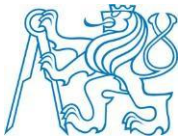

Czech Technical University in Prague

Faculty of Biomedical Engineering

Department of Biomedical and Clinical Technology

MOLECULAR BIOLOGY METHODS IN STUDIES OF GENETIC INFLUENCE AND IDENTIFICATION OF CONTROLING GENES IN EXPERIMENTAL MODEL OF INFECTIOUS DISEASES

Doctoral Thesis

Ing. Lucie Mrázková

Kladno, August 2024

Ph.D Programme: Biomedical and Clinical Technology
Branch of study: Biomedical and Clinical Technology

Supervisor: ***doc. Marie Lipoldová, CSc.***

Supervisor-Specialist: ***RNDr. Taťána Jarošíková, CSc.***

I declare that I carried out this research report independently, and only with the cited sources, literature and other professional sources. I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Czech Technical University in Prague has the right to conclude a license agreement on the use of this work as a school work according to Section 60 paragraph 1 of the Copyright Act.

Ing. Lucie Mrázková

August 2024

This work was performed at the Faculty of Biomedical Engineering of the Czech Technical University in Prague (FBE, CTU), Sítná 3105, 272 01, Kladno 2, with cooperation with the Institute of Molecular Genetics of the Czech Academy of Sciences (IMG, CAS), Laboratory of Molecular and Cellular Immunology (LMCI), Vídeňská 1083, 142 20, Praha 4 and the Institute of Animal Physiology and Genetics of the Czech Academy of Sciences (IAPG, CAS), Laboratory of Anaerobic Microbiology, Vídeňská 1083, 142 20, Praha 4 and at the Institute of Parasitology, Biology Centre of the Czech Academy of Sciences (IP BC CAS), Branišovská 31, 37005, České Budějovice.

The thesis was prepared under the supervision of doc. Marie Lipoldová, CSc. from the LMCI, IMG CAS, where the main experimental work was carried out. Microbiome study was performed by collaboration of CTU, IMG CAS and IAPG CAS, where the experimental part was realized in laboratories of the CAS.

Ethics Approval Statement

All experimental protocols utilized in this study comply with the Czech Government Requirements under the Animal Protection Law Policy (No. 246/1992) and with the regulations of the Ministry of Agriculture of the Czech Republic (No. 207/2004), which are in agreement with all relevant European Union guidelines for work with animals, and were approved by the Institutional Animal Care Committee of the Institute of Molecular Genetics CAS (permissions Nr. 190/2010; 232/2012 – leishmaniasis experiments), Institutional Animal Care Committee of the Institute of Parasitology, Biology Centre of the Czech Academy of Sciences (permission Nr. 165/2010 – encephalitis experiment), and by the Institutional Animal Care Committee of the Institute of Animal Physiology and Genetics of the Czech Academy of Sciences (permission number: 16OZ30963/2015 – microbiome experiment). All these permissions were endorsed by the Departmental Expert Committee for the Approval of Projects of Experiments on Animals of the Academy of Sciences of the Czech Republic.

The author has acquired professional competence in the field of experimental animals according to §15d paragraph 2 letter a) of Act No. 246/1992 Coll. (license number: CZ 02946)

Abstract:

Biomedical engineering plays a crucial role in the development and improvement of methods used in molecular biology. These methods enable the manipulation and analysis of biological material and thus contribute to the advancement of new therapeutic strategies.

This thesis deals with the diagnostic-oriented part of biomedical engineering and includes development of innovative approaches and methods for the prevention and diagnosis of diseases. A key element in detection of genetic basis of disease is establishment of genotype(s) associated with the disease susceptibility. In this study, we present a newly modified, robust, cheap and harmless method for genotyping. We have identified the conditions under which the resolution of the genotyping method can be increased to a difference of 6 nucleotides or even less. We performed proof of concept of this method in studies of two different infectious diseases: leishmaniasis and tick borne encephalitis (TBEV).

We determined and tested experimental populations in a study of the microbiome associated with leishmaniasis. This analysis confirmed genetic background of mouse strains that differ in their susceptibility to *Leishmania major* and enabled us to impact of genotype and infection on the microbiota. Microbial communities in the ileum and colon of the tested strains were compared by Linear discriminant analysis Effect Size (LefSe) and cluster analysis. We shown that *Leishmania* infection affects mainly the ileum microbiota, whereas the colon bacterial community was more stable. Obtained results indicate differences of microbiota composition among resistant and susceptible strains.

We also successfully used this method for mapping a new TBEV survival-controlling locus on chromosome 7. Combination of bioinformatics and a systems analysis based on the definition of gene expression patterns, the classification of individual genes into ontological pathways and the use of specific genetic polymorphisms affecting disease led to detection of nine candidate genes in this locus. Consequently, we proved usefulness of the improved method of genetic typing in studies of susceptibility to diseases. The obtained results could have a great potential for translational medicine and thus for improving human healthcare.

Content

List of abbreviations

A: adenine

Abcc6: (ATP-binding cassette, sub-family C (CFTR/MRP), member 6)

AD: Alzheimer's disease

ANOVA: analysis of variance

ARMS: amplification refractory mutation system

BLAST: basic local alignment search tool

bp : base pair

BWA: Burrows-Wheeler aligner

C: cytosine

Ccr5: chemokine (C-C motif) receptor 5

CcS/Dem: BALB/c-c-STS/Dem

Cd209: CD209 antigen

Cd33: CD33 antigen

cM: centimorgan

CNS: central nervous system

DGGE: denaturing gradient gel electrophoresis

DNA: deoxyribonucleic acid

EC: endemic controls

ELISA: enzyme linked immunosorbent assay

Fut2: fucosyltransferase 2

G: guanine

Grwd1: glutamate-rich WD repeat containing 1

HapMap: haplotype mapping

HGP: human genome project

HUGO: human genome organization

Ifnl3: interferon lambda 3 (synonym *Il28b*)

Il10: interleukin 10

ITS: internal transcribed spacer

KEGG: Kyoto encyclopedia of genes and genomes

Klk1b16: kallikrein 1-related peptidase b16

Klk1b22: kallikrein 1-related peptidase b22

KW: Kruskal-Wallis

LDA: linear discriminant analysis

LefSe: Linear discriminant analysis Effect Size

Mkrn3: makorin, ring finger protein, 3

MISA: MicroSAtellite

NCBI: national center for biotechnology information

NGS: next-generation sequencing

Oas1b: 2'-5' oligoadenylate synthetase 1B

OcB/Dem: O20-c-B10.O20/Dem

Otog: otogelin

OTUs: operational taxonomic units

PCoA: principal coordinate analysis

PCR: polymerase chain reaction

pfu: plaque-forming unit

PGM: personal genome machine

PICRUSt: phylogenetic investigation of communities by reconstruction of unobserved states

QIIME: quantitative insights into microbial ecology

QTLs: quantitative trait loci

RFLP: restriction fragment length polymorphism

s.c: Subcutaneous

SBS: sequencing by synthesis

Siglece: (sialic acid binding Ig-like lectin E)

SNP: single nucleotide polymorphism

STRs: short tandem repeats

T: thymine

TBE: Tick-borne encephalitis

TBEV: tick-borne encephalitis virus

Tlr3: toll-like receptor 3

TRF: tandem repeats finder

UV: ultra violet

VL: visceral leishmaniasis

1. Theory

In view of the progress made in the development and application of new technologies, the demands on the methods used in basic and applied research are increasing. Newly introduced methods must meet higher requirements in terms of work safety, user-friendliness, robustness of the data obtained as well as uniformity and reproducibility of the results. To facilitate the dissemination of this method, we have decided to use only standard laboratory equipment in order to keep costs low for laboratories using this method. Nowadays, more attention is also paid to the environmental impact of the chosen approaches. The aim of the present work is therefore to develop a simple, inexpensive, accurate and non-toxic method for genotyping and to prove the accuracy of the method using a practical example.

Genotyping, the process of determining an individual's genotype at specific locations in the genome, involves identifying the combination of alleles of a particular DNA variant. This unique genetic make-up, which is inherited from both parents, plays a decisive role in the expression of individual characteristics and susceptibility to diseases. Eukaryotic genomes contain a large number of short tandem repeats (STRs). These variable deoxyribonucleic acid (DNA) sequences exhibit polymorphism, which is mainly characterized by length variability. Their high frequency in combination with their polymorphism and the possibility of amplification make them a useful tool for genetic studies.

In the context of disease models, tick-borne encephalitis (TBE) stands out as a major viral infection in Eurasia, with the virus being transmitted mainly by ticks. The disease has a broad spectrum of symptoms, ranging from asymptomatic cases to severe encephalitis, which can be fatal. The underlying genetic factors that influence this spectrum of symptoms remain largely unknown, but it is likely that the genetic makeup of the host is a contributing factor. Leishmaniasis, a parasitic disease that affects over 12 million people in 98 countries, poses a significant health threat. The disease is transmitted by biting insects and manifests itself in three main forms: dermal, mucocutaneous, and visceral. The specific type of the pathology is influenced by both the parasite type and the genotype of the host, emphasizing the importance of genetic factors in the progression and severity of the disease.

Recent advances in molecular methods and quantitative techniques have provided powerful tools to study the polygenic control of complex and quantitative traits. These tools have improved our understanding of how genes influence phenotypic variation, particularly through systems analysis. This approach, which involves the definition of gene expression patterns, the classification of genes into ontological pathways and the identification of genetic polymorphisms associated with disease, allows a more detailed understanding of genetic regulation in molecular networks involved in specific processes or diseases. The development of new methods in biomedicine, including advanced genotyping techniques, microbiome research and gene mapping, promises a better understanding of complex diseases and the development of targeted treatments. These innovations have the potential to revolutionize healthcare and lead to more personalized and effective interventions.

To test the concept of the novel method on the models of different infectious diseases, we performed genotyping in two independent studies, the complex comparative microbiome study on the effects of leshmaniasis and the mapping and identification of candidate genes for tick-borne encephalitis (TBE). We decided to link the phenotypic manifestation of model diseases to specific sites on the chromosomes.

2. Current state of knowledge

1. 1 Short tandem repeats and the genotyping

STRs also known as microsatellites are DNA sequences consisting of repeating units (adenine (A), thymine (T), guanine (G), cytosine (C)) ranging from one base pair (bp) to thousands of bp. The main types of microsatellites consist of 1-6 bp motifs [1, 2]. Among these, dinucleotide repeats are the most prevalent, with (CA)n repeats being the most common, followed by (AT)n, (GA)n and (GC)n repeats, although the latter are relatively rare. Most of these simple repeats are located in non-coding DNA, either in intergenic sequences or introns [3]. However, many human [4], mouse [5] as well as plant [6] genes have STRs in their open reading frames, including promoters [7].

Eukaryotic genomes are characterized by a high number of STRs [8, 9, 10, 11, 12], while prokaryotes also contain these repetitive sequences, albeit in smaller numbers [13]. Microsatellites, which are among the most variable DNA sequence types in the genome, owe their polymorphism primarily to length variability. The abundance of microsatellites, combined with their polymorphic and hypervariable nature and the ability to amplify them by polymerase chain reaction (PCR), enabled the construction of genetic maps and the molecular tagging of genes [14, 15, 16]. First comprehensive dense genetic maps of human [17] and mouse [18] genomes based on STRs were published in 1996. Introduction of the more advanced techniques led to increase of STR density in both human [19], and mouse [20] genomes. This had wide-ranging applications, including the study of genetic susceptibility to diseases. Identification of the causative genetic mutation, especially the dysfunctional protein and its associated biological mechanism, is critical for effective disease management and evaluation of treatment options. While genotyping techniques remain consistent across species, the specific sequences of primers and probes are often tailored to individual species. The correlation between genotype and phenotype is evaluated by genotyping the animals. Once associations are identified, they can be back-translated and tested in the human genome [21].

Microsatellites can be identified from sequence data using various computer programmes such as Tandem Repeats Finder (TRF) [3, 22], MicroSAtellite (MISA) [23] MsDetector [24], WebSTR [25], Lu developed STR toolkit (LUSTR) [26] and many others [27, 12]. After identification, the flanking DNA sequences can be analysed for the presence of suitable forward and reverse PCR primers to test the STR loci. Currently, numerous computational tools are available to identify STRs in sequence data and design PCR primers suitable for amplification of specific loci [28, 29, 30, 31].

Data on STR sequences can be obtained from various public databases. STRBase (http://www.cstl.nist.gov/strbase/) provides information on human sequences [32], while the SNPSTR database (http://www.sbg.bio.ic.ac.uk/~ino/SNPSTRdatabase.html) contains data on STRs associated with SNPs in humans, mice, dogs, rats and chickens [33]. The Mouse Microsatellite Database of Japan (http://www.shigen.nig.ac.jp/mouse/mmdbj/top.jsp) and Mouse Genome Informatics (http://www.informatics.jax.org/marker/) list mouse microsatellites and the primers that flank each replicate. STR genotyping is used for a variety of purposes, including gene mapping, point mutation detection, marker-assisted breeding of congenic mouse strains, and many other applications.

2. 1. Genotyping

A genotype is the combination of alleles in a particular DNA variant within an individual. Each allele is inherited from one of the parents, resulting in ique genetic makeup for that individual [21]. Genotyping determines these DNA sequences, referred to as genotype, at positions within an individual's genome [34]. Early genotyping efforts that were used prior to the development of microsatellite markers utilized Southern blotting [35]. In the 1970s, pioneering work was carried out with DNA sequencing [36, 37], which made it possible to develop instruments for determining the genotype. Variations in the DNA sequence, which are recognized by bacterial restriction enzymes, cause the DNA to be cleaved at different locations, resulting in differences in DNA fragment length. These restriction fragment length polymorphisms (RFLPs) were first used for genotyping in the 1970s to 1980s. This early technique was labor intensive and often took several days and radioactive labeling was required [21]. The introduction of the polymerase chain reaction (PCR)

in 1985 revolutionized genetics and medicine [38, 39]. It enabled the amplification of an unlimited number of copies of specific DNA segments [21, 38, 39, 40].

In 1988, the Human Genome Organization (HUGO) was established and announced the plan to sequence the entire human genome [41]. The first sequences of human genome were published in 2001 by two groups, the publicly funded International Human Genome Project (HGP) Consortium [1] and Celera Genomics [42]. The second mammal to be sequenced was mouse; a high-quality draft of the mouse genome appeared in 2002 [20]. These achievements stimulated the programmes such as the Haplotype Mapping (HapMap) Project [43] and the 1000 Genomes Project [44]. The evolvement of the new high-throughput DNA sequencing platforms (first platform appeared on the market 2005) [45]: the next-generation sequencing (NGS) [21, 46] accelerated research. Another advancements included introduction of novel computational methods using long reads in 2014 [47] or the further development of DNA microarray technology [46]. This enabled the simultaneous interrogation of a large number of genetic variants and led to cost declines of human genome sequencing from $ 100,000,000 in 2001 to less than $1,000 in 2022 [48]. Whole genome sequencing (WGS) enables the determination of the complete DNA sequence of an individual's genome in a single experiment. This method offers the highest resolution for genotyping [21].

DNA variations, known as nucleotide polymorphisms, play a crucial role in understanding genetic diversity and linking specific DNA variants to phenotypic traits. In animal research, genotyping helps to identify animals for breeding, maintaining colonies, conducting experimental protocols [34] and especially in the study of genetic factors underlying complex diseases. The high density and mutational stability of single nucleotide polymorphisms (SNPs) make them particularly useful DNA markers for population genetics and for mapping susceptibility genes to complex diseases. A simple and inexpensive method for SNP genotyping involving a single polymerase chain reaction (PCR) followed by gel electrophoresis on polyacrylamide gel was first described in the technique called tetraprimer Amplification Refractory Mutation System (ARMS) PCR [40, 49]. This method was a breakthrough in laboratory practice, although it required the use of two specific primers and a control primer. The problem with this method lies in the preparation of the polyacrylamide gel, which is suitable as a molecular sieve for the resolution of a nucleotide in electrophoresis. Acrylamide is neurotoxic,

mutagenic, carcinogenic and teratogenic. Therefore, the preparation of polyacrylamide gel involves certain risks, especially with daily use, and there are also problems with unpolymerized residue and degradation after application. Polyacrylamide should be disposed of as toxic waste [50, 51]. The tetraprimer ARMS-PCR technique was used to analyze multiple SNPs, and the results were fully consistent with those obtained using an independent method, restriction fragment length polymorphism (RFLP) [40, 49].

For our genotyping requirements, we were looking for a cost-effective, low-risk method with a throughput resolution of around 6 base pairs (bp) that is suitable for many samples. Other important parameters are the time required for set-up, equipment and staff time. Our protocol includes DNA typing of STRs with a difference of at least 6 bp difference, using PCR and optimized high-resolution electrophoresis on agarose gel. This method is ideal for rapid testing of interval-specific congenic strains, marker-assisted breeding of congenic mouse strains, evaluation of the presence of transgenes and genotyping of intraspecific crosses, especially of parents with limited genomic differences [52].

2. 2. Disease models

2. 2. 1. Tick-borne encefalitis

Tick-borne encephalitis (TBE) is considered the most important tick-borne viral infection in Eurasia [53] with an incidence of 0.9 cases per 100,000 people. The TBE virus (tick-borne encephalitis virus) belongs to the genus of flaviviruses and is mainly transmitted to humans by infected ticks [54]. The disease can present in a variety of ways, from unnoticed infections and fevers that fully recover to severe or even fatal encephalitis. The factors contributing to this broad spectrum of symptoms are largely unknown, but it is likely that the genetic makeup of the host plays a role [53, 55].

2. 2. 2. Leishmaniasis

Leishmaniasis is a current problem in more than 98 countries, threatening the lives, health and quality of life of more than 12 million patients

and is one of the most common parasitic diseases in the world. The *Leishmania* parasites are transmitted to the host's body by stinging insects (*Lutzomnia*, *Psychodopygus*, *Phlebotomus*). Leishmaniasis causes major problems in tropical and subtropical regions of the world, but its occurrence has also been noted in southern European countries [56]. Leishmaniasis has three main clinical manifestations: dermal (see Fig. 1), mucocutaneous and visceral. Among the most important factors influencing the type of pathology are the type of parasite and the genotype of the host [57].

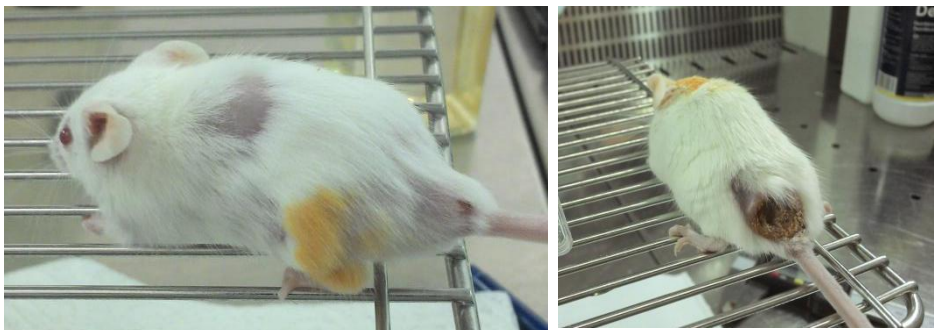

Figure 1: Dermal manifestation of leishmaniasis in different mouse individuals a) early stage of the lesion, b) lesion in 6th week of infection (the lesions are formed around the site of inoculation) [authors photo]

2. 3. Microbiome

Over the past decade, the microbiome has been shown to play a critical role in mammal health and disease, paving the way for innovative treatments and strategies. A variety of factors, including host genetics, age, gender, antibiotic use, diet and lifestyle, can influence the distribution of microbiota among individuals. These factors can significantly disrupt the microbial balance of the gut, often leading to disruption or 'dysbiosis'. The dysbiosis can potentially have a significant impact on an individual's health [58]. The gut microbiota potentially influences the maturation, development and functionality of key elements of the host immune system. In addition, the gut microbiota can adapt the host's immunity and even influence the host's immune response to parasites [59]. On the other hand, the host microbiota can make an individual more susceptible to parasite infections, potentially affecting the outcome of diseases [60]. Advances in biology and bioinformatics have enabled microbiome analysis, which has highlighted the crucial role of the host

microbiota and its metabolites in the onset and development of major human diseases and in regulating the function of the immune system [59, 61, 62, 63, 64]. Advances in next-generation sequencing technologies have enabled us to identify and quantify all microorganisms in studied samples. The huge amount of data generated in microbiome studies requires the use of advanced computational tools and technologies. The gut harbours a complex community of microbes, the so-called gut microbiome, which differs from person to person and depends on both genetics [65] and the environment [65]. A healthy gut microbiota is a complex ecological community composed of a large number of microorganisms, including bacteria, viruses, protozoa and fungi. The gut microbiota is highly diverse and varies between different regions of the gut [67].

The basic functions of microbial populations include the development of the immune system, protection against pathogens, the breakdown of nutrients, the synthesis of vitamins and amino acids, the metabolism of various substances and drugs, the influence on the central nervous system (CNS) and many others. Any imbalance (dysbiosis) has a significant impact on the health and well-being of the host. However, we have realised that microbiota imbalance is associated with various diseases. A decrease in microbial diversity and the proliferation of certain species can often lead to adverse effects such as inflammation or infection and contribute to diseases such as obesity, diabetes, Alzheimer's and others [68]. The present work was performed to characterise the impact of parasite infection on the composition and diversity of the microbiome in two parts of the digestive tract: the ileum and the colon of *Leishmania major*-infected and uninfected mice. *L. major* causes an infection in which the genotype of the host significantly influences the manifestation of the disease. We used our genotyping method to control allelic inheritance and to prepare an animal model to study the effects on gut microbiota composition using modern molecular microbiology tools and statistical methods (marker-assisted breeding of congenic mouse strains, denaturing gradient gel electrophoresis, Sanger sequencing, next-generation sequencing, principal coordinate analysis, LEfSe analysis), which could provide new insights into this severe infection and clarify the role of the host genotype. Clarifying the role of the microbiome in parasitic infections could provide new insights into this topic [69]. Many studies have shown the protective role of the intestinal microbiota against parasitic infections.

The composition of the intestinal bacterial population also modulates the course of parasitic disease [60, 62].

2. 3. 1. Animal studies

In order to gain an objective insight into the issue, the interactions between the microbiome and the host must be tested and extrapolated using model systems. Animal models are frequently used in human microbiome studies because they offer the advantages of easy experimental manipulation, control over certain variables, scalability and reproducibility that are largely unattainable in human studies. Gnotobiotic or germ-free mice are often used to study the effects of the absence of the microbiome or the administration of certain communities or strains of interest to the animal and the subsequent effects on the host. The advantage of animal models is the elimination of some variables that persist in humans (same age, diet, genotype and phenotype) [68].

Studies on the effects of *L. major* on the gut microbiome are limited. Only a few studies have investigated the role of the gut microbiota on the outcome of leishmaniasis. The only study investigating the relationship between leishmaniasis and the human microbiota was conducted by Lappan and his team. They applied a metataxonomic method to identify the prokaryotic and eukaryotic composition of fecal samples from individuals from an area in India due to *L. donovani* [70]. The study in germ-free mice is important to demonstrate the impact of the microbiota on a successful host immune response to *L. major* infection. Germ-free Swiss mice infected with *L. major* developed larger lesions and higher parasitism than conventional Swiss mice [71]. The use of germ-free C57BL/6 mice showed that the skin microbiota can also influence the modulation of the host immune response and the activation of macrophages [72, 73]. During infection with *L. major*, changes in the composition of faecal bacteria were observed: The content of members of the *Clostridia* class was higher in the non-healing strain BALB/c, and members of the *Gammaproteobacteria* class were higher in self-healing mice C57BL/6 [74]. The study of the microbiome can contribute significantly to the description of this infection and make it possible to link the results to a specific genotype.

2. 3. 2. Microbiome analyses

Marker gene sequencing techniques have revealed the function of the microbiome by focusing on a small part of the microbial genome. Identification of bacteria relies heavily on the 16S ribosomal RNA gene, particularly regions V3 and V4, and these genes have evolved to provide characteristic information that can be associated with specific taxa. The Operational Taxonomic Units (OTU) method is used to classify sequences with a divergence threshold of 97-99 %. Standard microbiome analysis tools such as QIIME provide a platform for taxonomic assignment. RNA-seq data are aligned to different genomes and pathways (e.g. KEGG) to determine the taxonomy of transcriptionally active organisms and the role of the genes they express. Bioinformatics software is used to compare and compile data from microbiome samples. Comparisons are made between different groups or variables to identify metabolic pathways or disease states [68].

Differences in the microbiome are assessed by comparing alpha and beta diversity metrics. Alpha diversity metrics measure the diversity within a sample and can be compared between different groups, e.g. between infected and uninfected individuals. Beta diversity, compares diversity between samples and is often calculated by comparing differences in traits, resulting in a distance matrix for all pairs of samples. A common calculation of beta diversity is Bray-Curtis dissimilarity, a quantitative measure that takes into account the abundance of taxa when comparing two communities. The Jaccard index or similarity coefficient is another qualitative measure that considers the presence or absence of traits rather than their relative abundance. Software for calculating alpha and beta diversity is included in common bioinformatics pipelines such as QIIME or Mothur.

In most microbiome studies, the analytical strategy is to look for differences in microbial diversity, abundance of taxa or functional elements (such as genes or pathways) between the groups being compared (e.g. infected vs. uninfected). Given the complexity of microbiome data, ordering methods based on dimensionality reduction, such as principal coordinate analysis (PCoA), are often used for visualization. These techniques convert the distance matrices into two- or three-dimensional visual representations of sample distances. These samples can then be conveniently labeled according to various categories (such as color, shape, etc.) to provide meaningful clinical metadata. In this way, the researcher can visualize

potential clustering by clinical variables in an unsupervised manner. Hierarchical clustering can be integrated with heatmaps to further categorize samples with similar bacterial profiles into branches of a dendrogram. Additionally, clinical metadata can be overlaid on the heatmaps to identify potential clinical cofactors associated with specific bacterial profiles [68].

Conventional statistical methods such as ANOVA or the Kruskal-Wallis test are often used to compare simpler traits between groups, e.g. differences in alpha diversity or frequencies of individual known specific taxa associated with disease phenotypes. However, these tests may provide inaccurate results due to the large number of variables. An alternative is linear discriminant analysis of effect sizes (LEfSe), a method that was developed specifically for microbiome data and is widely used. In this method, the Kruskal-Wallis rank sum p-value is first calculated to identify significantly differentially occurring features between groups, and then a linear discriminant analysis is performed to determine the effect size of these specific traits. Ideally, all associations should be confirmed by experimental tests.

Graphical networks are often used to contrast interactions in different states (e.g. healthy or diseased) or to visualise which organisms coexist or repel each other [68].

2. 4. Mapping and identification of candidate genes

Recent advances in molecular methods and quantitative techniques have provided us with tools to study the polygenic control of complex and quantitative traits. These tools provide a more detailed understanding of how genes influence phenotypic variation [75]. System analysis is a powerful tool for the study of genetic regulation of molecular networks that are involved in specific processes or diseases. It is based on a definition of gene expression patterns, classification of individual genes into ontological pathways, and use of distinct genetic polymorphisms that are influencing a disease [53].

There are two main approaches to deciphering the genetics of complex and quantitative traits: genome-wide scanning and the candidate gene approach. Genome-wide scanning does not rely on assumptions about the functional significance of specific traits, but is quite expensive. This method identifies broad

chromosomal regions with quantitative trait loci (QTLs) at the centimorgan (cM) level using DNA markers in the population that contain many candidate genes. The candidate gene approach is highly effective and economical for direct gene discovery. It has proven to be extremely powerful in studying the genetic architecture of complex traits [75]. The definition of global gene expression can be monitored using DNA microarrays or RNA-seq, with results validated by real-time PCR [76]. Cellular patterns of expression of mRNA of interest can be also revealed by *in silico* studies [53].

Our previous studies have shown that BALB/c mice exhibit moderate susceptibility to TBE virus (TBEV) infection, whereas STS mice show high resistance. Interestingly, the recombinant congenic strain CcS-11, which contains 12.5% of the STS genome on a BALB/c background, is more susceptible to TBEV than BALB/c mice [77]. Remarkably, the mouse equivalents of the human genes controlling TBE (2'-5' oligoadenylate synthetase 1B (*Oas1b)*, CD209 antigen (*Cd209*), toll-like receptor 3 (*Tlr3*), C-C motif chemokine receptor 5 (*Ccr5*), interferon lambda 3 (*Ifnl3*) and interleukin 10 (*Il10*)) are present in CcS-11 on segments derived from the BALB/c strain. This indicates that these genes are identical in BALB/c and CcS-11 mice. Since these genes cannot be responsible for the phenotypic differences between the two strains, we searched for the responsible gene locus originating from STS. We genotyped F_2 offspring from a cross between BALB/c and CcS-11 strains to identify the STS allele responsible for the phenotypic trait. In addition, it is worth noting that the STS-derived genes in CcS-11 may function by regulating or epigenetically modifying these non-polymorphic genes of BALB/c [53].

2. 5. Summary

Sequencing methods offer the highest resolution for genotyping (with a resolution of one nucleotide), but they are very expensive, especially when a large number of samples are analysed [21, 46]. DNA microarrays can be used to genotype thousands of markers simultaneously, but their use for mass testing is very expensive [21]. Cheaper methods based on electrophoresis (e.g. TETRA ARMS) are usually associated with a certain risk (working with acrylamide, working with radioactive labelling), require a lot of time or are labour intensive (e.g. RFLP) or have a low

resolution (microsatellites) [21]. For our genotyping purposes, we needed to find a simple, low-cost and low-risk method that is intended for daily routine use and where the achievable throughput is at a resolution of about 6 bp, which is suitable for many samples. The idea was to use microsatellite markers and improve the resolution of detection on agarose gels. There are significant differences in genotyping methods in terms of the time required for set-up, equipment and labour, which contribute to the cost of the method and the overall time of the experiment [52].

3. Aims of work

Engineering and molecular genetics are two disciplines that can complement each other effectively in the field of biotechnology. Engineering focuses on the application of scientific and mathematical principles to evaluate results, while molecular genetics is concerned with the study of the structure and function of genes at the molecular level. The newly developed methods could stimulate a revolution in areas such as genetic engineering and gene mapping. These methods offer new insights into the development and progression of infectious diseases. The aim of this work is to deepen the understanding of molecular biology methods as a tool for biomedical technology.

Aim of this project:

1. To develop a suitable method for estimation of the length of short tandem repeats in DNA, establish the discriminatory capacity of the method and determine the optimal conditions (molecular sieve density, voltage, pH and temperature). Establish a protocol for the subsequent implementation and publication of this protocol for the general application of this method.
2. To perform proof of the concept of the novel method in the models of different infectious diseases such as
    a. the influence of the genetic background and the infection with the parasite *Leishmania major* on the microbiome of the host
    b. genetic control of the survival after infection with the tick-borne encephalitis virus (TBEV)

4. Materials and methods

4. 1. Materials

4. 1. 1. Parasite

In our complex comparative microbiome study, we infected mice with the protozoan parasite *Leishmania major* LV 561 (MHOM/IL/67/LRC-L137 JERICHO II). Amastigotes were transformed into promastigotes using SNB-9. 10^7 promastigotes from the 6-day-old subculture 2 were inoculated subcutaneously into the tail base in 50 µl sterile saline [78]. The inoculum was prepared according to the protocol for the preparation of highly infectious *Leishmania* promastigotes [79]. Uninfected control mice were injected with 50 µl sterile saline solution.

4. 1. 2. Virus infection

The gene mapping and identification experiments were performed with the European prototypic TBEV strain Neudoerfl (a generous gift from Professor F. X. Heinz, Medical University of Vienna). This strain was passaged five times in the brains of infant mice prior to use in this study [77]. The mice were infected subcutaneously with 10^4 a plaque-forming unit (pfu) of the virus.

4. 1. 3. Mice

4.1.3.1. Microbiome

The complex comparative microbiome experiment was divided into two parts due to the large number of participating strains. These strains originate from two groups that are genetically different but internally related: CcS/Dem (BALB/c, STS, CcS-5, CcS-12, CcS-20) and OcB/Dem (O20, C57BL/10 B10, B10.O20). All participating animals were genotyped with appropriate markers to reveal/verify their genetic background. The experiment included only female mice aged 8-16 weeks (sexual and immune maturity). After weighing and tagging, the animals were used for inoculation. Infected animals were inoculated with parasites according to the protocol [79], the uninfected animals were inoculated with sterile saline

into a base tail. All experimental animals were fed with LASQCdiet® Rod18 (LASvendi, Soest, Germany). The mice were housed in plastic cages with wood-chip bedding, situated in a specific pathogen-free room with a constant temperature of 22°C (Figure 2). The size of skin lesions was measured weekly using an electronic digital professional LCD calliper (Shenzhen Xtension Technology Co., Ltd. Guangdong, China), which has an accuracy of 0.02 mm. The infection lasted for 8 weeks, then the experimental animals were weighed, euthanized and dissected. The dissected organs (liver, spleen) were weighed, cut, immediately frozen in liquid nitrogen and stored in a freezer until processing. The digestive tract was removed and separated into ileum and colon sections, frozen and subjected to microbiota analysis. A section of each organ was prepared for histological preparations. A higher number of individuals is required for the statistical significance of the entire study.

Figure 2: Individualy ventilated cages [autors photo]

The first part of the microbiome experiment included 113 experimental animals from 8 mouse strains (BALB/c, CcS-1, CcS-4, CcS-5, CcS-12, CcS-18, CcS-20, STS) to analyse in detail the genetic effects on the composition of the microbiota in health and disease. These strains were selected as they respond differently

to infection with *L. major* [80]. The strain BALB/c is susceptible to immunopathologic manifestations of leishmaniasis, whereas the strain STS is resistant to these manifestations. CcS strains are inbred strains created by crossing BALB/c and STS strains (see Fig. 3), Each CcS strain carries different donor segments of the strain STS (12.5 %) on the genetic background of BALB/c (87.5 %). The individual CcS strains vary greatly, and in some of these strains was observed even higher resistance to cancer than in the strain STS [81].

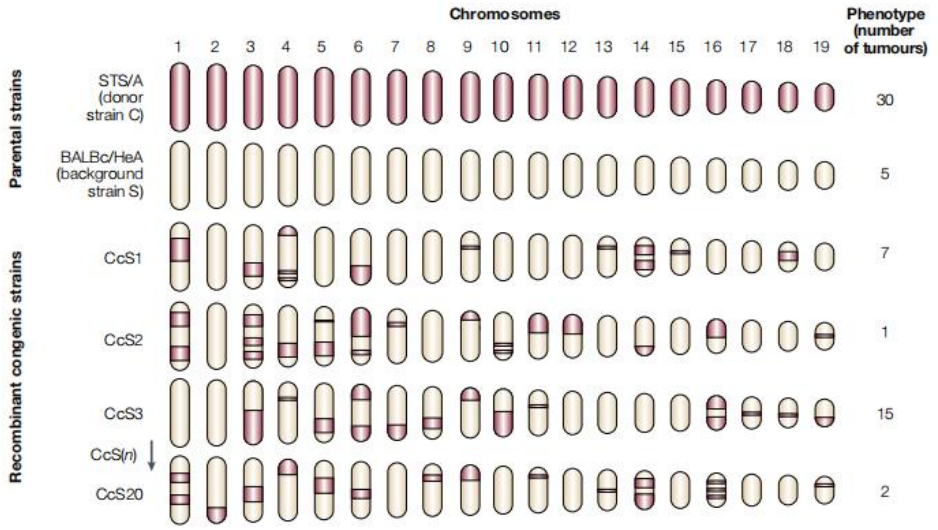

Figure 3: Composition of recombinant congenic strains [81]

The second part of the microbiome study consisted of 98 experimental animals from 5 mouse strains (O20, C57BL/10 [B10], B10.O20). These strains are also interesting due to the various responses to infection with *L. major*. The two strains B10 and O20 are resistant to *L. major*, which makes the susceptible strain B10.O20 even more interesting. B10.O20 is an inbred strain formed by crossing strains B10 and O20, which carries short donor segments of the strain O20 (4%) on the genetic background of B10 (96%) [82].

4.1.3.2. Gene mapping

The gene mapping and identification study consisted of 417 female F_2 offspring of an intercross between strains CcS-11 and BALB/c generated at the Institute of Molecular Genetics CAS. Mice were tested in three experimental groups at the Institute of Parasitology, CAS. The age of the mice varied between 8 and 16 weeks (at the time of infection). The first part of the experiment consisted of 120 F_2 mice, the second part of 121 F_2 mice and the third part of 176 F_2 mice.

Sterilized pellet food and water were provided ad libitum. The mice were housed in plastic cages with wood shavings in a special pathogen-free room with a constant temperature of 22°C and a relative humidity of 65%. Mice were examined for mortality and for the presence of ruffled fur and paresis 35 days post-infection (p.i.) with TBEV in three independent experiments at the Institute of Parasitology AS CR. The mice were euthanized 35 days after virus infection, the organs were removed and processed for detailed analysis [53].

4. 2. Methods

4. 2. 1. Genotyping

The presented genotyping method is based on DNA isolation, PCR amplification, optimized high-resolution electrophoresis and detection of the results (Figure 4), where marker selection is crucial for this method. This approach is suitable for rapid testing of interval-specific congenic strains, marker-assisted breeding of congenic mouse strains, evaluation of the presence of transgenes, and genotyping of intraspecific crosses, especially those derived from parents with limited genomic differences. The diagram (Figure 4) consists of the methods described below in chapter 4.2 and is intended to facilitate navigation in the following text. A detailed description of the methods used can be found in the published genotyping protocol [52].

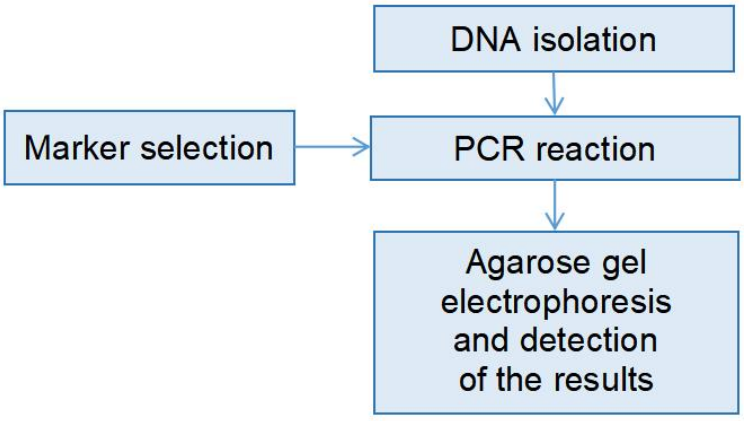

Figure 4: Diagram of the methods used in the genotypization method

4.2.1.1. DNA isolation

DNA was isolated from the tails using a standard proteinase method [83]. Tails were placed in 1.5 ml Eppendorf tubes for digestion in 500 µl SE buffer/proteinase K, 55°C overnight. 250 µl of pre-warmed, saturated 6M NaCl solution was added and the tubes containing the solution were gently mixed and the tubes were cooled on ice for 10 minutes. The tubes were spun at 4°C for 30 minutes at low speed (3220 rpm) and the supernatant were transferred to new 1.5 ml Eppendorf tubes. 650 µl of isopropanol was added for percipitation to each tube and the tubes were inverted for mixing. The samples were incubated for 20 minutes at room temperature. DNA was extracted by centrifugation at maximum speed (12000 rpm) for 10 minutes at room temperature. The supernatant were discarded and to the pellet 500 µl of 70% ethanol were added and allowed to stand at room temperature for 60 minutes to wash the resulting DNA. The washed DNA was collected by centrifugation at maximum speed (12000 rpm) for 10 minutes at room temperature. The supernatant were discarded and the pellet were air dried in inverted tubes for 5 minutes. 200µl of TE buffer pH 7.5 was added to each tube and the pellet were resuspended by pipetting up and down several times. The DNA concentration was then measured with the NanoDrop [52].

4.2.1.2. PCR amplification

In each well of the 96-well plates, 20 µl of the reaction mixture was prepared for each sample, then the plate was sealed with a foil to prevent evaporation and placed in the thermocycler. 20 µl of the reaction mixture contained 0.11 µM forward and reverse primers, 0.2 mM concentration of each dNTP, 1.5 mM $MgCl_2$, 50 mM KCl 10 mM Tris-HCl (pH 8.3), 0.01% gelatin, 0.4 U REDTaq® DNA polymerase and 40 ng sample DNA. The PCR reaction was performed using the DNA Dyad Peltier Thermal Cycler (Figure 5, other thermal cyclers can also be used) under the following conditions: Hot start DNA denaturation for 3 minutes at 94°C, followed by 40 cycles of denaturation for 30 seconds at 94°C, annealing for 60 seconds at 55°C and elongation for 60 seconds at 72°C. The final extension was carried out for 3 minutes at 72°C [52].

Figure 5: DNA Dyad Peltier Thermal Cycler (Bio-rad) [autors photo]

4.2.1.3. Agarose gel electroforesis

A 3% agarose gel was prepared for electrophoresis (the optimal concentration varies between 2% and 4%, depending on the size of the products, using a denser molecular sieve for smaller products). In 125 ml of 0.5 × TBE buffer, 3 g of Metaphore agarose and 0.75 g of UltraPure™ Agarose were added and the mixture was boiled in the microwave until melting, then 7 µl of EtBr from the stock solution (1 mg/ml) was added. After cooling for 10 minutes, the mixture was transferred to the prepared tray and a comb was inserted. The tray was left at room temperature for one hour to cool and solidify.

The tray with the gel was placed in an electrophoresis bath and the gel was overlaid with 0.5 × TBE buffer (in our case about 1.5 l buffer). 75 µl EtBr from the stock solution (1 mg/ml) was added to the buffer. After the comb was removed, the samples could be loaded. 3 µl of the standard (50 bp ladder) was loaded into the first well and 10 µl of the samples were loaded into the following wells. After loading the samples, a constant voltage of 150 V was applied (the optimal voltage varies between 75 V and 170 V, depending on the size of the products, with a lower voltage being used for smaller products) (Figure 6). The duration of the process depends on the size of the products and varies between 30 minutes and several hours. Immediately after turning, the voltage of the gel was transferred to the photo documentation system for photographing gels in visible and UV light,

e.g. Gel Doc XR+ (Bio-Rad) (Figure 7), and a photograph was taken to avoid blurring. We used different gel documentation systems with different degrees of resolution [52].

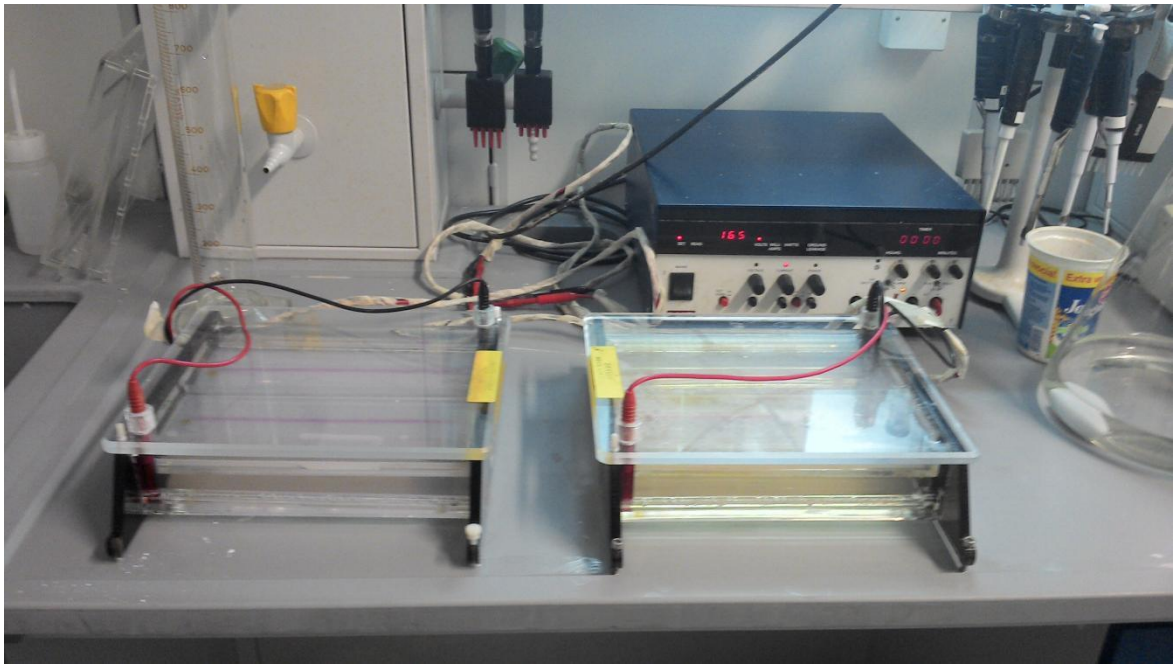

Figure 6: Aparatur for electrophoresis (voltage source, electrophoresis baths, gels with loaded samples [autors photo]

Figure 7: Gel Doc XR+ (Bio-Rad) [autors photo]

4.2.1.4. Primers used for genotyping

Genotyping for the complex comparative microbiome study

In the complex comparative microbiome study, 13 microsatellite markers (Generi Biotech, Hradec Králové, Czech Republic) were used for genotyping basic strains: D1Mit17, D2Mit52, D3Mit17, D3Mit49, D3Mit160, D4Mit149, D5Mit55, D5Mit114, D8Mit125, D9Mit2, D10Mit46, D11Mit62, D16Mit7 as described in articles [82, 84].

Genotyping of F_2 mice for the gene mapping

In the gene mapping study, 16 microsatellite markers (Generi Biotech, Hradec Králové, Czech Republic) were used for genotyping in the F_2 hybrid mice between CcS-11 and BALB/c: D1Mit403, D3Mit45, D7Mit25, D7Nds5, D7Mit18, D7Nds1, D7Mit282, D7Mit259, D8Mit85, D10Mit12, D10Mit46, D12Mit37, D16Mit73, D19Mit51, D19Mit60, D19Mit46 as described in article [85].

4. 2. 2. The complex comparative microbiome study

The complex comparative microbiome study consisted of a pilot study (Figure 8) and a main study (Figure 9). The diagrams (Figure 8, 9) are intended to facilitate navigation in the following text.

In the pilot microbiome study (Figure 8), not all experimental strains were processed and different methods were used to determine the microbiome composition than in the main microbiome study (Figure 9). Pathophysiological manifestations (lesion size, parasite load in the organs, hepatomegaly and splenomegaly) were investigated in both the pilot and the main study and related to the tested genotypes using statistical analyses. In the pilot study, total microbiome DNA was isolated from the ileum and colon of selected mouse strains and further analysed by DGGE in the amplified V4 - V5 16S rDNA region. A principal coordinate analysis was performed using the electropherograms to compare the composition of the microbiome. PCoA was performed to compare strains to each other as well as within individual strains for infected and uninfected individuals. Bands of interest were selected from the dendrograms and identified

by Sanger sequencing. The sequences obtained in FASTA format were processed and phylogenetic trees were generated. Some samples from the pilot study were used for the training run of Ion torrent PGM and processed in the same way as in the main study. The result of this training run is presented below in chapter 5.2.1.5, but not all samples from the pilot study were included in this training run, which is why this approach is not shown in the schematic (Figure 8). The pilot study is described in detail in chapter 4.2 of the dissertation. The pilot study has shown that the microbiome of the digestive tract (ileum, colon) changes during a leishmaniasis infection depending on the genotype of the host.

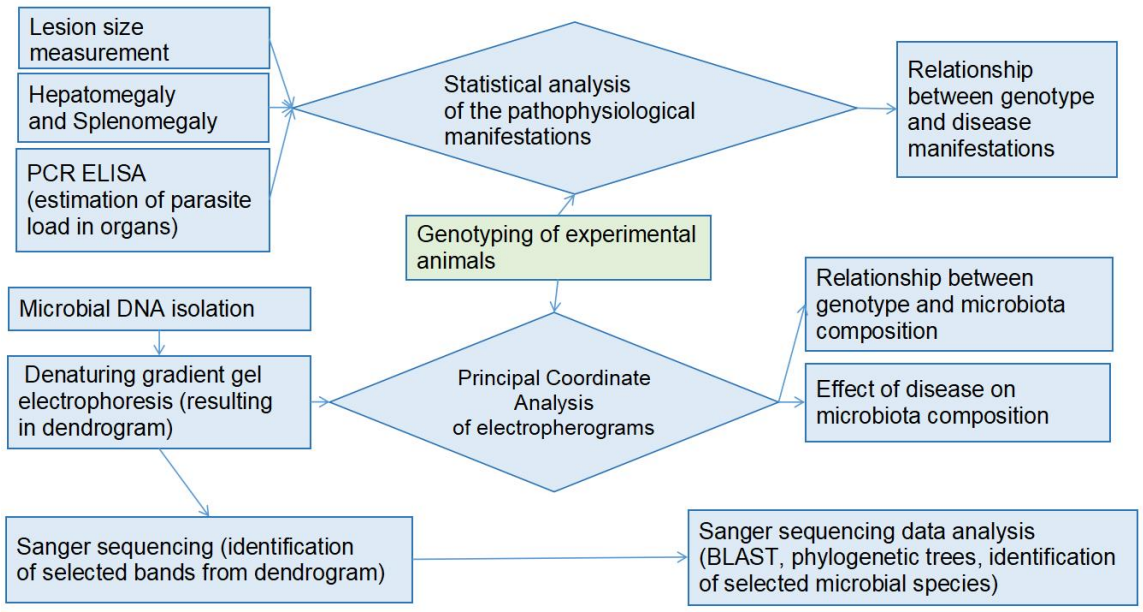

Figure 8: Diagram of the methods used in the pilot microbiome study

Based on the results of the pilot study, a detailed main microbial study was carried out using more expensive methods (Figure 9). In the main microbiome study, total microbiome DNA was isolated from the ileum and colon of all mouse strains. The DNA was amplified in the V4 - V5 region of the 16S rRNA and sequenced using Ion torrent PGM. The sequences were obtained in FASTQ format. These sequences were subjected to NGS data analysis. The results of NGS data analysis showed alpha diversities (Shannon index) between different strains of mice, plots of principal coordinate analysis (PCoA) showed distinct clusters between different groups (susceptible/resistant) of infected or uninfected mice and relative frequencies

of microbial population at family level between different infected or uninfected strains of mice. The results of the linear discriminant analysis were presented in two ways: by an LDA score diagram and by a cladogram.

A description of the methods used in the main microbiome study can be found below in chapter 4.2 of the thesis and the details in the article [86].

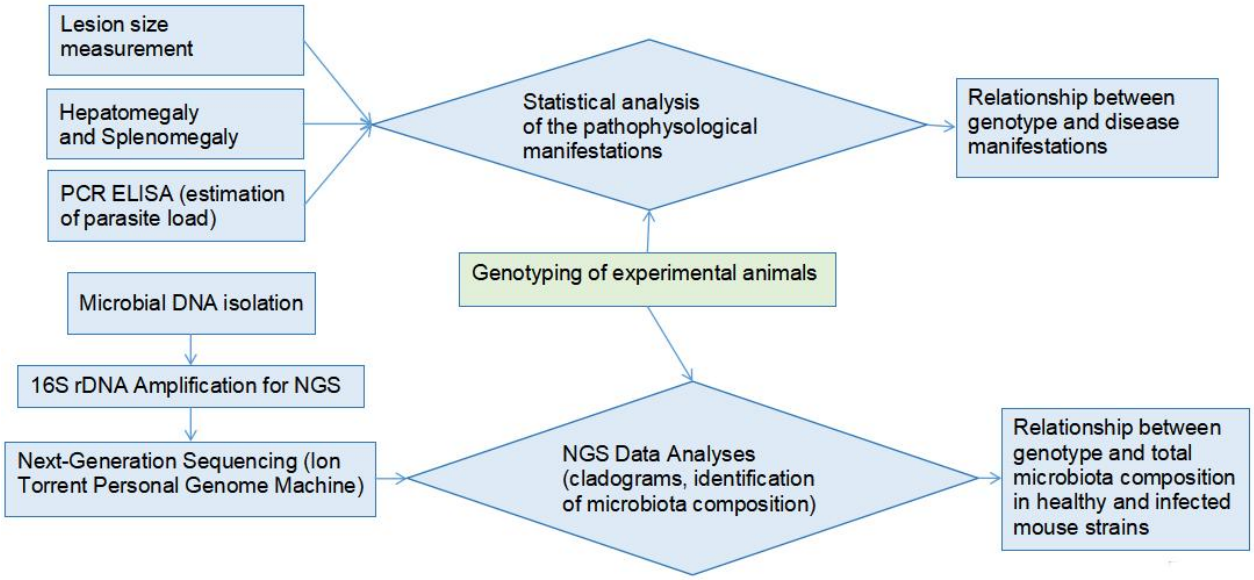

Figure 9: Diagram of the methods used in the main microbiome study

4.2.2.1. Lesion size measurement

The measurement of lesion size is a technique used to observe and compare the pathophysiological parameters of infected mice. The development of the lesions was observed once a week from the second week after inoculation. The diameter of the skin lesions was measured with a Profi LCD Electronic Digital Caliper (Shenzhen Xtension Technology Co., Ltd. Guangdong, China) with an accuracy of 0.02 mm. The area of lesions was estimated as eclipse according to equation (1). The lesion areas were monitored weekly to estimate the growth dynamics.

$$A = \pi \cdot \frac{w \cdot h}{4} \quad (1)$$

where A is lesion area, w is width, h is height and π is Archimedes constant (3.14)

4.2.2.2. Hepatomegaly and Splenomegaly

Hepatomegaly and splenomegaly are the parameters used to compare the degree of pathophysiology of infected mice. Hepatomegaly is an enlargement

of the liver and splenomegaly is an enlargement of the spleen, often caused by persistent infection. These signs are among the basic pathophysiologic manifestations of leishmaniasis. The values for the statistical evaluation of splenomegaly and hepatomegaly were estimated using the experimental equation (2). These values must be optimized for the weight of the mouse, as the mouse strains also differ in size and weight. To calculate these values, the spleens and livers were weighed immediately after dissection and the weight of the mouse was determined the day before the end of the experiment.

$$V_0 = 1000 \cdot \frac{m_0}{m} \qquad (2)$$

where V_0 is a representative value (for splenomegaly/hepatomegaly), m_0 is the weight of the organ (spleen/liver), m is the weight of a mouse

4.2.2.3. PCR ELISA

PCR ELISA is a molecular genetic method for the qualitative and quantitative estimation of the number of parasites in organs (parasite load). In our complex comparative microbiome study, we measured organs (liver, spleen) of infected mice. The PCR-ELISA was performed according to the protocol [87]. DNA for PCR ELISA was isolated with proteinase K and PCR mixtures were prepared under sterile conditions according to the protocol. The thermal profile consisted of an initial denaturation for 90 seconds at 94 °C, followed by 27 cycles for spleen and 24 cycles for livers, respectively. Each cycle consisted of 30 seconds at 94 °C, 45 seconds at 53 °C and 60 seconds at 72 °C and a final elongation of 10 minutes at 72 °C. The PCR products were added to the streptavidin-coated wells used for the ELISA reaction. The absorbance values were read in a spectrophotometer at a wavelength of 405 nm with a reference filter of 620 nm. The concentration of *L. major* DNA in the samples was estimated using a linear least squares– regression analysis [87].

4.2.2.4. Statistical analysis of the pathophysiological manifestations

The differences between strains in metabolic parameters were evaluated by the analysis of variance (ANOVA) and Newman-Keuls multiple comparison test at 95% significance using the program Statistica for Windows 13.0 (StatSoft, Inc.,

Tulsa, Oklahoma, U.S.A.). Analysis of variance (ANOVA) itself was used for comparing infected and uninfected mice for the parameters mentioned above. Strain and infection were fixed factors, individual experiments were considered as a random factor, age was taken as a covariate.

We used the R statistical environment [88] to visualize differences between different experimental groups using the principal component analysis (PCA). We sought to display differences in gross anatomical changes (Lesion.wk2, Lesion.wk4, Lesion.wk5, Lesion.wk6, Lesion.wk7, Lesion.wk8, Weight.change, Splenomegaly.wk8, Hepatomegaly.wk8), in changes in metabolic activity, and in all studied characteristics together. One sample (S_149_K90) was removed from the analysis, as it contained too many missing values. Other missing values were imputed by the average of the respective values measured in other mice of the same treatment, strain, and experimental batch. The values for each characteristic were then centred to have zero mean and unit standard deviation. PCA was performed using singular value decomposition [88].

4.2.2.5. Microbial DNA isolation

Total microbial DNA was isolated from colon and caecal samples with a QIAamp PowerFecal DNA kit (QIAGEN) according to the manufacturer's protocol, with the following exceptions. The disintegration of the cells was done with a FastPrep 24 device (Figure 10) at a vortexing speed of 6.5 m/s for 30 seconds, followed by incubation at 70 °C for 5 minutes. The vortexing and incubation were repeated once. The elution of DNA was done with 100 µl of elution buffer and stored at -20 °C until used.

Figure 10: Homogenization system FastPrep 24 (MP Biomedicals) [autors photo]

4.2.2.6. Denaturing gradient gel electrophoresis (DGGE)

Denaturing gradient gel electrophoresis (DGGE) is a molecular biology technique used for profiling microbial communities. We have used DGGE in pilot microbiome study. DGGE separates DNA according to its primary structure (sequence). DGGE was performed using DCode Universal Mutation Detection System (Figure 11). 20 µg of total microbial DNA was used for V4-V5 16S rDNA amplification according to the DGGE protocol [89] using the OneTaq 2X Master Mix (NEB). PCR amplicons were separated using a DCode Universal Mutation Detection System (Bio-Rad) on a 35-60 % DGGE gel at 60 °C for 18 hours.

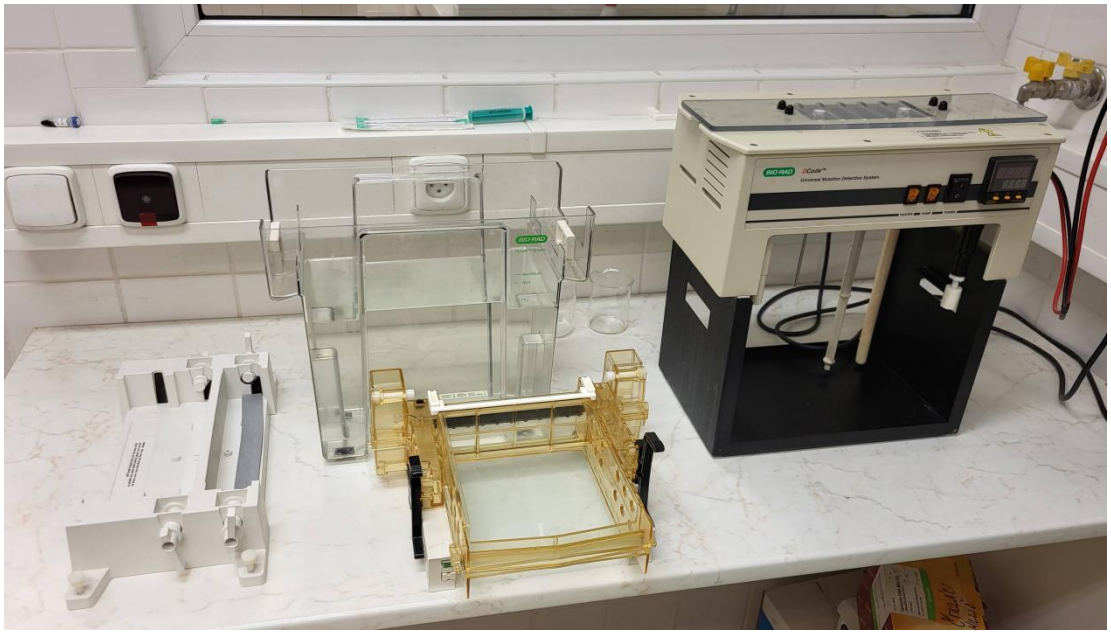

Figure 11: DCode Universal Mutation Detection System (Bio-rad) [autors photo]

4.2.2.7. Principal Coordinate Analysis (PCoA) of electroforeograms

Electrophorograms were analyzed using the BioNumerics software package (AppliedMaths) [90] to calculate a similarity matrix that was used to generate a dendrogram representing the similarities between the microbial profiles. The similarity matrix was also used to calculate a principal coordinate analysis (PCoA) for grouping.

4.2.2.8. Sanger sequencing

Sanger sequencing is also known as the chain termination method. In this technique, the DNA polymerase incorporates fluorescently labeled dideoxynucleotides to terminate chains during DNA replication, resulting in DNA fragments of different lengths [36]. The sequence of the nucleotides is determined by separating the fragments according to their size. Sanger sequencing is very accurate with an error rate of approx. 0.01%, making it the gold standard for DNA sequencing [91].

We have used Sanger sequencing in pilot microbiome study. Selected significant (strong signal, unique position) bands from the DGGE were prepared for sequencing. The bands containing DNA were cut from the gel and the DNA was eluted in 100 µl ultra pure dH_2O. The remnants of the gel were

homogenized with a vortex and subsequent centrifugation allowed us to recover the selected part of DNA. The DNA thus obtained was amplified using the PCR reaction. The thermal profile consisted of initial denaturation for 5 minutes at 95 °C, followed by 35 cycles of 30 seconds at 95 °C, 20 seconds at 61 °C and 40 seconds at 72 °C, and a final elongation for 5 minutes at 72 °C. The reaction parameters were selected according to the publications [89, 92]. PCR amplicons were verified by electrophoresis in 1.5% agarose, purified using the QIAquick PCR Purification Kit (QIAGEN) according to the protocol and quantified using Nanodrop OneC. DNA was diluted to a concentration of 50 ng/µl and sent to the SeqMe facility for sequencing.

4.2.2.9. Sanger sequencing data analysis

For the pilot microbial study, the sequences were obtained in the FASTA format. For each sample, we obtained two data sets (sequence with forward and reverse primers). We merged these complementary sequences to generate a complete sequence thread using the Geneious R9 software [93]. The resulting sequences were identified using the nucleotide BLAST ® algorithm of the NCBI database [94].

4.2.2.10. 16S rDNA Amplification for NGS

In the complex comparative microbiome study, 20 ng of total microbial DNA isolated from different parts of gut was used to prepare PCR amplicons of the V4-V5 region of 16S rRNA according to [95]. The mixture contained One*Taq* DNA Polymerase (New England Biolabs) and the thermal profile consisted of initial denaturation for 5 minutes at 95 °C, followed by 35 cycles of 30 seconds at 95 °C, 30 seconds at 57 °C and 30 seconds at 72 °C, and a final elongation for 5 minutes at 72 °C. The PCR amplicons were checked by electrophoresis in 1.5% agarose (25 minutes at 90 V), purified with a QIAquick PCR Purification Kit (QIAGEN) according to the protocol and quantified by Nanodrop OneC (Figure 12).

Figure 12: NanoDrop OneC (ThermoFisher Scientific) [autors photo]

4.2.2.11. Next-Generation Sequencing (NGS)

NGS is a molecular genetic tool that uses high-throughput parallel sequencing of immobilised templates to generate large amounts of data. In the microbiome study, we used semiconductor sequencing to obtain a detailed composition of the microbial communities. The Ion torrent platform was used in collaboration with the Institute of Animal Physiology and Genetics. The obtained PCR products were used to prepare amplicon libraries for diversity analyses by a next-generation sequencing approach on a Personal Genome Machine (Life Technologies) according to [96].

Semiconductor sequencing

The Ion Torrent Personal Genome Machine (PGM) is a next-generation sequencing platform that utilizes a novel approach to DNA sequencing called semiconductor sequencing (Figure 13). This technology is unique in that it converts chemically encoded information (A, C, G, T) directly into digital information (0, 1) on a semiconductor chip [97].

Figure 13: Ion Torrent Personal Genome Machine (Life Technologies) [autors photo]

The principle of semiconductor sequencing is based on the natural process of DNA replication. In nature, when a nucleotide is incorporated into a DNA strand by a polymerase, a hydrogen ion is released as a by-product [97, 98]. The Ion Torrent PGM makes use of this natural process and detects the release of these hydrogen ions.

The sequencing proces s begins with the construction of the library, which includes fragmentation of the DNA, polishing of the ends with enzymes and ligation of adapters [99]. The prepared DNA library is then loaded onto an Ion Torrent chip. Ion Sphere particles containing the DNA template are located in each well of the chip [100].

The Ion Torrent sequencer then sequentially floods the chip with one type of nucleotide at a time. If the nucleotide introduced is complementary to the leading template strand, it is incorporated by the polymerase and a hydrogen ion is released. The charge from the released ion changes the pH of the solution in the well, which is detected by the proprietary ion sensor. This change in pH is converted into a voltage change which is recorded by the chip [97, 98].

If there are two identical bases on the DNA strand, the voltage is twice as high and the chip registers two identical bases. If the next nucleotide that floods the chip does not match, no voltage change is recorded and no base is called. This process

is repeated with each nucleotide type in a predetermined order until the entire DNA fragment is sequenced [97, 98, 99].

One of the main advantages of Ion Torrent's semiconductor sequencing is its speed. Because it is direct detection — no scanning, no cameras, no light — each nucleotide insertion is recorded in seconds. In addition, the technology is scalable, with the number of sensors on the chip increasing from ~1 million for the first generation Ion 314 chips to ~7 million for the second generation Ion 316 chips and 11 million for the Ion 318 chips, which equates to up to 5.5 million reads for 400-base sequencing [100].

In summary, Ion Torrent's semiconductor sequencing technology represents a revolutionary approach to DNA sequencing. By creating a direct link between chemical and digital information, it provides a fast, simple, scalable sequencing solution that any laboratory can afford. This technology is also expected to enable more cost-effective and reliable diagnoses, improving the health of people worldwide [97, 98].

200 ng of DNA from each sample was used to prepare sequencing libraries using a NEBNext® Fast DNA Library Prep Set kit(New England Biolabs) according to the manufacturer's protocol. Ion Xpress barcode adapters (Thermo Fisher Scientific) were used to label each sample. The libraries obtained were used to prepare a sequencing template using an Ion PGMTM Hi-QTM View OT2 kit (Thermo Fisher Scientific). The template was then sequenced on an Ion 316TM Chip Kit v2 (Thermo Fisher Scientific) using an Ion PGMTM Hi-QTM View Sequencing kit (Thermo Fisher Scientific). Two independent sequencing runs were performed.

4.2.2.12. NGS Data Analyses

Microbiome statistical analysis

The obtained bacterial 16S rDNA sequences in FASTQ format were selected for analysis. These sequences were analysed using the next-generation microbiome bioinformatics platform QIIME 2 2020.2 pipeline as described by [101]. Quality control, filtering and trimming were performed using the DADA2 processing method [102]. Subsequently, the resulting amplicon sequence variants (ASVs) were clustered and taxonomically classified using an open-source metagenomics tool

VSEARCH with the SILVA database release132, which contains 99% OTUs (operational taxonomic unit) reference sequences [103].

Alpha diversity (Shannon index): The Shannon index of diversity was calculated using the Kruskal–Wallis test. This assesses the diversity of microbial communities within each group. Beta diversity: After rarefaction of the samples, a principal coordinate analysis (PCoA) was performed based on the Bray-Curtis distance. This visualises the dissimilarity between the microbial communities in the different groups. The visualisations of alpha diversity in boxplots for the Shannon index and 2-dimensional PCoA plots were created using the packages qiime2R and ggplot2 in R-Studio (version 3.6.3). The ellipses of statistical significance around each group represent 95% confidence intervals. A p-value ≤ 0.05 was considered statistically significant. The adonis plugin in QIIME2 evaluated the influence of factors (such as susceptibility to infection and host genetics) on the relative composition of the microbiomes. Adonis/PERMANOVA analysis with the Bray-Curtis distance matrix assessed dissimilarity between samples using a permutation set of 999 [104].

The linear discriminant analysis (LDA) algorithm with effect size (LefSe) [105] in the Galaxy module http://huttenhower.sph.harvard.edu/galaxy was used to detect biomarkers. The factorial Kruskal-Wallis (KW) test and the pairwise Wilcoxon test were used to evaluate the different relative abundances of bacterial families between susceptible and infected resistant groups in the colon and ileum. A significance threshold of 0.05 and a logarithmic LDA score threshold of 2.0 were applied.

Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt 2) was used for the prediction of metabolic functions [106]. PICRUSt 2 predicted the functional composition of the metagenome based on marker gene data. The predictions were categorized into KEGG pathways at levels 2 and 3. The resulting abundance table was analyzed using STAMP v2.1.3 [107] using the uncorrected two-sided Welch's t-test to compare resistant and susceptible strains ($p < 0.05$) for statistical significance. In addition, PCA examined the relationships between functional capacities.

LEfSe analysis

Conventional statistical methods such as ANOVA or the Kruskal-Wallis test are often used to compare simpler traits between groups, e.g. differences in alpha diversity or frequencies of individual known specific taxa associated with disease phenotypes. However, these tests may provide inaccurate results due to the large number of variables. An alternative is linear discriminant analysis of effect sizes (LEfSe), a method that was developed specifically for microbiome data and is widely used. In this method, the Kruskal-Wallis rank sum p-value is first calculated to identify significantly differentially occurring features between groups, and then a linear discriminant analysis is performed to determine the effect size of these specific traits [68].

Linear Discriminant Analysis Effect Size (LEfSe) is a tool for high-dimensional biomarker mining [108]. It identifies genomic features (such as genes, pathways and taxonomies) that significantly characterize two or more groups in microbiome data. LEfSe determines the features that are most likely to explain the differences between classes by coupling standard tests for statistical significance with additional tests for biological consistency and effect relevance [105, 108].
The LEfSe method begins by identifying traits that are statistically different between biological classes. It uses the nonparametric factorial Kruskal-Wallis (KW) sum-rank test to identify traits with significantly different frequencies with respect to the class of interest [109]. Biological consistency is then examined within each subclass (i.e. for each individual subject) using a series of pairwise tests between subclasses [105, 108]. The Wilcoxon rank sum test is used for this purpose [109].

The final step of the LEfSe method is to perform a Linear Discriminant Analysis (LDA) to estimate the effect size of each differentially frequent feature. LDA is a dimensionality reduction method used to find the linear combinations of features that characterize two or more classes [109].

The LEfSe results can be visualized in two ways: by an LDA score diagram and by a cladogram. The LDA score diagram ranks the features according to their effect size and thus enables quick identification of the features that show the greatest differences between the classes [105, 108]. The cladogram, on the other hand, provides a taxonomic representation of the microbial communities and highlights the phylogenetic branches that are enriched in each clade [105, 108].

To summarize, LEfSe is a powerful tool for biomarker discovery in microbiome data. It combines robust statistical tests with biological consistency and effect size estimation and provides a comprehensive overview of the features that characterize the differences between microbial communities [105, 108].

4. 2. 3. Mapping and identification of candidate genes

The candidate genes were mapped and identified using the methods shown in Figure 14. The schematic representation of the method (Figure 14) is intended to facilitate navigation in the following text.

The pathophysiological data were collected from infected animals and the genotyping method was applied to experimental animals. The information obtained was processed with a statistical analysis and the linkage for a marker D7Nds5 was found. Whole genome sequencing with NGS was performed on the parental strains BALB/c and STS, from which the CcS/Dem series originated. These strains were processed and polymorphisms were found. The regions identified by linkage analysis were searched in detail and nine polymorphic genes were found in this region. A more detailed description of the methods used can be found below in chapter 4.2.3 of the dissertation or in a published article [53].

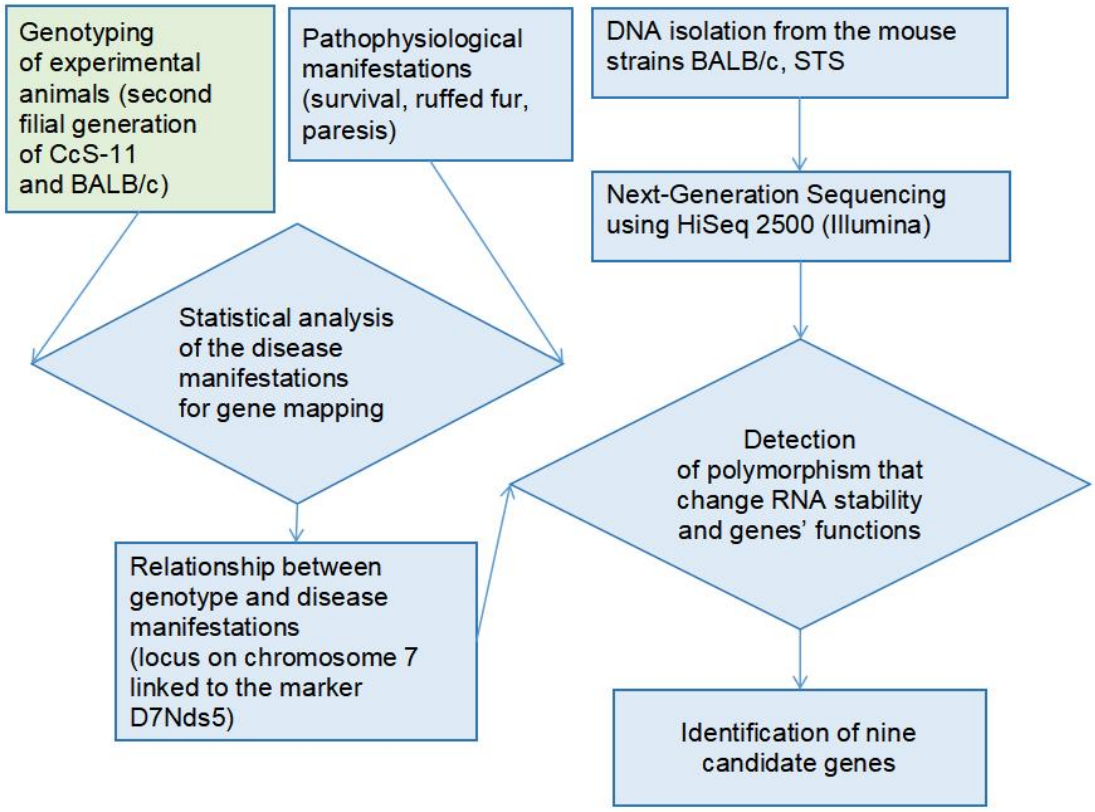

Figure 14: The schematic representation of the methods used in the mapping and identification of candidate genes

4.2.3.1. Statistical analysis for gene mapping

Survival, ruffed fur and paresis were treated as binary phenotypes (death/survival; presence/absence of symptom), and binary trait interval mapping was performed [110, 111]. A permutation test [112] was used to assess significance. This takes account of the limited genetic difference between the two strains. Based on 10,000 permutation replicates, the 5% significance LOD threshold was 2.56; the 10% threshold was 2.23.

4.2.3.2. Detection of polymorphisms that change RNA stability and genes' functions

In the gene mapping study, we have sequenced the genomes of strains BALB/c and STS using next-generation sequencing (NGS) system HiSeq 2500 (Illumina) (12x coverage). NGS data was preprocessed using software Trimmomatic [113] and overlapped paired reads were joined by software Flash [114]. Alignment -

reference mouse sequence mm10 (build GRCm38) - was performed using BWA (Burrows-Wheeler Aligner) program [115]. Mapped reads were sorted and indexed, duplicated reads were marked. Segment covering the peak of linkage on chromosome 7 from 36.2 Mb to 74.5 Mb was inspected for polymorphisms between BALB/c and STS that change RNA stability and genes' functions. Local realignment around indels, base recalibration and variants filtration was performed using software GATK (The Genome Analysis Toolkit) [116]. Variant annotation and effect prediction was performed by software SnpEff [117]. IGV (Integrated Genome Viewer) was used for visualization of results [118].

5. Results

5. 1. The improved method for genotyping

A new method for genotyping has been introduced. The development of the method began in 2009 [119] and the method was improved in 2011 [85] before being finalised in 2015 [52]. The method consists of DNA isolation, a PCR reaction, agarose gel electrophoresis stained with ethidium bromide and subsequent visualization with a transillumination device. Optimal conditions for the application were established (voltage magnitude, molecular sieve density, time required for band separation and dye concentration). The method was published in Protocols Exchange [52]. We performed two independent experiments, which are described below as a proof of concept of this method.

In the experiment focused on the study of the microbiome, the genotyping method was used to control inherited alleles in selected mouse strains [86]. In the experiment focused on mapping and identification of candidate genes, the method of genotyping was used to map the locus controlling the phenotypic manifestations of tick-borne encephalitis and to identify individual candidate genes [53].

5. 1. 1. The protocol for genotyping

A protocol for DNA typing of short tandem repeats (STR) differing in at least 6 bp is described. We optimized the analysis of PCR products in agarose electrophoresis by using a 4:1 mixture of Methaphore (Cambrex) or UltraPure™ (Invitrogen) agarose. This allowed us to separate PCR products with 6 and more bp difference in length (Figure 15).

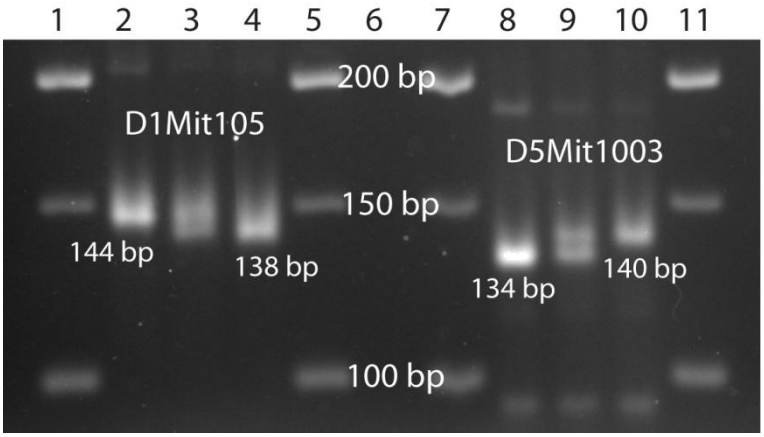

Figure 15: DNA typing using STR markers with PCR products length difference 6 bp. Picture was captured by author on GelDoc system (Bio-Rad): 1 - 50 bp ladder, 2 - BALB/c homozygote in the marker D1Mit105 (144 bp), 3 - heterozygote, 4 - STS homozygote (138 bp), 5 - 50 bp ladder, 6 - empty, 7 - 50 bp ladder, 8 - BALB/c homozygote in the marker D5Mit1003 (134 bp), 9 - heterozygote, 10 - STS homozygote (140 bp), 11 - 50 bp ladder. Gel size 23.8 x 25 cm [52]

The method has been successfully used to map mouse genes controlling susceptibility to *L. major* [120, 121] and to tick-borne encephalitis [53] and in the experiment focused on studying the microbiome to control inherited alleles in selected mouse strains [86]. During the development of the presented method, the proto-method was used to map mouse genes controlling susceptibility to *Trypanosoma brucei brucei* [85] and *L. tropica* [122].

The quality of DNA obtained by NaOH extraction is suitable for typing most markers and the whole procedure can be performed within one day (Figure 16).

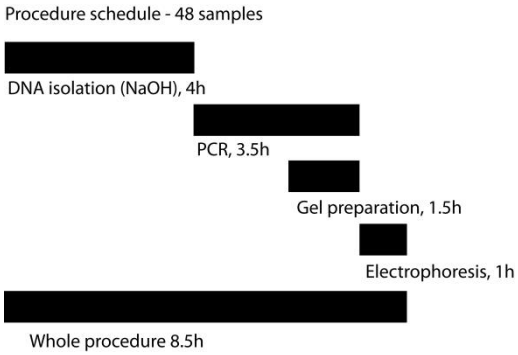

Figure 16: Time arrangement of the method [52]

If a higher DNA quality is required, the extraction can be performed with TRI reagent procedure or Proteinase K procedure.

The presented method is simple, robust, inexpensive, fast, and suitable for mass testing of a large number of samples and can also handle low quality DNA. The method has a wide range of applications, such as testing of interval-specific congenic strains, detection of point mutations, marker-assisted breeding of congenic mouse strains, a test for the presence of transgenes, knock-out or knock-in alleles in segregating experimental and breeding crosses, and for typing intraspecific crosses, especially those derived from parents differing in a limited percentage of their genome [52].

5. 2. Complex comparative microbiome study

To clarify the role of genotype in the composition of microbial populations, the method of genotyping was used to verify the genetic background of the individuals included in the microbiome experiment. The microbiome experiments on colon and ileum samples were performed as two independent studies using different methods. First, we conducted the pilot study to see if there was a reason to conduct a more elaborate study. The first study consisted of DNA isolation, denaturing gradient gel electrophoresis (DGGE), Sanger sequencing of the 16S rDNA region and PCoA analysis [123]. After evaluating the results of the pilot study, a second, more detailed study was conducted using advanced and more expensive methods. The second study included DNA isolation, NGS sequencing of the 16S rDNA region on an Ion Torrent PGM platform, PCoA analysis and LDA analysis (LEfSe) [86].

5. 2. 1. Pilot microbiome study

The microbiome pilot study on colon and ileum samples consisted of DNA isolation, denaturing gradient gel electrophoresis (DGGE), Sanger sequencing of the 16S rDNA region (cut from the gel), PCoA analysis and ion torrent sequencing. The pilot study consisted of a small number of test animals compared to the main study. The cost of this method is approximately 30,000 Kč for 100 samples, excluding ion torrent sequencing. The pilot study samples were used for training with the ion torrent PGM device, therefore the cost of ion torrent sequencing was not included in the cost of the pilot study. The pilot study served as a test study to determine if there was a reason to conduct a more detailed and expensive study.

5.2.1.1. Genotyping of animal model for pilot study

For the pilot study, we used an animal model consisting of three mouse strains that were uninfected or infected with *L. major*. These strains comprised the parental strains BALB/c and STS, and the recombinant congenic strain CcS-20. Each CcS/Dem strain consists of a unique, random set of genes, with about 12.5% coming from the donor strain STS and about 87.5% from the background strain BALB/c [124, 125]. The strains were classified as resistant or susceptible based on the pathology of their organs and the parasite load in the organs [80, 82]. Strain BALB/c is highly susceptible, shows extensive infiltration of parasites in its organs and develops significant skin lesions as well as splenomegaly and hepatomegaly [80]. The resistant strain STS has a low parasite count in its organs and shows no skin lesions, splenomegaly or hepatomegaly [80, 82]. The intermediate strain CcS-20 has a moderate number of parasites in the skin, develops no or only small lesions and shows no signs of splenomegaly or hepatomegaly [80, 82]. To confirm the quality of the breeding colony the strains were genotyped with microsatellite markers (Generi Biotech, Hradec Králové, Czech Republic): D1Mit17, D5Mit55, D5Mit114, D8Mit125, D9Mit2 as described in [84].

5.2.1.2. Leishmaniasis Development

Manifestations of leishmaniasis infection were most severe in the susceptible mouse strain BALB/c, followed by the recombinant congenic strain CcS-20. No clinical symptoms of infection were observed in the resistant strain STS (Table 1).

Table 1. Leishmaniasis: weight and lesion sizes in tested mice strains

Strain	condition	weight change (g)	lesion size (mm2)
BALB/c	infected	-1,09 ± 0,96	79,15 ± 34,48
BALB/c	uninfected	2,3 ± 1,49	0
CcS-20	infected	-0,62 ± 1,34	20,16 ± 34,35
CcS-20	uninfected	1,76 ± 1,13	0
STS	infected	2,03 ± 0,53	5 ± 14,14
STS	uninfected	1,97 ± 0,47	0

5.2.1.3. Ileum Microbiota

The standard ileum microbiota of healthy animals was dominated mostly by the *Lactobacillales* order in the strain BALB/c (79.65 %), while it was less frequent in the other two strains (about 50 %). The CcS-20 and STS strains had high levels of the S24-7 family (8.33 % and 22 %, respectively), while BALB/c mice lacked this otherwise common microbial group (Figure 17a). *Propionibacterium acnes* was also detected in the ileum tract of healthy CcS-20 mice. The BALB/c mice had the lowest microbiota diversity, as shown by the number of operational taxonomic units (OTU) (Figure 18a). However, those differences were not sufficient to distinctly separate microbiomes as tested by unweighted PCoA analyses of UniFrac distances (Figure 19a).

Leishmaniasis infection was associated with a decrease in *Bacteroidales* and *Propionibacterium* numbers in the strains CcS-20 and STS. The numbers of observed OTU were relatively low, and a slight decrease in diversity was observed in CcS-20 and STS mice (Figure 18a). Those changes affected microbiota profile groupings, by separating the resistant STS strain from the susceptible strain BALB/c and the intermediate strain CcS-20 (Figure 19b).

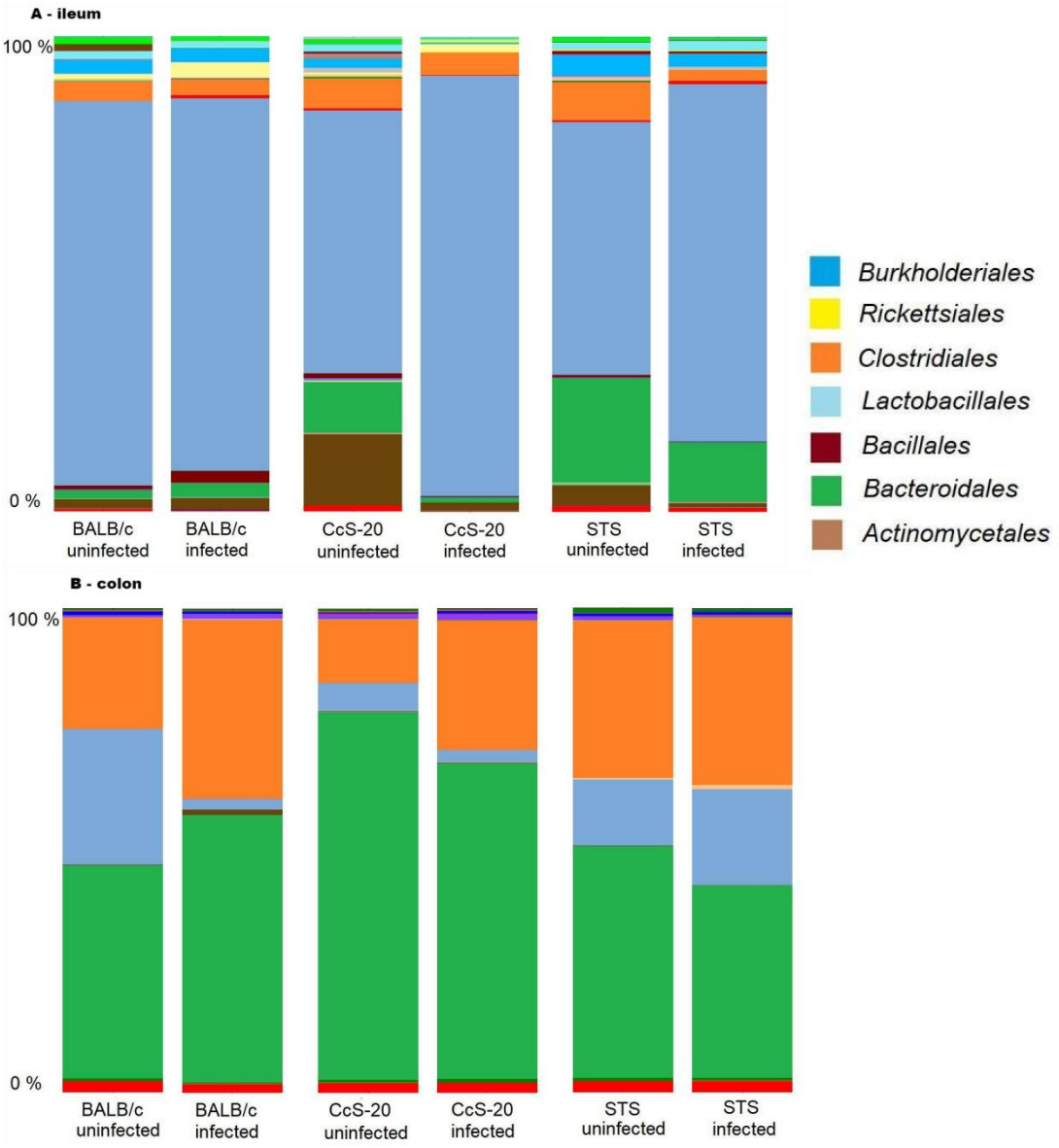

Figure 17: Relative bacteria abundance at the order level, in the ileum (a) and colon (b) samples

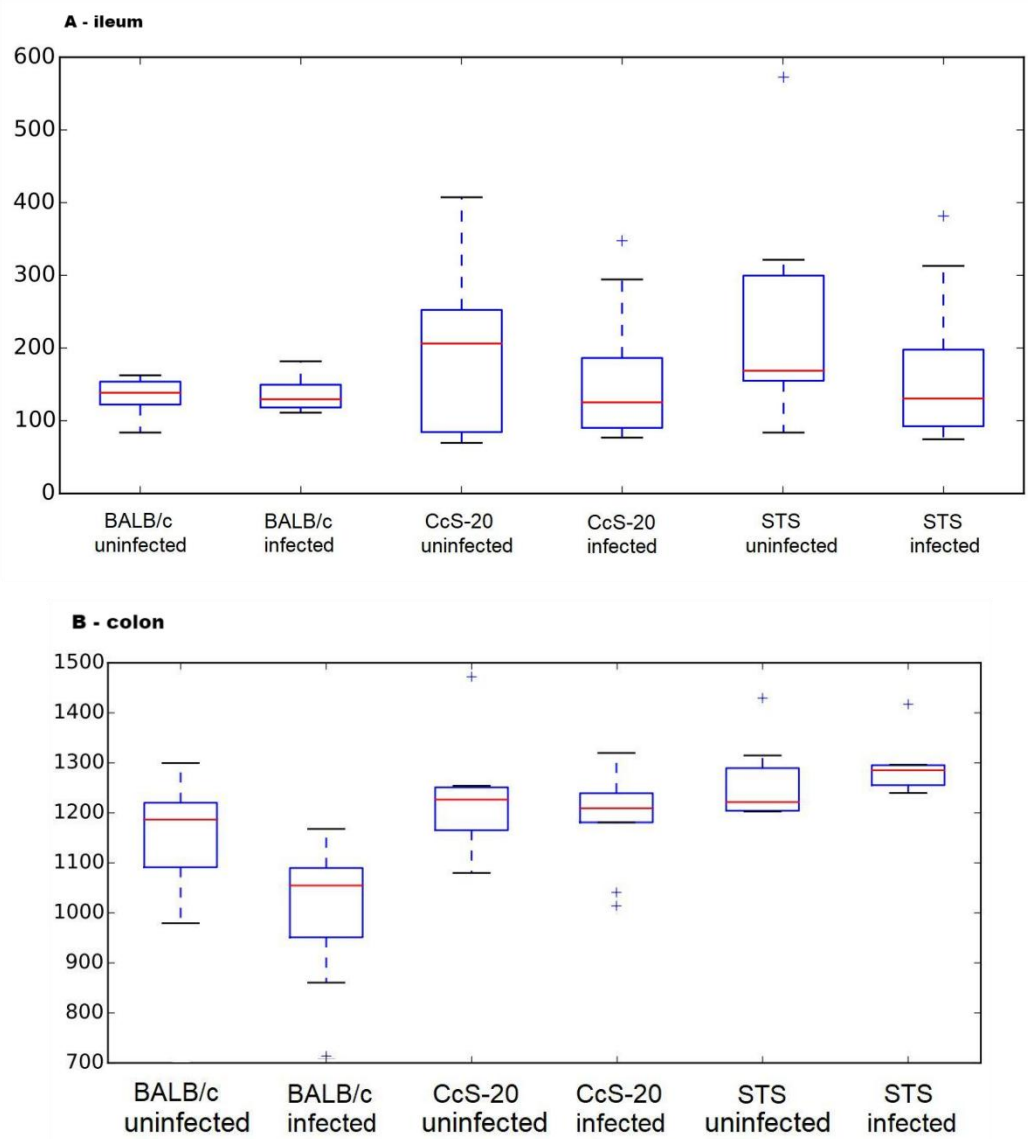

Figure 18: Boxplots demonstrating observed OTU in the ileum (a) and colon (b) samples

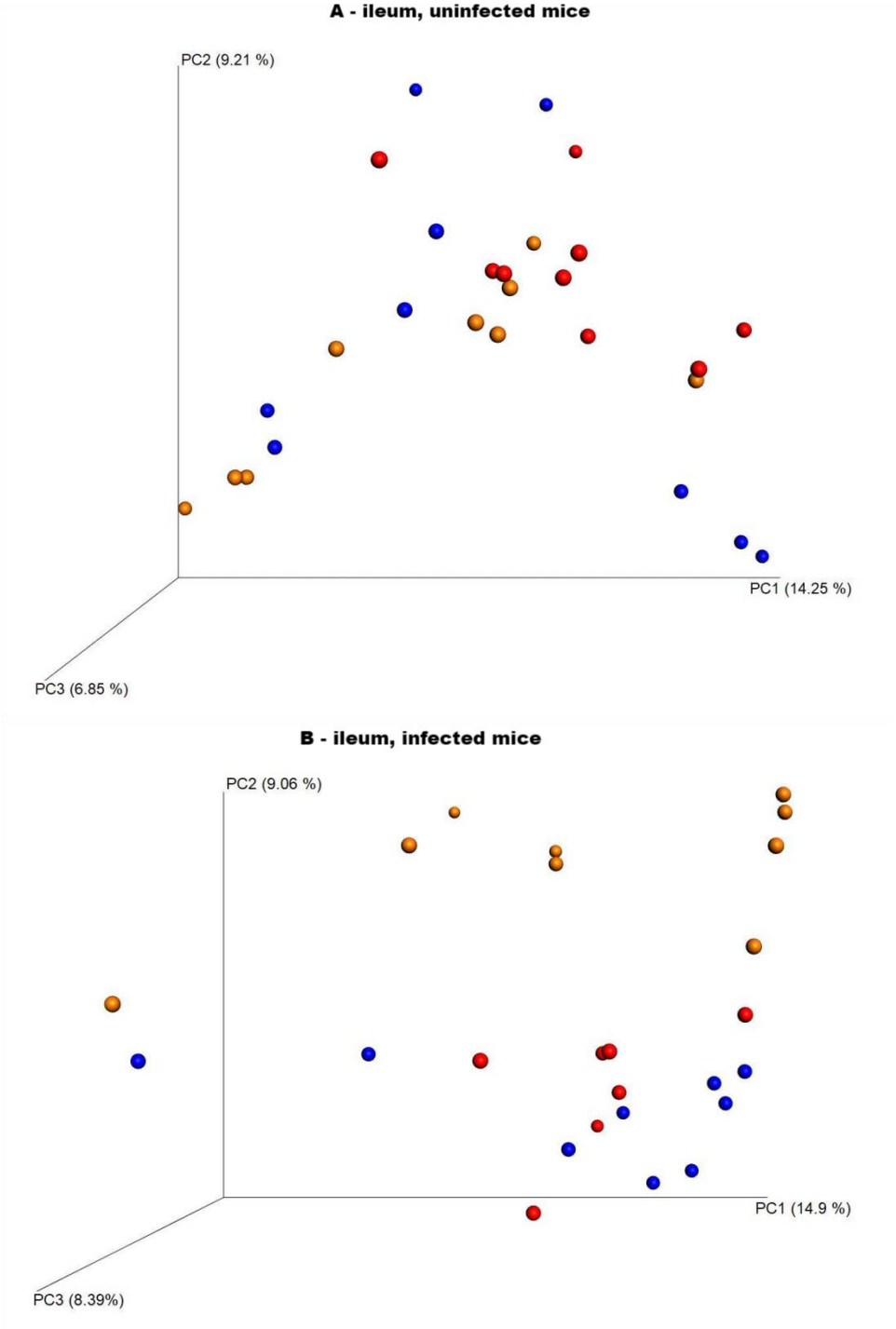

Figure 19: Variations in gut microbiota diversities in the ileum of uninfected (a) and infected (b) mice demonstrated by PCoA plots of unweighted UniFrac distances. (Red –BALB/c strain, blue –CcS-20 strain, and orange –STS strain)

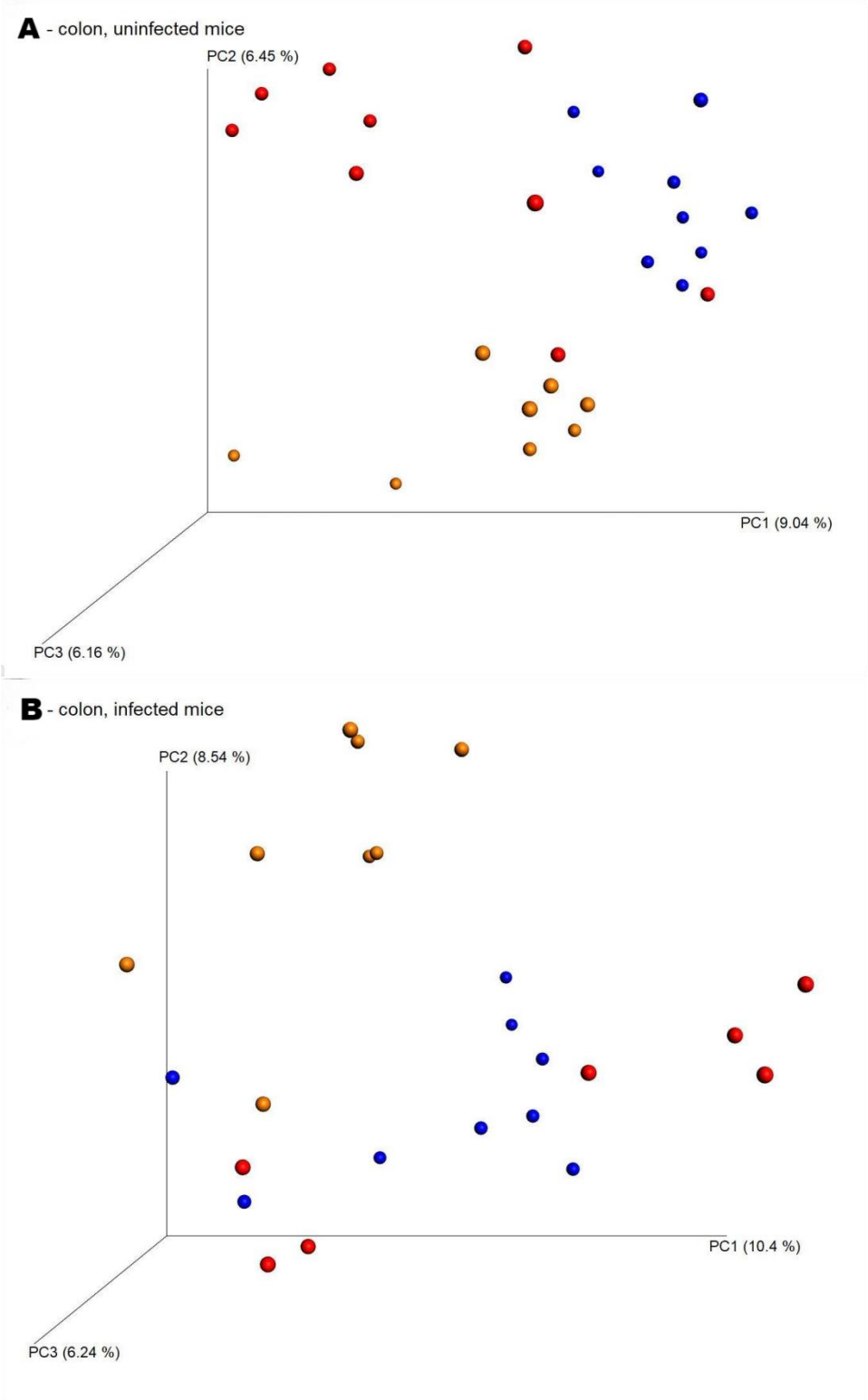

Figure 20: Variations in gut microbiota diversities in the colon of uninfected (a) and infected (b) mice, demonstrated by PCoA plots of unweighted UniFrac distances. (Red –BALB/c strain, blue –CcS-20 strain, and orange –STS strain)

5.2.1.4. Ileum and colon microbial profiles from PCR-DGGE data

For each of the mouse strain used in the study, a comparison was made between healthy and infected individuals. Dendrograms of microbial profiles from the ileum and colon were performed (Figure 21a, b). Selected bands were identified by Sanger sequencing as 1. *Bacteroides oleiciplenus* (with an accuracy of 89.23%), 2. *Lactobacillus johnsonii* (with an accuracy of 94.21%) 3. *Lactobacillus animalis* (with an accuracy of 97.06%), 4. *Mucispirillum schaedleri* (with an accuracy of 98.44%) and 5. *Anaerocolumna xylanovorans* (with an accuracy of 89.23%).

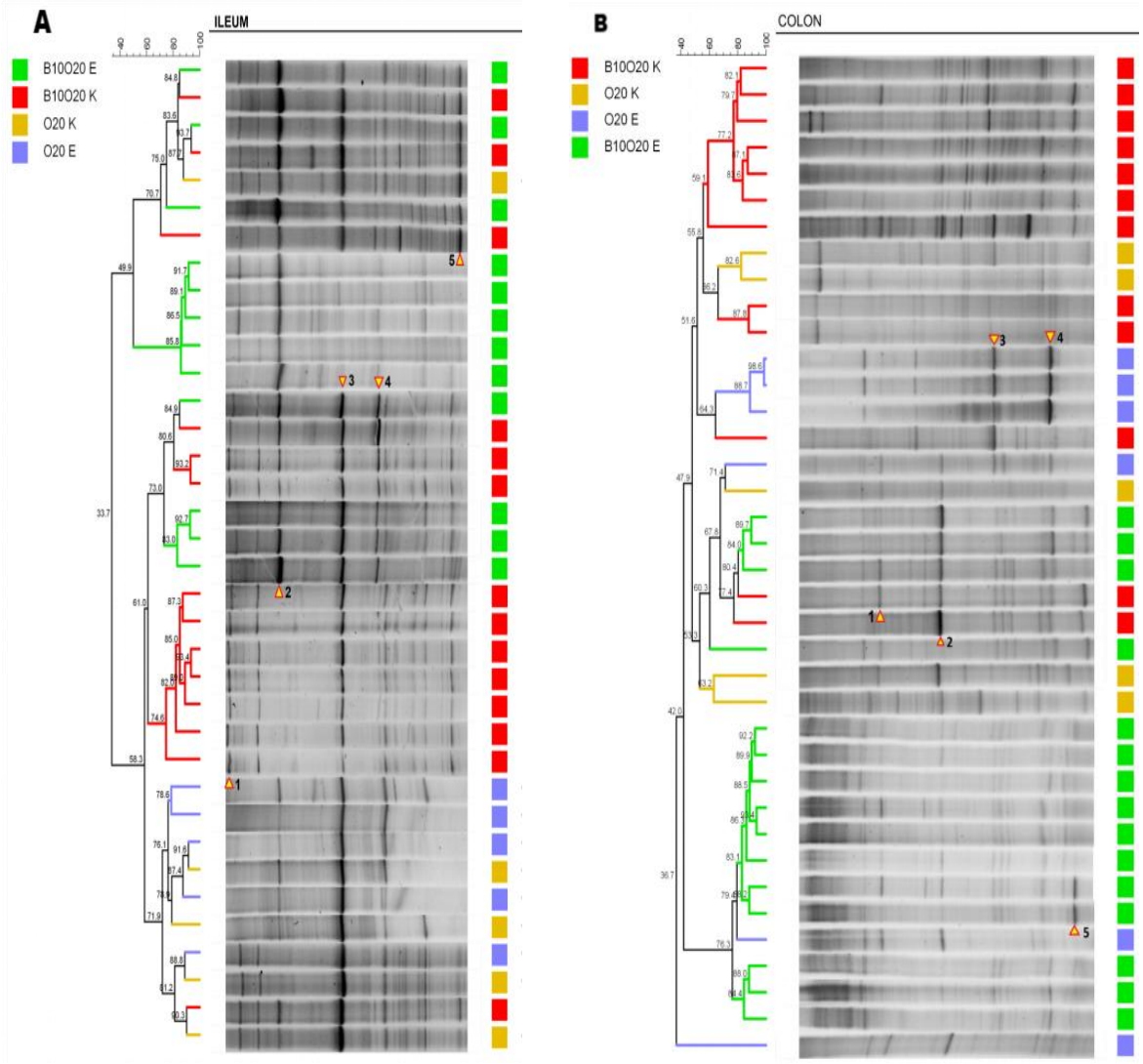

Figure 21: Dendrogram of microbial profiles from the ileum (a) and the colon (b) of mouse strains B10.O20 and O20. “E” stands for infected and “K” is a control group

Identification of microbiome composition of the gastrointestinal tract gave us a comparison between healthy and infected individuals in 6 mouse strains that differ in susceptibility to *L. major*. The PCoA analysis of preliminary data based on dendrograms (Figure 22a, b) shows a clear separation of infected B10.O20 from infected O20 and control samples, especially in colon samples. This data proves the influence of parasite infection on host gut microbiota, depending on the host genome. These results are consistent with [57].

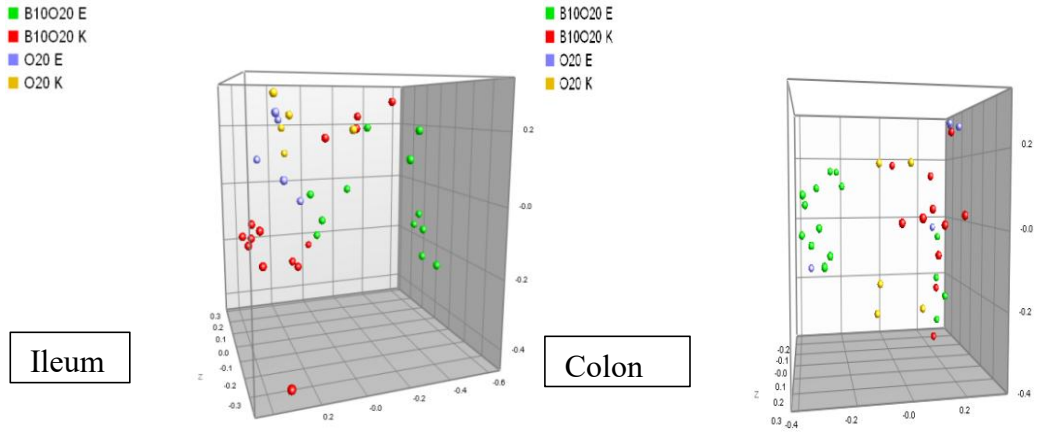

Figure 22: PCoA analysis of microbial profiles from the ileum (a) and colon (b) of mouse strains B10.O20 and O20. "E" stands for infected and "K" is a control group. Comparison of dendrogram data was performed in BioNumerics (AppliedMaths)

5.2.1.5. Colon Microbiota results by NGS data from pilot study

Colon samples of the microbiota of healthy animals were mostly comprised of the orders *Bacteroidales* and *Clostridiales*, followed by *Lactobacillales* (Figure 17b). Susceptible strain BABLB/c had the highest numbers of *Lactobacillales* (28 %). The numbers of detected OTUs were also significantly higher than in ileum samples (Figure 18b). Differences in colon microbiota among the tested strains were confirmed by unweighted PCoA analysis, with distinct groupings of each strain (Figure 20a).

The *Leishmania* infection caused notable microbiota changes in both susceptible strains, most notably a decrease in *Lactobacillales* numbers: from 28 % to 2.17 % in the strain BALB/c, and from 5.75 % to 2.64 % in the strain CcS-20 (Figure 17b). The strain BALB/c also exhibited a significant decrease in bacterial diversity, as expressed by the boxplots of OTU (Figure 18b). Those changes caused

a distinct separation of profiles in both groups after *Leishmania* infection, as tested by Unweighted PCoA analysis (Figure 20b).

5. 2. 2. Main microbiome study

The main microbiome study on colon and ileum samples involved DNA isolation, NGS sequencing of the 16S rDNA region on an Ion Torrent PGM platform, PCoA analysis, LDA analysis (LEfSe). The cost of this method is approximately 450 000 Kč for 100 samples [86].

5.2.2.1. Genotyping of animal model for main study

For main microbiome study we used two animal models consisting of eight mouse strains that were both uninfected and infected with *L. major*. These strains originate from two groups that are genetically different but internally related: CcS/Dem (BALB/c, STS, CcS-5, CcS-12, CcS-20) and OcB/Dem (O20, C57BL/10 (B10), C57BL/10-$H2^{pz}$ (B10.O20)). Each CcS/Dem strain consists of a unique, random set of genes, with approximately 12.5% coming from the donor strain STS and approximately 87.5% from the background strain BALB/c [124, 125]. B10.O20 carries 3.6% of the genes of the strain O20 on the genetic background of B10 [126]. The strains were categorized as resistant or susceptible based on the pathology of their organs and the parasite load in the organs [80, 82]. The strains BALB/c and CcS-12, which are highly susceptible, show extensive infiltration of the parasites into their organs and develop considerable skin lesions as well as splenomegaly and hepatomegaly. The CcS-12 strain in particular harbors an even greater number of parasites in its lymph nodes compared to BALB/c [80]. The susceptible strain B10.O20 carries a relatively high load of parasites in the skin and develops skin lesions [82]. The resistant strains STS, O20, B10 and CcS-5 have a low parasite count in their organs and show no skin lesions, splenomegaly or hepatomegaly [80, 82]. The intermediate strain CcS-20 has a moderate number of parasites in the skin, develops no or only small lesions and shows no signs of splenomegaly or hepatomegaly [80, 82]. To confirm the quality of the breeding colony the regions in CcS/Dem strains were genotyped using microsatellite markers (Generi Biotech, Hradec Králové, Czech Republic): D1Mit17, D2Mit52, D3Mit49,

D4Mit149, D5Mit55, D5Mit114, D8Mit125, D9Mit2, D10Mit46, D11Mit62, D16Mit7 as described in [84]. The regions in OcB/Dem strains were genotyped using microsatellite markers (Generi Biotech, Hradec Králové, Czech Republic): D3Mit160 and D3Mit17 as described in [82].

5.2.2.2. Impact of *Leishmania* infection on gut microbiome diversity

A total of 13,671,855 sequences were obtained from the samples of different mice. Most of these sequences (9,762,045) originated from the colon. The average sequence length was 260 base pairs. The analysis was performed independently for two groups of mouse strains, CcS/Dem and OcB/Dem, with ileum and colon samples treated separately.

The alpha diversity of samples from different parts of the gastrointestinal tract (ileum and colon) of the mouse strains was assessed to determine the bacterial diversity within each animal group (susceptible/resistant and infected/uninfected). The Shannon index, a commonly used measure to determine the diversity and abundance of species within a community, was used. The bacterial community was found to be most diverse in the colon and least diverse in the ileum (Figure 23).

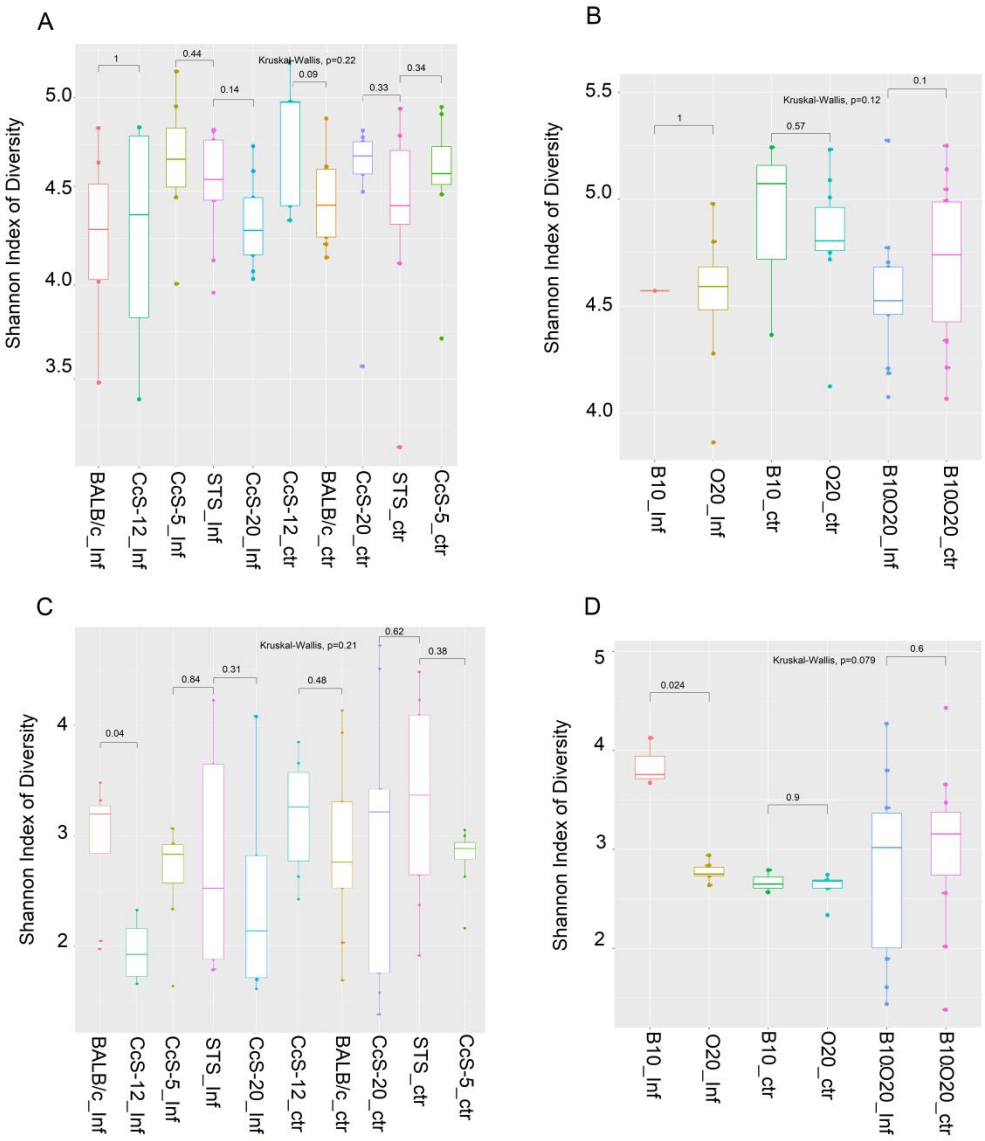

Figure 23: Microbiome alpha diversity (shannon index) among different mouse strains (susceptibles/ resistants) of infected or non-infected mice in: a) colon of CcS/Dem, b) colon of OcB/Dem, c) ileum of CcS/Dem and d) ileum of OcB/Dem. (p-value ≤ 0.05) was considered statistically significant. [Inf: infected, ctr: control (non-infected)] [86]

The diversity of the bacterial community was more uniform in the CcS/Dem series than in the OcB/Dem group. No significant differences (p>0.05) were observed in the CcS/Dem strains between infected and uninfected animals or between resistant and susceptible mice in both the ileum and colon microbiome (Figure 23a, c). In the OcB/Dem strains, a notable difference was found between infected and uninfected resistant mice in the microbiome of the colon (p=0.02) and ileum (p=0.01) (Figure 23b). The diversity of the infected mice was lower

than that of the uninfected mice in the colon of the OcB/Dem group (Figure 23b). No significant statistical differences were found in the colonic microbiome of all animals between the different strains of mice.

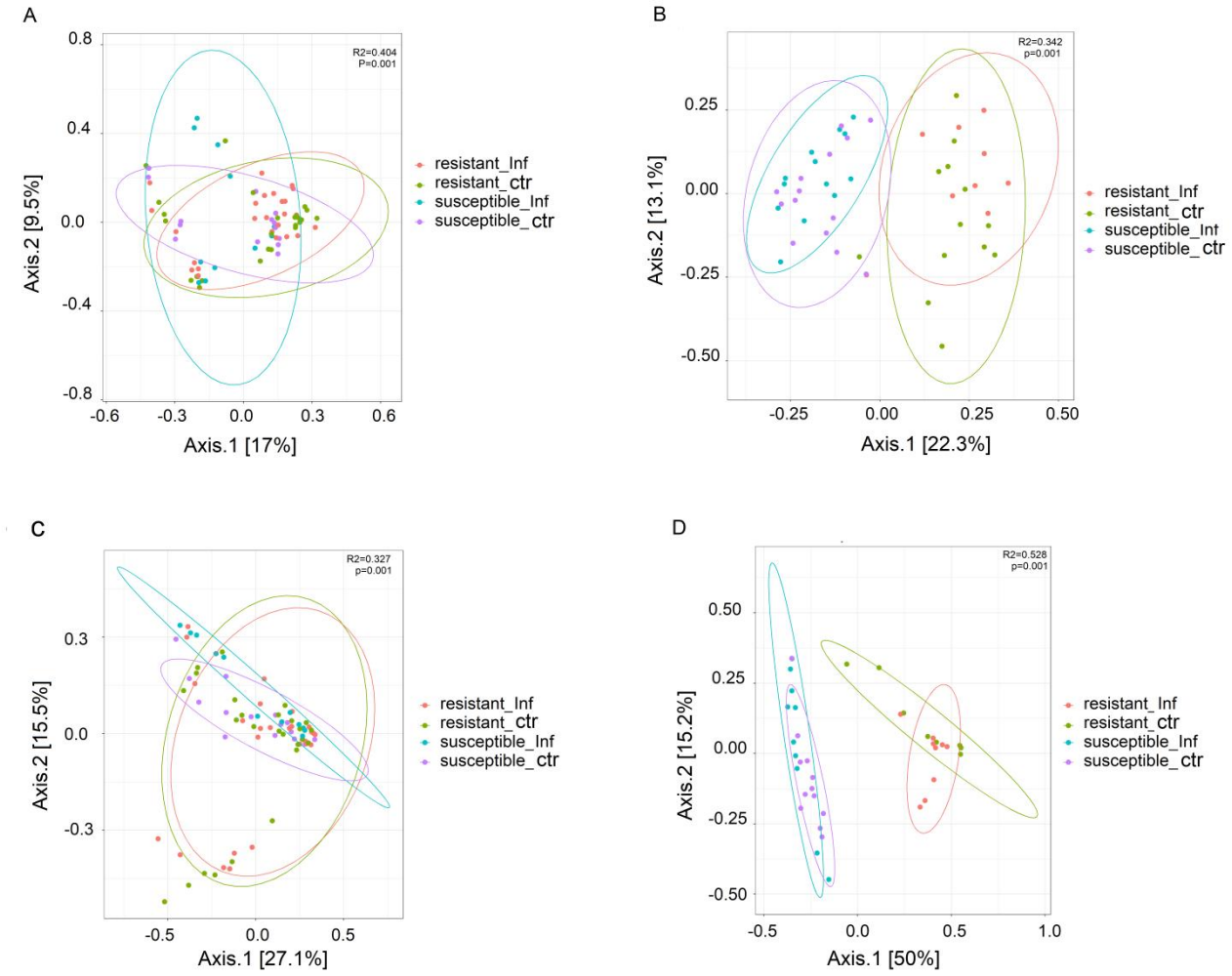

Figure 24: Principal coordinate analysis (PCoA) plots showing distinct clusters for different groups (susceptibles/ resistants) of infected or non-infected mice in: a) colon of CcS/Dem, b) colon of OcB/Dem, c) ileum of CcS/Dem and d) ileum of OcB/Dem. Ellipses mark 95% confidence ellipses around each group and (p-value $\leq$ 0.05) was considered statistically significant. [Resistant_Inf: infected resistant, Resistant_ctr: resistant control (non-infected), Susceptible_Inf: infected susceptible, Susceptible_ctr: susceptible control (non-infected)] [86]

The diversity in the ileum microbiome of the susceptible BALB/c strain was significantly greater (p=0.04) than in the infected susceptible CcS-12 strain. Diversity in the ileum microbiome was significantly greater (p=0.02) in the infected resistant B10 strain than in the infected resistant O20 strain.

To investigate the similarities between the different samples, beta diversity was used to independently analyze the composition of the bacterial communities in the colon and ileum. A principal coordinate analysis (PCoA) based on Bray-Curtis

distance was performed to compare the microbiome diversity between resistant and susceptible mice and between infected and uninfected mice of different strains (Figure 24).

Diversity of the colonic microbiome was found to be widely spread in samples from CcS/Dem strains (R2= 0.404, p=0.001) (Figure 24a). In the ileum microbiome of samples from CcS/Dem strains, beta diversity was more concentrated in the susceptible mouse strains and scattered in the resistant mouse strains (R2= 0.327, p=0.001) (Figure 24c). The beta diversity of the different strains from the CcS/Dem mouse series was examined individually (Figure 25).

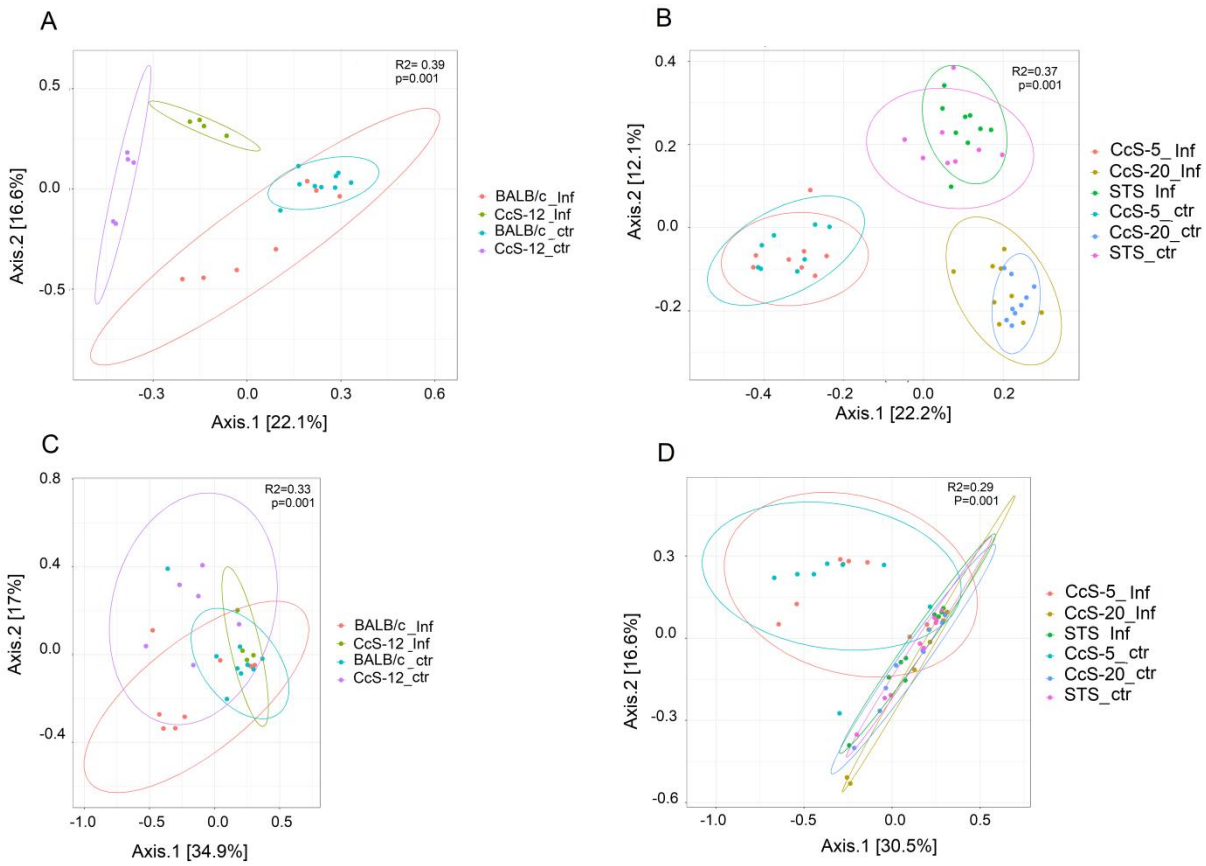

Figure 25: Principal coordinate analysis (PCoA) plots showed distinct clusters for infected or non-infected mouse strains of CcS/Dem group in: a) colon of susceptible CcS/Dem, b) colon of resistant CcS/Dem, c) ileum of susceptible CcS/Dem and d) ileum of resistant CcS/Dem. Ellipses mark 95% confidence ellipses around each group and (p-value ≤ 0.05) was considered statistically significant. [Inf: infected, ctr: control (non-infected)] [86]

In the colon microbiome, a larger significant difference was found between susceptible (R2= 0.39, p=0.001) and resistant mouse strains (R2= 0.37, p=0.001), with different clusters identified for different strains and separated from each other (Figure 25a, b). However, in the ileum microbiome, all resistant samples formed a cluster that was separated from the CcS-20 strain (R2= 0.29, p=0.001) (Figure 25d). The results showed that beta diversity was similar in infected and uninfected mice, with the exception of the samples of the susceptible CcS-12 strain, which were divided into an infected and uninfected group in both the colon and ileum microbiomes (R2= 0.33, p=0.001) (Figure 25a, c).

In addition, two main clusters were identified in the OcB/Dem mice in both the colon and ileum microbiome: one cluster for resistant mice and another cluster for susceptible mice (R2= 0.342, p=0.001) (Figure 24b, d). However, no significant differences were found between infected and uninfected mice, except in the ileum microbiome of the resistant strains, where the clusters of infected and uninfected mice were separated (R2= 0.528, p=0.001) (Figure 24d).

Colon microbiota

In general, the abundance of gut bacteria was different in the different mouse strains and varied. In the CcS/Dem mice, the microbiota of the colon samples of all animals was mainly composed of *Bacteroidetes* and *Firmicutes*. The presence of *Bacteroidetes* was slightly more pronounced in the infected susceptible strain CcS-12 (49.2 %), with *Muribaculaceae* being the most abundant family (39.5 %). In the infected intermediate strain CcS-20, *Muribaculaceae* was the dominant family (46.39 %). *Lachnospiraceae* were more prevalent in the infected resistant strain CcS-5 (31.9 %) than in the infected susceptible strain CcS-12 (10.3 %). However, in the colonic microbiome of the uninfected strain CcS-12, the *Lachnospiraceae* were the dominant bacterial family (59.1 %) (Figure 26a).

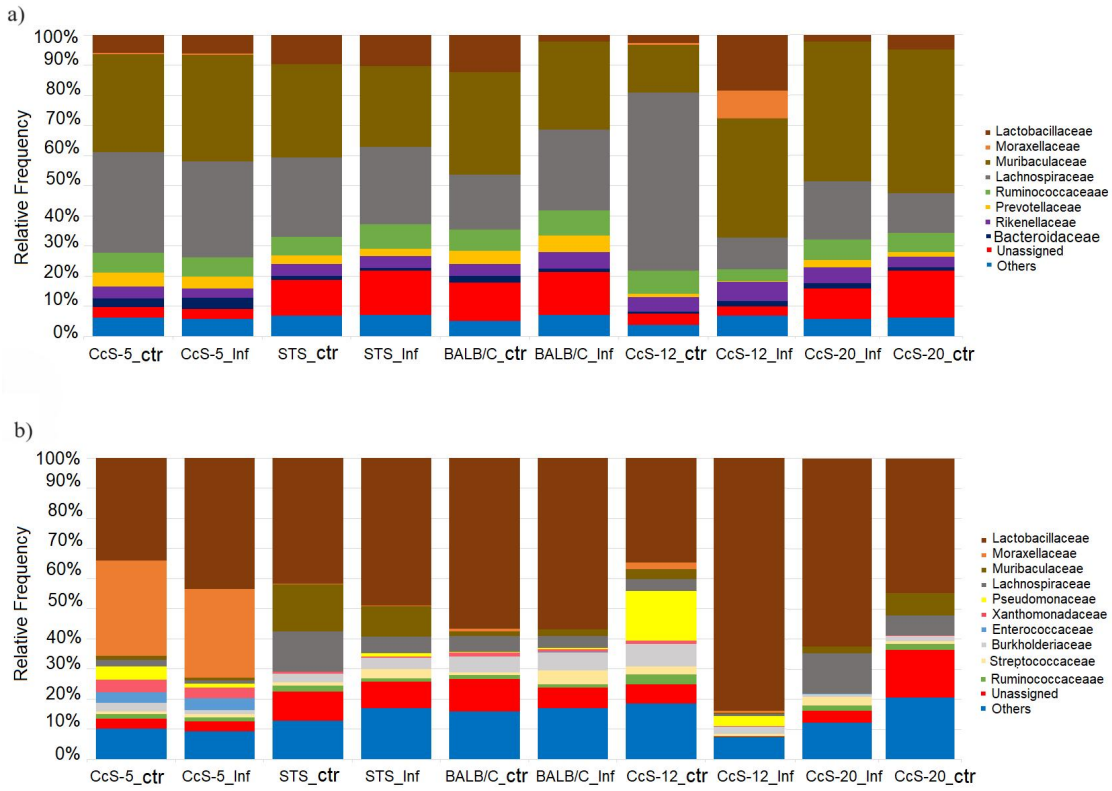

Figure 26: Relative abundance of the microbial population at the family level among different infected or non-infected CcS/Dem mouse strains in: a) colon, b) ileum. [Inf: infected, ctr: control (non-infected)] [86]

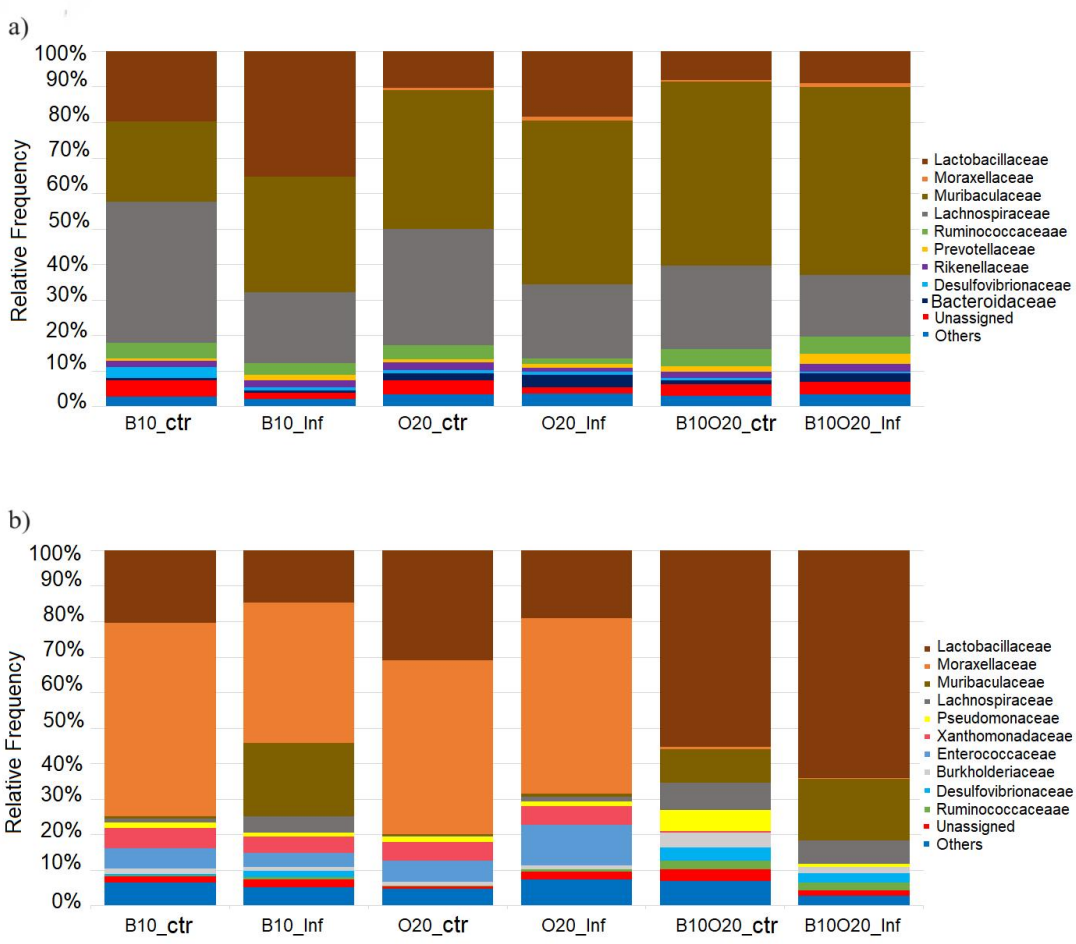

Figure 27: Relative abundance of the microbial population at the family level among different infected or non-infected OcB/Dem mouse strains in: a) colon, b) ileum. [Inf: infected, ctr: control (non-infected)] [86]

LEfSe analysis identified *Lactobacillaceae* as the major biomarker associated with the colonic microbiota of the infected resistant CcS/Dem mice, along with Family XIII (Clostridiales), *Peptococcaceae, Tannerellaceae* and *Burkholderiaceae.* In contrast, *Rikenellaceae, Deferribacteriaceae* and a family of uncultured *Firmicutes* bacteria were identified as biomarkers associated with the infected susceptible CcS/Dem mice in their colon (Figure 28, 29).

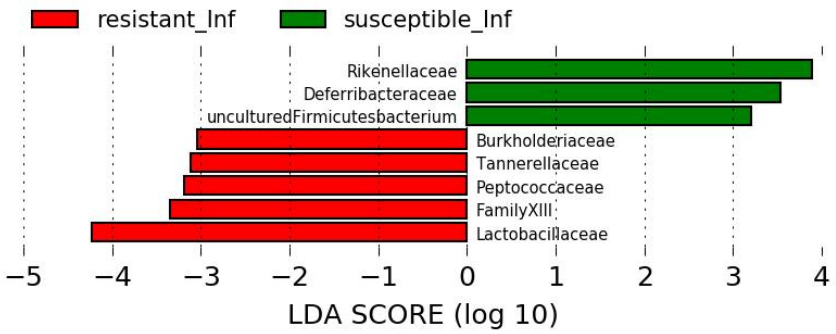

Figure 28: Linear discriminant analysis (LDA) effect size (LEfSe) of taxa at family level in infected susceptible CcS/Dem mice (in green) and infected resistant CcS/Dem mice (in red) in colon with alpha values of 0.05 and a threshold value of 2.0. [Resistant_Inf: infected resistant, Susceptible_Inf: infected susceptible] [86]

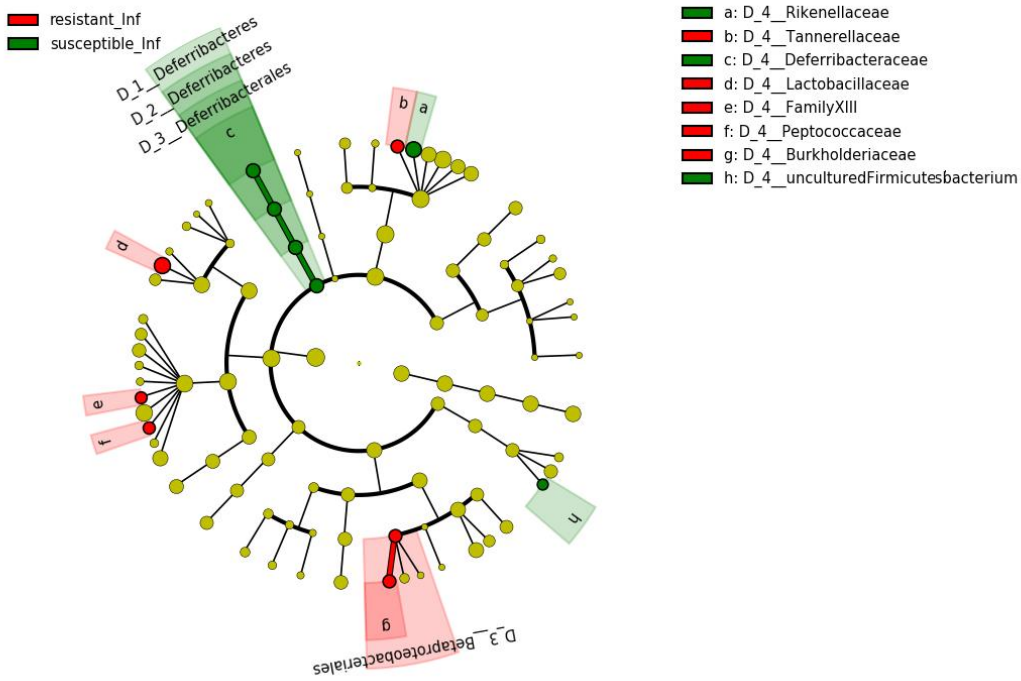

Figure 29: Cladogram of colon microbiota present a phylogenetic plot of LEfSe and taxa: D1 for Phylum level, D2 for Class level, D3 for Order level and D4 for Family level). [Resistant_Inf: infected resistant, Susceptible_Inf: infected susceptible] [86]

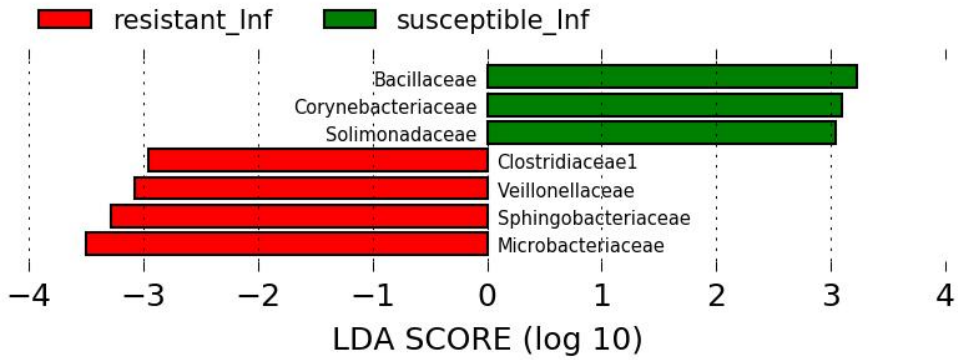

Figure 30: Linear discriminant analysis (LDA) effect size (LEfSe) of taxa at family level in infected susceptible CcS/Dem mice (in green) and infected resistant CcS/Dem mice (in red) in ileum with alpha values of 0.05 and a threshold value of 2.0. [Resistant_Inf: infected resistant, Susceptible_Inf: infected susceptible] [86]

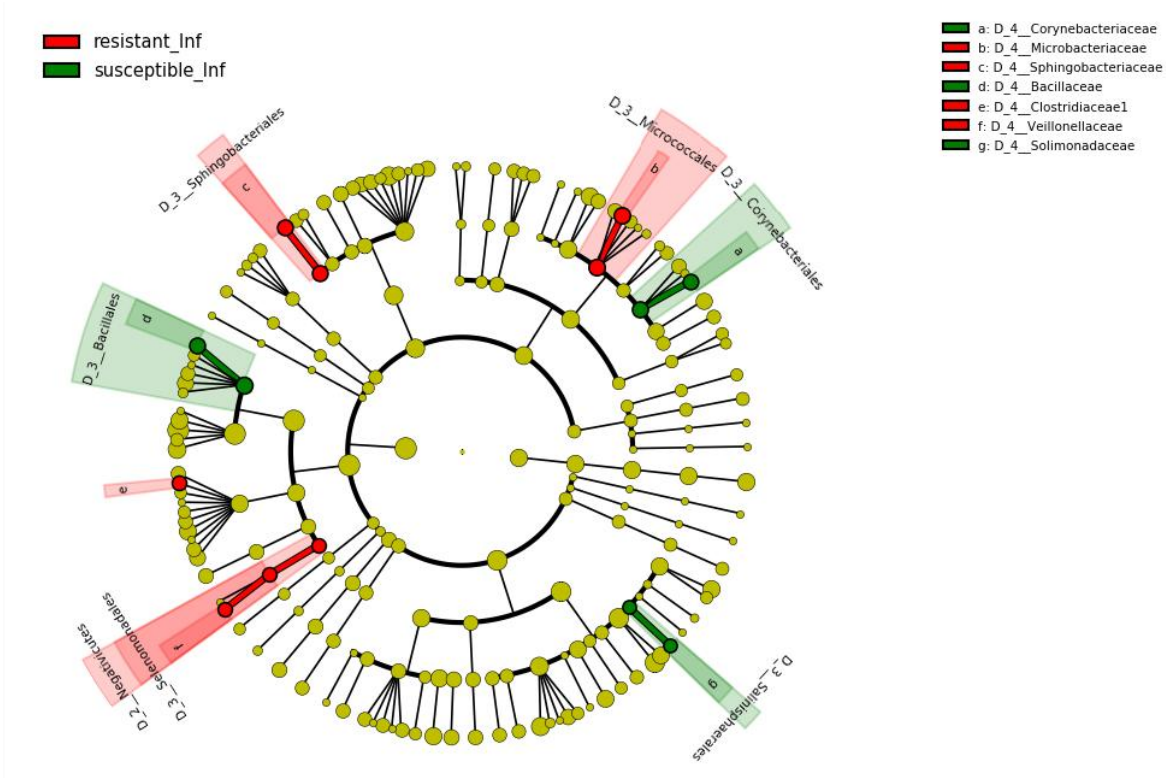

Figure 31: Cladogram of ileum microbiota present a phylogenetic plot of LEfSe and taxa: D1 for Phylum level, D2 for Class level, D3 for Order level and D4 for Family level). [Resistant_Inf: infected resistant, Susceptible_Inf: infected susceptible] [86]

In the OcB/Dem mouse strains (Figure 27a), the colon microbiota of all animals was mainly composed of the phyla *Bacteroidetes* and *Firmicutes*. *Bacteroidales* was the dominant order in all animals, but it was slightly more prevalent (61.8 %) in the infected susceptible strain B10.O20, where *Muribaculaceae* was the most abundant family (52.9 %). The relative abundance of *Lactobacillaceae* and *Lactobacillus* was higher in the infected resistant strain B10 (35.2 %) than in the infected susceptible strain B10.O20 (8.9 %). *Lachnospiraceae* was present in all animals, but was slightly higher in the uninfected resistant strain B10 (39.8 %) (Figure 27a).

The LEfSe results (Figure 28, 29) identified two bacterial families (*Lactobacillaceae* and *Clostridiaceae*) associated with the infected resistant OcB/Dem mouse group, while *Ruminococcaceae*, *Rikenellaceae*, *Moraxellaceae*, *Tennerellaceae* and *Christensenellaceae* were associated with the infected susceptible strain B10.O20.

Ileum microbiota

The ileum microbiota of all mice consisted mainly of the phyla *Firmicutes*, *Proteobacteria* and *Bacteroidetes*. In the CcS/Dem mice (Figure 26b), *Firmicutes* was the predominant phylum in the ileum microbiota of the infection-resistant animals, with a frequency of 52.3 % in the strain CcS-5 and 61.4 % in the strain STS. *Firmicutes* was the most prevalent and dominant phylum in the ileum microbiota of infected susceptible mice, with frequencies of 86.3 % in the strain CcS-12, 69.7 % in the strain BALB/c and 81.2 % in the intermediate strain CcS-20. In the ileum microbiota of the infection-resistant strain CcS-5, *Proteobacteria* were most frequently represented at 38.2 %. In the ileum microbiota of the infection-resistant strain STS, *Bacteroidetes* were more frequently represented with a proportion of 12.9 %.

In the microbiota of the ileum of the infected resistant strain CcS-5, gammaproteobacteria were the most prevalent at 36.1%. Within this bacterial class, the family *Moraxellaceae* and the genus *Acinetobacter* were the most frequently represented with 29.5 %. In the ileum microbiome of the resistant strain STS, however, the *Moraxellaceae* family and the *Acinetobacter* genus were less frequently represented. In the infected resistant mice, the family *Lactobacillaceae* and the genus

Lactobacillus were most abundant, with a frequency of 43.4 % for the CcS-5 strain and 49.0 % for the strain STS. In susceptible and intermediate infected animals, the family *Lactobacillaceae* and the genus *Lactobacillus* were more prevalent and dominant with a frequency of 83.8% for the strain CcS-12, 56.9% for the strain BALB/c and 62.36% for the strain CcS-20. In the resistant STS mice, the *Muribaculaceae* family was more prevalent with a frequency of 10.1 % in the ileum microbiota of infected STS mice and 15.7 % in the ileum microbiota of uninfected mice (Figure 26a).

LEfSe analysis (Figure 30, 31) identified four bacterial families associated with resistance in the ileum microbiota of infected CcS/Dem mice: *Microbacteriaceae*, *Sphingobacteriaceae*, *Veillonellaceae* and *Clostridiaceae1*. In contrast, three bacterial families (*Bacillaceae*, *Corynebacteriaceae* and *Solimonadaceae*) were associated with susceptibility in the ileum microbiota of infected susceptible CcS/Dem mice.

In the OcB/Dem mouse group (Figure 27b), the ileum microbiota of infected resistant animals was dominated by *Proteobacteria*, with relative abundances of 48.3% in the B10 strain and 57.8% in the O20 strain. *Gammaproteobacteria* was the predominant class in the ileum microbiota of resistant strains, represented by 39.5% *Moraxellaceae* and *Acinetobacter* in infected B10 mice and 49.4% in infected O20 mice (Figure 27b). *Bacteroidetes* was more prevalent in infected resistant B10 mice (22.1%) compared to uninfected B10 mice (2.6%) and infected resistant O20 mice (3.2%).

In infected susceptible B10.O20 mice, the microbiota of the ileum consisted predominantly of *Firmicutes*, with relative abundances of 66.4% in uninfected mice and 73.1% in infected mice. *Lactobacillaceae* was the most abundant family (63.9%) in the ileum microbiota of infected susceptible B10.O20 mice, represented by the genus *Lactobacillus*. In the ileum microbiota of infected susceptible B10.O20 mice, *Bacteroidetes* was more common (17.8%) and *Proteobacteria* less common (6.1%). *Muribaculaceae* was more prevalent in the ileum microbiota of infected susceptible mice (17.4%) and infected resistant B10 mice (20.52%) compared to infected resistant O20 mice (0.7%) (Figure 27b).

The LEfSe analysis (Figure 32, 33) of the ileum samples revealed that several bacterial families were linked to the infected resistant OcB/Dem mouse group, including *Moraxellaceae*, *Veillonellaceae*, *Family XI (Clostridiales)*, *Enterococcaceae*, *Leuconostocaceae*, *Xanthomonadaceae*, *Deferribacteriaceae*,

Tannerellaceae, *Beijerinckiaceae*, *Sphingomonadaceae*, *Clostridiaceae1*, *Sphingobacteriaceae*, *Micrococcaceae*, *Prevotellaceae*, *Microbacteriaceae*, *Weeksellaceae*, *Streptococcaceae*, *Coriobacteriaceae*, *Enterobacteriaceae*, *Propionibacteriaceae*, and *Bacteroidaceae*. In contrast, the ileum microbiota of the infected susceptible strain B10.O20 was associated with *Lactobacillaceae*, *Caulobacteraceae* and *Eggerthellaceae*.

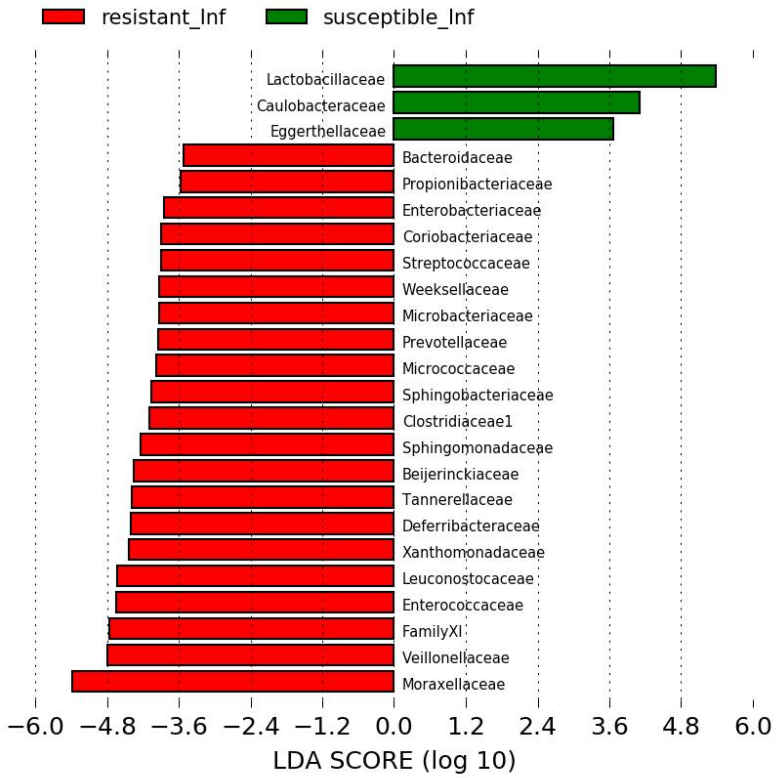

Figure 32: Linear discriminant analysis (LDA) effect size (LEfSe) of taxa at family level in infected susceptible OcB/Dem mice (in green) and infected resistant OcB/Dem mice (in red) in ileum with alpha values of 0.05 and a threshold value of 2.0. [Resistant_inf: infected resistant, Susceptible_inf: infected susceptible] [86]

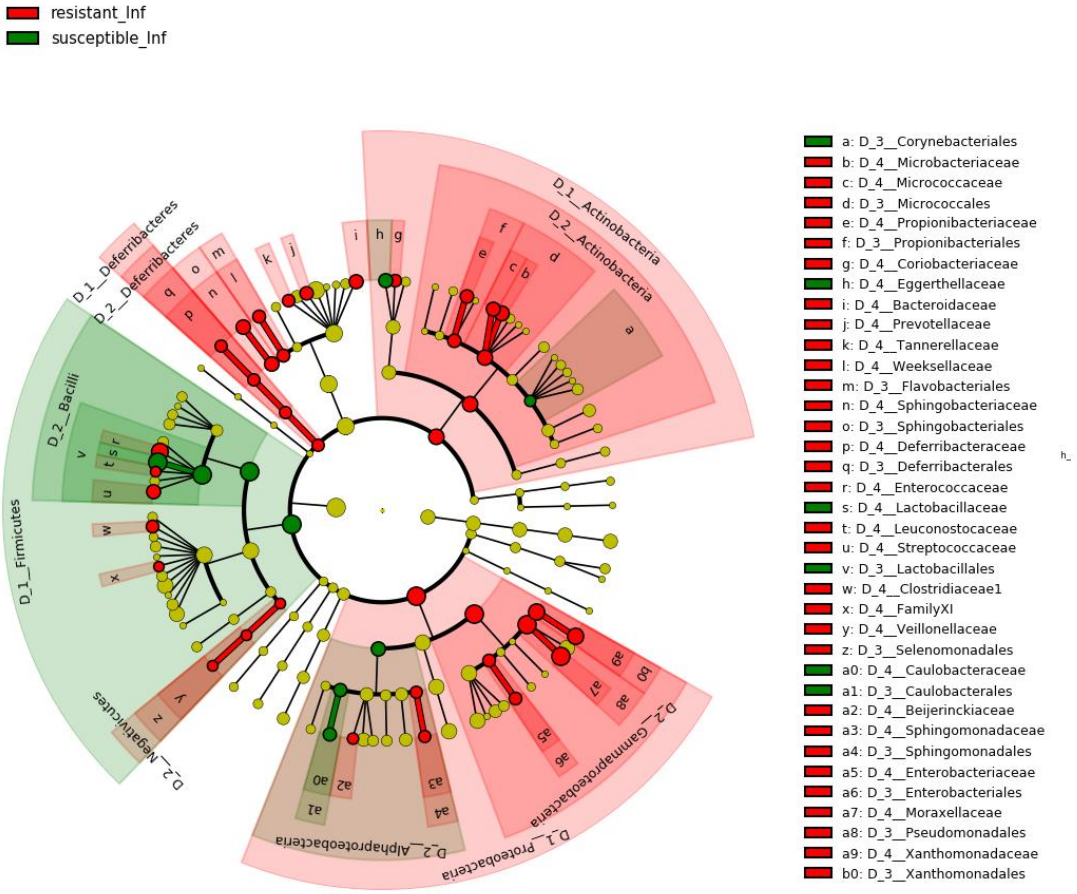

Figure 33: Cladogram of ileum microbiota present a phylogenetic plot of LEfSe and taxa: D1 for Phylum level, D2 for Class level, D3 for Order level and D4 for Family level). [Resistant_inf: infected resistant, Susceptible_inf: infected susceptible] [86]

5.2.2.3. Impact of host genetics on gut microbiome of non-infected mice

Analysis of alpha diversity revealed that there was no significant difference between the different uninfected mouse strains used in this study. However, Bray-Curtis dissimilarity showed a significant correlation with host genetics ($p = 0.001$) (Figure 24). The relative abundance of taxa at different bacterial taxonomic levels varied greatly between the different mouse strains (Figure 25).

5.2.2.4. Metagenomic functional prediction

Functional inference analyses using PICRUSt were conducted separately for the colonic and ileal microbiota to identify significant differences

in the abundance of potential functions between the two groups of infected mice: resistant and susceptible. Histograms were created to represent the relative abundance of functions. The results indicated that the functional genes in both groups were mainly associated with “cellular processes", “environmental information processing", “genetic information processing", “human diseases", “metabolism” and “organismal systems” at KEGG level 1.

Summarizing the data at KEGG level 2, the most significant differences between the two groups were observed in the ileum samples of the OcB/Dem strains. Thirty-one pathways showed significantly different abundances, indicating different functional profiles between the infected resistant and susceptible strains. Functional prediction revealed that the ileum microbiota of infected resistant OcB/Dem strains was primarily associated with “xenobiotic degradation and metabolism", “amino acid metabolism", “lipid metabolism” and “cellular processes and signaling”. In contrast, the ileum microbiota of infected susceptible OcB/Dem strains was mainly associated with “replication and repair", “translation", “nucleotide metabolism", “carbohydrate metabolism", “transcription", “processing of genetic information", “energy metabolism", “metabolism of cofactors and vitamins” and “enzyme families”. In the colon microbiome of the OcB/Dem strains, only nine KEGG pathways showed significant differences between the two groups. Most of the predicted functions were significantly more abundant in the infected susceptible B10.O20 mice, including “glycan biosynthesis and metabolism", “transport and catabolism", “biosynthesis of other secondary metabolites", “metabolism", “enzyme families", “endocrine system", “neurodegenerative diseases” and “digestive system”. The only significantly abundant function associated with the colonic microbiome of resistant OcB/Dem strains was “environmental adaptation”.

Functional prediction for the ileum microbiome of CcS/Dem strains identified eight significantly abundant KEGG pathways. In the ileum microbiome of infected susceptible CcS/Dem mice, “membrane transport” and “carbohydrate metabolism” were significantly more abundant. In contrast, the ileum microbiome of infected resistant CcS/Dem mice showed a significant abundance of “amino acid metabolism", “poorly characterised” functions, “glycan biosynthesis and metabolism” and “metabolism of terpenoids and polyketides”. In the colon microbiome of the CcS/Dem strains, only three KEGG pathways were significantly different. These pathways - "poorly characterised” functions, “metabolism”

and “neurodegenerative diseases" - were significantly more prevalent in the colonic microbiome of infected susceptible CcS/Dem strains.

PCA analysis showed that the functions of the ileal microbiota of the same group of mice (infected resistant or infected susceptible) were clustered together. In the ileum of the OcB/Dem mouse strains, the two clusters were more distinct, with the first two components explaining 92.7% of the variation. This indicates that the functional KOs varied mainly in the microbiota of the ileum. There was no significant difference between the two groups of mice (infected resistant and infected susceptible) in the predicted functions of the colonic microbiota in both strains.

5. 3. Mapping and identification of candidate genes

5. 3. 1. Mapping genes controlling susceptibility to TBEV

For the gene mapping study, an animal model consisting of three mouse strains was used: the parental strains BALB/c and STS and the recombinant congenic strain CcS-11, which carries about 12.5% genes from the donor strain STS and about 87.5% genes from the background strain BALB/c [124, 125]. We have previously found that BALB/c mice exhibit intermediate susceptibility to TBE infection, whereas STS mice are highly resistant, while the recombinant congenic strain CcS-11 is even more susceptible than BALB/c [77]. Strain CcS-11 differs from BALB/c at STS-derived regions on eight chromosomes (precisely: chromosomes 1, 3, 7, 8, 10, 12, 16 and 19) [85]. These different regions were genotyped in the F_2 hybrid mice between CcS-11 and BALB/c infected with TBEV using 16 microsatellite markers (Generi Biotech, Hradec Králové, Czech Republic): D1Mit403, D3Mit45, D7Mit25, D7Nds5, D7Mit18, D7Nds1, D7Mit282, D7Mit259, D8Mit85, D10Mit12, D10Mit46, D12Mit37, D16Mit73, D19Mit51, D19Mit60, D19Mit46 as described in [85]. The linkage with survival was tested by binary feature interval mapping.

5. 3. 2. A novel locus on mouse chromosome 7

Linkage analysis revealed a novel suggestive survival-controlling locus on chromosome 7 linked to the marker D7Nds5 (44.2 Mb) (Figure 34).

We sequenced the genomes of BALB/c and STS by next-generation sequencing and performed a bioinformatic analysis of the chromosomal segment linked to TBEV survival.

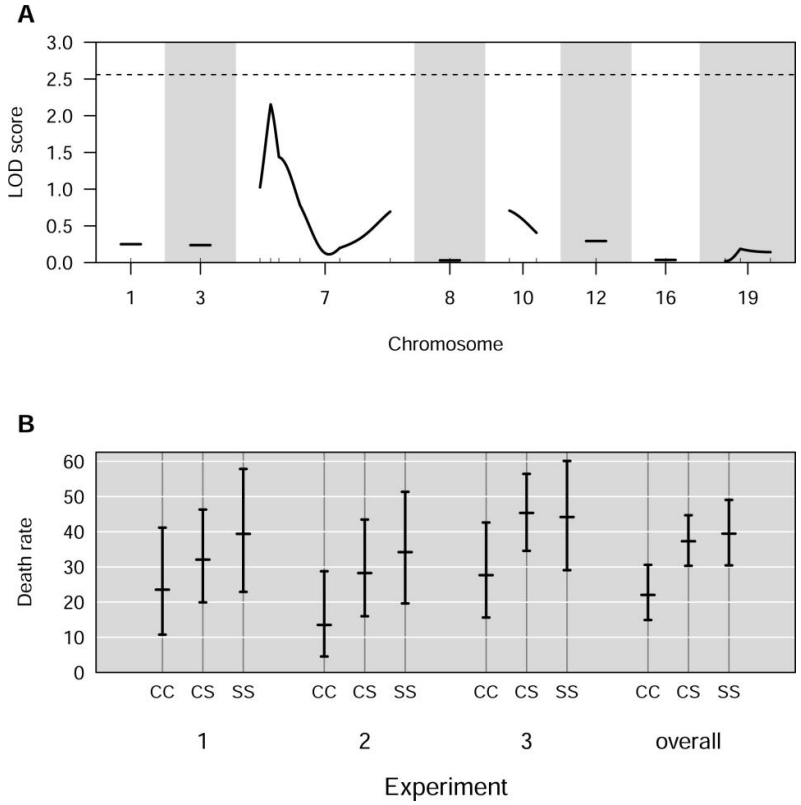

Figure 34: Genetic influence on susceptibility to TBEV in an F_2 intercross between BALB/c and CcS-11. A) LOD curves from binary trait interval mapping for death/survival. A dashed horizontal line is plotted at the 5% significance threshold, adjusting for the genome scan. B) A plot of the death rate as a function of genotype at marker D7Nds5 and experiment, with 95% confidence intervals. C and S indicate the presence of BALB/c and STS allele, respectively. The S allele is associated with a higher death rate [53]

Analysis of this locus for polymorphisms between BALB/c and STS that change RNA stability and genes' functions led to the detection of 9 potential candidate genes: *Cd33* (CD33 antigen), *Klk1b22* (kallikrein 1-related peptidase b22), *Siglece* (sialic acid binding Ig-like lectin E), *Klk1b16* (kallikrein 1-related peptidase b16), *Fut2* (fucosyltransferase 2), *Grwd1* (glutamate-rich WD repeat containing 1), *Abcc6* (ATP-binding cassette, sub-family C (CFTR/MRP), member 6), *Otog* (otogelin), and *Mkrn3* (makorin, ring finger protein, 3). One of them *Cd33*, carried a nonsense mutation in the strain STS (Table 2).

Table 2. List of candidate genes in TBEV susceptibility locus. The table shows differences between BALB/c and STS in DNA and protein sequences in potential candidate genes. The table shows also sequences of the reference mouse strain C57BL/6 [53]

Position Bp	Reference genotype C57BL/6	Genotype BALB/c	Genotype STS	Protein position of amino acid	Reference amino acid	Alteration	Type of change	Gene symbol	Transcription status	Gene name	Gene ID: MGI	Gene ID: NCBI
43,528,893	C/C	C/C	T/T	353	G	K	Single AA Change	***Cd33***	KNOWN	CD33 antigen	99440	12489
43,532,167	G/G	G/G	A/A	190	R	*	Nonsense Mutation	***Cd33***	KNOWN	CD33 antigen	99440	12489
43,659,827	G/G	G/G	T/T	102	D	E	Single AA Change	***Siglece***	KNOWN	sialic acid binding Ig-like lectin E	1932475	83382
44,115,970	A/A	A/A	C/A	115	L	Y	Single AA Change	***Klk1b22***	KNOWN	kallikrein 1-related peptidase b22	95291	13646
44,140,534	G/G	G/G	C/C	76	G	A	Single AA Change	***Klk1b16***	KNOWN	kallikrein 1-related peptidase b16	891982	16615
45,650,779	G/G	G/G	A/A	190	R	W	Single AA Change	***Fut2***	KNOWN	fucosyltransferase 2	109374	14344
45,830,054	CTCTTCA/ CTCTTCA	C/C	CTCTTCA/CTC TTCA	129	ED	.	Deletion	***Grwd1***	KNOWN	glutamate-rich WD repeat containing 1	2141989	101612
45,977,290	C/C	A/A	C/C	1448	V	L	Single AA Change	***Abcc6***	KNOWN	ATP-binding cassette, sub-family C (CFTR/MRP), member 6	1351634	27421
46,262,804	C/C	C/C	T/T	748	R	W	Single AA Change	***Otog***	KNOWN	otogelin	1202064	18419
62,419,214	C/C	CGGCATTGG CACT/CGGCA TTGGCACT	C/C	275	P	PVPMP	Insertion	***Mkrn3***	KNOWN	makorin, ring finger protein, 3	2181178	22652

6. Discussion

6. 1. Genotyping

We can now routinely genotype polymorphisms and mutations and link them to specific traits. Genotyping identifies genetic variants that lead to specific phenotypes, such as disease manifestations or drug responses. In addition, there is a wide range of applications [80, 127]. The genetic information required for large-scale characterization or discovery studies differs significantly from that required for targeted studies of specific genetic variants, such as genetic typing. When selecting a genotyping method, consider the number of markers, the precision of the genetic information required, the number of individuals to be typed, the capacity for pre- and post-processing, analysis and financial constraints (see table 3 and 4). Sequencing provides the most detailed genetic information, but is more expensive and computationally intensive. In contrast, low-resolution genotyping methods are easier to perform and less costly. The method chosen should strike a balance between coverage and detail to fulfil the purpose of the study with regard of available resources [21].

According to publications Wu (2019), Meng (2021), Wang (2020), Zhao (2019) and Kim (2022), the presented genotyping method in combination with enzymatic cleavage could also be used to detect the influence of a genetic polymorphism on the risk of complications after the use of mechanical lung ventilation [128, 129, 130, 131, 132]. A list of the specific SNPs used for this purpose can be found in the articles.

Table 3. Comparison of the used genotyping methods

	Genotyping of STR	**Sanger sequencing**	**Ion Torrent PGM**	**HiSeq 2500**
Principle	length diference (presence-absence of STR)	chain terminating sequencing	semiconductor sequencing	sequencing by synthesis
Number of samples per run/reaction	182	1	96	depends on size of cartridge and data trouhgput e.g. 150 human exomes or 8 genomes
Number of markers	1	1	96	96
Resolution of method	6 bp length difference	1 base identification	1 base identification	1 base identification
DNA quality	poor- high	High	high	high with no fragmentation
Time required	4 hours	24 hours	2-10 hours	1-6 days
Applications	allele inheritance detection, forensic fingerprinting, gene mapping, point mutation detection, marker-assisted breeding of congenic mouse strains	single gene identification, validation of NGS data, genotyping of microsatelite markers, identification plasmids,	viral and bacterial typing, targeted sequencing, microbial research, exome sequencing	de novo sequencing, whole genome sequencing, exome sequencing, transcriptome sequencing, metagenomics, targeted gene sequencing
Relative Price per Sample	low cost	1-20 targets low, more than 20 targets high	high cost	very high cost
Relative Instrument Price	low cost	low cost	intermediate cost	intermediate cost
Total cost	low	intermediate cost	high	very high

Table 4. Comparison of sequencing used

	Ion Torrent PGM	HiSeq 2500	Sanger sequencing
Sequencing principle	Semiconductor sequencing	Sequencing by synthesis	Chain terminating dideoxynucleotides
Read lenght (bp)	2x200, 2x 400	2x75, 2x100	400 - 1100
Accuracy (%)	98	98	99.99
Time per run	Depends on size of chip used 2-8 hours	Depends on size of cartridge used 24-56 hours	24 hours
Data format	BAM, FASTQ, SFF, VCF	FASTQ	FASTA, .ab1, .phd.1
Advantages	Constant quality during whole read lenght, Long read lenght, Fast	High accuracy in reading DNA, Fast processing time, Provides detailed genetic information	High accuracy ("golden standard"), Cost-effective when sequencing 1-20 targets,
Disadvantages	Read of homopolymers (problem with 5 and more homopolymers), Low read output capacity, Requires high-quality DNA	High cost of equipment and materials, Short read length, Requires high-quality DNA, Time-consuming process, Potential for sequencing errors	Sequencing one fragment in a time, The quality of sequence degrades after 700 bp, Low sensitivity (limit of detection ~15–20%), Low throughput, Not cost-effective for >20 targets

6. 2. Microbiome

The result of the complex comparative study demonstrated that standard methods used in pilot study produce incomparably less information about microbiome composition than methods used in main study. In the pilot study, an older and less expensive method was used to study the microbiome, while in the main study a new, more expensive and more detailed method using next-generation sequencing (NGS) was used. The accuracy of the results in the individual microbiome experiments depends on the methods used, so the introduction and utilization of new and modern technologies in applied research is crucial. Although both methods lead to a bacterial profile of the digestive tract of the examined individuals, the new method provides incomparably more information, which leads to more accurate interpretations and a better understanding of the issue. The more information we obtain about the sample under investigation, the more targeted the therapy can be.

6. 2. 1. Pilot study

We focused on the analysis of microbial populations in healthy subjects and individuals infected with protozoan *Leishmania major*. In the first part of our study, the effect of *Leishmania* infection on bacterial populations in the small and large intestines was investigated in 3 mouse strains (STS, BALB/c, CcS-20). These strains are particularly interesting due to the extensively studied manifestations of various disease caused by the parasite *L. major* [80, 82]. With the susceptible strain BALB/c, there were no changes in the microbiome diversity of the ileum between infected and uninfected mice (Figure 19a, b) with the *Lactobacillales* order of bacteria predominating, while for the two other mice strains, we observed a decrease in *Bacteroidales* and *Clostridiales* levels and increase in *Lactobacillales* levels in the infected mice (Figure 17a). The difference in the number of OTU was not significant for the BALB/c strain (Figure 18a).

However, in the colon of the susceptible strain, we observed a large change in microbiome diversity (Figure 17b, 20b) with a drop in *Lactobacillales* level and increase in *Bacteroidales* and *Clostridiales* levels after infection with *L. major*. The same was observed for the CcS-20 strain, while with the resistant strain STS, an increase in *Lactobacillales* level and reduction in *Clostridiales* and *Bacteroidales* levels were noted (Figure 18b). A significant difference in the number of OTU was observed in the BALB/c strain (Figure 18b. Also, symptoms of *Leishmania* development were proved in the infected BALB/c strain large lesions were observed (see Table 1.). However, the resistant strains, none or very small lesions observed. Other pathophysiological parameters were monitored and examined, but the results from the analysis are not available yet.

In conclusion, the susceptibility is correlated with the abundance of *Clostridiales* and *Bacteriodales* orders, which was also found by [74], while the resistance was correlated with the relative abundance of the *Lactobacillales* order. Infection by *L. major* affects the microbiota of the two different regions of the intestine (colon and ileum) differently. The ileum region seems to be more resistant because *Lactobacillales* dominate this region of the intestine, for this reason, the susceptible strains have no changes in diversity. In contrast, the colon is characterized

by the presence of *Clostridiales* and *Bacteriodales*, which are obligate anaerobic bacteria, and in this region, the level of *Lactobacillales* is lower. For this reason, the strains BALB/c and CcS-20 were affected by *L. major* infection.

In summary, this study has shown that *L. major* infection causes a modification in the gut microbiome of the host, but this modification is poorly understood. It is clear that there are interactions between microbiota, the parasite and cells of the immune system, but the mechanisms have yet to be elucidated.

6. 2. 2. Main study

Different bacterial communities colonize different sections of the gastrointestinal tract of mice [133, 134, 135]. Oliveira (1999) demonstrated the crucial role of intestinal microbiota in effective resistance to *L. major* infection and found that germ-free mice infected with *L. major* developed significantly larger lesions compared to conventional controls [136]. Therefore, this study investigates the microbiome and bacterial diversity in the ileum and colon of two groups of mouse models of leishmaniasis using a next-generation sequencing approach.

The data indicated that the intestinal microbiota of OcB/Dem mice exhibited greater variability after infection than the more stable microbiota of CcS/Dem strains (Figure 23, 24, 25). The bacterial composition in the gastrointestinal tract of the two groups of mice was diverse and differed at various taxonomic levels. Three major phyla were predominant in the gut of all mouse strains: *Firmicutes*, *Bacteroidetes* and *Proteobacteria*, which is consistent with previous studies [133, 135].

In resistant strains B10 and O20, *Proteobacteria* were the dominant phylum in the ileum microbiota. *Firmicutes* was the most abundant phylum in the ileum of the susceptible strain B10.O20 and in the colon of the resistant strains B10 and O20 as well as in the ileum and colon of all CcS/Dem strains. The *Bacteroidetes* content was higher in the colon of the infected susceptible CcS/Dem strains and dominated in the colon of the susceptible strain B10.O20 and the infected resistant strains B10 and O20. The ileum was the most affected part of the gastrointestinal tract (Figure 24, 25), while the microbiota remained more stable in the colon. This stability in the large intestine compared to the small intestine has already been observed in several studies [133, 134, 135]. Furthermore, no significant difference in the diversity of the human

stool microbiome was observed between cases of visceral leishmaniasis (VL) and endemic controls (EC) [70].

Using LEfSe analysis, several biomarkers were identified in the gut microbiota of both the infected resistant and susceptible mice (Figure 26, 27, 28, 29, 30, 31). Common biomarkers in the ileum microbiota of infected resistant mice included *Clostridiaceae1*, *Sphingobacteriaceae*, *Veillonellaceae* and *Microbacteriaceae*. *Moraxellaceae*, a bacterial family from the *Gammaproteobacteria* class, was the predominant biomarker in the ileum microbiota of infected resistant OcB/Dem strains, but was not identified in infected resistant CcS/Dem strains. Consequently, the class of *Gammaproteobacteria* has only been associated with resistance to *L. major* infection in C57BL/6 [74] and O20 strains (this work). *Bacilli* were linked to susceptibility in the ileum microbiota of infected OcB/Dem mouse strains. Similarly, a study of the fecal microbiota of two experimental models of *L. major* infection, BALB/c and C57BL/6, found that *Gammaproteobacteria* were strongly associated with the self-healing C57BL/6 strain, while the *Bacilli* class was associated with the non-healing phenotype in BALB/c [74]. *Rikenellaceae* were identified as a common biomarker in the colon microbiota of all infected susceptible mouse strains. *Tannerellaceae* was linked to resistance to leishmaniasis in the colon microbiota of infected resistant CcS/Dem mice and in the ileum microbiota of infected OcB/Dem strains. However, it was also significantly abundant in the colonic microbiota of infected susceptible OcB/Dem strains. *Clostridiaceae* was conspicuously abundant in the colonic microbiota of infected resistant OcB/Dem strains and in the ileum microbiota of all infected resistant strains. The *Clostridia* class was associated with the non-healing phenotype [74]. *Lactobacillaceae* was correlated with resistance in the colonic microbiota of all infected resistant mouse strains, whereas it was significantly abundant in the ileum microbiota of infected susceptible OcB/Dem strains. These divergent results are due to differences in the diversity of the fecal microbiota and the microbiota of the small and large intestine.

Although parasites of the genus *Leishmania* (*infantum, braziliensis, donovani*) have been detected in the gastrointestinal tract of rodents [137, 138, 139, 140, 141, 142], there are no reports on the detection of *Leishmania major* in the digestive tract of mice, and the analysis of parasitic spread in the ileum or colon was beyond the scope of our study. However, we cannot exclude the possibility that *L. major* parasites could be present in the ileum or colon and influence the composition of the microbiome.

PICRUSt was used to analyze the functional composition and identify differences in the gut microbiota of infected resistant and susceptible mice. Cluster analysis revealed significant differences between the ileum and colon. In particular, the microbiota of the ileum showed significant differences between the clusters of infected resistant and susceptible mice, suggesting that the ileum is the part of the intestine most affected by *Leishmania* infection. Functional prediction analysis indicates a great diversity in the gut microbiota, with numerous microbial functional genes related to “metabolism”, “environmental information processing”, “genetic information processing”, “human diseases” and “cellular processes".

The microbiome of the ileum of resistant OcB/Dem strains was found to be significantly enriched with genes related to the biodegradation and metabolism of xenobiotics, in particular the metabolism of xenobiotics by cytochrome P450 (KEGG level 3). The gut microbiota plays a crucial role in xenobiotic metabolism [143, 144, 145, 146]. The gut microbiome regulates the metabolic outcomes of xenobiotics and host gene expression of CYP450s [146]. Experimental studies have shown that leishmaniasis affects drug metabolism by decreasing cytochrome P450 (CYP) levels. An early study suggested that L. donovani infection affects xenobiotic-metabolising enzymes in the liver of mice [147]. It was reported that the phenotypic activities of CYP3A4 and CYP2C19 were significantly reduced in Brazilian patients during the acute phase of visceral leishmaniasis [148]. Reduced function of xenobiotic metabolism leads to altered drug clearance rates, which can have serious consequences for individuals infected with *Leishmania*. Genes related to amino acid metabolism were upregulated in the ileum microbiome of infected resistant mice in all model strains. Amino acids such as arginine, asparagine and tryptophan are considered mediators of metabolic cross-talk between host and pathogen [149]. The administration of glutamine to mice infected with *L. donovani* may act as a promising adjuvant during miltefosine treatment and improve the anti-*Leishmania* immune response by significantly reducing the parasite load [150]. The L-arginine pathway is essential for the regulation of iNOS-mediated parasite killing and polyamine-mediated parasite growth. In addition, amino acids play a role in the regulation of immune cells during leishmaniasis [151].

The gut microbiome of infected susceptible mice showed significantly higher abundance of genes related to “carbohydrate metabolism” and “glycan biosynthesis and metabolism”. *L. major* promastigotes were found to stimulate macrophages to enhance anaerobic glycolysis and lead to cholesterol accumulation [152].

Bodhale (2018) observed a correlation between cytokine secretion profiles, *Leishmania* susceptibility and the expression of various enzymes of the glycolytic pathway in the spleen of *L. donovani* infected mice [153]. In addition, a recent study suggests that diet-induced obesity reduces the resistance of C57BL/6 mice to infection with *L. major* [154].

This study investigated the differences in cutaneous leishmaniasis in the colon and ileum using different experimental mouse models. The experimental models differ in genetic background to identify the genes involved in the pathophysiology of the disease. The bacterial composition and diversity of the intestinal microbiota was analyzed in both susceptible and resistant leishmaniasis mouse models. Infection with *L. major* significantly altered the composition of the microbiome in the ileum, whereas no significant changes were observed in the microbiome of the colon. The results also suggest that host genetics play a crucial role in shaping and modulating the composition of the gut microbiome. In addition, significant differences in the beta diversity of the colon microbiome were observed in all strains of mouse.

The study identified biomarkers in the gut microbiome associated with susceptibility or resistance to infection with *L. major*. *Rikenellaceae* were common biomarkers associated with susceptibility in the gut microbiota of all infected susceptible mouse strains. Several biomarkers were identified in the ileum microbiota of infected resistant OcB/Dem mice, with *Moraxellaceae* being the dominant biomarker. The study also revealed possible gene functions in the gut microbiome of infected mice. Xenobiotic biodegradation and metabolism and amino acid metabolism pathways were mainly associated with the ileum microbiome of infected resistant strains. In contrast, genes related to carbohydrate metabolism as well as glycan biosynthesis and glycan metabolism were significantly higher in the gut microbiome of infected susceptible mice. The identification of these biomarkers and the prediction of the functional pathways involved in susceptibility or resistance could provide valuable insights for the development of preventive microbiome-based medicine and new drugs against leishmaniasis by identifying new targets.

6. 3. Gene mapping and identification

The robust genetic system of recombinant congenic strains of mice enabled the detection of a novel suggestive locus on chromosome 7. This locus contains 9 candidate genes: *Cd33*, *Klk1b22*, *Siglece*, *Klk1b16*, *Fut2*, *Grwd1*, *Abcc6*, *Otog*, and *Mkrn3*.

Mapping of TBEV controlling genes in mice is not easy due to presence of a strong TBEV controlling gene *Oas1b*, which is identical both in BALB/c and CcS-11, as well as in majority of laboratory mouse strains and masks effects of other controlling genes, therefore we used a powerful genetic system of recombinant congenic strains [53].

Gene *Cd33* carried in the strain STS a nonsense mutation (Table 2). Product of this gene is in mouse expressed on myeloid precursors and cells of myeloid origin [155] and on microglial cells [156]. It can inhibit response to amyloid plaques and its deletion leads to protection in the mouse model of Alzheimer disease (AD) [156] and in humans some *CD33* genetic variants are associated with late-onset AD [157]; its potential role in pathology of TBE might be associated with its regulatory role in inflammatory responses.

The genes detected and their influence on disease outcome in mouse will be in focus for futher studies. The candidate genes will be subjected for search of human orthologs and their role on susceptibility to TBE in humans [53].

7. Contribution to biomedical engineering

The implementation of a new, simple, robust, non-toxic and inexpensive method for the mass testing of laboratory samples could significantly benefit to researchers and technicians. This genotyping method could also be used to select samples for more costly and detailed analysis. This new typing method has proven to be functional and there are many more applications than those presented. The principles of biomedical engineering emphasise the importance of using new technologies in applied research. The development of novel method that enable mass and faster diagnosis of disease consequences could be crucial for effective treatment of patients. Proof of concept of this method also led to the novel original findings in studies of genetic impact on development of leishmaniasis and encephalitis.

In addition to the benefits for biomedical engineering described above, our method opens up new possibilities for application in other projects, such as the artificial lung ventilation project. The genotyping method in combination with enzymatic cleavage can help to identify the influence of a genetic polymorphism on the risk of developing inflammation during artificial lung ventilation.

8. Conclusion

The newly developed DNA typing method that we have presented is simple, inexpensive and easy to implement in the laboratory. The instruments used in this method are common in all standard laboratories and the method is robust enough to be performed by regular laboratory personnel. Researchers and technicians using this genotyping method are not exposed to the risk of working with radioactive labeling or acrylamide. The entire method is non-toxic and poses no risks for routine use. It has proven to be effective in two disease models: leishmaniasis and encephalitis.

The new genotyping method has been used to test genetically defined animal models to decipher the contribution of individual genes to the phenotypic manifestations of specific diseases. In the complex comparative microbiome study, a different composition of the microbiome in the ileum was observed depending on the genotype of the tested strains during *L. major* infection.

This genotyping method also enabled the gene mapping and identification of a new TBE susceptibility locus on chromosome 7 and the identification of 9 potential candidate genes. The products of some of these genes have been described as being involved in defense against flaviviruses, but the role of others is unknown. The genes detected here will be the focus of future studies that will include characterization of the products of the candidate genes.

Due to the 92% homology between the mouse and human genome, these original findings could have a major impact on translational medicine and thus on improving human healthcare.

9. List of outputs

9. 1. Publications

Mrázek J, Mrázková L, Mekadim C, Jarošíková T, Krayem I, Sohrabi Y, Demant P, Lipoldová M (2024). Effects of *Leishmania major* infection on the gut microbiome of resistant and susceptible mice. *Applied Microbiology and Biotechnology,* 108(1), 1-16. Available from: https://doi.org/10.1007/s00253-024-13002-y

Palus M, Sohrabi Y, Broman KW, Strnad H, Šíma M, Růžek D, Volkova V, Slapničková M, Vojtíšková J, Mrázková L, Salát J, Lipoldová M. A novel locus on mouse chromosome 7 that influences survival after infection with tick-borne encephalitis virus. *BMC Neuroscience* 2018; 19(1):39. doi: 10.1186/s12868-018-0438-8.

Šíma M, Kocandová L, Lipoldová M. Genotyping of short tandem repeats (STRs) markers with 6 bp or higher length difference using PCR and high resolution agarose electrophoresis. *Protocol Exchange.* 2015, 2015 ISSN 2043-0116. DOI 10.1038/protex.2015.054. Available from: http://www.nature.com/protocolexchange/protocols/3973

9. 2. Conference proceedings

Mrázková L, Štěpánová A, Zavoloková B, Jarošíková T, Krayem I, Mrázek J, Mekadim Ch, Slapničková M, Demant P, Lipoldová M. Molecular genetics methods in use. In: JAROŠÍKOVÁ, T. and R. ŠIROKÁ, eds. Instruments and Methods for Biology and Medicine 2020. 10th Student Scientific Conference on Instruments and Methods for Biology and Medicine, Kladno, 2021-10-08. Praha: Czech Technical University in Prague, 2020. p. 31-35. ISBN 9788001067963.

Štěpánová A, Jarošíková T, Mrázková L, Lipoldová M. Methods of the genetic engineering for mapping of genes that influence response to parasite *Leishmania major*. In: Instruments and Methods for Biology and Medicine 2016. Conference Proceedings. Kladno: Faculty of Biomedical Engineering, Department of Natural sciences, 2016. p. 39-43. ISBN 978-80-01-06017-9.

Zavoloková B, Mrázek J, Mrázková L, Jarošíková T, Slapničková M, Mekadim Ch, Lipoldová M. Estimation of influence of *Leishmania major* infection on mouse gut microbiome by tools of molecular genetics. In: Instruments and Methods for Biology and Medicine 2018. Conference Proceedings. Prague: Czech Technical University, 2018. ISBN 978-80-01-06502-0.

9. 3. Microbiome data availability

Sequences were deposited into the SRA database of NCBI under accession number: PRJNA973043.

9. 4. Support

The thesis was supported by GAČR GA14-30186S "Hidden relevant functional pathways in host response to *Leishmania* infection revealed in focused genomic constructions", SGS15/173/OHK4/2T/17 “The methods of molecular biology in study of RNA expression in experimental model”, GAČR GA16-22346S "Strong epistatic control of development of leishmaniasis - identification of genes and mechanisms", SGS18/202/OHK4/3T/17 “The methods of molecular biology in study of microbiome in experimental model”, NV19-05-00457 „Genetic predisposition to different clinical course and severity of tick-borne encephalitis“ and by the Charles University Program COOPERATIO: the scientific project “Medical Diagnostics and Basic Medical Sciences” (the field “Medical Genetics”).

10. References

[1] Lander ES et al. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860-921. doi: 10.1038/35057062. Erratum in: Nature 2001 Aug 2;412(6846):565. Erratum in: Nature 2001 Jun 7;411(6838):720. Szustakowki, J [corrected to Szustakowski, J]. PMID: 11237011.

[2] Ziaei JH et al. A deep population reference panel of tandem repeat variation. Nat Commun. 2023 Oct 23;14(1):6711. doi: 10.1038/s41467-023-42278-3. PMID: 37872149; PMCID: PMC10593948.

[3] Ellegren, H. Microsatellites: simple sequences with complex evolution. *Nature reviews. Genetics* 5, 435-445, doi:10.1038/nrg1348 (2004).

[4] Gemayel, R., Vinces, M. D., Legendre, M. & Verstrepen, K. J. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu Rev Genet* 44, 445-477, doi:10.1146/annurev-genet-072610-155046 (2010).

[5] Arslan A. Compendious survey of protein tandem repeats in inbred mouse strains. BMC Genom Data. 2022 Aug 5;23(1):62. doi: 10.1186/s12863-022-01079-1. PMID: 35931961; PMCID: PMC9354378.

[6] Press MO, McCoy RC, Hall AN, Akey JM, Queitsch C. Massive variation of short tandem repeats with functional consequences across strains of Arabidopsis thaliana. Genome Res. 2018 Aug;28(8):1169-1178. doi: 10.1101/gr.231753.117. Epub 2018 Jul 3. PMID: 29970452; PMCID: PMC6071631.

[7] Sawaya, S. et al. Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. *PLoS One* 8, e54710, doi:10.1371/journal.pone.0054710 (2013).

[8] Hamada, H., Petrino, M. G. & Kakunaga, T. A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc Natl Acad Sci U S A* 79, 6465-6469 (1982).

[9] Bell GI, Selby MJ, Rutter WJ. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature. 1982 Jan 7;295(5844):31-5. doi: 10.1038/295031a0. PMID: 7035959.

[10] Goodbourn SE, Higgs DR, Clegg JB, Weatherall DJ. Molecular basis of length polymorphism in the human zeta-globin gene complex. Proc Natl Acad Sci U S A. 1983 Aug;80(16):5022-6. doi: 10.1073/pnas.80.16.5022. PMID: 6308667; PMCID: PMC384180.

[11] Richard, G. F., Kerrest, A. & Dujon, B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev* 72, 686-727, doi:10.1128/MMBR.00011-08 (2008).

[12] Tanudisastro HA, Deveson IW, Dashnow H, MacArthur DG. Sequencing and characterizing short tandem repeats in the human genome. Nat Rev Genet. 2024 Jul;25(7):460-475. doi: 10.1038/s41576-024-00692-3. Epub 2024 Feb 16. PMID: 38366034.

[13] Metzgar, D., Thomas, E., Davis, C., Field, D. & Wills, C. The microsatellites of Escherichia coli: rapidly evolving repetitive DNAs in a non-pathogenic prokaryote. *Mol Microbiol* 39, 183-190 (2001).

[14] Tautz, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res* 17, 6463-6471 (1989).

[15] Love, J. M., Knight, A. M., McAleer, M. A. & Todd, J. A. Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites. *Nucleic Acids Res* 18, 4123-4130 (1990).

[16] Weber, J L, and P E May. "Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction." *American journal of human genetics* vol. 44,3 (1989): 388-96.

[17] Dib C et al. A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature. 1996 Mar 14;380(6570):152-4. doi: 10.1038/380152a0. PMID: 8600387.

[18] Dietrich WF et al. A comprehensive genetic map of the mouse genome. Nature. 1996;380:149–52.

[19] Willems T, Gymrek M, Highnam G; 1000 Genomes Project Consortium, Mittelman D, Erlich Y. The landscape of human STR variation. *Genome Res.* 2014 Nov;24(11):1894-904. doi: 10.1101/gr.177774.114. Epub 2014 Aug 18. PMID: 25135957; PMCID: PMC4216929.

[20] Mouse Genome Sequencing Consortium., Genome Sequencing Center:., Chinwalla, A. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562 (2002). https://doi.org/10.1038/nature01262

[21] Kockum, I., Huang, J., & Stridh, P. (2023). Overview of genotyping technologies and methods. *Current Protocols*, 3, e727. doi: 10.1002/cpz1.727

[22] Gary Benson, Tandem repeats finder: a program to analyze DNA sequences, *Nucleic Acids Research*, Volume 27, Issue 2, 1 January 1999, Pages 573–580, https://doi.org/10.1093/nar/27.2.573

[23] Thiel, T., Michalek, W., Varshney, R. K. & Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). *Theor Appl Genet* 106, 411-422, doi:10.1007/s00122-002-1031-0 (2003).

[24] Girgis, H. Z. & Sheetlin, S. L. MsDetector: toward a standard computational tool for DNA microsatellites detection. *Nucleic Acids Res* 41, e22, doi:10.1093/nar/gks881 (2013).

[25] Lundström OS et al. WebSTR: A Population-wide Database of Short Tandem Repeat Variation in Humans. J Mol Biol. 2023 Oct 15;435(20):168260. doi: 10.1016/j.jmb.2023.168260. Epub 2023 Sep 7. PMID: 37678708.

[26] Lu J et al. LUSTR: a new customizable tool for calling genome-wide germline and somatic short tandem repeat variants. BMC Genomics. 2024 Jan 26;25(1):115. doi: 10.1186/s12864-023-09935-9. PMID: 38279154; PMCID: PMC10811831.

[27] Gymrek M. A genomic view of short tandem repeats. Curr Opin Genet Dev. 2017 Jun;44:9-16. doi: 10.1016/j.gde.2017.01.012. Epub 2017 Feb 16. PMID: 28213161.

[28] Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289-1291, doi:10.1093/bioinformatics/btm091 (2007).

[29] Untergasser, A. et al. Primer3--new capabilities and interfaces. *Nucleic Acids Res* 40, e115, doi:10.1093/nar/gks596 (2012).

[30] Robinson, A. J., Love, C. G., Batley, J., Barker, G. & Edwards, D. Simple sequence repeat marker loci discovery using SSR primer. *Bioinformatics* 20, 1475-1476, doi:10.1093/bioinformatics/bth104 (2004).

[31] da Maia, L. C. et al. SSR Locator: Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation. Int *J Plant Genomics* 2008, 412696, doi:10.1155/2008/412696 (2008).

[32] Ruitberg, C. M., Reeder, D. J. & Butler, J. M. STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res* 29, 320-322 (2001).

[33] Agrafioti, I. & Stumpf, M. P. SNPSTR: a database of compound microsatellite-SNP markers. *Nucleic Acids Res* 35, D71-75, doi:10.1093/nar/gkl806 (2007).

[34] Jacquot, S et al. Optimizing PCR for mouse genotyping: Recommendations for reliable, rapid, cost effective, robust and adaptable to high-throughput genotyping protocol for any type of mutation. *Current Protocols in Mouse Biology*, 9, e65. 2019. doi: 10.1002/cpmo.65

[35] Southern E. Gel electrophoresis of restriction fragments. Methods Enzymol. 1979;68:152-76. doi: 10.1016/0076-6879(79)68011-4. PMID: 232210.

[36] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977 Dec;74(12):5463-7. doi: 10.1073/pnas.74.12.5463. PMID: 271968; PMCID: PMC431765.

[37] Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S A. 1977 Feb;74(2):560-4. doi: 10.1073/pnas.74.2.560. PMID: 265521; PMCID: PMC392330.

[38] Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science. 1985 Dec 20;230(4732):1350-4. doi: 10.1126/science.2999980. PMID: 2999980.

[39] Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol. 1986;51 Pt 1:263-73. doi: 10.1101/sqb.1986.051.01.032. PMID: 3472723.

[40] Holloway JW et al. Comparison of three methods for single nucleotide polymorphism typing for DNA bank studies: Sequence-specific oligonucleotide probe hybridisation, TaqMan liquid phase hybridisation, and microplate array diagonal gel electrophoresis (MADGE). *Human Mutation*, 14(4), 340–347. 1999. https://doi.org/10.1002/(sici)1098-1004(199910)14:4<340:aid-humu10>3.0.co;2-z

[41] Lee C, Antonarakis SE, Hamosh A, Burn J. Three decades of the Human Genome Organization. Am J Med Genet A. 2021 Nov;185(11):3314-3321. doi: 10.1002/ajmg.a.62512. Epub 2021 Sep 28. PMID: 34581472; PMCID: PMC9293206.

[42] Venter JC et al. The sequence of the human genome. Science. 2001 Feb 16;291(5507):1304-51. doi: 10.1126/science.1058040. Erratum in: Science 2001 Jun 5;292(5523):1838. PMID: 11181995.

[43] International HapMap Consortium. The International HapMap Project. *Nature*. 2003 Dec 18;426(6968):789-96. doi: 10.1038/nature02168. PMID: 14685227.

[44] 1000 Genomes Project Consortium et al. “A map of human genome variation from population-scale sequencing.” *Nature* vol. 467,7319 (2010): 1061-73. doi:10.1038/nature09534

[45] Kircher M, Kelso J. High-throughput DNA sequencing--concepts and limitations. Bioessays. 2010 Jun;32(6):524-36. doi: 10.1002/bies.200900181. PMID: 20486139.

[46] Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016 May 17;17(6):333-51. doi: 10.1038/nrg.2016.49. PMID: 27184599; PMCID: PMC10373632.

[47] Kuleshov V, Xie D, Chen R, Pushkarev D, Ma Z, Blauwkamp T, Kertesz M, Snyder M. Whole-genome haplotyping using long reads and statistical methods. Nat Biotechnol. 2014 Mar;32(3):261-266. doi: 10.1038/nbt.2833. Epub 2014 Feb 23. PMID: 24561555; PMCID: PMC4073643.

[48] https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost (Assessed August 11, 2024)

[49] Ye S, Dhillon S, Ke X, Collins AR, Day IN. An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res.* 2001 Sep 1;29(17):E88-8. doi: 10.1093/nar/29.17.e88. PMID: 11522844; PMCID: PMC55900.

[50] Havlenová S, *Akrylamid*. Bakalářská práce, RNDr. Jiří Totušek, CSc. vedoucí práce, Brno Masarykova univerzita, fakulta lékařská, 2009, dostupné z: https://is.muni.cz/th/kdzu0/Bakalaska_prace_Akrylamid.pdf

[51] Bušová M, Bencko V, Veszelits Laktičová K, Holcátová I, Vargová M. Risk of exposure to acrylamide. *Cent Eur J Public Health*. 2020;28(Supplement):S43-46. doi: 10.21101/cejph.a6177. PubMed PMID: 33069180.

[52] Šíma M, Kocandová L, Lipoldová M. Genotyping of short tandem repeats (STRs) markers with 6 bp or higher length difference using PCR and high resolution agarose electrophoresis. *Protocol Exchange*. 2015, 2015 ISSN 2043-0116. DOI 10.1038/protex.2015.054. Available from: http://www.nature.com/protocolexchange/protocols/3973

[53] Palus M, Sohrabi Y, Broman KW, Strnad H, Šíma M, Růžek D, Volkova V, Slapničková M, Vojtíšková J, Mrázková L, Salát J, Lipoldová M. A novel locus on mouse chromosome 7 that influences survival after infection with tick-borne encephalitis virus. *BMC Neurosci.* 2018; 19(1):39. doi: 10.1186/s12868-018-0438-8.

[54] Pustijanac E, Buršić M, Talapko J, Škrlec I, Meštrović T, Lišnjić D. Tick-Borne Encephalitis Virus: A Comprehensive Review of Transmission, Pathogenesis, Epidemiology, Clinical Manifestations, Diagnosis, and Prevention. Microorganisms. 2023 Jun 22;11(7):1634. doi: 10.3390/microorganisms11071634. PMID: 37512806; PMCID: PMC10383662.

[55] Bogovic P, Strle F. Tick-borne encephalitis: A review of epidemiology, clinical characteristics, and management. *World J Clin Cases*. 2015 May 16;3(5):430-41. doi: 10.12998/wjcc.v3.i5.430. PMID: 25984517; PMCID: PMC4419106.

[56] Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, den Boer M; WHO Leishmaniasis Control Team. Leishmaniasis worldwide and global estimates of its incidence. *PLoS One*. 2012; 7(5):e35671.

[57] Lipoldová M, Demant P. Genetic susceptibility to infectious disease: lessons from mouse models of leishmaniasis. *Nat Rev Genet.* 2006; 7(4):294-305.

[58] Chang CS, Kao CY. (2019) Current understanding of the gut microbiota shaping mechanisms. *J Biomed Sci* 26: 59. doi: 10.1186/s12929-019-0554-5

[59] Partida-Rodríguez O, Serrano-Vázquez A, Nieves-Ramírez M E, Moran P, Rojas L, Portillo T, González E, Hernández E, Finlay BB, Ximenez C (2017) Human intestinal microbiota: interaction between parasites and the host immune response. *Arch Med Res* 48:690-700. doi:10.1016/j.arcmed.2017.11.015

[60] Bhattacharjee S, Kalbfuss N, Prazeres da Costa C (2017) Parasites, microbiota and metabolic disease. *Parasite Immunol* 39. doi:10.1111/pim.12390

[61] Kamada N, Seo SU, Chen GY, Núñez G (2013) Role of the gut microbiota in immunity and inflammatory disease. *Nat Rev Immunol* 13:321-335. doi:10.1038/nri3430

[62] Allen, J.E. (2016). Microbiota, parasites and immunity. *Parasite Immunology*. 38, 3–4. doi: 10.1111/pim.12298

[63] Wang B, Yao M, Lv L, Ling Z, Li L (2017) The Human Microbiota in Health and Disease. *Engineering* 3:71-82. doi:10.1016/J.ENG.2017.01.008

[64] Ding RX, Goh WR, Wu RN, Yue XQ, Luo X, Khine WWT, Wu JR, Lee YK (2019) Revisit gut microbiota and its impact on human health and disease. *J Food Drug Anal* 27:623-631. doi: 10.1016/j.jfda.2018.12.012

[65] Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE. Human genetics shape the gut microbiome. *Cell*. 2014; 159(4):789-99. doi: 10.1016/j.cell.2014.09.053.

[66] Rothschild D et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature*. 2018; 555(7695):210-215. doi: 10.1038/nature25973.

[67] Caballero,S.,Pamer, E.G. (2015). Microbiota-mediatedinflammationandantimi-crobialdefense in the intestine. *Annu. Rev. Immunol*. 33, 227-256. doi:10.1146/annurev-immunol-032713-120238

[68] Galloway-Peña J, Hanson B. Tools for Analysis of the Microbiome. *Dig Dis Sci*. 2020 Mar;65(3):674-685. doi: 10.1007/s10620-020-06091-y. PMID: 32002757; PMCID: PMC7598837.

[69] Hand TW, Vujkovic-Cvijin I, Ridaura VK, Belkaid Y. Linking the microbiota, chronic Disease, and the immune system. *Trends Endocrinol Metab.* 2016; 27(12):831-843.

[70] Lappan R, Classon C, Kumar S, Singh OP, De Almeida RV, Chakravarty J, Kumari P, Kansal S, Sundar S, Blackwell JM (2019) Meta-taxonomic analysis of prokaryotic and eukaryotic gut flora in stool samples from visceral leishmaniasis cases and endemic controls in Bihar State India. *PLoS Negl Trop Dis* 13:1-28. doi:10.1371/journal.pntd.0007444

[71] Oliveira, M.R., Tafuri, W.L., Afonso, L.C., Oliveira M.A.P.,Nicolo, J. R., Vieira, E. C., Scott, P., Melo, M. N., Vieira, L. Q. (2005). Germ-free mice produce high levels of inter-feron-gamma in response to infection with *Leishmania major* but fail to heal lesions. *Parasitology.* 131, 477–488. doi:10.1017/S0031182005008073

[72] Naik, S., Bouladoux, N., Wilhelm, C., Molloy, M.J., Salcedo, R., Kastenmuller, W., Deming, C., Quinones, M., Koo, L., Conlan, S., Spencer, S., Hall, J.A., Dzutsev, A., Kong, H., Campbell, D.J., Trinchieri, G., Segre, J.A., Belkaid, Y. (2012). Compartmentalized control of skin immunity by resident commensals. *Science*. 337, 1115–1119. doi: 10.1126/science.1225152.

[73] Canesso, M.C., Vieira, A.T., Castro, T.B., Schirmer, B.G., Cisalpino, D., Martins, F.S., Rachid, M.A., Nicoli, J.R., Teixeira, M.M., Barcelos, L.S. (2014). Healing is accelerated and scarless in the absence of commensal microbiota. *J. Immunol.* 193, 5171–5180. doi:10.4049/jimmunol.1400625.

[74] Lamour SD, Veselkov KA, Posma JM, Giraud E, Rogers ME, Croft S, Marchesi JR, Holmes E, Seifert K, Saric J. Metabolic, immune, and gut microbial signals mount a systems response to *Leishmania major* infection. *J Proteome* Res. 2015; 14(1):318-29.

[75] Zhu M, Zhao S. Candidate Gene Identification Approach: Progress and Challenges. Int *J Biol Sci* 2007; 3(7):420-427. doi:10.7150/ijbs.3.420. https://www.ijbs.com/v03p0420.htm

[76] Segundo-Val IS, Sanz-Lozano CS. Introduction to the Gene Expression Analysis. Methods Mol Biol. 2016;1434:29-43. doi: 10.1007/978-1-4939-3652-6_3. PMID: 27300529.

[77] Palus M, Vojtíšková J, Salát J, Kopecký J, Grubhoffer L, Lipoldová M, Demant P, Růžek D: Mice with different susceptibility to tick-borne encephalitis virus infection show selective neutralizing antibody response and inflammatory reaction in the central nervous system. J *Neuroinflammation* 2013, 10:77.

[78] Lipoldová M, Svobodová M, Krulová M, Havelková H, Badalová J, Nohýnková E, Holán V, Hart AA, Volf P, Demant P. Susceptibility to Leishmania major infection in mice: multiple loci and heterogeneity of immunopathological phenotypes. Genes Immun. 2000 Feb;1(3):200-6. doi: 10.1038/sj.gene.6363660. PMID: 11196712.

[79] Grekov I, Svobodová M, Nohýnková E, Lipoldová M. Preparation of highly infective *Leishmania* promastigotes by cultivation on SNB-9 biphasic medium. *Journal of Microbiological Methods* 87 (2011): 273–277.

[80] Lipoldová M, Svobodová M, Havelková H, Krulová M, Badalová J, Nohýnková E, Hart AA, Schlegel D, Volf P, Demant P. Mouse genetic model for clinical and immunological heterogeneity of leishmaniasis. *Immunogenetics*. 2002; 54(3):174-83.

[81] Demant, Peter. (2003). Cancer susceptibility in the mouse: Genetics, biology and implications for human cancer. *Nature reviews. Genetics*. 4. 721-34. 10.1038/nrg1157.

[82] Sohrabi Y, Volkova V, Kobets T, Havelková H, Krayem I, Slapničková M, Demant P, Lipoldová M. Genetic regulation of guanylate-binding proteins 2b and 5 during leishmaniasis in mice. *Front Immunol.* 2018 Feb 7;9:130. doi: 10.3389/fimmu.2018.00130.

[83] Laird PW, Zijderveld A, Linders K, Rudnicki MA, Jaenisch R, Berns A. Simplified mammalian DNA isolation procedure. Nucleic Acids Res. 1991 Aug 11;19(15):4293. doi: 10.1093/nar/19.15.4293. PMID: 1870982; PMCID: PMC328579.

[84] Havelková H, Badalová J, Svobodová M, Vojtíková J, Kurey I, Vladimirov V, Demant P, Lipoldová M. Genetics of susceptibility to leishmaniasis in mice: four novel loci and functional heterogeneity of gene effects. Genes Immun. 2006 Apr;7(3):220-33. doi: 10.1038/sj.gene.6364290. PMID: 16511555.

[85] Šíma M, Havelková H, Quan L, Svobodová M, Jarošíková T, Vojtíšková J, Stassen AP, Demant P, Lipoldová M. Genetic control of resistance to Trypanosoma brucei brucei infection in mice. *PLoS Negl Trop Dis.* 2011; 5(6):e1173. doi: 10.1371/journal.pntd.0001173.

[86] Mrázek J, Mrázková L, Mekadim C, Jarošíková T, Krayem I, Sohrabi Y, Demant P, Lipoldová M (2024). Effects of *Leishmania major* infection on the gut microbiome of resistant and susceptible mice. *Applied Microbiology and Biotechnology*, 108(1), 1-16. Available from: https://doi.org/10.1007/s00253-024-13002-y

[87] Kobets T, Badalová J, Grekov I, Havelková H, Svobodová M, Lipoldová M. *Leishmania* parasite detection and quantification using PCR-ELISA. *Nature Protocol* 5, 1074–1080 (20 May 2010). https://doi.org/10.1038/nprot.2010.68.

[88] R Core Team (2017). [online] R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [21-03-15] URL https://www.R-project.org/.

[89] Muyzer G, de Waal EC, Uitterlinden a G (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*, 59(3): 695–700.

[90] BioNumerics at https://www.bionumerics.com

[91] CD genomics at https://www.cd-genomics.com/blog/sanger-sequencing-introduction-principle-and-protocol/

[92] Rinttilä T., Kassinen A., Malinen E., Krogius L., Palva A., Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR, *Journal of Applied Microbiology*, Volume 97, Issue 6, 1 December 2004, Pages 1166–1177, https://doi.org/10.1111/j.1365-2672.2004.02409.x

[93] Geneious at https://www.geneious.com

[94] NCBI database at https://blast.ncbi.nlm.nih.gov

[95] Fliegerova, K., Tapio, I., Bonin, A., Mrazek, J., Callegari, M.L., Bani, P., Bayat, A., Vilkki, J., Kopecny, J., Shingfield, K. J., Boyer, F., Coissac, E., Taberlet, P., Wallace, J. R. (2014) Effect of DNA extraction and sample preservation method on rumen bacterial population. *Anaerobe* 29: 80-84. doi: 10.1016/j.anaerobe.2013.09.015.

[96] Milani, Ch., Hevia, A., Foroni, E., Duranti, S., Turroni, F., Lugli, G. A., Sanchez, B., Martin, R., Gueimonde, M., van Sinderen, D., Margolles, A., Ventura, M. (2013) Assessing the Fecal Microbiota: An

Optimized Ion Torrent 16S rRNA Gene-Based Analysis Protocol. *PLoS ONE* 8 (7): e68739. doi: 10.1371/journal.pone.0068739.

[97] Ion Torrent at https://www.thermofisher.com/cz/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html

[98] Rothberg M R, Hinz W, Rearick T M, Schulz J, Mileski W, Davey M, Leamon J H, et al. (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475(7356):348-52. doi: 10.1038/nature10242.

[99] Amplicon Sequencing at https://tools.thermofisher.com/content/sfs/brochures/cms_094273.pdf

[100] The PGM sequencer at https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_094139.pdf

[101] Bolyen E et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* 2019 Aug;37(8):852-857. doi: 10.1038/s41587-019-0209-9. Erratum in: Nat Biotechnol. 2019 Sep;37(9):1091. doi: 10.1038/s41587-019-0252-6. PMID: 31341288; PMCID: PMC7015180.

[102] Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581-583. doi:10.1038/nmeth.3869

[103] Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: A versatile open source tool for metagenomics. *PeerJ* 4:e2584. doi:10.7717/peerj.2584.

[104] Anderson MJ (2017) Permutational multivariate analysis of variance (PERMANOVA). In: Balakrishnan N, Colton T, Everitt B, Piegorsch W, Ruggeri F and Teugels JL (ed) StatsRef: Statistics Reference Online. Wiley, New York. doi:10.1002/9781118445112.stat07841

[105] Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett W S, Huttenhower C (2011) Metagenomic biomarker discovery and explanation. *Genome Biol* 12(6):R60. doi: 10.1186/gb-2011-12-6-r60.

[106] Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 31:814-821. doi:10.1038/nbt.2676

[107] Parks DH, Tyson GW, Hugenholtz P, Beiko RG (2014) STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30:3123-3124. doi:10.1093/bioinformatics/btu494

[108] LEfSe at https://huttenhower.sph.harvard.edu/lefse/

[109] LEfSe in Microbiome Data at https://www.jove.com/v/61715/assisted-selection-biomarkers-linear-discriminant-analysis-effect doi: 10.3791/61715-v.

[110] Xu S, Atchley WR: Mapping quantitative trait loci for complex binary diseases using line crosses. *Genetics* 1996, 143:1417–1424.

[111] Broman KW: Mapping quantitative trait loci in the case of a spike in the phenotype distribution. *Genetics* 2003, 163(3):1169-1175.

[112] Churchill GA, Doerge RW: Empirical threshold values for quantitative trait mapping. *Genetics* 1994, 138:963–971.

[113] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014, 30(15):2114-2120.

[114] Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 2011, 27(21):2957-2963.

[115] Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010, 26(5):589-595.

[116] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010, 20(9):1297-1303.

[117] Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* (Austin) 2012, 6(2):80-92.

[118] Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol* 2011, 29(1):24-26.

[119] Kurey I, Kobets T, Havelková H, Slapnicková M, Quan L, Trtková K, Grekov I, Svobodová M, Stassen AP, Hutson A, Demant P, Lipoldová M. Distinct genetic control of parasite elimination, dissemination, and disease after *Leishmania major* infection. *Immunogenetics*. 2009;61(9):619-33. doi: 10.1007/s00251-009-0392-9

[120] Slapničková M, Volkova V, Čepičková M, Kobets T, Šíma M, Svobodová M, Demant P, Lipoldová M. Gene-specific sex effects on eosinophil infiltration in leishmaniasis. Biol Sex Differ. 2016 Nov 22;7:59. doi: 10.1186/s13293-016-0117-3. PMID: 27895891; PMCID: PMC5120444.

[121] Kobets T, Čepičková M, Volkova V, Sohrabi Y, Havelková H, Svobodová M, Demant P, Lipoldová M. Novel Loci Controlling Parasite Load in Organs of Mice Infected With Leishmania major, Their Interactions and Sex Influence. Front Immunol. 2019 Jun 7;10:1083. doi: 10.3389/fimmu.2019.01083. PMID: 31231359; PMCID: PMC6566641.

[122] Sohrabi Y, Havelková H, Kobets T, Šíma M, Volkova V, Grekov I, Jarošíková T, Kurey I, Vojtíšková J, Svobodová M, Demant P, Lipoldová M. Mapping the genes for susceptibility and response to *Leishmania tropica* in mouse. *PLoS Negl Trop Dis.* 2013;7(7):e2282. doi: 10.1371/journal.pntd.0002282

[123] Mrázková L, Štěpánová A, Zavoloková B, Jarošíková T, Krayem I, Mrázek J, Mekadim Ch, Slapničková M, Demant P, Lipoldová M. Molecular genetics methods in use. In: JAROŠÍKOVÁ, T. and R. ŠIROKÁ, eds. Instruments and Methods for Biology and Medicine 2020. 10th Student Scientific Conference on Instruments and Methods for Biology and Medicine, Kladno, 2021-10-08. Praha: Czech Technical University in Prague, 2020. p. 31-35. ISBN 9788001067963.

[124] Demant P, Hart AA (1986) Recombinant congenic strains - a new tool for analyzing genetic traits determined by more than one gene. *Immunogenetics* 24:416–422.

[125] Stassen AP, Groot PC, Eppig JT, Demant P (1996) Genetic composition of the recombinant congenic strains. *Mamm Genome* 7:55-8. doi: 10.1007/s003359900013

[126] Krayem I, Sohrabi Y, Javorková E, Volkova V, Strnad H, Havelková H, Vojtíšková J, Aidarova A, Holáň V, Demant P, Lipoldová M (2022) Genetic influence on frequencies of myeloid-derived cell subpopulations in mouse. *Front Immunol* 12:760881. doi:10.3389/fimmu.2021.760881

[127] Lee H. K., Tsongalis G. J. Chapter 31 - Pharmacogenomics and Personalized Medicine in the Treatment of Human Diseases, Essential Concepts in Molecular Pathology, 2010, Academic Press, ISBN 9780123744180, https://doi.org/10.1016/B978-0-12-374418-0.00040-2

[128] Wu D, Wu C, Zhang S, Zhong Y. Risk Factors of Ventilator-Associated Pneumonia in Critically III Patients. Front Pharmacol. 2019 May 9;10:482. doi: 10.3389/fphar.2019.00482. PMID: 31143118; PMCID: PMC6521332.

[129] Meng W, Cao X, Sun W, Zheng L, Fan B, Zhou S, Liu H, Wang H, Wang W, Liu X. A functional polymorphism at the miR-491-5p binding site in the 3'-untranslated region of the MMP-9 gene increases the risk of developing ventilator-associated pneumonia. Int J Mol Med. 2021 Dec;48(6):217. doi: 10.3892/ijmm.2021.5050. Epub 2021 Oct 19. PMID: 34664683.

[130] Wang SP, Li W, Li C, Duan XY, Duan J. Effect of rs4719839 polymorphism on risk of ventilator-associated pneumonia, expression of microRNA-148 and autophagy-related 16-like 1 (ATG16L1). J Cell Mol Med. 2020 Nov;24(21):12599-12607. doi: 10.1111/jcmm.15824. Epub 2020 Sep 17. PMID: 32940422; PMCID: PMC7686989.

[131] Zhao X, Feng J, Zhang L, Zhao F, Li M, Du Y, Li Y, Wu Q, Li G. One functional variant in the 3'-untranslated region of TLR4 is associated with the elevated risk of ventilator-associated pneumonia in the patients with chronic obstructive pulmonary disease. J Cell Physiol. 2019 Aug;234(10):18879-18886. doi: 10.1002/jcp.28526. Epub 2019 Apr 10. PMID: 30972742.

[132] Kim EJ, Kim MS, Kim MS, Nam J, Choi SH. Association of HLA-DPA1 polymorphism with prolonged mechanical ventilation in patients undergoing liver transplantation. Korean J Anesthesiol. 2022 Oct;75(5):397-406. doi: 10.4097/kja.22014. Epub 2022 May 3. PMID: 35505457; PMCID: PMC9539425

[133] Gu S, Chen D, Zhang JN, Lv X, Wang K, Duan LP, Nie Y, Wu XL (2013) Bacterial community mapping of the mouse gastrointestinal tract. *PLoS ONE* 8:e74957. doi:10.1371/journal.pone.0074957

[134] Wang J, Lang T, Shen J, Dai J, Tian L, Wang X (2019) Core gut bacteria analysis of healthy mice. *Front Microbiol* 10:1-14. doi:10.3389/fmicb.2019.00887

[135] Han X, Shao H, Wang Y, Hu A, Chen R, Chen Q (2020) Composition of the bacterial community in the gastrointestinal tract of Kunming mice. *Electron J Biotechnol* 43:16–22. doi:10.1016/j.ejbt.2019.11.003

[136] Oliveira MR, Tafuri WL, Nicoli JR, Vieira EC, Melo MN, Vieira LQ (1999) Influence of microbiota in experimental cutaneous leishmaniasis in Swiss mice. *Rev Inst Med Trop Sao Paulo* 41:87-94. doi:10.1590/s0036-46651999000200005

[137] Cavallone IN, de Lima SKS, Oliveira KS, Passero LFD, Laurenti MD, Jesus JA, Marinsek GP, Chucri TM, Mari RB (2022) Histological and neuronal changes in the duodenum of hamsters infected with *Leishmania* (*Leishmania*) *infantum*. *Exp Parasitol* 239:108315. https:// doi. org/ 10. 1016/j. exppa ra. 2022. 108315

[138] de Lima SKS, Cavallone IN, Oliveira KS, Passero LFD, Laurenti MD, Jesus JA, Marinsek GP, Chucri TM, Mari RdB (2021) Infection with *Leishmania* (*Leishmania*) *infantum* Changes the Morphology and Myenteric Neurons of the Jejunum of Golden Hamsters. *Parasitologia* 1:225–237. doi:10.3390/ parasitologia1040024

[139] Passos FC, Gois MB, Sousa AD, de Marinho AIL, Corvo L, Soto M, Barral-Netto M, Barral A, Baccan GC (2020) Investigating associations between intestinal alterations and parasite load according to *Bifidobacterium* spp. and *Lactobacillus* spp. abundance in the gut microbiota of hamsters infected by *Leishmania infantum*. *Mem Inst Oswaldo Cruz.* 115:e200377. doi: 10.1590/0074-02760200377

[140] Santos AGA, da Silva MGL, Carneiro EL, de Lima LL, Fernandes ACBS, Silveira TGV, Sant'Ana DMG and Nogueira-Melo GA (2021) A New Target Organ of *Leishmania* (Viannia) *braziliensis* Chronic Infection: The Intestine. *Front. Cell. Infect. Microbiol*. 11:687499. doi: 10.3389/fcimb.2021.687499

[141] Lewis MD, Paun A, Romano A, Langston H, Langner CA, Moore IN, Bock KW, Francisco AF, Brenchley JM, Sacks DL (2020) Fatal progression of experimental visceral leishmaniasis is associated with intestinal parasitism and secondary infection by commensal bacteria, and is delayed by antibiotic prophylaxis. *PLoS Pathog* 16(4):e1008456. doi: 10.1371/journal.ppat.1008456

[142] Chattopadhyay A, Mittal S, Gupta K, Dhir V, Jain S (2020) Intestinal leishmaniasis. *Clin Microbiol Infect*. 10:1345-1346. doi: 10.1016/j.cmi.2020.05.003

[143] Björkholm B, Bok CM, Lundin A, Rafter J, Hibberd ML, Pettersson S (2009) Intestinal microbiota regulate xenobiotic metabolism in the liver. *PLoS ONE* 4:e6958 doi:10.1371/journal.pone.0006958

[144] Toda T, Saito N, Ikarashi N, Ito K, Yamamoto M, Ishige A, Watanabe K, Sugiyama K (2009) Intestinal flora induces the expression of Cyp3a in the mouse liver. *Xenobiotica* 39:323–334. doi:10.1080/00498250802651984

[145] Li CY, Lee S, Cade S, Kuo LJ, Schultz IR, Bhatt DK, Prasad B, Bammler TK, Cui JY (2017) Novel interactions between gut microbiome and host drug-processing genes modify the hepatic metabolism of the environmental chemicals polybrominated diphenyl ethers. *Drug Metab Dispos* 45: 1197-1214. doi:10.1124/dmd.117.077024

[146] Collins SL, Patterson AD (2020) The gut microbiome: an orchestrator of xenobiotic metabolism. *Acta Pharm Sin B* 10:19-32. doi:10.1016/j.apsb.2019.12.001

[147] Coombs GH, Roland Wolf C, Morrison VM, Craft JA (1990) Changes in hepatic xenobiotic-metabolising enzymes in mouse liver following infection with *Leishmania donovani*. *Mol Biochem Parasitol* 41:17-24. doi:10.1016/0166-6851(90)90092-Z

[148] Lanchote VL, Almeida R, Barral A, Barral-Netto M, Marques MP, Moraes NV, da Silva AM, Souza TM, Suarez-Kurtz G (2015) Impact of visceral leishmaniasis and curative chemotherapy on cytochrome P450 activity in Brazilian patients. *Br J Clin Pharmacol* 80:1160-1168. doi:10.1111/bcp.12677

[149] Ren W, Rajendran R, Zhao Y, Tan B, Wu G, Bazer FW, Zhu G, Peng Y, Huang X, Deng J, Yin Y (2018) Amino acids as mediators of metabolic cross talk between host and pathogen. *Front Immunol* 9:319 doi:10.3389/fimmu.2018.00319

[150] Ferreira C, Mesquita I, Barbosa AM, Osório NS, Torrado E, Beauparlant CJ, Droit A, Cunha C, Carvalho A, Saha B, Estaquier J, Silvestre R (2020) Glutamine supplementation improves the efficacy of miltefosine treatment for visceral leishmaniasis. *PLoS Negl Trop Dis* 14:1-17. doi:10.1371/journal.pntd.0008125

[151] Wanasen N, Soong L (2008) L-arginine metabolism and its impact on host immunity against *Leishmania* infection. *Immunol Res* 41:15-25. doi:10.1007/s12026-007-8012-y

[152] Rabhi I, Rabhi S, Ben-Othman R, Rasche A, Daskalaki A, Trentin B, Piquemal D, Regnault B, Descoteaux A, Guizani-Tabbane L; Sysco Consortium. (2012) Transcriptomic signature of *Leishmania* infected mice macrophages: A metabolic point of view. *PLoS Negl Trop Dis* 6:e1763. doi:10.1371/journal.pntd.0001763

[153] Bodhale NP, Pal S, Kumar S, Chattopadhyay D, Saha B, Chattopadhyay N, Bhattacharyya M (2018) Inbred mouse strains differentially susceptible to *Leishmania donovani* infection differ in their immune cell metabolism. *Cytokine*112:12–15. doi:10.1016/j.cyto.2018.06.003

[154] Martins VD, Silva FC, Caixeta F, Carneiro MB, Goes GR, Torres L, Barbosa SC, Vaz L, Paiva NC, Carneiro CM, Vieira LQ, Faria AMC, Maioli TU (2020) Obesity impairs resistance to *Leishmania major* infection in C57BL/6 mice. *PLoS Negl Trop Dis* 14:1-20. doi:10.1371/journal.pntd.0006596.

[155] Brinkman-Van der Linden EC, Angata T, Reynolds SA, Powell LD, Hedrick SM, Varki A. CD33/Siglec-3 binding specificity, expression pattern, and consequences of gene deletion in mice. Mol Cell Biol 2003, 23(12):4199-4206.

[156] Griciuc A, Serrano-Pozo A, Parrado AR, Lesinski AN, Asselin CN, Mullin K, Hooli B, Choi SH, Hyman BT, Tanzi RE. Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. Neuron 2013, 78(4):631-643.

[157] Jiang T, Yu JT, Hu N, Tan MS, Zhu XC, Tan L. CD33 in Alzheimer's disease. Mol Neurobiol 2014; 49(1):529-535.

11. Supplements

11. 1. List of figures

Figure 20: Variations in gut microbiota diversities in the colon of uninfected (a) and infected (b) mice, demonstrated by PCoA plots of unweighted UniFrac distances. (Red –BALB/c strain, blue - CcS-20 strain, and orange –STS strain) [unpublished results]

Figure 21: Dendrogram of microbial profiles from the ileum (a) and the colon (b) of mouse strains B10.O20 and O20. “E” stands for infected and “K” is a control group [unpublished results]

Figure 22: PCoA analysis of microbial profiles from the ileum (a) and colon (b) of mouse strains B10.O20 and O20. “E” stands for infected and “K” is a control group. Comparison of dendrogram data was performed in BioNumerics (AppliedMaths) [unpublished results]

Figure 23: Microbiome alpha diversity (shannon index) among different mouse strains (susceptibles/ resistants) of infected or non-infected mice in: a) colon of CcS/Dem, b) colon of OcB/Dem, c) ileum of CcS/Dem and d) ileum of OcB/Dem. (p-value ≤ 0.05) was considered statistically significant. [Inf: infected, ctr: control (non-infected)] [86]

Figure 24: Principal coordinate analysis (PCoA) plots showed distinct clusters among different groups (susceptibles/ resistants) of infected or non-infected mice in: a) colon of CcS/Dem, b) colon of OcB/Dem, c) ileum of CcS/Dem and d) ileum of OcB/Dem. Ellipses mark 95% confidence ellipses around each group and (p-value ≤ 0.05) was considered statistically significant. [Resistant_Inf: infected resistant, Resistant_ctr: resistant control (non-infected), Susceptible_Inf: infected susceptible, Susceptible_ctr: susceptible control (non-infected)] [86]

Figure 25: Principal coordinate analysis (PCoA) plots showed distinct clusters among infected or non-infected mouse strains of CcS/Dem group in: a) colon of susceptible CcS/Dem, b) colon of resistant CcS/Dem, c) ileum of susceptible CcS/Dem and d) ileum of resistant CcS/Dem. Ellipses mark 95% confidence ellipses around each group and (p-value ≤ 0.05) was considered statistically significant. [Inf: infected, ctr: control (non-infected)] [86]

Figure 26: Relative abundance of the microbial population at the family level among different infected or non-infected CcS/Dem mouse strains in: a) colon, b) ileum. [Inf: infected, ctr: control (non-infected)] [86]

Figure 27: Relative abundance of the microbial population at the family level among different infected or non-infected OcB/Dem mouse strains in: a) colon, b) ileum. [Inf: infected, ctr: control (non-infected)] [86]

Figure 28: Fig. 23: Linear discriminant analysis (LDA) effect size (LEfSe) of taxa at family level in infected susceptible CcS/Dem mice (in green) and infected resistant CcS/Dem mice (in red) in colon with alpha values of 0.05 and a threshold value of 2.0. [Resistant_Inf: infected resistant, Susceptible_Inf: infected susceptible] [86]

Figure 29: Cladogram of colon microbiota present a phylogenetic plot of LEfSe and taxa: D1 for Phylum level, D2 for Class level, D3 for Order level and D4 for Family level). [Resistant_Inf: infected resistant, Susceptible_Inf: infected susceptible] [86]

Figure 30: Linear discriminant analysis (LDA) effect size (LEfSe) of taxa at family level in infected susceptible CcS/Dem mice (in green) and infected resistant CcS/Dem mice (in red) in ileum with alpha values of 0.05 and a threshold value of 2.0. [Resistant_Inf: infected resistant, Susceptible_Inf: infected susceptible] [86]

Figure 31: Cladogram of ileum microbiota present a phylogenetic plot of LEfSe and taxa: D1 for Phylum level, D2 for Class level, D3 for Order level and D4 for Family level).

11. 2. List of tables