CZECH TECHNICAL UNIVERSITY IN PRAGUE

DISSERTATION THESIS

Autor: Ing. Lucie Mrázková

Název díla: **Molecular biology methods in studies of genetic influence and identification of controling genes in experimental model of infectious diseases**

Vydalo: České vysoké učení technické v Praze

Zpracovala: Fakulta biomedicínského inženýrství

Kontaktní adresa: Nám. Sítná 3105, 272 01 Kladno

Počet stran: 34

Pořadí vydání: 1.

Tisk: JAMACOPY s.r.o., Kotorská 1569/14, 140 00 Praha 4

ISBN 978-80-01-07335-3

Czech Technical University in Prague
Faculty of Biomedical Engineering
Biomedical and Clinical Technology

Ing. Lucie Mrázková

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

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

Dissertation proposal for the award of the academic title "Doctor", abbreviated to "Ph.D.".

Kladno, August 2024

The dissertation was elaborated in the combined form of doctoral studies at the Department of Natural Sciences, Faculty of Biomedical Engineering, Czech Technical University in Prague.

Applicant: Ing. Lucie Mrázková

Department of Natural Sciences
Faculty of Biomedical Engineering, CTU in Prague
Nám. Sítná 3105, 272 01 Kladno

Supervisor: doc. Marie Lipoldová, CSc.

Department of Natural Sciences
Faculty of Biomedical Engineering,CTU in Prague
nám Sítná 3105, 272 01 Kladno

Laboratory of Molecular and Cellular Immunology (LMCI)
Institute of Molecular Genetics, CAS
Vídeňská 1083, 142 20, Praha 4

Supervisor-Specialist: RNDr. Taťána Jarošíková, CSc.

Department of Natural Sciences
Faculty of Biomedical Engineering, CTU in Prague
nám Sítná 3105, 272 01 Kladno

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 (FBME, 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.

List of abbreviations

A: adenine
ARMS: amplification refractory mutation system
BLAST: basic local alignment search tool
bp : base pair
BWA: Burrows-Wheeler aligner
C: cytosine
DGGE: denaturing gradient gel electrophoresis
DNA: deoxyribonucleic acid
ELISA: enzyme linked immunosorbent assay
G: guanine
HapMap: haplotype mapping
HGP: human genome project
HUGO: human genome organization
LDA: linear discriminant analysis
LefSe: Linear discriminant analysis Effect Size
MISA: MicroSAtellite
NCBI: national center for biotechnology information
NGS: next-generation sequencing
OTUs: operational taxonomic units
PCoA: principal coordinate analysis
PCR: polymerase chain reaction
PGM: personal genome machine
QIIME: quantitative insights into microbial ecology
QTLs: quantitative trait loci
RFLP: restriction fragment length polymorphism
SNP: single nucleotide polymorphism
STRs: short tandem repeats
T: thymine
TBE: Tick-borne encephalitis
TBEV: tick-borne encephalitis virus
TRF: tandem repeats finder
WGS: whole genome sequencing

1. CURRENT STATE OF KNOWLEDGE

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. As technology advances, so do expectations of the methods used in basic and applied research. New methods must now meet higher standards in terms of safety, user-friendliness, robustness of data, consistency and reproducibility of results. In addition, increasing attention is being paid to the environmental impact of the chosen methods.

Short tandem repeats (STR)

STRs, so-called microsatellites, are DNA sequences with repeating units of A, T, G and C, ranging from 1 to thousands of bp. The most important types are 1-6 bp long motifs [1, 2], with dinucleotide repeats such as (CA)n being the most common. Most STRs are located in non-coding DNA [3], but many human [4], mouse [5] and plant [6] genes have STRs in their open reading frames, including promoters [7].

Eukaryotic genomes contain a large number of STRs [8, 9, 10, 11, 12]. Microsatellites are highly variable due to differences in length. Their occurrence and variability, together with PCR amplification, enabled genetic mapping and gene tagging [13, 14, 15]. The first dense genetic maps of the human [16] and mouse genomes [17] based on STRs were published in 1996. Modern techniques have increased STR density in the human [18]

and mouse [19] genomes, facilitating disease susceptibility studies. Identifying the causative genetic mutations is crucial for the treatment of disease. Genotyping techniques are the same for all species, but primers and probes are species-specific. Correlations between genotype and phenotype are evaluated in animal models and identified associations are tested in humans after back-translation [20].

Microsatellites can be identified from sequence data using various computer programmes such as Tandem Repeats Finder [3, 21], MicroSAtellite [22] MsDetector [23], WebSTR [24], Lu developed STR toolkit [25] and many others [26, 12]. After identification, the flanking DNA sequences can be analyzed for suitable PCR primers to test the STR loci. Many computational tools are available to identify STRs and design PCR primers for specific loci [27, 28, 29, 30].

Data on STR sequences can be obtained from various public databases. STRBase (http://www.cstl.nist.gov/strbase/) provides information on human sequences [31], 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 [32]. 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 suitable for variety of purposes, including gene

mapping, point mutation detection, marker-assisted breeding of congenic mouse strains, and many other applications.

Genotyping

A genotype is the combination of alleles that are inherited from each parent and form a unique genetic makeup [20]. Genotyping determines these DNA sequences, referred to as genotype, at positions within an individual's genome [33]. Early methods such as Southern blotting [34] and DNA sequencing [35, 36] led to the development of genotyping tools in the 1970s. Initially, restriction fragment length polymorphisms (RFLPs) were used, but the process was labor intensive, time consuming and radioactive labeling was required [20]. The polymerase chain reaction (PCR), introduced in 1985 [37, 38], revolutionized genotyping by allowing unlimited amplification of DNA segments [20, 37, 38, 39].

In 1988, the Human Genome Organization (HUGO) was founded to sequence the entire human genome [40]. The first human genome sequences were published in 2001 by the International Human Genome Project [1] and Celera Genomics [41]. The mouse genome was sequenced in 2002 [19]. These milestones led to projects such as HapMap [42] and the 1000 Genomes Project [43]. With the advent of high-throughput DNA sequencing in 2005 [20, 44, 45] and advances such as long reads [46] and DNA microarrays [45], sequencing costs have been significantly reduced [47]. Whole genome sequencing (WGS) now enables the complete determination of the genome

in a single experiment and offers the highest resolution in genotyping [20].

DNA variations, known as nucleotide polymorphisms, are the key to understanding genetic diversity and linking DNA variants to phenotypic traits. In animal research, genotyping is used to identify animals for breeding and to study genetic factors in various diseases [33]. Single nucleotide polymorphisms (SNPs) are valuable markers for population genetics and disease mapping due to their high density and mutational stability. The tetraprimer ARMS PCR method [39, 48], which includes PCR and gel electrophoresis, was a breakthrough for SNP genotyping, despite the risks associated with working with polyacrylamide gel [49, 50]. The results of this method were consistent with those of restriction fragment length polymorphism (RFLP) analysis [39, 48].

We were looking for a cost-effective, low-risk method with a throughput resolution of around 6 base pairs (bp), which 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 [51].

Summary

Next-generation 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 [20, 45]. DNA microarrays can be used to genotype thousands of markers simultaneously, but their use

for mass testing is very expensive [20]. 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) [20]. 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 [51].

2. AIMS OF THE THESIS

The aim of this work is to deepen the understanding of molecular biology methods as a tool for biomedical technology. Engineering applies scientific and mathematical principles to evaluate results, while molecular genetics studies the structure and function of genes. New methods could revolutionize genetic engineering and gene mapping and provide insights into the development and progression of various infectious diseases.

Goals 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).

3. PROCESSING METHODS

Genotyping of short tandem repeats

The presented genotyping method is based on DNA isolation, PCR amplification, optimized high-resolution electrophoresis and detection of the results (Figure 1), 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. A detailed description of the methods used can be found in chapter 4.2 of the dissertation and are published in detail in the genotyping protocol [51].

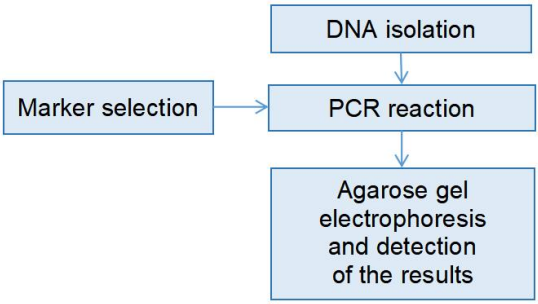

Figure 1: Diagram of the methods used in the genotypization method

The complex comparative microbiome study

The study consisted of a pilot microbiome study (Figure 2) and a main microbiome study (Figure 3). In the pilot microbiome study (Figure 2), not all experimental strains were processed and different methods were used to determine the microbiome composition than in the main microbiome study. 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 were obtained in FASTA format and processed using Geneious R9 software [52]. Phylogenetic trees were generated from the processed sequences using the BLAST® algorithm of the NCBI database [53], which are presented in the results of the dissertation. 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.

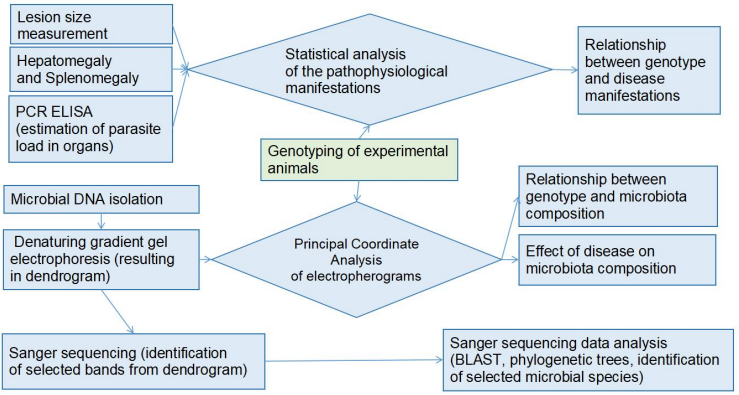

Figure 2: Diagram of the methods used in the pilot microbiome study

Based on the results of the pilot study, a detailed main microbial study was carried out using more expensive methods (Figure 3). 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 (using platform QIIME 2 2020.2 pipeline [54] 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 among different infected or uninfected strains of mice. The results of the linear discriminant analysis (LDA) LEfSe [55] 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 is included in chapter 4.2 of the dissertation and the details can be found in the article [56].

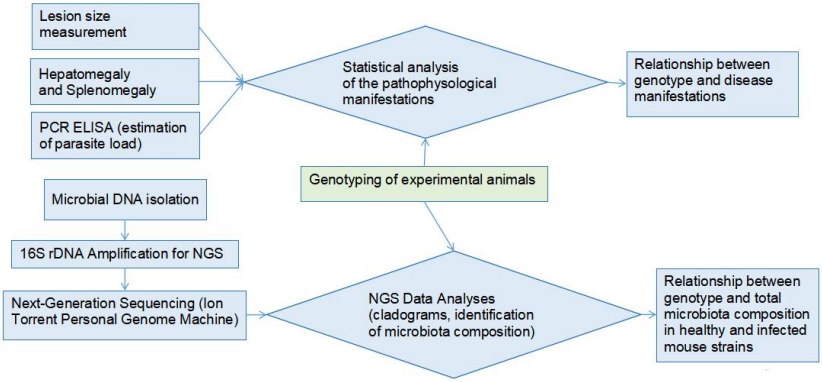

Figure 3: Diagram of the methods used in the main microbiome study

Mapping and identification of candidate genes

The candidate genes were mapped and identified using the methods shown in (Figure 4). First, 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 in chapter 4.2 of the dissertation or in a published article [57].

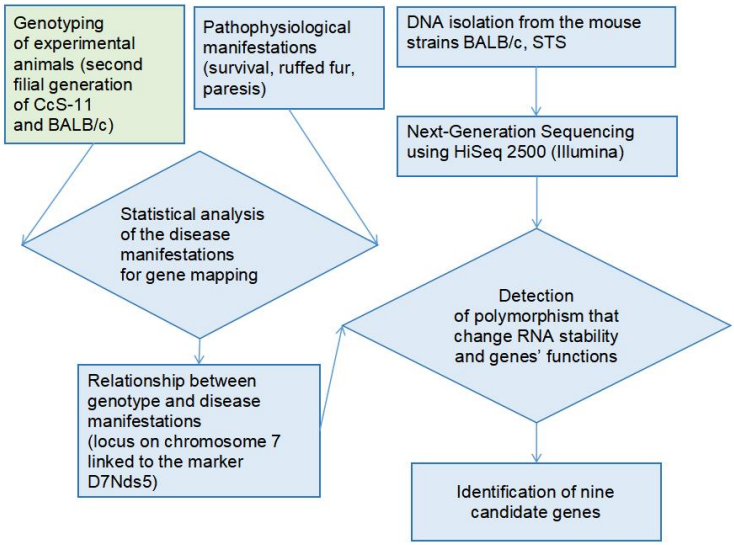

Figure 4: The schematic representation of the methods used in the mapping and identification of candidate genes

4. RESULTS

The improved method for genotyping

A new method for genotyping was introduced. The development of the method started in 2009 [58], where the proto-method was used for the first time. The method was improved in 2011 [59] before it was finalised in 2015 [51]. The method consists of DNA isolation, PCR reaction, agarose gel electrophoresis stained with ethidium bromide and subsequent visualisation 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 [51]. We conducted two independent experiments, which are described below as a proof of concept for 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 [56]. 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 [57].

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 5).

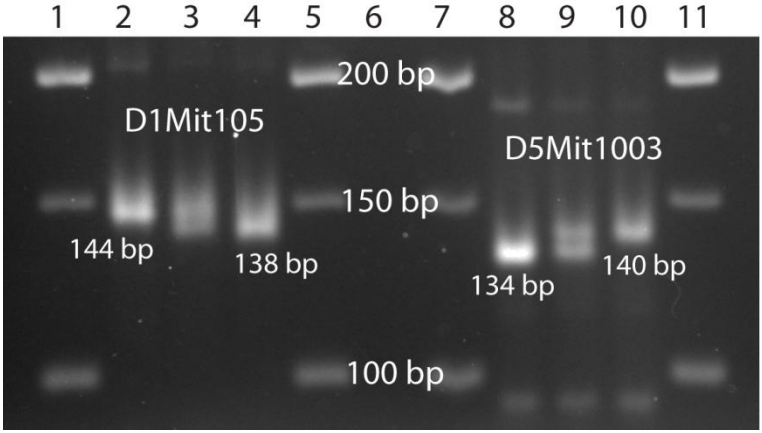

Figure 5: DNA typing using STR markers with PCR products length difference 6 bp. Picture was captured by author on GelDoc system (Bio-Rad): 1 - 50 bp ladder, 2 - BALB/c homozygote in the marker D1Mit105 (144 bp), 3 - heterozygote, 4 - STS homozygote (138 bp), 5 - 50 bp ladder, 6 - empty, 7 - 50 bp ladder, 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 [51]

The method has been successfully used to map mouse genes controlling susceptibility to *L. major* [60, 61], to map mouse genes controlling susceptibility to tick-borne encephalitis [57] and in the experiment focused on studying the microbiome to control inherited alleles in selected mouse strains [56].

The presented method is safe, robust, 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. The details of this protocol can be found in Chapter 5.1 of the dissertation and in the published protocol [51]. Proof of concept of this new genotyping method not only confirmed its usefulness, but also led to the novel original findings in studies of genetic impact on development of leishmaniasis and encephalitis. The results of all experiments can be found in Chapter 5 of the dissertation, the results of the main study on the microbiome are published in the article in Applied Microbiology and Biotechnology [56] and the results of the study on gene mapping and identification are published in the article in BMC Neuroscience [57].

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. According to publications Wu (2019), Meng (2021), Wang (2020), Zhao (2019) and Kim (2022) [62, 63, 64, 65, 66] 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.

5. 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.

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 Jam H 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 et al. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu Rev Gene*t 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 et al. 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 et al. 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 et al. 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 et al. 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 GF et al. 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 et al. 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] Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res* 17, 6463-6471 (1989).

[14] Love JM et al. Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites. *Nucleic Acids Res* 18, 4123-4130 (1990).

[15] Weber JL and May PE. “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.
[16] 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.
[17] Dietrich WF et al. A comprehensive genetic map of the mouse genome. Nature. 1996;380:149–52.
[18] 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.
[19] 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
[20] Kockum, I., Huang, J., & Stridh, P. (2023). Overview of genotyping technologies and methods. *Current Protocols*, 3, e727. doi: 10.1002/cpz1.727
[21] 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
[22] 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).
[23] 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).
[24] Lundström OS, Adriaan Verbiest M, Xia F, Jam HZ, Zlobec I, Anisimova M, Gymrek M. 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.
[25] Lu J, Toro C, Adams DR; Undiagnosed Diseases Network; Moreno CAM, Lee WP, Leung YY, Harms MB, Vardarajan B, Heinzen EL. 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.
[26] 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.
[27] Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289-1291, doi:10.1093/bioinformatics/btm091 (2007).
[28] Untergasser, A. et al. Primer3--new capabilities and interfaces. *Nucleic Acids Res* 40, e115, doi:10.1093/nar/gks596 (2012).

[29] 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).
[30] 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).
[31] 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).
[32] 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).
[33] Jacquot, S., Chartoire, N., Piguet, F., Hérault, Y., & Pavlovic, G. (2019). 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. doi: 10.1002/cpmo.65
[34] Southern E. Gel electrophoresis of restriction fragments. Methods Enzymol. 1979;68:152-76. doi: 10.1016/0076-6879(79)68011-4. PMID: 232210.
[35] 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.
[36] 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.
[37] 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.
[38] 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.
[39] Holloway, J. W., Beghe, B., Turner, S., Hinks, L. J., Day, I. N. M., & Howell, W. M. (1999). 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. doi: 10.1002/(SICI)1098-1004(199910)14:4<340::AID-HUMU10>3.0.CO;2-Z. PMID: 10502782.
[40] 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.
[41] 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.

[42] International HapMap Consortium. The International HapMap Project. Nature. 2003 Dec 18;426(6968):789-96. doi: 10.1038/nature02168. PMID: 14685227.
[43] 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
[44] 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.
[45] 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.
[46] 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.
[47] https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost (Assessed August 11, 2024)
[48] 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.
[49] 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
[50] 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.
[51] 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
[52] Geneious at https://www.geneious.com
[53] NCBI database at https://blast.ncbi.nlm.nih.gov
[54] 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.
[55] LEfSe at https://huttenhower.sph.harvard.edu/lefse/
[56] 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
[57] 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.
[58] 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
[59] 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.
[60] 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.
[61] 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.
[62] 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.
[63] 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.
[64] 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.
[65] 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.
[66] 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

Publications of the author on the topic and the grant support

1. Journal articles

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. (2018) A novel locus on mouse chromosome 7 that influences survival after infection with tick-borne encephalitis virus. BMC *Neuroscience* 19(1):39. doi: 10.1186/s12868-018-0438-8.

2. Online protocol

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

3. 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.

4. Grant 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”).

SUMMARY

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.

RESUMÉ

Biomedicínské inženýrství hraje klíčovou roli ve vývoji a zlepšování metod používaných v molekulární biologii. Tyto metody umožňují práci s biologickým materiálem a jeho analýzu. Tím tyto metody přispívají k pokroku nových terapeutických strategií.

Tato práce se zabývá diagnostickou částí biomedicínského inženýrství a zahrnuje vývoj inovativních přístupů a metod pro prevenci a diagnostiku nemocí. Klíčovým prvkem při detekci genetického základu nemocí je stanovení genotypu (genotypů) spojených s vnímavostí k nemocem. V této studii představujeme nově modifikovanou, robustní, levnou a bezpečnou metodu pro genotypizaci. Identifikovali jsme podmínky, za kterých lze zvýšit rozlišení genotypizační metody až na rozdíl 6 nukleotidů. Provedli jsme důkaz konceptu této metody ve studiích pro dvě různá infekční onemocnění: leishmanióza a klíšťová encefalitida (TBEV).

Vybrali a genotypovali jsme experimentální populace ve studii mikrobiomu spojeného s leishmaniózou. Tato analýza potvrdila genetické pozadí myších kmenů, které se liší svou vnímavostí k infekci parazity *Leishmania major*, a umožnila nám zhodnotit vliv genotypu a infekce na složení mikrobioty. Mikrobiální komunity v ileu a tlustém střevě testovaných kmenů byly porovnány pomocí lineární diskriminační analýzy efektu velikosti (LefSe) a shlukové analýzy. Ukázali jsme, že infekce parazity rodu *Leishmania* ovlivňuje hlavně mikrobiotu ilea, zatímco bakteriální komunita v tlustém střevě byla stabilnější. Získané výsledky naznačují rozdíly ve složení mikrobioty mezi rezistentními a vnímavými kmeny.

Tuto genotypizační metodu jsme také úspěšně použili pro mapování nového lokusu na chromozomu 7, který kontroluje přežití po infekci TBEV. Kombinace bioinformatiky a systémové analýzy založené na definici vzorců genové exprese, klasifikaci jednotlivých genů do ontologických drah a použití specifických

genetických polymorfismů ovlivňujících nemoc vedla k detekci devíti kandidátních genů v tomto lokusu.

Prokázali jsme užitečnost vylepšené metody genetického typování ve studiích náchylnosti k nemocem. Získané výsledky by mohly mít velký potenciál pro translační medicínu a tím pro zlepšení lidské zdravotní péče.