T-test is to be chosen [74].

3.3.2.2.4 Wilcoxon hypothesis test

Unless the Shapiro Normality Test returns p-value greater than the alpha value, one of the non-parametric pairwise tests is to be applied. These are reported as Wilcoxon Signed Rank Test or Mann Whitney U Test, otherwise known as Wilcoxon Rank Sum Test [74]. When the dataset consists of paired data, similar to when applied T-test options, Wilcoxon Signed Rank Test is to be applied.

The OTUs of the set groups were evaluated for normality with Shapiro-Wilk's normality test, and the comparisons were carried out with pairwise t-test or Mann-Whitney U (paired Wilcoxon signed rank where applicable) test, depending on the results of the normality test ($p<0.05$). The pairwise comparisons were carried out between the 1st step and 2nd, 2nd and 3rd, 3rd and 4th, and finally, 1st and 3rd steps.

3.3.2.2.5 ANOVA hypothesis test

Otherwise known as Analysis of Variance, it is applied when the datasets are compliant to normal distribution. Although ANOVA has the name "variance" in it, the test compares the variances of means of groups with each other. If the groups' variances are sufficiently apart from each other, according to the given alpha value, the null hypothesis is rejected [75]. The test is only applied when more than two groups are being compared with each other and can be considered as a multi-group version of t-test.

3.3.2.2.6 Kruskal-Wallis hypothesis test

ANOVA's non-parametric counterpart is known as the Kruskal-Wallis test. It is known as one way ANOVA on ranks, as well. Kruskal-Wallis test, similar to Wilcoxon tests, is applied when the dataset distribution doesn't comply with the normal, or gaussian distribution. The test applies the Wilcoxon Rank Sum test to more than two groups.

3.3.2.3 P-value adjustment

When multiple tests are applied to a dataset, there remains a possibility to obtain a false positive result. Since the alpha value was selected to be 0.05, the random positive results increase, as the number of applied hypothesis tests increase. There are a few methods to combat this problem. Most common ones are the Bonferroni and FDR (Benjamini-Hochberg false discovery rate) p-value correction methods [76]. While the Bonferroni correction is one of the stricter methods and may cause the false negative rate to increase slightly, the Benjamini-Hochberg method mostly prevents the false negatives.

Feature selection was carried out by eliminating the statistically insignificant results ($p>0.05$). Through the Kruskal-Wallis hypothesis test, the statistical significance values of the OTUs at species level were obtained. Afterwards, the obtained p-values were adjusted using the FDR (false discovery rate) approach.

3.3.2.4 Regression

Regression is used to determine the relationship between an outcome or response, which is the dependent variable, of an action and its independent variables. It has linear and non-linear forms to create the best fitting models for the variables included in the regression processes [77].

3.3.2.4.1 Linear regression

Being the most common application of regression models, linear regression [77]. Linear regression aims to find a linear relationship between the data and fit them to a line, otherwise described as the best fit line. The formula below represents the linear model created after the regression application. While "*y*" demonstrates the response values to the best line, the "*β*" and "*x*"s represent the regression slope coefficients and the variables, respectively. Additionally, the "*c*" is the constant intersection value at the y-axis of the two-dimensional coordinate plane or cartesian plane.

$$y=\beta*x+c \qquad \text{or} \qquad y=\beta_1*x_1+\beta_2*x_2+\beta_3*x_3\ldots+c$$

Linear regression has various types in which the dataset distribution type determines the regression type to be used. For example, if the data is composed of count data, abiding by the Poisson distribution, Poisson generalized regression, and logistic generalized linear regression is used for binary data.

Through an existing study, patients' appetite and 23 different metabolite abundance values of the 8 individuals at each of the treatment steps were procured: 2-methylcitrate, 3-hydroxybutyrate, 3-hydroxypropionate, albumin, ammonium, vitamin B12, carnitine, ferritin, folate, isoleucine, isoleucine ratio, lactate, leucine, methionine, methylmalonic acid, prealbumin, propionylcarnitine, propionylglycine, pyruvate, threonine, tiglylglycine, valine, zinc and appetite values. Incorporating the relative abundances of only the significantly found 20 statistically significant OTUs ($p<0.05$) with the lowest adjusted p-values as variables, linear regression models were created for each of the metabolite abundance values obtained using the R statistical software base stats package lm function.

3.3.2.4.2 Nonlinear regression

Nonlinear regression is used to achieve exponential best fit lines. Any exponential function, like power and polynomial functions can be used as the best fit

line. The dataset is assumed to be parametrically distributed in nonlinear regression. Otherwise, with non-parametric datasets, machine learning approaches are preferred [78].

3.3.3 Diversity analysis

The alpha and beta metrics are used to measure the diversity within and without the sample groups. The calculation starts after the rarefaction of the sample OTUs with the appropriate depth settings [79].

3.3.3.1 Alpha diversity

There are two important factors in alpha diversity calculations: richness and evenness. Richness is evaluated by counting different features in a given OTU table while evenness is calculated by measuring the distribution of the features based on their relative abundance values.

There are various metrics used to analyze the alpha diversity of a dataset. Most commonly used three of them are the observed features, Shannon diversity index and Simpson's index. "Observed features" measures the number of features and focuses only on the richness of a dataset. Shannon diversity index explains richness and evenness by providing equal weight to both in its calculation. Lastly, the Simpson's index can explain both richness and evenness, but giving evenness more weight during its calculation [80].

The created OTUs at species level were treated with downstream exploratory and confirmatory statistical analyses with R (v. 4.1.3) using Rstudio IDE (v. 1.4.1717). The time periods with treatment and pre-study were each separated into 4 paired groups and statistical analyses were carried out at phylum and species level. The alpha diversity of the set groups was calculated using three different diversity metrics, "observed features", "Shannon" and "Inverse Simpson" with the R's "phyloseq" (v. 1.44.0) package.

3.3.3.2 Beta diversity

Being another ecological diversity measure, beta diversity aims to analyze the ratio between within and without the group diversity [81].

In microbial community beta diversity analyses, three of the most commonly used metrics are Jaccard's distance, Bray-Curtis dissimilarity, and UniFrac distances. Jaccard distance metric incorporates only the existence of the features while Bray-Curtis dissimilarity and UniFrac distance include the abundances in their calculations [82-84]. UniFrac distance is more commonly used in short-read NGS applications, because it integrates the distances between OTUs through including a phylogenetic tree, which is created using a clustering technique of the high-quality reads, in its calculation [85].

To evaluate the differences between the pre-study and treatment groups, beta diversity analysis was carried out with Bray-Curtis dissimilarity index calculation for each sample bacterial relative abundances at both phylum and species level.

4 RESULTS

After the OTU relative abundance creation, exploratory statistical analyses were carried out and the averages, medians, maximums and minimums belonging to each of the OTUs at the phylum and species level were calculated (Appendix Table 1).

4.1 Phyla Level OTU Calculation Results

The total number of phyla observed in the assay was 30, and among them, the most abundant 4 were Firmicutes, Proteobacteria, Bacteroidetes and Actinobacteria in the order of relative abundance (%) (Figure 4a).

Eleven of the phyla were determined to be common among the treatment steps. 23 different phyla were discovered to inhabit the individuals' gut microbiota in the duration of the treatment 1. The total phyla level bacteria count was seen to be 24 while the patients were going through the step 2 of the diet-based protein supplement treatment. 22 different phyla were observed in the 3rd step of the treatments. Lastly, the 4th step showed only 17 different phyla. On the other hand, step 1 had two specific phyla, which are Caldiserica and Kiritimatiellaeota that were present only in the gut microbiota during the 1st step of treatments. There was only one phylum observed in each the second treatment step and the third, which are Calditrichaeota and Gemmatimonadetes, respectively (Figure 4b).

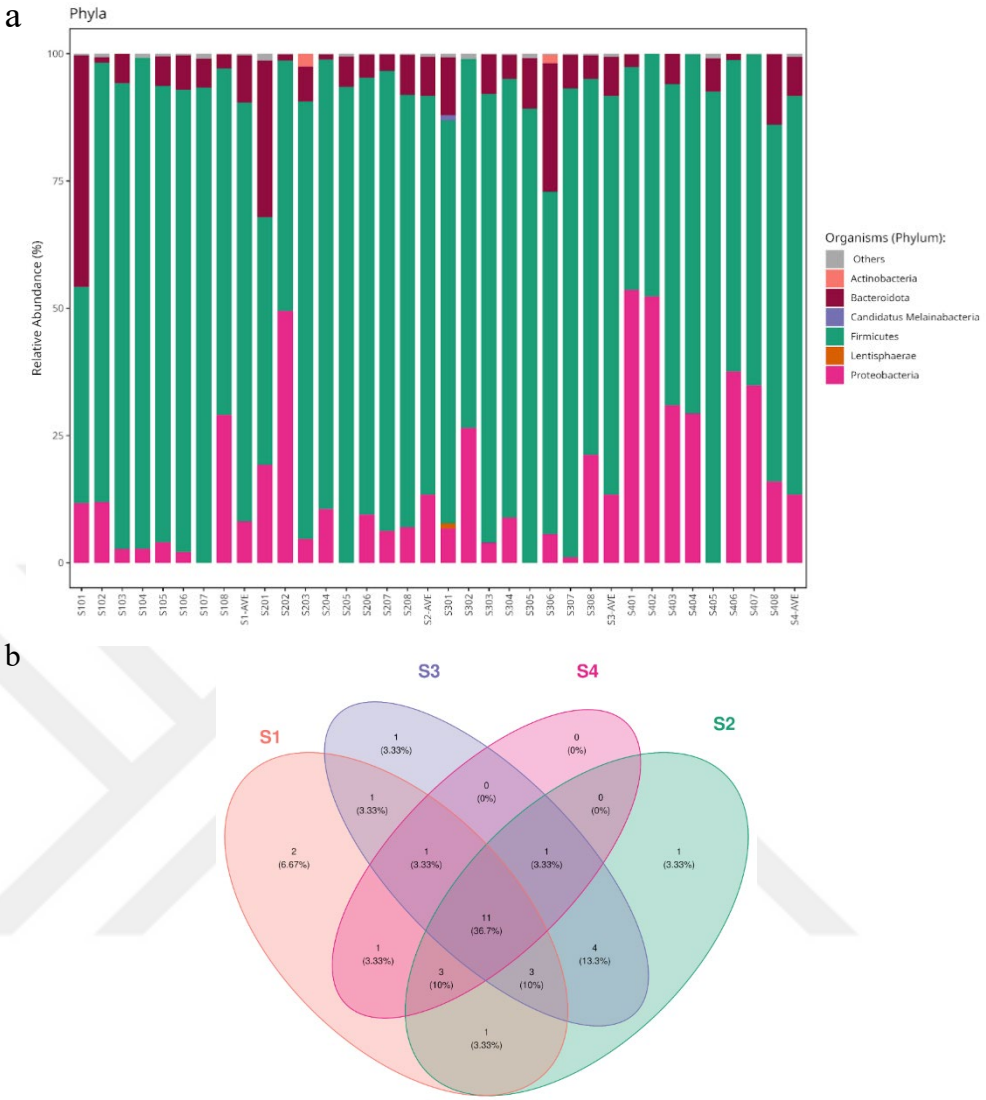

Figure 4. (a) Stacked bar plot of all phyla and their relative abundances. Phyla with relative abundance more than 1% are depicted in the plot above. The rest of the phyla relative abundances were added together and labelled as “Others” and marked with “grey”. Enteric microbiota composition of each sample and the averages of each group was shown in the plot. (b) The distribution of phyla in the cohort in the separate treatment steps. In total, 30 different phyla were found, and while some are shared between the treatment steps, some are observed to be belonging to a specific treatment stage. The treatment step 1 is depicted with S1 (orange) while S2 represents Step 2 (green), Step3 is shown as S3 (purple) and, lastly, S4 is the treatment step 4 (pink).

4.2 Species Level OTU Calculation Results

There were 2744 species in total found in the assay. Among these, the most dominant first 10 species were found to be *Faecalibacterium spp.*, *Blautia spp.*, *Ruminococcus spp.*, *Escherichia spp.*, *Roseburia spp.*, *Megamonas spp.*, *Bacteroides spp.*, *Clostridium spp.*, *Anaerostipes spp.* and *Veillonella spp*. (Figure 5a).

Among the 2744 total species level bacteria, 519 of them were found common at all of the treatment steps while 502 at only the step 1, 336 during only step 2, 192 in only step 3 and 295 species were only observed during the treatment step 4 was being carried out (Figure 5b).

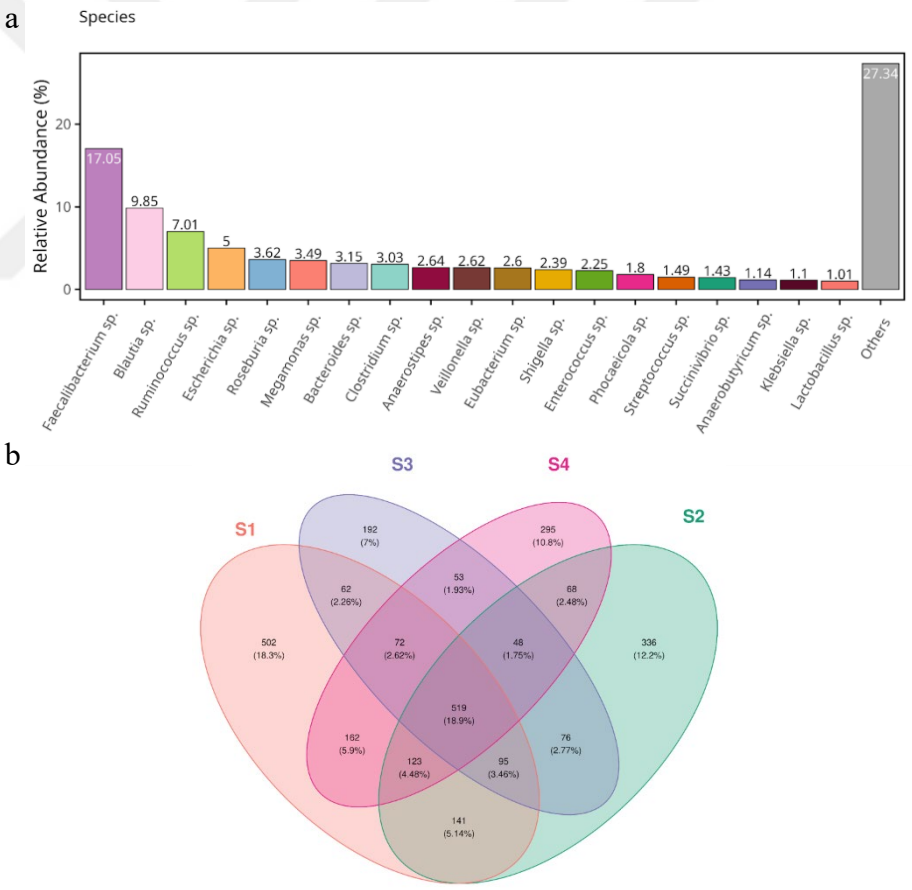

Figure 5. (a) OTUs at species level with relative abundance (%) values. The OTUs which had relative abundance values lower than 1% were summed and depicted under the name “Others”. (b) The distribution of species in the cohort in the separate treatment steps. In total, 2744 different species were found, and while some are shared between the treatment steps, some are observed to be belonging to a specific

treatment stage. The treatment step 1 is depicted with S1 (orange) while S2 represents Step 2 (green), Step3 is shown as S3 (purple) and, lastly, S4 is the treatment step 4 (pink).

4.3 OTUs Phyla & Species Level Alpha - Beta Diversity Calculations

OTUs of the steps were compared for enteric bacterial diversity, in terms of richness, evenness and sample distances, at the phylum and species level (Figure 6).

Alpha diversity analysis, which focuses on richness and evenness of the samples, showed a decrease in richness, albeit statistically insignificant at the observed richness at phylum level ($p>0.05$). However, the increase in the other indexes, especially, the Inverse Simpson index indicates that evenness of the samples has increased significantly ($p=0.02$) after the change of the treatment at step 4 (Figure 6a/Phylum).

According to the alpha diversity analysis at the species level, the statistically significant difference was observed between the treatment step 1 and step 3 ($p=0.042$). The other changes weren't found significant with the Wilcoxon Signed Rank test, albeit the observed richness at species level has decreased slightly. However, due to the increase in evenness for Shannon and Inverse Simpson metrics, their diversity values increase, too (Figure 6a).

The beta-diversity analysis was visualized with a PCoA plot and the ellipses were drawn at the 95% confidence range. The PCoA plot at the phylum level was able to show the variance between the data with two axes, however the gut microbiome compositions didn't show a significant separation between the treatment groups in both phylum and species levels (Figure 6b).

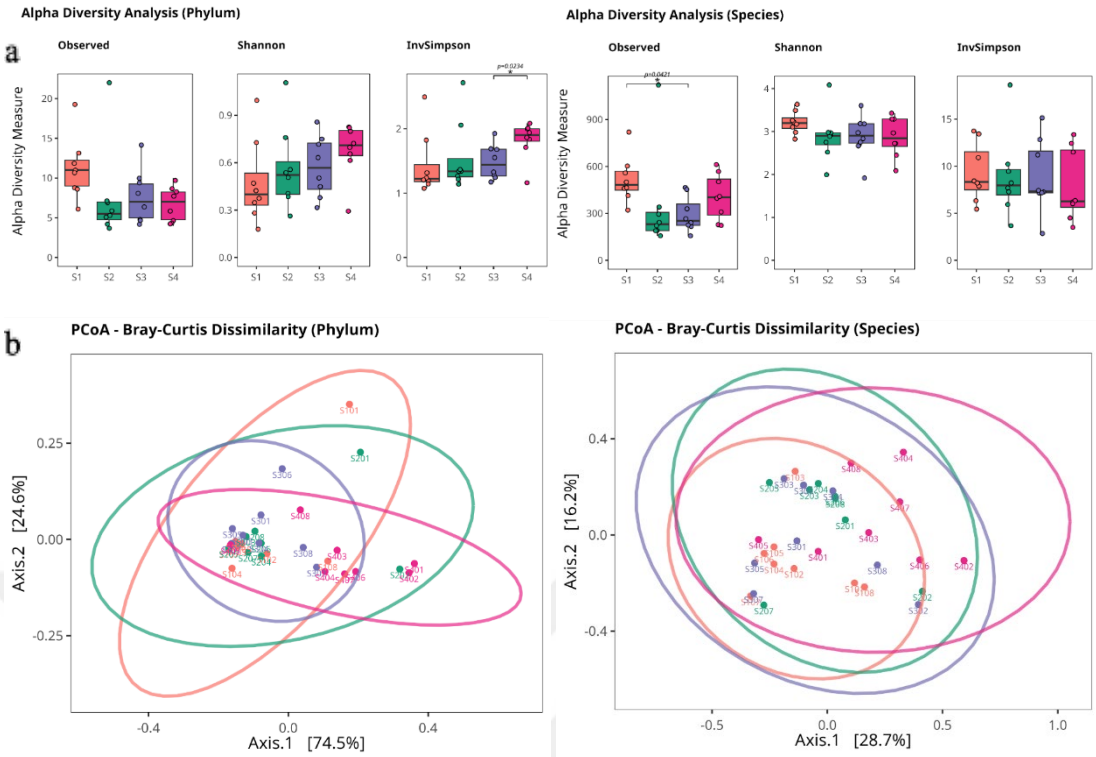

Figure 6. The figure above shows the alpha and beta diversity visualizations with box and PCoA plots comparing the first step of the treatment and the latter steps. While S1 represents the first step of the treatment (orange), S2, the second step (green), S3, the third (purple), and, finally, S4 indicates the fourth treatment step (pink).

4.4 Species Level Propionate Producing OTUs

Methylmalonic acidemia patients are known to have been unable to process some amino acids and fatty chains due to an impairment in energy production metabolism. This pathway involves a propionic acid-derivative, Propionyl-CoA, which is a precursor for the other crucial intermediates, such as D-methylmalonyl-CoA, Methylmalonyl-CoA, and succinyl-CoA. Furthermore, propionic acid is obtained from enteric microbiota as its conjugate base, propionate, in which certain bacterial species are known to produce. propionate as their metabolites, such as *Bifidobacterium longum*, *Bacteroides vulgatus*, *Roseburia inulinivorans*, *Ruminococcus obeum*, *Blautia obeum*, *Alistipes putredinis*, *Prevotella copri, Ruminococcus bromii, Bifidobacterium bifidum*, *Coprococcus catus*, *Akkermansia muciniphila*, *Veillonella parvula*, *Bacteroides thetaiotaomicron*, *Bacteroides uniformis*, etc. Although at species level, no statistically significantly different OTUs

were discovered, slight decreases in relative abundances of *Akkermansia muciniphila, Ruminococcus bromii, Bacteroides uniformis and Bacteroides thetaiotaomicron* when compared to the step 1 of the treatment sets were observed (Figure 7).

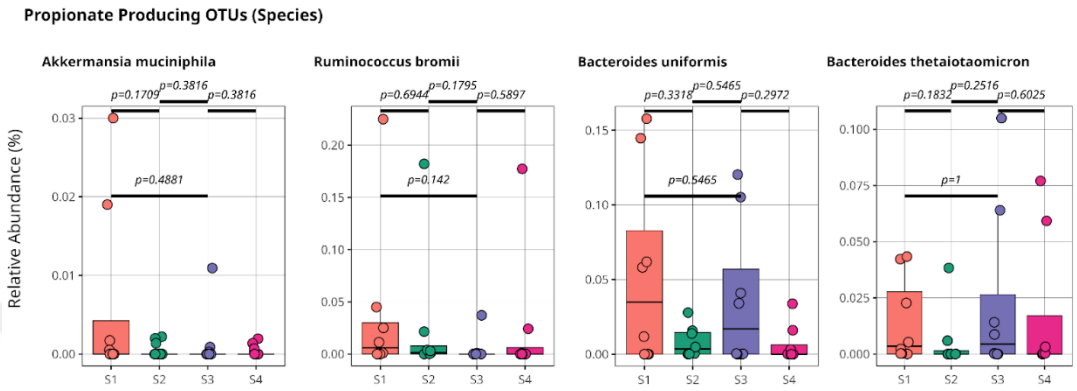

Figure 7. Detected propionate producing bacteria with relative abundance values in a decreasing pattern are given. The S1 (orange) stands for step 1 of the treatment, S2 (green) depicts the treatment step 2 while S3 (purple) is the step 3 and S4 (pink) represents the last step of the treatment duration.

4.5 Linear Regression Results

To determine the impact of the methylmalonic acidemia treatments on the gut bacteria, and visa-versa, regression analyses were carried out at genus and species levels. Since the high presence of MMA in the urea extract is known to be one of the signs of the metabolic disorder, it was used to determine the response value (y) of the regression model while the variables were set as the bacterial OTUs at genus (Table 5) and species (Table 6) levels. The one of the most commonly seen symptoms of the methylmalonic acidemia are vomiting and drastic decrease in appetite. Hence, a regression model based on the individuals' appetite levels was developed at genus level. The appetite values were selected to be the response values (y) to the linear regression model while the variables (features) were OTUs at genus levels.

The linear regression results may include errors which may be caused by collinearity of the relative abundance values of OTUs, which can be detected by the correlation results. In the separate regression models, this case occurred for 8 OTUs

at the species level, *Butyribacter spp., Desulforamulus spp., Enterocloster spp., Hungatella xylanolytica, Ruminococcoides spp., Petrocella spp., Bautia argi,* and *Caproicibacterium spp.*, which are highly correlated to *Variimorphobacter spp., Butyribacter spp., Lacrimispora spp., Anaerosaccharophilus spp.* while at the genus level, *Solibacterium, Ruminococcoides, Lawsonibacter, Peptacetobacter, Coproicibacterium, Virgibacillus* and *Hydrogenispora* were found highly correlated to *Tepidibacterium, Petrocella, Variimorphobacter, Lacrimispora, Anaerosaccharophilus* and *Gluceribacter* (Figure 8).

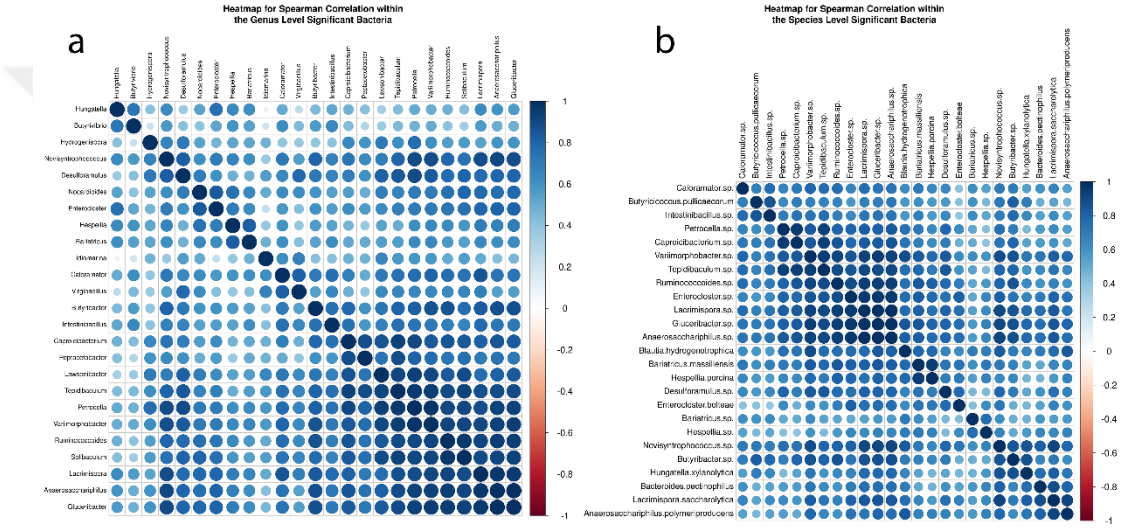

Figure 8. Correlation plot for best significant OTUs at the genus (a) and species (b) level. Significant OTUs for the first 3 treatment steps were incorporated while calculating the correlations. Dark colors (dark blue or dark red) represent high correlation values along with the circles with a large size.

The created linear regression models at the genus and species levels for MMA, showed very few bacteria that were found to be statistically significant ($p<0.05$). The adjusted R^2 values were found to be 0.9091 at genus level and 0.8945 at species level. The p-values are 0.02986 at genus level and 0.00252 at species level. Therefore, the model shows high accuracy with a linear change at both genus and species regression. While at the genus level, only *Butyribacter,* along with small contributions from *Bariatricus* and *Petrocella*, was found to be a statistically significant contributor. According to the genus level regression model, 0.001 increase of *Butyribacter* relative abundance causes an increase in MAA

concentration by 1187.073 units (Table 5). However, at the species level, 0.001 increase of *Bariatricus spp.* relative abundance causes 34.519 units increase in the MAA concentration in the patients' urea (Table 6).

Another linear regression model was created for the appetite levels of the patients, as it is well known that organic acidemia patients suffer from vomiting and loss of appetite [16]. The adjusted R^2 value was found to be 0.9756 and the p-value 0.004362. Hence, it could be said that the OTU data and appetite values could be represented linearly. The changes of the OTUs at the genus level was shown to be statistically significant ($p<0.05$) at the genera of *Gluceribacter*, *Anaerosacchariphilus*, *Nocardioides*, *Variimorphobacter*, *Novisyntrophococcus*, *Petrocella*, *Bariatricus*, *Lacrimispora*, *Butyribacter*, *Tepidibaculum*, *Idiomarina*, *Caloramator* and *Syntrophococcus*. Among them, *Anaerosacchariphilus, Nocardioides, Variimorphobacter, Novisyntrophococcus, Bariatricus, Butyribacter* and *Idiomarina* are found to be inversely correlated with the appetites of the individuals. Thus, an increase in the relative abundances of these, especially *Novisyntrophococcus* and *Variimorphobacter*, causes a descent in the appetite values of the patients. On the other hand, increase of the *Gluceribacter*, *Petrocella*, *Lacrimispora*, *Tepidibaculum, Caloramator* and *Syntrophococcus* show a positive correlation with the appetite values (Table 7).

Table 5: Linear regression results for the methylmalonic acid contents of patients' urine and relative abundance of OTUs at the genus level. The coefficients show the contribution of the changes in the relative abundance of the respective bacteria. Positive values indicate that methylmalonic acid levels in urea increase as the respective bacteria is increasing while negative values show that appetite levels and the respective bacteria are negatively correlated. The p-value shows the importance of the respective bacteria while the significance column indicates the importance level of the bacteria in the created model. .: <0.1/not significant; *: <0.05/significant; **: <0.01/very significant; ***: <0.001/very very significant.

METHYLMALONIC ACID ~	**Coefficients**	**P-value**	**Significance**
(Intercept)	593.953	0.4333	
Lacrimispora	-117.954	0.3269	
Novisyntrophococcus	22016.38	0.1782	
Caloramator	-1285.37	0.2793	
Petrocella	-8934.21	0.06	.
Variimorphobacter	8921.133	0.2077	
Anaerosacchariphilus	458.563	0.2305	
Gluceribacter	-775.89	0.1638	
Hespellia	2.376	0.8264	
Bariatricus	18.837	0.0686	.
Tepidibaculum	75.734	0.5992	
Butyribacter	1187.073	0.0355	*
Desulforamulus	93.265	0.9766	
Intestinibacillus	-77.129	0.4847	
Hungatella	-16.539	0.1072	
Idiomarina	60.039	0.6409	
Nocardioides	154.538	0.619	
Enterocloster	19.693	0.4855	
Butyrivibrio	-3.564	0.8876	
Syntrophococcus	288.062	0.2368	
Lachnoanaerobaculum	-80.028	0.5701	

Table 6. Linear regression results for the methylmalonic acid contents of patients' urine and OTU relative abundance at species level. The coefficients show the contribution of the changes in the relative abundance of the respective bacteria. Positive values indicate that methylmalonic acid in urea increases as the respective bacteria is increasing while negative values show that methylmalonic acid and the respective bacteria are negatively correlated. The p-value shows the importance of the respective bacteria while the significance column indicates the importance level of the bacteria in the created model. *: <0.05/significant; **: <0.01/very significant; ***: <0.001/very very significant.

METHYLMALONIC ACID ~	**Coefficients**	**P-value**	**Significance**
(Intercept)	407.206	0.612	
Anaerosacchariphilus spp.	296.211	0.879	
Bariatricus massiliensis	720.198	0.401	
Blautia hydrogenotrophica	-16833.972	0.758	
Caloramator spp.	-1762.314	.576	
Gluceribacter spp.	-21.850	0.989	
Lacrimispora saccharolytica	90.524	0.977	
Lacrimispora spp.	-115.087	0.776	
Novisyntrophococcus spp.	8640.075	0.662	
Variimorphobacter spp.	2999.966	0.910	
Butyricicoccus pullicaecorum	-105.317	0.939	
Bariatricus spp.	34.519	0.004	**
Hespellia spp.	-15.205	0.272	
Tepidibaculum spp.	-134.417	0.942	
Anaerosacchariphilus polymeriproducens	-7002.199	0.913	
Hespellia porcina	-963.495	0.579	
Intestinibacillus spp.	-365.588	0.391	
Bacteroides pectinophilus	-123.232	0.582	

Table 7. Linear regression results for the appetites of patients and relative abundance of OTUs at the genus level. The coefficients show the contribution of the changes in the relative abundance of the respective bacteria. Positive values indicate that appetite levels increase as the respective bacteria is increasing while negative values show that appetite levels and the respective bacteria are negatively correlated. The p-value shows the importance of the respective bacteria while the significance column indicates the importance level of the bacteria in the created model. .: <0.1/not significant; *: <0.05/significant; **: <0.01/very significant; ***: <0.001/very very significant

APPETITE ~	**Coefficients**	**P-value**	**Significance**
(Intercept)	9.093694	1.67E-05	***
Lacrimispora	0.187254	0.00642	**
Novisyntrophococcus	-27.4337	0.00403	**
Caloramator	1.280314	0.01692	*
Petrocella	6.592673	0.00406	**
Variimorphobacter	-13.3633	0.00307	**
Anaerosacchariphilus	-0.77981	0.00256	**
Gluceribacter	1.131331	0.00223	**
Hespellia	0.006251	0.10358	
Bariatricus	-0.01333	0.00540	**
Tepidibaculum	0.202872	0.01033	*
Butyribacter	-0.59853	0.00658	**
Desulforamulus	0.590023	0.51215	
Intestinibacillus	0.061544	0.10172	
Hungatella	0.003447	0.17894	
Idiomarina	-0.18096	0.01053	*
Nocardioides	-0.68969	0.00282	**
Enterocloster	-0.01475	0.11657	
Butyrivibrio	0.014839	0.09984	.
Syntrophococcus	0.236326	0.02109	*
Lachnoanaerobaculum	-0.08924	0.07973	.

5 DISCUSSION

Creation of a 16S long-read bioinformatics pipeline for propionic acidemia patients was the primary aim of this thesis study. A database was developed for the reference-based alignment by adding newly discovered human gut related 16S sequences from the HMP to the NCBI 16S BLAST database. Additionally, consensus sequences were created to almost compensate for the low-quality reads obtained from ONT MinION and assigning the correct bacteria to the OTUs.

After the 4 steps of dietary treatment during the study, it was observed that the symptoms that the study cohort showed to have been alleviated. Since the aggregation of propionate derivatives in the human body is known to be the main reason for methylmalonic acidemia, decrease of propionate producing bacteria, therefore, the slight decrease may indicate that the dietary treatment worked by targeting the human gut microbiota [16].

The genus and species level MMA based regression model resulted with high adjusted R-squared values. For the genus level model, it was 0.9091 (p-value=0.02986), and species level model, 0.8945 (p-value=0.00252), which concludes that the models successfully fit both, the genus and species level OTU relative abundance and MMA concentration data sets. From the genus level model, it could be inferred that *Butyribacter* abundance was negatively impacted the patients' organic acidemia, due to the alongside increase of the MMA in the patient's urea with *Butyribacter*. Sahu *et al*. and Zou *et al.* both mentioned that this bacterial genus is capable of producing SCFAs [110, 111]. On the other hand, for the species level, the regression results indicate that since *Bariatricus spp.* is positively correlated with the MMA concentration in urea, *Bariatricus spp.* may show a negative effect on this metabolic disease. Being the only very significant (p=0.004) bacteria found after the regression modeling, *Bariatricus spp.* was not studied in-depth with regards to its possible effects on inherited metabolic defects. However, in a 2018 study, it was shown that *Bariatricus massiliensis* might take place in a histone deacetylase (HDAC) inhibition. Known for its involvement in various diseases, such as cancer,

neurodegeneration, cardiovascular diseases, etc., HDAC was shown to be inhibited more in presence of *Bariatricus massiliensis* [112]. Moreover, Manchia *et al.,* in a 2021 paper, discussed that an underrepresentation of *Bariatricus massiliensis*, along with other few gut bacteria, may increase an individual's susceptibility to schizophrenia [113]. Aside from that, relative abundance of *Bariatricus* at the genera level was found to be increased statistically significantly, after treatment with an herbal medicine (licorice) [114].

The genus level linear model focusing on the appetite levels, showed a better fit model than the models created for MMA concentrations (adjusted R-squared value=0.9756, p-value=0.004362). Many previous studies have shown that there are many species belonging to the Lachnospiraceae family capable of producing short-chained fatty acids (SCFAs) [110, 111, 115-124]. For instance, Koller et al emphasized that the Lachnospiraceae family is known for SCFA producing bacteria, capable of acetate, propionate and butyrate production, along with biogas generating [116]. One of the members of the Lachnospiraceae family, *Gluceribacter* was found positively correlated with the patients' appetite levels. Being one of the Firmicutes phylum, *Gluceribacter canis* species was tested for produced acids by Kawata *et al*. in 2018 [117]. It was shown that the species isolated from dogs produced long-Carbon-chain acids [117]. In 2022, Asanuma's study had focused on this species in mice with IBD, and a decrease in body weight and beneficial effects on the gut health was observed after dietary ceramide and *Gluceribacter* were given to the mice [118]. Being another member of Lachnospiraceae family, *Anaerosacchariphilus* was observed to have a negative correlation with the appetite levels. In a 2022 study, this genus was observed to have produced butyric acid, similar to the other Lachnospiraceae bacteria [115]. SCFA production by *Novisyntrophococcus, Variimorphobacter, Butyribacter, Nocardioides* and *Idiomarina* relative abundance values were all found correlated with SCFA concentrations, also were found to be negatively affecting the patient's appetite levels [119-129]. The genera not belonging to the Firmicutes phylum, a member of the Actinobacteria phylum, *Nocardioides* was reported to be found in the human gut in a 2015 paper, isolated from an intensive care unit patient and is found to be prevalent in obese, IBD, Crohn's disease patients

[130-131]. In another study, *Nocardioides* relative abundance values were found positively correlated significantly with acetate, butyrate and isovaleric acids [125]. Additionally, *Idiomarina* genus belongs to the Proteobacteria phylum. In a 2022 study, Idiomarina genus was found to have been more abundant in healthy control individuals compared to the breast cancer survivors [141].

The SCFAs, especially the ones known to be produced by the Lachnospiraceae family, butyrate and acetate, were shown to adversely affect the organic acidemia patients in two studies published in 2009 and 2014 via other metabolic pathways, for instance, not activating the ketogenesis pathway and creating acetyl-CoA [138, 139].

6 CONCLUSION

Last few years, human enteric microbiota has been found to be one of the most prominent indicators of major diseases, from inflammatory bowel disease (IBS) to colorectal cancer (CRC) and even some neurodegenerative diseases. Being one of the inherited metabolic disorders and organic acidemias, methylmalonic acidemia can cause various symptoms, such as vomiting, loss of appetite, hypotonia (loss of muscle tone), and more seriously, long-term possessors of the condition may experience delayed development (physical and mental), cardiac abnormalities, seizures, coma and even death. Showing its symptoms early after birth, methylmalonic acidemia is known to be a metabolic defect in which the body cannot process proteins and lipids properly. This study aimed to distinguish between the samples received from disease possessing individuals before and after treatment with a protein supplement through the changes in their enteric bacterial composition.

The created bioinformatics analysis pipeline and the linear regression model showed that *Bariatricus spp.* the distinct separation of the gut microbiota composition between the treatment steps. This species was found to be positively correlated with the methylmalonic acid concentrations in the urine obtained from the measurements from another study on the same patients. Due to high amount of OTUs found at the species level, another linear regression model was created that tried to find the MMA levels only depending on the genus level OTUs. Though, *Bariatricus* was not found significant, its change impacted the result by a small margin. On the other hand, *Butyribacter* was found to be statistically significant, and its change impacted the MMA levels positively. Hence, high *Butyribacter* levels indicated high MMA amounts in urea.

At the genus level, one more approach was implemented using the appetite survey (based on numerical values). Due to the drastic loss of appetite in methylmalonic acidemia patients, a linear regression model was developed that can predict the numerical value for the appetite of the patients depending only on the OTUs at the genus level. After the removal of the collinear OTUs, *Gluceribacter*,

Anaerosacchariphilus, *Nocardioides*, *Variimorphobacter*, *Novisyntrophococcus*, *Petrocella*, *Bariatricus*, *Lacrimispora*, *Butyribacter*, *Tepidibaculum*, *Idiomarina*, *Caloramator* and *Syntrophococcus* genera were found to have impact on the change in the appetite levels. Increase of these SCFA producer bacteria may exacerbate methylmalonic acidemia patients' symptoms.

Being a unique study that focused on dietary nutrient intake targeting the human intestinal bacterial composition in methylmalonic acidemia patients, the future studies could focus on both including more patients and, consequently, improving the bioinformatics analysis pipeline using a more varied sample set with allowing usage of more features for linear modelling. Also, the significant bacteria found were novel when compared to the past methylmalonic acidemia studies. Hence, the future studies may focus on these bacteria and their effect on the outset of organic acidemias.

7 REFERENCES

1. Shennar HK, Al-Asmar D, Kaddoura A, Al-Fahoum S. Diagnosis and clinical features of organic acidemias: A hospital-based study in a single center in Damascus, Syria. Qatar Medical Journal. 2015 Jul 4;2015(1):9.
2. Fraser JL, Venditti CP. Methylmalonic and propionic acidemias: clinical management update. Current opinion in pediatrics. 2016 Dec;28(6):682.
3. Schreiber J, Chapman KA, Summar ML, Mew NA, Sutton VR, MacLeod E, Stagni K, Ueda K, Franks J, Island E, Matern D. Neurologic considerations in propionic acidemia. Molecular genetics and metabolism. 2012 Jan 1;105(1):10-5.
4. Ugarte M, Pérez-Cerdá C, Rodríguez-Pombo P, Desviat LR, Pérez B, Richard E, Muro S, Campeau E, Ohura T, Gravel RA. Overview of mutations in the PCCA and PCCB genes causing propionic acidemia. Human mutation. 1999 Oct;14(4):275-82.
5. Upton AM, McKinney JD. Role of the methylcitrate cycle in propionate metabolism and detoxification in Mycobacterium smegmatis. Microbiology. 2007 Dec 1;153(12):3973-82.
6. Wongkittichote P, Mew NA, Chapman KA. Propionyl-CoA carboxylase–a review. Molecular genetics and metabolism. 2017 Dec 1;122(4):145-52. https://doi.org/10.1016/j.ymgme.2017.10.002
7. Reszko AE, Kasumov T, Pierce BA, David F, Hoppel CL, Stanley WC, Des Rosiers C, Brunengraber H. Assessing the reversibility of the anaplerotic reactions of the propionyl-CoA pathway in heart and liver. Journal of Biological Chemistry. 2003 Sep 12;278(37):34959-65.
8. Shchelochkov OA, Carrillo N, Venditti C. Propionic acidemia.
9. Baumgartner MR, Hörster F, Dionisi-Vici C, Haliloglu G, Karall D, Chapman KA, Huemer M, Hochuli M, Assoun M, Ballhausen D, Burlina A. Proposed guidelines for the diagnosis and management of methylmalonic and propionic acidemia. Orphanet journal of rare diseases. 2014 Dec;9(1):1-36.
10. Rosenblatt D, Watkins, D. Vitamin B12-responsive methylmalonic acidemia. Orphanet encyclopedia, May, 2012, https://www.orpha.net/consor/cgi-bin/Disease_Search.php?lng=EN&data_id=3260&Disease_Disease_Search_diseaseGroup=Methylmalonic-acidemia&Disease_Disease_Search_diseaseType=Pat&Disease(s)/group%20of%20diseases=Vitamin-B12-responsive-methylmalonic-acidemia&title=Vitamin%20B12-responsive%20methylmalonic%20acidemia&search=Disease_Search_Simple
11. Tanpaiboon, P. (2005). Methylmalonic acidemia (MMA). *Molecular genetics and metabolism*, *85*(1), 2-6.
12. Bikker H, Bakker HD, Abeling NG, Poll-The BT, Kleijer WJ, Rosenblatt DS, Waterham HR, Wanders RJ, Duran M. A homozygous nonsense mutation in the methylmalonyl-CoA

epimerase gene (MCEE) results in mild methylmalonic aciduria. Human mutation. 2006 Jul;27(7):640-3.

13. Gradinger AB, Bélair C, Worgan LC, Li CD, Lavallée J, Roquis D, Watkins D, Rosenblatt DS. Atypical methylmalonic aciduria: frequency of mutations in the methylmalonyl CoA epimerase gene (MCEE). Human mutation. 2007 Oct;28(10):1045-.
14. Froese DS, Kochan G, Muniz JR, Wu X, Gileadi C, Ugochukwu E, Krysztofinska E, Gravel RA, Oppermann U, Yue WW. Structures of the human GTPase MMAA and vitamin B12-dependent methylmalonyl-CoA mutase and insight into their complex formation. Journal of Biological Chemistry. 2010 Dec 3;285(49):38204-13.
15. Hörster F, Baumgartner MR, Viardot C, Suormala T, Burgard P, Fowler B, Hoffmann GF, Garbade SF, Kölker S, Baumgartner E. Long-term outcome in methylmalonic acidurias is influenced by the underlying defect (mut0, mut−, cblA, cblB). Pediatric research. 2007 Aug;62(2):225-30.
16. Manoli I, Sloan JL, Venditti CP. Isolated methylmalonic acidemia.
17. Gavrilov, D.K., Piazza, A.L., Pino, G., Turgeon, C., Matern, D., Oglesbee, D., Raymond, K., Tortorelli, S. and Rinaldo, P., 2020. The combined impact of CLIR post-analytical tools and second tier testing on the performance of newborn screening for disorders of propionate, methionine, and cobalamin metabolism. *International Journal of Neonatal Screening*, *6*(2), p.33.
18. Fowler, B., Leonard, J. V., & Baumgartner, M. R. (2008). Causes of and diagnostic approach to methylmalonic acidurias. *Journal of Inherited Metabolic Disease: Official Journal of the Society for the Study of Inborn Errors of Metabolism*, *31*(3), 350-360.
19. Aldubayan SH, Rodan LH, Berry GT, Levy HL. Acute Illness Protocol for Organic Acidemias: Methylmalonic Acidemia and Propionic Acidemia. Pediatr Emerg Care. 2017 Feb;33(2):142-146. doi: 10.1097/PEC.0000000000001028.
20. Coelho D, Suormala T, Stucki M, Lerner-Ellis JP, Rosenblatt DS, Newbold RF, Baumgartner MR, Fowler B. Gene identification for the cblD defect of vitamin B12 metabolism. New England Journal of Medicine. 2008 Apr 3;358(14):1454-64.
21. Lempp TJ, Suormala T, Siegenthaler R, Baumgartner ER, Fowler B, Steinmann B, Baumgartner MR. Mutation and biochemical analysis of 19 probands with mut0 and 13 with mut– methylmalonic aciduria: identification of seven novel mutations. Molecular genetics and metabolism. 2007 Mar 1;90(3):284-90.
22. Worgan LC, Niles K, Tirone JC, Hofmann A, Verner A, Sammak AA, Kucic T, Lepage P, Rosenblatt DS. Spectrum of mutations in mut methylmalonic acidemia and identification of a common Hispanic mutation and haplotype. Human mutation. 2006 Jan;27(1):31-43.
23. Acquaviva C, Benoist JF, Pereira S, Callebaut I, Koskas T, Porquet D, Elion J. Molecular basis of methylmalonyl-CoA mutase apoenzyme defect in 40 European patients affected by mut° and mut–forms of methylmalonic acidemia: Identification of 29 novel mutations in the MUT gene. Human mutation. 2005 Feb;25(2):167-76.

24. Acquaviva C, Benoist JF, Callebaut I, Guffon N, Ogier de Baulny H, Touati G, Aydin A, Porquet D, Elion J. N219Y, a new frequent mutation among mut° forms of methylmalonic acidemia in Caucasian patients. European Journal of Human Genetics. 2001 Aug;9(8):577-82.
25. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics. 2016 Jan 1;107(1):1-8.
26. Kchouk M, Gibrat JF, Elloumi M. Generations of sequencing technologies: from first to next generation. Biology and Medicine. 2017;9(3).
27. Janitz M, editor. Next-generation genome sequencing: towards personalized medicine. John Wiley & Sons; 2011 Aug 24.
28. Watts GS, Hurwitz BL. Metagenomic next-generation sequencing in clinical microbiology. Clinical Microbiology Newsletter. 2020 Apr 1;42(7):53-9.
29. Mignardi M, Nilsson M. Fourth-generation sequencing in the cell and the clinic. Genome medicine. 2014 Dec;6:1-4.Courtesy: National Human Genome Research Institute, https://www.genome.gov/genetics-glossary/Exon
30. Courtesy: National Human Genome Research Institute, https://www.genome.gov/genetics-glossary/Intron
31. Genetic Testing. https://www.cdc.gov/genomics/gtesting/genetic_testing.htm. June 14, 2023.
32. Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. Genome Res. 2013 Apr;23(4):628-37. doi: 10.1101/gr.146985.112. Epub 2012 Dec 26. PMID: 23269663; PMCID: PMC3613580.
33. Martin GT, Seymour DK, Gaut BS. CHH methylation islands: a nonconserved feature of grass genomes that is positively associated with transposable elements but negatively associated with gene-body methylation. Genome Biology and Evolution. 2021 Aug;13(8):evab144.
34. Dalakouras A, Dadami E, Zwiebel M, Krczal G, Wassenegger M. Transgenerational maintenance of transgene body CG but not CHG and CHH methylation. Epigenetics. 2012 Sep 7;7(9):1071-8.
35. Nakato R, Sakata T. Methods for ChIP-seq analysis: a practical workflow and advanced applications. Methods. 2021 Mar 1;187:44-53.
36. "The Nobel Prize in Chemistry 1993". *NobelPrize.org*.
37. Silverman JD, Bloom RJ, Jiang S, Durand HK, Dallow E, Mukherjee S, David LA. Measuring and mitigating PCR bias in microbiota datasets. PLoS computational biology. 2021 Jul 6;17(7):e1009113.
38. Jordan IK, Rogozin IB, Wolf YI, Koonin EV. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. Genome research. 2002 Jun 1;12(6):962-8.
39. Wang M, Lemos B. Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. Genome research. 2019 Mar 1;29(3):325-33.

40. Murray, R.G.E., Holt, J.G. (2005). The history of *Bergey's Manual*. In: Garrity, G.M., Boone, D.R. & Castenholz, R.W. (eds., 2001). *Bergey's Manual of Systematic Bacteriology*, 2nd ed., vol. 1, Springer-Verlag, New York, p. 1-14. link. [See p. 2.]
41. Nouioui I, Carro L, García-López M, Meier-Kolthoff JP, Woyke T, Kyrpides NC, Pukall R, Klenk HP, Goodfellow M, Göker M. Genome-based taxonomic classification of the phylum Actinobacteria. Frontiers in microbiology. 2018 Aug 22;9:2007.
42. Espejo RT, Plaza N. Multiple ribosomal RNA operons in bacteria; their concerted evolution and potential consequences on the rate of evolution of their 16S rRNA. Frontiers in microbiology. 2018 Jun 8;9:1232.
43. Ree HK, Zimmermann RA. Organization and expression of the 16S, 23S and 5S ribosomal RNA genes from the archaebacterium Thermoplasma acidophilum. Nucleic Acids Res. 1990 Aug 11;18(15):4471-8. doi: 10.1093/nar/18.15.4471. PMID: 1697064; PMCID: PMC331267.
44. Johnson A, Lewis J, ALBERTS B. Molecular biology of the cell.
45. Byrgazov K, Vesper O, Moll I. Ribosome heterogeneity: another level of complexity in bacterial translation regulation. Current opinion in microbiology. 2013 Apr 1;16(2):133-9.
46. Clarridge III JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clinical microbiology reviews. 2004 Oct;17(4):840-62.
47. Dos Santos HR, Argolo CS, Argôlo-Filho RC, Loguercio LL. A 16S rDNA PCR-based theoretical to actual delta approach on culturable mock communities revealed severe losses of diversity information. BMC microbiology. 2019 Dec;19(1):1-4.
48. Pearman WS, Freed NE, Silander OK. Testing the advantages and disadvantages of short-and long-read eukaryotic metagenomics using simulated reads. BMC bioinformatics. 2020 Dec;21(1):1-5.
49. Jeong J, Yun K, Mun S, Chung WH, Choi SY, Nam YD, Lim MY, Hong CP, Park C, Ahn YJ, Han K. The effect of taxonomic classification by full-length 16S rRNA sequencing with a synthetic long-read technology. Scientific reports. 2021 Jan 18;11(1):1727.
50. Brown CG, Clarke J. Nanopore development at Oxford nanopore. Nature biotechnology. 2016 Aug;34(8):810-1.
51. Nygaard AB, Tunsjø HS, Meisal R, Charnock C. A preliminary study on the potential of Nanopore MinION and Illumina MiSeq 16S rRNA gene sequencing to characterize building-dust microbiomes. Scientific Reports. 2020 Feb 21;10(1):1-0.
52. Matsuo Y, Komiya S, Yasumizu Y, Yasuoka Y, Mizushima K, Takagi T, Kryukov K, Fukuda A, Morimoto Y, Naito Y, Okada H. Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution. BMC microbiology. 2021 Dec;21(1):1-3.
53. Baker M. De novo genome assembly: what every biologist should know. Nature methods. 2012 Apr;9(4):333-7.

54. Cole, J. R., Q. Wang, J. A. Fish, B. Chai, D. M. McGarrell, Y. Sun, C. T. Brown, A. Porras-Alfaro, C. R. Kuske, and J. M. Tiedje. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis Nucl. Acids Res. 42(Database issue):D633-D642; doi: 10.1093/nar/gkt1244 [PMID: 24288368]
55. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and environmental microbiology. 2006 Jul;72(7):5069-72.
56. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. The SILVA and "all-species living tree project (LTP)" taxonomic frameworks. Nucleic acids research. 2014 Jan 1;42(D1):D643-8.
57. Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, Mcveigh R, O'Neill K, Robbertse B, Sharma S. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database. 2020 Jan 1;2020.
58. Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. GenBank. Nucleic acids research. 2019 Jan 8;47(D1):D94-9.
59. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome biology. 2019 Dec;20:1-0.
60. Xu Z, Mai Y, Liu D, He W, Lin X, Xu C, Zhang L, Meng X, Mafofo J, Zaher WA, Koshy A. Fast-bonito: a faster deep learning based basecaller for nanopore sequencing. Artificial Intelligence in the Life Sciences. 2021 Dec 1;1:100011.
61. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
62. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States); 2014 Mar 17.
63. Boratyn GM, Thierry-Mieg J, Thierry-Mieg D, Busby B, Madden TL. Magic-BLAST, an accurate RNA-seq aligner for long and short reads. BMC bioinformatics. 2019 Dec;20(1):1-9.
64. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9.
65. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC bioinformatics. 2009 Dec;10:1-9.
66. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, Ravel J. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome. 2014 Dec;2(1):1-7.
67. Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. PloS one. 2014 Apr 10;9(4):e94249.

68. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016 Jul;13(7):581–3.
69. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature biotechnology. 2019 Aug;37(8):852-7.
70. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. Deblur rapidly resolves single-nucleotide community sequence patterns. MSystems. 2017 Apr 21;2(2):e00191-16.
71. Baker GC, Smith JJ, Cowan DA. Review and re-analysis of domain-specific 16S primers. Journal of microbiological methods. 2003 Dec 1;55(3):541-55.
72. Shapiro S. The Shapiro-Wilk and related test for normality. Statistics (Ber). 2015.
73. ĐUROVIĆ ŽM, KOVAČEVIĆ BD. QQ-plot approach to robust Kalman filtering. International Journal of Control. 1995 Apr 1;61(4):837-57.
74. Xia Y. Correlation and association analyses in microbiome study integrating multiomics in health and disease. Progress in Molecular Biology and Translational Science. 2020 Jan 1;171:309-491.
75. St L, Wold S. Analysis of variance (ANOVA). Chemometrics and intelligent laboratory systems. 1989 Nov 1;6(4):259-72.
76. Jafari M, Ansari-Pour N. Why, when and how to adjust your P values?. Cell Journal (Yakhteh). 2019;20(4):604.
77. Freedman DA. Statistical models: theory and practice. cambridge university press; 2009 Apr 27.
78. Ke C, Wang Y. Smoothing spline nonlinear nonparametric regression models. Journal of the American Statistical Association. 2004 Dec 1;99(468):1166-75.
79. Freedman DA. Statistical models: theory and practice. Cambridge university press; 2009 Apr 27.
80. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one. 2013 Apr 22;8(4):e61217.
81. Morris EK, Caruso T, Buscot F, Fischer M, Hancock C, Maier TS, Meiners T, Müller C, Obermaier E, Prati D, Socher SA. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. Ecology and evolution. 2014 Sep;4(18):3514-24.
82. Whittaker RH. Vegetation of the Siskiyou mountains, Oregon and California. Ecological monographs. 1960 Jul 1;30(3):279-338.
83. Bray JR, Curtis J. T.(1957). An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr.;27:325-49.
84. Jaccard P. The distribution of the flora in the alpine zone. 1. New phytologist. 1912 Feb;11(2):37-50.

85. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Applied and environmental microbiology. 2007 Mar 1;73(5):1576-85.
86. Sender R, Fuchs S, Milo R. Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. Cell. 2016 Jan 28;164(3):337-40.
87. Thursby E, Juge N. Introduction to the human gut microbiota. Biochemical journal. 2017 Jun 1;474(11):1823-36.
88. Gargaud M, Amils R, editors. Encyclopedia of astrobiology. Springer Science & Business Media; 2011 May 26.
89. Ellis SR, Nguyen M, Vaughn AR, Notay M, Burney WA, Sandhu S, Sivamani RK. The skin and gut microbiome and its role in common dermatologic conditions. Microorganisms. 2019 Nov;7(11):550.
90. Wexler HM. Bacteroides: the good, the bad, and the nitty-gritty. Clinical microbiology reviews. 2007 Oct;20(4):593-621.
91. Yu Y, Zhu S, Li P, Min L, Zhang S. Helicobacter pylori infection and inflammatory bowel disease: a crosstalk between upper and lower digestive tract. Cell death & disease. 2018 Sep 20;9(10):961.
92. Das S, Dash HR, editors. Microbial diversity in the genomic era. Academic Press; 2018 Sep 20.
93. Parada Venegas D, De la Fuente MK, Landskron G, González MJ, Quera R, Dijkstra G, Harmsen HJ, Faber KN, Hermoso MA. Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. Frontiers in immunology. 2019:277.
94. Hosseini E, Grootaert C, Verstraete W, Van de Wiele T. Propionate as a health-promoting microbial metabolite in the human gut. Nutrition reviews. 2011 May 1;69(5):245-58.
95. Blaak EE, Canfora EE, Theis S, Frost G, Groen AK, Mithieux G, Nauta A, Scott K, Stahl B, Van Harsselaar J, van Tol R. Short chain fatty acids in human gut and metabolic health. Beneficial microbes. 2020 Sep 1;11(5):411-55.
96. Asadpoor M, Ithakisiou GN, Henricks PA, Pieters R, Folkerts G, Braber S. Non-digestible oligosaccharides and short chain fatty acids as therapeutic targets against enterotoxin-producing bacteria and their toxins. Toxins. 2021 Feb 25;13(3):175.
97. Ríos-Covián D, Ruas-Madiedo P, Margolles A, Gueimonde M, De Los Reyes-gavilán CG, Salazar N. Intestinal short chain fatty acids and their link with diet and human health. Frontiers in microbiology. 2016 Feb 17;7:185.
98. Dalile B, Van Oudenhove L, Vervliet B, Verbeke K. The role of short-chain fatty acids in microbiota–gut–brain communication. Nature reviews Gastroenterology & hepatology. 2019 Aug;16(8):461-78.
99. van der Hee B, Wells JM. Microbial regulation of host physiology by short-chain fatty acids. Trends in Microbiology. 2021 Aug 1;29(8):700-12.

100. Markowiak-Kopeć P, Śliżewska K. The effect of probiotics on the production of short-chain fatty acids by human intestinal microbiome. Nutrients. 2020 Apr 16;12(4):1107.
101. LeBlanc JG, Chain F, Martín R, Bermúdez-Humarán LG, Courau S, Langella P. Beneficial effects on host energy metabolism of short-chain fatty acids and vitamins produced by commensal and probiotic bacteria. Microbial cell factories. 2017 Dec;16(1):1-0.
102. Louis P, Flint HJ. Formation of propionate and butyrate by the human colonic microbiota. Environmental microbiology. 2017 Jan;19(1):29-41.
103. Ratajczak W, Rył A, Mizerski A, Walczakiewicz K, Sipak O, Laszczyńska M. Immunomodulatory potential of gut microbiome-derived short-chain fatty acids (SCFAs). Acta Biochimica Polonica. 2019 Mar 4;66(1):1-2.
104. Silva YP, Bernardi A, Frozza RL. The role of short-chain fatty acids from gut microbiota in gut-brain communication. Frontiers in endocrinology. 2020 Jan 31;11:25.
105. Reichardt, N., Duncan, S. H., Young, P., Belenguer, A., McWilliam Leitch, C., Scott, K. P., Flint, H. J., & Louis, P. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. The ISME journal, 8(6). 2014; 1323–1335. https://doi.org/10.1038/ismej.2014.14
106. El Hage, R., Hernandez-Sanabria, E., Calatayud Arroyo, M., Props, R., & Van de Wiele, T. Propionate-producing consortium restores antibiotic-induced dysbiosis in a dynamic in vitro model of the human intestinal microbial ecosystem. Frontiers in microbiology. 2019; 1206.
107. Aryal S, Alimadadi A, Manandhar I, Joe B, Cheng X. Machine learning strategy for gut microbiome-based diagnostic screening of cardiovascular disease. Hypertension. 2020 Nov;76(5):1555-62.
108. Beura S, Kundu P, Das AK, Ghosh A. Metagenome-scale community metabolic modelling for understanding the role of gut microbiota in human health. Computers in Biology and Medicine. 2022 Aug 19:105997.
109. Deodato, F., Boenzi, S., Santorelli, F. M., & Dionisi-Vici, C. (2006, May). Methylmalonic and propionic aciduria. In *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* (Vol. 142, No. 2, pp. 104-112). Hoboken: Wiley Subscription Services, Inc., A Wiley Company.
110. Sahu S, Kaushik SR, Chaudhary S, MOHAPATRA AK, Kappa R, Kapfo W, Saha S, Das R, Das A, Khamo V, NANDA R. Gut microbiota dysbiosis observed in tuberculosis patients resolves partially with anti-tuberculosis therapy. medRxiv. 2023:2023-06.
111. Zou Y, Xue W, Lin X, Lv M, Luo G, Dai Y, Sun H, Liu SW, Sun CH, Hu T, Xiao L. Butyribacter intestini gen. nov., sp. nov., a butyric acid-producing bacterium of the family Lachnospiraceae isolated from human faeces, and reclassification of Acetivibrio ethanolgignens as Acetanaerobacter ethanolgignens gen. nov., comb. nov. Systematic and Applied Microbiology. 2021 May 1;44(3):126201.

112. Yuille S, Reichardt N, Panda S, Dunbar H, Mulder IE. Human gut bacteria as potent class I histone deacetylase inhibitors in vitro through production of butyric acid and valeric acid. Nie D, editor. PLOS ONE [Internet]. 2018 Jul 27 [cited 2019 Aug 4];13(7):e0201073.
113. Manchia M, Fontana A, Panebianco C, Paribello P, Arzedi C, Cossu E, et al. Involvement of Gut Microbiota in Schizophrenia and Treatment Resistance to Antipsychotics. Biomedicines. 2021 Jul 23;9(8):875.
114. Peterson CT, Sharma V, Uchitel S, Denniston K, Chopra D, Mills PJ, et al. Prebiotic Potential of Herbal Medicines Used in Digestive Health and Disease. Journal of Alternative and Complementary Medicine [Internet]. 2018 Jul 1;24(7):656–65.
115. Abdugheni R, Wang WZ, Wang YJ, Du MX, Liu FL, Zhou N, Jiang CY, Wang CY, Wu L, Ma J, Liu C. Metabolite profiling of human-originated Lachnospiraceae at the strain level. iMeta. 2022 Dec;1(4):e58.
116. Köller N, Hahnke S, Zverlov V, Wibberg D, Klingl A, Busche T, Klocke M, Pühler A, Schlüter A, Liebl W, Maus I. Anaeropeptidivorans aminofermentans gen. nov., sp. nov., a mesophilic proteolytic salt-tolerant bacterium isolated from a laboratory-scale biogas fermenter, and emended description of Clostridium colinum. International Journal of Systematic and Evolutionary Microbiology. 2022 Dec 20;72(12):005668.
117. Kawata M, Tsukamoto A, Isozaki R, Nobukawa S, Kawahara N, Akutsu S, Suzuki M, Asanuma N. Glucerabacter canisensis gen. nov., sp. nov., isolated from dog feces and its effect on the hydrolysis of plant glucosylceramide in the intestine of dogs. Archives of microbiology. 2018 Apr;200:505-15.
118. Asanuma N. Effect of dietary ceramide and glucosylceramide on the alleviation of experimental inflammatory bowel disease in mice. Journal of oleo science. 2022;71(9):1397-402.
119. Doré J, Bryant MP. Metabolism of one-carbon compounds by the ruminal acetogen Syntrophococcus sucromutans. Applied and environmental microbiology. 1990 Apr;56(4):984-9.
120. Bordugo A, Salvetti E, Rodella G, Piazza M, Dianin A, Amoruso A, Piacentini G, Pane M, Torriani S, Vitulo N, Felis GE. Assessing gut microbiota in an infant with congenital propionic acidemia before and after probiotic supplementation. Microorganisms. 2021 Dec 16;9(12):2599.
121. Fan H, Li J, Wu W, Chen R, Yang M, Zhang Y, Cong L, Dai L, Deng Y, Cheng L, Ma S. Description of a moderately acidotolerant and aerotolerant anaerobic bacterium Acidilutibacter cellobiosedens gen. nov., sp. nov. within the family Acidilutibacteraceae fam. nov., and proposal of Sporanaerobacteraceae fam. nov. and Tepidimicrobiaceae fam. nov. Systematic and Applied Microbiology. 2023 Jan 1;46(1):126376.
122. Chai LJ, Fang GY, Xu PX, Zhang XJ, Lu ZM, Zhang SY, Wang ST, Shen CH, Shi JS, Xu ZH. Novisyntrophococcus fermenticellae gen. nov., sp. nov., isolated from an anaerobic

fermentation cellar of Chinese strong-flavour baijiu. International Journal of Systematic and Evolutionary Microbiology. 2021 Sep 9;71(9):004991.

123. Lin CJ, Cheng YC, Chen HC, Chao YK, Nicholson MW, Yen EC, Kamp TJ, Hsieh PC. Commensal gut microbiota-derived acetate and propionate enhance heart adaptation in response to cardiac pressure overload in mice. Theranostics. 2022;12(17):7319.
124. Holst AQ, Jois H, Laursen MF, Sommer MO, Licht TR, Bahl MI. Human milk oligosaccharides induce acute yet reversible compositional changes in the gut microbiota of conventional mice linked to a reduction of butyrate levels. Microlife. 2022;3:uqac006.
125. Liu S, Yu H, Li P, Wang C, Liu G, Zhang X, Zhang C, Qi M, Ji H. Dietary nano-selenium alleviated intestinal damage of juvenile grass carp (Ctenopharyngodon idella) induced by high-fat diet: Insight from intestinal morphology, tight junction, inflammation, anti-oxidization and intestinal microbiota. Animal Nutrition. 2022 Mar 1;8:235-48.
126. Park YH, Lee DH, Lee YS, Jung JS, Kahng HY. Idiomarina taeanensis sp. nov., a Novel Marine Bacterium Isolated from Crude Oil-Contaminated Seawater. 한국미생물학회 학술대회논문집. 2010 May:155-.
127. Daroonpunt R, Tanaka N, Uchino M, Tanasupawat S. Characterization and screening of lipolytic bacteria from Thai fermented fish. Sains Malaysiana. 2018 Jan 1;47(1):91-7.
128. Xia J, Lv L, Liu B, Wang S, Zhang S, Wu Z, Yang L, Bian X, Wang Q, Wang K, Zhuge A. Akkermansia muciniphila ameliorates acetaminophen-induced liver injury by regulating gut microbial composition and metabolism. Microbiology spectrum. 2022 Feb 23;10(1):e01596-21.
129. Yang Y, Zhao X, Xie Y, Wu C. Modulative effect of Physalis alkekengi on both gut bacterial and fungal micro-ecosystem. Chinese Herbal Medicines. 2023 May 9.
130. Lagier JC, Hugon P, Khelaifia S, Fournier PE, La Scola B, Raoult D. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clinical microbiology reviews. 2015 Jan;28(1):237-64.
131. Mandal RS, Saha S, Das S. Metagenomic surveys of gut microbiota. Genomics, proteomics & bioinformatics. 2015 Jun 1;13(3):148-58.
132. Hwang YJ, Son JS, Lee SY, He Y, Jo Y, Shin JH, Ghim SY. Nocardioides sambongensis sp. nov., isolated from Dokdo Islands soil. International Journal of Systematic and Evolutionary Microbiology. 2020 Jan;70(1):16-22.
133. Sharpton TJ. An introduction to the analysis of shotgun metagenomic data. Frontiers in plant science. 2014 Jun 16;5:209.
134. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nature biotechnology. 2017 Sep;35(9):833-44.
135. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome biology. 2017 Dec;18(1):1-5.

136. Colonetti K, Roesch LF, Schwartz IV. The microbiome and inborn errors of metabolism: Why we should look carefully at their interplay?. Genetics and molecular biology. 2018 Jul;41:515-32.
137. Bordugo A, Salvetti E, Rodella G, Piazza M, Dianin A, Amoruso A, Piacentini G, Pane M, Torriani S, Vitulo N, Felis GE. Assessing gut microbiota in an infant with congenital propionic acidemia before and after probiotic supplementation. Microorganisms. 2021 Dec 16;9(12):2599.
138. Tims S, Marsaux C, Pinto A, Daly A, Karall D, Kuhn M, Santra S, Roeselers G, Knol J, MacDonald A, Scholl-Bürgi S. Altered gut microbiome diversity and function in patients with propionic acidemia. Molecular Genetics and Metabolism. 2022 Nov 1;137(3):308-22.
139. Saito T, Saito O, Maeda T, Ito C, Ando Y, Yamagata T, Muto S, Momoi M, Kusano E. Metabolic and hemodynamic advantages of an acetate-free citrate dialysate in a uremic case of congenital methylmalonic acidemia. American journal of kidney diseases. 2009 Oct 1;54(4):764-9.
140. Baumgartner MR, Hörster F, Dionisi-Vici C, Haliloglu G, Karall D, Chapman KA, Huemer M, Hochuli M, Assoun M, Ballhausen D, Burlina A. Proposed guidelines for the diagnosis and management of methylmalonic and propionic acidemia. Orphanet journal of rare diseases. 2014 Dec;9(1):1-36.
141. Caleça T, Ribeiro P, Vitorino M, Menezes M, Sampaio-Alves M, Mendes AD, Vicente R, Negreiros I, Faria A, Costa DA. Breast Cancer Survivors and Healthy Women. Cancers. 2023;15(3).

8 APPENDIX

8.1 Appendix 1

Appendix Table 1: Abundance and statistical values for observed gut bacteria at phylum level. Relative abundance for 4 different treatment steps, significance levels, Kruskal-Wallis p-value, medians, means, minimum and maximum values for the treatment steps of each organism are represented below.

Organism	S101	S102	S103	S104	S105
Acidobacteria	0.002	0.005	0	0	0.003
Actinobacteria	0.27	0.524	0.02	0.083	0.064
Aquificae	0.005	0.003	0	0.002	0
Bacteroidota	45.538	1.088	5.787	0.712	5.841
Balneolaeota	0	0	0	0	0
Caldiserica	0	0.003	0	0	0
Calditrichaeota	0	0	0	0	0
Candidatus Melainabacteria	0	0	0	0	0
Chlamydiae	0	0.003	0	0	0
Chlorobi	0	0.003	0	0	0
Chloroflexi	0.005	0.008	0	0.002	0
Chrysiogenetes	0	0.003	0	0	0
Cyanobacteria	0.013	0.018	0.007	0.012	0.078
Deferribacteres	0	0	0	0	0
Deinococcus-Thermus	0	0.003	0	0	0
Fibrobacteres	0	0	0	0	0.003
Firmicutes	42.417	86.332	91.437	96.379	89.666
Fusobacteria	0.002	0.01	0.002	0.005	0
Gemmatimonadetes	0	0	0	0	0
Ignavibacteriae	0	0	0	0	0
Kiritimatiellaeota	0.002	0	0	0	0
Lentisphaerae	0	0	0	0	0
Nitrospinae	0	0	0	0	0
Proteobacteria	11.738	11.918	2.748	2.803	4.026
Rhodothermaeota	0	0	0	0	0.003
Spirochaetes	0.002	0.003	0	0.002	0
Synergistetes	0	0.038	0	0	0.043
Tenericutes	0.005	0.035	0	0	0.003
Thermotogae	0.002	0.005	0	0	0.014
Verrucomicrobia	0	0.003	0	0	0.255

Appendix Table 1 continued

S106	S107	S108	S201	S202	S203	S204
0	0.002	0	0.005	0	0	0
0.242	0.099	0.037	0.265	0.006	2.495	0.12
0	0.003	0	0.002	0	0	0
6.758	5.713	2.787	30.793	1.218	6.845	1.061
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0.002	0	0	0
0	0	0	0.476	0	0	0
0	0	0	0.002	0	0	0
0	0	0.004	0.005	0	0	0
0.002	0	0	0.002	0	0	0
0	0	0	0.002	0	0	0
0.011	0.021	0.004	0.52	0	0	0
0	0	0	0.002	0	0	0
0	0	0	0.007	0	0	0
0	0.002	0	0.003	0	0	0
90.778	93.336	68.044	48.614	49.134	85.905	88.233
0.062	0	0	0.002	0.143	0.042	0
0	0	0	0	0	0	0
0	0	0	0.002	0	0	0
0	0	0	0	0	0	0
0	0	0	0.008	0	0	0
0	0.002	0	0.003	0	0	0
2.126	0.59	29.081	19.256	49.499	4.684	10.585
0	0	0	0	0	0	0
0.002	0	0	0	0	0	0
0.008	0.002	0.007	0.02	0	0	0
0.005	0	0.007	0.01	0	0.014	0
0.006	0	0	0	0	0	0
0	0.231	0.029	0	0	0.014	0

Appendix Table 1 continued

S205	S206	S207	S208	S301	S302	S303
0	0	0	0	0	0	0.006
0.051	0.165	0.14	0.199	0.331	0.012	0.102
0	0	0	0	0	0	0
5.968	4.536	3.183	7.888	11.389	0.993	7.735
0	0	0.007	0	0	0.004	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	1.01	0	0
0	0	0	0	0	0	0.003
0	0	0	0	0	0	0.006
0	0	0	0	0	0	0.003
0	0	0	0	0	0	0.003
0	0	0	0	0.331	0.004	0.011
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0.003	0	0	0	0
93.486	85.8	90.357	84.888	79.144	72.467	88.167
0	0.012	0	0	0.045	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0.011
0	0	0	0	0	0	0
0	0	0	0	1.01	0	0
0	0	0	0	0	0	0
0.484	9.483	6.303	7.025	6.732	26.513	3.945
0	0	0	0	0	0	0
0	0	0	0	0	0	0.003
0	0	0	0	0	0.004	0
0	0.004	0.007	0	0.009	0.004	0.003
0	0	0	0	0	0	0.003
0.01	0	0	0	0	0	0

Appendix Table 1 continued

S304	S305	S306	S307	S308	S401	S402
0	0	0	0	0	0.004	0
0.261	0.172	1.61	0.193	0.269	0.135	0.022
0	0	0	0	0	0	0
4.69	9.904	25.263	6.617	4.643	2.478	0.014
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0.006	0.004	0
0	0	0	0	0	0	0
0	0	0	0	0.018	0	0.002
0	0	0	0	0.006	0	0
0	0	0	0	0	0.004	0
0	0	0	0	0	0	0
86.186	89.203	67.226	92.12	73.801	43.771	47.622
0	0	0	0.009	0	0	0
0	0	0	0	0.006	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
8.862	0.616	5.662	1.062	21.24	53.598	52.338
0	0	0	0	0	0	0
0	0	0	0	0.006	0	0
0	0.005	0	0	0.006	0.004	0
0	0	0	0	0	0.004	0.002
0	0	0	0	0	0	0
0	0.1	0.24	0	0	0	0

Appendix Table 1 continued

S403	S404	S405	S406	S407	S408	significance
0.002	0	0	0	0	0	N
0.053	0.034	0.05	0.028	0.05	0.071	N
0.002	0	0	0	0	0	Y *
5.912	0.072	6.514	1.202	0.038	13.831	N
0	0	0	0	0	0	N
0	0	0	0	0	0	N
0	0	0	0	0	0	N
0	0	0	0	0	0	N
0	0	0	0	0	0	N
0	0	0	0	0	0	N
0	0	0	0	0	0	N
0	0	0	0	0	0	N
0.002	0.004	0	0	0.009	0.003	Y **
0	0	0	0	0	0	N
0	0	0	0	0	0	N
0	0	0	0	0	0	N
63.109	70.575	92.578	61.208	65.036	70.119	N
0	0	0	0	0.002	0.003	N
0	0	0	0	0	0	N
0	0	0	0	0	0	N
0	0	0	0	0	0	N
0	0	0	0	0.002	0	N
0	0	0	0	0.002	0	N
30.915	29.315	0.858	37.563	34.859	15.967	Y *
0	0	0	0	0	0.003	N
0	0	0	0	0	0	Y *
0	0	0	0	0.002	0	N
0	0	0	0	0.002	0	N
0	0	0	0	0	0.003	N
0.004	0	0	0	0	0	N

Appendix Table 1 continued

Kruskal_ p-value	S1_median	S2_median	S3_median	S4_median	S1_mean	S2_mean
0.41147	0.001	0	0	0	0.0015	0.000625
0.05673	0.091	0.1525	0.227	0.05	0.167375	0.430125
0.04614	0.001	0	0	0	0.001625	0.00025
0.26798	5.75	5.252	7.176	1.84	9.278	7.6865
0.5583	0	0	0	0	0	0.000875
0.39163	0	0	0	0	0.000375	0
0.39163	0	0	0	0	0	0.00025
0.5583	0	0	0	0	0	0.0595
0.78311	0	0	0	0	0.000375	0.00025
0.58186	0	0	0	0	0.000875	0.000625
0.26583	0.001	0	0	0	0.002125	0.00025
0.78311	0	0	0	0	0.000375	0.00025
0.00912	0.0125	0	0.002	0.002	0.0205	0.065
0.5583	0	0	0	0	0	0.00025
0.78263	0	0	0	0	0.000375	0.000875
0.21932	0	0	0	0	0.000625	0.00075
0.09425	90.222	85.8525	82.665	64.0725	82.29863	78.30213
0.33702	0.002	0.001	0	0	0.010125	0.024875
0.39163	0	0	0	0	0	0
0.5583	0	0	0	0	0	0.00025
0.39163	0	0	0	0	0.00025	0
0.78263	0	0	0	0	0	0.001
0.78311	0	0	0	0	0.00025	0.000375
0.03033	3.4145	8.254	6.197	32.887	8.12875	13.41488
0.55869	0	0	0	0	0.000375	0
0.04544	0.001	0	0	0	0.001125	0
0.12717	0.0045	0	0	0	0.01225	0.0025
0.37579	0.004	0.002	0	0	0.006875	0.004375
0.05349	0.001	0	0	0	0.003375	0
0.34245	0.0015	0	0	0	0.06475	0.003

Appendix Table 1 continued

S3_mean	S4_mean	S1_min	S2_min	S3_min	S4_min	S1_max
0.00075	0.00075	0	0	0	0	0.005
0.36875	0.055375	0.02	0.006	0.012	0.022	0.524
0	0.00025	0	0	0	0	0.005
8.90425	3.757625	0.712	1.061	0.993	0.014	45.538
0.0005	0	0	0	0	0	0
0	0	0	0	0	0	0.003
0	0	0	0	0	0	0
0.12625	0	0	0	0	0	0
0.000375	0	0	0	0	0	0.003
0.00075	0	0	0	0	0	0.004
0.001125	0.0005	0	0	0	0	0.008
0.000375	0	0	0	0	0	0.003
0.0455	0.0025	0.004	0	0	0	0.078
0.00075	0	0	0	0	0	0
0	0.0005	0	0	0	0	0.003
0	0	0	0	0	0	0.003
81.03925	64.25225	42.417	48.614	67.226	43.771	96.379
0.00675	0.000625	0	0	0	0	0.062
0.00075	0	0	0	0	0	0
0.001375	0	0	0	0	0	0
0	0	0	0	0	0	0.002
0.12625	0.00025	0	0	0	0	0
0	0.00025	0	0	0	0	0.002
9.329	31.92663	0.59	0.484	0.616	0.858	29.081
0	0.000375	0	0	0	0	0.003
0.001125	0	0	0	0	0	0.003
0.001875	0.00075	0	0	0	0	0.043
0.002	0.001	0	0	0	0	0.035
0.000375	0.000375	0	0	0	0	0.014
0.0425	0.0005	0	0	0	0	0.255

Appendix Table 1 continued

S2_max	S3_max	S4_max
0.005	0.006	0.004
2.495	1.61	0.135
0.002	0	0.002
30.793	25.263	13.831
0.007	0.004	0
0	0	0
0.002	0	0
0.476	1.01	0
0.002	0.003	0
0.005	0.006	0
0.002	0.006	0.004
0.002	0.003	0
0.52	0.331	0.009
0.002	0.006	0
0.007	0	0.004
0.003	0	0
93.486	92.12	92.578
0.143	0.045	0.003
0	0.006	0
0.002	0.011	0
0	0	0
0.008	1.01	0.002
0.003	0	0.002
49.499	26.513	53.598
0	0	0.003
0	0.006	0
0.02	0.006	0.004
0.014	0.009	0.004
0	0.003	0.003
0.014	0.24	0.004

9 CURRICULUM VITAE