T.C.
DOKUZ EYLÜL UNIVERSITY
IZMIR INTERNATIONAL BIOMEDICINE AND GENOMEINSTITUTE

# DEVELOPMENT OF AN APPLICATION PROGRAMMING INTERFACE FOR DYNAMICALLY ESTIMATING REFERENCE INTERVALS FOR FUNCTIONAL ABUNDANCES OF GUT MICROBIOTA

LEMAN BİNOKAY

DEPARTMENT OF BIOMEDICINE AND HEALTH TECHNOLOGIES

**MASTER OF SCIENCE THESIS**

İZMİR – 2023

THESIS CODE: DEU.IBG.MSc-2021850004

T.C.
DOKUZ EYLÜL UNIVERSITY
IZMIR INTERNATIONAL BIOMEDICINE AND GENOME INSTITUTE

# DEVELOPMENT OF AN APPLICATION PROGRAMMING INTERFACE FOR DYNAMICALLY ESTIMATING REFERENCE INTERVALS FOR FUNCTIONAL ABUNDANCES OF GUT MICROBIOTA

DEPARTMENT OF BIOMEDICINE AND HEALTH TECHNOLOGIES

**MASTER OF SCIENCE THESIS**

LEMAN BİNOKAY

SUPERVISOR: ASSOC. PROF. DR. GÖKHAN KARAKÜLAH
CO-SUPERVISOR: ASST. PROF. DR. YAVUZ OKTAY

THESIS CODE: DEU.IBG.MSc-2021850004

Dokuz Eylül Üniversitesi İzmir Uluslararası Biyotıp ve Genom Enstitüsü Biyotıp ve Sağlık Teknolojileri Anabilim Dalı,

Biyotıp ve Sağlık Teknolojileri Yüksek Lisans programı öğrencisi Leman Binokay

**"Bağırsak Mikrobiyotası Fonksiyonel Bolluğu Referens Aralıklarının Dinamik Kestirimi İçin Bir Uygulama Programlama Arayüzü Geliştirilmesi"** konulu

Yüksek Lisans tezini 19 / 01 / 2023 tarihinde başarılı olarak tamamlamıştır.

Doç. Dr. Gökhan Karakülah
Dokuz Eylül Üniversitesi
BAŞKAN

Dr. Öğr. Üye. Cihangir Yandım
İzmir Ekonomi Üniversitesi
ÜYE

Doç.Dr. Evren Koban Baştanlar
Ege Üniversitesi
ÜYE

Dokuz Eylül University İzmir International Biomedicine and Genome Enstitute

Department of Biomedicine and Health Technologies,

Biomedicine and Health Technologies graduate program Master of Science student Leman Binokay has successfully completed her Master of Science thesis titled

**'Development Of An Application Programming Interface For Dynamically Estimating Reference Intervals For Functional Abundances Of Gut Microbiota'** on the date

of 19 / 01 / 2023.

Assoc. Prof. Dr. Gökhan Karakülah

Dokuz Eylül University

CHAIR

Asst. Prof. Dr. Cihangir Yandım

İzmir University of Economics

MEMBER

Assoc. Prof. Dr. Evren Koban Baştanlar

Ege University

MEMBER

**TABLE OF CONTENTS: Page Number:**

**TABLE OF CONTENTS ..... i**

**INDEX OF TABLES ..... v**

**INDEX OF FIGURES ..... viii**

**LIST OF ABBREVIATIONS ..... xii**

**ACKNOWLEDGEMENTS ..... xiii**

**ABSTRACT ..... 1**

**ÖZET ..... 2**

**1. INTRODUCTION AND AIM ..... 4**

**1.1. Statement and Importance of the Problem ..... 4**

**1.2. Aim of the Study ..... 4**

**1.3. The hypothesis of the Study… ..... 4**

**2. GENERAL INFORMATION ..... 5**

**2.1. Human Microbiota ..... 6**

*2.1.1. Human microbiota and health ..... 6*

*2.1.2. Gut microbiota and metabolic pathways ..... 6*

**2.2. Metagenomics ..... 8**

*2.2.1. 16S rRNA sequencing ..... 9*

**2.3. Metagenomics Data Analyses ..... 10**

*2.3.1. Pathway analysis (Ampliseq/Picrust2) ..... 10*

*2.3.2. Taxonomic classification (Kraken2) ..... 11*

*2.3.3. Reference interval estimation (RefineR) ..... 11*

**2.4. Application Programming Interfaces (API) ..... 11**

*2.4.1. The use of restful API in bioinformatic..... 12*

**3. MATERIALS AND METHODS…..... 14**

**3.1. Type of the Study… ..... 14**

**3.2. Time and Place of the Study..... 14**

**3.3. Materials of the Study..... 14**

**3.4. Variables of the Study…..... 17**

**3.5. Tools for Data Collection..... 17**

**3.6. Study Plan and Calendar..... 18**

**3.7. Data Evaluation ..... 18**

*3.7.1. 16S rRNA sequencing… ..... 18*

*3.7.2. Bioinformatic analysis..... 19*

*3.7.2.1. Taxonomic profiling ..... 19*

*3.7.2.2. Pathway abundance analysis..... 19*

*3.7.3. Generating restful API..... 20*

*3.7.4. Application programming interface dockerizing… ..... 20*

**3.8. Limitations of the Study ..... 20**

**3.9. Ethics Committee Approval ..... 21**

**4. RESULTS ..... 22**

**4.1. Taxonomic Profiling…..... 22**

**4.2. Pathways Abundance Analysis..... 25**

**4.3. Reference Intervals Estimation ..... 27**

*4.3.1. Anxiety..... 27*

*4.3.2. Bacterial chemotaxis..... 28*

*4.3.3. Base excision repair* ..... *30*

*4.3.4. Bile secretion* ..... *32*

*4.3.5. Biosynthesis of other secondary metabolites* ..... *34*

*4.3.6. Butanoate metabolism* ..... *36*

*4.3.7. Carbohydrate digestion and absorption* ..... *38*

*4.3.8. Depression* ..... *40*

*4.3.9. Fat digestion and absorption* ..... *42*

*4.3.10. Gut microbiome health* ..... *44*

*4.3.11. Inflammation* ..... *46*

*4.3.12. Inflammatory bowel disease* ..... *48*

*4.3.13. Insulin resistance* ..... *50*

*4.3.14. Irritable bowel syndrome* ..... *52*

*4.3.15. Lipid and atherosclerosis* ..... *54*

*4.3.16. Lipopolysaccharide biosynthesis* ..... *56*

*4.3.17. Metabolism of xenobiotics* ..... *58*

*4.3.18. Methane metabolism* ..... *60*

*4.3.19. Mineral absorption* ..... *62*

*4.3.20. Nitrogen metabolism* ..... *64*

*4.3.21. Non-alcoholic fatty liver disease* ..... *66*

*4.3.22. Oxidative phosphorylation* ..... *68*

*4.3.23. Pancreatic secretion* ..... *70*

*4.3.24. Propionate metabolism* ..... *72*

*4.3.25. Protein digestion and absorption* ..... *74*

*4.3.26. Sulfur metabolism* ..... *76*

*4.3.27. Type II diabetes mellitus* ..... *78*

*4.3.28. Apoptosis* **.....** *80*

**4.4. Application Programming Interface ..... 82**

*4.4.1. Data… ..... 82*

*4.4.2. Database ..... 82*

*4.4.3. Server ..... 84*

*4.4.4. Installation… ..... 85*

*4.4.5. Restful API for reference interval estimation… ..... 85*

**5. DISCUSSION ..... 98**

**6. CONCLUSION AND SUGGESTIONS ..... 101**

**7. REFERENCES ..... 102**

**8. APPENDIX I ..... 105**

**INDEX OF TABLES**

**Table 1: Reference interval estimation results for anxiety metabolic pathway…..... 27**

**Table 2: Reference interval estimation results for bacterial chemotaxis metabolic pathway…..... 29**

**Table 3: Reference interval estimation results for base excision repair metabolic pathway…..... 31**

**Table 4: Reference interval estimation results for bile secretion metabolic pathway…..... 33**

**Table 5: Reference interval estimation results for biosynthesis of other secondary metabolites metabolic pathway… ..... 35**

**Table 6: Reference interval estimation results for butanoate metabolism metabolic pathway…..... 37**

**Table 7: Reference interval estimation results for carbohydrate digestion and absorption metabolic pathway…..... 39**

**Table 8: Reference interval estimation results for depression metabolic pathway…..... 41**

**Table 9: Reference interval estimation results for fat digestion and absorption metabolic pathway…..... 43**

**Table 10: Reference interval estimation results for gut microbiome health metabolic pathway…..... 45**

**Table 11: Reference interval estimation results for inflammation metabolic pathway…..... 47**

**Table 12: Reference interval estimation results for inflammatory bowel disease metabolic pathway…..... 49**

**Table 13: Reference interval estimation results for insulin resistance metabolic pathway…..... 51**

**Table 14: Reference interval estimation results for irritable bowel syndrome metabolic pathway…..... 53**

**Table 15: Reference interval estimation results for lipid and atherosclerosis metabolic pathway…..... 55**

**Table 16: Reference interval estimation results for lipopolysaccharide biosynthesis metabolic pathway…..... 57**

**Table 17: Reference interval estimation results for Metabolism of xenobiotics by cytochrome P450metabolic pathway…..... 59**

**Table 18: Reference interval estimation results for methane metabolism metabolic pathway…..... 61**

**Table 19: Reference interval estimation results for mineral absorption metabolic pathway…..... 63**

**Table 20: Reference interval estimation results for nitrogen metabolism metabolic pathway..... 65**

**Table 21: Reference interval estimation results for non-alcoholic fatty liver disease metabolic pathway…..... 67**

**Table 22: Reference interval estimation results for oxidative phosphorylation metabolic pathway..... 69**

**Table 23: Reference interval estimation results for pancreatic secretion metabolic pathway..... 71**

**Table 24: Reference interval estimation results for propionate metabolism metabolic pathway..... 73**

**Table 25: Reference interval estimation results for protein digestion and absorption metabolic pathway..... 75**

**Table 26: Reference interval estimation results for sulfur metabolism metabolic pathway…..... 77**

**Table 27: Reference interval estimation results for type II diabetes mellitus metabolic pathway..... 79**

**Table 28: Reference interval estimation results for apoptosis metabolic pathway…..... 81**

**INDEX OF FIGURES**

**Figure 1: Stacked bar plot showing the top 10 most abundant bacterial phylum. ……...……...…….....22**

**Figure 2: Stacked bar plot showing the top 20 most abundant bacterial genus. ……...……23**

**Figure 3: Heatmap of bacterial genus with significant differences between samples. ……..……...……..…….....24**

**Figure 4: Stacked bar plot showing the least top 10 bacterial genera. ……...…….…….....24**

**Figure 5: Principal component analysis (PCA) of the profiling data from the gut microbiome ..... 25**

**Figure 6: Heatmap of microbial signatures in 30 different metabolic pathways of interest in 16S rRNA sequencing data from samples collected from volunteers ..... 26**

**Figure 7: Graph of reference interval estimation result for anxiety metabolic pathway…..... 28**

**Figure 8: Graph of reference interval estimation result for bacterial chemotaxis metabolic pathway..... 30**

**Figure 9: Graph of reference interval estimation result for base excision repair metabolic pathway..... 32**

**Figure 10: Graph of reference interval estimation result for bile secretion metabolic pathway..... 34**

**Figure 11: Graph of reference interval estimation result for biosynthesis of other secondary metabolites metabolic pathway..... 36**

**Figure 12: Graph of reference interval estimation result for butanoate metabolism metabolic pathway..... 38**

**Figure 13: Graph of reference interval estimation result for carbohydrate digestion and**

**absorption metabolic pathway. ..... 40**

**Figure 14: Graph of reference interval estimation result for depression metabolic pathway..... 42**

**Figure 15: Graph of reference interval estimation result for fat digestion and absorption metabolic pathway..... 44**

**Figure 16: Graph of reference interval estimation result for gut microbiome health metabolic pathway..... 46**

**Figure 17: Graph of reference interval estimation result for inflammation metabolic pathway…..... 48**

**Figure 18: Graph of reference interval estimation result for inflammatory bowel disease metabolic pathway…..... 50**

**Figure 19: Graph of reference interval estimation result for insulin resistance metabolic pathway..... 52**

**Figure 20: Graph of reference interval estimation result for irritable bowel syndrome metabolic pathway..... 54**

**Figure 21: Graph of reference interval estimation result for lipid and atherosclerosis metabolic pathway..... 56**

**Figure 22: Graph of reference interval estimation result for lipopolysaccharide biosynthesis metabolic pathway. ..... 58**

**Figure 23: Graph of reference interval estimation result for the metabolism of xenobiotics by cytochrome P450 metabolic pathway..... 60**

**Figure 24: Graph of reference interval estimation result for methane metabolism metabolic pathway..... 62**

**Figure 25: Graph of reference interval estimation result for mineral absorption metabolic pathway..... 64**

**Figure 26: Graph of reference interval estimation result for nitrogen metabolism**

**metabolic pathway. ..... 66**

**Figure 27: Graph of reference interval estimation result for nonalcoholic fatty liver disease metabolic pathway. ..... 68**

**Figure 28: Graph of reference interval estimation result for oxidative phosphorylation metabolic pathway. ..... 70**

**Figure 29: Graph of reference interval estimation result for pancreatic secretion metabolic pathway. ..... 72**

**Figure 30: Graph of reference interval estimation result for propionate metabolism metabolic pathway. ..... 74**

**Figure 31: Graph of reference interval estimation result for protein digestion and absorption metabolic pathway. ..... 76**

**Figure 32: Graph of reference interval estimation result for sulfur metabolism metabolic pathway. ..... 78**

**Figure 33: Graph of reference interval estimation result for type II diabetes mellitus metabolic pathway. ..... 80**

**Figure 34: Graph of reference interval estimation result for apoptosis metabolic pathway… ..... 82**

**Figure 35. SQLITE3 database browser interface image ..... 83**

**Figure 36: Entity relationship diagram of the database ..... 84**

**Figure 37: Screenshot of Restful API home page ..... 85**

**Figure 38: Workflow of Restful API for reference interval estimation ..... 86**

**Figure 39: Restful API documentation screenshot ..... 87**

**Figure 40: Restful API's Register page ..... 88**

**Figure 41: Restful API's Login page ..... 88**

**Figure 42: List Pathway Abundance page ..... 89**

**Figure 43: JSON output of functional abundance results in metabolic pathways for the selected sample ID. ..... 90**

**Figure 44: HTML output of functional abundance results in metabolic pathways for the selected sample ID. ..... 91**

**Figure 45: PDF output of functional abundance results in metabolic pathways for the selected sample ID. ..... 92**

**Figure 46: Edit page of functional abundance results in metabolic pathways for the selected sample ID. ..... 93**

**Figure 47: Add Pathway Abundance Results page ..... 94**

**Figure 48: Reference Interval Estimation page ..... 95**

**Figure 49: JSON output of selected sample ID and metabolic pathway reference interval analysis results ..... 95**

**Figure 50: HTML output of selected sample ID and metabolic pathway reference interval analysis results ..... 96**

**Figure 51: PDF output of selected sample ID and metabolic pathway reference interval analysis results ..... 97**

## LIST OF ABBREVIATIONS

API: Application Programming Interface

HGP: Human Genome Project

HMP: Human Microbiome Project

HTML: Hyper Text Markup Language

HTTP: Hyper Text Transfer Protocol

JSON: JavaScript Object Notation

KEGG: Kyoto Encyclopedia of Genes and Genomes

KO: KEGG Ortholog

MetaHIT: METAgenomics of the Human Intestinal Tract

NGS: Next Generation Sequencing

OTU: Operational Taxonomic Unit

PCoA: Principal Component Analysis

PCR: Polymerase Chain Reaction

Picrust: Phylogenetic Investigation of Communities by Reconstruction of Unobserved States

RDP: Ribosomal Database Project

RefineR: Reference Interval Estimation using Real-World Data

SQL: Structured Query Language

**ACKNOWLEDGEMENTS**

I would like to thank the following persons for their academic contributions to the preparation of my dissertation:

Associate Professor Gökhan KARAKÜLAH, for his constant support, encouragement, and patience during the creation and completion of this dissertation, as well as his mentoring in the area of computational biology;

Assistant Professor Yavuz OKTAY, for his help during the design and completion of this dissertation, in addition to his mentoring in the area of genomics;

IBG Bioinformatics Platform members who are always supportive. However, I want to express my special thanks to Alirıza ARIBAŞ for his invaluable efforts in technical assistance and high motivation.

Oktay Lab members Burcu EKİNCİ and Tutku YARAŞ helped me with wet lab experiments.

My family, for their emotional and financial support.

# DEVELOPMENT OF AN APPLICATION PROGRAMMING INTERFACE FOR DYNAMICALLY ESTIMATING REFERENCE INTERVALS FOR FUNCTIONAL ABUNDANCES OF GUT MICROBIOTA

**Leman Binokay, İzmir International Biomedicine and Genome Institute, Dokuz Eylul University Health Campus, Balcova 35340 - Izmir / TURKEY**

## ABSTRACT

Humans interact with microbes in the environment and their bodies throughout their lives. The microbiota of the human gut is the biggest and most significant component of the human microbiome, comprised mostly of bacteria. Dysbiosis, which occurs when the healthy structure of the microbiome is disrupted, has been linked to various diseases. Therefore, gut microbes play an essential role in disease diagnosis and therapy. Recent studies of the human gut microbiota have used next-generation sequencing (NGS) technology to detect, compare, and research the functional properties of the microbiota's composition. This research aimed to develop an application programming interface (API) tool that allows the functional abundances of gut microbiota in defined and annotated metabolic pathways in publicly available databases (KEGG) to be dynamically estimate reference intervals in a specific population. The API will also help to establish population-specific healthy human gut microbiota profiles. The V3-V4 region of the 16SrRNA gene from fecal samples taken from 57 healthy individuals was sequenced according to the inclusion-exclusion criteria. The data were analyzed using Kraken2 and Picrust2 tools, and the mapping of microbial signatures and composition data to metabolic pathways associated with the gut microbiota was performed. Population-specific reference intervals for each metabolic pathway were obtained using the refineR package. Our results of the healthy human gut microbiota composition are comparable with that of the healthy human gut microbiota in the literature. An application programming interface that dynamically estimates reference intervals using Flask framework and Python programming language has been developed.

**Keywords:** Human Microbiome, NGS, 16S rRNA, Metagenome, Reference Interval, Restful API.

BAĞIRSAK MİKROBİYOTASI FONKSİYONEL BOLLUĞU REFERANS ARALIKLARININ DİNAMİK KESTRİMİ İÇİN BİR UYGULAMA PROGRAMLAMA ARAYÜZÜ GELİŞTİRİLMESİ

**Leman Binokay, İzmir Uluslararası Biyotıp ve Genom Enstitüsü, Dokuz Eylul Üniversitesi Sağlık Yerleşkesi, Balcova 35340 - Izmir / Türkiye**

## ÖZET

İnsanlar yaşamları boyunca çevrelerindeki ve vücutlarındaki mikroorganizmalar ile etkileşim halindedirler. İnsan bağırsak mikrobiyotası, çoğunlukla bakterilerden oluşan insan mikrobiyomunun en büyük ve en önemli bileşenidir. İnsan sağlığı, insan bağırsağı mikrobiyotası olan homeostatik ilişkiyle yakından bağlantılıdır. Mikrobiyomun sağlıklı yapısı bozulduğunda ortaya çıkan disbiyoz, metabolik, nöropsikiyatrik ve kanser gibi çok çeşitli hastalıklarla ilişkilendirilmiştir. Bu nedenle insan bağırsak mikrobiyotası, hastalıkların tanı ve tedavisinde önemli bir rol oynamaktadır. İnsan bağırsağı mikrobiyotasına ilişkin son araştırmalar, insan bağırsağı mikrobiyotasının bileşimini tanımlamak, karşılaştırma yapmak ve işlevsel olarak incelemek için yeni nesil dizileme teknolojisi kullanılarak gerçekleştirilmiştir. Yeni nesil dizileme teknolojisi, çok sayıda bireyin büyük genomik bölgelerinin (tüm genom, ekzom,amplikon) çeşitli yöntemlerle kısa sürede ve birim başına düşük maliyetle çalışılmasına olanak sağlar. Bu araştırmada, açık veritabanlarında (örn. KEGG) tanımlanmış ve manuel olarak seçilenmetabolik yolaklardaki bağırsak mikrobiyotasının fonksiyonel bolluğunun belirli birpopülasyondaki referans aralıklarını dinamik olarak kestirimini sağlayan bir uygulama programlama arayüzü (API) geliştirmeyi amaçlandı. Bu amaçla, dahil etme-dışlama kriterlerine göre 57 sağlıklı bireyden alınan dışkı örneklerinden 16S rRNA geninin V3-V4 bölgesi sekanslanmış; veriler Kraken2 ve Picrust2 araçları kullanılarak analiz edildi ve mikrobiyal imzaların ve bileşim verilerinin bağırsak mikrobiyotası ile ilişkili metabolik yollara haritalanması gerçekleştirildi. Her metabolik yolak için popülasyona özgü referans aralıkları, refineR paketi kullanılarak kestirimi yapıldı. Sağlıklı insan bağırsağı mikrobiyota kompozisyonunun sonuçları, literatürdeki sağlıklı insan bağırsağı mikrobiyotasının sonuçlarıyla uyumludur. Özetle, Flask ve Python programlama dili kullanılarak dinamik olarak referans aralıklarını kestirimini yapan bir uygulama programlama arayüzü geliştirilmiştir.

**Anahtar Sözcükler:** İnsan Mikrobiyomu, NGS, 16S rRNA, Metagenom, Referans Aralık, Restful API.

# 1. INTRODUCTION AND AIM

## 1.1. Statement and Importance of the Problem

The intestinal microbiota consists of many diverse microorganisms that play important roles in various metabolic pathways. Recent research has shown conclusively that the metabolic pathways influenced by gut bacteria are intimately associated with a variety of disorders. To completely comprehend the intricate interaction between the human gut microbiota, human health, and disease risk, further research on the influence of gut bacteria on metabolic pathways is required.

## 1.2. Aim of the Study

This research aimed to develop an application programming interface (API) tool that allows the functional abundances of gut microbiota in defined and annotated metabolic pathways in publicly available databases (e.g., KEGG) to be dynamically estimated with reference intervals in a specific population. It is aimed that this tool will be open-source, platform-independent, and amenable to integration into different systems. In addition, it is aimed to obtain a basal gut microbiota profile of a healthy Turkish population with a data set of 57 healthy individuals by integrating the 16S sequencing data into this tool.

## 1.3. Hypothesis of the Study

The study hypothesizes that an API that determines the bacteria involved in a healthy gut microbiome and their functional abundances in various metabolic pathways by estimating reference intervals can be developed.

## 2. GENERAL INFORMATION

### 2.1. Human Microbiome

Joshua Lederberg introduced the concept of the human microbiome, which he characterized as "the ecological community of commensal, symbiotic, and pathogenic bacteria and their genes that share our internal space" (Ursell et al., 2012). Microbiota is the ecological community of microorganisms such as bacteria, viruses, fungi, and archaea that reside in the human body, whereas the microbiome is the entire genome of microorganisms residing in this environment. Various anatomical body locations, including the skin, gastrointestinal system, respiratory system, urogenital system, and oral, have colonized microorganisms. They develop a distinct and sophisticated ecosystem that adapts to the needs of each niche. This community of microorganisms, known as the microbiota, contains around 100 trillion bacteria, accounting for approximately 2% to 3% of human body weight. Our microbiome contains ten times more cells than our bodies (Berg et al., 2020; Lloyd-Price et al., 2016b; Reynoso-García et al., 2022).

The Human Genome Project (HGP) and the Human Microbiome Project (HMP) have given the area of health a new viewpoint throughout the last two decades. As a continuation of the HGP, the United States National Institute of Health started the HMP to investigate the function of microorganisms in health and illness and to comprehend the microbiota and microbiome diversity in distinct anatomical locations. The Human Microbiome Project has sequenced the genome of the human microbiota, with an emphasis on bacteria in the skin, mouth, nose, digestive system, and genitalia. On the dynamics of the quantity and variety of bacteria in thegut microbiota ecosystem, a consensus is still being determined. Today, breakthroughs in molecular biology methods have significantly aided in the comprehension of gut bacteria.Recent research and HMP suggest that there are about 35,000 bacterial species in the gastrointestinal system (Pace, 2007; Proctor et al., 2019; Turnbaugh et al., 2007a).

In the microbiota of the gastrointestinal system, there are anaerobic, facultatively anaerobic, and aerobic microorganisms. However, the most significant component of the gut microbiota consists primarily of anaerobic bacteria, such as Bacteroidetes and Firmicutes. In a healthy gut microbiome, the predominant bacterial phyla are Actinobacteria, Firmicutes, Bacteroidetes, Verrucomicrobia, and Proteobacteria. The Firmicutes phylum is the most abundant in the gut's flora, followed by the Bacteroidetes phylum (King et al., 2019; Lloyd-Price et al., 2016b).

Birth marks the beginning of a symbiotic relationship between the human body and its natural microbiota. These interactions play vital roles in the maintenance of health and well-being on a global scale. As a consequence of their bioactivity, these microorganisms are regarded to be a component of the human body and are responsible for a range of alterations from birth to death (Kho & Lal, 2018; Lloyd-Price et al., 2016a). Age, diet, lifestyle, genetics, environment, and underlying disease are major factors in the human microbiome. In response to host characteristics such as genetics, diet, environment, and lifestyle, the human microbiome is continuously evolving and developing. Nonetheless, dysbiosis of gut microbiota may result in life-threatening diseases (Ogunrinola et al., 2020; Reynoso-García et al., 2022).

### *2.1.1. Gut microbiota and health*

Hippocrates, considered the father of modern medicine, said 2000 years ago that "all diseases begin in the gut." This statement has been proven to be true thanks to recent human microbiome studies. It is becoming increasingly clear that there may be problems with the gut microbiome under many seemingly unrelated diseases. In other words, the health of our gut affects all body systems. It has been shown that a balanced microbiota is essential for maintaining health. The gut has the highest concentration of the human microbiome. These creatures are essential for maintaining human health. Past research on the human microbiome has shown that changes in the immune system are closely related to unhealthy gut flora. A shift in the gut microbial ecology has a significant impact on human health and the development of illness. These alterations are the result of an unhealthy lifestyle and an underlying illness. Dysbiosis has also been related to life-threatening disorders such as cancer, cardiovascular disease, intestinal disease, and bacterial infections due to antibiotic resistance (Kho & Lal, 2018; Lloyd-Price etal., 2016b; Ogunrinola et al., 2020).

### *2.1.2. Gut microbiota and metabolic pathways*

The gut microbiota comprises many various bacteria that influence human physiology through several metabolic pathways. A diverse and metabolically active microbial ecology exists in the human gut. Recent research demonstrates conclusively that the metabolic pathways influenced by gut bacteria are intimately associated with a variety of disorders (Pascal Andreu et al., 2021).

Microorganisms are able to colonize many different locations and across the human body, and once there, they are able to adapt to the unique characteristics of each niche. The distinct metabolic activities and roles of these bacteria in each bodily location may be explained by the unique variety of the human microbiota. For instance, bacteria in the human gut regulatefundamental metabolic, physiological, and immunological activities through several pathways, and disturbances in this ecosystem have direct effects on host health (Kho & Lal, 2018; Visconti et al., 2019). Consequently, it is essential to comprehend the microbial makeup and activity of the human microbiome in metabolic pathways that contribute to health and illness. In Table 1, thirty different metabolic pathways used in this study are listed. Gut microbiota has the ability tocarry out a variety of metabolic functions that humans are either unable to do at all or can only do to a minor degree. Bacteria found in gut microbiota are capable of producing many vitamins, minerals, and compounds, as well as synthesizing all necessary and non-essential amino acids (Pascal Andreu et al., 2021.). Consequently, any change in microbiota may compromise hemostasis by altering metabolic pathways. Acetate, propionate, and butyrate are three different short-chain fatty acids (SCFAs) produced mostly by anaerobic fermentation of gut bacteria. SCFAs are essential for regulating energy supply and metabolism and maintaining the gut's homeostasis (He et al., 2020). Abnormalities in the microbiota of the gut may, through the gut-brain axis, also have a substantial influence on the functioning of the brain. Neurotransmitter disturbances, such as those involving serotonin, may have a role in anxiety and depression (Cryan et al., 2019; Füllinget al., 2019). Metabolic pathways found to be associated with human gut microbiota in the literature are listed below.

1. Apoptosis
2. Anxiety
3. Bacterial chemotaxis
4. Base excision repair
5. Bile secretion
6. Biosynthesis of other secondary metabolites
7. Butanoate metabolism
8. Carbohydrate digestion and absorption

9. Depression
10. Fat digestion and absorption
11. Gut microbiome health
12. Inflammation
13. Inflammatory bowel disease
14. Insulin resistance
15. Irritable bowel syndrome
16. Lipid and atherosclerosis
17. Lipopolysaccharide biosynthesis
18. Metabolism of xenobiotics
19. Methane metabolism
20. Mineral absorption
21. Nitrogen metabolism
22. Non-alcoholic fatty liver disease
23. Oxidative phosphorylation
24. Pancreatic secretion
25. Propionate metabolism
26. Protein digestion and absorption
27. Sulfur metabolism
28. Type II diabetes mellitus

**2.2.Metagenomics**

The term meta-genomics, first used by Jo Handelsman in 1998, can be defined as obtaining DNA content without the need to isolate and culture the entire microbiome in an environment (Handelsman, 2004). It is a technique applied in the form of amplification and sequencing of DNA by fragmentation or amplification of genes (Lepage et al., 2013). 16S rRNA and 18S rRNA genes obtained from the metagenome are most often utilized to define microbial populations comprehensively. Additionally, several conserved genes, including recA, rpoB, HSP70, and EF-Tu, provide phylogenetic classifications. Conventional molecular techniques, such as denaturing gradient gel electrophoresis, terminal restriction fragment length polymorphism analysis, and Sanger sequencing, may be used to explore these genes. The fact

that Sanger sequencing depends on the investigation of microbial composition requires a significant investment of both time and effort. It involves the development of clone libraries is a significant drawback (Sanger F. et al., 1977). Recently, systems for next-generation sequencing (NGS) have been used to decode meta-genomic DNA. Sequencing by synthesis (Solexa, Illumina), semiconductor chip sequencing (Ion Torrent/Life Technologies), pyro-sequencing (454/ROCHE), and long-read sequencing are the most used methods (Oxford Nanopore Technologies and SMRT-Pacific Biosciences). Comprehensive knowledge of the evolutionary history and functional diversity of complex human microbial communities may be obtained via the use of metagenomic approaches (Behjati & Tarpey, 2013). Thanks to the advent of sequencing technology, new bioinformatics algorithms and tools, and the high accuracy rates afforded by metagenomics investigations, human microbiome research has exploded in recent years.

### *2.2.1. 16S rRNA sequencing*

The gut microbiome is very dynamic and varies among people. It is continually changed by variables which are diet, lifestyle, environment, and genetics. NGS has substantially improved our knowledge of the human microbiome since it enables culture-independent investigation of differences in the variety and quantity of bacteria. Shotgun sequencing and 16S rRNA sequencing are these approaches. As a direct consequence of recent developments in bioinformatics approaches and the availability of data, the process of metagenome assembly has made it possible to identify hundreds of novel microorganisms in the host and nature. Since the 16S small subunit ribosomal gene is highly conserved across bacteria and archaea, it has been widely applied to infer microbial phylogenetic relationships. The 16S rRNA gene has nine hyper-variable regions (V1-V9) with species-specific variations that are conserved across bacterial species. The amplification of 16S rRNA hyper-variable areas may thus be used to identify genus-level microbial communities in a sample, and identification of species-level microbial communities is also conceivable if complete 16S sequences are identified. Nevertheless, conserved areas are not similar between archaeal and bacterial species, which might affect the PCR (Polymerase Chain Reaction) amplification process. According to an *in silico* study, V4-V6 sections are more effective in modeling the whole taxonomic distribution of the 16S gene, , although V3-V4 regions are still largely acknowledged as the most informative

and most regularly utilized regions in gut microbiome investigations (Johnson et al., 2019). Fifty-seven healthy individuals' feces were sequenced using V3-V4 primers with the help of Illumina iSeq 100 equipment to collect 16S rRNA sequencing data for this thesis (Escobar-Zepeda et al., 2015; Lepage et al., 2013; Nakatsu et al., 2019; O'Callaghan et al.,2021).

### *2.3. Metagenomics Data Analysis*

Bioinformatics is an interdisciplinary science that integrates biological data with tools of information storage, distribution, and analysis to serve several scientific disciplines, such as biomedicine, particularly molecular biology and genetics, computer science, mathematics, and statistics. 16S small subunit rRNA gene sequence datasets may be processed using several instruments, software packages, analytical web servers, and virtual samples, such as Picrust2 and Kraken2.

#### *2.3.1. Ampliseq/Picrust2*

To understand how the human gut microbiota affects human health, it is necessary to establish a functional profile of the microbiota. For this purpose, the activity of gut microbiota data obtained from healthy individuals in metabolic pathways should be determined. One of the mostwidely used tools for functional profiling of the microbiome is PICRUST2 (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States), developed by the Huttenhower Lab (Douglas et al., 2020). In this thesis, analysis of 16S rRNA sequencing data and mapping of obtained microbial signatures and composition to metabolic pathways associated with gut microbiota were performed. For this purpose, the ampliseq workflow developed by nf-core using Nextflow was used for the bioinformatic analysis of 16S sequencingdata. By using the Picrust2 tool included in this workflow, functional abundances were obtained in previously identified metabolic pathways based on marker gene sequences. The activities of the microbiota were determined for 30 manually selected metabolic pathways in our analyses of the sequencing data; a customized data table was used by determining the KEGG orthologous (KO) gene families for each metabolic pathway, which was prepared manually from the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa et al., 2016).

### *2.3.2. Kraken2*

Kraken2 is a taxonomy classification method that uses accurate k-mer matching to achieve high precision and rapid classification rates. This classifier matches each k-mer inside a querysequence to the most recent common ancestor of all genomes that include the supplied k-mer. The assignments of k-mers guide the classification method. In this thesis, the Kraken2 tool was utilized to assess the profile of the gut microbiota in healthy people. The readings were aligned to the target species defined in the RDP (Ribosomal Database Project) database using Kraken2 for taxonomic profiling (Wood et al., 2019).

### *2.3.4. Reference interval estimation (RefineR)*

For any result to be interpreted, its position in the reference interval is required. In patient health management, the analysis results of the patient are compared with the data of healthy people. Normal values for clinical laboratory data are the reference interval. Terminologically, the use of the expression "normal interval" was not found sufficient; instead, the terms "reference values obtained from healthy people" or "reference interval" are used. Thus, the comparison is essential for any subject to make sense of the results obtained. To compare biological data, reference intervals created with data obtained from healthy individuals are used. RefineR (Reference Interval Estimation with Real-World Data) is a method for estimating reference intervals using real-world data. This technique employs complex statistical approaches to construct a model describing the distribution of routinely measured samples for diagnostic testing. In order to make a highly accurate interpretation of the gut microbiota health of the individuals, it is essential to determine the reference intervals for the data obtained from healthy individuals. For this purpose, reference intervals were determined for each metabolic pathway using the refineR package developed in the R environment (Ammer et al., 2021).

## *2.4.* Application Programming Interface (API)

APIs are methods that enable two software components to interact via the use of specified definitions and protocols. In application programming interfaces, "Application" refers to any software with a distinct purpose. An interface is a service agreement between apps. This

contract specifies how the apps will exchange requests and answers to communicate. The documentation for the associated API describes how developers should configure these requests and answers. Client and server are often used to describe the architecture of an application programming interface. The application that is responsible for sending the request is describedto as the "client," while the program that is responsible for sending the response is described to as the "server." APIs are categorized based on their design and application scope. Restful API is the most prevalent and versatile API type. REST is an abbreviation for "Representational State Transfer." REST employs operations such as GET, PUT, and DELETE so that clients may access server data. Rest is an architectural concept that permits server-client data communication. Through HTTP, clients and servers exchange data (Hyper Text Transfer Protocol). In brief, it is the protocol that allows web pages to be displayed. HTTP establishes thestandards for data flow between the client and server. Restful API is an application interface thatallows web services to communicate data using the Rest architecture. APIs are services that disseminate data and execute the software's functionality (Alonso, 2015; Tarkowska et al.,2018).

### *2.4.1. The use of Restful API in bioinformatics*

REST APIs have several advantages, such as integration, innovation, expansion, and maintainability. APIs are used to incorporate new software applications into existing software systems. There is no need to create each function from scratch, which increases the rate of development. By utilizing APIs, existing code can be utilized even further. APIs can be utilized in altering systems or applications by making API-level changes instead of reworking the code. This allows work to continue without any interruption to the ongoing applications. APIs work on multiple platforms. Additionally, the API enables two systems to communicate with each other. Each system must implement internal modifications such that the API remains unaffected.Thus, future code modifications with one side will not impact the other (Alonso, 2015; Tarkowska et al., 2018). In bioinformatics studies, APIs are widely used because biological data can be distributed in the desired and needed format. Moreover, developed APIs can be integrated into different pipelines, run on different platforms, and are open to development.After analyzing the 16S rRNA sequencing data and mapping the obtained microbial taxonomic data to the metabolic pathways associated with the gut microbiota, reference intervals of each

metabolic pathway specific to the samples taken from healthy volunteers were estimated using the refineR package in the R environment. In this thesis, an application programming interface has been developed to share these data with authorized users with the help of an interface and to estimate reference intervals dynamically.

# 3.MATERIALS AND METHODS

## 3.1. Type of the Study

This study is based on computational and statistical calculations applied on biological samples.

## 3.2. Time and Place of the Study

This study was conducted at Izmir International Biomedicine and Genome Institute and Izmir Biomedicine and Genome Center between November 2022 and January 2023.

## 3.3. Materials of the Study

In this study, samples of feces were taken from 57 different volunteers. A sample of feces approximately the size of a chickpea was collected using a disposable spoon and transferred into a screwcap tube containing 10 volume:volume of DNA/RNA Shield (ZymoResearch). The following conditions were used as exclusion criteria:

1. Presence of any chronic medical disease,

2. Having had an acute medical illness in the last three months,

3. To have taken a regular drug treatment (antibiotic, antiviral, steroids, stomach protectors, etc.) in the last three months,

4. To have followed a special diet program in the past three months (history of unbalanced eating defined by major changes in diet that has eliminated or significantly increased a major food group from the diet during the previous month),

5. To have taken nutritional supplements (probiotics, vitamins) regularly in the last three months,

6. Pregnancy and breastfeeding,

7. Body mass index being above 35 or below 18,

8. Recent history of chronic alcohol consumption (consumption of more than five servings of alcohol (1 serving 225 ml), defined as five servings of beer or 750 ml of wine per day),

9. To have received a flu vaccine administered nasally in the last 28 days,

10. Exercising at a very high intensity or being a professional athlete.

11. The inclusion criterion was to be between the ages of 18-65.

To isolate DNA from these samples, a "ZymoBIOMICS™ DNA Miniprep Kit" (Zymo Research, #D4300) was used. As a result, 1000 µl of fecal sample was vortexed for 40 minutes in the ZR BashingBead™ lysis tube. ZR BashingBead™ lysis tubes were centrifuged at 10,000x g for 1 minute, and then 400 µl of the supernatant was placed in the Zymo-Spin™ III-F filter in the collection tube and centrifuged at 8000 x g for 1 minute. The Zymo-Spin™ III-F filter was then discarded. The filtrate was added to 200 µl of ZymoBIOMICS™ DNA Binding Buffer and carefully mixed. 800 µl of the mixture was transferred to the Zymo-Spin™ IICR in a new collecting tube and centrifuged at 10,000 x g for 1 minute. The liquid in the collection tube was discarded, and the remaining mixture was transferred back to the Zymo-Spin™ IICR and centrifuged at 10,000 x g for 1 minute. In a new collection tube, 400 µl of ZymoBIOMICS™ DNA wash buffer one was placed to the ZymoSpin™ IICR Column and centrifuged at 10,000 x g for 1 minute. The liquid in the collection tube was discarded. 700 µl of ZymoBIOMICS™ DNA wash buffer 2 was added to the ZymoSpin™ IICR column inside the collection tube and centrifuged at 10,000 x g for 1 minute. The liquid in the collection tube was discarded. 200 µl of ZymoBIOMICS™ DNA wash buffer 2 was added to the ZymoSpin™IICR Column in the collection tube and centrifuged at 10,000 x g for 1 minute. 100 µl of ZymoBIOMICS™ DNase/RNase-Free water was added by transferring the Zymo-Spin™ IICRColumn to a clean 1.5 ml microcentrifuge tube. After 1 minute of incubation at room temperature, it was centrifuged at 10,000 x g for 1 minute. A Zymo Spin™ III-HRC filter was placed in the collection tube, and 600 µl of ZymoBIOMICS™ HRC preparation solution was added. It was centrifuged at 8,000 x g for 3 minutes. After transferring the Zymo-Spin™ III-HRC filter to a new 1.5 ml microcentrifuge tube, the eluted DNA was transferred to the Zymo-Spin™ III-HRC in a newly prepared 1.5 ml microcentrifuge tube. It was centrifuged at 16,000 x g for 3 minutes. Thus, bacterial DNA was obtained. Quality and quantity measurements were made with the Nanodrop 2000 (Thermo) UV Spectrophotometer device.

Bacterial DNA was stored at -20C for 16S rRNA sequencing. Bacterial DNA was amplified in the V3-V4 regions of the 16S rRNA gene using the Illumina-specific primers.. The sequences of forward and reverse primers used for library preparation are:

Forward Primer:

5’-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3’

Reverse Primer:

5’-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3’

2.5 µl Microbial DNA (5 ng/µl), 5 µl amplicon PCR forward primer (1 µM), 5 µl amplicon PCR reverse primer (1 µM) and 12.5 µl KAPA HiFi HotStart ReadyMix were mixed for PCR. PCR conditions were as follows: 95°C for 3 minutes then 25 cycles of 95°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds, followed by a final extension of 72°C for 5 minutes.

After the PCR amplification, clean-up of PCR products was performed. Amplicon PCR plate was centrifuged at 1,000 × g at 20°C for 1 minute. AMPure XP beads was vortexed for 30 seconds and 20 µl of AMPure XP beads was added to each well of the amplicon PCR plate. Then, pipetting was performed gently 10 times. Plates containing Amplicon PCR were incubated at room temperature for 5 minutes without shaking. Two minutes later, an Amplicon PCR plate was put on a magnetic platform. Supernatants were removed.

After removal of supernatant, beads were washed with freshly prepared 80% ethanol and incubated for 30 seconds. Supernatants were removed carefully. After that, 200 µl of freshly prepared 80% ethanol were added to each sample well and incubated for 30 seconds. Supernatants were removed carefully and excess ethanol was removed. With the amplicon PCR plate still on the magnetic platform, the beads were allowed to air-dry for 10 minutes. After the removal of amplicon PCR plate from magnetic platform, 52.5 µl of 10 mM Tris pH 8.5 was added to each well. Then, pipetting was performed gently 10 times and the plate was incubated at room temperature for 2 minutes. Amplicon PCR plate was placed on the magnetic platform for 2 minutes After that, 50 µl of the supernatant was transferred to a new 96-well PCR plate.

5 µl sample from each well was transferred to a new 96-well plate. 5 µl DNA, 5 µl Nextera XT Index Primer 1 (N7xx), 5 µl Nextera XT Index Primer 2 (S5xx), 25 µl 2x KAPA HiFi HotStart ReadyMix, and 10 µl PCR-grade water were mixed and pipetting was performed 10 times. After that, 96-well plate was centrifuged at 1,000 × g at 20°C for 1 minute. PCR was

performed on a thermal cycler. PCR conditions were 95°C for 3 minutes, followed by 8 cycles of 95°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds, after that 72°C for 5 minutes.

PCR clean-up was performed as described below. Concentration and size distribution of the library was measured using a Qubit fluorometer and a Bioanalyzer. The library should have a size range of ~550 bp and a concentration of at least 10 nM. After the library preparation with 16S Metagenomics Sequencing Library Preparation protocol supplied from Illumina, 5 ng DNA solution from each sample was sequenced with the iSeq 100 sequencing machine with paired-end (2x150 cycles) reading method using a iSeq i1 reagent v2.

### 3.4. Variables of the Study

Pathway abundance analysis outputs are independent variables. Estimated reference intervals are dependent variables.

### 3.5. Tools for Data Collection

All data collection and processing operations were carried out on a WorkStation which has Intel® Core™ i7- 9750H CPU @ 2.60GHz × 12 Processor, 64 GB RAM, 1 TB hard disk with Ubuntu release 22.04 LTS operating system installed and IBG HPC environments.

## 3.6. Study Plan and Calendar




## 3.7. Data Evaluation

*3.7.1. 16S rRNA sequencing*

In this thesis, it was aimed to determine the bacterial profile of 57 different samples with meta-genomic approaches. For this purpose, next generation sequencing was carried out by targeting 16S rRNA V3-V4 gene regions. In this approach, the 16S rRNA gene region is amplified and sequenced. The 1500 base long gene, which is usually called 16S rRNA andcodes for ribosomal RNA, is found in all bacteria. While there are highly conserved regions in the gene, it also contains regions that are variable enough to allow phylogenetic classification of different bacterial species. Sequence differences in 16S rRNA regions between different species increase in proportion to evolutionary distance. For this reason, the 16S rRNA sequence is used as a marker for the identification of different bacterial species. The 16S rRNA sequences of numerous bacterial species obtained from culture and environmental samples have been extracted by numerous studies so far. These sequences have been registered in various databases, including GreenGenes, Ribosomal Database Project (RDP), and SILVA. High-volume

sequence data obtained by methods such as next-generation sequencing can be compared with these databases, and the type of DNA fragments detected can be defined proportionally. 16S rRNA sequencing results were obtained on the Illumina iSeq 100 next generation sequencing platform (Illumina, CA, USA) by paired-end reading with the iSeq 100 i1 v2 reagent kit (2x150 bp) and following the manufacturer's instructions.

*3.7.2. Bioinformatic analysis*

After the application of sequencing, the FastQC software was utilized for quality control stages. According to the findings of the quality control, data quantities, read quality, GC distributions, kmer distributions, and potential adapter contaminations were analyzed for each sample. Poor quality readings (Phred Score Q20) were eliminated from all data after quality control. Additionally, the Trimmomatic tool was used to remove low-quality base readings and probable adaptor contamination. For taxonomic profiling, reads were aligned to target organism genomes using the KRAKEN2 method and the RDP (2022) database. After alignment, OTU groupings were established for each sample. According to FastQC, high-quality readings were analyzed using two methods: taxonomic classification and pathway abundance analysis.

*3.7.2.1. Taxonomic profiling*

For taxonomic profiling of 16S rRNA sequencing data, the Kraken2 (https://ccb.jhu.edu/software/kraken2/) classifier algorithm was applied to classify each high-throughput sequence read directly against the RDP. (Ribosomal Database Project, 2020) database with default parameters and return their taxa. After alignment, OTU (Operational Taxonomic Unit) groups in each sample were determined. For each sample, a Kraken 2 format output and report file of microbiome compositions with properties varying from kingdom to species were obtained. The Kraken2 report of each sample was combined using the combine_kreports.py function in KrakenTools (https://github.com/jenniferlu717/KrakenTools). Afterward, heatmaps were obtained by using the pheatmap package for the report of 57 samples that were manipulated in the R environment.

*3.7.2.2. Pathway abundance analysis*

Analysis of 16S sequencing data and mapping of obtained microbial signatures and composition data to metabolic pathways associated with the gut microbiota were performed. In this way,

population-specific reference intervals of each reported metabolic pathway were obtained. For this purpose, the ampliseq v2.3.1 workflow developed by nf-core using Nextflow v21.10.6 was used for bioinformatic analysis of 16S sequencing data. Thanks to the Picrust2 tool included in this workflow, functional abundances were obtained in previously defined metabolic pathways based on marker gene sequences. Using the data table, which is a customized data table manually prepared from the KEGG database and determining the ortholog gene families foreach metabolic pathway, results were obtained for 30 selected metabolic pathways in theanalyzes we performed with the sequencing data. Reference intervals were determined for each metabolic pathway using the refineR v1.5.1 (Reference Interval Estimation using Real-World Data) package developed in the R environment.

#### *3.7.3. Generating Restful API*

The gut microbiota profile and functional abundances in metabolic pathways results were transferred to the SQLITE3 database and structured in accordance with SQL (Structured Query Language) query, an Application Programming Interface (API) that can work locally and be moved to different platforms has been developed in order to organize the access of authorized users to the database and to dynamically calculate reference intervals. The Restful API was developed using the open-source Python programming language and the Flask software development platform.

#### *3.7.4. Application Programming Interface Dockerizing*

It is aimed to be ready for immediate use without the need for compilation after the installation of the libraries required for the API to work and the configuration of the web server. If desired, the Docker container (lemanbinokay/referenceintervalapi) can be run without installation, and inthis case, it is configured to select only the output port for Docker. Any operating system can be selected as the server, and it is a recommended Linux-based operating system.

### 3.8. Limitations of the Study

A small number of samples limit the scope of this study since the number of samples is not enough to represent the population. Read-length of 2X150 bp is another limitation, as more detailed information can be gained with longer reads with the Illumina MiSeq (i.e., 2x300 bp)

or ONT Nanopore or PacBio systems (both capable of sequencing the whole V1-V9 as a single read).

### 3.9. Ethics Committee Approval

Izmir Biomedicine and Genome Center Noninvasive Research Ethics Committee Approval carrying protocol number “2022-037” is granted for this thesis (see Appendix I).

# 4. RESULTS

## 4.1 Taxonomic Profiling

A total of 57 Turkish adults were included in this study. As a result of fecal microbiome analysis, a total of 89 phyla, 188 classes, 416 orders, 603 families, and 1464 genera were detected in all samples based on 16S rRNA sequences.

The dominant phyla were Bacteroidetes, Firmicutes, Acidobacteria, and Chloroflexi, of which medians of the relative abundance (% OTUs) were 84.82%, 12.87%, 1.49%, and 0.07%, respectively, and encompassed a total of 99.25% of the overall microbiota, as shown in Figure 1.

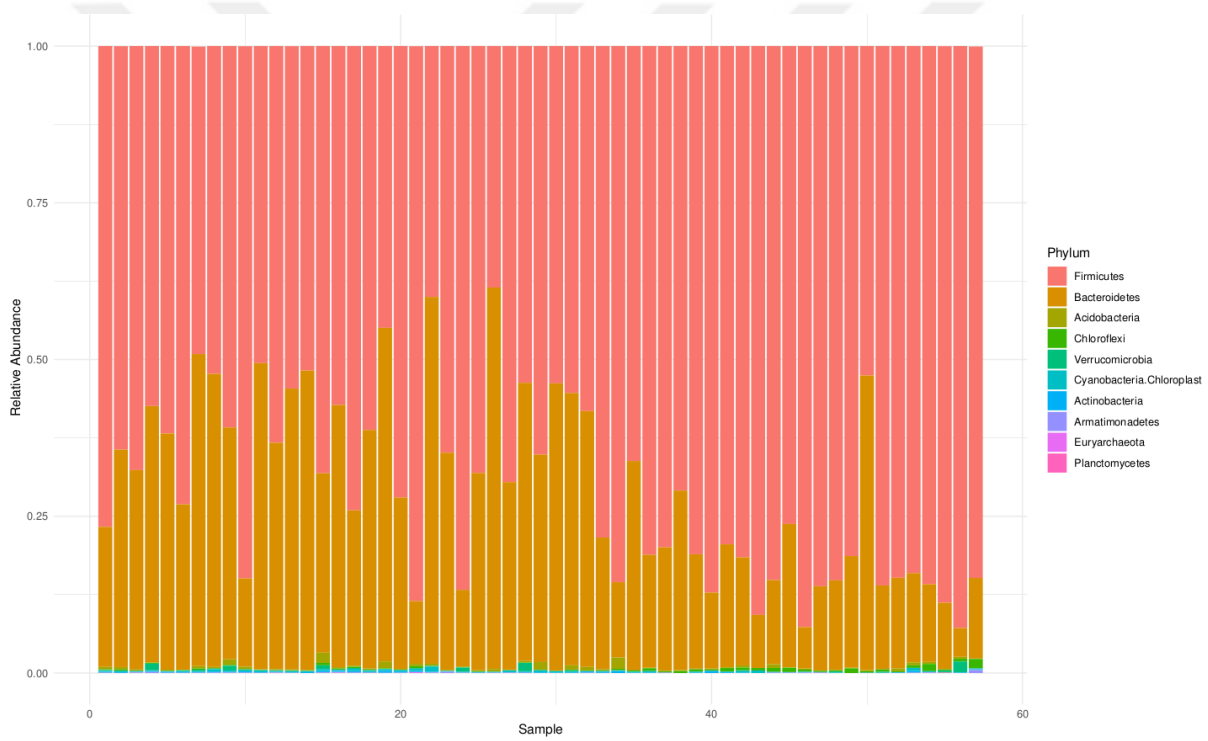


**Figure 1: Stacked bar plot showing the top 10 most abundant bacterial phylum.**

The list and distribution of the 20 bacterial genera with the highest incidence in all samples are shown in Figure 2. The dominant genera were Bacteroides, Prevotella, Faecalibacterium, Lachnospiracea, Blautia, Asteroleplasma, Porphyromonas, Alistipes, Roseburia, and Collinsella, of which medians of the relative abundance (% OTUs) were 24.27%, 23.54%, 16.63%, 3.95%, 3.47%, 2.59%, 2.37%, 2.26%, 2.22%, and 2.14%, respectively, and encompassed a total of 82.96% of the overall microbiota, as shown in Figure 2. Moreover, a heatmap of bacterial genera with significant differences between samples is shown in Figure 3.

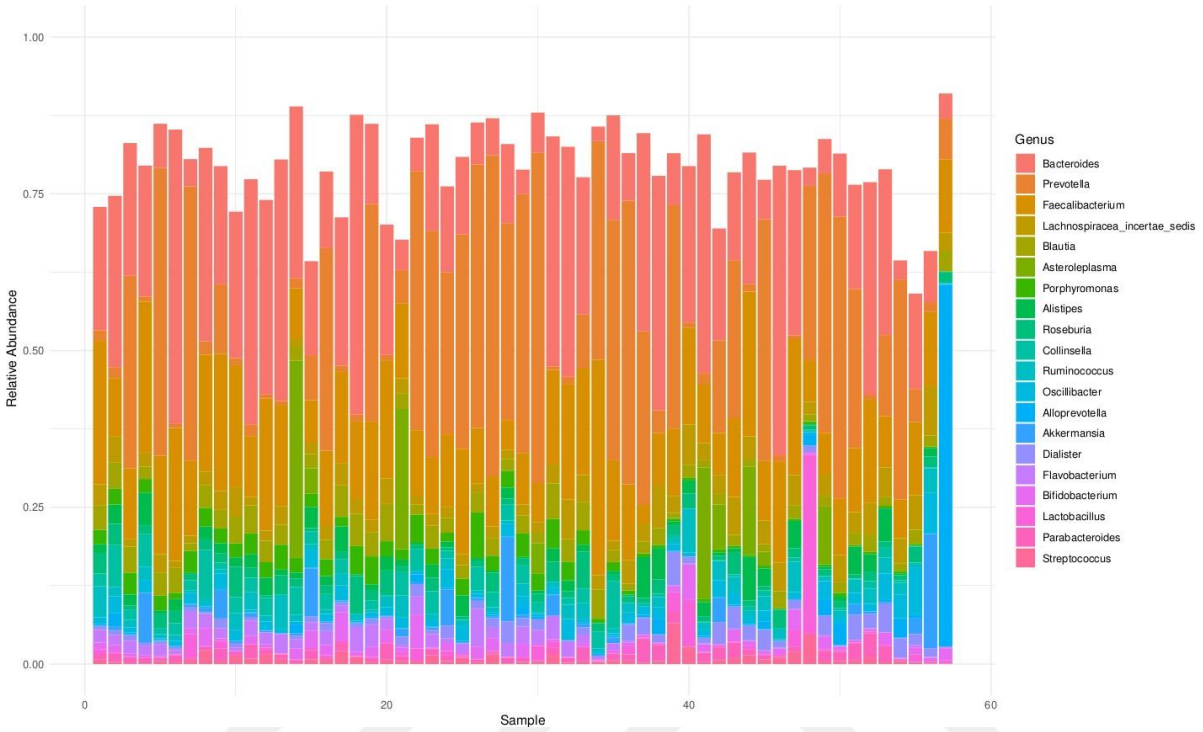


**Figure 2: Stacked bar plot showing the top 20 most abundant bacterial genera.**

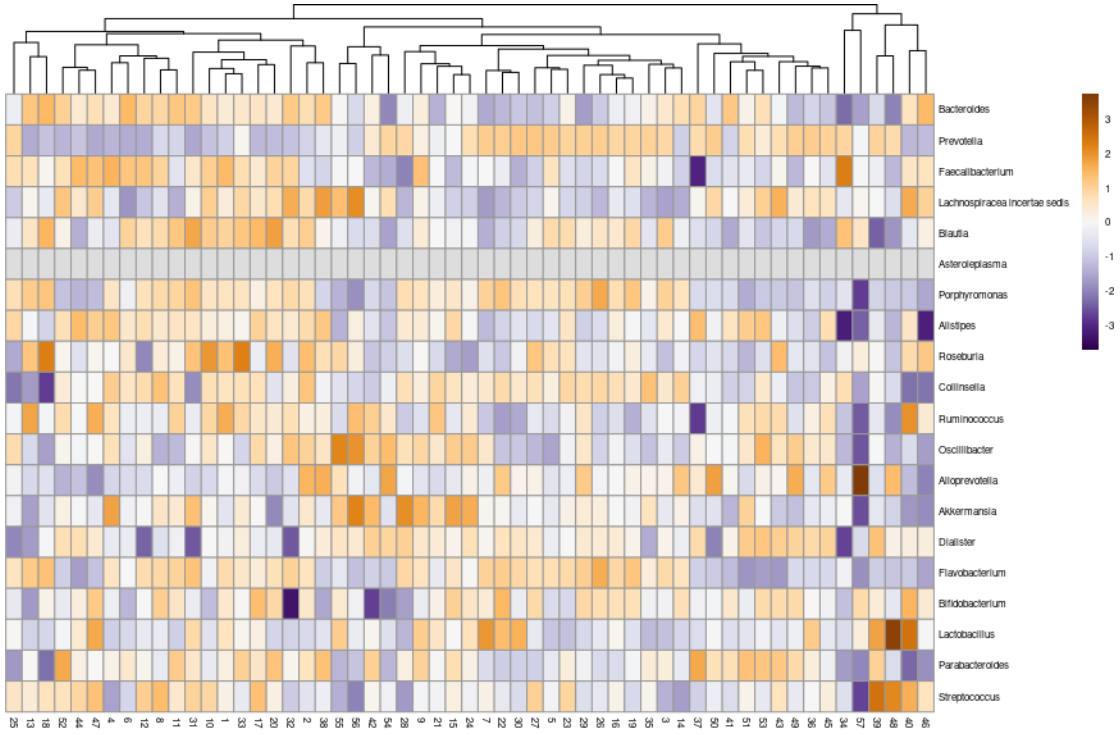


**Figure 3: Heatmap of bacterial genera with significant differences between samples.**

The z-score was used to standardize the data on relative abundance by row. The plot wasgenerated using the pheatmap R package with the default settings.

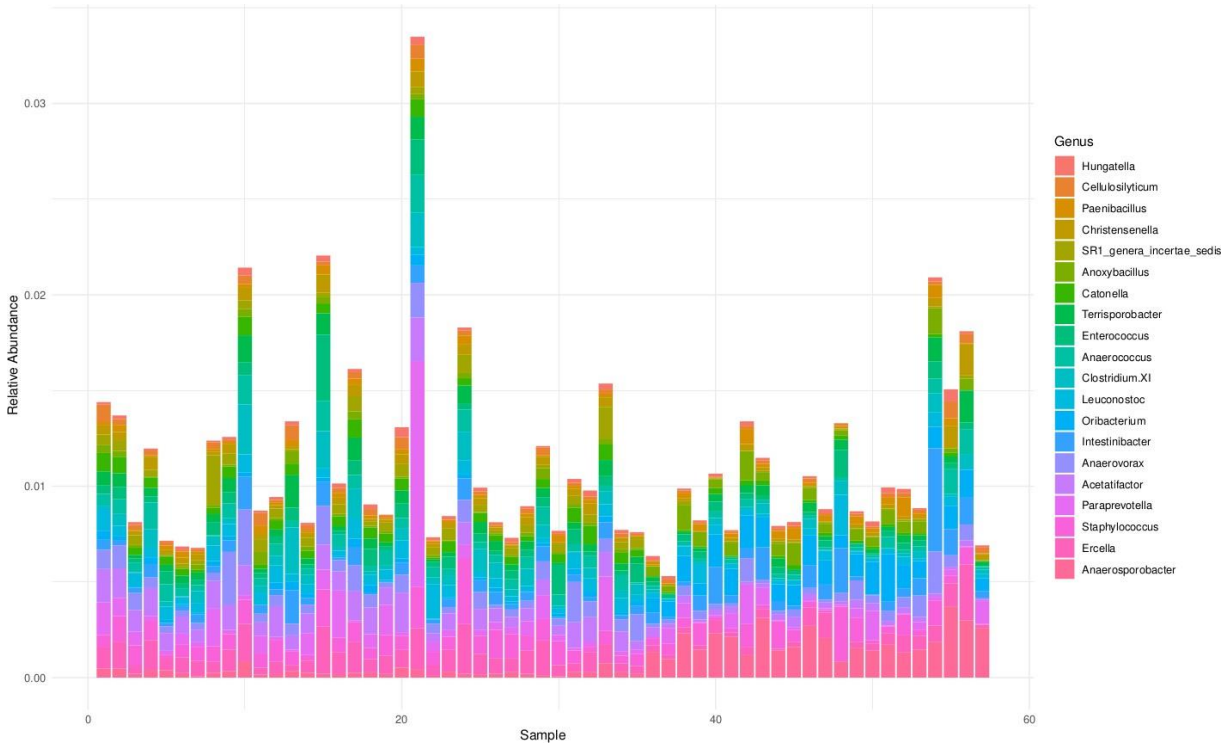


**Figure 4: Stacked bar plot showing the least 20 bacterial genera**.

A list of the 20 bacterial genera with the lowest incidence in all samples is shown in Figure 4. These genera were Hungatella, Cellulosilyticum, Paenibacillus, Christensenalla and SR1_genera_incertae_sedis, Anoxybacillus, Catonalla, Terrisporobacter, Enterococcus, and Anaerococcus, which encompassed a total of 2% of the overall microbiota.

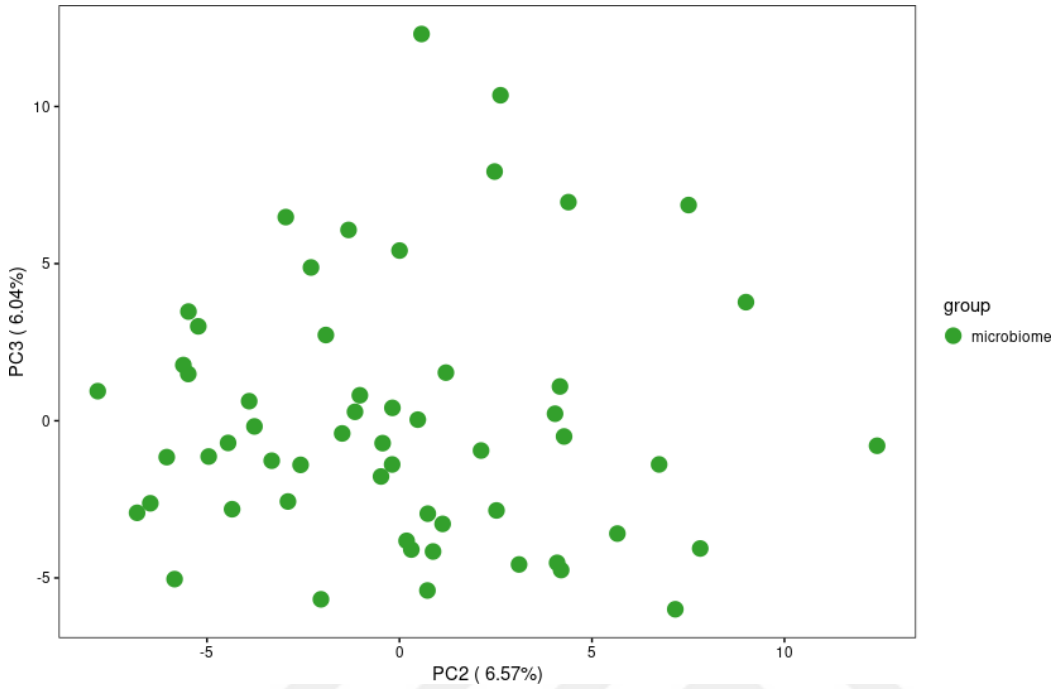


**Figure 5: Principal component analysis of the profiling of the gut microbiome at genus level.**

Figure 5 shows the normalized relative microbial diversity at the genus level of all samples using the Principal Component Analysis (PCA) approach. PCA evaluates variations in the distribution of taxonomic classifications among samples up to a predetermined taxonomic level. Clustered samples exhibit a similar microbial distribution throughout the population. The PCA estimate was based on the genus-level microbial profile.,

### 4.2 Pathway Abundance Analysis

Microbial data obtained by analysis of 16S sequencing data were mapped to metabolic pathways associated with the gut microbiota. In this way, functional abundances specific to the population of each metabolic pathway were obtained. For this purpose, the ampliseq v2.3.1 workflow developed by nf-core using Nextflow v21.10.6 was used for bioinformatics analysis of 16S sequencing data. By using the Picrust2 tool included in this workflow, functional abundances were obtained based on marker gene sequences in previously defined metabolic pathways.

While analyzing the functional abundance of gut microbiota in metabolic pathways, a custom database was created using the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. While creating the custom database, first of all, the relevant pathways and their ko ids were obtained manually, and all KO (KEGG orthology) ids associated with these ko ids were detected using the Python programming language. A Heatmap of microbial signatures in 30 different metabolic pathways of interest in 16S rRNA sequencing data from samples collected from volunteers is shown in Figure 6.

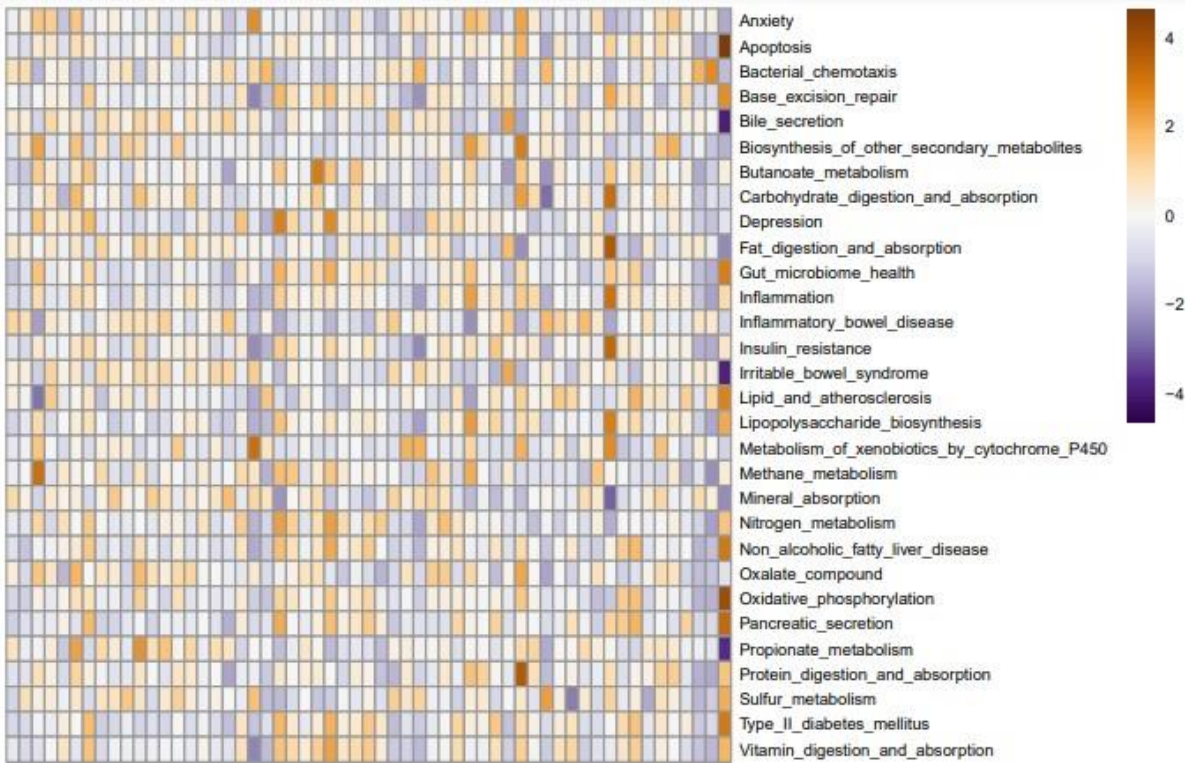


**Figure 6: Heatmap of microbial signatures in 30 different metabolic pathways of interest in 16S rRNA sequencing data from samples collected from volunteers.**

The horizontal axis represents the samples, and the vertical axis represents the metabolic pathways in Figure 6.

### 4.3 Rereference Interval Estimation

Reference intervals were determined for each metabolic pathway using the refineR (Reference Interval Estimation using Real-World Data) package developed in the R environment. The reference interval table and graph for the 28 different metabolic pathways are shown below.

#### 4.3.1 Anxiety

The reference interval result table and graph for the anxiety metabolic pathway are shown in Table 1 and Figure 7. The lower limit of the anxiety pathway is 0.00254, and the upper limit is 0.00262, for which reference interval estimation was made for 57 samples.

**Table 1: Reference interval estimation results for anxiety metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.00254 |
| upper limit [97.5% perc] | 0.00262 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 1.36 |
| mu | -0.736 |
| sigma | 2.39e-06 |
| shift | 0 |
| cost | 5.17 |
| NP fraction | 0.4 |

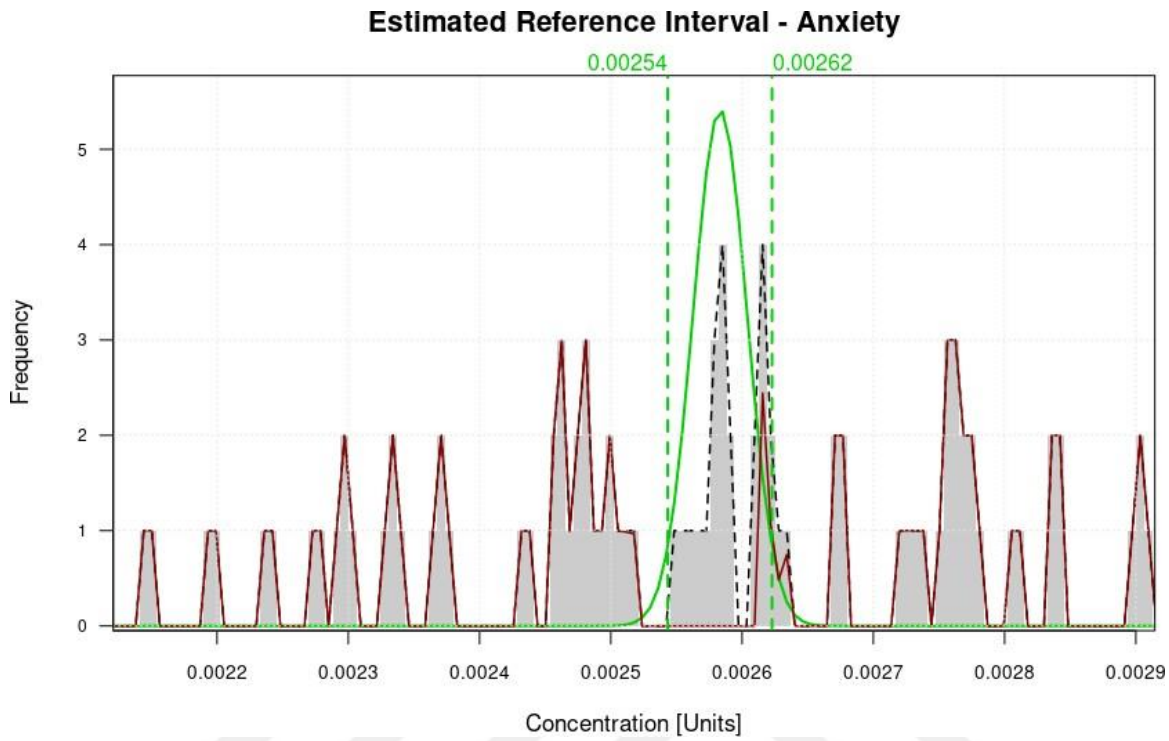


**Figure 7: Graph of reference interval estimation result for anxiety metabolic pathway.**

### 4.3.2 Bacterial chemotaxis

The reference interval result table and graph for bacterial chemotaxis metabolic pathway are shown in Table 2 and Figure 8. The lower limit of the bacterial chemotaxis pathway is 0.0977 and the upper limit is 0.114 for which reference interval estimation was made for 57 samples.

**Table 2: Reference interval estimation results for bacterial chemotaxis metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0977 |
| upper limit [97.5% perc] | 0.114 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 1.09 |
| mu | -0.836 |
| sigma | 0.00339 |
| shift | 0 |
| cost | 2.5 |
| NP fraction | 0.41 |

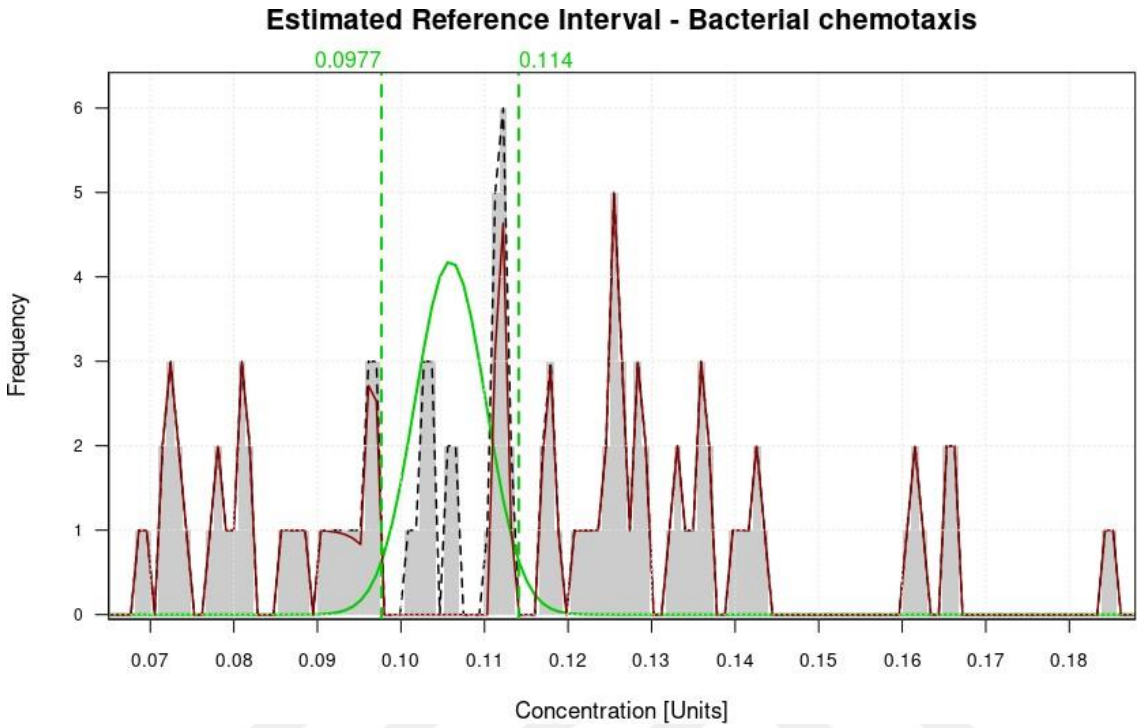


**Figure 8: Graph of reference interval estimation result for bacterial chemotaxis metabolic pathway.**

### 4.3.3. Base excision repair

The reference interval result table and graph for base excision repair metabolic pathway are shown in Table 3 and Figure 9. The lower limit of the base excision pathway is 0.0711 and the upper limit is 0.0727 for which reference interval estimation was made for 57 samples.

**Table 3: Reference interval estimation results for base excision repair metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0711 |
| upper limit [97.5% perc] | 0.0727 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.799 |
| mu | -1.1 |
| sigma | 0.000731 |
| shift | 0 |
| cost | 5.36 |
| NP fraction | 0.4 |

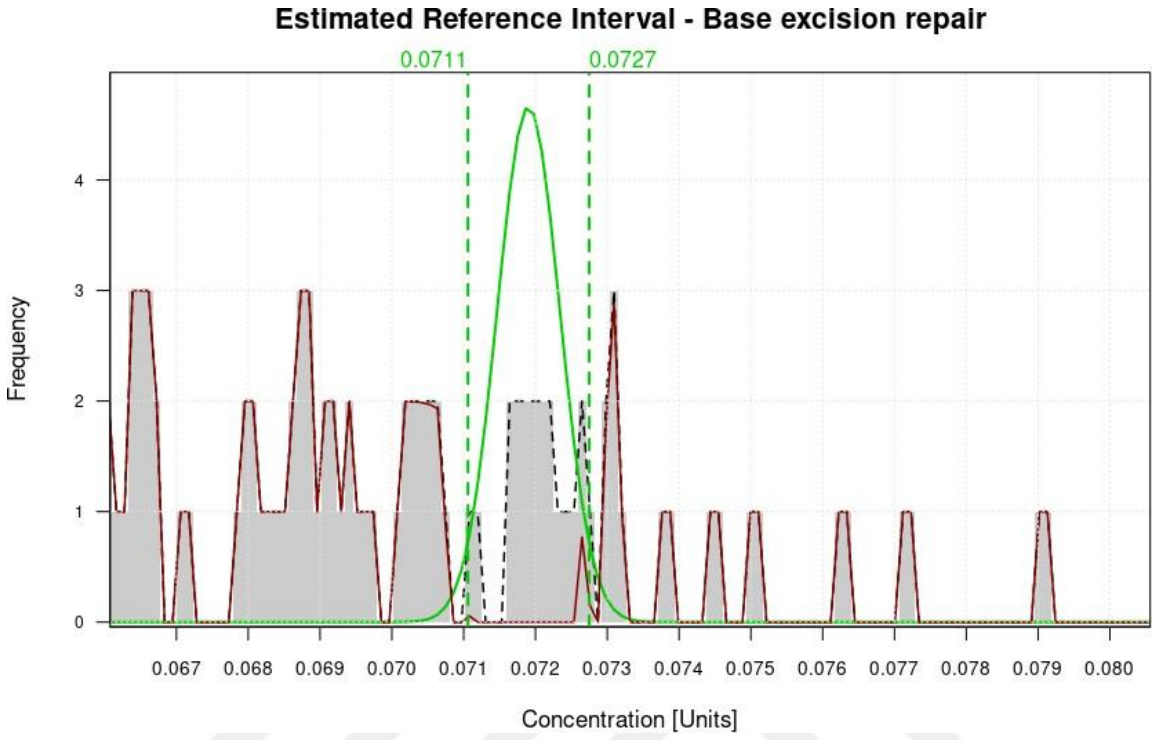


**Figure 9: Graph of reference interval estimation result for base excision repair metabolic pathway.**

### 4.3.4. Bile secretion

The reference interval result table and graph for bile secretion metabolic pathway are shown in Table 4 and Figure 10. The lower limit of the bile secretion pathway is 0.00351 and the upper limit is 0.0382 for which reference interval estimation was made for 57 samples.

**Table 4: Reference interval estimation results for bile secretion metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.00351 |
| upper limit [97.5% perc] | 0.00382 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.00875 |
| mu | -5.47 |
| sigma | 0.02 |
| shift | 0 |
| cost | -2.03 |
| NP fraction | 0.4 |

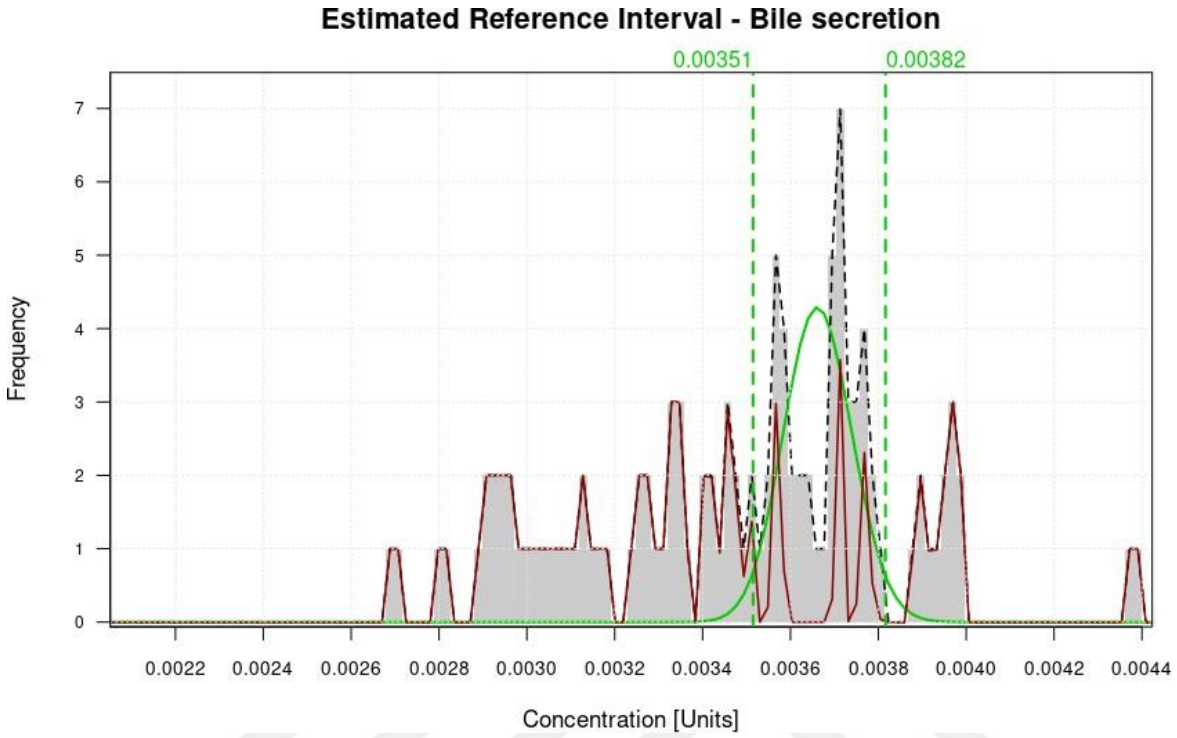


**Figure 10: Graph of reference interval estimation result for bile secretion metabolic pathway.**

### 4.3.5. Biosynthesis of other secondary metabolites

The reference interval result table and graph for biosynthesis of other secondary metabolites metabolic pathway are shown in Table 5 and Figure 11. The lower limit of the biosynthesis ofm other secondary metabolites pathway is 0.0103 and the upper limit is 0.0123 for which reference interval estimation was made for 57 samples.

**Table 5: Reference interval estimation results for biosynthesis of other secondary metabolites metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0103 |
| upper limit [97.5% perc] | 0.0123 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.417 |
| mu | -2.03 |
| sigma | 0.00684 |
| shift | 0 |
| cost | -1.97 |
| NP fraction | 0.4 |

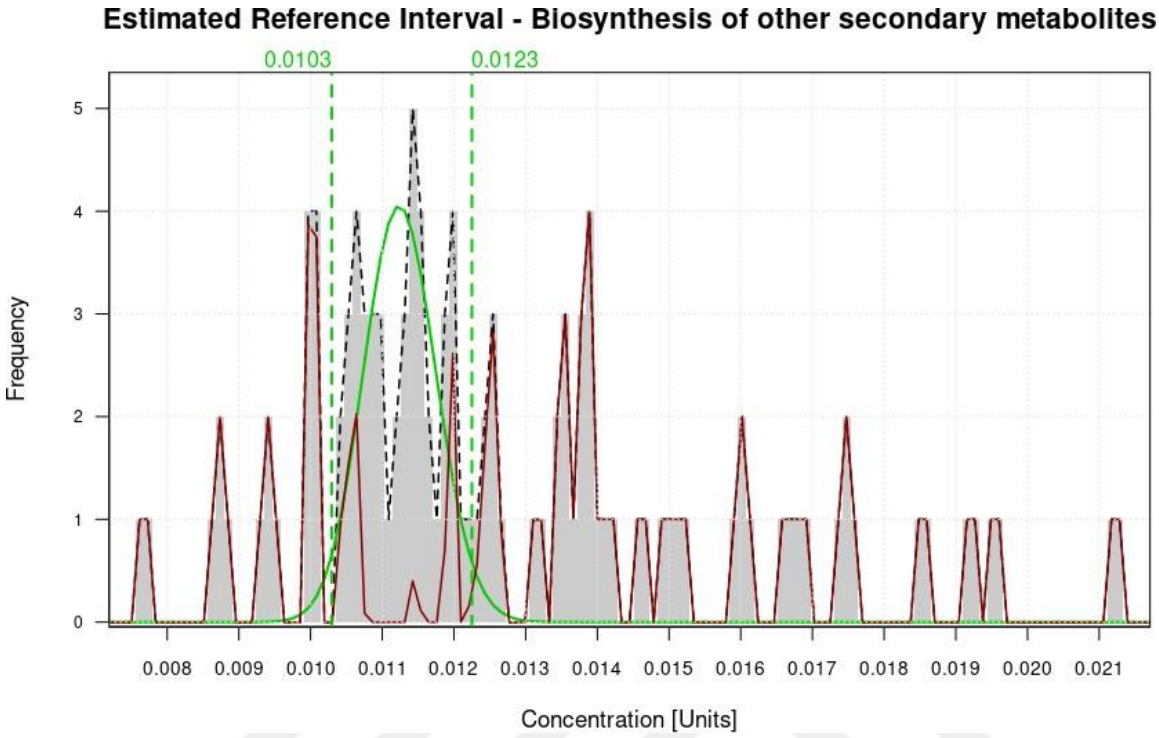


**Figure 11: Graph of reference interval estimation result for biosynthesis of other secondary metabolites metabolic pathway.**

### 4.3.6. Butanoate metabolism

The reference interval result table and graph for butanoate metabolism metabolic pathway are shown in Table 6 and Figure 12. The lower limit of the butanoate metabolism pathway is 0.0601 and the upper limit is 0.0621 for which reference interval estimation was made for 57 samples.

**Table 6: Reference interval estimation results for butanoate metabolism metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0601 |
| upper limit [97.5% perc] | 0.0621 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.0721 |
| mu | -2.53 |
| sigma | 0.00704 |
| shift | 0 |
| cost | -1.76 |
| NP fraction | 0.4 |

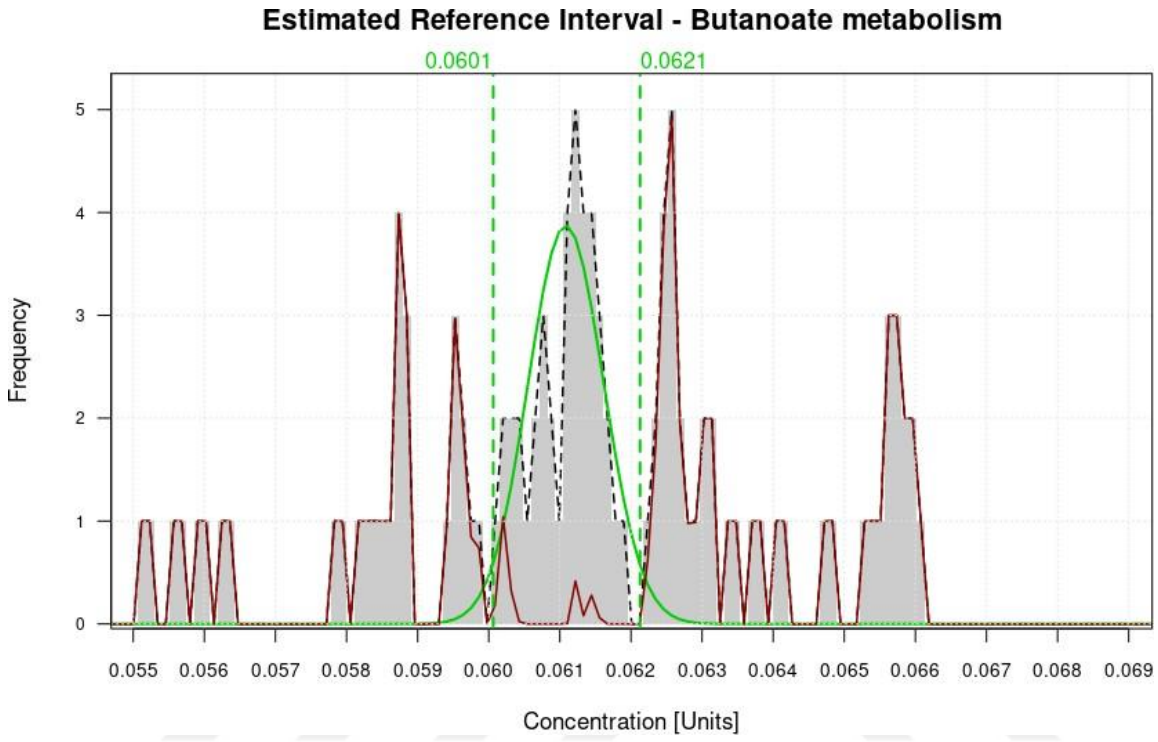


**Figure 12: Graph of reference interval estimation result for butanoate metabolism metabolic pathway.**

### 4.3.7. Carbohydrate digestion and absorption

The reference interval result table and graph for carbohydrate digestion and absorption metabolic pathway are shown in Table 7 and Figure 13. The lower limit of the carbohydrate digestion and absorption pathway is 0.032 and the upper limit is 0.0324 for which reference interval estimation was made for 57 samples.

**Table 7: Reference interval estimation results for carbohydrate digestion and absorption metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.032 |
| upper limit [97.5% perc] | 0.0324 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.136 |
| mu | -2.75 |
| sigma | 0.00211 |
| shift | 0 |
| cost | 5.87 |
| NP fraction | 0.4 |

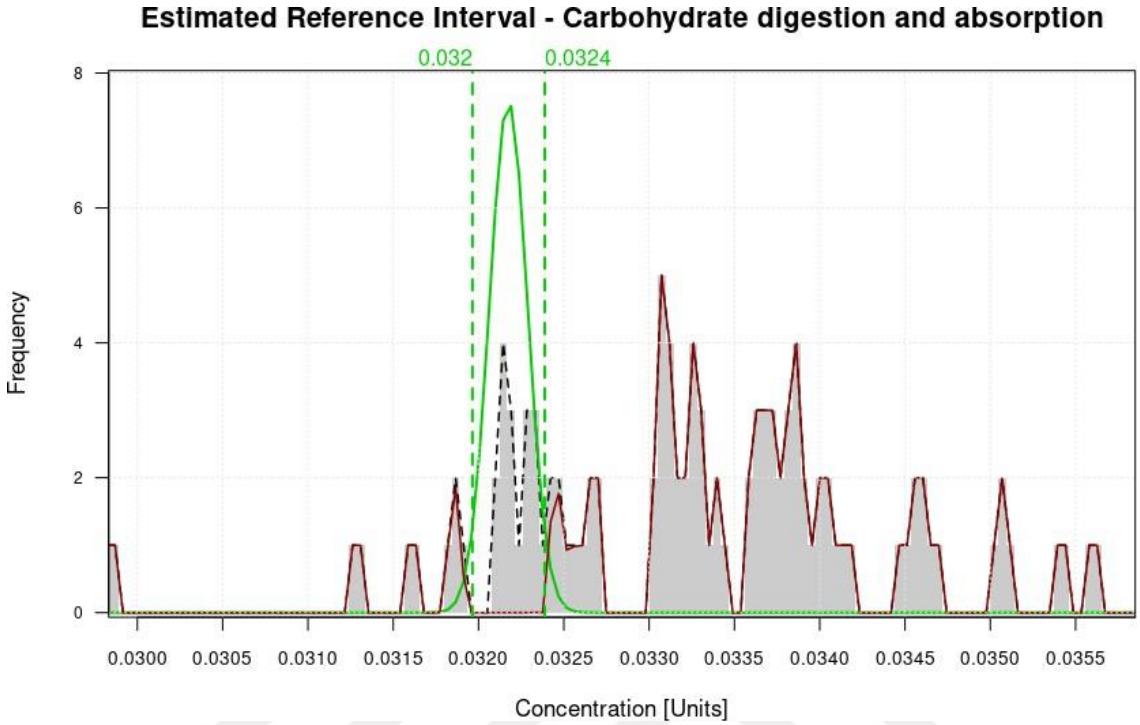


**Figure 13: Graph of reference interval estimation result for carbohydrate digestion and absorption metabolic pathway.**

### 4.3.8. Depression

The reference interval result table and graph for depression metabolic pathway are shown in Table 8 and Figure 14. The lower limit of the depression pathway is 0.00164 and the upper limit is 0.00172 for which reference interval estimation was made for 57 samples.

**Table 8: Reference interval estimation results for depression metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.00164 |
| upper limit [97.5% perc] | 0.00172 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 1 |
| mu | -0.998 |
| sigma | 2e-05 |
| shift | 0 |
| cost | 4.84 |
| NP fraction | 0.4 |

,

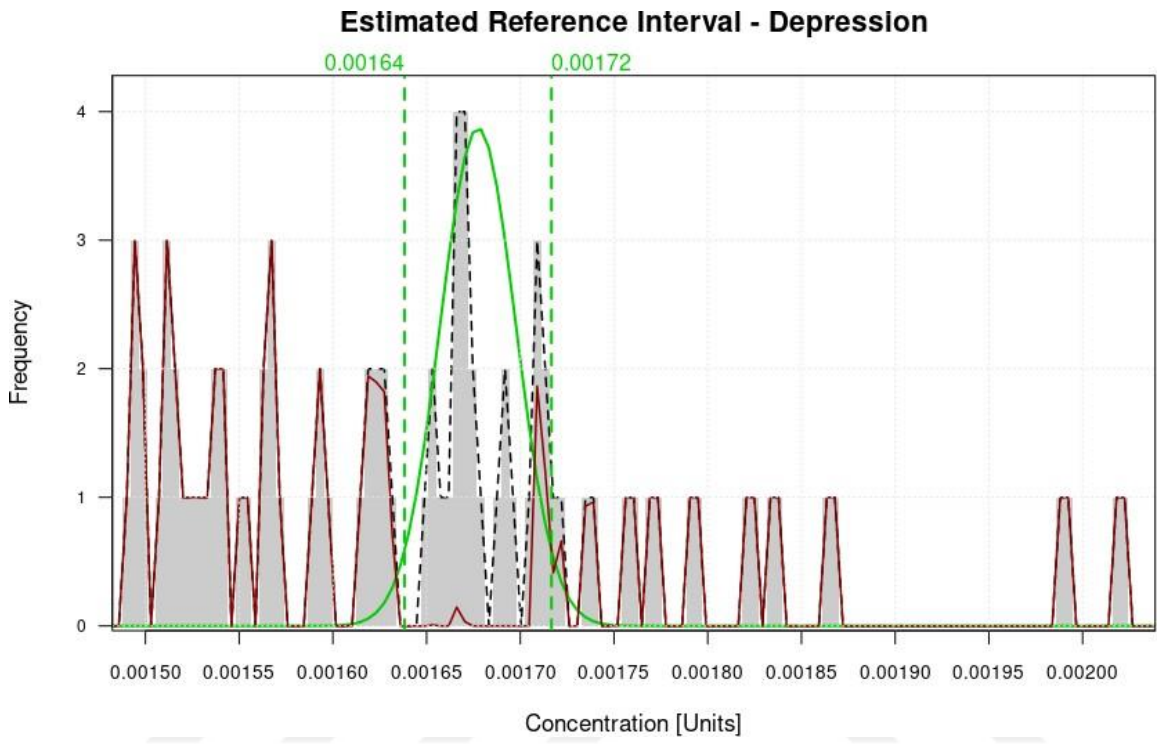


**Figure 14: Graph of reference interval estimation result for depression metabolic pathway.**

### 4.3.9. Fat digestion and absorption

The reference interval result table and graph for fat digestion and absorption metabolic pathway are shown in Table 9 and Figure 15. The lower limit of the fat digestion and absorption pathway is 0.0143 and the upper limit is 0.0154 for which reference interval estimation was made for 57 samples.

**Table 9: Reference interval estimation results for fat digestion and absorption metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0143 |
| upper limit [97.5% perc] | 0.0154 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.479 |
| mu | -1.81 |
| sigma | 0.00257 |
| shift | 0 |
| cost | -1.66 |
| NP fraction | 0.4 |

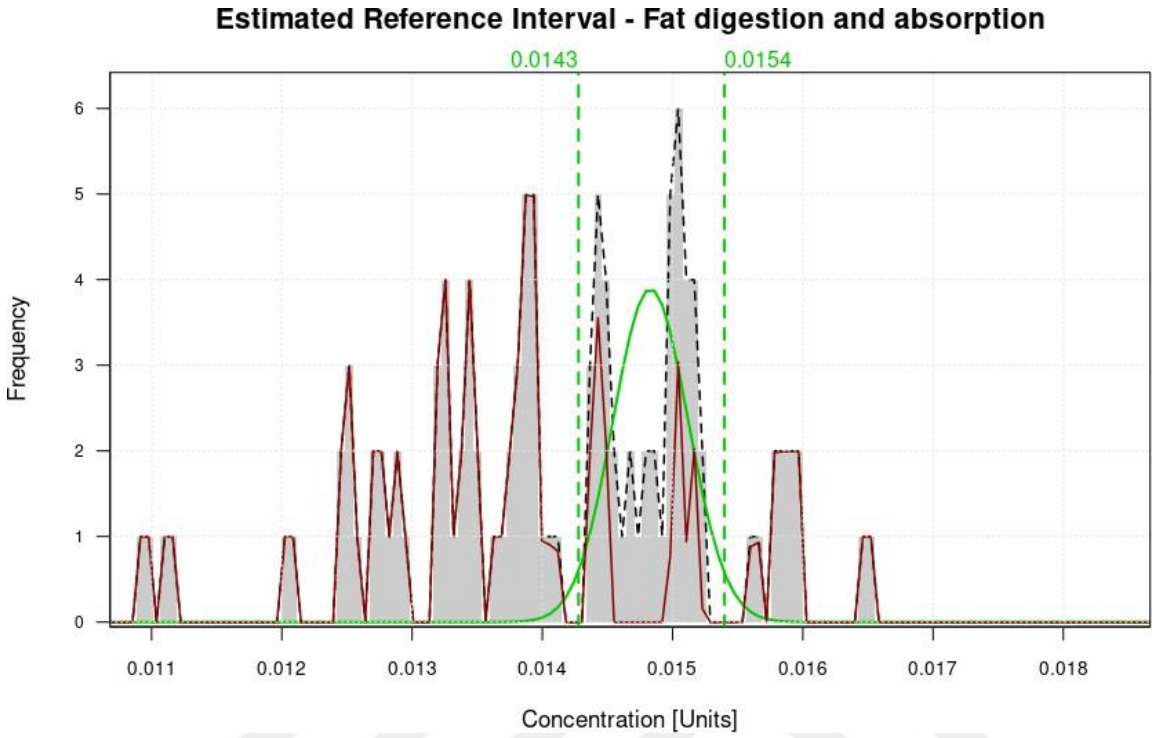


**Figure 15: Graph of reference interval estimation result for fat digestion and absorption metabolic pathway.**

### 4.3.10. Gut microbiome health

The reference interval result table and graph for gut microbiome health metabolic pathway are shown in Table 10 and Figure 16. The lower limit of the gut microbiome health pathway is 0.0181 and the upper limit is 0.0185 for which reference interval estimation was made for 57 samples.

**Table 10: Reference interval estimation results for gut microbiome health metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0181 |
| upper limit [97.5% perc] | 0.0185 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.32 |
| mu | -2.25 |
| sigma | 0.00146 |
| shift | 0 |
| cost | 6.76 |
| NP fraction | 0.4 |

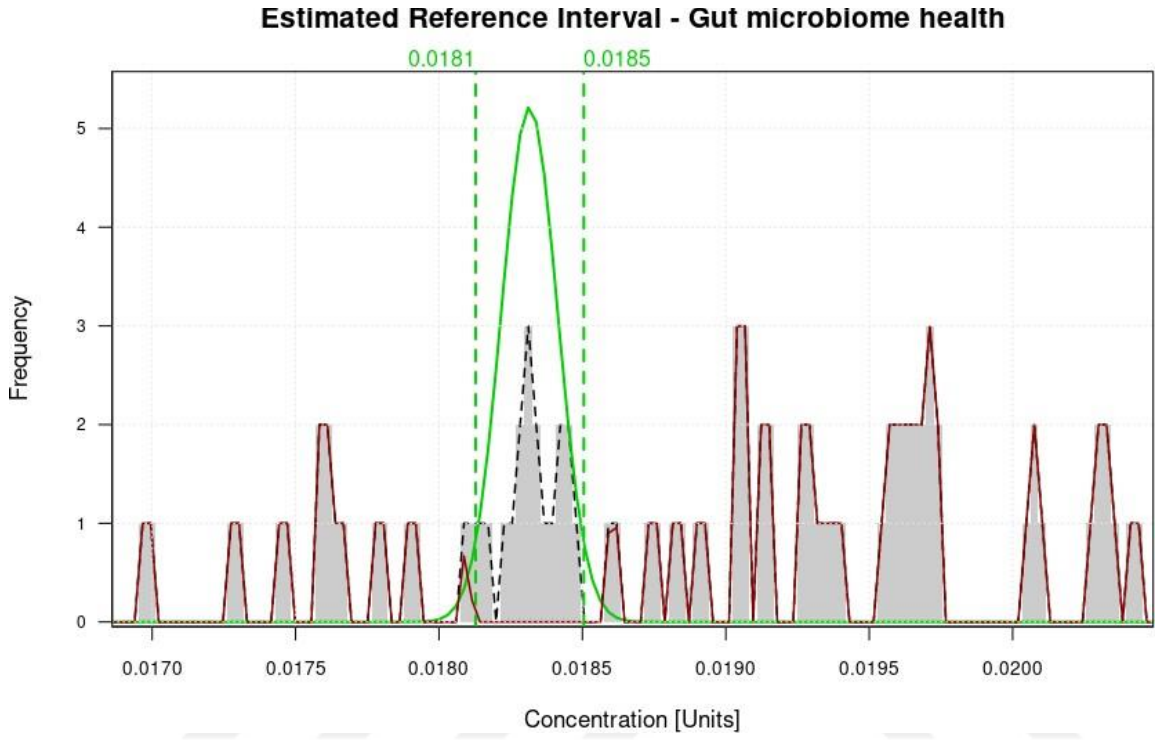


**Figure 16: Graph of reference interval estimation result for gut microbiome health metabolic pathway.**

### 4.3.11. Inflammation

The reference interval result table and graph for inflammation metabolic pathway are shown in Table 11 and Figure 17. The lower limit of the inflammation pathway is 0.0452 and the upper limit is 0.0469 for which reference interval estimation was made for 57 samples.

**Table 11: Reference interval estimation results for inflammation metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0452 |
| upper limit [97.5% perc] | 0.0469 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.0109 |
| mu | -3.03 |
| sigma | 0.0092 |
| shift | 0 |
| cost | 0.0897 |
| NP fraction | 0.4 |

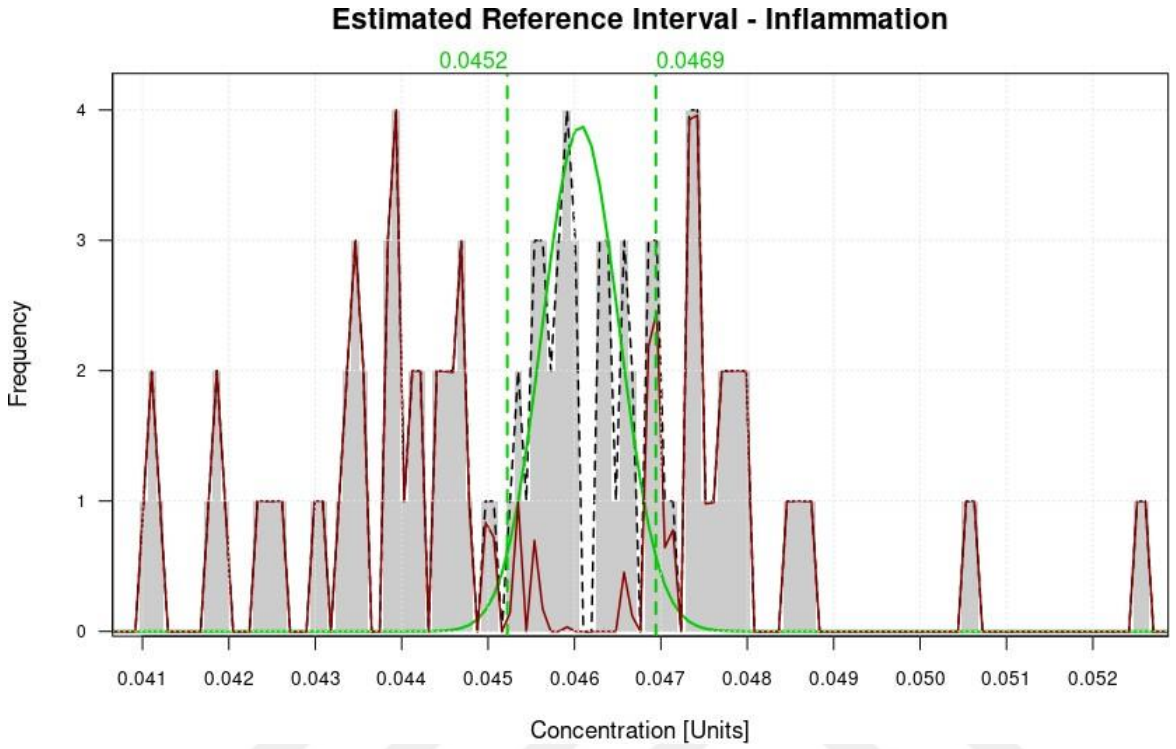


**Figure 17: Graph of reference interval estimation result for inflammation metabolic pathway.**

### 4.3.12. Inflammatory bowel disease

The reference interval result table and graph for inflammatory bowel disease metabolic pathway are shown in Table 12 and Figure 18. The lower limit of the inflammatory bowel disease pathway is 0.00333 and the upper limit is 0.0034 for which reference interval estimation was made for 57 samples.

**Table 12: Reference interval estimation results for inflammatory bowel disease metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.00333 |
| upper limit [97.5% perc] | 0.0034 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.00219 |
| mu | -5.66 |
| sigma | 0.00541 |
| shift | 0 |
| cost | 5.03 |
| NP fraction | 0.4 |

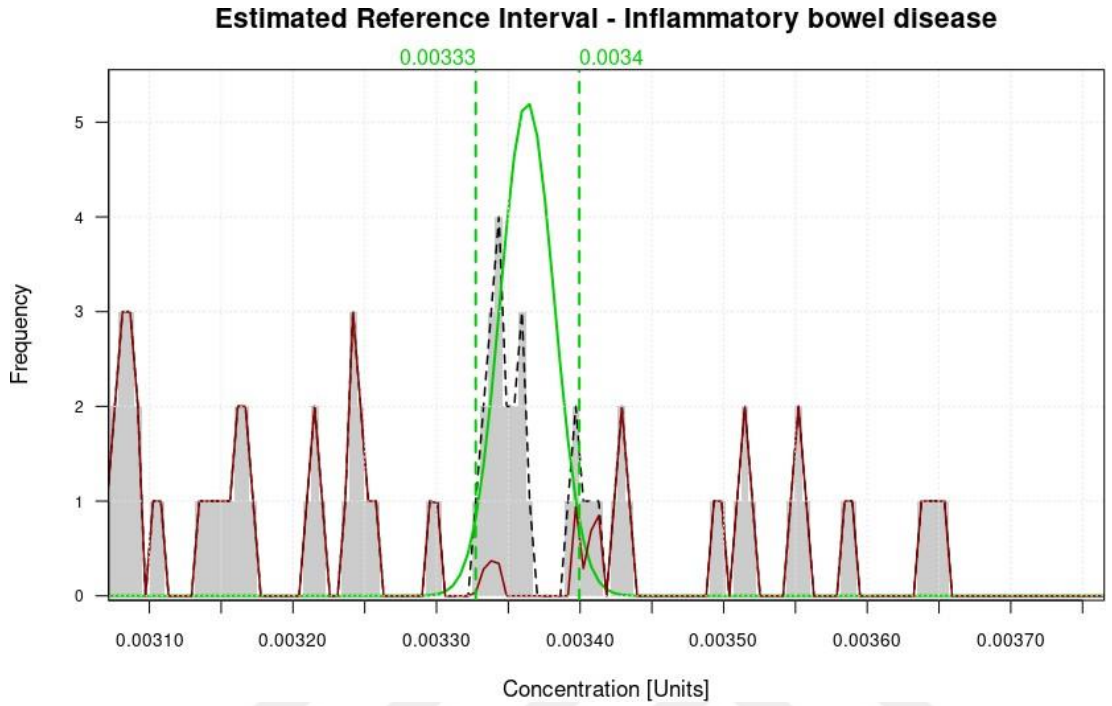


**Figure 18: Graph of reference interval estimation result for inflammatory bowel disease metabolic pathway.**

### 4.3.13. Insulin resistance

The reference interval result table and graph for insulin resistance metabolic pathway are shown in Table 13 and Figure 19. The lower limit of the insulin resistance pathway is 0.0421 and the upper limit is 0.0435 for which reference interval estimation was made for 57 samples.

**Table 13: Reference interval estimation results for insulin resistance metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0421 |
| upper limit [97.5% perc] | 0.0435 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.0109 |
| mu | -3.1 |
| sigma | 0.00787 |
| shift | 0 |
| cost | -2.04 |
| NP fraction | 0.4 |

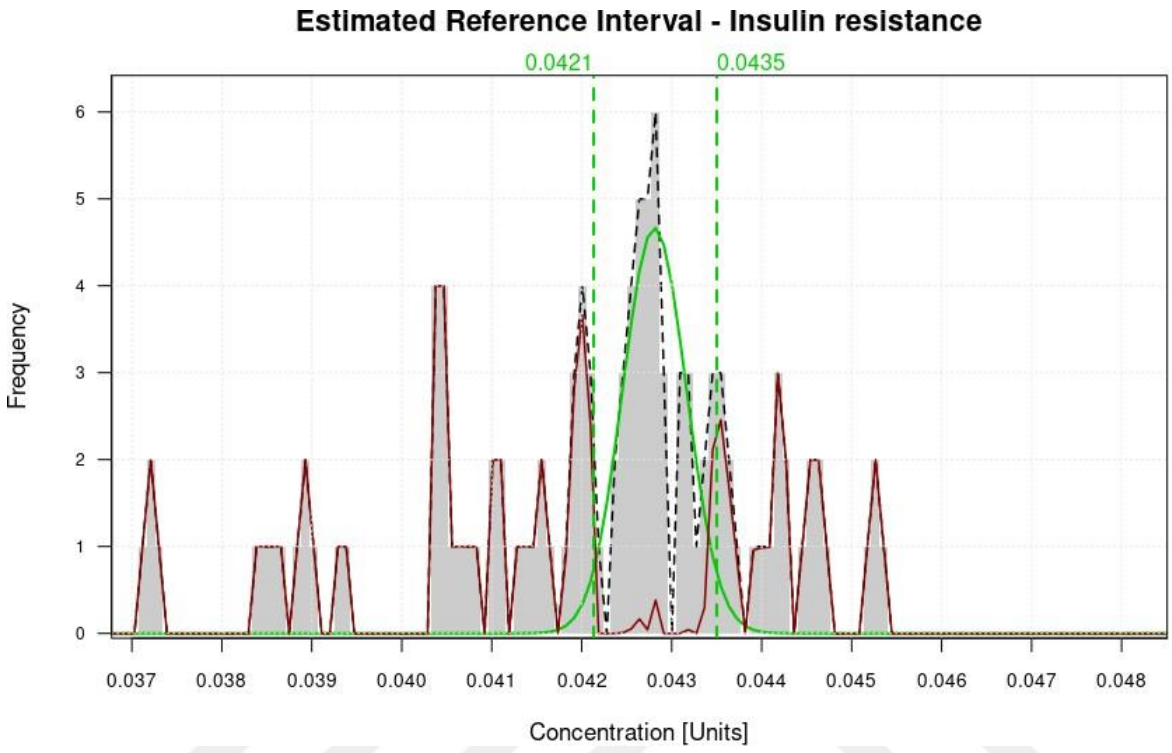


**Figure 19: Graph of reference interval estimation result for insulin resistance metabolic pathway.**

### 4.3.14. Irritable bowel syndrome

The reference interval result table and graph for irritable bowel syndrome metabolic pathway are shown in Table 14 and Figure 20. The lower limit of the irritable bowel syndrome pathway is 0.00129 and the upper limit is 0.00141 for which reference interval estimation was made for 57 samples.

**Table 14: Reference interval estimation results for irritable bowel syndrome metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.00129 |
| upper limit [97.5% perc] | 0.00141 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.772 |
| mu | -1.29 |
| sigma | 0.000138 |
| shift | 0 |
| cost | -1.31 |
| NP fraction | 0.4 |

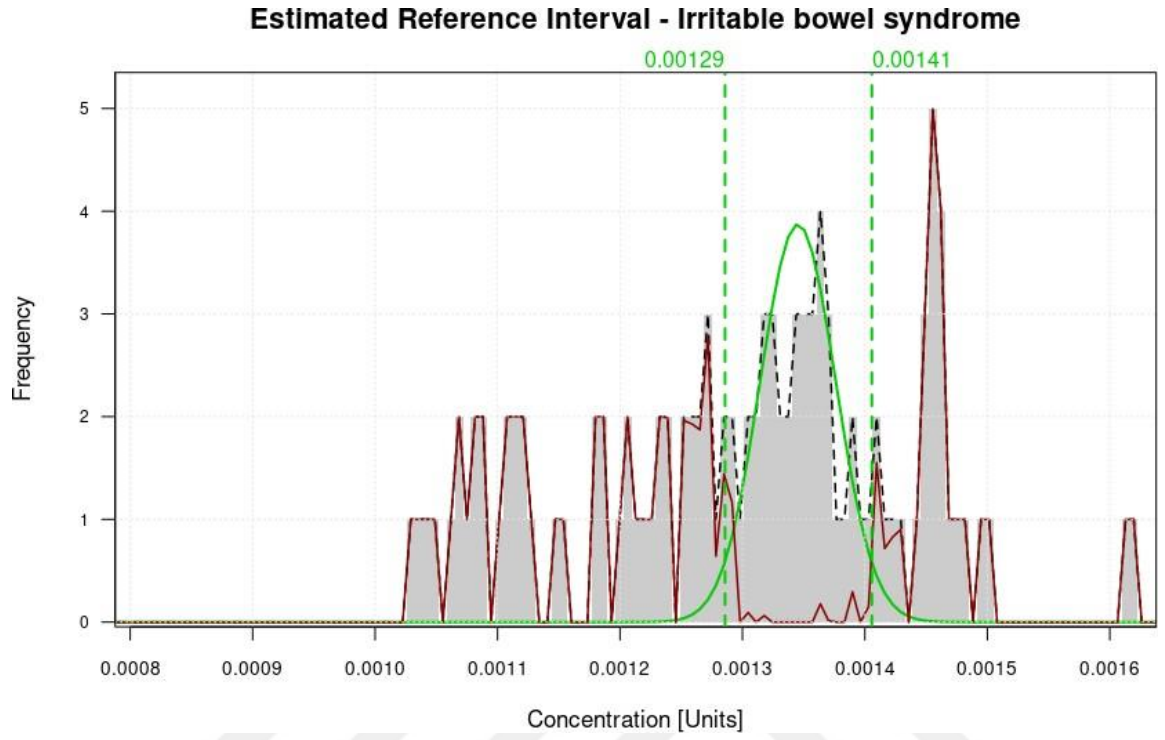


**Figure 20: Graph of reference interval estimation result for irritable bowel syndrome metabolic pathway.**

### 4.3.15. Lipid and atherosclerosis

The reference interval result table and graph for lipid and atherosclerosis metabolic pathway are shown in Table 15 and Figure 21. The lower limit of the lipid and atherosclerosis pathway is 0.00371 and the upper limit is 0.00389 for which reference interval estimation was made for 57 samples.

**Table 15: Reference interval estimation results for lipid and atherosclerosis metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.00371 |
| upper limit [97.5% perc] | 0.00389 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.942 |
| mu | -1.06 |
| sigma | 6.46e-05 |
| shift | 0 |
| cost | -0.452 |
| NP fraction | 0.4 |

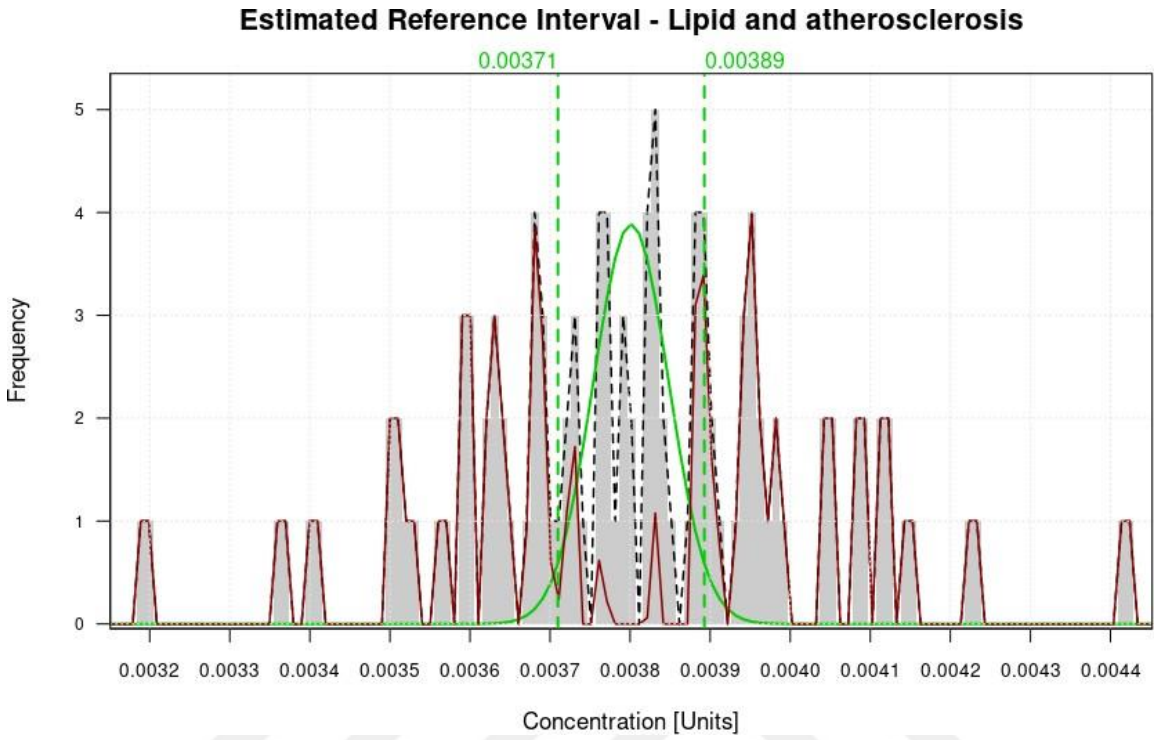


**Figure 21: Graph of reference interval estimation result for lipid and atherosclerosis metabolic pathway.**

### 4.3.16. Lipopolysaccharide biosynthesis

The reference interval result table and graph for lipopolysaccharide biosynthesis metabolic pathway are shown in Table 16 and Figure 22. The lower limit of the lipopolysaccharide biosynthesis pathway is 0.0819 and the upper limit is 0.0859 for which reference interval estimation was made for 57 samples.

**Table 16: Reference interval estimation results for lipopolysaccharide biosynthesis metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0819 |
| upper limit [97.5% perc] | 0.0859 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.502 |
| mu | -1.42 |
| sigma | 0.00356 |
| shift | 0 |
| cost | 0.1 |
| NP fraction | 0.4 |

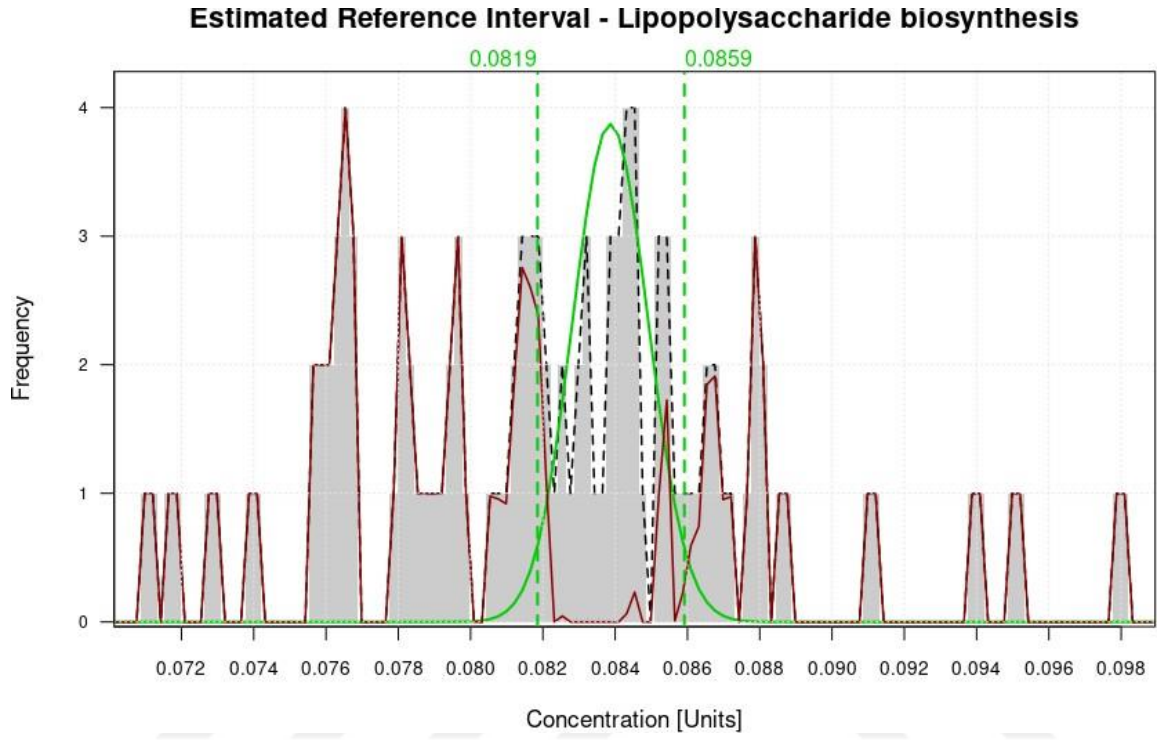


**Figure 22: Graph of reference interval estimation result for lipopolysaccharide biosynthesis metabolic pathway.**

### 4.3.17. Metabolism of xenobiotics by cytochrome P450

The reference interval result table and graph for metabolism of xenobiotics by cytochrome P450 metabolic pathway are shown in Table 17 and Figure 23. The lower limit of the metabolism of xenobiotics by cytochrome P450 metabolic pathway is 0.00194 and the upper limit is 0.00353 for which reference interval estimation was made for 57 samples.

**Table 17: Reference interval estimation results for Metabolism of xenobiotics by cytochrome P450metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.00194 |
| upper limit [97.5% perc] | 0.00353 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 1.5 |
| mu | -0.667 |
| sigma | 2.11e-05 |
| shift | 0 |
| cost | -2.25 |
| NP fraction | 0.4 |

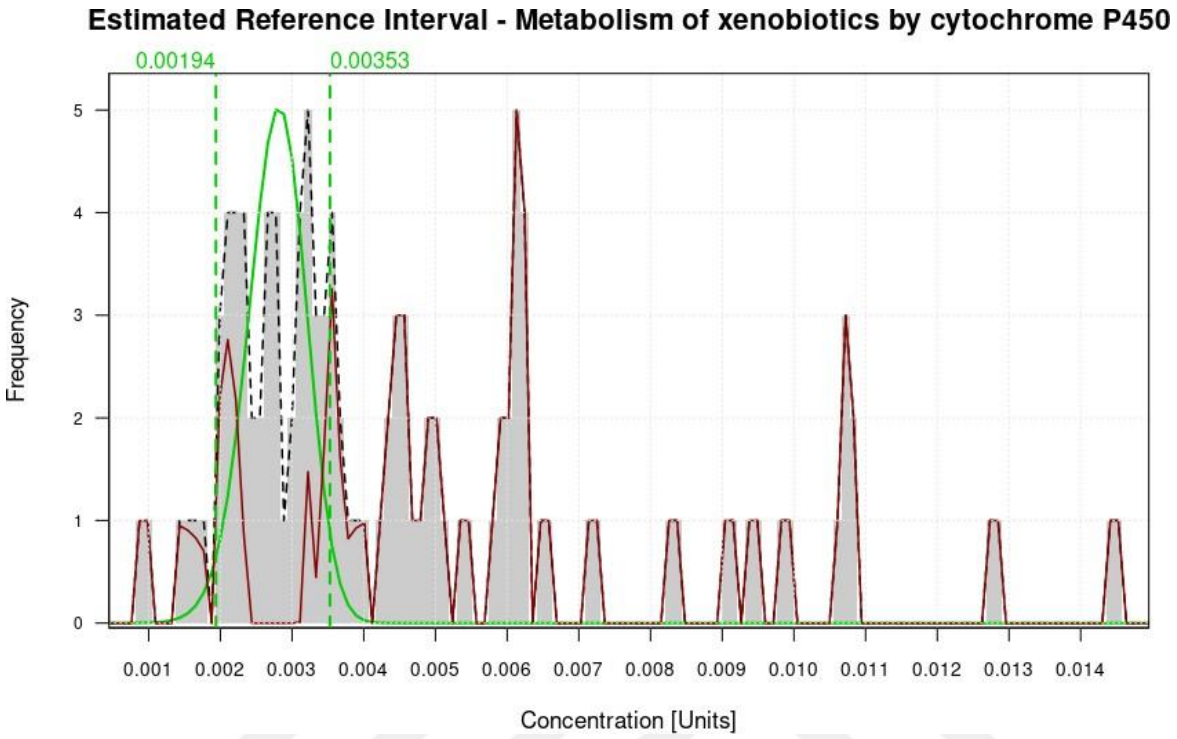


**Figure 23: Graph of reference interval estimation result for metabolism of xenobiotics by cytochrome P450 metabolic pathway.**

### 4.3.18. Methane metabolism

The reference interval result table and graph for methane metabolism metabolic pathway are shown in Table 18 and Figure 24. The lower limit of the methane metabolism pathway is 0.0472 and the upper limit is 0.048 for which reference interval estimation was made for 57 samples.

**Table 18: Reference interval estimation results for methane metabolism metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0472 |
| upper limit [97.5% perc] | 0.048 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.0386 |
| mu | -2.87 |
| sigma | 0.00371 |
| shift | 0 |
| cost | 6.3 |
| NP fraction | 0.401 |

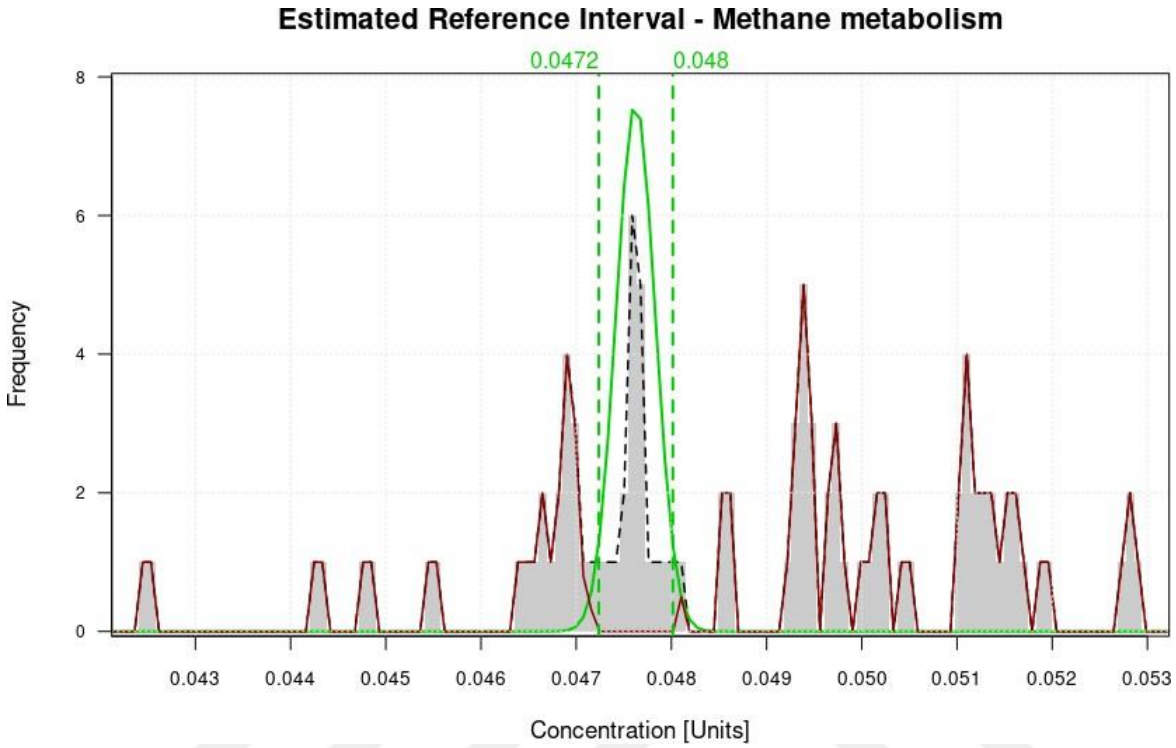


**Figure 24: Graph of reference interval estimation result for methane metabolism metabolic pathway.**

### 4.3.19. Mineral absorption

The reference interval result table and graph for mineral absorption metabolic pathway are shown in Table 19 and Figure 25. The lower limit of the mineral absorption pathway is 0.0289 and the upper limit is 0.0306 for which reference interval estimation was made for 57 samples.

**Table 19: Reference interval estimation results for mineral absorption metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0289 |
| upper limit [97.5% perc] | 0.0306 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 1.19 |
| mu | -0.828 |
| sigma | 0.000221 |
| shift | 0 |
| cost | 1.31 |
| NP fraction | 0.4 |

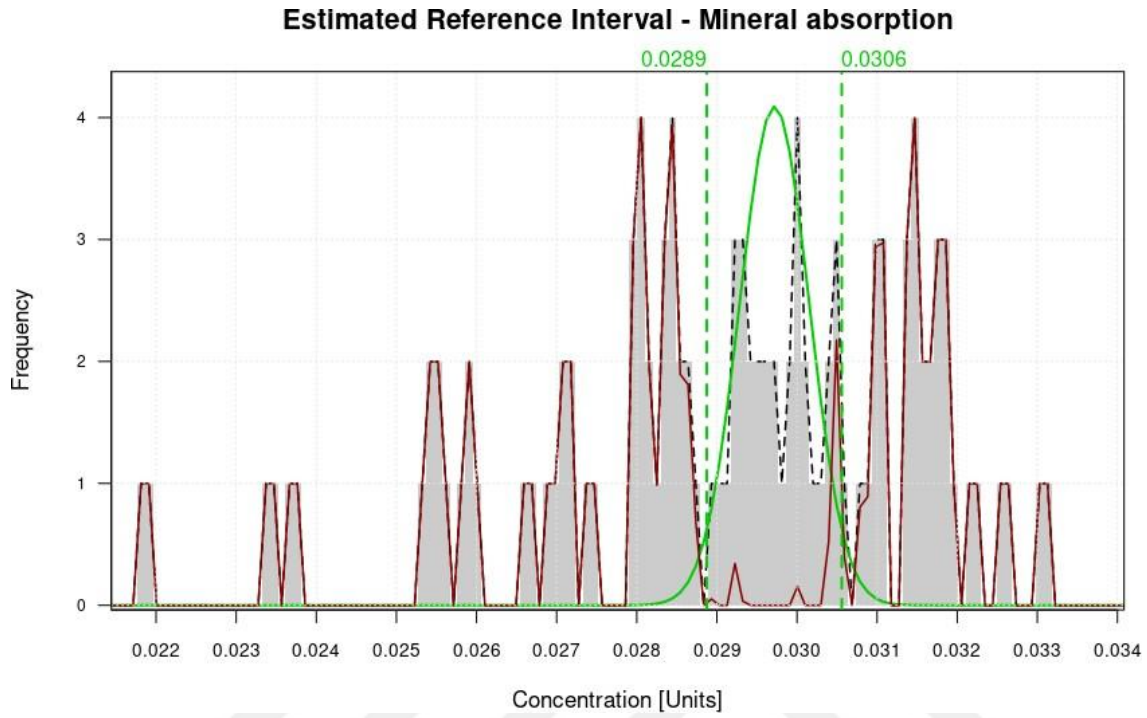


**Figure 25: Graph of reference interval estimation result for mineral absorption metabolic pathway.**

### 4.3.20. Nitrogen metabolism

The reference interval result table and graph for nitrogen metabolism metabolic pathway are shown in Table 20 and Figure 26. The lower limit of the nitrogen metabolism pathway is 0.0636 and the upper limit is 0.0652 for which reference interval estimation was made for 57 samples.

**Table 20: Reference interval estimation results for nitrogen metabolism metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0636 |
| upper limit [97.5% perc] | 0.0652 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 1 |
| mu | -0.936 |
| sigma | 0.000401 |
| shift | 0 |
| cost | 3.94 |
| NP fraction | 0.4 |

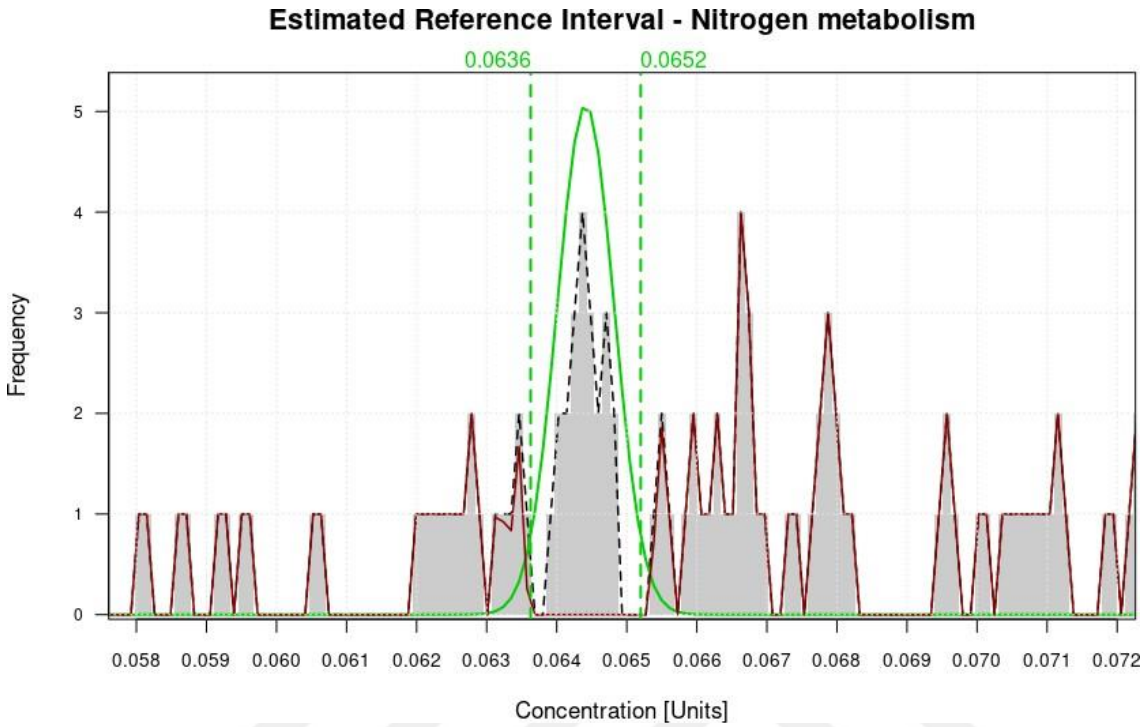


**Figure 26: Graph of reference interval estimation result for nitrogen metabolism metabolic pathway.**

### 4.3.21 Non alcoholic fatty liver disease

The reference interval result table and graph for non alcoholic fatty liver disease metabolic pathway are shown in Table 21 and Figure 27. The lower limit of the non alcoholic fatty liver disease pathway is 0.0225 and the upper limit is 0.0236 for which reference interval estimation was made for 57 samples.

**Table 21: Reference interval estimation results for non alcoholic fatty liver disease metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0225 |
| upper limit [97.5% perc] | 0.0236 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.204 |
| mu | -2.63 |
| sigma | 0.00585 |
| shift | 0 |
| cost | 1.16 |
| NP fraction | 0.4 |

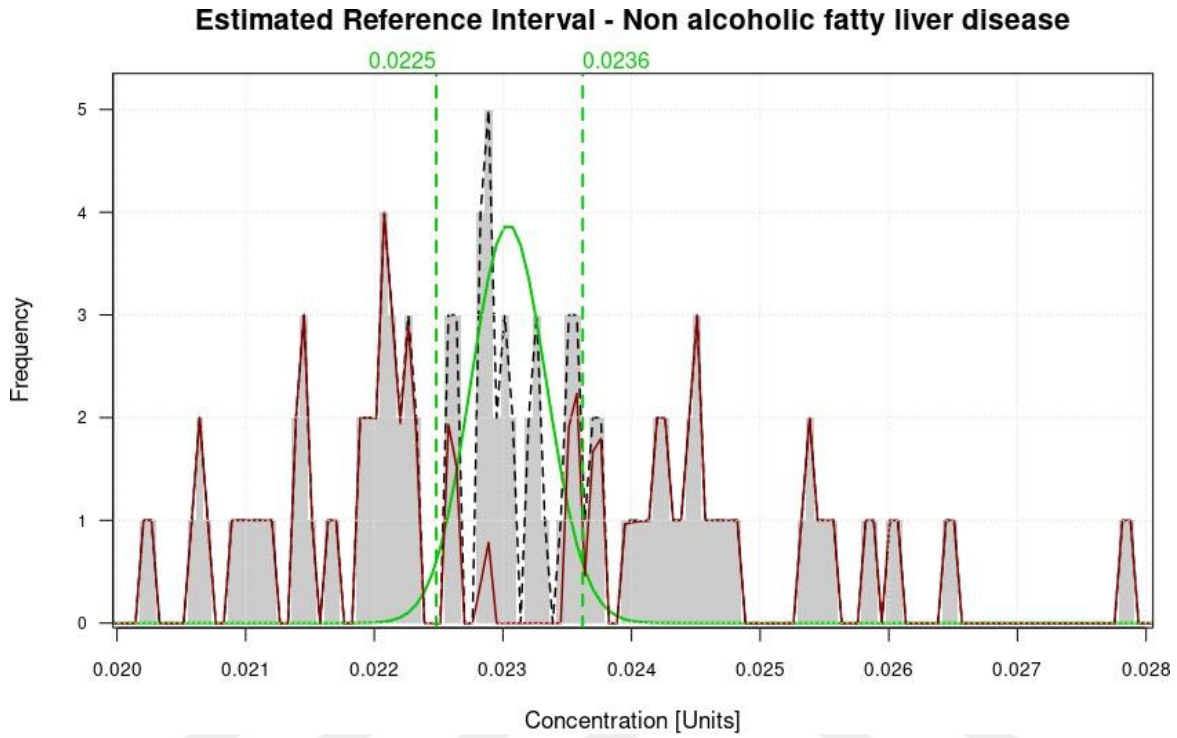


**Figure 27: Graph of reference interval estimation result for non alcoholic fatty liver disease metabolic pathway.**

### 4.3.22. Oxidative phosphorylation

The reference interval result table and graph for oxidative phosphorylation metabolic pathway are shown in Table 22 and Figure 28. The lower limit of the oxidative phosphorylation pathway is 0.0531 and the upper limit is 0.0559 for which reference interval estimation was made for 57 samples.

**Table 22: Reference interval estimation results for oxidative phosphorylation metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0531 |
| upper limit [97.5% perc] | 0.0559 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.396 |
| mu | -1.73 |
| sigma | 0.0042 |
| shift | 0 |
| cost | 0.181 |
| NP fraction | 0.4 |

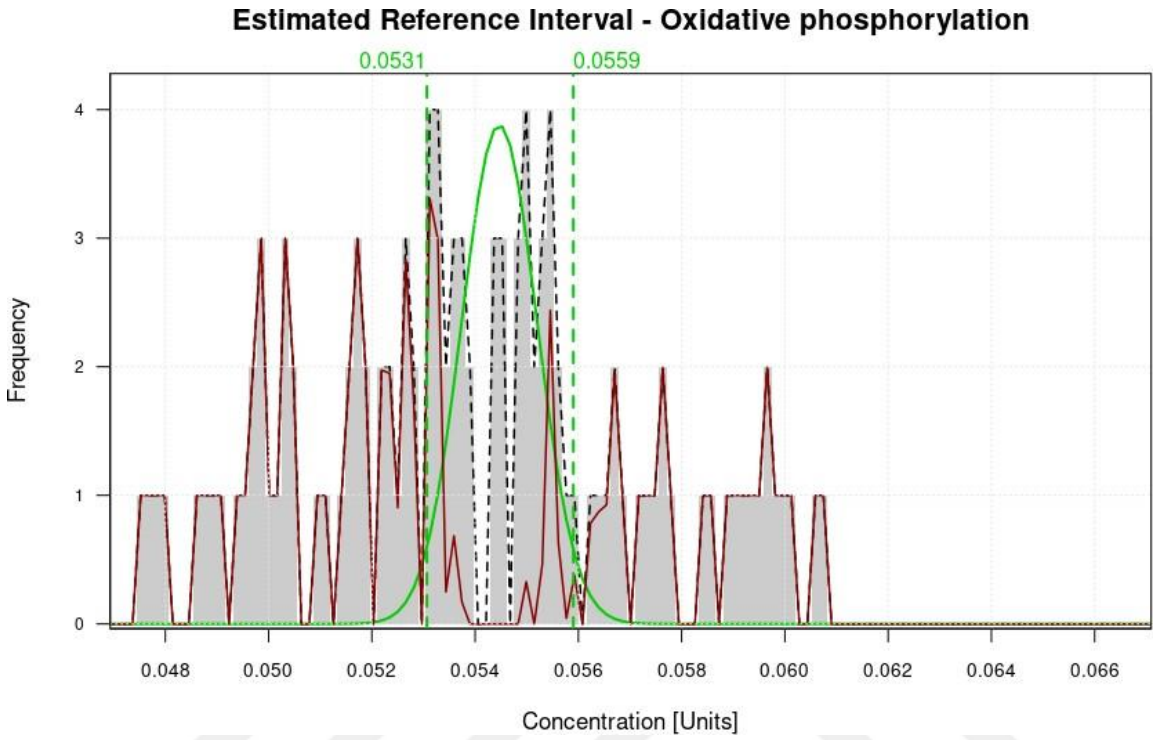


**Figure 28: Graph of reference interval estimation result for oxidative phosphorylation metabolic pathway.**

### 4.3.23. Pancreatic secretion

The reference interval result table and graph for pancreatic secretion metabolic pathway are shown in Table 23 and Figure 29. The lower limit of the pancreatic secretion pathway is 0.00103and the upper limit is 0.00138 for which reference interval estimation was made for 57 samples.

**Table 23: Reference interval estimation results for pancreatic secretion metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.00103 |
| upper limit [97.5% perc] | 0.00138 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.502 |
| mu | -1.93 |
| sigma | 0.00254 |
| shift | 0 |
| cost | -1.93 |
| NP fraction | 0.4 |

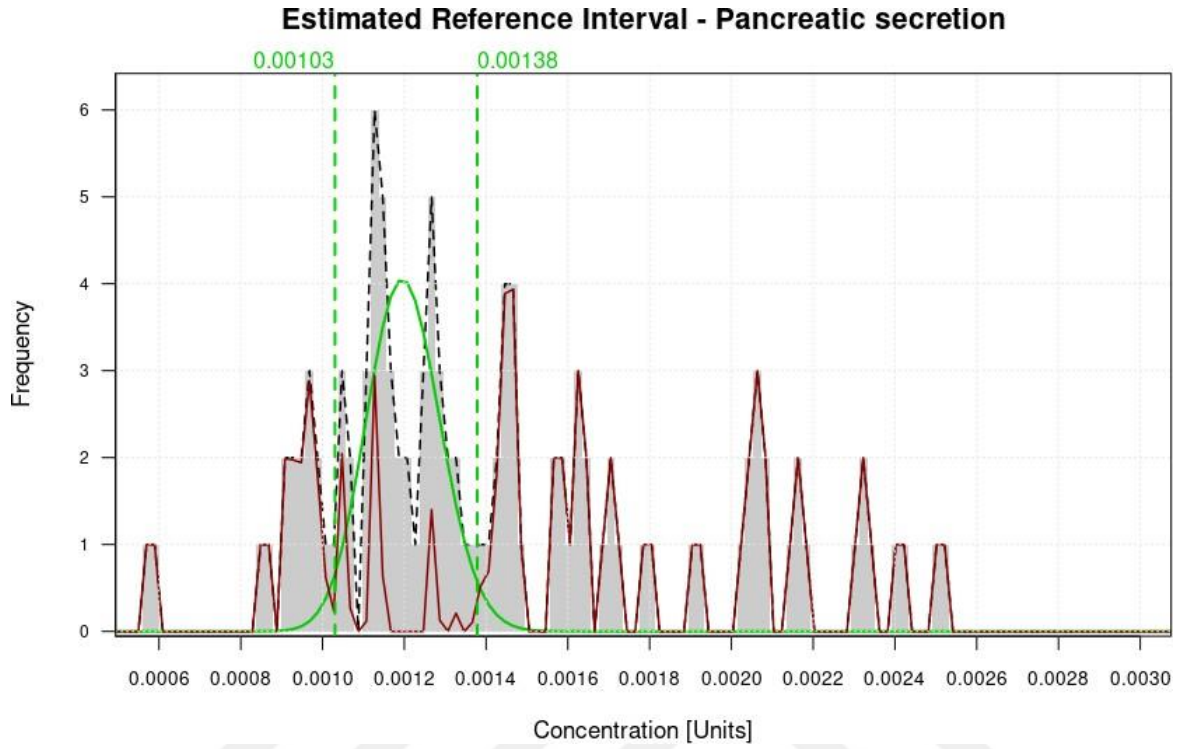


**Figure 29: Graph of reference interval estimation result for pancreatic secretion metabolic pathway.**

### 4.3.24. Propionate metabolism

The reference interval result table and graph for propionate metabolism metabolic pathway are shown in Table 24 and Figure 30. The lower limit of the propionate metabolism pathway is 0.0615 and the upper limit is 0.065 for which reference interval estimation was made for 57 samples.

**Table 24: Reference interval estimation results for propionate metabolism metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0615 |
| upper limit [97.5% perc] | 0.065 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.149 |
| mu | -2.26 |
| sigma | 0.00951 |
| shift | 0 |
| cost | -1.69 |
| NP fraction | 0.4 |

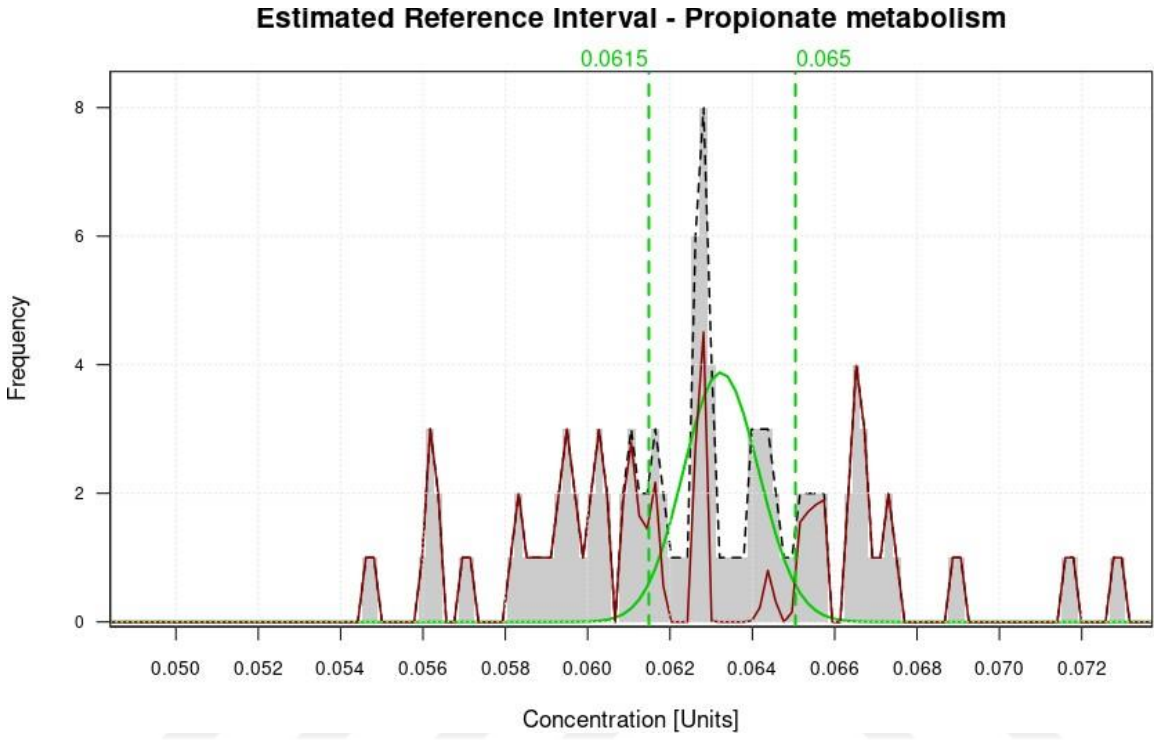


**Figure 30: Graph of reference interval estimation result for propionate metabolism metabolic pathway.**

### 4.3.25. Protein digestion and absorption

The reference interval result table and graph for protein digestion and absorption metabolic pathway are shown in Table 25 and Figure 31. The lower limit of the protein digestion and absorption pathway is 0.0294 and the upper limit is 0.0302 for which reference interval estimation was made for 57 samples.

**Table 25: Reference interval estimation results for protein digestion and absorption metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0294 |
| upper limit [97.5% perc] | 0.0302 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.57 |
| mu | -1.52 |
| sigma | 0.000972 |
| shift | 0 |
| cost | 0.641 |
| NP fraction | 0.4 |

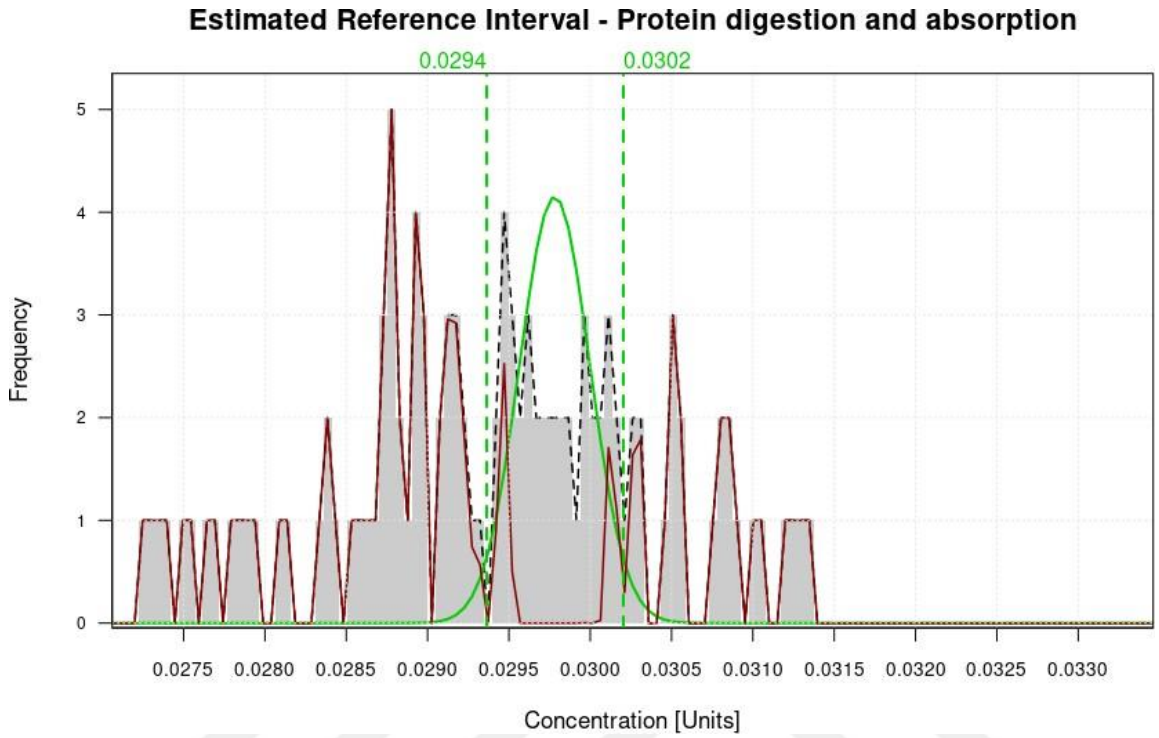


**Figure 31: Graph of reference interval estimation result for protein digestion and absorption metabolic pathway.**

### 4.3.26. Sulfur metabolism

The reference interval result table and graph for sulfur metabolism metabolic pathway are shown in Table 26 and Figure 32. The lower limit of the sulfur metabolism pathway is 0.0328 and the upper limit is 0.034 for which reference interval estimation was made for 57 samples.

**Table 26: Reference interval estimation results for sulfur metabolism metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.0328 |
| upper limit [97.5% perc] | 0.034 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.718 |
| mu | -1.27 |
| sigma | 0.000794 |
| shift | 0 |
| cost | 1.91 |
| NP fraction | 0.4 |

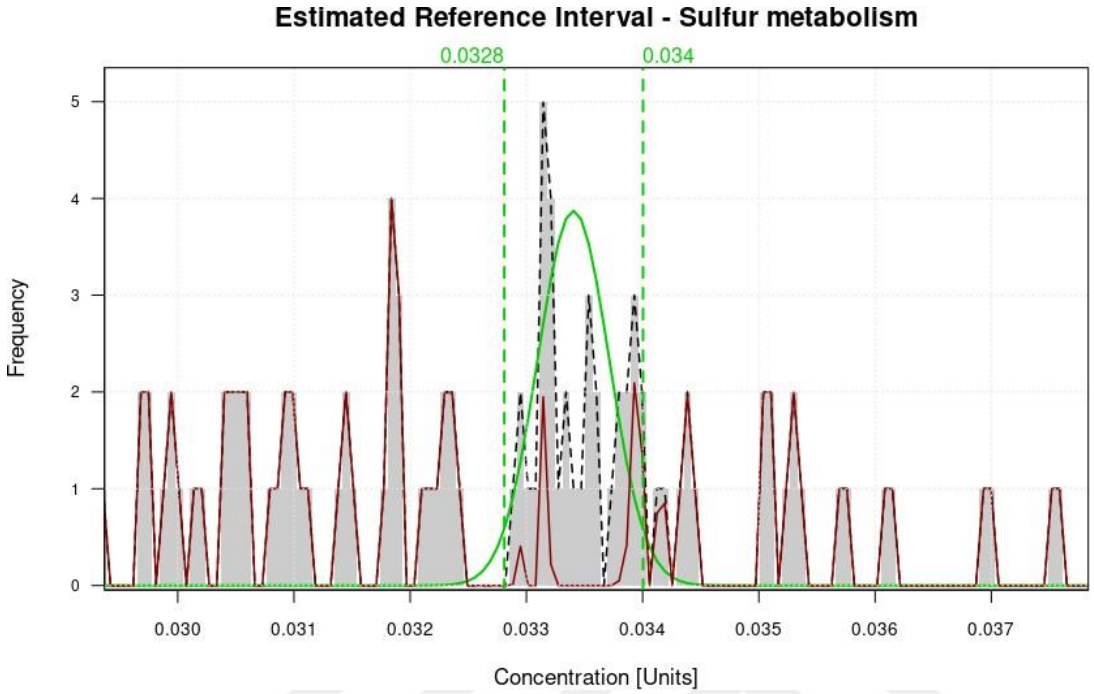


**Figure 32: Graph of reference interval estimation result for sulfur metabolism metabolic pathway.**

### 4.3.27. Type II diabetes mellitus

The reference interval result table and graph for type II diabetes mellitus metabolic pathway are shown in Table 27 and Figure 33. The lower limit of the type II diabetes mellitus pathway is 0.00832 and the upper limit is 0.00864 for which reference interval estimation was made for 57 samples.

**Table 27: Reference interval estimation results for type II diabetes mellitus metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.00832 |
| upper limit [97.5% perc] | 0.00864 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.284 |
| mu | -2.61 |
| sigma | 0.0025 |
| shift | 0 |
| cost | 4.17 |
| NP fraction | 0.401 |

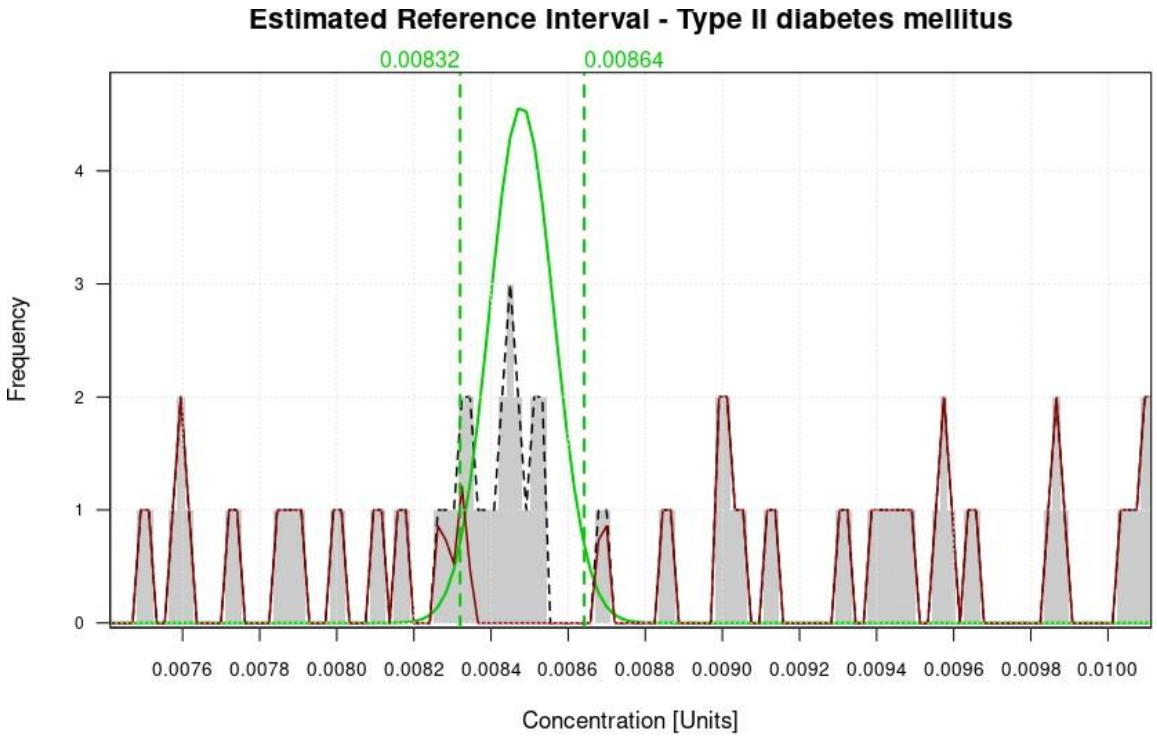


**Figure 33: Graph of reference interval estimation result for type II diabetes mellitus metabolic pathway.**

### 4.3.28. Apoptosis

The reference interval result table and graph for apoptosis pathway are shown in Table 28 and Figure 34. The lower limit of the apoptosis pathway is 0.00733 and the upper limit is 0.00759 for which reference interval estimation was made for 57 samples.

**Table 28: Reference interval estimation results for apoptosis metabolic pathway.**

| Reference Intervals | |
|---|---|
| lower limit [2.5% perc] | 0.00733 |
| upper limit [97.5% perc] | 0.00759 |
| Model Parameters | |
| method | refineR (v1.5.1) |
| model | BoxCox |
| N data | 57 (data points) |
| rounded | no |
| lambda | 0.854 |
| mu | -1.15 |
| sigma | 0.000136 |
| shift | 0 |
| cost | -0.941 |
| NP fraction | 0.4 |

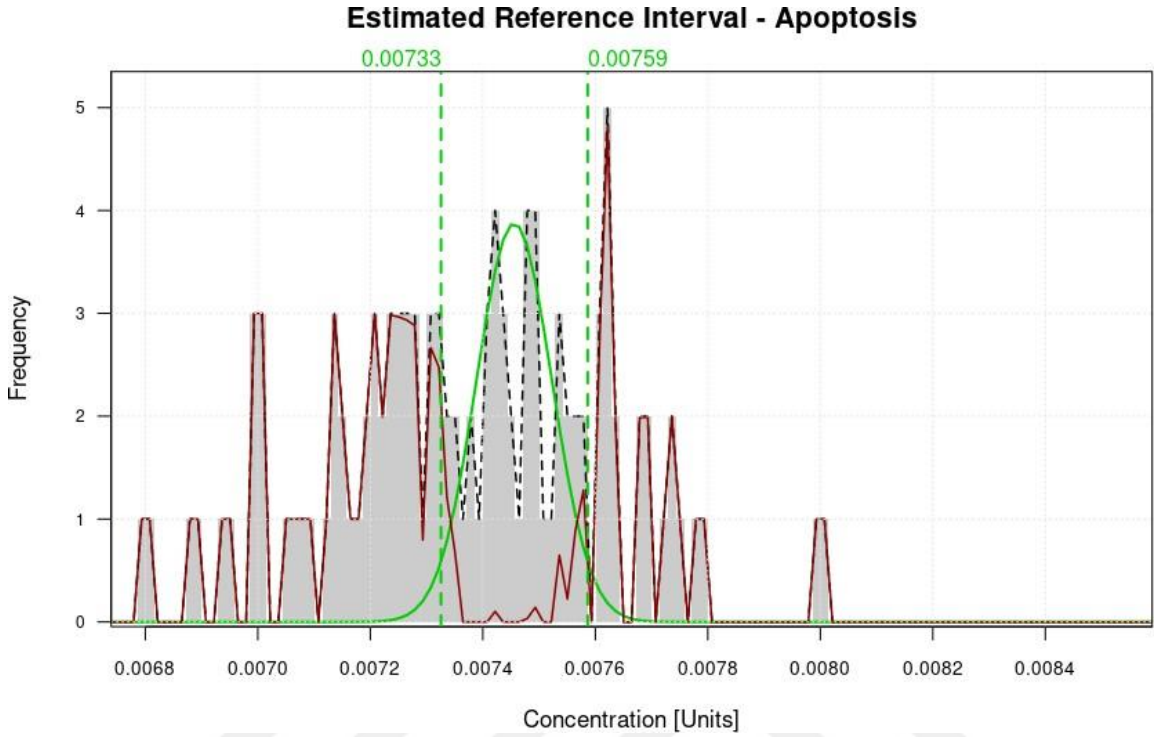


**Figure 34: Graph of reference interval estimation result for apoptosis metabolic pathway.**

## 4.4 Application Programming Interface

### 4.4.1 Data

16S sequencing data can be found in the metabolic pathways identified based on the marker gene sequences thanks to the Picrust2 tool in the ampliseq v2.3.1 workflow developed by nf-core using Nextflow v21.10.6 (The data table for Picrust 2 of the defined metabolic pathways is from the KEGG database, manually prepared and each individual After obtaining the functional abundances of the orthologous gene families related to the metabolic pathway (obtained with customized data), the population-specific reference intervals of each metabolic pathway were estimated using the refineR package in R environment.

### 4.4.2 Database

The data, including the functional abundances of the intestinal microbiota in the metabolic pathways, were structured in accordance with SQL (Structured Query Language) queries and transferred to the SQLITE3 database. A user table was also imported into the database to control

users' access to the database. The db browser interface image of the database is shown in Figure 35, and the entity relationship diagram of the database is shown in Figure 36.

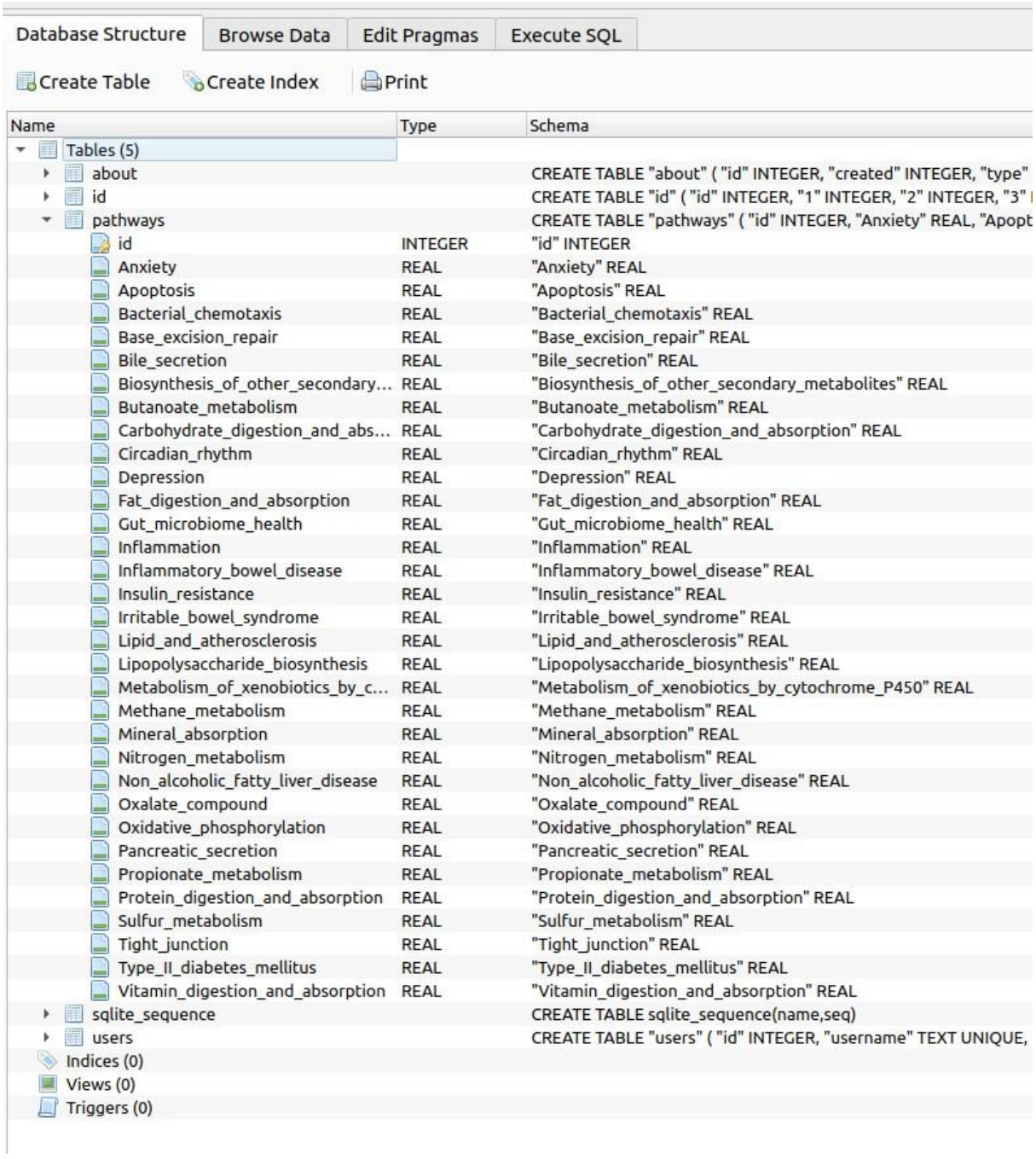


**Figure 35. SQLITE3 database browser interface image.**

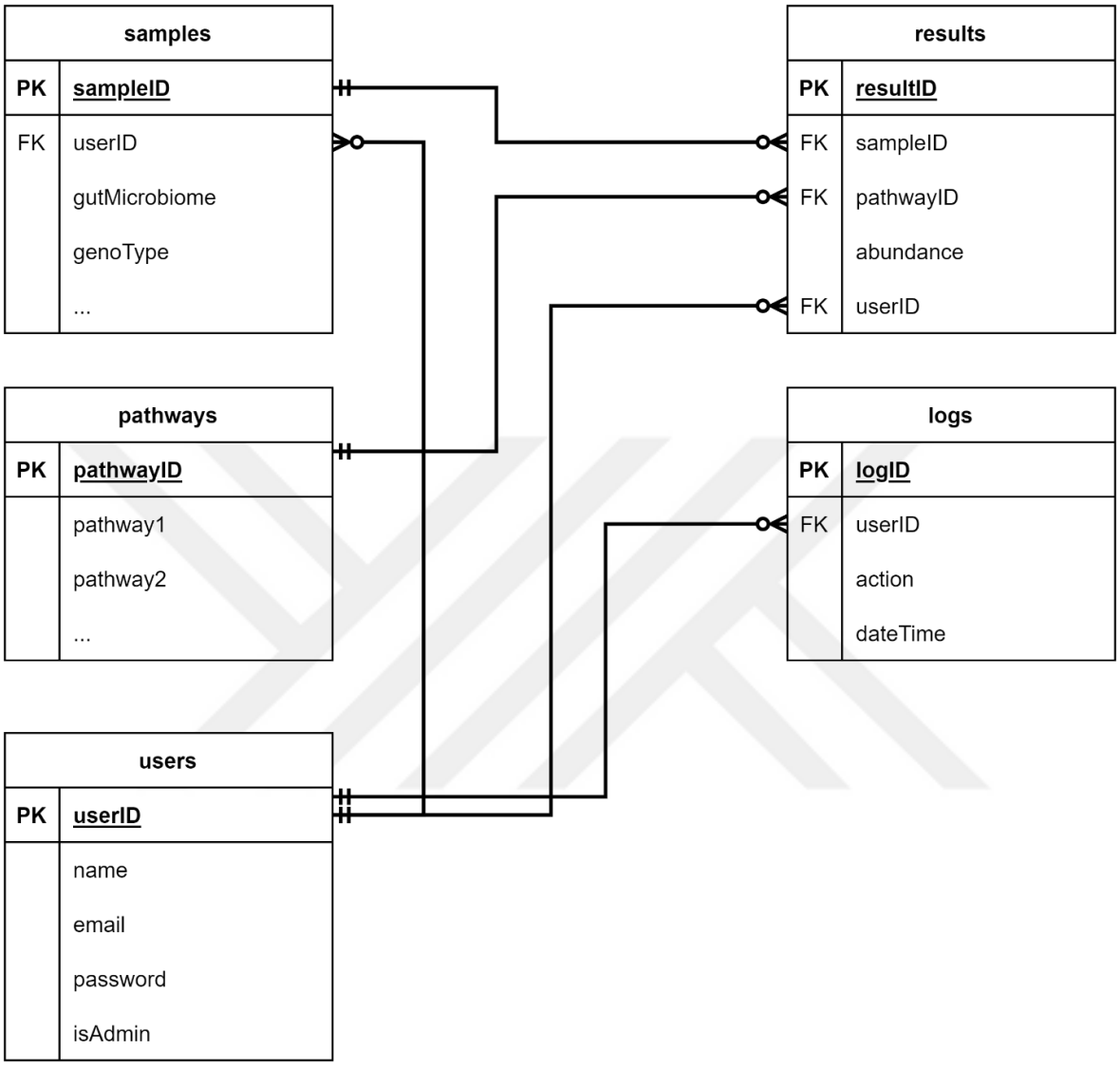


**Figure 36: Entity relationship diagram of the database.**

### 4.4.3 Server

The application programming interface (API) was developed using the open-source Python programming language and the Flask framework as the software development platform. Any operating system can be selected as the server, and it is recommended to choose a Linux-based operating system.

### 4.4.4 Installation

It is aimed to be ready for immediate use without the need for compilation after the installation of the libraries required for the API to work and the configuration of the web server. If desired, a Docker container can be preferred to run without installation. In this case, it will be sufficient to simply select the output port for Docker.

### 4.4.5 Restful API for Reference Interval Estimation

Taxonomic profiles were mapped to annotated metabolic pathways in the KEGG database associated with the gut microbiota, and reference intervals of each metabolic pathway specific to the healthy population included in the thesis were estimated by using the refineR package in the R environment. An API was developed that allows dynamic estimation of gut microbiota functional abundance reference intervals. The screenshot of the homepage of the developed APIis shown in Figure 37, and the workflow of the API is shown in Figure 38.

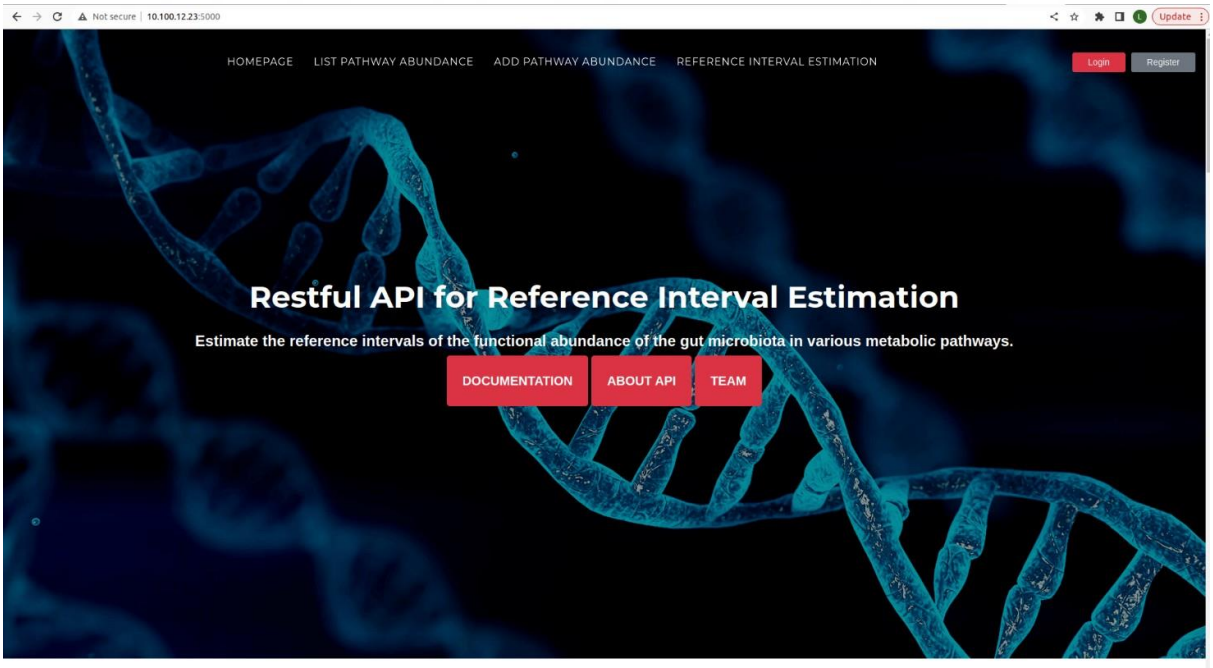


**Figure 37. Restful API home page.**

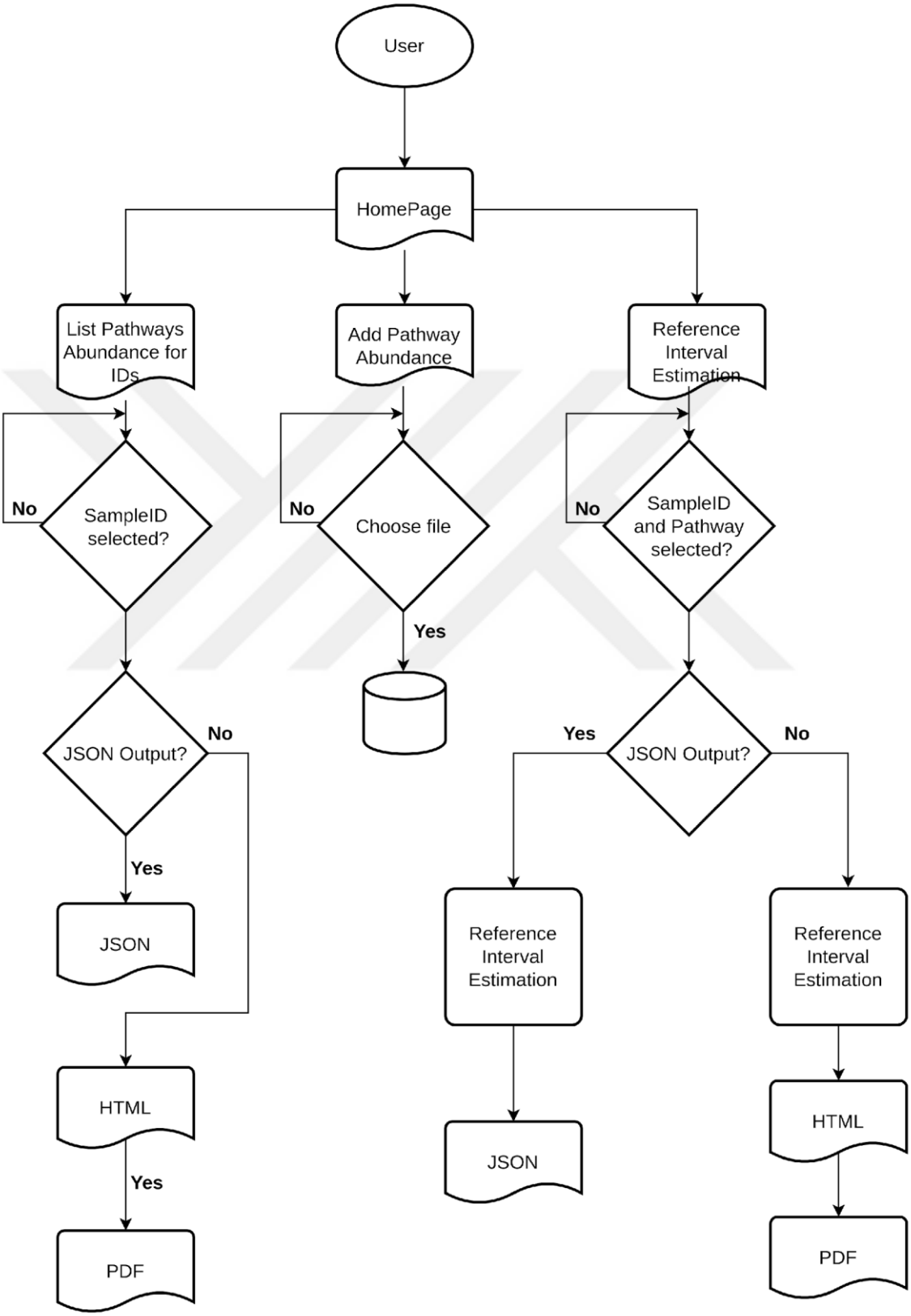


**Figure 38. Workflow of Restful API for reference interval estimation.**

The guide for each API endpoints is provided by the API, and the necessary documentation has been created to communicate with other applications such as postman (Figure 39).

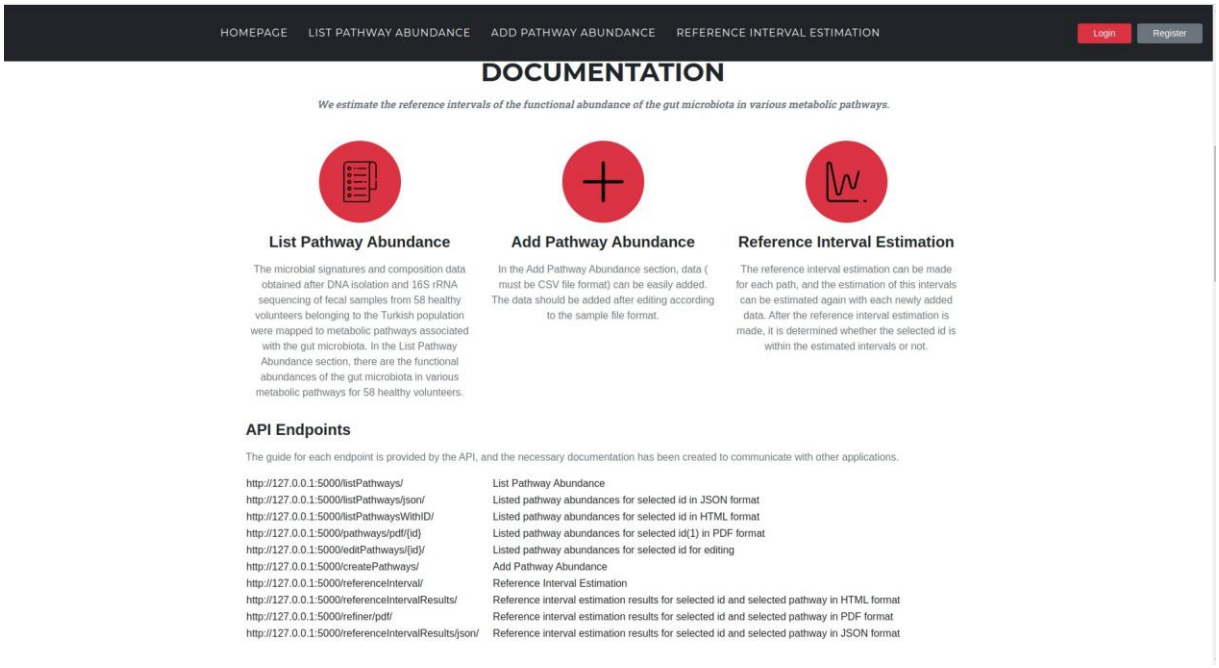


**Figure 39. Restful API documentation.**

To use the API's functions, it is necessary to register first. After clicking the register section onthe upper right in Figure 40, the registration page will open, and the registration will be completed after entering the necessary information. The register page is shown in Figure 41. When the user logs in from the login section, the pages with the functions will be activated. The login page is shown in Figure 40.

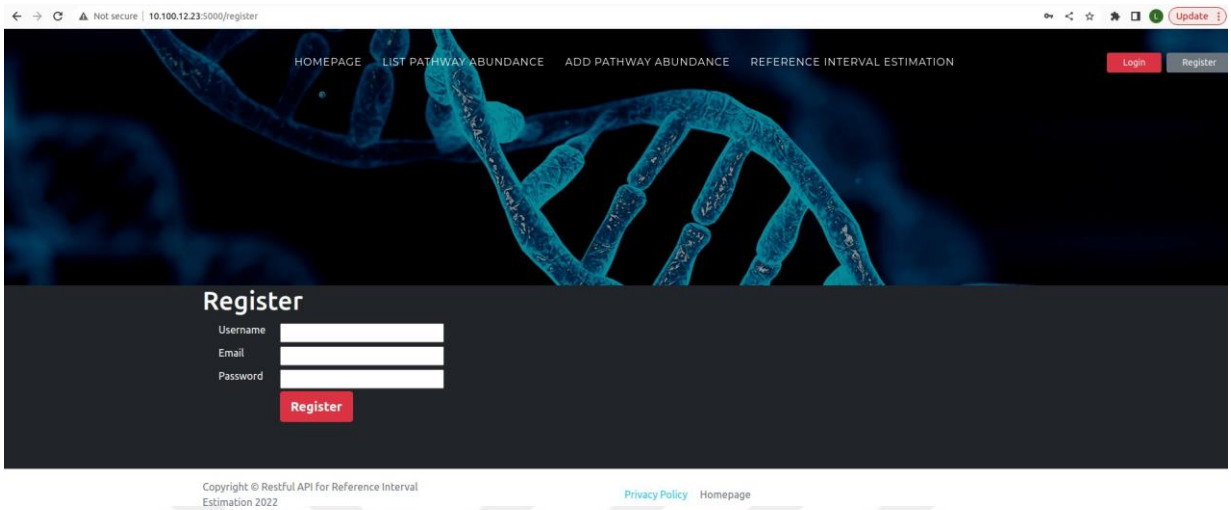


**Figure 40. Restful API's Register page.**

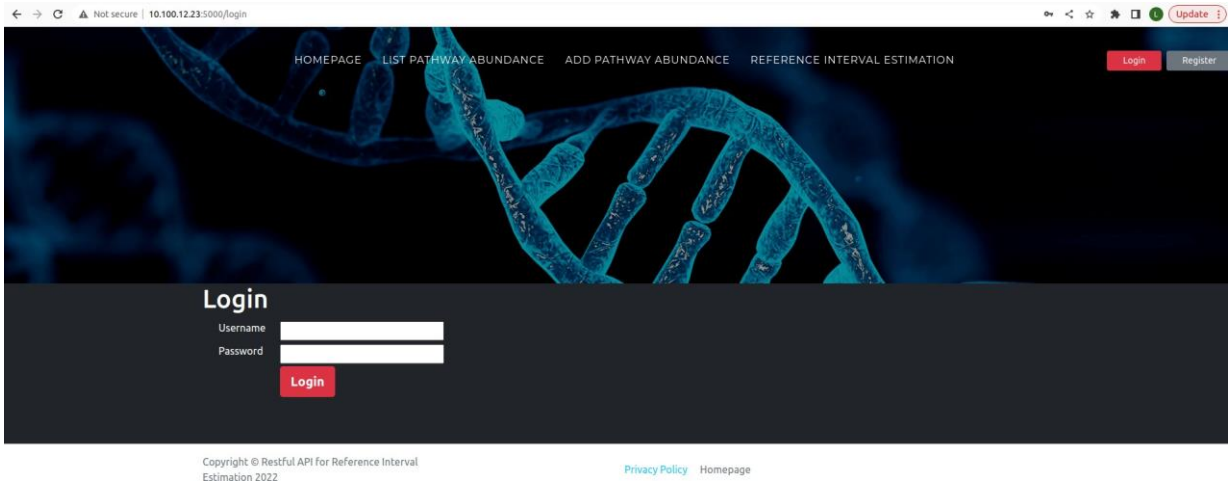


**Figure 41. Restful API's Login page**.

Through the API, the user is provided with the opportunity to get output in the web interface (HTML), as JSON output for other programs, or as PDF output for the end user. In order to facilitate the usage and to instantly access the desired data, the "List Pathway Abundance" page provides the opportunity to generate output by querying the functional abundance analysis resultsaccording to sample IDs (Figure 42). The outputs can be taken in JSON format (Figure 43) with the "JSON_output" button, in HTML format (Figure 44), or as PDF (Figure 45) with the "PDF" button; data can be edited or deleted (Figure 46).

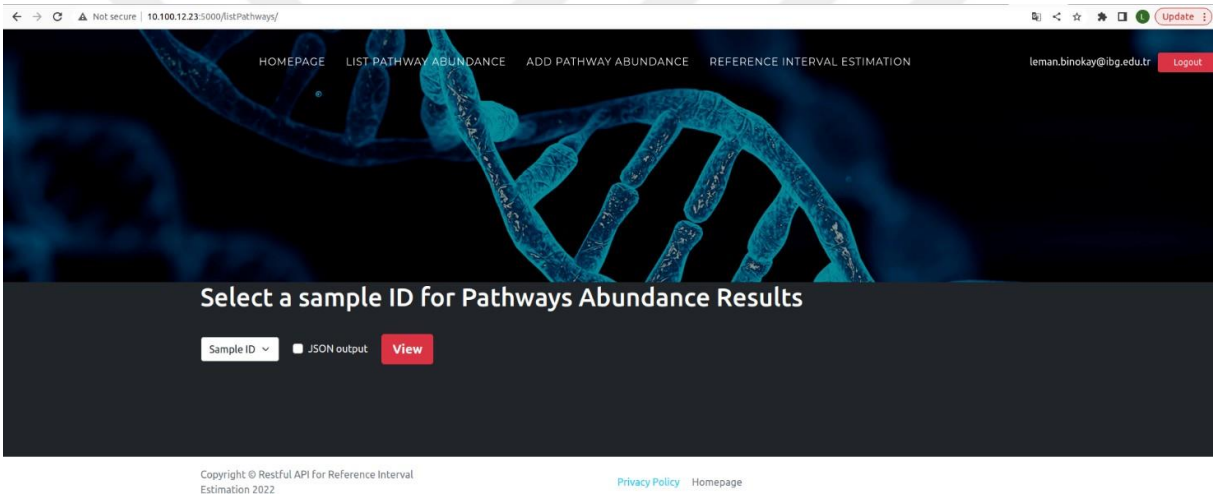


**Figure 42. List Pathway Abundance page.**

```
[
  {
    "Anxiety": 315.75,
    "Apoptosis": 883.8,
    "Bacterial_chemotaxis": 16628.84,
    "Base_excision_repair": 8473.95,
    "Bile_secretion": 445.27,
    "Biosynthesis_of_other_secondary_metabolites": 1310.28,
    "Butanoate_metabolism": 7133.98,
    "Carbohydrate_digestion_and_absorption": 4036.87,
    "Circadian_rhythm": 0.07,
    "Depression": 190.75,
    "Fat_digestion_and_absorption": 1833.82,
    "Gut_microbiome_health": 2147.24,
    "Inflammation": 5301.41,
    "Inflammatory_bowel_disease": 426.51,
    "Insulin_resistance": 5012.07,
    "Irritable_bowel_syndrome": 164.43,
    "Lipid_and_atherosclerosis": 473.9,
    "Lipopolysaccharide_biosynthesis": 9358.96,
    "Metabolism_of_xenobiotics_by_cytochrome_P450": 401.27,
    "Methane_metabolism": 5928.46,
    "Mineral_absorption": 3859.38,
    "Nitrogen_metabolism": 7901.91,
    "Non_alcoholic_fatty_liver_disease": 2697.26,
    "Oxalate_compound": 9721.99,
    "Oxidative_phosphorylation": 6153.39,
    "Pancreatic_secretion": 112.52,
    "Propionate_metabolism": 8017.97,
    "Protein_digestion_and_absorption": 3511.63,
    "Sulfur_metabolism": 3660.86,
    "Tight_junction": 0.0,
    "Type_II_diabetes_mellitus": 1013.35,
    "Vitamin_digestion_and_absorption": 4908.55,
    "id": 1
  }
]
```


**Figure 43. JSON output of functional abundance results in metabolic pathways for the selected sample ID.**

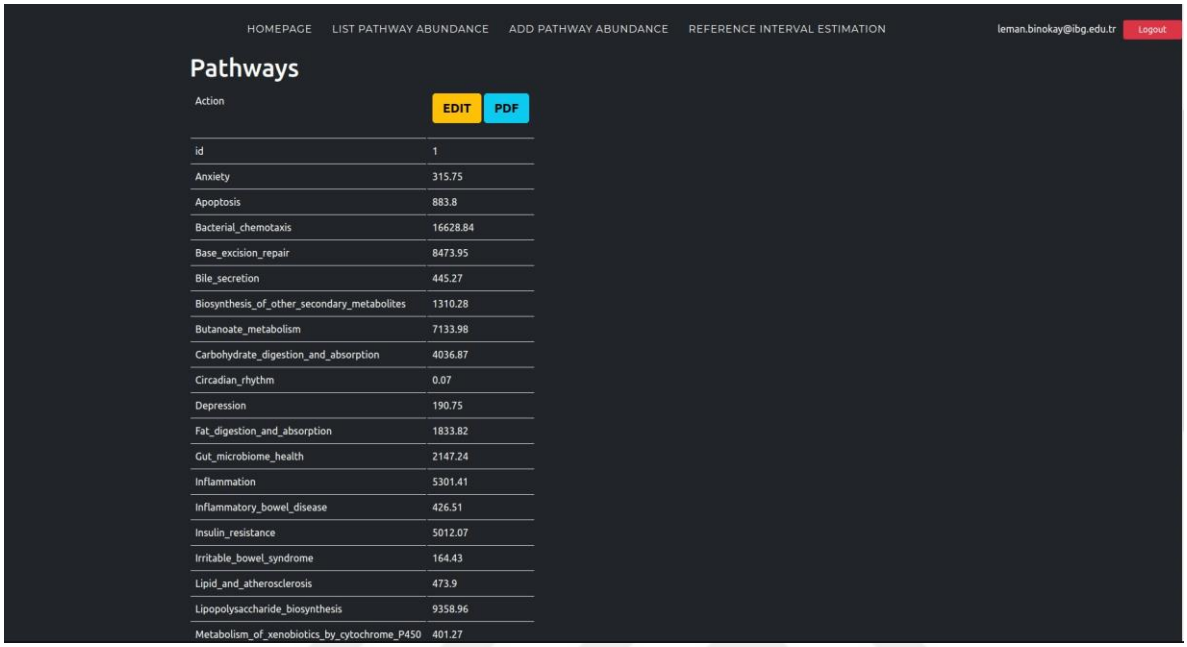


**Figure 44. HTML output of functional abundance results in metabolic pathways for the selected sample ID.**

Pathway Abundance Result for ID=1

| | |
|---|---|
| id | 1 |
| Anxiety | 315.75 |
| Apoptosis | 883.8 |
| Bacterial_chemotaxis | 16628.84 |
| Base_excision_repair | 8473.95 |
| Bile_secretion | 445.27 |
| Biosynthesis_of_other_secondary_metabolites | 1310.28 |
| Butanoate_metabolism | 7133.98 |
| Carbohydrate_digestion_and_absorption | 4036.87 |
| Circadian_rhythm | 0.07 |
| Depression | 190.75 |
| Fat_digestion_and_absorption | 1833.82 |
| Gut_microbiome_health | 2147.24 |
| Inflammation | 5301.41 |
| Inflammatory_bowel_disease | 426.51 |
| Insulin_resistance | 5012.07 |
| Irritable_bowel_syndrome | 164.43 |
| Lipid_and_atherosclerosis | 473.9 |
| Lipopolysaccharide_biosynthesis | 9358.96 |
| Metabolism_of_xenobiotics_by_cytochrome_P450 | 401.27 |
| Methane_metabolism | 5928.46 |
| Mineral_absorption | 3859.38 |
| Nitrogen_metabolism | 7901.91 |
| Non_alcoholic_fatty_liver_disease | 2697.26 |
| Oxalate_compound | 9721.99 |
| Oxidative_phosphorylation | 6153.39 |
| Pancreatic_secretion | 112.52 |
| Propionate_metabolism | 8017.97 |
| Protein_digestion_and_absorption | 3511.63 |
| Sulfur_metabolism | 3660.86 |
| Tight_junction | 0.0 |
| Type_II_diabetes_mellitus | 1013.35 |
| Vitamin_digestion_and_absorption | 4908.55 |

**Figure 45. PDF output of functional abundance results in metabolic pathways for the selected sample ID.**

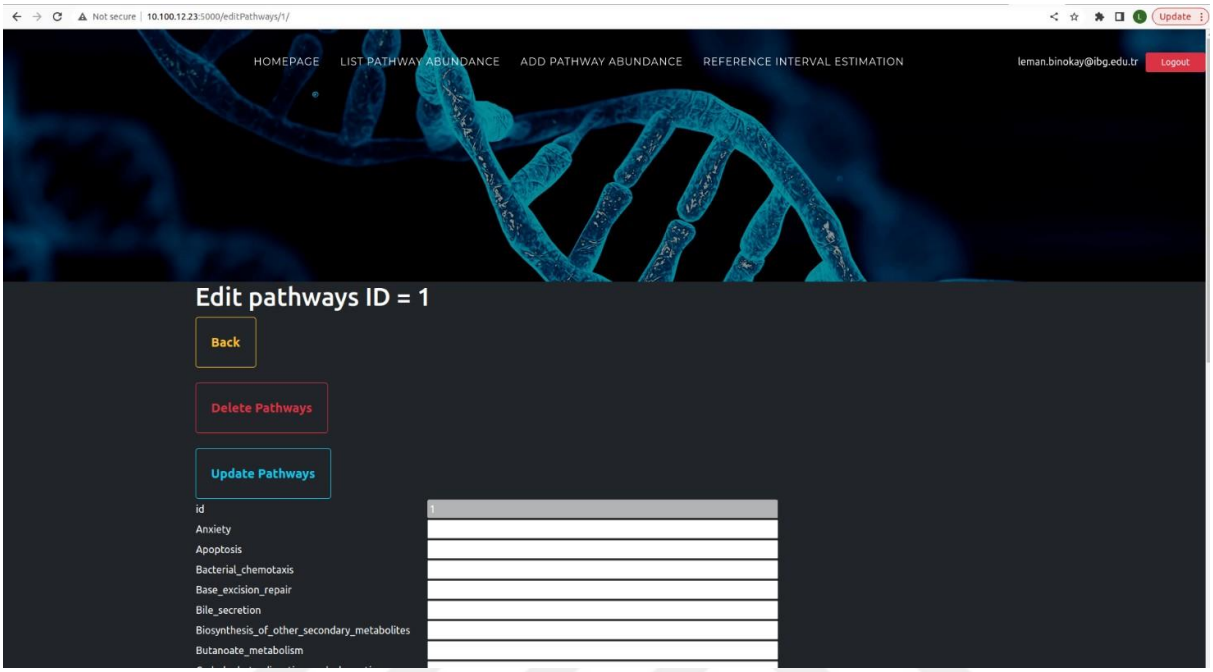


**Figure 46. Edit page of functional abundance results in metabolic pathways for the selected sample ID.**

Data in CSV file format can be easily added to the database with the help of the “Upload csv” button from the “Add Pathway Abundance Results” page. The data should be added after editing in the downloaded file format by using the “Sample File” button on the same page (Figure 47).

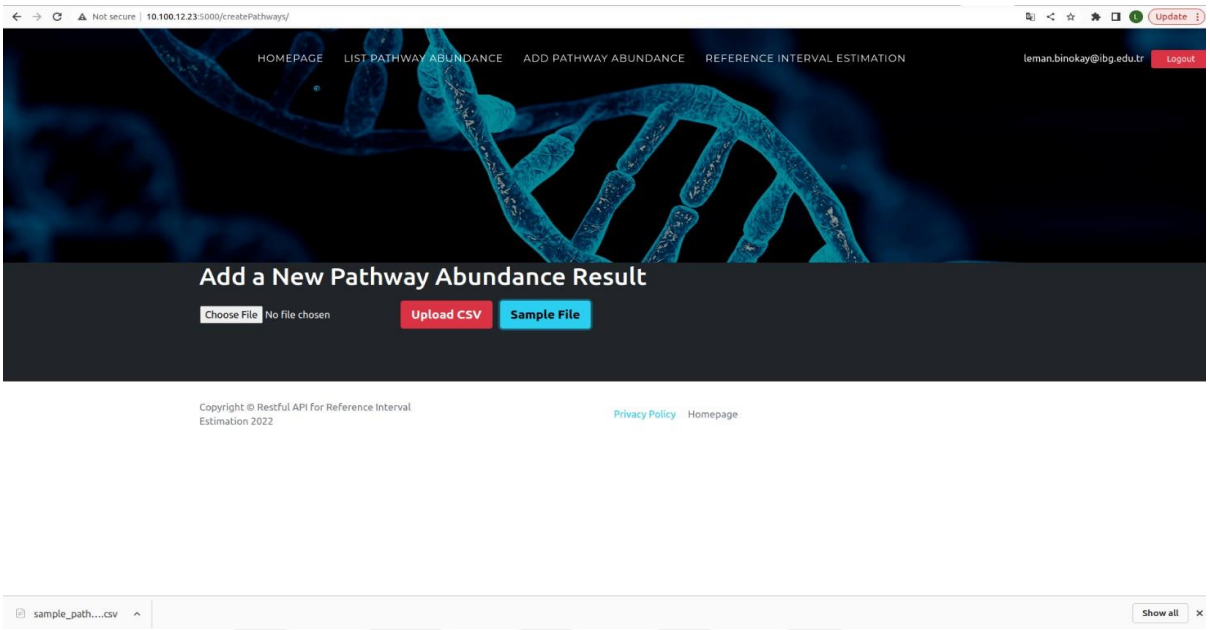


**Figure 47. Add Pathway Abundance Results page.**

In the "Reference Interval Estimation" page, the reference interval estimation can be made according to the selected pathway (Figure 48), and the estimation of this interval can be re-estimated with each newly added data. While estimating the reference interval, it is determined whether the selected sample ID is within the estimated interval or not. Outputs can be received in JSON format (Figure 49) with the "JSON_output" button, in HTML format (Figure 50), or in PDF (Figure 51) with the "PDF" button.

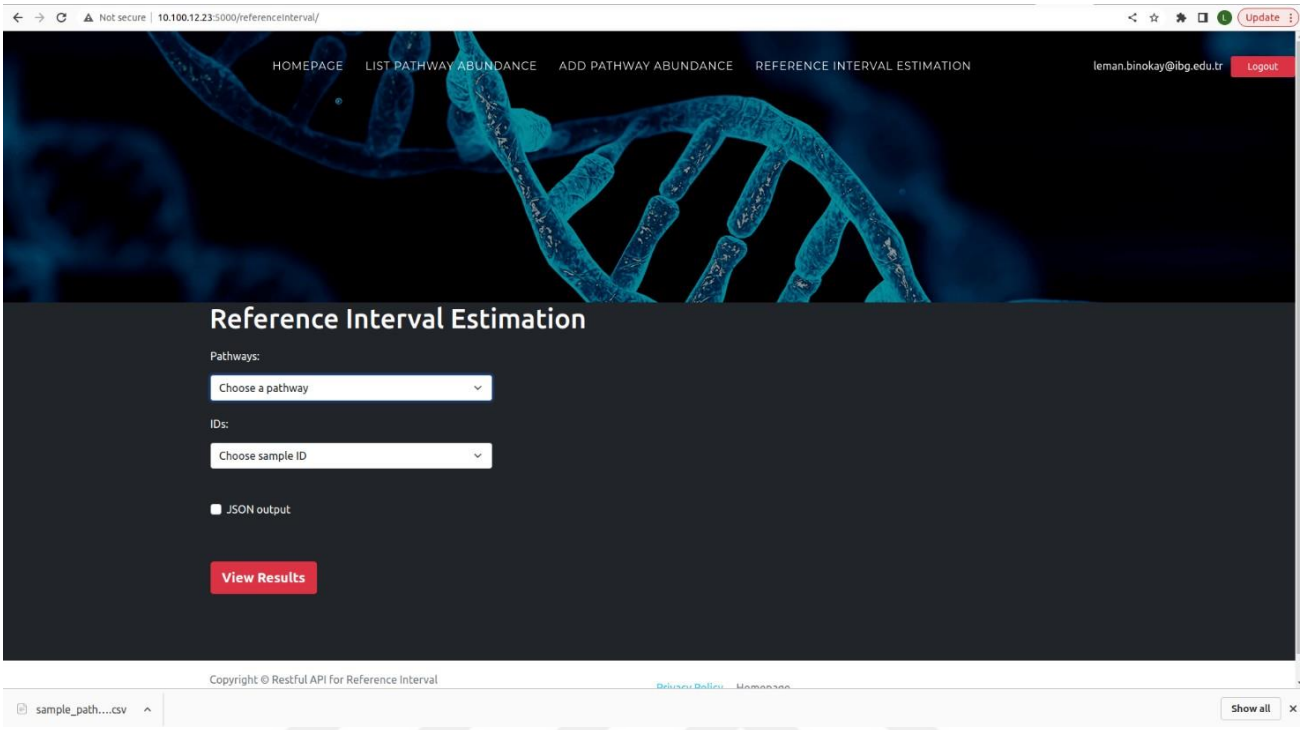


**Figure 48. Reference Interval Estimation page.**

127.0.0.1:5000/referenceIntervalResults/json


```
{
  "1": "Reference Intervals",
  "3": "lower limit [ 2.5% perc]: 0.0104",
  "4": "upper limit [97.5% perc]: 0.0161",
  "6": "Model Parameters",
  "8": "method: refineR (v1.5.1)",
  "9": "model: BoxCox",
  "10": "N data: 57 (data points)",
  "11": "rounded: no",
  "12": "lamda: 0.284",
  "13": "mu: 2.49",
  "14": "sigma: 0.0324",
  "15": "shift: 0",
  "16": "cost: 0.49",
  "17": "NP fraction: 0.4",
  "18": "[1] The sample is not in estimated reference interval"
}
```


**Figure 49. JSON output of selected sample ID and metabolic pathway reference interval**

**analysis results.**

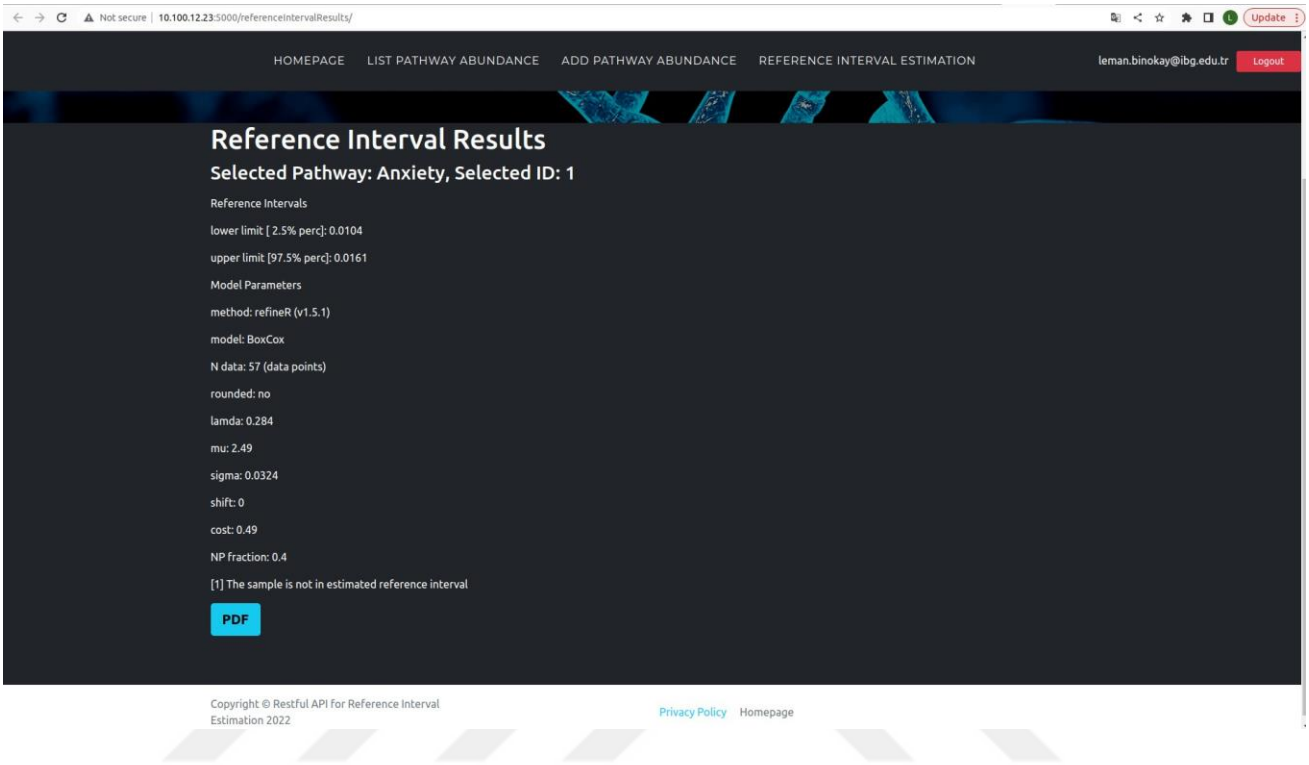


**Figure 50. HTML output of selected sample ID and metabolic pathway reference interval analysis results.**

# Reference Interval Estimation Results

**Selected Pathway: Anxiety**
**Selected ID: 1**

Reference Intervals

lower limit 2.5% perc: 0.0104

upper limit 97.5% perc: 0.0161

Model Parameters

method: refineR (v1.5.1)

model: BoxCox

N data: 57 (data points)

rounded: no

lamda: 0.284

mu: 2.49

sigma: 0.0324

shift: 0

cost: 0.49

NP fraction: 0.4

1 The sample is not in estimated reference interval

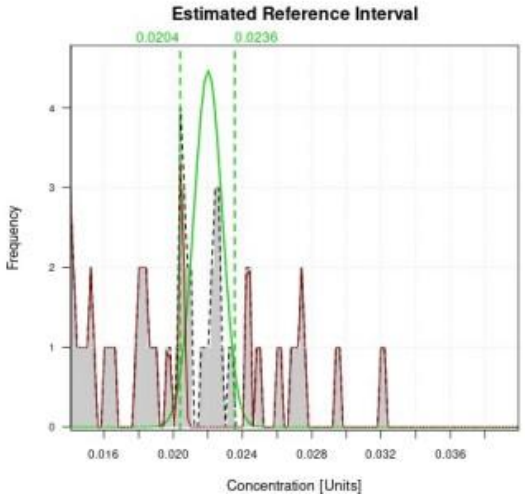


**Figure 51. PDF output of selected sample ID and metabolic pathway reference interval analysis results.**

## 5. DISCUSSION

In light of recent discoveries, the significance of the human gut microbiome has been better appreciated, and research in this field has gained speed. Since the HGP (Human Genome Project), studies such as the HMP (Human Microbiome Project) and the MetaHIT (METAgenomics of the Human Intestinal Tract) consortium have been established to find out significant microorganism-disease correlations (Pace, 2007; Turnbaugh et al., 2007b). A typical individual contains roughly 1-2 kg of bacteria, 90 percent of which are found in the gut, according to the most recent data (Sender et al., 2016).

It is recognized that the human gut microbiota plays crucial roles in human health and disease susceptibility. Recent research on the relationship between illnesses and the gut microbiota has found that the microbiota undergoes major changes in disease states and that certain diseases have a distinct and unique microbiome. Included in this group are metabolic and inflammatory illnesses, cancer, neurologic and neuropsychiatric disorders (Kho & Lal, 2018; Ogunrinola et al., 2020; Sędzikowska & Szablewski, 2021; Sepich-Poore et al., 2021; Visconti et al., 2019). Due to the fact that the gut microbiota encodes substantially more transcripts than the human host, they are able to perform a number of metabolic processes that humans cannot or can only do to a limited degree. Consequently, any alteration or degradation in the gut microbiota may result in several metabolic disorders (Vyas & Ranganathan, 2012). By modulating severalbiological pathways of the host, the bacterial microbiome drives carcinogenesis through different methods. Various proteins, enzymes, and toxins of pathogenic bacteria attach to host cells, altering its normal function and promoting the development of a variety of tumors and malignancies (Sepich-Poore et al., 2021). According to growing clinical and experimentalfindings, the microbiome may be a significant risk factor for neurological illnesses such as Alzheimer's disease, Parkinson's disease, stroke, autism spectrum disorder, and multiple sclerosis. These conditions can be caused by a disruption in the balance of the microbiome. Clinical cross-sectional investigations support the notion that changing microbial composition contributes to the pathogenesis of these disorders (Cryan et al., 2019; Szablewski, 2022). In the light of all this information, microbiome studies should be continued to improve human health, to understand disease mechanisms and to develop various therapeutic treatments.

The gut microbiome is dynamic and in flux from the mother's womb to the moment we die. Our genetics, lifestyle, the environment, and eating habits determine our gut microbiota. For these reasons, there are fundamental differences in the gut microbiota of human populations. A major strength of our study, a data set of the healthy Turkish population was created for the first time, to our knowledge. In order to create this data set, it was aimed to obtain a healthy data set by selecting the volunteers according to well-defined inclusion-exclusion criteria. Thus, the obtained data has the feature of being a baseline data set and can be used for comparison in future studies.

In order to investigate the variety of gut microbiota in healthy persons, this research used feces samples from 57 healthy volunteers and taxonomic profiling by 16S rRNA sequencing. Firmicutes is the most prevalent phylum and Bacteroides is the second most prevalent genus in samples obtained from healthy persons in Spain, China, Sweden, the United States, Denmark, and France, according to the scientific literature. Bifidobacterium has been shown to be present in virtually all healthy human feces samples. In addition, our taxonomic categorization results are similar to those published in the scientific literature (King et al., 2019).

Another notable strength of our study is development of an application programming interface (Restful API) for reference interval estimation. One of the most important advantages of Restful API, which has been developed to share these data with authorized users with the help of an interface and to dynamically estimate reference intervals, is that it provides a great deal of flexibility for users who want to access data. In addition, it can be integrated into workflows. The data has been made accessible to the user by any platform or method. The API is capable of handling multiple types of request and data is convertible to the different formats provided by the API. Through this API interface we have developed, data that are outputs of workflows are presented to users as HTML output in the web interface and as JSON output for other programs. Thanks to the API being locally accessible, new data can be added or desired data can be deleted. For each new data added, the reference interval estimation is updated dynamically andit can be determined whether the selected data is within the reference interval or not. Thus, with the increase in the number of samples, it will be possible to determine the reference intervals of the population with higher accuracy. With the determination of population-specific reference

intervals, more accurate interpretations about the gut health of individuals will be possible. During the study, version control was used so that the new interface ports that were added/removed while the API was being developed, and the database management system, which was edited, would not affect the user. While the user always accesses the data via themost up-to-date API in the web interface, the data of users accessing from other programs will not be affected by version control. In API structure, database management is under the control of API developers only. A different API port has been developed for each function, ensuringthat the product cycle remains more sustainable. The API developer is able to add or remove new versions or new ports in line with the future requirements of the study.

This study's results and conclusions should be considered in light of the following restrictions. This study's sample size (n=57) was suboptimal for estimating reference intervals. However, owing to the tool that we have built, future users will be able to enhance the data we have already included by adding their own. Whole genome shotgun sequencing (WGS) and amplicon sequencing are the most often utilized technologies for sequencing microbiota. In this investigation, 16S rRNA sequencing was used. The sequencing approach for 16S rRNA is based on amplicon sequencing. As stated in the Introduction section, the 16S rRNA gene has nine hyper-variable regions (V1-V9) that are highly conserved with bacterial species-specific changes, allowing species-level identification of microbial communities if full-length 16S sequences are identified. Due to logistic constraints, V3-V4 areas were used in this investigation, resulting in suboptimal findings at the species level (O'Callaghan et al., 2021).

How the gut microbiota may affect human health and illness risk/progression is an issue that remains un-answered. Before we can greatly expand our understanding of the microbiome and use this knowledge to modify the microbiome to benefit human health, we must first overcome some limitations. As a result of the rapid developments in sequencing technologies, which has yielded novel insights into the processes underlying host-microbiome interactions, it is probable that such constraints will be largely eliminated in the near future.

## 6. CONCLUSION AND SUGGESTIONS

Human gut bacteria, which are directly associated to human health, are influenced by a variety of variables, including as diet, lifestyle, and the environment, and their features and functions are not entirely understood. The microbiome is a highly dynamic and complex structure, as well as its sensitivity to external stimuli necessitates a combination of several elements to define healthy microbiota. It is important to carry out detailed functional and taxonomic studies, in particular, to determine whether dysbiosis develops as a cause or consequence of microbiota-disease relationships, and to explain the type of microbiota structure necessary for a healthy life. An avalanche of new knowledge on the nature of the human microbiota and its interactions with the host has gathered as a result of the widespread usage of NGS technology atsubstantially reduced costs. However, only after large-scale and comprehensive studies are completed, it will enable us to clarify the concept of a healthy microbiome and to develop microbiota-focused diagnostic and therapeutic approaches. Through this study,population-specific healthy human gut microbiota taxonomic profile and functional abundances were determined. Nonetheless, we believe a bigger impact of this study will be through the developed application programming interface (Restful API) that will allow dynamic estimation of reference intervals for gut microbiota functional abundance for a wide audience.

## 7. REFERENCES


Ammer, T., Schützenmeister, A., Prokosch, H. U., Rauh, M., Rank, C. M., & Zierk, J. (2021). refineR: A Novel Algorithm for Reference Interval Estimation from Real-World Data. *Scientific Reports*, *11*(1), 1–17. https://doi.org/10.1038/s41598-021-95301-2

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood: Education and Practice Edition*, *98*(6), 236–238. https://doi.org/10.1136/archdischild-2013-304340

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C. C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G. H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J. A., Maguin, E., Mauchline, T., McClure, R., … Schloter, M. (2020). Correction to: Microbiome definition re-visited: old concepts and new challenges. *Microbiome*, *8*(1), 1–22. https://doi.org/10.1186/s40168-020-00905-x

Cryan, J. F., O'riordan, K. J., Cowan, C. S. M., Sandhu, K. V., Bastiaanssen, T. F. S., Boehme, M., Codagnone, M. G., Cussotto, S., Fulling, C., Golubeva, A. V., Guzzetta, K. E., Jaggar, M., Long-Smith, C. M., Lyte, J. M., Martin, J. A., Molinero-Perez, A., Moloney, G., Morelli,E., Morillas, E., … Dinan, T. G. (2019). The microbiota-gut-brain axis. *PhysiologicalReviews*, *99*(4), 1877–2013. https://doi.org/10.1152/physrev.00018.2018

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., & Langille, M. G. I. (2020). PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*, *38*(6), 685–688. https://doi.org/10.1038/s41587-020-0548-6

Escobar-Zepeda, A., De León, A. V. P., & Sanchez-Flores, A. (2015). The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics*, *6*(DEC), 1–15. https://doi.org/10.3389/fgene.2015.00348

Fernando Somoza Alonso. (2015). *Development of a RESTful API - HATEOAS & Driven API*. https://core.ac.uk/display/38126503

Fülling, C., Dinan, T. G., & Cryan, J. F. (2019). Gut Microbe to Brain Signaling: What Happens in Vagus…. *Neuron*, *101*(6), 998–1002. https://doi.org/10.1016/j.neuron.2019.02.008

Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. Microbiology and Molecular Biology Reviews, 68(4), 669–685. https://doi.org/10.1128/mmbr.68.4.669-685.2004

He, J., Zhang, P., Shen, L., Niu, L., Tan, Y., Chen, L., Zhao, Y., Bai, L., Hao, X., Li, X., Zhang, S., & Zhu, L. (2020). Short-chain fatty acids and their association with signalling pathways in inflammation, glucose and lipid metabolism. International Journal of Molecular Sciences, 21(17), 6356. https://doi.org/10.3390/ijms21176356

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

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2016). KEGG: New Perspectives on Genomes, Pathways, diseases and Drugs. Nucleic Acids Research, 45(D1). https://doi.org/10.1093/nar/gkw1092

Kho, Z. Y., & Lal, S. K. (2018). The human gut microbiome - A potential controller of wellness and disease. In *Frontiers in Microbiology* (Vol. 9, Issue AUG). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2018.01835

King, C. H., Desai, H., Sylvetsky, A. C., LoTempio, J., Ayanyan, S., Carrie, J., Crandall, K. A., Fochtman, B. C., Gasparyan, L., Gulzar, N., Howell, P., Issa, N., Krampis, K., Mishra, L., Morizono, H., Pisegna, J. R., Rao, S., Ren, Y., Simonyan, V., … Mazumder, R. (2019). Baseline human gut microbiota profile in healthy people and standard reporting template. *PLoS ONE*, *14*(9), 1–25. https://doi.org/10.1371/journal.pone.0206484

Lepage, P., Leclerc, M. C., Joossens, M., Mondot, S., Blottière, H. M., Raes, J., Ehrlich, D., & Doré, J. (2013). A metagenomic insight into our gut's microbiome. *Gut*, *62*(1), 146–158. https://doi.org/10.1136/gutjnl-2011-301805

Lloyd-Price, J., Abu-Ali, G., & Huttenhower, C. (2016a). The healthy human microbiome. *Genome Medicine*, *8*(1), 1–11. https://doi.org/10.1186/s13073-016-0307-y

Lloyd-Price, J., Abu-Ali, G., & Huttenhower, C. (2016b). The healthy human microbiome. In *Genome Medicine* (Vol. 8, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13073-016-0307-y

Nakatsu, C. H., Byappanahalli, M. N., & Nevers, M. B. (2019). Bacterial community 16S rRNA gene sequencing characterizes riverine microbial impact on Lake Michigan. *Frontiers in Microbiology*, *10*(MAY), 1–12. https://doi.org/10.3389/fmicb.2019.00996

O'Callaghan, J. L., Willner, D., Buttini, M., Huygens, F., & Pelzer, E. S. (2021). Limitations of 16S rRNA Gene Sequencing to Characterize Lactobacillus Species in the Upper Genital Tract. *Frontiers in Cell and Developmental Biology*, *9*(July). https://doi.org/10.3389/fcell.2021.641921

Ogunrinola, G. A., Oyewale, J. O., Oshamika, O. O., & Olasehinde, G. I. (2020). The Human Microbiome and Its Impacts on Health. *International Journal of Microbiology*, *2020*. https://doi.org/10.1155/2020/8045646

Pace, B. S. (2007). The human genome project. *Renaissance of Sickle Cell Disease Research in the Genome Era*, 27–44. https://doi.org/10.1142/9781860947964_0003

Pascal Andreu, V., Augustijn, H. E., Chen, L., Zhernakova, A., Fischbach, M. A., Dodd, D., & Medema, M. H. (n.d.). *A systematic analysis of metabolic pathways in the human gut microbiota 1*. https://doi.org/10.1101/2021.02.25.432841

Proctor, L. M., Creasy, H. H., Fettweis, J. M., Lloyd-Price, J., Mahurkar, A., Zhou, W., Buck, G. A., Snyder, M. P., Strauss, J. F., Weinstock, G. M., White, O., & Huttenhower, C. (2019). The Integrative Human Microbiome Project. *Nature*, *569*(7758), 641–648. https://doi.org/10.1038/s41586-019-1238-8

Reynoso-García, J., Miranda-Santiago, A. E., Meléndez-Vázquez, N. M., Acosta-Pagán, K., Sánchez-Rosado, M., Díaz-Rivera, J., Rosado-Quiñones, A. M., Acevedo-Márquez, L., Cruz-Roldán, L., Tosado-Rodríguez, E. L., Figueroa-Gispert, M. D. M., & Godoy-Vitorino, F. (2022). A complete guide to human microbiomes: Body niches, transmission,development, dysbiosis, and restoration. *Frontiers in Systems Biology*, *2*. https://doi.org/10.3389/fsysb.2022.951403

Sędzikowska, A., & Szablewski, L. (2021). Human gut microbiota in health and selected cancers. In *International Journal of Molecular Sciences* (Vol. 22, Issue 24). MDPI. https://doi.org/10.3390/ijms222413440

Sender, R., Fuchs, S., & Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. PLOS Biology, 14(8). https://doi.org/10.1371/journal.pbio.1002533

Sepich-Poore, G. D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J. A., & Knight, R. (2021). The microbiome and human cancer. *Science*, *371*(6536). https://doi.org/10.1126/science.abc4552

Szablewski, L. (2022). Human Gut Microbiota in Health and Alzheimer's Disease. *Advances in Alzheimer's Disease*, *9*, 217–228. https://doi.org/10.3233/AIAD220015

Tarkowska, A., Carvalho-Silva, D., Cook, C. E., Turner, E., Finn, R. D., & Yates, A. D. (2018). Eleven quick tips to build a usable REST API for life sciences. *PLoS ComputationalBiology*, *14*(12), 1–8. https://doi.org/10.1371/journal.pcbi.1006542

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007a). The Human Microbiome Project. *Nature*, *449*(7164), 804–810. https://doi.org/10.1038/nature06244

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007b). The Human Microbiome Project. In *Nature* (Vol. 449, Issue 7164, pp. 804–810). Nature Publishing Group. https://doi.org/10.1038/nature06244

Ursell, L. K., Metcalf, J. L., Parfrey, L. W., & Knight, R. (2012). Defining the human microbiome. *Nutrition Reviews*, *70*(SUPPL. 1). https://doi.org/10.1111/j.1753-4887.2012.00493.x

Visconti, A., Le Roy, C. I., Rosa, F., Rossi, N., Martin, T. C., Mohney, R. P., Li, W., de Rinaldis, E., Bell, J. T., Venter, J. C., Nelson, K. E., Spector, T. D., & Falchi, M. (2019). Interplay between the human gut microbiome and host metabolism. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-12476-z

Vyas, U., & Ranganathan, N. (2012). Probiotics, prebiotics, and synbiotics: Gut and beyond. *Gastroenterology Research and Practice*, *2012*. https://doi.org/10.1155/2012/872716

Whitehead I., M. I., Hillard, T. C., & Crook, D. (1990). The role and use of progestogens. *Obstetrics and Gynecology*, *75*(4), 59S-76S. https://doi.org/10.1097/00006250-199004001-00013

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 1–13. https://doi.org/10.1186/s13059-019-1891-0

## APPENDIX I

**İZMİR BİYOTIP VE GENOM MERKEZİ**
**GİRİŞİMSEL OLMAYAN ARAŞTIRMALAR ETİK KURULU (İBG-GOEK)**
**KARARI**

| | | | |
|---|---|---|---|
| **Toplantı Tarihi** | : 27/10/2022 | **Toplantı Günü** | : Perşembe |
| **Toplantı Sayısı** | : 13 | **Toplantı Saati** | : 10:30 |

**Sayın Doç. Dr. Gökhan KARAKÜLAH,**

**2022-037 Protokol No'lu;** sorumlusu olduğunuz "Bağırsak mikrobiyotası fonksiyonel bolluğu referans aralıklarının dinamik kestirimi için bir uygulama programlama arayüzü geliştirilmesi" başlıklı araştırmanın uygulanmasında etik açıdan sakınca olmadığına oy birliği ile karar verilmiştir.

Bilgilerinizi ve gereğini rica ederiz.

| | |
|---|---|
| **Prof. Dr. H. Alper BAĞRIYANIK**<br>**Başkan** | **Dr. Serap ERKEK ÖZHAN**<br>**Başkan Yardımcısı** |
| **Prof. Dr. Sedef AKGÜNGÖR**<br>**Üye** (Katılmadı) | **Prof. Dr. İnci ALACACIOĞLU**<br>**Üye** |
| **Prof. Dr. Gülgün OKTAY**<br>**Üye** | **Prof. Dr. Hilmi ORHAN**<br>**Üye** |
| **Prof. Dr. Şermin GENÇ**<br>**Üye** (Katılmadı) | **Doç. Dr. Canan Aslı YILDIRIM**<br>**Üye** |
| **Doç. Dr. Can KÜÇÜK**<br>**Üye** (Katılmadı) | (Projede araştırmacı)<br>**Dr. Öğr. Üyesi Yavuz OKTAY**<br>**Üye** |