DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
406

OLIVER AASMETS

The importance of microbiome in human health

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
406

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
406

OLIVER AASMETS

The importance of microbiome in human health

UNIVERSITY OF TARTU
Press
1632

Institute of Molecular and Cell Biology, Institute of Genomics, University of Tartu, Estonia

This dissertation is accepted for the commencement of the degree of Doctor of Philosophy in Molecular Biomedicine on September 12, 2022 by the Council of the Institute of Molecular and Cell Biology, University of Tartu.

Supervisor: Elin Org, PhD
Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia

Reviewer: Prof. Maido Remm, PhD
Chair of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu, Estonia

Opponent: Thomas Sebastian B. Schmidt, PhD
European Molecular Biology Laboratory

Commencement: Room No. 105, 23B Riia St., Tartu, on October 17, 2022, at 14:15.

The publication of this dissertation is granted by the Institute of Genomics and Institute of Molecular and Cell Biology at the University of Tartu, Estonia.

This research was funded by Estonian Research Council grants PUT 1371, PRG687 and PRG1076; EMBO Installation grant 3573; European Regional Development Fund project no.15-0012 GENTRANSMED; Estonian Center of Genomics/ Roadmap II project no.16-0125; Archimedes Foundation scholarships in smart specialization growth and Dora Plus scholarship for short-term study mobility; the Doctoral School of Biomedicine and Biotechnology scholarship. Data analyses were in part carried out at the High-Performance Computing Center of the University of Tartu.

ISSN 1024-6479 (print)
ISBN 978-9916-27-023-3 (print)
ISSN 2806-2140 (pdf)
ISBN 978-9916-27-024-0 (pdf)

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, referred to in the text by Roman numerals (Ref. I to Ref. III):

I **Aasmets, Oliver**; Lull, Kreete; Lang, Jennifer M.; Pan, Calvin; Kuusisto, Johanna; Fischer, Krista; Laakso, Markku; Lusis, Aldons J.; Org, Elin (2021). Machine Learning Reveals Time-Varying Microbial Predictors with Complex Effects on Glucose Regulation. mSystems, 6 (1), ARTN e01191-20. https://doi.org/10.1128/mSystems.01191-20.

II **Aasmets, Oliver***; Krigul, Kertu Liis*; Lüll, Kreete; Metspalu, Andres; Org, Elin (2022). Gut metagenome associations with extensive digital health data in a volunteer-based Estonian microbiome cohort. Nature Communications, 13 (1), 1–11. https://doi.org/10.1038/s41467-022-28464-9.

III **Aasmets, Oliver**; Krigul, Kertu Liis; Org, Elin (2022). Evaluating the clinical relevance of the enterotypes in the Estonian Microbiome cohort. Frontiers in Genetics, 1943. https://doi.org/10.3389/fgene.2022.917926.

My contributions to the listed publications were as follows:

Ref. I Designed the study, performed the data analysis, and wrote the manuscript.

Ref. II Performed the data analysis, interpreted the data, prepared the figures, and wrote the manuscript.

Ref. III Designed the study, performed the data analysis, interpreted the data, prepared the figures, and wrote the manuscript.

LIST OF ABBREVIATIONS

CT	Community Type
DA	Differential Abundance
EstBB	Estonian Biobank
EstMB	Estonian Microbiome Cohort
ET	Enterotype
FINRISK	Finland Cardiovascular Risk Study
FMT	Fecal Microbiota Transplantation
HbA1c	Glycosylated Hemoglobin
METSIM	METabolic Syndrome In Men
ML	Machine Learning
OGTT	Oral Glucose Tolerance Test
qPCR	Quantitative Polymerase Chain Reaction
RMSE	Root Mean Square Error
SCFA	Short Chain Fatty Acids
TMAO	Trimethylamine N-oxide
T2D	Type 2 Diabetes
WGS	Whole Genome Sequencing
WHO	World Health Organization
rRNA	Ribosomal Ribonucleic Acid

INTRODUCTION

The technological revolution has provided us an opportunity to study the world beyond the limits that were holding us back only a couple of decades ago. One of such fields enjoying increasing popularity is the study of the human microbiome. Tiny microorganisms making up the microbiome, such as bacteria and viruses have been known to intervene with our health for centuries, but the ecosystem of the human microbiome and their functionality has turned out to be perhaps more complex than previously thought. It is rather clear though that the extent of the role of the microbiome to our own functioning and well-being is just starting to unravel. Nevertheless, microbiome has already been associated with a large variety of intrinsic and extrinsic factors, including various complex diseases. The accumulating evidence is leading a slow but steady progress towards clinical applications, which again feeds the public interest and popularity of the field with never ending gasp for new knowledge.

This thesis uses the value of the comprehensive phenotyping data from the Estonian Biobank and Finnish METSIM (METabolic Syndrome In Men) cohort to expand the understanding of the factors influencing our microbiome composition and assesses the possibility and challenges in using the microbiome composition for clinical applications. In the first part I will take the reader into the world of human microbiome by summarizing the key knowledge from the scientific literature. I will introduce the elements that give rise to the variability of the gut microbiome profile and describe the role of microbiome in human health together with a few examples about how microbiome can be used to improve our well-being. Additionally, I will introduce some challenges that need to be accounted for when studying the human microbiome. In the second part I will present the results of the three original publications. I will show that the gut microbiome can be used to predict changes in glucose regulation, which is important for handling the increasing prevalence of type 2 diabetes. Further, I will demonstrate, how population-based biobanks and especially electronic health records can expand our knowledge about the gut microbiome.

1. REVIEW OF THE LITERATURE

1.1. Human as a superorganism

We are not alone. Even without giving context, the statement can be considered true, because we are surrounded, inhabited, and vastly outnumbered by microorganisms in every way. These microorganisms include predominantly bacteria, viruses, fungi, and archaea and are collectively termed "microbiome", which refers to their collective genome. Perhaps the perception about life around us is easier to grasp, but surprisingly microbial life analogously dominates our body considering the number of cells and genes we carry each day. The most recent estimates show that for each human cell, we have about 1.3 microbial cells and for each human gene, around 100 microbial genes, which in total make up approximately 200 grams of our body weight (Sender et al., 2016; Turnbaugh et al., 2007). This is likely a surprise for many, considering that a negative scent has often been laid on bacteria and viruses ranging from children's books to advertisements, characterized for example by "Karius and Bactus" by Thorbjörn Egner. However, we live in a microbial world, not *vice versa*.

It is natural to assume that the microbial genes carry out a function in our body similarly to differences in human genome can correspond to differences in eye color and height. Thus, a reasonable question follows – does microbiome have a part in why we are different and why will one person develop a disease and the other does not. This is a natural starting point for human microbiome studies.

1.1.1. Landscape of the human microbiome

Characterization of the microbiome composition is a rather new skill for the human. In the early phases of the research, cultivation-based techniques were used to isolate and study certain bacteria. However, in 1996, a sequencing-based characterization of the fecal microbiota was first performed that revolutionized the field (Wilson & Blitchington, 1996). Sequencing allowed to analyze uncultured, anaerobic taxa, which was previously inconceivable. Although today the technological advantages do allow the characterization and cultivation of anaerobic bacteria (Lagier et al., 2018), the sequencing still forms the backbone of the research. Moreover, sequencing widens the scope by allowing to characterize viruses and other members of the ecosystem. As a result, the next-generation sequencing techniques are widely used and continuously developed. Two common strategies for sequencing the microbiome must be highlighted. Firstly, amplicon sequencing aims to sequence only a specific region of the microbial genome, such as the 16S rRNA gene. In contrary, "shotgun sequencing" aims to sequence the whole microbial genome. These two strategies have predominantly relied on reading small fragments of the DNA using so called short-read sequencing, but the field is already looking towards long-read sequencing (Karst et al., 2021; Tedersoo et al., 2021). Thanks to these technological breakthroughs,

we have obtained quite some knowledge about the human microbiome, but undoubtedly, we are still at the beginning of the journey. Here is what we do know.

Like distant geographic regions of the world are inhabited by different mammalian species, the microbiome composition is different depending on the body site we examine. For example, there is a unique composition on the skin, in the nasal cavity, in the gut and in the oral cavity (Huttenhower et al., 2012). Thus, differences in the availability of resources, such as oxygen, flow of nutritional ingredients and other environment properties such as pH and temperature give rise to differences in the microbial composition. In an ideal world with limitless research interest and funding, the different body sites would be studied in parallel to really gasp the idea of human as a "superorganism", which could be used to improve health care and clinical applications. However, with limited resources, the focus has so far concentrated primarily on the gut, skin, and oral microbiomes, which are the easiest to sample. The body sites can additionally have local sub-communities. For example, sections of the gut have different properties in terms of oxygen content and pH and therefore also different microbial content, with the large intestine being the most densely populated part of the colon (Leshem et al., 2020). As the stool sample largely characterizes the microbiome composition of the distal part of the gastrointestinal (GI) tract, it has alternatively referred to as "fecal" microbiome to not oversimplify the gut. It is noteworthy that the performance of microbiome-based applications might differ depending on the body site examined. For example, the gut microbiome has shown to provide better discriminability for pancreatic ductal adenocarcinoma detection when compared to saliva samples (Kartal et al., 2022). Therefore, although ideally the human microbial landscape should be studied as a whole, this thesis focuses on the microbiome, which is characterized by sampling stool and referred to as "gut microbiome" from now on. Nevertheless, it is necessary to emphasize that similar questions can be asked about the other body sites.

Let's change scales. There can also be remarkable differences in the microbiome composition when different human populations are compared. The extremes of this phenomena are said to follow western and non-western lifestyles. The most well-known example is about the Hadza hunter-gatherers who display a remarkably different gut microbiome when compared to western populations (Schnorr et al., 2014). It is clear that the Hadza live in a quite a different environment and follow a different lifestyle than an average westerner does, which is also reflected by the microbial composition. Moreover, the Hadza microbiome exhibits remarkable seasonal differences due to the availability of food, a property less evident for other populations (Smits et al., 2017). Differences in microbiome composition have even been shown between agricultural and urban regions within populations (Ayeni et al., 2018). Due to the large differences in the microbiome of western or "industrialized" populations, concerns have been raised about the "lost diversity" and even "rewilding" has been called upon to rediversify our microbiomes (Blaser, 2018). However, it is unclear, how these observations and suggestions translate to different populations. For example, Estonians have been largely dependent on "porridge and potatoes" for centuries and the diet is

far from the one of the Hadza. Adding that the horizontal and vertical transfer of the microbial strains over generations can yield demographic signatures with possibly population-specific health benefits (Suzuki et al., n.d.), population-based cohorts become fundamental to expand our knowledge and establish baselines and checkpoints for the future.

The story becomes even more complex. One of the most fascinating properties of the human microbiome is that it is not a characteristic we either have or don't have. Our gut microbiome is a complex ecosystem consisting of around 160–400 species (Lloyd-Price et al., 2016) and the system is in continuous movement. Although a topic of active discussion, we are not colonized by microbes before birth (de Goffau et al., 2019; Enav et al., 2022). The first microbes we obtain are usually from the birth canals and vaginal fluids and thereafter we are continuously being colonized by microbes from the environment, diet and social interactions (Enav et al., 2022). During the first years of life, the gut microbiome continues maturation and at 3–4 years, the microbiome starts to resemble the one of an adult (Derrien et al., 2019). But the composition isn't "ready", it is still open to development and changes. A large number of factors have been shown to be associated with the gut microbiome or shape the gut microbiome composition throughout life. For example, medication usage (S. K. Forslund et al., 2021; Jackson et al., 2018; Maier et al., 2018; Zhernakova et al., 2016), physical activity (Dziewiecka et al., 2022), host genetics (Kurilshikov et al., 2021), smoking (Gui et al., 2021), alcohol consumption (Segovia-Rodríguez et al., 2022) and time of the day (Nobs et al., 2019) among many other factors as summarized in **Figure 1**. Importantly, various diseases have been associated with the gut microbiome, which will be discussed closely in the following chapters. These results are to some extent expected as the human gut microbiome is a dynamic and open ecosystem, which needs to adapt to the changing environment. Largely due to this property, the human microbiome is highly personalized, and the microbiome composition can go through relatively large temporal changes (Olsson et al., 2022; Vandeputte et al., 2021). This dynamic nature turns out to be both, a curse, and a blessing for microbiome science.

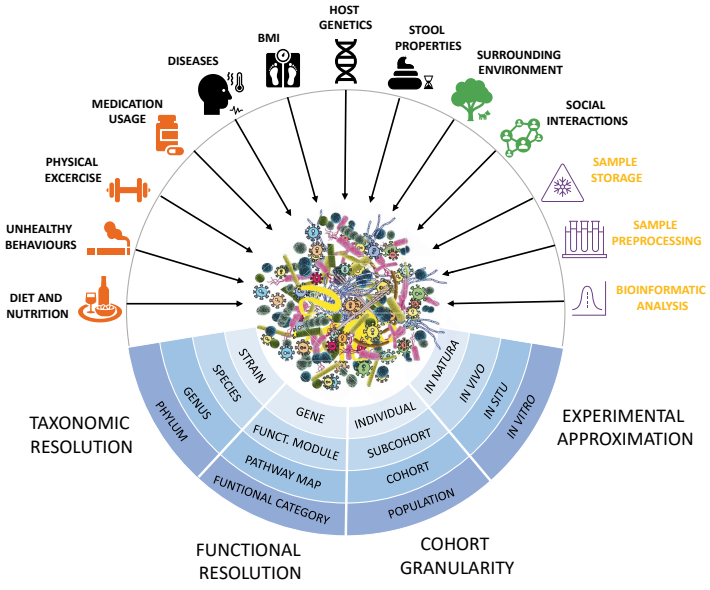

Figure 1. Intrinsic and extrinsic factors influencing the microbiome composition and considerations for study design in microbiome studies.

1.1.2. Microbiome in human health

There is nothing new in the history of microbiome research focusing on using the microbiome for inferring and treating health conditions. The first implications date back to the second part of the 19th century, when pioneers such as Pasteur and Koch carried out their landmark research. Pasteur's research, in addition to trying to improve the quality of French beer, led to the disapprovement of the "spontaneous generation" theory and development of the vaccine for rabies. Koch was able to show that in addition to viruses, bacteria can cause diseases, with examples about tuberculosis, gonorrhea and tetanus among many others confirmed by him or his coworkers (Keen, 2007). His postulates for determining the causal effect of microbes to diseases stand today, in the era of computational biology (Neville et al., 2018; Singh et al., 2016). Thus, the field has gone a long way considering that we have just learned that microscopic life doesn't come from thin air. However, we are only now starting to understand the applicational potential and role of the whole microbiome composition in complex diseases.

Modern times yield modern diseases and there are a few that have gained a lot of attention from the microbiome perspective. The common denominator of the microbiome-associated diseases is usually a significant rise in the prevalence during the last century (Bach, 2002). These diseases include metabolic diseases such as type 2 diabetes (Gurung et al., 2020; J. Wang et al., 2012; Wu et al., 2010) and obesity (Alili et al., 2021), gastrointestinal disorders such as colorectal cancer (Wirbel et al., 2019; Zeller et al., 2014), Crohn's disease and ulcerative colitis

(Kostic et al., 2014; Pascal et al., 2017), but also psychiatric disorders like anxiety disorder and depression (Simpson et al., 2021). Such diseases are often combined under a term "complex diseases" because they are caused by a combination of genetic, environmental and lifestyle factors. Likewise, a key conclusion from the microbiome studies focusing on the complex diseases is that they don't have a single clear responsible member of the microbiome that is causal for the disease but rather there are changes in the whole community or subcommunities that contribute to the disease development. To indicate and cover this complexity in disease-related microbial signals, the term "microbial dysbiosis" is often being used, however argued to be too vague (Shanahan et al., 2021; Shanahan & Hill, 2019). We don't yet know, how to define neither a "healthy" nor "unhealthy" microbiome – various states can refer to similar outcomes. Therefore, we firstly need come up with a baseline to define the dysbiotic state.

The motivation for further health-focused microbiome studies, however, is not based on mere associations and correlations. It is established that the gut microbiome carries out fundamental functions for our physiology. The presence of an active and diverse ecosystem itself protects us from the colonization of pathogens (Belkaid & Hand, 2014). The gut microbiome is responsible for the fermentation of non-digestible carbohydrates such as dietary fibers, production of a vast array of metabolites such as vitamins and short chain fatty acids (SCFAs) (Krautkramer et al., 2021; Morrison & Preston, 2016) and serving as an important mediator for our immune system (Belkaid & Hand, 2014). Nevertheless, the understanding about the functions gut microbiome carries out is constantly updated (Shine & Crawford, 2021). The microbiome holds a great metabolic potential and there is active research going on.

Therefore, our microbial pals are not mere passengers, but we are depending on them. Taken together, an intriguing concept forms: we are born as empty vessels, and we need to form ourselves a symbiotic microbial community that protects us and benefits us. However, if something goes wrong, health complications may follow. Considering that during the last few generations a lot has changed in the society, for example the proportion of vaginally born children is decreasing and the consumption of processed foods and drugs is increasing, we might run into trouble (Sonnenburg & Sonnenburg, 2019). These are remarkable times for microbiome research. We have technology to study the microbiome in a rapidly changing world and we are fueled by the already existing knowledge. Besides satisfying our desire for knowledge, the microbiome science can provide a new perspective for improving our health and clinical practice, which will be discussed in the following sections.

1.2. Microbiome in real-world applications

1.2.1. Microbiome for disease diagnostics and prognostics

Using the microbiome to improve the decision making in clinical applications is a natural next step following the accumulation of evidence for its biological mechanisms and disease associations.

Firstly, microbiome can be considered a marker for diagnosing a disease. It is especially reasonable to bear in mind the diseases that are difficult to diagnose. A notorious and probably most-studied of such diseases is colorectal cancer. The routine procedure for diagnosing colorectal cancer consists of detecting blood from the feces, which is followed by colonoscopy. The invasiveness of the colonoscopy is the logical target for microbiome-based alternatives: can we diagnose the disease by analyzing the fecal microbiome instead? It has been shown that there is a colorectal cancer specific microbial signal that can improve the detection of the disease when compared to traditional colorectal cancer risk factors (Wirbel et al., 2019). Similarly, microbiome can be used as a means to diagnose Pancreatic cancer (Kartal et al., 2022), and Crohn's disease and ulcerative colitis are also being actively researched in this regard (Kostic et al., 2014; Pascal et al., 2017). A related aim is to use microbiome for estimating the risk of developing a disease. The development of risk models based on microbiome data is currently restricted due to the availability of large-scale prospective cohorts. As one of the few, Finnish FINRISK cohort contains long follow-up data that has been used to identify microbial risk factors for cause-specific mortality, liver disease and type 2 diabetes (Y. Liu et al., 2022; Ruuskanen et al., 2022; Salosensaari et al., 2021). In other cases, the research is restricted by short follow-up or limited sample size (Gou et al., 2021; Leung et al., 2022).

There are several challenges to overcome before putting those models to work in clinical practice, some of which will be discussed in the next sections, but a principal problem needs to be highlighted. To accompany the development of the diagnostic and prognostic models with hope, trust, and cost-effectiveness, we need to be sure that the contribution of the microbiome is truly independent of the factors that are potentially easier and cheaper to measure. So far, only a few cost-effectiveness analyses have been carried out (Bajaj et al., 2020; Padula et al., 2020). Importantly, for many clinical applications we already have a "baseline" which microbiome analysis effectiveness can be compared to. For example, for diagnosing colorectal cancer, we are testing for blood in the fecal sample. So, the question is not whether the microbiome can be used to diagnose colorectal cancer, but whether the microbiome can provide some extra information to the fecal occult blood test to increase the sensitivity of the diagnostic test. Similar considerations must accompany all microbiome-based applications. For clinical trials, often not placebo is considered as a comparison for a new drug, but an existing drug or procedure that is the best performing intervention for the disease at the current moment. We should be aware of the simpler alternatives because microbiome today is not a simple alternative in those terms, although there are obvious benefits and successful use-cases.

1.2.2. Microbiome-guided health interventions

Perhaps the most influential microbiome-guided application for an individual today is personalized nutrition (Leshem et al., 2020; Shilo et al., 2022; Zeevi et al., 2015). Changing one's dietary preferences is possibly one of the easiest lifestyle interventions, which, turns out, can be individually directed to benefit health as needed. Luckily, this field has taken huge leaps in the last few years. The microbiome takes part in the degradation of multiple dietary ingredients, while being almost solely responsible for the degradation of dietary fiber (Cantarel et al., 2012). However, there is large inter-individual variability in the degrading capacity of the macro-nutrients, which can even lead to differences in energy harvest (Turnbaugh et al., 2006). Therefore, also the capability to produce the products of macronutrient degradation such as metabolites trimethylamine N-oxide (TMAO) and short-chain fatty acids (SCFAs), is personalized. This is a good motivation for the personalized nutrition as the SCFAs and TMAO are known to have physiological effects (Krautkramer et al., 2021). Interventional studies exposing our microbiomes to different diets have learned that our body responds differently to the same foods depending on the baseline microbiome composition (Suez et al., 2022; Zeevi et al., 2015). Successful applications have been built that leverage this concept to predict the changes in postprandial glucose levels, which can help to manage obesity and type 2 diabetes – two major threats to public health (Zeevi et al., 2015).

A rising field that displays similarities with the personalized nutrition approach is pharmacomicrobiomics – the study of how drugs work on a subject depending on the gut microbiome composition. It has been recognized that the microbiome can influence drug pharmacokinetics and pharmacodynamics in various ways (Zimmermann et al., 2021). Most alarmingly, microbiome can push a common drug to produce toxic metabolic subproducts, which result in adverse drug events (Zimmermann et al., 2019a). Similarly, the drug effects can be suppressed or even enhanced based on the available composition. For example, the response to statins has been shown to be more intense when there is an enrichment of the genus *Bacteroides* (Wilmanski et al., 2022). Remarkably, a systematic approach studying common host-targeted drugs showed that up to 65% of the drugs were metabolized by the microbial community (Zimmermann et al., 2019b). Thus, it can be expected that the current studies are just the start and future microbiome studies can advance the development of novel therapeutics, reduce the variation in drug-response and explain or even avoid adverse events. It is definitely one of the most exciting ways microbiome can be used to give a personal touch, however, the field is still in its infancy.

1.2.3. Interventions via changing the microbiome

As previously described, one possibility to take advantage of the gut microbiome would be to recognize the immense information and functional potential present and come up with applications that can benefit from it. This perspective is shared with the human genetics research, but the microbiome has a big advantage – it can be rapidly and easily modified. This is the beneficial part of the personalized and dynamic nature of the microbiome. Thus, if we would know what is "bad", we would then turn it to "good" and we are done with increased disease risks, we can optimize our drug consumption and consume the diet we like with preferrable outcomes. This is understandably an appealing perspective.

First option for the better could be to introduce beneficial microbes into the ecosystem or enhance the conditions of the already existent "good bacteria". Probiotics and prebiotics respectively are perhaps the most well-known 16ight16ent-tations of the two ideas. Probiotics are by definition live microorganisms that, when administered in adequate amounts, confer a health benefit on the host (Hill et al., 2014) and prebiotics is a substrate that is selectively utilized by host microorganisms conferring a health benefit (Gibson et al., 2017). Although numerous products on the market are advertised as probiotics or prebiotics, there are key challenges ahead for the field to make the next steps. For example, the major criticism is the lack of evidence for their favorable effect on health (Binda et al., 2020; McCoubrey et al., 2022). Currently, no probiotics have been licensed as medicines. Fortunately, the health claims are being addressed through ongoing clinical trials that follow the standard drug development pipelines, which are undoubtedly vital for the trustworthiness and development of the field in general (McCoubrey et al., 2022). Further considerations can be necessary. Conventional drugs are normally considered to dispense from the organism after there is no further addition of the molecule. However, probiotics can in theory permanently colonize and shift the whole composition, which therefore might have long-term consequences to the health when compared to standard drugs. Also, as with traditional drugs, individuals can respond to the probiotics differently. So, a moment of consideration – if we know that the human microbiome is highly personalized, can we assume that there is a uniform solution to a problem in a form of probiotics, prebiotics or postbiotics? An alternative solution can be to use fecal microbiota transplantation (FMT), which in essence aims to replace the whole existing community. Successful applications of FMT are already in clinical practice, oral FMT therapies are in stage III clinical trials (McCoubrey et al., 2022) and novel applications proposed. For example, FMT is currently the most efficient intervention to treating *C. difficile* infection (Quraishi et al., 2017).

Another solution for driving the microbiome towards the desired composition would be to eliminate the harmful ones. Antibiotics is the most well-known example of such approach. The conscious production and usage of antibiotics in medicine started in 1928, which led to the decrease in formerly deadly diseases. When earlier communicable infectious diseases were the main concern, non-communicable diseases became the main problem. Antibiotics have undoubtedly

made a revolution in medicine and Fleming truly deserved his Nobel prize. However, modern times raise modern problems and two major threats related to antibiotics usage have been raised. Firstly, there is a significant rise in the antibiotic resistance, referred to as the rise of the "super-bugs", which were associated with 4.5 million deaths worldwide and the problem is intensifying (Murray et al., 2022). Secondly, evidence is building that the antibiotics are associated with higher risk for complex diseases such as type 1 diabetes and asthma, with microbiome being the likely mediator (Blaser, 2016). Thus, the concept and practice of using antibiotics might need a reevaluation. One alternative possibility for fighting the pathogen would be to "targeted elimination" approach. Analogously, if antibiotics usually target a large proportion of the microbiome composition like FMT, bacteria-specific antibiotics or phages can be used to eliminate the pathogens (Federici et al., 2021), as probiotics are used for adding a few beneficial members.

Despite the many possibilities to nudge the microbiome composition, achieving stable changes tends to be difficult. The microbiome composition tends to return to its initial state resulting in a need for repeated manipulation. This is at least partly common for antibiotics, diet, probiotics and FMT (Leeming et al., 2019; Li et al., 2016; Palleja et al., 2018; Suez et al., 2019). Interestingly, probiotics may reduce the time of recovery after the antibiotic treatment and autologous FMT could be a better solution (Suez et al., 2018), further highlighting the complexities of community manipulation. Although it is agreed that probiotics do not have to colonize the gut, changing the "state" of the ecological composition is a whole another field, which will likely have a great influence in the upcoming years. Putting the pieces together – identification of the microbial signals responsible for the disease occurrence, unfavorable reactions after dietary or drug administration and making the necessary adjustments for changing the effects are largely unresolved. These challenges make the field highly interesting as every piece of new information builds the knowledge needed to master the human as a superorganism. We still have some work to do.

1.3. Challenges in microbiome research

Although the microbiome science is enjoying a lot of attention and novel findings are being reported daily, the field is open to challenges and pitfalls like any other field. There are challenges lurking in every step of the research: selection of the study design, data collection, sample preprocessing, bioinformatical analysis, statistical analysis, and interpretation of the results among many others. However, perhaps one of the first, often subconscious decision point in the research is whether there is intention to draw biological conclusions or is there some sort of application kept in mind, that just takes advantage of the extra information available. The choice leads to different paths with partly overlapping, but partly distinct challenges as the knowledge searched for is different. For example, we can try to understand, which bugs and how influence the progression of colorectal

cancer or, we might be interested, whether we can use the microbiome for improving the diagnostics of colorectal cancer. For the former, a more detailed characterization of the microbiome profile might be needed whereas the diagnostic application relies more on the homogeneity and standardization of the data collection and modelling process. Then, the correctness of the approach is a secondary priority – if it works for diagnosing a disease, the primary aim is fulfilled. For example, returning to the discussion about whether the microbiome obtained from stool should be called "fecal" or "gut" microbiome, this is again relevant for only the one of the frameworks. For applications, using some smart capsules to characterize the microbiome of the small intestine is and will likely be unfeasible for a while, especially if no gain in performance is obtained.
Often, however, no distinction is made and admirably the two ideals are simultaneously chased. The next sections aim to address some of the common challenges in microbiome research, with the focus towards prognostic modelling and biomarker detection. The aim of the following is to guide the reader into thinking about the relative standpoint a researcher takes and recognize the potential pitfalls in hope that the future studies can reduce the unknowns and mischiefs.

1.3.1. Considerations before the data analysis

The questions asked in research are largely driven by the study design and data availability. For example, even for the same question, the strength and extent of the claims made is understandably different for an interventional study and for a cross-sectional observational study. Perhaps an ideal study would collect highly precise information about one's health, lifestyle, and the surrounding environment in addition to microbiome samples from various body sites and other multiomics-data throughout the life, each day and minute. This would be done to infinite number of subjects who have gone through a myriad of various interventions. Unfortunately, it is a dream that will remain a dream. Therefore, it is necessary to consider that there are limitations depending on the study design and cohort granularity (**Figure 1**). For example, referring to increased disease risks has more firmness, when it is calculated on large number of incident cases when compared to a small sample of prevalent cases, where medication usage could shadow the disease effects. When weighting evidence, the study design has its part, and it must be accounted for. However, there is no rule of thumb for study design as they are all shadowed by feasibility and the ideal plan.

Microbiome studies have additional considerations and decisions that give context to the results. Firstly, the choices for sample collection and DNA extraction can also have a fingerprint on the results (Fernández-Pato et al., 2022; Shaffer et al., 2022). Secondly, the taxonomic or functional resolution for characterizing the microbiome composition needs to be chosen (**Figure 1**). The decision in favor of species-level resolution, enterotypes (Arumugam et al., 2011), functional profiles, or even strain level resolution and studying structural variation (Zeevi et al., 2019) in the microbiome can largely direct the next steps that need to be taken.

Such selection can be attributable to the aim of the study as mentioned before – whether the goal is to build biological knowledge or use it in an application that searches for cost-efficiency among other aspects. For example, strains within a species can have a varying functional output (Leimbach et al., 2013). Then again, it19ightt be more cost-efficient to use 16S rRNA sequencing if this is sufficient for the aim of the study. For example, the enterotypes based on genus-level taxonomic profile have been evaluated for directing personalized nutrition (Christensen et al., 2018). Either way, the decision can lead to the usage of 16S rRNA sequencing, shotgun metagenomics or quantitative PCR (qPCR) and the downstream analysis depends on the selection.

A challenge in microbiome studies has been the lack of reproducibility and replicability of the reported results. This includes results reporting biomarkers for various diseases, results showing how microbiome can improve the prediction of a diagnostic or prognostic models etc. For example, going through a database of microbe-disease interactions, one can easily stumble upon associations with effects in opposite directions for the same outcome. Moreover, every prevalent bacterial species and genus seems to be associated with a variety of diseases. The problem is complex and multifaceted. Answering the aforementioned challenges ranging from study design selection to biostatistics is part of the solution, but the first action point is more philosophical. In the era of "low-hanging fruits" and enthusiasm in microbiome research, optimistic bias is likely to pop up. Therefore, predefining the analysis plan as is done in clinical trials can alleviate the issues with repeatability and luckily such studies have become more prevalent (Eckermann et al., 2022; Sowah et al., 2022). The reasons might not be that the researcher consciously modified the results to match his/her interests, but it might be a subconscious tweak for the better. Thus, paying attention to the study design and sticking to the plan can save us. A parallel solution can be the propagation of "open science". However, although ideally the phenotype and microbiome data would be publicly available for the common good, the jurisdiction and data privacy issues together with the competitive edge will not likely cross the borders in near future.

1.3.2. Challenges in data analysis

Reaching from the initial material to sequences is not an easy task, but the story continues. Broadly speaking, the sequences need to be given a meaning and the variability in the microbiome composition needs to be associated with the phenotype or trait of interest.

The first step involves bioinformatic analysis. Although the world of microbiome bioinformatics covers way more than taxonomic assignment, this task will be discussed to illustrate the depth of the challenges. There is a rich set of tools and software that are up for the task and are advocated for and against by different researchers. Many have also aimed to benchmark these common taxonomic profilers such as mOTUs (Milanese et al., 2019), MetaPhLan (Beghini et al., 2021)

and Kraken (Wood et al., 2019), but the results have been conflicting (Sczyrba et al., 2017; Ye et al., 2019). However, the benchmarks can be misleading because the performance is evaluated uniformly for all profilers although the algorithms in principle seek to characterize different aspects of the composition (Sun et al., 2021). **Figure 2** summarizes the caveats that have been discussed so far. Even if we consider the most prominent bacterial species such as *Akkermansia muciniphila*, *Bacteroides uniformis*, *Faecalibacterium prausnitzii* and *Prevotella copri*, we can obtain large differences in relative abundance estimates when different sequencing techniques and bioinformatic tools are used. Thus, it is crucial to emphasize that the final results are always dependent on the previous steps taken. We are essentially building a framework in which we are working in and the results can depend on road chosen.

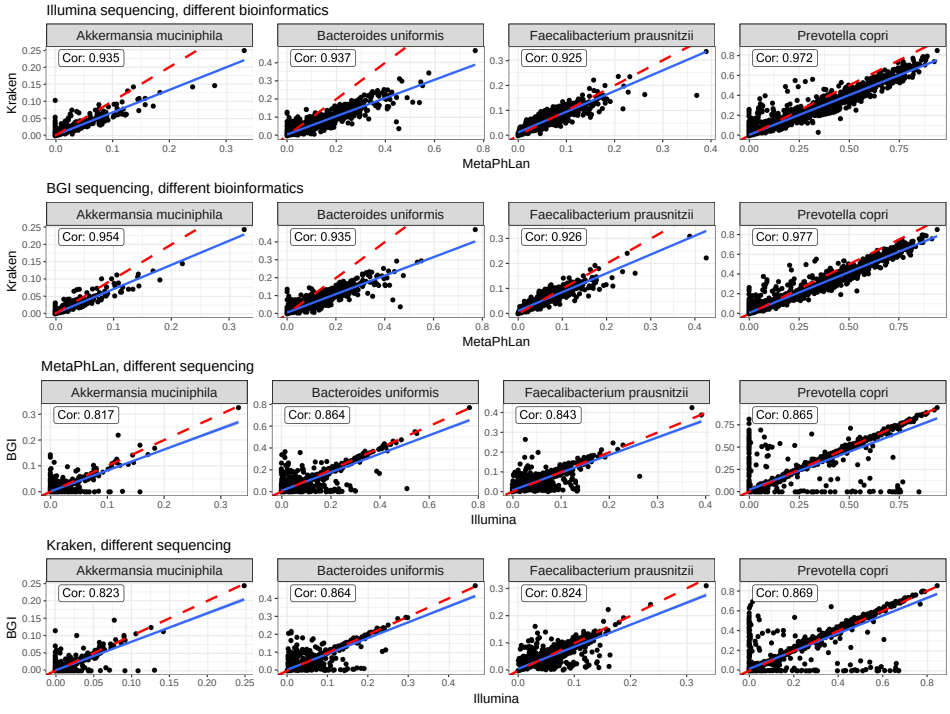

Figure 2. Comparison of the relative abundances of four gut commensals according to the sequencing method (DNBSEQ vs Illumina) and bioinformatics method used (Kraken vs MetaPhLan). The red line indicates the situation where x = y, the blue line represents the relationship in the data. Pearson correlation coefficient is shown. (*not published*)

Once reaching to the point where microbiome composition is characterized and quantified, the inter-individual variation of the composition can be linked to the trait of interest. The traditional strategies can be broadly divided into three groups based on the set of microbial features used for the association analysis (Bastiaanssen et al., 2022). Firstly, diversity analysis, a concept from general ecology, aims to characterize and analyze the whole microbiome composition

without paying direct interest on one single element of the composition. Similarly, clustering and stratification approaches such as "enterotyping" commonly take advantage of the information carried by the whole composition. Normally, the diversity analysis is followed by a differential abundance analysis that puts the focus on single species, functional pathway, or other univariate element of the composition, and associates it to the trait of interest. Thirdly, a subset of microbial features from the full set of composition elements can be used to derive more complex features. This might include data-driven features such as balances, amalgamations and (sub)networks of taxa, but also knowledge-based features such as ecological guilds and functional modules (Bastiaanssen et al., 2022). Each of these analyses associating microbial features with the trait come with their own challenges. For example, it is tricky to estimate microbial alpha-diversity (Willis, 2019), the beta-diversity analysis and clustering are dependent on the distance measure used for calculating the between-sample differences (Koren et al., 2013) and the differential abundance methods are showing disturbing inconsistencies (Nearing et al., 2022).

Differential abundance (DA) analysis in particular has become a topic of active discussion and development because of its great scientific and practical value. However, the DA algorithms come with remarkably different assumptions for declaring differential abundance, which are in practice rarely discussed. Hence, there is even a call to drop the idea and forget differential abundance analysis as it is impossible to test these assumptions (T. P. Quinn et al., 2021). Large fraction of the discrepancies can be attributed to the properties of the microbiome data. The relative abundance data available is mathematically defined as compositional data, which can lead to spurious results when not accounted for (Aitchison, 1982; Gloor & Reid, 2016; T. P. Quinn et al., 2019). For example, negative correlation bias arises due to the properties of the data, which renders the identification of correlations between the microbial abundances using traditional methods unreliable (Kurtz et al., 2015). Similarly, only some of the DA methods account for the compositional nature of the data. Luckily, compensating for the compositional nature of the data through log-ratio analysis has become more prevalent in the microbiome research. Nevertheless, although the mathematical properties of compositional data and possible pitfalls are thoroughly discussed, there are second thoughts about the impact of these problems due to the high dimensionality of modern microbiome data (Greenacre et al., 2022). Interestingly, there is a conflict between the mathematical excellence and practical solutions in several analysis steps, such as rarefaction for compensating for differences in sequencing depth (Hong et al., 2022), zero-imputation for log-ratio transformations (Baruzzo et al., 2021), and estimating alpha-diversity (Willis, 2019). Taken together, there are trade-offs during the analysis phase that a researcher is facing, and the results, conclusions and claims are dependent on those decisions. It is up for the researcher to apply as much rigor as possible and to be aware of the possible limitations.

Machine learning (ML) lies in between the three realms of microbial feature sets, and its impact and popularity makes it a whole another subtopic with tremendous amount of research and effort put into. Using the microbiome composition to predict a trait, for example diagnosing subjects with colorectal cancer as described in chapter 1.2.1 or predicting changes in glucose measurements are common tasks often attributed to the machine learning algorithms. Without a doubt, the results leveraging machine learning have been promising. Variety of the algorithms can work with the high dimensionality of the microbiome data, that is a relative strength as the complexities of interactions in the ecosystem can be taken into account by the learning system. There are concerns, however. Most importantly, the seeming simplicity of applying machine learning models on whatever data has its setbacks. Firstly, returning to the curse of the personalization and temporal dynamics of the gut microbiome composition, the huge variability complicates the construction of models which can generalize within and between cohorts (Wirbel et al., 2019). There are additional concerns with model evaluation with overpromised performances (T. P. Quinn, 2021), and with the lack of transparency and explainability of the models, that can hamper their implementation to clinical practice (T. P. Quinn et al., 2022). Also, there is no consensus about an algorithm that can provide the best performance (Marcos-Zambrano et al., 2021). Several steps can be taken to confront the aforementioned challenges. Firstly, considering the compositional nature of the microbiome data can improve the performance of the prediction models and lead to sparse models. For example, log-ratio transformations can improve the performance of common machine learning algorithms (Tolosana-Delgado et al., 2019) and significantly increase the sparsity (Coenders & Greenacre, 2021; Gordon-Rodriguez et al., 2021; T. Quinn & Erb, 2019; Rivera-Pinto et al., 2018). Then again, it has been shown that even the presence-absence of the taxa can be a viable alternative (Giliberti et al., 2022). Secondly, due to the microbiome datasets still being rather small, the model evaluation requires the estimation of variability of the performance, for example through rigorous testing schemes such as nested-cross-validation (Vabalas et al., 2019). Thus, balancing performance, robustness and interpretability will need to be resolved for the models to take the step from research to practice. Step ahead, the microbiome is prone to changes and so is the microbiome in the population-scale, which has been proposed for Western populations (Schnorr et al., 2014). This can lead to relatively rapid "distributional shifts" in the data, that can hamper the performance of the production-ready microbiome-based prediction tools raising the need for constant development and quality assessment (T. P. Quinn et al., 2022).

Finally, several larger-scale studies have revealed an interesting concept, which makes up a challenge on its own. Namely, several common diseases seem to show similar changes in the microbiome composition when compared to healthy controls (Armour et al., 2019; Jackson et al., 2018). A phenomenon often referred to as "common dysbiosis" or "shared dysbiosis". When thinking back about the diagnostic and predictive applications we are interested in, identifying disease-specific signals becomes important. Thus, the analysis, either univariate or multivariate,

should be aware of the common part of the dysbiosis and try to dissect it. It might be that some common covariate leads to the common dysbiosis and the confounding effect of such confounder should be taken into account.

1.4. Perspectives for the human microbiome research

Despite the many challenges ahead, the next years of microbiome research will be undoubtedly interesting. The field will pick up pace with the number of sequenced microbiome samples rapidly increasing and private sector being more and more curious (Eisenstein, 2020). Here are some perspectives, dreams and wishes for the microbiome research for the upcoming years.

Perhaps the standardization of the microbiome analysis from sample collection to taxonomic and functional profiles would be a welcomed gift by and to the scientific community. This would allow to analyze the data globally with less variability attributed to technical errors and batch effects (Y. Wang & LêCao, 2020). This again can help us in the "replicability" crisis allowing direct comparisons of the results and larger sample sizes for the analysis. Also, applications such as disease diagnostics and prediction could be finalized into market-ready products. This would raise interest and funding, that could exponentially increase the pace of obtaining new knowledge.

The amount of research being published in the microbiome field is doubling approximately in every 3 years according to PubMed. Although this promises a fertile ground for progress, dissecting the results has become increasingly difficult. The research includes functional studies in animal models, population-based observational studies, interventional studies and clinical trials to name a few. In the end, all the results aim to conclude something about one system – human. Setting aside the challenges in more technical aspects, combining the available and upcoming results into that one framework is and will be a challenge. In the same line, the quest for causality is and will be an active research question. Luckily, there are studies paving the way, which have studied the causal role of microbiome through Mendelian randomization analysis (García-Santisteban et al., 2020; X. Liu et al., 2022), there are promising randomized double-blinded clinical trials ongoing, and we have a framework to modulate Koch's postulates for human microbiome research as described. Then again, the mendelian randomization studies are limited in their applicability due to the shortage of the identified human genetics-microbiome interactions, and the clinical trials as a gold standard are heavily time-consuming. Therefore, there are high hopes for the upcoming years for methodological development that would help to expand the causal understanding of the seen associations.

The thesis has so far been entirely focused on the microbiome and in particular, gut microbiome. Human, however, is a complex system and microbiome is a team player in the game of taking care of the body and mind. Thus, the gut microbiome needs to be integrated and analyzed in the context of a multi-site, multi-omic, multi-factorial system incorporating human genetics, metabolomics,

but also environmental factors. As previously mentioned, this can serve two aims: the integration can help to digest the biological mechanisms or help to improve the performance and compare cost-efficiency of clinical applications. Luckily integrating these layers or viewpoints is ongoing and the stage is set for new discoveries with initiations such as the 10k project (Shilo et al., n.d.).

We are standing in an interesting point in time. On one hand, the scientists are aware of the potential microbiome can have for clinical practice and we can almost taste the success, but the microbiome just keeps on surprising. These are exciting years ahead.

2. AIMS OF THE STUDY

The general aims of this thesis were to study the host factors influencing human gut microbiome composition and how could the microbiome data be used to address the needs of the clinical applications. The specific aims were following:

- To study the potential for using gut microbiome for predicting changes in glucose regulation and to understand the robustness of the prediction models
- To characterize the lifestyle and health parameters that are associated with the gut microbiome composition using the value of the extensive biobank data and electronic health records
- To evaluate the relevance of the enterotypes for disease classification, prognostics, and subtype identification

3. RESULTS AND DISCUSSION

3.1. Using microbiome for predicting changes in glycose regulation (Ref. I)

Type 2 diabetes (T2D) is a great example of a modern lifestyle disease: the prevalence of T2D has more than doubled since 1980 and it has a large burden on the health care system, which makes it one of the most popular target diseases for microbiome studies (Gurung et al., 2020; World Health Organization, 2016). The characterizing nature of the disease is the body's inability to handle glucose, which elevates the glucose levels in the system, which in turn can lead to decreased quality of life, blindness, amputations, kidney failure etc (World Health Organization, 2016). Although seemingly simple, the mechanisms that help to control the glucose are rather difficult with multiple pathophysiological pathways involved making T2D a heterogenous disease. That said, type 2 diabetes is in a sense a "human defined" disease (Gale, 2013). The diagnosis of the type 2 diabetes relies, depending on the region, on two or three continuous measurements: fasting glucose, 2-hour glucose and glycosylated hemoglobin (HbA1c), which are turned into the disease using a human-defined threshold. Thus, case-control studies, which consider T2D a binary health state, might not capture the full complexity of the disease. Notably, the T2D is preceded by prediabetes, which is a condition that is characterized by higher-than-normal glucose levels, which haven't crossed the threshold yet. Importantly, the progress to T2D can be reversed during the prediabetic state, which emphasizes the need for detecting disease progression in the early phases (Tabák et al., 2012). Analysis of the continuous measurements directly can improve the understanding of the role of microbiome in T2D, which are the focus of the first manuscript of the thesis (Aasmets et al., 2021).

3.1.1. Description of the cohort and methods

This analysis took advantage of the Metabolic Syndrome In Men (METSIM) study, which is a cohort of Finnish men aged 45–73 years, who have been carefully phenotyped for metabolic diseases (Laakso et al., 2017). Specifically, a subset of the METSIM cohort were analyzed, who took part in the METSIM follow-up study and from whom stool samples were collected and sequenced using 16S rRNA amplicon sequencing (V4 hypervariable region). The subjects (N = 608) gave samples at three different timepoints- at baseline, 18-months after the baseline and 48-months after the baseline. At each time point, the subjects went through a 1-day outpatient visit during which they provided blood samples after an overnight fast and performed an oral glucose tolerance test (OGTT), which gives an extensive overview of the glucose regulation. The study design and modelling procedure is shown in **Figure 3**.

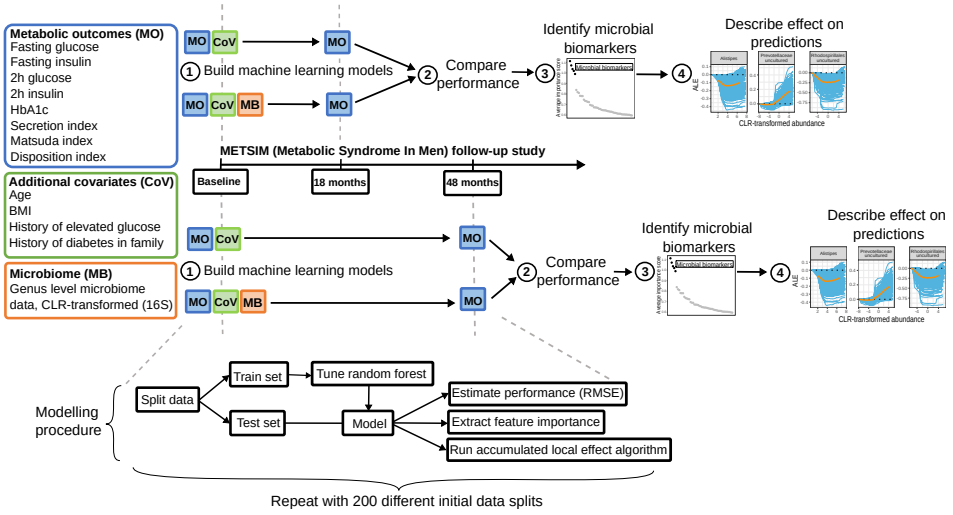

Figure 3. Study design and the modelling procedure.

Random forest models were built to predict the continuous "metabolic outcomes" (MO) in the follow-up timepoints using the baseline measures of MOs and centered log-ratio (CLR) transformed microbial genera abundances as predictors. Genera that appeared in at least 50% of the samples were included in the final modeling task, 172 in total. The MOs included parameters measured during the OGTT (fasting glucose, 2h glucose, HbA1c, fasting insulin, 2h insulin) and three glucose regulation indexes calculated based on the OGTT results (Matsuda insulin sensitivity index (Matsuda & DeFronzo, 1999), insulin secretion index and disposition index). The models including microbial predictors were compared to the models excluding microbial predictors to assess the added predictive value of the gut microbiome. The modelling procedure was repeated 200 times with different data splits for model training and evaluation to evaluate the robustness of the approach. Permutational feature importance together with accumulated local effect plots were used to identify and characterize the most significant microbial predictors for each MO and both follow-up periods.

3.1.2. Microbiome composition predicts changes in insulin secretion and glycated hemoglobin

Firstly, we asked whether including the microbiome data could increase the prediction accuracy for predicting changes in glucose regulation. To answer this question, we compared the performances measured by root-mean-square-error (RMSE) of the random forest models including and excluding microbial predictors. Results showed that for the 18-month follow-up time point, the microbiome can improve the prediction accuracy for predicting 2h insulin levels, insulin secretion index and HbA1c. In the 48-month time frame, the microbiome improved

the prediction accuracy for insulin secretion, fasting insulin and 2h insulin (**Table 1**). Remarkably, none of the direct glucose measures turned out to benefit from the microbial predictors, highlighting that the microbiome is likely involved in the insulin regulation process. These results were consistent to the mendelian randomization study, which showed a causal effect of short chain fatty acids (SFCA) to various insulin measures, primarily insulin secretion (Sanna et al., 2019). The next question after identifying the outcomes that can take advantage of the microbiome was to identify the microbial genera that are responsible for the improved accuracy. The most significant microbial predictors for the relevant metabolic outcomes according to the permutational variable importance are shown in **Figure 4**. For each MO, several prominent genera could be detected. For example, unclassified *Rhodospirillales* was identified as a predictor for fasting insulin and 2h insulin in the 48-month setting. Interestingly, bacteria from the *Rhodospirillales* order can produce acetic acid (Mamlouk & Gullo, 2013), which can improve insulin sensitivity (Johnston et al., 2004; Mitrou et al., 2015). Nevertheless, only a few of the identified genera had been previously reported nor did the findings include the most frequently reported T2D-associated taxa such as *Roseburia* or *Bifidobacterium* (Gurung et al., 2020). The inconsistent patterns identified can be attributable to study design. Cross-sectional study designs, which have been the source for a majority of the findings, can confuse disease-effects with the effects of other covariates such as medication-effects as has been shown for metformin, a common drug for T2D (K. Forslund et al., 2015).

Table 1. Model stability and generalizability.

	18-month time frame		**48-month time frame**	
Trait	Mean (sd) difference in RMSE	# models including microbiome performing better	Mean (sd) difference in RMSE	# models including microbiome performing better
Fasting glucose	0.001 (0.0594)	99 (49.5%)	–0.006 (0.0641)	112 (56%)
2h glucose	–0.02 (0.217)	118 (59%)	0.07 (0.332)	73 (36.5%)
Fasting insulin	0.20 (1.04)	73 (36.5%)	–0.29 (1.080)	137 (68.5%) *
2h insulin	–3.23 (10.840)	141 (70.5%) *	–1.42 (12.304)	122 (61%) *
HbA1c	–0.005 (0.0305)	129 (64.5%) *	–0.002 (0.0360)	111 (55.5%)
Secretion index	–0.36 (4.949)	122 (61%) *	–0.77 (3.254)	138 (69%) *
Matsuda index	0.07 (0.573)	90 (45%)	–0.01 (0.569)	103 (51.5%)
Disposition index	4.42 (26.590)	77 (38.5%)	2.01 (16.251)	86 (43%)

Mean differences in root-mean-square error **(**RMSE) between models including microbial predictors and models excluding microbial predictors. Negative value indicates a model including microbial predictors outperforming the model excluding microbial predictors. * shows statistically significant results according to the binomial test after Bonferroni correction.

Recently, several large-scale prospective studies have published their results, which allow a more comprehensive comparison. A prospective study of 2772 Chinese individuals incorporating similar follow-up period and continuous glucose and insulin measures confirmed the predictive properties of *Para-prevotella*, *[Ruminococcus] torques group* and *Family XIII AD3011 group* (H. Wang et al., 2022). Furthermore, analysis of the prospective data of FINRISK cohort identified several bacterial species associated with the T2D incidence with several species from the family *Lachnospiraceae* and genus *Alistipes* confirming the results (Ruuskanen et al., 2022). Interestingly, the FINRISK data indicated that a short follow-up time might not be sufficient to identify the biomarkers for increased disease risk. Our results showed that genus *Alistipes* and genera from the family *Lachnospiraceae* are predictive of changes in insulin parameters, but not for glucose-parameters directly. Thus, it might be that although the signals are too weak for predicting the disease as a binary outcome, analyzing the continuous glucose regulation markers can identify the first signal of disease progression.

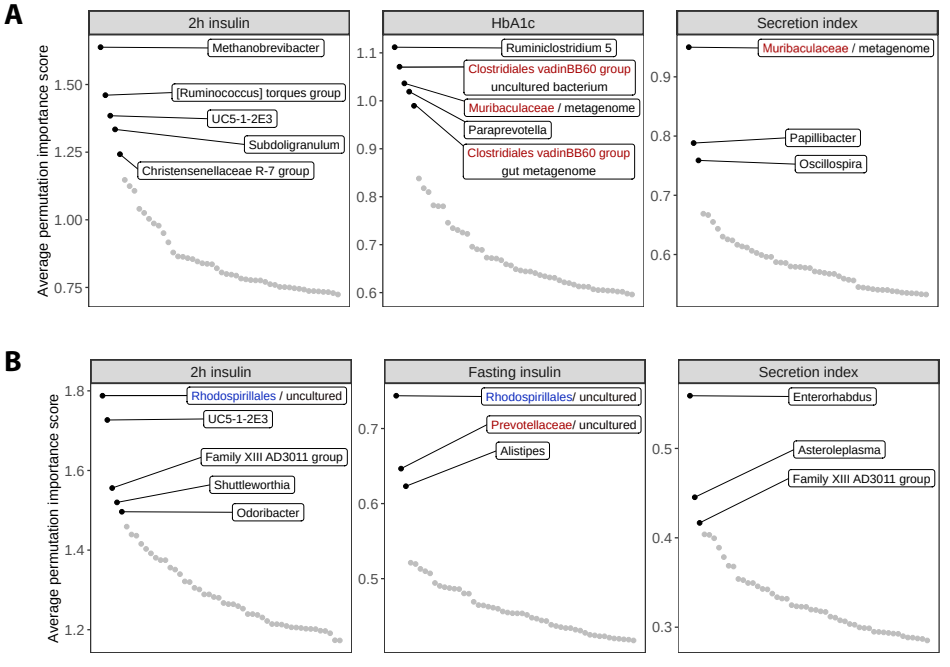

Figure 4. Average feature importance scores for top 50 microbial markers. Highlighted taxa are considered the most significant biomarkers. (**A**) Predictors for 18-month follow-up. (**B**) Predictors for 48-month follow-up.

Taking together, these results emphasize the need for detailed phenotyping of complex diseases to identify the first signals of disease progression and to understand the potential mechanisms microbiome is involved in. The WHO allows T2D to be analyzed using different measurements. Although clinically relevant, it might be that scientifically this definition "blurs" the question we are asking and the answer depends on the way T2D is truly diagnosed in clinical practice, which can be different by countries. This can be one piece of the puzzle for the difficulties researchers encounter when replicating the results of previous studies. It is possible that we try to compare slightly different phenotypes without realizing the "distributional shift" in the data as previously described. Therefore, ideally the two virtues of the METSIM and FINRISK studies would be combined with a large sample size, long follow-up, and deep phenotyping to take the most of the T2D studies.

3.1.3. Modelling microbiome data is challenging

The most significant takeaway from the analysis was the difficulty in obtaining a robust model and estimate of the model performance. The root-mean-squared-error (RMSE) of the model was largely dependent on the initial train-test split. In the best-case scenario, for 2h insulin in a 18-month time frame, the model including microbial predictors outperformed the model excluding microbial predictors in 70.5% of the splits (**Table 1**). The sample sizes in microbiome field today are small, which makes it problematic to evaluate the performance of the machine learning models correctly and robustly. In that scenario, without an understanding about the variability in the performance of such predictive models, we can get lucky in performance estimates just by chance. Such high variability arising from data split the had been also shown before (Topçuoğlu et al., 2020). The large inconsistencies in model performance estimates can be observed even when applying a model developed on one population on the data from another population (Wirbel et al., 2019) or even within a population (Ruuskanen et al., 2022). This is an important problem that requires further investigation, and it is likely an underestimated challenge that the microbiome community faces (T. P. Quinn et al., 2022). On one hand, the selection and efficacy of the algorithms that could best leverage the information hidden in the sequences can be improved. There is no strong consensus on the best-performing approaches, but there are several ideas in the air. For example, accounting for the data compositionality by using log-ratio transformations (Tolosana-Delgado et al., 2019) or data augmentation (Gordon-Rodriguez et al., 2022) can improve the model performance, sparsity and thereby interpretability. On the other hand, the populations can have specific demographic-signatures, which can hamper the development of universal and global tools (Suzuki et al., n.d.). Thereby, clinical applications exploiting the microbiome data might need to be developed for or adapted to each population to accompany reliable performance estimates.

3.2. Estonian Microbiome cohort gives novel insights into the microbiome-associated factors (Ref. II, III)

During the last decades, population-based biobanks have become important resources for building scientific knowledge, especially for the human genetics field. Relatively recently, large-scale collection of samples for microbiome research started, with new microbiome-based biobanks initialized and the existing biobanks expanding their data collection (Falony et al., 2016; Gacesa et al., 2022; Shilo et al., n.d.; Turnbaugh et al., 2007; Zhernakova et al., 2016). The biobanks generally contain rich phenotyping and multi-omics data, which allows to identify factors associated with the microbiome composition and study host-microbiome interactions. Such large-scale cohorts have contributed significantly to the collective knowledge and have been instrumental in generating hypothesis, validating experimental findings etc. However, as the number of microbiome cohorts is increasing, the scale of the cohorts can become limiting for tracking the participants in time. Most of the cohorts characterize the microbiome and the health state in one, cross-sectional manner, which limits the number of questions asked. This includes the clinical cohorts, which generally focus on only specific diseases. That said, the next section introduces the Estonian Microbiome cohort that takes advantage of the availability of electronic health records (EHR). The results of two studies will be shown, which highlight the added value of the EHR for microbiome studies.

3.2.1. Description of the cohort and methods

As part of the Estonian Biobank (EstBB), the Estonian Microbiome project (EstMB) was initiated in 2017, with more than 2500 individuals providing stool and oral samples for microbiome studies (1764 females and 745 males, aged 23–89, samples collected from the whole Estonia). Estonian Biobank is a volunteer-based cohort consisting of approximately around 20% of the Estonian adult population, that was initiated in 1999 with the objective to investigate the genetic, environmental, and behavioral background of common diseases and traits (Leitsalu et al., 2015). The biobank takes advantage of the linkage to the electronic health records (EHR), Human Genes Research Act and a wide array of biological samples. This study took advantage of the rich phenotype data to improve our knowledge of the microbiome-associated factors.

The electronic health records are one of the core strengths of the Biobank. They allow a detailed characterization of one's health through time as the diseases, drug prescriptions bought out, and medical procedures carried out are registered by the medical professionals with great accuracy. Adding the extensive self-reported questionnaires about the dietary preferences and lifestyle habits, a total of 71 diseases, 136 medications, 21 dietary items, 5 medical procedures and 19 additional lifestyle factors were questioned from the microbiome perspective – is the microbiome composition associated with these factors? This pheno-

typic variation was in **Ref II** associated with the microbial alpha diversity (observed richness and Shannon index with Spearman correlation), beta diversity (Aitchison distance and PERMANOVA) and with the abundance of each species separately (ALDEx2). Additionally, machine learning models (elastic net regression) were built using different predictor sets to characterize the potential of using microbiome data for disease classification and diagnostics. The shotgun metagenomic paired-end sequencing was performed by Novogene Bioinformatics Technology Co., Ltd. using the Illumina NovaSeq6000 platform, resulting in 4.62 ± 0.44 Gb of data per sample. As shotgun metagenomics data is available, the analysis was carried out for the taxonomic profile and for the functional profile characterized by the Kyoto Encyclopedia of Genes and Genomes (KEGG).

Ref III categorized the samples into distinct clusters based on their taxonomic profile using the Dirichlet-Multinomial Mixture model (DMM) (Holmes et al., 2012). A 5-cluster "community type" model (CT) and a 3-cluster "enterotype" model (ET) were considered for downstream analysis based on the Laplace approximation. The clusters were associated with the phenotype data using generalized linear models and chi-squared tests. Furthermore, incident diseases were analyzed with the Cox proportional hazards models to understand, whether the CT or ET models could be used to assess the risk for disease progression.

3.2.2. Microbiome composition reflects our lifestyle and health

The initial aim when analyzing the EstMB dataset was to get an understanding of the factors associated with the gut microbiome composition. Analysis identified a large number of phenotypic factors associated with either taxonomic and functional-level alpha-diversity, beta-diversity or with certain species or KEGG orthologs. In total, 136 out of 252 dietary-, disease-, medications usage- and other lifestyle factors were associated with the microbiome composition in some way. These factors together explained around 10.14% of the interindividual variation in the gut microbiota compositions. Firstly, the results of the community-scale analysis allow to make several observations (**Figure 5**). For example, the stool characteristics are the major drivers of the microbiome variation dominating the host-intrinsic factors including BMI and disease states. Secondly, factors generally associated with an unhealthy lifestyle: smoking, no physical exercise, high BMI, high consumption of soft drinks and processed meat products with low consumption of vegetables and berries, are associated with lower alpha-diversity. Same can be said about the prevalence of nearly all common diseases which, importantly, describe more variance in the microbiome composition when compared to the medications. These results have helped to expand the set of microbiome-associated factors by medical procedures such as removal of the cecum and medications such as glucocorticoids. Regardless, many of these results are not necessarily novel to the scientific community. For example, the phenotypic factors with the largest impact on the microbial variation including gut emptying frequency, stool consistency, BMI and gender have been previously observed in several population-based cohorts (Falony et al., 2016; Zhernakova et al., 2016).

Figure 5. Statistically significant associations with species-level microbiome alpha and beta diversity. The bar plot indicates the explained variance in the interindividual variation of the microbial composition (based on the Euclidean distance on the centered log-ratio-transformed data). The heatmap shows the Spearman correlation coefficients of each factor with the Shannon's index of diversity and the observed species richness. Blue indicates a negative correlation, and red indicates a positive correlation. *FDR < 0.05.

The overarching virtue of the Estonian Microbiome cohort data relies on the possibility to use the electronic health records (EHR). As previously noted for T2D, the more detailed information we can analyze, the better. Most importantly, the EHR allow to pinpoint the drug usage and disease occurrences with great accuracy both prospectively and retrospectively, which gives rise to a range of questions. Our study took advantage of the EHR by studying the accumulative effects of drug usage. It turned out that the history of antibiotics usage characterized by the number of prescriptions over the last 10 years before the sample collection had a significant effect on microbiome composition, independent of the recent usage (**Figure 6**). The number of antibiotics prescriptions was significantly associated with the first principal components of the species-level microbiome composition and remarkably, this effect was already evident in subjects having used antibiotics at least 3 times in the previous 10 years. Similar observation with antibiotics was recently made by others (S. K. Forslund et al., 2021), but we were able to extend this idea to show also a weak accumulative effect of antidepressant usage.

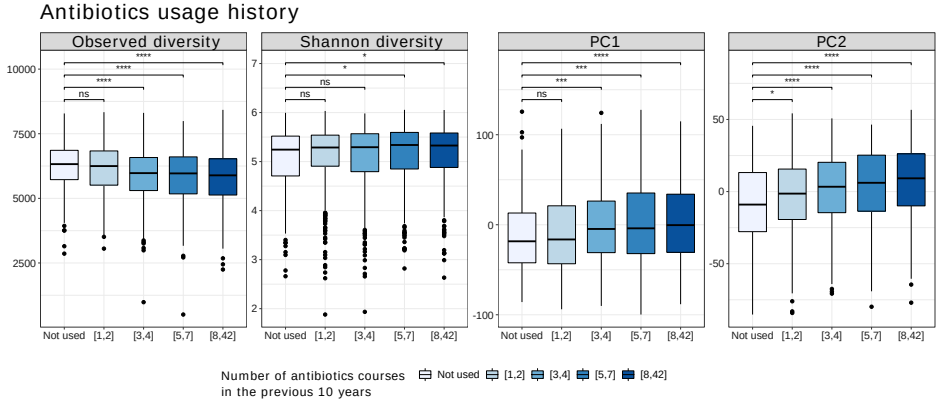

Figure 6. Associations with antibiotics usage history and the observed number of species (the y-axis represents the number of species), Shannon diversity (the y-axis represents the Shannon's diversity index), or the first two principal components (PCs) of the species-level microbial composition. Asterisks indicate statistically significant differences between the drug usage history groups using Wilcoxon test (FDR < 0.05*, FDR < 0.01**, FDR < 0.001***, FDR < 0.0001****), and *ns* notes statistically nonsignificant results. Color key indicates the five distinct classes of medication users, the non-users and four additional classes based on the quartiles of the number of prescriptions filled over the 10-year period. The sample size for antibiotics were the following: nonusers n = 243; [1,2] n = 549; [3,4] n = 440; [5,7] n = 395; [8,42] n = 400.

This long-term effect of antibiotics was perhaps anticipated, but the extent and scope of the association turned out to be a surprise. Usage of antibiotics is a significant perturbation that asks for the community to rebuild itself and our results show that this restoration might not end in returning to the baseline state as thought (Palleja et al., 2018). We clearly saw that if more antibiotics were used in the past, the microbiome composition was more likely to be dominated by the genus *Bacteroides*. This finding has strong implication for the future studies.

Firstly, the history of antibiotics usage should be accounted for in microbiome studies. For example, adjusting for the usage of antibiotics in the differential abundance analysis can help to identify disease-specific microbial markers as shown in **Figure 7**. Remarkably, the number of univariate associations is significantly lower after the adjustment and the decrease in the number of associations is dependent on the disease. Secondly, the history of antibiotics usage should be considered in clinical applications. Asking about the number of prescriptions for antibiotics can be a simple solution to account for the microbiome component without the extra cost of sample collection and sequencing. Thus, the efficacy of a diagnostic test, which is based on characterizing microbiome composition should supersede the test that is based on the data about antibiotics consumption. Furthermore, following up on the "common dysbiosis" idea (Jackson et al., 2018; Xu et al., 2020), the long-term antibiotic usage can be a partial reason behind why we see such results (**Figure 7**). This question should be further investigated, and luckily the Estonian Microbiome cohort provides a great resource for answering questions alike.

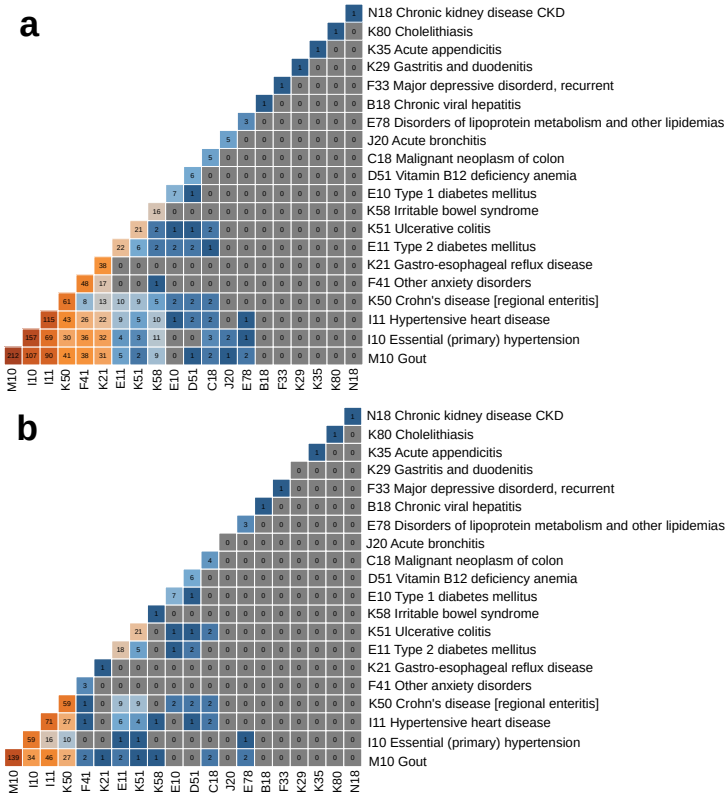

Figure 7. Effect of adjusting for antibiotic usage on the number of overlapping associations between various diseases. **a** – Heatmap of overlapping associations between various complex diseases before adjusting for antibiotic usage. **b** – Heatmap of overlapping associations between various complex diseases after taking long-term antibiotic usage into account.

3.2.3. The clinical relevance of enterotyping remains fragile

Human gut microbiome is an extremely difficult "organ" to analyze because of the large variability of the microbiome composition within and between individuals. When we look at two people, their microbiome compositions are profoundly different and when looking at the microbiome of one individual in time, we encounter significant fluctuations in the composition. Thus, instead of focusing on a set of species, strains or other taxonomic units, a clustering approach has been proposed, which divides the subjects into a small number of distinct clusters based on their microbiome composition. Often referred to as "enterotyping" (Arumugam et al., 2011), the clustering approach has been rigorously applied in the microbiome field (Christensen et al., 2018; Costea et al., 2017; Vandeputte et al., 2016). Nevertheless, the clinical relevance of such clustering has remained questionable, which this part of the thesis aimed to share a light on (Aasmets, Krigul, & Org, 2022). The microbiome genus-level composition was used to cluster the EstMB subjects into two competing set of clusters using the Dirichlet-Multinomial Mixture model (Holmes et al., 2012). An "enterotype" model (ET) consisting of 3 clusters and "community type" (CT) model consisting of 5 clusters were considered. Firstly, several associations between the identified clusters and phenotypic factors were identified (**Figure 8**).

From the clinical perspective, numerous diseases were found to be associated with the cluster composition giving hope for potential applications. Firstly, the enterotype composition could be considered for disease diagnostics. However, as the cluster composition is also associated with common confounders and risk factors such as age, gender and BMI, the diagnostic properties of the clustering should exceed the discriminability of the covariates only. After adjusting for age, gender, and BMI, several of the associations between diseases and microbial clusters were not statistically significant. Only associations with gout, disorders of lipoprotein metabolism, essential hypertension and chronic tubulo-interstitial nephritis for enterotype-model and gout and anxiety disorders for community type model remained statistically significant. Also, as we and others have shown that drug usage can confound the microbiome-disease interactions, we further adjusted the models for drug usage. After adjusting for drug usage, associations with anxiety disorder and tubulo-interstitial nephritis were not detected. Although the enterotype composition remained associated with different diseases, the enterotyping alone might not be sufficient for diagnostic purposes because different diseases are overrepresented in the same enterotypes. Next, we took advantage of the electronic health records and analyzed incident diseases to see whether the clustering approach would allow to estimate the risk of developing a condition. However, the survival analysis diseases didn't identify any significant results, only suggestive evidence for higher risk of migraine in the *community type (CT) 3* cluster. As the median follow-up time of the cohort was 3.1 years, it is possible that the follow-up is too short to characterize the predictive ability of the enterotype composition. For example, it has been shown for type 2 diabetes that the differences in disease incidence among the relative abundance quartiles of the

identified biomarkers appears only after 5 years (Ruuskanen et al., 2022). Also, the enterotype of the human gut microbiome has been considered a relatively stable characteristic for an individual, but a recent study suggest that we can change our enterotypes more often than previously thought (Olsson et al., 2022). Therefore, although the enterotyping can paint a nice picture about the lifestyle and physical health of the subject, the evidence for the clinical relevance remains fragile.

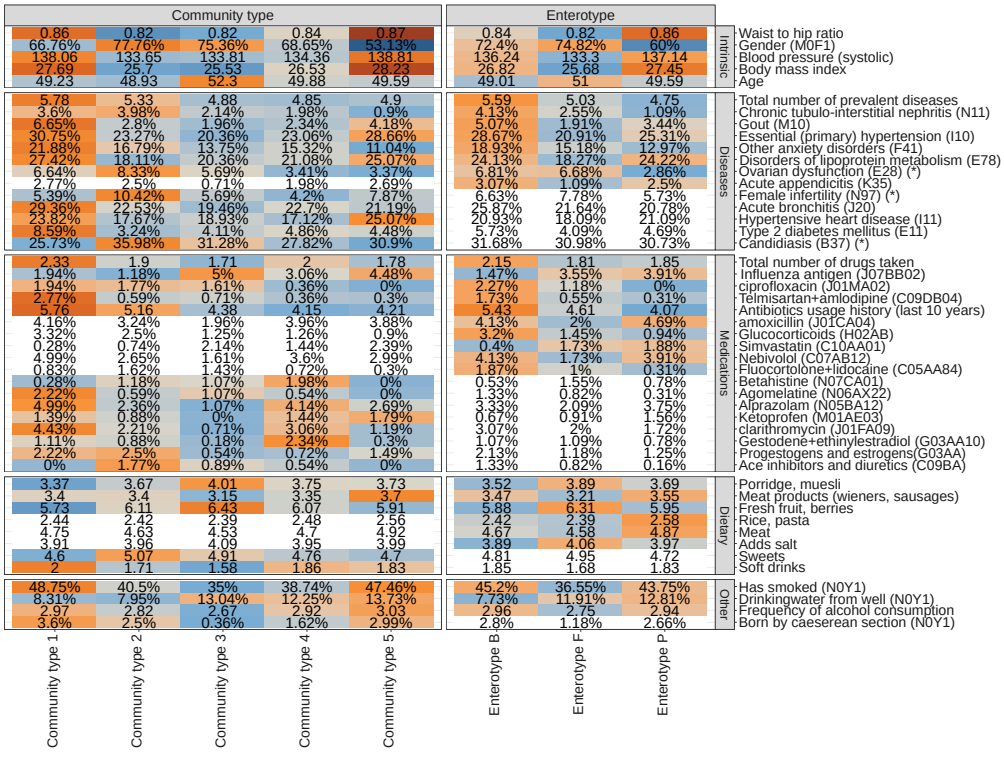

Figure 8. Associations with the enterotype (*ET*) model and community type (*CT*) model. Colored cells represent factor associated with *CT* and *ET* models respectively (FDR <= 0.1), white cells indicate no statistically significant association (FDR > 0.1). Blue colors indicate lower mean values or proportions for the cluster and orange color indicate higher values. Mean values or proportions (indicated by %) per cluster are shown. Asterix (*) in the names of the factors indicate that a subpopulation consisting of women was used for calculating the displayed value.

3.2.4. We are still in the exploratory phase with microbiome studies

We are aware of numerous phenotypic factors that are associated with the microbiome composition, but it seems we are still missing a lot. For example, the well-characterized population-based cohorts have been able to characterize around 10% to 19% of the variation in the taxonomic composition (Aasmets, Krigul, Lüll,

et al., 2022; Zhernakova et al., 2016). It has been suggested that the properties of the microbial community, such as interactions between the community members, growth rates and immigration can have a larger effect (Vandeputte et al., 2021). For example, the intra-individual variance was estimated to be 23% in a Swedish cohort (Olsson et al., 2022).However, we now know that the drugs can have an "accumulative" effect on the gut microbiome composition, which is a totally new player. This knowledge can help to further pinpoint disease-specific signals, raise new hypothesis, and direct the microbiome-targeted applications. In this study we showed that in addition to antibiotics, the host-targeted drugs such as antidepressants can have similar effects. This is remarkable as nearly 24% of common host-targeted have shown an effect on the microbiome *in vitro* (Maier et al., 2018) raising the possibility that such accumulation effects are prevalent for various other drug classes. It is tempting to say that we are far from being finished with the exploratory studies. It also shows that information about the current state of the microbiome composition can be replaced by asking a simple question from the participant such as "how many antibiotics have you consumed in the last 2 years?". Therefore, some applications that take advantage of the microbiome composition could potentially be replaced by a more cost-efficient questionnaire.

The accumulation of antibiotics and perhaps other host-targeted drugs in the human gut microbiome raises another possibility. As the antibiotic usage rates (drug usage in general) and distributions of consumed drug subclasses differ by populations, we might end up with a large-scale *in natura* experiment. It might be that this can create population-specific "dysbiotic" signatures that make it difficult to obtain similar signals in seemingly similar studies, hence another piece of puzzle to the replicability issue. The idea could be expanded by considering the horizontal or vertical transfer of microbes, which plays a role in diversifying our microbiomes. If we now live in a population with altered microbiome pool, it can become problematic – we have less to share, and the wheel keeps on spinning...

CONCLUSIONS

The evidence piled up by the literature leaves no doubt for the author of this thesis that human microbiome science is and will stay a rapidly advancing field in the coming years. Beyond the knowledge about the dynamics and host-targeted functions of the ecosystem itself, we will develop novel diagnostic measures and microbiome-based therapies that will improve the public health in a convenient and cost-efficient manner. Nevertheless, there is a lot of work to do and several challenges ahead that need to be resolved.

This thesis aimed to expand the knowledge about the factors influencing the gut microbiome composition and evaluate several ways gut microbiome data could be used for future applications. Several key findings from this thesis can show a way forward. Firstly, a detailed phenotypic characterization in microbiome studies is highly desirable. This thesis took advantage of two well-characterized cohorts by highlighting the part of glucose regulation, that microbiome is likely involved in and showing an accumulative effect of the history of antibiotics usage on the gut microbiome composition. The identified accumulative effect of long-term antibiotics usage highlights the need for further exploratory studies to identify the factors associated with the gut microbiome composition. Secondly, using microbiome data in data-driven applications will need to balance between the explainability, performance and robustness of the approach. Trying to generalize the microbiome variability by a clustering approach has its advantages for implementation, but the evidence of diagnostic or prognostic performance is weak. Using a more complex approach such as machine learning can potentially improve the performance but needs to be thoroughly evaluated to ascertain reliability and usefulness. Either way, the microbiome-based clinical applications should be tested against a simple alternative which doesn't include microbiome sampling. Nevertheless, these are not challenges that can't be overcome.

An interim conclusion about the microbiome field would be that we are currently simultaneously enjoying the virtues of new knowledge and opening potential for improving our everyday life. The challenges ahead do not overshadow the enthusiasm.

SUMMARY IN ESTONIAN

Mikrobioomi väärtus terviseuuringutes

Tehnoloogia meeletu areng on andnud inimesele võimaluse uurida ümbritsevat maailma nurkade alt, mille jaoks veel mõned kümnendid tagasi võimalused puudusid. Üks selliseid teadusvaldkondi on inimese mikrobioomi ehk meie kehal ja kehas elavate mikroorganismide nagu näiteks bakterite ja viiruste uurimine. On märkimisväärne, et mikroorganisme inimese kehas on rohkem kui inimesel enda rakke ning mikrobioom koondab endas kordades rohkem geene kui on inimesel vastu panna. Selline meeletu mikroobide mitmekesisus, millest enamus pesitseb inimese soolestikus, omab inimese tervisele ning edukale toimimisele olulist rolli. Näiteks lagundatakse mikrobioobide poolt toodetud ensüümide abil kiudained, mida inimene ise ei suuda seedida ning millest sünteesitakse inimese ainevahetusele vajalikke ühendeid nagu näiteks vitamiine ja lühikese ahelaga rasvhappeid. Mikrobioomil on oluline roll organismi immuunsüsteemi arengus ning funktsioneerimisel, mis on meie tervise oluline alustala.

Mikrobioomi kooslust omakorda mõjutab suurel määral meie elustiil, toitumisharjumused, ümbritsev keskkond ning tervislik seisund. Suurimad mikrobioomi koosluse mõjutajad on inimese kehakaal, väljaheite konsistents, sugu, suitsetamine ning ka mitmete komplekshaiguste esinemine ja ravimite tarbimine. On näidatud, et tervetel inimestel on mikrobioomi kooslus erinev kui inimestel, kellel esineb haigusi nagu teist tüüpi diabeet, depressioon, soolehaigused või isegi vähk. Just seosed haigustega on tekitanud huvi mikrobioomi kasutamiseks meditsiinis, milleks on mitu võimalust. Esiteks, inimese mikrobioomi on võimalik üsna lihtsasti muuta, mistõttu saaks potentsiaalselt ka läbi koosluse muutmise tervist parandada. Näiteks võiks saada inimesele kasulike bakterite ehk probiootikumide või bakteritele sobivate toitainete ehk prebiootikumidega suunata mikrobioomi kooslust sobivasse seisu. Teiseks saab inimese mikrobioomi kooslust kasutada selleks, et diagnoosida haigusi, prognoosida haiguste riski või ennustada, milline dieet või ravim inimesele sobib. Doktoritöö uuribki, mis mõjutab meie soolestiku mikrobioomi kooslust ning kuidas on seda võimalik kasutada meditsiinilistes rakendustes.

Väitekirja esimene pool annab teaduskirjandusele toetudes ülevaate inimese mikrobioomist, valdkonna põhilistest uurimissuundadest ning kirjeldab väljakutseid, millega tuleb mikrobioomi uurimisel arvestada. Põhjalikumalt käsitletakse soolestiku mikrobioomi kooslust mõjutavaid tegureid ning andmeanalüüsiga seonduvaid väljakutseid. Töö praktiline osa kirjeldab kolme doktoritöö osaks oleva teadusartikli tulemusi.

Esimene artikkel uurib teist tüüpi diabeeti, mis on krooniline haigus, mida iseloomustab normist kõrgem veresuhkru tase. Teist tüüpi diabeeti (T2D) haigestumus on nii Eestis kui ülemaailmselt tõusutrendis ning samuti on diabeedi ravile kuluv ressurss ja koormus meditsiinisüsteemile kasvamas. Haigusele eelneb eeldiabeedi seisund, kus glükoositase veres on kõrgenenud, kuid ei saavuta veel diabeeti defineerivat taset. Märkimisväärne on, et eeldiabeedi seisundis on võimalik

haiguse progresseerumist tagasi pöörata. See aitaks oluliselt parandada rahva tervist ning vähendada koormust meditsiinisüsteemile, mistõttu on tähtis tuvastada haiguse progresseerumine võimalikult varajases staadiumis. Eelnevalt on teada, et soolestiku mikrobioomi kooslus on T2D põdevatel inimestel erinev kui tervetel, kuid vajalike andmete puudumise tõttu on jäänud selgusetuks, kas muutused mikrobioomis toimuvad juba enne haiguse avaldumist. Doktoritöös analüüsiti Soome METSIM kohorti, mis hõlmab endas andmeid tervete meeste kohta, kellel on detailselt mõõdetud T2D-ga seotud veremarkerid kolmes ajapunktis nelja aasta jooksul. Uuringu tulemused näitasid, et mikrobioomi aitab täpsemalt ennustada muutusi mitmetes parameetrites, milleks olid eelkõige insuliini eritamisega seotud näitajaid nagu insuliini väärtus paastuveres ning insuliini sekretsiooni indeks. Töös kirjeldati ka bakteriperekondi, kelle arvukusel on suurim roll mudeli ennustustäpsuse parandamisel. Töö tulemused viitavad selgelt mikrobioomi rollile veresuhkru regulatsioonis ning rõhutavad mitmeid tähelepanupunkte keeruka füsioloogiaga haiguste nagu T2D uurimisel ja analüüsimisel.

Teadustöö järgmises kahes artiklis uuriti soolestiku mikrobioomi mõjutavaid tegureid Eesti Geenivaramu (Estonian Biobank, EstBB) osana loodud Eesti Mikrobioomi kohordi (EstMB) näitel. Eesti Mikrobioomi kohort on populatsioonipõhine kohort geenidoonoritest, kes on andnud väljaheiteproovi soolestiku mikrobioomi analüüsimiseks. Tänu osalusele Eesti Geenivaramus on võimalik iga uuritava kohta lisaks põhjalikele elustiiliküsimustikele kasutada ka elektroonilise registrite andmeid, mis võimaldavad detailselt jälgida inimese tervist ja ravimite tarbimist ning nende seoseid mikrobioomiga analüüsida. Rikkalik terviseandmestik on tugevaks eeliseks võrreldes seniste suuremate mikrobioomi uuringutega, mis võimaldavad analüüsida vaid küsimustikel põhinevaid andmeid. Kasutades EstMB põhjalikku andmestikku, leiti 136 mikrobioomi kooslusega seotud tunnust, sealjuures mitmeid seni rapoteerimata seoseid. Analüüsitavad tunnused hõlmasid inimese elustiili kirjeldavaid tunnuseid nagu suitsetamine ning alkoholitarbimine, haiguste esinemine, tarvitatavad ravimid, toidueelistused ning muid inimest kirjeldavaid tunnuseid nagu sugu, vanus ning kehamassiindeks. Suurimad mikrobioomi varieeruvust kirjeldavad tunnused Eesti Mikrobioomi kohordis olid sarnaselt mitmetele eelnevalt publitseeritud töödele soole tühjendamise sagedus, kehamassiindeks, suitsetamine ning inimese sugu. Kõige märkimisväärsem leid oli antibiootikumide pikaajalise kasutamise akkumuleeruv mõju mikrobioomi kooslusele. Selgus, et korduv antibiootikumide tarvitamine viimase 10 aasta jooksul mõjutab olulisel määral soolestiku mikrobioomi olenemata sellest, kas antibiootikume on kasutatud hiljuti (viimase kuue kuu jooksul). Analüüsides pikaajalise antibiootikumide mõju arvesse võtmine võimaldab omakorda täpsustada haigusspetsiifilisi muutusi mikrobioomis. Lisaks uuriti, kas inimeste grupeerimine soolestiku mikroobide esinemissageduste põhjal kolmeks või viieks grupiks võimaldaks kasutust kliinilistes rakendustes. Eelnevalt on näidatud, et selline grupeerimine võib aidata suurt inimestevahelist varieeruvust mikrobioomi kooslustes arvesse võtta ning sellisel kujutamisel oleksid ka sobivad omadused tulemuste interpreteerimiseks. Meie uuringu tulemused näitasid, et selliselt mikrobioomi kooslust lihtsustades on võimalik küll anda hinnang inimese

üldisele elustiilile (nt toitumine), kuid tõendid gruppide kasutamiseks haiguste diagnoosimisel või haiguse riski hindamiseks on nõrgad.

Töö tulemused andsid suuna edasisteks uuringuteks. Esiteks, võimalikult täpse kliinilise informatsiooni kasutamine aitab mõista mikrobioomi olulisust ning kasutamisvõimalusi meditsiinis. Doktoritöös analüüsiti kahe andmerikka kohordi andmeid, millest selgus, et mikrobioom aitab prognoosida muutusi glükoosi regulatsioonis väga kindlate markerite järgi ning pikaaegne antibiootikumide kasutamine kajastub soolestiku mikrobioomi koosluses. Teiseks, mikrobioomi kasutamine andmetel põhinevates rakendustes nõuab valiku tegemist metoodika läbipaistvuse, efektiivsuse ning robustsuse vahel. Mikrobioomi varieeruvuse üldistamine üksikutesse klastritesse omab eeliseid kliinilistes rakendustes, aga tõendusmaterjal sellise lähenemise rakendamiseks haiguste diagnoosimiseks või haigusriskide hindamiseks on vähene. Keerukamad lähenemised nagu masinõppe kasutamine võib aidata rakenduste efektiivsust tõsta, aga vajavad põhjalikku hindamist, et tagada usaldusväärsus ning kasutatavus.

Tõendite hulk, mis mikrobioomi inimese tervisega seostab, ei jäta töö autorile kahtlustki, et inimese mikrobioomi valdkond on ja jääb järgnevatel aastatel kiirelt arenevaks teadusvaldkonnaks. Lisaks mikrobioomi kui ökosüsteemi dünaamika tundma õppimisele on järgnevatel aastatel oodata mikrobioomil põhinevate diagnostiliste meetmete ning testide ja mikrobioomil põhinevate teraapiate kasutuselevõtt, mis võimaldaks senisest efektiivsemalt ning kuluefektiivsemalt rahvatervist parandada. Sellegipoolest on sellel teekonnal mitmeid väljakutseid, mis vajavad eelnevalt uurimist ning lahendamist.

REFERENCES

Aasmets, O., Krigul, K. L., Lüll, K., Metspalu, A., & Org, E. (2022). Gut metagenome associations with extensive digital health data in a volunteer-based Estonian microbiome cohort. *Nature Communications*, *13*(1), 869. https://doi.org/10.1038/s41467-022-28464-9

Aasmets, O., Krigul, K. L., & Org, E. (2022). Evaluating the clinical relevance of the enterotypes in the Estonian microbiome cohort. *Frontiers in Genetics*, *13*. https://doi.org/10.3389/fgene.2022.917926

Aasmets, O., Lüll, K., Lang, J. M., Pan, C., Kuusisto, J., Fischer, K., Laakso, M., Lusis, A. J., & Org, E. (2021). Machine Learning Reveals Time-Varying Microbial Predictors with Complex Effects on Glucose Regulation. *MSystems*, *6*(1). https://doi.org/10.1128/mSystems.01191-20

Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society: Series B (Methodological)*. https://doi.org/10.1111/j.2517-6161.1982.tb01195.x

Alili, R., Belda, E., Fabre, O., Pelloux, V., Giordano, N., Legrand, R., Bel Lassen, P., Swartz, T. D., Zucker, J.-D., & Clément, K. (2021). Characterization of the Gut Microbiota in Individuals with Overweight or Obesity during a Real-World Weight Loss Dietary Program: A Focus on the Bacteroides 2 Enterotype. *Biomedicines*, *10*(1), 16. https://doi.org/10.3390/biomedicines10010016

Armour, C. R., Nayfach, S., Pollard, K. S., & Sharpton, T. J. (2019). A Metagenomic Meta-analysis Reveals Functional Signatures of Health and Disease in the Human Gut Microbiome. *MSystems*, *4*(4). https://doi.org/10.1128/mSystems.00332-18

Arumugam, M., Raes, J., Pelletier, E., Paslier, D. le, Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J. M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., … Zeller, G. (2011). Enterotypes of the human gut microbiome. *Nature*, *473*(7346), 174–180. https://doi.org/10.1038/nature09944

Ayeni, F. A., Biagi, E., Rampelli, S., Fiori, J., Soverini, M., Audu, H. J., Cristino, S., Caporali, L., Schnorr, S. L., Carelli, V., Brigidi, P., Candela, M., & Turroni, S. (2018). Infant and Adult Gut Microbiome and Metabolome in Rural Bassa and Urban Settlers from Nigeria. *Cell Reports*, *23*(10), 3056–3067. https://doi.org/10.1016/j.celrep.2018.05.018

Bach, J.-F. (2002). The Effect of Infections on Susceptibility to Autoimmune and Allergic Diseases. *New England Journal of Medicine*, *347*(12), 911–920. https://doi.org/10.1056/NEJMra020100

Bajaj, J. S., Acharya, C., Sikaroodi, M., Gillevet, P. M., & Thacker, L. R. (2020). Cost-effectiveness of integrating gut microbiota analysis into hospitalisation prediction in cirrhosis. *GastroHep*, *2*(2), 79–86. https://doi.org/10.1002/ygh2.390

Baruzzo, G., Patuzzi, I., & di Camillo, B. (2021). Beware to ignore the rare: how imputing zero-values can improve the quality of 16S rRNA gene studies results. *BMC Bioinformatics*, *22*(S15), 618. https://doi.org/10.1186/s12859-022-04587-0

Bastiaanssen, T. F. S., Quinn, T. P., & Loughman, A. (2022). *Treating Bugs as Features: A compositional guide to the statistical analysis of the microbiome-gut-brain axis*.

Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., & Segata, N. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *ELife*, *10*. https://doi.org/10.7554/eLife.65088

Belkaid, Y., & Hand, T. W. (2014). Role of the Microbiota in Immunity and Inflammation. *Cell*, *157*(1), 121–141. https://doi.org/10.1016/j.cell.2014.03.011

Binda, S., Hill, C., Johansen, E., Obis, D., Pot, B., Sanders, M. E., Tremblay, A., & Ouwehand, A. C. (2020). Criteria to Qualify Microorganisms as "Probiotic" in Foods and Dietary Supplements. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.01662

Blaser, M. J. (2016). Antibiotic use and its consequences for the normal microbiome. *Science*, *352*(6285), 544–545. https://doi.org/10.1126/science.aad9358

Blaser, M. J. (2018). The Past and Future Biology of the Human Microbiome in an Age of Extinctions. *Cell*, *172*(6), 1173–1177. https://doi.org/10.1016/j.cell.2018.02.040

Cantarel, B. L., Lombard, V., & Henrissat, B. (2012). Complex Carbohydrate Utilization by the Healthy Human Microbiome. *PLoS ONE*, *7*(6), e28742. https://doi.org/10.1371/journal.pone.0028742

Christensen, L., Roager, H. M., Astrup, A., & Hjorth, M. F. (2018). Microbial enterotypes in personalized nutrition and obesity management. *The American Journal of Clinical Nutrition*, *108*(4), 645–651. https://doi.org/10.1093/ajcn/nqy175

Coenders, G., & Greenacre, M. (2021). *Three approaches to supervised learning for compositional data with pairwise logratios*. http://arxiv.org/abs/2111.08953

Costea, P. I., Hildebrand, F., Manimozhiyan, A., Bäckhed, F., Blaser, M. J., Bushman, F. D., de Vos, W. M., Ehrlich, S. D., Fraser, C. M., Hattori, M., Huttenhower, C., Jeffery, I. B., Knights, D., Lewis, J. D., Ley, R. E., Ochman, H., O'Toole, P. W., Quince, C., Relman, D. A., … Bork, P. (2017). Enterotypes in the landscape of gut microbial community composition. *Nature Microbiology 2017 3:1*, *3*(1), 8–16. https://doi.org/10.1038/s41564-017-0072-8

de Goffau, M. C., Lager, S., Sovio, U., Gaccioli, F., Cook, E., Peacock, S. J., Parkhill, J., Charnock-Jones, D. S., & Smith, G. C. S. (2019). Human placenta has no microbiome but can contain potential pathogens. *Nature*, *572*(7769), 329–334. https://doi.org/10.1038/s41586-019-1451-5

Derrien, M., Alvarez, A.-S., & de Vos, W. M. (2019). The Gut Microbiota in the First Decade of Life. *Trends in Microbiology*, *27*(12), 997–1010. https://doi.org/10.1016/j.tim.2019.08.001

Dziewiecka, H., Buttar, H. S., Kasperska, A., Ostapiuk–Karolczuk, J., Domagalska, M., Cichoń, J., & Skarpańska-Stejnborn, A. (2022). Physical activity induced alterations of gut microbiota in humans: a systematic review. *BMC Sports Science, Medicine and Rehabilitation*, *14*(1), 122. https://doi.org/10.1186/s13102-022-00513-2

Eckermann, H. A., Ou, Y., Lahti, L., & Weerth, C. (2022). Can gut microbiota throughout the first 10 years of life predict executive functioning in childhood? *Developmental Psychobiology*, *64*(3). https://doi.org/10.1002/dev.22226

Eisenstein, M. (2020). Early investments powering the ascent of microbiome therapeutics. *Biopharma Dealmakers*. https://doi.org/10.1038/d43747-020-01178-x

Enav, H., Bäckhed, F., & Ley, R. E. (2022). The developing infant gut microbiome: A strain-level view. *Cell Host & Microbe*, *30*(5), 627–638. https://doi.org/10.1016/j.chom.2022.04.009

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M. J., Valles-Colomer, M., Vandeputte, D., Tito, R. Y., Chaffron, S., Rymenans, L., Verspecht, C., Sutter, L. de, Lima-Mendez, G., D'hoe, K., Jonckheere, K., Homola, D., … Raes, J. (2016). Population-level analysis of gut microbiome variation. *Science*, *352*(6285), 560–564. https://doi.org/10.1126/science.aad3503

Federici, S., Nobs, S. P., & Elinav, E. (2021). Phages and their potential to modulate the microbiome and immunity. *Cellular & Molecular Immunology*, *18*(4), 889–904. https://doi.org/10.1038/s41423-020-00532-4

Fernández-Pato, A., Sinha, T., Gacesa, R., Gois, M. F. B., Gelderloos-Arends, J., Jansen, D. B. H., Jaeger, M., Joosten, L. A. B., Weersma, R. K., Harmsen, H. J. M., Fu, J., Zhernakova, A., & Kurilshikov, A. (2022). *Choice of DNA extraction method affects stool microbiome recovery and subsequent phenotypic association analyses*. https://doi.org/10.21203/rs.3.rs-1967940/v1

Forslund, K., Hildebrand, F., Nielsen, T., Falony, G., le Chatelier, E., Sunagawa, S., Prifti, E., Vieira-Silva, S., Gudmundsdottir, V., Krogh Pedersen, H., Arumugam, M., Kristiansen, K., Yvonne Voigt, A., Vestergaard, H., Hercog, R., Igor Costea, P., Roat Kultima, J., Li, J., Jørgensen, T., … Pedersen, O. (2015). Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature*, *528*(7581), 262–266. https://doi.org/10.1038/nature15766

Forslund, S. K., Chakaroun, R., Zimmermann-Kogadeeva, M., Markó, L., Aron-Wisnewsky, J., Nielsen, T., Moitinho-Silva, L., Schmidt, T. S. B., Falony, G., Vieira-Silva, S., Adriouch, S., Alves, R. J., Assmann, K., Bastard, J.-P., Birkner, T., Caesar, R., Chilloux, J., Coelho, L. P., Fezeu, L., … Bork, P. (2021). Combinatorial, additive and dose-dependent drug–microbiome associations. *Nature*, *600*(7889), 500–505. https://doi.org/10.1038/s41586-021-04177-9

Gacesa, R., Kurilshikov, A., Vich Vila, A., Sinha, T., Klaassen, M. A. Y., Bolte, L. A., Andreu-Sánchez, S., Chen, L., Collij, V., Hu, S., Dekens, J. A. M., Lenters, V. C., Björk, J. R., Swarte, J. C., Swertz, M. A., Jansen, B. H., Gelderloos-Arends, J., Jankipersadsing, S., Hofker, M., … Weersma, R. K. (2022). Environmental factors shaping the gut microbiome in a Dutch population. *Nature*, *604*(7907), 732–739. https://doi.org/10.1038/s41586-022-04567-7

Gale, E. A. M. (2013). Is type 2 diabetes a category error? In *The Lancet*. https://doi.org/10.1016/S0140-6736(12)62207-7

García-Santisteban, I., Cilleros-Portet, A., Moyua-Ormazabal, E., Kurilshikov, A., Zhernakova, A., Garcia-Etxebarria, K., Fernandez-Jimenez, N., & Bilbao, J. R. (2020). A Two-Sample Mendelian Randomization Analysis Investigates Associations Between Gut Microbiota and Celiac Disease. *Nutrients*, *12*(5), 1420. https://doi.org/10.3390/nu12051420

Gibson, G. R., Hutkins, R., Sanders, M. E., Prescott, S. L., Reimer, R. A., Salminen, S. J., Scott, K., Stanton, C., Swanson, K. S., Cani, P. D., Verbeke, K., & Reid, G. (2017). Expert consensus document: The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. *Nature Reviews Gastroenterology & Hepatology*, *14*(8), 491–502. https://doi.org/10.1038/nrgastro.2017.75

Giliberti, R., Cavaliere, S., Mauriello, I. E., Ercolini, D., & Pasolli, E. (2022). Host phenotype classification from human microbiome data is mainly driven by the presence of microbial taxa. *PLOS Computational Biology*, *18*(4), e1010066. https://doi.org/10.1371/journal.pcbi.1010066

Gloor, G. B., & Reid, G. (2016). Compositional analysis: A valid approach to analyze microbiome high-throughput sequencing data. *Canadian Journal of Microbiology*. https://doi.org/10.1139/cjm-2015-0821

Gordon-Rodriguez, E., Quinn, T. P., & Cunningham, J. P. (2021). Learning sparse log-ratios for high-throughput sequencing data. *Bioinformatics*, *38*(1), 157–163. https://doi.org/10.1093/bioinformatics/btab645

Gordon-Rodriguez, E., Quinn, T. P., & Cunningham, J. P. (2022). *Data Augmentation for Compositional Data: Advancing Predictive Models of the Microbiome*.

Gou, W., Ling, C., He, Y., Jiang, Z., Fu, Y., Xu, F., Miao, Z., Sun, T., Lin, J., Zhu, H., Zhou, H., Chen, Y., & Zheng, J.-S. (2021). Interpretable Machine Learning Framework Reveals Robust Gut Microbiome Features Associated With Type 2 Diabetes. *Diabetes Care*, *44*(2), 358–366. https://doi.org/10.2337/dc20-1536

Greenacre, M., Grunsky, E., Bacon-Shone, J., Erb, I., & Quinn, T. (2022). *Aitchison's Compositional Data Analysis 40 Years On: A Reappraisal*.

Gui, X., Yang, Z., & Li, M. D. (2021). Effect of Cigarette Smoke on Gut Microbiota: State of Knowledge. *Frontiers in Physiology*, *12*. https://doi.org/10.3389/fphys.2021.673341

Gurung, M., Li, Z., You, H., Rodrigues, R., Jump, D. B., Morgun, A., & Shulzhenko, N. (2020). Role of gut microbiota in type 2 diabetes pathophysiology. In *EBioMedicine*. https://doi.org/10.1016/j.ebiom.2019.11.051

Hill, C., Guarner, F., Reid, G., Gibson, G. R., Merenstein, D. J., Pot, B., Morelli, L., Canani, R. B., Flint, H. J., Salminen, S., Calder, P. C., & Sanders, M. E. (2014). The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nature Reviews Gastroenterology & Hepatology*, *11*(8), 506–514. https://doi.org/10.1038/nrgastro.2014.66

Holmes, I., Harris, K., & Quince, C. (2012). Dirichlet Multinomial Mixtures: Generative Models for Microbial Metagenomics. *PLoS ONE*, *7*(2), e30126. https://doi.org/10.1371/journal.pone.0030126

Hong, J., Karaoz, U., de Valpine, P., & Fithian, W. (2022). To rarefy or not to rarefy: robustness and efficiency trade-offs of rarefying microbiome data. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btac127

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., Fitzgerald, M. G., Fulton, R. S., Giglio, M. G., Hallsworth-Pepin, K., Lobos, E. A., Madupu, R., Magrini, V., Martin, J. C., Mitreva, M., Muzny, D. M., Sodergren, E. J., … White, O. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207–214. https://doi.org/10.1038/nature11234

Jackson, M. A., Verdi, S., Maxan, M.-E., Shin, C. M., Zierer, J., Bowyer, R. C. E., Martin, T., Williams, F. M. K., Menni, C., Bell, J. T., Spector, T. D., & Steves, C. J. (2018). Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nature Communications*, *9*(1), 2655. https://doi.org/10.1038/s41467-018-05184-7

Johnston, C. S., Kim, C. M., & Buller, A. J. (2004). Vinegar Improves Insulin Sensitivity to a High-Carbohydrate Meal in Subjects with Insulin Resistance or Type 2 Diabetes [10]. In *Diabetes Care*. https://doi.org/10.2337/diacare.27.1.281

Karst, S. M., Ziels, R. M., Kirkegaard, R. H., Sørensen, E. A., McDonald, D., Zhu, Q., Knight, R., & Albertsen, M. (2021). High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nature Methods*, *18*(2), 165–169. https://doi.org/10.1038/s41592-020-01041-y

Kartal, E., Schmidt, T. S. B., Molina-Montes, E., Rodríguez-Perales, S., Wirbel, J., Maistrenko, O. M., Akanni, W. A., Alashkar Alhamwe, B., Alves, R. J., Carrato, A., Erasmus, H.-P., Estudillo, L., Finkelmeier, F., Fullam, A., Glazek, A. M., Gómez-Rubio, P., Hercog, R., Jung, F., Kandels, S., … Bork, P. (2022). A faecal microbiota signature with high specificity for pancreatic cancer. *Gut*, gutjnl-2021-324755. https://doi.org/10.1136/gutjnl-2021-324755

Keen, M. G. (2007). Microbial Life, Second Edition. James T. Stanley, Robert P. Gunsalus, Stephen Lory, and Jerome J. Perry. *Integrative and Comparative Biology*, *47*(6), 896–898. https://doi.org/10.1093/icb/icm096

Koren, O., Knights, D., Gonzalez, A., Waldron, L., Segata, N., Knight, R., Huttenhower, C., & Ley, R. E. (2013). A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets. *PLoS Computational Biology*, *9*(1), e1002863. https://doi.org/10.1371/journal.pcbi.1002863

Kostic, A. D., Xavier, R. J., & Gevers, D. (2014). The Microbiome in Inflammatory Bowel Disease: Current Status and the Future Ahead. *Gastroenterology*, *146*(6), 1489–1499. https://doi.org/10.1053/j.gastro.2014.02.009

Krautkramer, K. A., Fan, J., & Bäckhed, F. (2021). Gut microbial metabolites as multi-kingdom intermediates. *Nature Reviews Microbiology*, *19*(2), 77–94. https://doi.org/10.1038/s41579-020-0438-4

Kurilshikov, A., Medina-Gomez, C., Bacigalupe, R., Radjabzadeh, D., Wang, J., Demirkan, A., le Roy, C. I., Raygoza Garay, J. A., Finnicum, C. T., Liu, X., Zhernakova, D. v., Bonder, M. J., Hansen, T. H., Frost, F., Rühlemann, M. C., Turpin, W., Moon, J.-Y., Kim, H.-N., Lüll, K., … Zhernakova, A. (2021). Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nature Genetics*, *53*(2), 156–165. https://doi.org/10.1038/s41588-020-00763-1

Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., & Bonneau, R. A. (2015). Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLoS Computational Biology*, *11*(5), e1004226. https://doi.org/10.1371/journal.pcbi.1004226

Laakso, M., Kuusisto, J., Stančáková, A., Kuulasmaa, T., Pajukanta, P., Lusis, A. J., Collins, F. S., Mohlke, K. L., & Boehnke, M. (2017). The Metabolic Syndrome in Men study: A resource for studies of metabolic & cardiovascular diseases. *Journal of Lipid Research*. https://doi.org/10.1194/jlr.O072629

Lagier, J.-C., Dubourg, G., Million, M., Cadoret, F., Bilen, M., Fenollar, F., Levasseur, A., Rolain, J.-M., Fournier, P.-E., & Raoult, D. (2018). Culturing the human microbiota and culturomics. *Nature Reviews Microbiology*, *16*(9), 540–550. https://doi.org/10.1038/s41579-018-0041-0

Leeming, E. R., Johnson, A. J., Spector, T. D., & le Roy, C. I. (2019). Effect of Diet on the Gut Microbiota: Rethinking Intervention Duration. *Nutrients*, *11*(12), 2862. https://doi.org/10.3390/nu11122862

Leimbach, A., Hacker, J., & Dobrindt, U. (2013). *E. coli as an All-Rounder: The Thin Line Between Commensalism and Pathogenicity* (pp. 3–32). https://doi.org/10.1007/82_2012_303

Leitsalu, L., Haller, T., Esko, T., Tammesoo, M. L., Alavere, H., Snieder, H., Perola, M., Ng, P. C., Mägi, R., Milani, L., Fischer, K., & Metspalu, A. (2015). Cohort profile: Estonian biobank of the Estonian genome center, university of Tartu. *International Journal of Epidemiology*, *44*(4), 1137–1147. https://doi.org/10.1093/ije/dyt268

Leshem, A., Segal, E., & Elinav, E. (2020). The Gut Microbiome and Individual-Specific Responses to Diet. *MSystems*, *5*(5). https://doi.org/10.1128/mSystems.00665-20

Leung, H., Long, X., Ni, Y., Qian, L., Nychas, E., Siliceo, S. L., Pohl, D., Hanhineva, K., Liu, Y., Xu, A., Nielsen, H. B., Belda, E., Clément, K., Loomba, R., Li, H., Jia, W., & Panagiotou, G. (2022). Risk assessment with gut microbiome and metabolite markers in NAFLD development. *Science Translational Medicine*, *14*(648). https://doi.org/10.1126/scitranslmed.abk0855

Li, S. S., Zhu, A., Benes, V., Costea, P. I., Hercog, R., Hildebrand, F., Huerta-Cepas, J., Nieuwdorp, M., Salojärvi, J., Voigt, A. Y., Zeller, G., Sunagawa, S., de Vos, W. M., & Bork, P. (2016). Durable coexistence of donor and recipient strains after fecal microbiota transplantation. *Science*, *352*(6285), 586–589. https://doi.org/10.1126/science.aad8852

Liu, X., Tong, X., Zou, Y., Lin, X., Zhao, H., Tian, L., Jie, Z., Wang, Q., Zhang, Z., Lu, H., Xiao, L., Qiu, X., Zi, J., Wang, R., Xu, X., Yang, H., Wang, J., Zong, Y., Liu, W., … Zhang, T. (2022). Mendelian randomization analyses support causal relationships between blood metabolites and the gut microbiome. *Nature Genetics*, *54*(1), 52–61. https://doi.org/10.1038/s41588-021-00968-y

Liu, Y., Méric, G., Havulinna, A. S., Teo, S. M., Åberg, F., Ruuskanen, M., Sanders, J., Zhu, Q., Tripathi, A., Verspoor, K., Cheng, S., Jain, M., Jousilahti, P., Vázquez-Baeza, Y., Loomba, R., Lahti, L., Niiranen, T., Salomaa, V., Knight, R., & Inouye, M. (2022). Early prediction of incident liver disease using conventional risk factors and gut-microbiome-augmented gradient boosting. *Cell Metabolism*. https://doi.org/10.1016/j.cmet.2022.03.002

Lloyd-Price, J., Abu-Ali, G., & Huttenhower, C. (2016). The healthy human microbiome. *Genome Medicine*, *8*(1), 51. https://doi.org/10.1186/s13073-016-0307-y

Maier, L., Pruteanu, M., Kuhn, M., Zeller, G., Telzerow, A., Anderson, E. E., Brochado, A. R., Fernandez, K. C., Dose, H., Mori, H., Patil, K. R., Bork, P., & Typas, A. (2018). Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature*, *555*(7698), 623–628. https://doi.org/10.1038/nature25979

Mamlouk, D., & Gullo, M. (2013). Acetic Acid Bacteria: Physiology and Carbon Sources Oxidation. In *Indian Journal of Microbiology*. https://doi.org/10.1007/s12088-013-0414-z

Marcos-Zambrano, L. J., Karaduzovic-Hadziabdic, K., Loncar Turukalo, T., Przymus, P., Trajkovik, V., Aasmets, O., Berland, M., Gruca, A., Hasic, J., Hron, K., Klammsteiner, T., Kolev, M., Lahti, L., Lopes, M. B., Moreno, V., Naskinova, I., Org, E., Paciência, I., Papoutsoglou, G., … Truu, J. (2021). Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.634511

Matsuda, M., & DeFronzo, R. A. (1999). Insulin sensitivity indices obtained from oral glucose tolerance testing: Comparison with the euglycemic insulin clamp. *Diabetes Care*. https://doi.org/10.2337/diacare.22.9.1462

McCoubrey, L. E., Elbadawi, M., & Basit, A. W. (2022). Current clinical translation of microbiome medicines. *Trends in Pharmacological Sciences*, *43*(4), 281–292. https://doi.org/10.1016/j.tips.2022.02.001

Milanese, A., Mende, D. R., Paoli, L., Salazar, G., Ruscheweyh, H.-J., Cuenca, M., Hingamp, P., Alves, R., Costea, P. I., Coelho, L. P., Schmidt, T. S. B., Almeida, A., Mitchell, A. L., Finn, R. D., Huerta-Cepas, J., Bork, P., Zeller, G., & Sunagawa, S. (2019). Microbial abundance, activity and population genomic profiling with mOTUs2. *Nature Communications*, *10*(1), 1014. https://doi.org/10.1038/s41467-019-08844-4

Mitrou, P., Petsiou, E., Papakonstantinou, E., Maratou, E., Lambadiari, V., Dimitriadis, P., Spanoudi, F., Raptis, S. A., & Dimitriadis, G. (2015). The role of acetic acid on glucose uptake and blood flow rates in the skeletal muscle in humans with impaired glucose tolerance. *European Journal of Clinical Nutrition*. https://doi.org/10.1038/ejcn.2014.289

Morrison, D. J., & Preston, T. (2016). Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes*, *7*(3), 189–200. https://doi.org/10.1080/19490976.2015.1134082

Murray, C. J., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., McManigal, B., … Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*, *399*(10325), 629–655. https://doi.org/10.1016/S0140-6736(21)02724-0

Nearing, J. T., Douglas, G. M., Hayes, M. G., MacDonald, J., Desai, D. K., Allward, N., Jones, C. M. A., Wright, R. J., Dhanani, A. S., Comeau, A. M., & Langille, M. G. I. (2022). Microbiome differential abundance methods produce different results across 38 datasets. *Nature Communications*, *13*(1), 342. https://doi.org/10.1038/s41467-022-28034-z

Neville, B. A., Forster, S. C., & Lawley, T. D. (2018). Commensal Koch's postulates: establishing causation in human microbiota research. *Current Opinion in Microbiology*, *42*, 47–52. https://doi.org/10.1016/j.mib.2017.10.001

Nobs, S. P., Tuganbaev, T., & Elinav, E. (2019). Microbiome diurnal rhythmicity and its impact on host physiology and disease risk. *EMBO Reports*, *20*(4). https://doi.org/10.15252/embr.201847129

Olsson, L. M., Boulund, F., Nilsson, S., Khan, M. T., Gummesson, A., Fagerberg, L., Engstrand, L., Perkins, R., Uhlén, M., Bergström, G., Tremaroli, V., & Bäckhed, F. (2022). Dynamics of the normal gut microbiota: A longitudinal one-year population study in Sweden. *Cell Host & Microbe*. https://doi.org/10.1016/j.chom.2022.03.002

Padula, W., Chen, Y., Reid, N., Tierce, J., Popper, C., & Walz, K. (2020). PCN67 using information from the microbiome to improve the efficiency of selecting therapeutic options for melanoma: A cost-effectiveness analysis. *Value in Health*, *23*, S35. https://doi.org/10.1016/j.jval.2020.04.1570

Palleja, A., Mikkelsen, K. H., Forslund, S. K., Kashani, A., Allin, K. H., Nielsen, T., Hansen, T. H., Liang, S., Feng, Q., Zhang, C., Pyl, P. T., Coelho, L. P., Yang, H., Wang, J., Typas, A., Nielsen, M. F., Nielsen, H. B., Bork, P., Wang, J., … Pedersen, O. (2018). Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nature Microbiology*, *3*(11), 1255–1265. https://doi.org/10.1038/s41564-018-0257-9

Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., Martinez, X., Varela, E., Sarrabayrouse, G., Machiels, K., Vermeire, S., Sokol, H., Guarner, F., & Manichanh, C. (2017). A microbial signature for Crohn's disease. *Gut*, *66*(5), 813–822. https://doi.org/10.1136/gutjnl-2016-313235

Quinn, T., & Erb, I. (2019). Using balances to engineer features for the classification of health biomarkers: a new approach to balance selection. *BioRxiv*. https://doi.org/10.1101/600122

Quinn, T. P. (2021). *Stool Studies Don't Pass the Sniff Test: A Systematic Review of Human Gut Microbiome Research Suggests Widespread Misuse of Machine Learning*.

Quinn, T. P., Erb, I., Gloor, G., Notredame, C., Richardson, M. F., & Crowley, T. M. (2019). A field guide for the compositional analysis of any-omics data. *GigaScience*. https://doi.org/10.1093/gigascience/giz107

Quinn, T. P., Gordon-Rodriguez, E., & Erb, I. (2021). *A Critique of Differential Abundance Analysis, and Advocacy for an Alternative*. https://arxiv.org/abs/2104.07266v2

Quinn, T. P., Jacobs, S., Senadeera, M., Le, V., & Coghlan, S. (2022). The three ghosts of medical AI: Can the black-box present deliver? *Artificial Intelligence in Medicine*, *124*, 102158. https://doi.org/10.1016/j.artmed.2021.102158

Quraishi, M. N., Widlak, M., Bhala, N., Moore, D., Price, M., Sharma, N., & Iqbal, T. H. (2017). Systematic review with meta-analysis: the efficacy of faecal microbiota transplantation for the treatment of recurrent and refractory *Clostridium difficile* infection. *Alimentary Pharmacology & Therapeutics*, *46*(5), 479–493. https://doi.org/10.1111/apt.14201

Rivera-Pinto, J., Egozcue, J. J., Pawlowsky-Glahn, V., Paredes, R., Noguera-Julian, M., & Calle, M. L. (2018). Balances: a New Perspective for Microbiome Analysis. *MSystems*, *3*(4). https://doi.org/10.1128/mSystems.00053-18

Ruuskanen, M. O., Erawijantari, P. P., Havulinna, A. S., Liu, Y., Méric, G., Tuomilehto, J., Inouye, M., Jousilahti, P., Salomaa, V., Jain, M., Knight, R., Lahti, L., & Niiranen, T. J. (2022). Gut Microbiome Composition Is Predictive of Incident Type 2 Diabetes in a Population Cohort of 5,572 Finnish Adults. *Diabetes Care*, *45*(4), 811–818. https://doi.org/10.2337/dc21-2358

Salosensaari, A., Laitinen, V., Havulinna, A. S., Meric, G., Cheng, S., Perola, M., Valsta, L., Alfthan, G., Inouye, M., Watrous, J. D., Long, T., Salido, R. A., Sanders, K., Brennan, C., Humphrey, G. C., Sanders, J. G., Jain, M., Jousilahti, P., Salomaa, V., … Niiranen, T. (2021). Taxonomic signatures of cause-specific mortality risk in human gut microbiome. *Nature Communications*, *12*(1), 1–8. https://doi.org/10.1038/s41467-021-22962-y

Sanna, S., van Zuydam, N. R., Mahajan, A., Kurilshikov, A., Vich Vila, A., Võsa, U., Mujagic, Z., Masclee, A. A. M., Jonkers, D. M. A. E., Oosting, M., Joosten, L. A. B., Netea, M. G., Franke, L., Zhernakova, A., Fu, J., Wijmenga, C., & McCarthy, M. I. (2019). Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. In *Nature Genetics*. https://doi.org/10.1038/s41588-019-0350-x

Schnorr, S. L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., Turroni, S., Biagi, E., Peano, C., Severgnini, M., Fiori, J., Gotti, R., de Bellis, G., Luiselli, D., Brigidi, P., Mabulla, A., Marlowe, F., Henry, A. G., & Crittenden, A. N. (2014). Gut microbiome of the Hadza hunter-gatherers. *Nature Communications*, *5*(1), 3654. https://doi.org/10.1038/ncomms4654

Sczyrba, A., Hofmann, P., Belmann, P., Koslicki, D., Janssen, S., Dröge, J., Gregor, I., Majda, S., Fiedler, J., Dahms, E., Bremges, A., Fritz, A., Garrido-Oter, R., Jørgensen, T. S., Shapiro, N., Blood, P. D., Gurevich, A., Bai, Y., Turaev, D., … McHardy, A. C. (2017). Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software. *Nature Methods*, *14*(11), 1063–1071. https://doi.org/10.1038/nmeth.4458

Segovia-Rodríguez, L., Echeverry-Alzate, V., Rincón-Pérez, I., Calleja-Conde, J., Bühler, K. M., Giné, E., Albert, J., Hinojosa, J. A., Huertas, E., Gómez-Gallego, F., Bressa, C., Rodríguez de Fonseca, F., & López-Moreno, J. A. (2022). Gut microbiota and voluntary alcohol consumption. *Translational Psychiatry*, *12*(1), 146. https://doi.org/10.1038/s41398-022-01920-2

Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biology*, *14*(8), e1002533. https://doi.org/10.1371/journal.pbio.1002533

Shaffer, J. P., Carpenter, C. S., Martino, C., Salido, R. A., Minich, J. J., Bryant, M., Sanders, K., Schwartz, T., Humphrey, G., Swafford, A. D., & Knight, R. (2022). A comparison of six DNA extraction protocols for 16S, ITS and shotgun metagenomic sequencing of microbial communities. *BioTechniques*, *73*(1), 34–46. https://doi.org/10.2144/btn-2022-0032

Shanahan, F., Ghosh, T. S., & O'Toole, P. W. (2021). The Healthy Microbiome—What Is the Definition of a Healthy Gut Microbiome? *Gastroenterology*, *160*(2), 483–494. https://doi.org/10.1053/j.gastro.2020.09.057

Shanahan, F., & Hill, C. (2019). Language, numeracy and logic in microbiome science. *Nature Reviews Gastroenterology & Hepatology*, *16*(7), 387–388. https://doi.org/10.1038/s41575-019-0163-5

Shilo, S., Bar, N., Keshet, A., Talmor-Barkan, Y., Rossman, H., Godneva, A., Aviv, Y., Edlitz, Y., Reicher, L., Kolobkov, D., Wolf, B. C., Lotan-Pompan, M., Levi, K., Cohen, O., Saranga, H., Weinberger, A., & Segal, E. (n.d.). *Cohort profile: 10K – a large-scale prospective longitudinal study in Israel*. https://doi.org/10.1101/2021.02.19.21251487

Shilo, S., Godneva, A., Rachmiel, M., Korem, T., Kolobkov, D., Karady, T., Bar, N., Wolf, B. C., Glantz-Gashai, Y., Cohen, M., Zuckerman Levin, N., Shehadeh, N., Gruber, N., Levran, N., Koren, S., Weinberger, A., Pinhas-Hamiel, O., & Segal, E. (2022). Prediction of Personal Glycemic Responses to Food for Individuals With Type 1 Diabetes Through Integration of Clinical and Microbial Data. *Diabetes Care*, *45*(3), 502–511. https://doi.org/10.2337/dc21-1048

Shine, E. E., & Crawford, J. M. (2021). Molecules from the Microbiome. *Annual Review of Biochemistry*, *90*(1), 789–815. https://doi.org/10.1146/annurev-biochem-080320-115307

Simpson, C. A., Diaz-Arteche, C., Eliby, D., Schwartz, O. S., Simmons, J. G., & Cowan, C. S. M. (2021). The gut microbiota in anxiety and depression – A systematic review. *Clinical Psychology Review*, *83*, 101943. https://doi.org/10.1016/j.cpr.2020.101943

Singh, V. P., Proctor, S. D., & Willing, B. P. (2016). Koch's postulates, microbial dysbiosis and inflammatory bowel disease. *Clinical Microbiology and Infection*, *22*(7), 594–599. https://doi.org/10.1016/j.cmi.2016.04.018

Smits, S. A., Leach, J., Sonnenburg, E. D., Gonzalez, C. G., Lichtman, J. S., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Elias, J. E., Dominguez-Bello, M. G., & Sonnenburg, J. L. (2017). Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science*, *357*(6353), 802–806. https://doi.org/10.1126/science.aan4834

Sonnenburg, J. L., & Sonnenburg, E. D. (2019). Vulnerability of the industrialized microbiota. *Science*, *366*(6464). https://doi.org/10.1126/science.aaw9255

Sowah, S. A., Milanese, A., Schübel, R., Wirbel, J., Kartal, E., Johnson, T. S., Hirche, F., Grafetstätter, M., Nonnenmacher, T., Kirsten, R., López-Nogueroles, M., Lahoz, A., Schwarz, K. v., Okun, J. G., Ulrich, C. M., Nattenmüller, J., von Eckardstein, A., Müller, D., Stangl, G. I., … Zeller, G. (2022). Calorie restriction improves metabolic state independently of gut microbiome composition: a randomized dietary intervention trial. *Genome Medicine*, *14*(1), 30. https://doi.org/10.1186/s13073-022-01030-0

Suez, J., Cohen, Y., Valdés-Mas, R., Mor, U., Dori-Bachash, M., Federici, S., Zmora, N., Leshem, A., Heinemann, M., Linevsky, R., Zur, M., Ben-Zeev Brik, R., Bukimer, A., Eliyahu-Miller, S., Metz, A., Fischbein, R., Sharov, O., Malitsky, S., Itkin, M., … Elinav, E. (2022). Personalized microbiome-driven effects of non-nutritive sweeteners on human glucose tolerance. *Cell*. https://doi.org/10.1016/j.cell.2022.07.016

Suez, J., Zmora, N., Segal, E., & Elinav, E. (2019). The pros, cons, and many unknowns of probiotics. *Nature Medicine*, *25*(5), 716–729. https://doi.org/10.1038/s41591-019-0439-x

Suez, J., Zmora, N., Zilberman-Schapira, G., Mor, U., Dori-Bachash, M., Bashiardes, S., Zur, M., Regev-Lehavi, D., Ben-Zeev Brik, R., Federici, S., Horn, M., Cohen, Y., Moor, A. E., Zeevi, D., Korem, T., Kotler, E., Harmelin, A., Itzkovitz, S., Maharshak, N., … Elinav, E. (2018). Post-Antibiotic Gut Mucosal Microbiome Reconstitution Is Impaired by Probiotics and Improved by Autologous FMT. *Cell*, *174*(6), 1406-1423.e16. https://doi.org/10.1016/j.cell.2018.08.047

Sun, Z., Huang, S., Zhang, M., Zhu, Q., Haiminen, N., Carrieri, A. P., Vázquez-Baeza, Y., Parida, L., Kim, H. C., Knight, R., & Liu, Y. Y. (2021). Challenges in benchmarking metagenomic profilers. *Nature Methods*, *18*(6), 618–626. https://doi.org/10.1038/s41592-021-01141-3

Suzuki, T. A., Fitzstevens, L., Schmidt, V. T., Enav, H., Huus, K., Mbong, M., Adegbite, B. R., Zinsou, J. F., Esen, M., Velavan, T. P., Adegnika, A. A., Song, L. H., Spector, T. D., Muehlbauer, A. L., Marchi, N., Blekhman, R., Ségurel, L., Youngblut, N. D., Kremsner, P., & Ley, R. E. (n.d.). *Codiversification of gut microbiota with humans*. https://doi.org/10.1101/2021.10.12.462973

Tabák, A. G., Herder, C., Rathmann, W., Brunner, E. J., & Kivimäki, M. (2012). Prediabetes: A high-risk state for diabetes development. In *The Lancet*. https://doi.org/10.1016/S0140-6736(12)60283-9

Tedersoo, L., Albertsen, M., Anslan, S., & Callahan, B. (2021). Perspectives and Benefits of High-Throughput Long-Read Sequencing in Microbial Ecology. *Applied and Environmental Microbiology*, *87*(17). https://doi.org/10.1128/AEM.00626-21

Tolosana-Delgado, R., Talebi, H., Khodadadzadeh, M., & Boogaart, K. (2019). On machine learning algorithms and compositional data. *Proceedings of the 8th International Workshop on Compositional Data Analysis (CoDaWork2019)*, *November*, 172–175.

Topçuoğlu, B. D., Lesniak, N. A., Ruffin, M. T., Wiens, J., & Schloss, P. D. (2020). A Framework for Effective Application of Machine Learning to Microbiome-Based Classification Problems. *MBio*, *11*(3). https://doi.org/10.1128/mBio.00434-20

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

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, *444*(7122), 1027–1031. https://doi.org/10.1038/nature05414

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLOS ONE*, *14*(11), e0224365. https://doi.org/10.1371/journal.pone.0224365

Vandeputte, D., de Commer, L., Tito, R. Y., Kathagen, G., Sabino, J., Vermeire, S., Faust, K., & Raes, J. (2021). Temporal variability in quantitative human gut microbiome profiles and implications for clinical research. *Nature Communications 2021 12:1*, *12*(1), 1–13. https://doi.org/10.1038/s41467-021-27098-7

Vandeputte, D., Falony, G., Vieira-Silva, S., Tito, R. Y., Joossens, M., & Raes, J. (2016). Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut*, *65*(1), 57–62. https://doi.org/10.1136/gutjnl-2015-309618

Wang, H., Gou, W., Su, C., Du, W., Zhang, J., Miao, Z., Xiao, C., Jiang, Z., Wang, Z., Fu, Y., Jia, X., Ouyang, Y., Jiang, H., Huang, F., Li, L., Zhang, B., & Zheng, J.-S. (2022). Association of gut microbiota with glycaemic traits and incident type 2 diabetes, and modulation by habitual diet: a population-based longitudinal cohort study in Chinese adults. *Diabetologia*. https://doi.org/10.1007/s00125-022-05687-5

Wang, J., Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., Peng, Y., Zhang, D., Jie, Z., Wu, W., Qin, Y., Xue, W., Li, J., Han, L., … Wang, J. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. https://doi.org/10.1038/nature11450

Wang, Y., & LêCao, K.-A. (2020). Managing batch effects in microbiome data. *Briefings in Bioinformatics*, *21*(6), 1954–1970. https://doi.org/10.1093/bib/bbz105

Willis, A. D. (2019). Rarefaction, Alpha Diversity, and Statistics. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02407

Wilmanski, T., Kornilov, S. A., Diener, C., Conomos, M. P., Lovejoy, J. C., Sebastiani, P., Orwoll, E. S., Hood, L., Price, N. D., Rappaport, N., Magis, A. T., & Gibbons, S. M. (2022). Heterogeneity in statin responses explained by variation in the human gut microbiome. *Med*, *3*(6), 388-405.e6. https://doi.org/10.1016/j.medj.2022.04.007

Wilson, K. H., & Blitchington, R. B. (1996). Human colonic biota studied by ribosomal DNA sequence analysis. *Applied and Environmental Microbiology*, *62*(7), 2273–2278. https://doi.org/10.1128/aem.62.7.2273-2278.1996

Wirbel, J., Pyl, P. T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J. S., Voigt, A. Y., Palleja, A., Ponnudurai, R., Sunagawa, S., Coelho, L. P., Schrotz-King, P., Vogtmann, E., Habermann, N., Niméus, E., Thomas, A. M., Manghi, P., Gandini, S., … Zeller, G. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nature Medicine*, *25*(4), 679–689. https://doi.org/10.1038/s41591-019-0406-6

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

World Health Organization. (2016). Global Report on Diabetes. *Isbn*. https://doi.org/ISBN 978 92 4 156525 7

Wu, X., Ma, C., Han, L., Nawaz, M., Gao, F., Zhang, X., Yu, P., Zhao, C., Li, L., Zhou, A., Wang, J., Moore, J. E., Cherie Millar, B., & Xu, J. (2010). Molecular characterisation of the faecal microbiota in patients with type II diabetes. *Current Microbiology*. https://doi.org/10.1007/s00284-010-9582-9

Xu, F., Fu, Y., Sun, T., Jiang, Z., Miao, Z., Shuai, M., Gou, W., Ling, C., Yang, J., Wang, J., Chen, Y., & Zheng, J.-S. (2020). The interplay between host genetics and the gut microbiome reveals common and distinct microbiome features for complex human diseases. *Microbiome*, *8*(1), 145. https://doi.org/10.1186/s40168-020-00923-9

Ye, S. H., Siddle, K. J., Park, D. J., & Sabeti, P. C. (2019). Benchmarking Metagenomics Tools for Taxonomic Classification. *Cell*, *178*(4), 779–794. https://doi.org/10.1016/j.cell.2019.07.010

Zeevi, D., Korem, T., Godneva, A., Bar, N., Kurilshikov, A., Lotan-Pompan, M., Weinberger, A., Fu, J., Wijmenga, C., Zhernakova, A., & Segal, E. (2019). Structural variation in the gut microbiome associates with host health. *Nature*, *568*(7750), 43–48. https://doi.org/10.1038/s41586-019-1065-y

Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., Suez, J., Mahdi, J. A., Matot, E., Malka, G., Kosower, N., Rein, M., Zilberman-Schapira, G., Dohnalová, L., Pevsner-Fischer, M., … Segal, E. (2015). Personalized Nutrition by Prediction of Glycemic Responses. *Cell*, *163*(5), 1079–1094. https://doi.org/10.1016/j.cell.2015.11.001

Zeller, G., Tap, J., Voigt, A. Y., Sunagawa, S., Kultima, J. R., Costea, P. I., Amiot, A., Böhm, J., Brunetti, F., Habermann, N., Hercog, R., Koch, M., Luciani, A., Mende, D. R., Schneider, M. A., Schrotz-King, P., Tournigand, C., Tran Van Nhieu, J.,

Yamada, T., … Bork, P. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. *Molecular Systems Biology*, *10*(11), 766. https://doi.org/10.15252/msb.20145645

Zhernakova, A., Kurilshikov, A., Bonder, M. J., Tigchelaar, E. F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A. V., Falony, G., Vieira-Silva, S., Wang, J., Imhann, F., Brandsma, E., Jankipersadsing, S. A., Joossens, M., Cenit, M. C., Deelen, P., Swertz, M. A., Weersma, R. K., … Fu, J. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science*, *352*(6285), 565–569. https://doi.org/10.1126/science.aad3369

Zimmermann, M., Patil, K. R., Typas, A., & Maier, L. (2021). Towards a mechanistic understanding of reciprocal drug–microbiome interactions. *Molecular Systems Biology*, *17*(3). https://doi.org/10.15252/msb.202010116

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., & Goodman, A. L. (2019a). Separating host and microbiome contributions to drug pharmacokinetics and toxicity. *Science*, *363*(6427). https://doi.org/10.1126/science.aat9931

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., & Goodman, A. L. (2019b). Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature*, *570*(7762), 462–467. https://doi.org/10.1038/s41586-019-1291-3

ACKNOWLEDGMENTS

Dear friends and senior scientists. It has been a pleasure. The four years has passed rather quickly, perhaps due to the coffee breaks, summer days and pizza seminars. There has been a lot of inspiration, discussions, and motivation happening during these seemingly inadequate events, which I value a lot. Thank you! I remember a late-night beer in Narva-Jõesuu in my first days as a doctorate student at Institute of Genomics with certain gentlemen from the fourth floor who were well-spirited and cherished the work and colleagues in every way. Nothing has changed in that time, and I can proudly say that I truly share the beliefs and amazement of these two gentlemen. I am lucky with the colleagues. It is wonderful what you all have established and achieved during the years, and I am proud to be part of such a team. I hope your microbiomes behave nicely.

Dear Elin. I am truly astonished, how enthusiastic, passionate and energic you are in everything you engage in. I wouldn't have decided to take on such a challenge in the first place if you hadn't given me hope and made a damn good pitch for science. The microbiome team was, is and will be grateful for the work you are doing to keep us safe, well-fed, and motivated. Help, one-on-one discussions, and inspiration is always there if needed (and sometimes just jumps in due to mixing up the offices). Not to mention the countless laughs we own you. I have learned a lot from you, and I am trying to implement several of those lessons into my own way of living as you read this section. Nevertheless, I still have a feeling that I should have written things down. I hope that our collaboration continues because I can only visualize success stories.

Dear Kertu-Liis. If we had put down 10% of the ideas and discussions we've had over the years, I guess we would have graduated several times by now, redefined science, and made Elin proud. Presumably and hopefully these times are still to come. You have raised me from the motivational grave and stood up for me more than you know. You are like mini-Elin (another compliment, but the resemblance is truly uncanny). Thank you for your support and energy and I hope that our retirement plan concerning publishing in Nature still holds.

My dear companions, the road is long and there are too many of you to properly thank, but all you have had a part in this thesis, I owe you! Dear Krista, thank you for supporting and inspiring me from the very beginning of my academic life, which started already 10 years ago. Discussions of my Bachelor thesis in the Omicum gave me the first ideas that this is the place to be. Kreete, I truly admire your determination and stubbornness in taking up challenges and achieving your goals (Saare värk vist). Merli and Nele, go for it, there is nothing To Louse – Hadley, Haaadley (good times)! The beginning of doctorate studies is never easy, and I am grateful that I was granted you as my office roommates – our whining and working routine was a perfect match in my opinion! Dear fellow PhD students, you know how a single atom can feel lonely and sad, but together it makes up a marvelous structure that doesn't hold up on its own? That's right,

I am talking about diamonds! Finally, I would especially want to highlight prof. Reedik Mägi, who doesn't like spotlight.

Dear Marit, Ants and Aime. You all have your part in this thesis. I realized that I have tried to mix up your lessons and know-how throughout my career and choices, which has resulted in this writing. Ants, your interest in numbers, Aime, your influence on a 5-year-old Oliver to enjoy the nature and Marit with the clinical side. Thank you! Dear Lisett, you have way more courage and stubbornness than I do, which can lead to absolutely anywhere. Thank you for the inspiration!

Dear Maria, Nils, Mona and Nuru. You are super cool, keep going! Thank you for providing me the opportunity to do the thing I am doing and meet the people I have met. Nils and Mona, I guess you hope that one day your dad has a job with a title that you can truly explain, but I promise you nothing. Nils, at the time of defending my PhD thesis, you are 4 years old, so not much more than my road in science. If you happen to read this, keep in mind that within four years you can achieve a lot, meet a bunch of wonderful people, live through tons of emotions, and enjoy the virtues of progress and learning. It is never too late to make a move or change, which can become a great story. Mona, you will soon celebrate your first birthday. By now you know that one year is enough to do something crazy – for example learn to walk or train for a half-ironman triathlon like you mother does. So, dream big, put your goals down on a paper and yell it out loud. This way everything is possible. Dear Nuru, woof woof! Dear Maria, this is just one chapter of our story, what next? Love you!

PUBLICATIONS

CURRICULUM VITAE

Name: Oliver Aasmets
Date of birth: 17.09.1993
Contact: Estonian Genome Center, Institute of Genomics, University of Tartu Riia 23b, 51010, Tartu
E-mail: oliver.aasmets@ut.ee

Education:

2021–	PhD in Molecular Biomedicine, Institute of Genomics, Faculty of Science and Technology, University of Tartu, Estonia (transferred due to a change in the structural unit)
2018–2021	PhD in Molecular Biomedicine, Faculty of Science and Technology, Institute of Molecular and Cell Biology, University of Tartu, Estonia
2016–2018	Master's Degree in Mathematical Statistics, Faculty of Mathematics and statistics, University of Tartu, Estonia
2012–2015	Bachelor's Degree in Mathematical Statistics, Faculty of Mathematics and statistics, University of Tartu, Estonia

Professional employment:

2019–	Estonian Genome Center, Institute of Genomics, University of Tartu, specialist
2016–2018	StatFinn OY/IQVIA, statistician, statistical programmer
2014–2015	Seesam Insurance AS, junior UW, sales associate

Administrative work:

2018–	Member of Estonian Society of Human Genetics

Publications:

Krigul, Kertu Liis; **Aasmets, Oliver**; Lüll, Kreete; Org, Tõnis; Org, Elin (2021). Using fecal immunochemical tubes for the analysis of the gut microbiome has the potential to improve colorectal cancer screening. Scientific Reports, 11 (1). https://doi.org/10.1038/s41598-021-99046-w.

Marcos-Zambrano, Laura Judith; Karaduzovic-Hadziabdic, Kanita; Loncar Turukalo, Tatjana; Przymus, Piotr; Trajkovik, Vladimir; **Aasmets, Oliver**; Berland, Magali; Gruca, Aleksandra; Hasic, Jasminka; Hron, Karel; Klammsteiner, Thomas; Kolev, Mikhail; Lahti, Leo; Lopes, Marta B.; Moreno, Victor; Naskinova, Irina; Org, Elin; Paciencia, Ines; Papoutsoglou, Georgios; Shigdel, Rajesh; Stres, Blaz; Vilne, Baiba; Yousef, Malik; Zdravevski , Eftim; Tsamardinos, Ioannis; Carrillo de Santa Pau, Enrique; Claesson, Marcus J; Moreno-Indias, Isabel; Truu, Jaak (2021). Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment. Frontiers in Microbiology, 12, 313. https://doi.org/10.3389/fmicb.2021.634511.

Lüll, Kreete; Arffman, Riikka K; Sola-Leyva, Alberto; Molina, Nerea M; **Aasmets, Oliver**; Herzig, Karl-Heinz; Plaza-Díaz, Julio; Franks, Stephen; Morin-Papunen, Laure; Tapanainen, Juha S; Salumets, Andres; Altmäe, Signe; Piltonen, Terhi T; Org, Elin (2021). The Gut Microbiome in Polycystic Ovary Syndrome and its Association with Metabolic Traits. The Journal of Clinical Endocrinology & Metabolism. https://doi.org/10.1210/clinem/dgaa848.

Aasmets, Oliver; Lull, Kreete; Lang, Jennifer M.; Pan, Calvin; Kuusisto, Johanna; Fischer, Krista; Laakso, Markku; Lusis, Aldons J.; Org, Elin (2021). Machine Learning Reveals Time-Varying Microbial Predictors with Complex Effects on Glucose Regulation. mSystems, 6 (1), ARTN e01191-20. https://doi.org/10.1128/mSystems.01191-20.

Aasmets, Oliver; Krigul, Kertu Liis; Lüll, Kreete; Metspalu, Andres; Org, Elin (2022). Gut metagenome associations with extensive digital health data in a volunteer-based Estonian microbiome cohort. Nature Communications, 13 (1), 1–11. https://doi.org/10.1038/s41467-022-28464-9.

Ratnik, Kaspar; Rull; Kristiina; **Aasmets, Oliver**; Kikas; Triin; Hanson, Ele; Kisand; Kalle; Fischer, Krista (2022). Novel Early Pregnancy Multimarker Screening Test for Preeclampsia Risk Prediction. Frontiers in Cardiovascular Medicine, 9 (932480), 1–11. https://doi.org/10.3389/fcvm.2022.932480.

Aasmets, Oliver; Krigul, Kertu Liis; Org, Elin (2022). Evaluating the clinical relevance of the enterotypes in the Estonian Microbiome cohort. Frontiers in Genetics, 1943. https://doi.org/10.3389/fgene.2022.917926.

Supervised dissertations:

Triin Bulõgina	Bachelor's Degree, Institute of Mathematics and Statistics, University of Tartu, Estonia, "Colorectal cancer prediction models using gut microbiome as a predictor"
Kristina Muhu	Master's Degree, Institute of Mathematics and Statistics, University of Tartu, Estonia," Prediction models based on log-ratio transformations for diagnosing colorectal cancer using microbiome data"
Cheung Wai Chan	Master's Degree, Institute of Technology, University of Tartu, "Gut microbiome in adjustment disorder"
Jennifer Neumaier	Master's Degree, Faculty of Mathematics and Natural Sciences, University of Kiel, "Benchmarking machine learning performances with microbiome data"

Awards and stipends:

2020 COST Short Term Scientific Mission scholarship
2018 Scholarship in smart specialization growth areas

ELULOOKIRJELDUS

Nimi: Oliver Aasmets
Sünniaeg: 17.09.1993
Aadress: Eesti Geenivaramu, Genoomika instituut, Tartu Ülikool
Riia 23b, 51010, Tartu
E-post: oliver.aasmets@ut.ee

Haridus:

2021–	Doktoriõpe, Biomeditsiin, Genoomika Instituut, Loodus- ja täppisteaduste valdkond, Tartu Ülikool (viidud üle seoses struktuuriüksuse muutumisega)
2018–2021	Doktoriõpe, Biomeditsiin, Molekulaar- ja rakubioloogia Instituut, Loodus- ja täppisteaduste valdkond, Tartu Ülikool
2016–2018	Magistriõpe, Matemaatiline statistika, Matemaatilise statistika Instituut, Loodus- ja täppisteaduste valdkond, Tartu Ülikool
2012–2015	Bakalaureuseõpe, Matemaatiline statistika, Matemaatilise statistika Instituut, Loodus- ja täppisteaduste valdkond, Tartu Ülikool

Teenistuskäik:

2019–	Eesti Geenivaramu, Genoomika instituut, Tartu Ülikool, Spetsialist
2016–2018	StatFinn OÜ/IQVIA, statistik, statistiline programmeerija
2014–2015	Seesam Kindlustus AS, noorem UW

Teadusorganisatsiooniline ja- administratiivne tegevus:

2018–	Eesti Inimesegeneetika ühingu liige

Teaduspublikatsioonid:

Krigul, Kertu Liis; **Aasmets, Oliver**; Lüll, Kreete; Org, Tõnis; Org, Elin (2021). Using fecal immunochemical tubes for the analysis of the gut microbiome has the potential to improve colorectal cancer screening. Scientific Reports, 11 (1). https://doi.org/10.1038/s41598-021-99046-w.

Marcos-Zambrano, Laura Judith; Karaduzovic-Hadziabdic, Kanita; Loncar Turukalo, Tatjana; Przymus, Piotr; Trajkovik, Vladimir; **Aasmets, Oliver**; Berland, Magali; Gruca, Aleksandra; Hasic, Jasminka; Hron, Karel; Klammsteiner, Thomas; Kolev, Mikhail; Lahti, Leo; Lopes, Marta B.; Moreno, Victor; Naskinova, Irina; Org, Elin; Paciencia, Ines; Papoutsoglou, Georgios; Shigdel, Rajesh; Stres, Blaz; Vilne, Baiba; Yousef, Malik; Zdravevski , Eftim; Tsamardinos, Ioannis; Carrillo de Santa Pau, Enrique; Claesson, Marcus J; Moreno-Indias, Isabel; Truu, Jaak (2021). Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment. Frontiers in Microbiology, 12, 313. https://doi.org/10.3389/fmicb.2021.634511.

Lüll, Kreete; Arffman, Riikka K; Sola-Leyva, Alberto; Molina, Nerea M; **Aasmets, Oliver**; Herzig, Karl-Heinz; Plaza-Díaz, Julio; Franks, Stephen; Morin-Papunen, Laure; Tapanainen, Juha S; Salumets, Andres; Altmäe, Signe; Piltonen, Terhi T; Org, Elin (2021). The Gut Microbiome in Polycystic Ovary Syndrome and its Association with Metabolic Traits. The Journal of Clinical Endocrinology & Metabolism. https://doi.org/10.1210/clinem/dgaa848.

Aasmets, Oliver; Lull, Kreete; Lang, Jennifer M.; Pan, Calvin; Kuusisto, Johanna; Fischer, Krista; Laakso, Markku; Lusis, Aldons J.; Org, Elin (2021). Machine Learning Reveals Time-Varying Microbial Predictors with Complex Effects on Glucose Regulation. mSystems, 6 (1), ARTN e01191-20. https://doi.org/10.1128/mSystems.01191-20.

Aasmets, Oliver; Krigul, Kertu Liis; Lüll, Kreete; Metspalu, Andres; Org, Elin (2022). Gut metagenome associations with extensive digital health data in a volunteer-based Estonian microbiome cohort. Nature Communications, 13 (1), 1–11. https://doi.org/10.1038/s41467-022-28464-9.

Ratnik, Kaspar; Rull; Kristiina; **Aasmets, Oliver**; Kikas; Triin; Hanson, Ele; Kisand; Kalle; Fischer, Krista (2022). Novel Early Pregnancy Multimarker Screening Test for Preeclampsia Risk Prediction. Frontiers in Cardiovascular Medicine, 9 (932480), 1–11. https://doi.org/10.3389/fcvm.2022.932480.

Aasmets, Oliver; Krigul, Kertu Liis; Org, Elin (2022). Evaluating the clinical relevance of the enterotypes in the Estonian Microbiome cohort. Frontiers in Genetics, 1943. https://doi.org/10.3389/fgene.2022.917926.

Juhendatud väitekirjad:

Triin Bulõgina	Bakalaureusetöö, Tartu Ülikool, Loodus- ja täppisteaduste valdkond, matemaatika ja statistika instituut, "Mikrobioomil põhinevad ennustusmudelid soolevähi diagnoosimiseks"
Kristina Muhu	Magistritöö, Tartu Ülikool, Loodus- ja täppisteaduste valdkond, matemaatika ja statistika instituut, "Log-suhte teisendustel põhinevad ennustusmudelid soolevähi diagnoosimiseks mikrobioomi andmetelt"
Cheung Wai Chan	Magistritöö, Tartu Ülikool, Loodus- ja täppisteaduste valdkond, Tehnoloogiainstituut, "Soolestiku mikrobioomi roll kohanemishäire puhul"
Jennifer Neumaier	Magistritöö, Kieli Ülikool, Matemaatika- ja loodusteaduste valdkond, "Masinõppemeetodite võrdlus mikrobioomi andmetel"

Stipendiumid:

2020	COST võrgustiku lähetusstipendium
2018	Kõrghariduse erialastipendium nutika spetsialiseerumise kasvuvaldkondades

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida.* Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand**. The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.
19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.

20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro*. Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous grassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.
40. **Arnold Kristjuhan**. Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.
41. **Sulev Ingerpuu**. Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.

42. **Veljo Kisand**. Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa**. Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa**. Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik**. Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo**. Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo**. Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots**. Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero**. Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees**. Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks**. Cholecystokinin (CCK) – induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and serotonin. Tartu, 1999, 123 p.
52. **Ebe Sild**. Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva**. Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna**. Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro**. Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane**. Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm**. Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg**. Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild**. The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu**. Studies of the TOL plasmid transcription factor XylS. Tartu, 2000, 88 p.
61. **Dina Lepik**. Modulation of viral DNA replication by tumor suppressor protein p53. Tartu, 2000, 106 p.
62. **Kai Vellak**. Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu, 2000, 122 p.

63. **Jonne Kotta**. Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu, 2000, 160 p.
64. **Georg Martin**. Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000, 139 p.
65. **Silvia Sepp**. Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira**. On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000, 96 p.
67. **Priit Zingel**. The role of planktonic ciliates in lake ecosystems. Tartu, 2001, 111 p.
68. **Tiit Teder**. Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu, 2001, 122 p.
69. **Hannes Kollist**. Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu, 2001, 80 p.
70. **Reet Marits**. Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu, 2001, 112 p.
71. **Vallo Tilgar**. Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002, 126 p.
72. **Rita Hõrak**. Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002, 108 p.
73. **Liina Eek-Piirsoo**. The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002, 74 p.
74. **Krõõt Aasamaa**. Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002, 110 p.
75. **Nele Ingerpuu**. Bryophyte diversity and vascular plants. Tartu, 2002, 112 p.
76. **Neeme Tõnisson**. Mutation detection by primer extension on oligonucleotide microarrays. Tartu, 2002, 124 p.
77. **Margus Pensa**. Variation in needle retention of Scots pine in relation to leaf morphology, nitrogen conservation and tree age. Tartu, 2003, 110 p.
78. **Asko Lõhmus**. Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003, 168 p.
79. **Viljar Jaks**. p53 – a switch in cellular circuit. Tartu, 2003, 160 p.
80. **Jaana Männik**. Characterization and genetic studies of four ATP-binding cassette (ABC) transporters. Tartu, 2003, 140 p.
81. **Marek Sammul**. Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003, 159 p
82. **Ivar Ilves**. Virus-cell interactions in the replication cycle of bovine papillomavirus type 1. Tartu, 2003, 89 p.
83. **Andres Männik**. Design and characterization of a novel vector system based on the stable replicator of bovine papillomavirus type 1. Tartu, 2003, 109 p.

84. **Ivika Ostonen**. Fine root structure, dynamics and proportion in net primary production of Norway spruce forest ecosystem in relation to site conditions. Tartu, 2003, 158 p.
85. **Gudrun Veldre**. Somatic status of 12–15-year-old Tartu schoolchildren. Tartu, 2003, 199 p.
86. **Ülo Väli.** The greater spotted eagle *Aquila clanga* and the lesser spotted eagle *A. pomarina*: taxonomy, phylogeography and ecology. Tartu, 2004, 159 p.
87. **Aare Abroi**. The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004, 135 p.
88. **Tiina Kahre**. Cystic fibrosis in Estonia. Tartu, 2004, 116 p.
89. **Helen Orav-Kotta**. Habitat choice and feeding activity of benthic suspension feeders and mesograzers in the northern Baltic Sea. Tartu, 2004, 117 p.
90. **Maarja Öpik**. Diversity of arbuscular mycorrhizal fungi in the roots of perennial plants and their effect on plant performance. Tartu, 2004, 175 p.
91. **Kadri Tali**. Species structure of *Neotinea ustulata*. Tartu, 2004, 109 p.
92. **Kristiina Tambets**. Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: a phylogeographic approach. Tartu, 2004, 163 p.
93. **Arvi Jõers**. Regulation of p53-dependent transcription. Tartu, 2004, 103 p.
94. **Lilian Kadaja**. Studies on modulation of the activity of tumor suppressor protein p53. Tartu, 2004, 103 p.
95. **Jaak Truu**. Oil shale industry wastewater: impact on river microbial community and possibilities for bioremediation. Tartu, 2004, 128 p.
96. **Maire Peters**. Natural horizontal transfer of the *pheBA* operon. Tartu, 2004, 105 p.
97. **Ülo Maiväli**. Studies on the structure-function relationship of the bacterial ribosome. Tartu, 2004, 130 p.
98. **Merit Otsus**. Plant community regeneration and species diversity in dry calcareous grasslands. Tartu, 2004, 103 p.
99. **Mikk Heidemaa**. Systematic studies on sawflies of the genera *Dolerus, Empria,* and *Caliroa* (Hymenoptera: Tenthredinidae). Tartu, 2004, 167 p.
100. **Ilmar Tõnno**. The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and N_2 fixation in some Estonian lakes. Tartu, 2004, 111 p.
101. **Lauri Saks**. Immune function, parasites, and carotenoid-based ornaments in greenfinches. Tartu, 2004, 144 p.
102. **Siiri Rootsi**. Human Y-chromosomal variation in European populations. Tartu, 2004, 142 p.
103. **Eve Vedler**. Structure of the 2,4-dichloro-phenoxyacetic acid-degradative plasmid pEST4011. Tartu, 2005. 106 p.
104. **Andres Tover**. Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 126 p.
105. **Helen Udras**. Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005, 100 p.

106. **Ave Suija**. Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005, 162 p.
107. **Piret Lõhmus**. Forest lichens and their substrata in Estonia. Tartu, 2005, 162 p.
108. **Inga Lips**. Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005, 156 p.
109. **Krista Kaasik**. Circadian clock genes in mammalian clockwork, metabolism and behaviour. Tartu, 2005, 121 p.
110. **Juhan Javoiš**. The effects of experience on host acceptance in ovipositing moths. Tartu, 2005, 112 p.
111. **Tiina Sedman**. Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005, 103 p.
112. **Ruth Aguraiuja**. Hawaiian endemic fern lineage *Diellia* (Aspleniaceae): distribution, population structure and ecology. Tartu, 2005, 112 p.
113. **Riho Teras**. Regulation of transcription from the fusion promoters generated by transposition of Tn*4652* into the upstream region of *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 106 p.
114. **Mait Metspalu**. Through the course of prehistory in India: tracing the mtDNA trail. Tartu, 2005, 138 p.
115. **Elin Lõhmussaar**. The comparative patterns of linkage disequilibrium in European populations and its implication for genetic association studies. Tartu, 2006, 124 p.
116. **Priit Kupper**. Hydraulic and environmental limitations to leaf water relations in trees with respect to canopy position. Tartu, 2006, 126 p.
117. **Heili Ilves**. Stress-induced transposition of Tn*4652* in *Pseudomonas Putida.* Tartu, 2006, 120 p.
118. **Silja Kuusk**. Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006, 126 p.
119. **Kersti Püssa**. Forest edges on medium resolution landsat thematic mapper satellite images. Tartu, 2006, 90 p.
120. **Lea Tummeleht**. Physiological condition and immune function in great tits (*Parus major* l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006, 94 p.
121. **Toomas Esperk**. Larval instar as a key element of insect growth schedules. Tartu, 2006, 186 p.
122. **Harri Valdmann**. Lynx (*Lynx lynx*) and wolf (*Canis lupus*) in the Baltic region: Diets, helminth parasites and genetic variation. Tartu, 2006. 102 p.
123. **Priit Jõers**. Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisia*. Tartu, 2006. 113 p.
124. **Kersti Lilleväli**. Gata3 and Gata2 in inner ear development. Tartu, 2007, 123 p.
125. **Kai Rünk**. Comparative ecology of three fern species: *Dryopteris carthusiana* (Vill.) H.P. Fuchs, *D. expansa* (C. Presl) Fraser-Jenkins & Jermy and *D. dilatata* (Hoffm.) A. Gray (Dryopteridaceae). Tartu, 2007, 143 p.

126. **Aveliina Helm**. Formation and persistence of dry grassland diversity: role of human history and landscape structure. Tartu, 2007, 89 p.
127. **Leho Tedersoo**. Ectomycorrhizal fungi: diversity and community structure in Estonia, Seychelles and Australia. Tartu, 2007, 233 p.
128. **Marko Mägi**. The habitat-related variation of reproductive performance of great tits in a deciduous-coniferous forest mosaic: looking for causes and consequences. Tartu, 2007, 135 p.
129. **Valeria Lulla**. Replication strategies and applications of Semliki Forest virus. Tartu, 2007, 109 p.
130. **Ülle Reier**. Estonian threatened vascular plant species: causes of rarity and conservation. Tartu, 2007, 79 p.
131. **Inga Jüriado**. Diversity of lichen species in Estonia: influence of regional and local factors. Tartu, 2007, 171 p.
132. **Tatjana Krama**. Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007, 112 p.
133. **Signe Saumaa**. The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in *Pseudomonas putida.* Tartu, 2007, 172 p.
134. **Reedik Mägi**. The linkage disequilibrium and the selection of genetic markers for association studies in european populations. Tartu, 2007, 96 p.
135. **Priit Kilgas**. Blood parameters as indicators of physiological condition and skeletal development in great tits (*Parus major*): natural variation and application in the reproductive ecology of birds. Tartu, 2007, 129 p.
136. **Anu Albert**. The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007, 95 p.
137. **Kärt Padari**. Protein transduction mechanisms of transportans. Tartu, 2008, 128 p.
138. **Siiri-Lii Sandre**. Selective forces on larval colouration in a moth. Tartu, 2008, 125 p.
139. **Ülle Jõgar**. Conservation and restoration of semi-natural floodplain meadows and their rare plant species. Tartu, 2008, 99 p.
140. **Lauri Laanisto**. Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008, 133 p.
141. **Reidar Andreson**. Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008, 105 p.
142. **Birgot Paavel**. Bio-optical properties of turbid lakes. Tartu, 2008, 175 p.
143. **Kaire Torn**. Distribution and ecology of charophytes in the Baltic Sea. Tartu, 2008, 98 p.
144. **Vladimir Vimberg**. Peptide mediated macrolide resistance. Tartu, 2008, 190 p.
145. **Daima Örd**. Studies on the stress-inducible pseudokinase TRB3, a novel inhibitor of transcription factor ATF4. Tartu, 2008, 108 p.
146. **Lauri Saag**. Taxonomic and ecologic problems in the genus *Lepraria* (*Stereocaulaceae*, lichenised *Ascomycota*). Tartu, 2008, 175 p.

147. **Ulvi Karu**. Antioxidant protection, carotenoids and coccidians in greenfinches – assessment of the costs of immune activation and mechanisms of parasite resistance in a passerine with carotenoid-based ornaments. Tartu, 2008, 124 p.
148. **Jaanus Remm**. Tree-cavities in forests: density, characteristics and occupancy by animals. Tartu, 2008, 128 p.
149. **Epp Moks**. Tapeworm parasites *Echinococcus multilocularis* and *E. granulosus* in Estonia: phylogenetic relationships and occurrence in wild carnivores and ungulates. Tartu, 2008, 82 p.
150. **Eve Eensalu**. Acclimation of stomatal structure and function in tree canopy: effect of light and CO_2 concentration. Tartu, 2008, 108 p.
151. **Janne Pullat**. Design, functionlization and application of an *in situ* synthesized oligonucleotide microarray. Tartu, 2008, 108 p.
152. **Marta Putrinš**. Responses of *Pseudomonas putida* to phenol-induced metabolic and stress signals. Tartu, 2008, 142 p.
153. **Marina Semtšenko**. Plant root behaviour: responses to neighbours and physical obstructions. Tartu, 2008, 106 p.
154. **Marge Starast**. Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
155. **Age Tats**. Sequence motifs influencing the efficiency of translation. Tartu, 2009, 104 p.
156. **Radi Tegova**. The role of specialized DNA polymerases in mutagenesis in *Pseudomonas putida.* Tartu, 2009, 124 p.
157. **Tsipe Aavik**. Plant species richness, composition and functional trait pattern in agricultural landscapes – the role of land use intensity and landscape structure. Tartu, 2009, 112 p.
158. **Kaja Kiiver**. Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.
159. **Meelis Kadaja**. Papillomavirus Replication Machinery Induces Genomic Instability in its Host Cell. Tartu, 2009, 126 p.
160. **Pille Hallast**. Human and chimpanzee Luteinizing hormone/Chorionic Gonadotropin beta (*LHB/CGB*) gene clusters: diversity and divergence of young duplicated genes. Tartu, 2009, 168 p.
161. **Ain Vellak**. Spatial and temporal aspects of plant species conservation. Tartu, 2009, 86 p.
162. **Triinu Remmel**. Body size evolution in insects with different colouration strategies: the role of predation risk. Tartu, 2009, 168 p.
163. **Jaana Salujõe**. Zooplankton as the indicator of ecological quality and fish predation in lake ecosystems. Tartu, 2009, 129 p.
164. **Ele Vahtmäe**. Mapping benthic habitat with remote sensing in optically complex coastal environments. Tartu, 2009, 109 p.
165. **Liisa Metsamaa**. Model-based assessment to improve the use of remote sensing in recognition and quantitative mapping of cyanobacteria. Tartu, 2009, 114 p.

166. **Pille Säälik**. The role of endocytosis in the protein transduction by cell-penetrating peptides. Tartu, 2009, 155 p.
167. **Lauri Peil**. Ribosome assembly factors in *Escherichia coli*. Tartu, 2009, 147 p.
168. **Lea Hallik**. Generality and specificity in light harvesting, carbon gain capacity and shade tolerance among plant functional groups. Tartu, 2009, 99 p.
169. **Mariliis Tark**. Mutagenic potential of DNA damage repair and tolerance mechanisms under starvation stress. Tartu, 2009, 191 p.
170. **Riinu Rannap**. Impacts of habitat loss and restoration on amphibian populations. Tartu, 2009, 117 p.
171. **Maarja Adojaan**. Molecular variation of HIV-1 and the use of this knowledge in vaccine development. Tartu, 2009, 95 p.
172. **Signe Altmäe**. Genomics and transcriptomics of human induced ovarian folliculogenesis. Tartu, 2010, 179 p.
173. **Triin Suvi**. Mycorrhizal fungi of native and introduced trees in the Seychelles Islands. Tartu, 2010, 107 p.
174. **Velda Lauringson**. Role of suspension feeding in a brackish-water coastal sea. Tartu, 2010, 123 p.
175. **Eero Talts**. Photosynthetic cyclic electron transport – measurement and variably proton-coupled mechanism. Tartu, 2010, 121 p.
176. **Mari Nelis**. Genetic structure of the Estonian population and genetic distance from other populations of European descent. Tartu, 2010, 97 p.
177. **Kaarel Krjutškov**. Arrayed Primer Extension-2 as a multiplex PCR-based method for nucleic acid variation analysis: method and applications. Tartu, 2010, 129 p.
178. **Egle Köster**. Morphological and genetical variation within species complexes: *Anthyllis vulneraria* s. l. and *Alchemilla vulgaris* (coll.). Tartu, 2010, 101 p.
179. **Erki Õunap**. Systematic studies on the subfamily Sterrhinae (Lepidoptera: Geometridae). Tartu, 2010, 111 p.
180. **Merike Jõesaar**. Diversity of key catabolic genes at degradation of phenol and *p*-cresol in pseudomonads. Tartu, 2010, 125 p.
181. **Kristjan Herkül**. Effects of physical disturbance and habitat-modifying species on sediment properties and benthic communities in the northern Baltic Sea. Tartu, 2010, 123 p.
182. **Arto Pulk**. Studies on bacterial ribosomes by chemical modification approaches. Tartu, 2010, 161 p.
183. **Maria Põllupüü**. Ecological relations of cladocerans in a brackish-water ecosystem. Tartu, 2010, 126 p.
184. **Toomas Silla**. Study of the segregation mechanism of the Bovine Papillomavirus Type 1. Tartu, 2010, 188 p.
185. **Gyaneshwer Chaubey**. The demographic history of India: A perspective based on genetic evidence. Tartu, 2010, 184 p.

186. **Katrin Kepp**. Genes involved in cardiovascular traits: detection of genetic variation in Estonian and Czech populations. Tartu, 2010, 164 p.
187. **Virve Sõber**. The role of biotic interactions in plant reproductive performance. Tartu, 2010, 92 p.
188. **Kersti Kangro**. The response of phytoplankton community to the changes in nutrient loading. Tartu, 2010, 144 p.
189. **Joachim M. Gerhold**. Replication and Recombination of mitochondrial DNA in Yeast. Tartu, 2010, 120 p.
190. **Helen Tammert**. Ecological role of physiological and phylogenetic diversity in aquatic bacterial communities. Tartu, 2010, 140 p.
191. **Elle Rajandu**. Factors determining plant and lichen species diversity and composition in Estonian *Calamagrostis* and *Hepatica* site type forests. Tartu, 2010, 123 p.
192. **Paula Ann Kivistik**. ColR-ColS signalling system and transposition of Tn*4652* in the adaptation of *Pseudomonas putida*. Tartu, 2010, 118 p.
193. **Siim Sõber**. Blood pressure genetics: from candidate genes to genome-wide association studies. Tartu, 2011, 120 p.
194. **Kalle Kipper**. Studies on the role of helix 69 of 23S rRNA in the factor-dependent stages of translation initiation, elongation, and termination. Tartu, 2011, 178 p.
195. **Triinu Siibak**. Effect of antibiotics on ribosome assembly is indirect. Tartu, 2011, 134 p.
196. **Tambet Tõnissoo**. Identification and molecular analysis of the role of guanine nucleotide exchange factor RIC-8 in mouse development and neural function. Tartu, 2011, 110 p.
197. **Helin Räägel**. Multiple faces of cell-penetrating peptides – their intracellular trafficking, stability and endosomal escape during protein transduction. Tartu, 2011, 161 p.
198. **Andres Jaanus**. Phytoplankton in Estonian coastal waters – variability, trends and response to environmental pressures. Tartu, 2011, 157 p.
199. **Tiit Nikopensius**. Genetic predisposition to nonsyndromic orofacial clefts. Tartu, 2011, 152 p.
200. **Signe Värv**. Studies on the mechanisms of RNA polymerase II-dependent transcription elongation. Tartu, 2011, 108 p.
201. **Kristjan Välk**. Gene expression profiling and genome-wide association studies of non-small cell lung cancer. Tartu, 2011, 98 p.
202. **Arno Põllumäe**. Spatio-temporal patterns of native and invasive zooplankton species under changing climate and eutrophication conditions. Tartu, 2011, 153 p.
203. **Egle Tammeleht**. Brown bear (*Ursus arctos*) population structure, demographic processes and variations in diet in northern Eurasia. Tartu, 2011, 143 p.
205. **Teele Jairus**. Species composition and host preference among ectomycorrhizal fungi in Australian and African ecosystems. Tartu, 2011, 106 p.

206. **Kessy Abarenkov**. PlutoF – cloud database and computing services supporting biological research. Tartu, 2011, 125 p.
207. **Marina Grigorova**. Fine-scale genetic variation of follicle-stimulating hormone beta-subunit coding gene (*FSHB*) and its association with reproductive health. Tartu, 2011, 184 p.
208. **Anu Tiitsaar**. The effects of predation risk and habitat history on butterfly communities. Tartu, 2011, 97 p.
209. **Elin Sild**. Oxidative defences in immunoecological context: validation and application of assays for nitric oxide production and oxidative burst in a wild passerine. Tartu, 2011, 105 p.
210. **Irja Saar**. The taxonomy and phylogeny of the genera *Cystoderma* and *Cystodermella* (Agaricales, Fungi). Tartu, 2012, 167 p.
211. **Pauli Saag**. Natural variation in plumage bacterial assemblages in two wild breeding passerines. Tartu, 2012, 113 p.
212. **Aleksei Lulla**. Alphaviral nonstructural protease and its polyprotein substrate: arrangements for the perfect marriage. Tartu, 2012, 143 p.
213. **Mari Järve**. Different genetic perspectives on human history in Europe and the Caucasus: the stories told by uniparental and autosomal markers. Tartu, 2012, 119 p.
214. **Ott Scheler**. The application of tmRNA as a marker molecule in bacterial diagnostics using microarray and biosensor technology. Tartu, 2012, 93 p.
215. **Anna Balikova**. Studies on the functions of tumor-associated mucin-like leukosialin (CD43) in human cancer cells. Tartu, 2012, 129 p.
216. **Triinu Kõressaar**. Improvement of PCR primer design for detection of prokaryotic species. Tartu, 2012, 83 p.
217. **Tuul Sepp**. Hematological health state indices of greenfinches: sources of individual variation and responses to immune system manipulation. Tartu, 2012, 117 p.
218. **Rya Ero**. Modifier view of the bacterial ribosome. Tartu, 2012, 146 p.
219. **Mohammad Bahram**. Biogeography of ectomycorrhizal fungi across different spatial scales. Tartu, 2012, 165 p.
220. **Annely Lorents**. Overcoming the plasma membrane barrier: uptake of amphipathic cell-penetrating peptides induces influx of calcium ions and downstream responses. Tartu, 2012, 113 p.
221. **Katrin Männik**. Exploring the genomics of cognitive impairment: whole-genome SNP genotyping experience in Estonian patients and general population. Tartu, 2012, 171 p.
222. **Marko Prous**. Taxonomy and phylogeny of the sawfly genus *Empria* (Hymenoptera, Tenthredinidae). Tartu, 2012, 192 p.
223. **Triinu Visnapuu**. Levansucrases encoded in the genome of *Pseudomonas syringae* pv. tomato DC3000: heterologous expression, biochemical characterization, mutational analysis and spectrum of polymerization products. Tartu, 2012, 160 p.
224. **Nele Tamberg**. Studies on Semliki Forest virus replication and pathogenesis. Tartu, 2012, 109 p.

225. **Tõnu Esko**. Novel applications of SNP array data in the analysis of the genetic structure of Europeans and in genetic association studies. Tartu, 2012, 149 p.
226. **Timo Arula**. Ecology of early life-history stages of herring *Clupea harengus membras* in the northeastern Baltic Sea. Tartu, 2012, 143 p.
227. **Inga Hiiesalu**. Belowground plant diversity and coexistence patterns in grassland ecosystems. Tartu, 2012, 130 p.
228. **Kadri Koorem**. The influence of abiotic and biotic factors on small-scale plant community patterns and regeneration in boreonemoral forest. Tartu, 2012, 114 p.
229. **Liis Andresen**. Regulation of virulence in plant-pathogenic pectobacteria. Tartu, 2012, 122 p.
230. **Kaupo Kohv**. The direct and indirect effects of management on boreal forest structure and field layer vegetation. Tartu, 2012, 124 p.
231. **Mart Jüssi**. Living on an edge: landlocked seals in changing climate. Tartu, 2012, 114 p.
232. **Riina Klais**. Phytoplankton trends in the Baltic Sea. Tartu, 2012, 136 p.
233. **Rauno Veeroja**. Effects of winter weather, population density and timing of reproduction on life-history traits and population dynamics of moose (*Alces alces*) in Estonia. Tartu, 2012, 92 p.
234. **Marju Keis**. Brown bear (*Ursus arctos*) phylogeography in northern Eurasia. Tartu, 2013, 142 p.
235. **Sergei Põlme**. Biogeography and ecology of *alnus*- associated ectomycorrhizal fungi – from regional to global scale. Tartu, 2013, 90 p.
236. **Liis Uusküla**. Placental gene expression in normal and complicated pregnancy. Tartu, 2013, 173 p.
237. **Marko Lõoke**. Studies on DNA replication initiation in *Saccharomyces cerevisiae*. Tartu, 2013, 112 p.
238. **Anne Aan**. Light- and nitrogen-use and biomass allocation along productivity gradients in multilayer plant communities. Tartu, 2013, 127 p.
239. **Heidi Tamm**. Comprehending phylogenetic diversity – case studies in three groups of ascomycetes. Tartu, 2013, 136 p.
240. **Liina Kangur**. High-Pressure Spectroscopy Study of Chromophore-Binding Hydrogen Bonds in Light-Harvesting Complexes of Photosynthetic Bacteria. Tartu, 2013, 150 p.
241. **Margus Leppik**. Substrate specificity of the multisite specific pseudouridine synthase RluD. Tartu, 2013, 111 p.
242. **Lauris Kaplinski**. The application of oligonucleotide hybridization model for PCR and microarray optimization. Tartu, 2013, 103 p.
243. **Merli Pärnoja**. Patterns of macrophyte distribution and productivity in coastal ecosystems: effect of abiotic and biotic forcing. Tartu, 2013, 155 p.
244. **Tõnu Margus**. Distribution and phylogeny of the bacterial translational GTPases and the Mqsr/YgiT regulatory system. Tartu, 2013, 126 p.
245. **Pille Mänd**. Light use capacity and carbon and nitrogen budget of plants: remote assessment and physiological determinants. Tartu, 2013, 128 p.

246. **Mario Plaas**. Animal model of Wolfram Syndrome in mice: behavioural, biochemical and psychopharmacological characterization. Tartu, 2013, 144 p.
247. **Georgi Hudjašov**. Maps of mitochondrial DNA, Y-chromosome and tyrosinase variation in Eurasian and Oceanian populations. Tartu, 2013, 115 p.
248. **Mari Lepik**. Plasticity to light in herbaceous plants and its importance for community structure and diversity. Tartu, 2013, 102 p.
249. **Ede Leppik**. Diversity of lichens in semi-natural habitats of Estonia. Tartu, 2013, 151 p.
250. **Ülle Saks**. Arbuscular mycorrhizal fungal diversity patterns in boreonemoral forest ecosystems. Tartu, 2013, 151 p.
251. **Eneli Oitmaa**. Development of arrayed primer extension microarray assays for molecular diagnostic applications. Tartu, 2013, 147 p.
252. **Jekaterina Jutkina**. The horizontal gene pool for aromatics degradation: bacterial catabolic plasmids of the Baltic Sea aquatic system. Tartu, 2013, 121 p.
253. **Helen Vellau**. Reaction norms for size and age at maturity in insects: rules and exceptions. Tartu, 2014, 132 p.
254. **Randel Kreitsberg**. Using biomarkers in assessment of environmental contamination in fish – new perspectives. Tartu, 2014, 107 p.
255. **Krista Takkis**. Changes in plant species richness and population performance in response to habitat loss and fragmentation.Tartu, 2014, 141 p.
256. **Liina Nagirnaja**. Global and fine-scale genetic determinants of recurrent pregnancy loss. Tartu, 2014, 211 p.
257. **Triin Triisberg**. Factors influencing the re-vegetation of abandoned extracted peatlands in Estonia. Tartu, 2014, 133 p.
258. **Villu Soon**. A phylogenetic revision of the *Chrysis ignita* species group (Hymenoptera: Chrysididae) with emphasis on the northern European fauna. Tartu, 2014, 211 p.
259. **Andrei Nikonov**. RNA-Dependent RNA Polymerase Activity as a Basis for the Detection of Positive-Strand RNA Viruses by Vertebrate Host Cells. Tartu, 2014, 207 p.
260. **Eele Õunapuu-Pikas**. Spatio-temporal variability of leaf hydraulic conductance in woody plants: ecophysiological consequences. Tartu, 2014, 135 p.
261. **Marju Männiste**. Physiological ecology of greenfinches: information content of feathers in relation to immune function and behavior. Tartu, 2014, 121 p.
262. **Katre Kets**. Effects of elevated concentrations of CO_2 and O_3 on leaf photosynthetic parameters in *Populus tremuloides*: diurnal, seasonal and interannual patterns. Tartu, 2014, 115 p.
263. **Külli Lokko**. Seasonal and spatial variability of zoopsammon communities in relation to environmental parameters. Tartu, 2014, 129 p.
264. **Olga Žilina**. Chromosomal microarray analysis as diagnostic tool: Estonian experience. Tartu, 2014, 152 p.

265. **Kertu Lõhmus**. Colonisation ecology of forest-dwelling vascular plants and the conservation value of rural manor parks. Tartu, 2014, 111 p.
266. **Anu Aun**. Mitochondria as integral modulators of cellular signaling. Tartu, 2014, 167 p.
267. **Chandana Basu Mallick**. Genetics of adaptive traits and gender-specific demographic processes in South Asian populations. Tartu, 2014, 160 p.
268. **Riin Tamme**. The relationship between small-scale environmental heterogeneity and plant species diversity. Tartu, 2014, 130 p.
269. **Liina Remm**. Impacts of forest drainage on biodiversity and habitat quality: implications for sustainable management and conservation. Tartu, 2015, 126 p.
270. **Tiina Talve**. Genetic diversity and taxonomy within the genus *Rhinanthus.* Tartu, 2015, 106 p.
271. **Mehis Rohtla**. Otolith sclerochronological studies on migrations, spawning habitat preferences and age of freshwater fishes inhabiting the Baltic Sea. Tartu, 2015, 137 p.
272. **Alexey Reshchikov**. The world fauna of the genus *Lathrolestes* (Hymenoptera, Ichneumonidae). Tartu, 2015, 247 p.
273. **Martin Pook**. Studies on artificial and extracellular matrix protein-rich surfaces as regulators of cell growth and differentiation. Tartu, 2015, 142 p.
274. **Mai Kukumägi**. Factors affecting soil respiration and its components in silver birch and Norway spruce stands. Tartu, 2015, 155 p.
275. **Helen Karu**. Development of ecosystems under human activity in the North-East Estonian industrial region: forests on post-mining sites and bogs. Tartu, 2015, 152 p.
276. **Hedi Peterson**. Exploiting high-throughput data for establishing relationships between genes. Tartu, 2015, 186 p.
277. **Priit Adler**. Analysis and visualisation of large scale microarray data, Tartu, 2015, 126 p.
278. **Aigar Niglas**. Effects of environmental factors on gas exchange in deciduous trees: focus on photosynthetic water-use efficiency. Tartu, 2015, 152 p.
279. **Silja Laht**. Classification and identification of conopeptides using profile hidden Markov models and position-specific scoring matrices. Tartu, 2015, 100 p.
280. **Martin Kesler**. Biological characteristics and restoration of Atlantic salmon *Salmo salar* populations in the Rivers of Northern Estonia. Tartu, 2015, 97 p.
281. **Pratyush Kumar Das**. Biochemical perspective on alphaviral nonstructural protein 2: a tale from multiple domains to enzymatic profiling. Tartu, 2015, 205 p
282. **Priit Palta**. Computational methods for DNA copy number detection. Tartu, 2015, 130 p.
283. **Julia Sidorenko**. Combating DNA damage and maintenance of genome integrity in pseudomonads. Tartu, 2015, 174 p.

284. **Anastasiia Kovtun-Kante**. Charophytes of Estonian inland and coastal waters: distribution and environmental preferences. Tartu, 2015, 97 p.
285. **Ly Lindman**. The ecology of protected butterfly species in Estonia. Tartu, 2015, 171 p.
286. **Jaanis Lodjak**. Association of Insulin-like Growth Factor I and Corticosterone with Nestling Growth and Fledging Success in Wild Passerines. Tartu, 2016, 113 p.
287. **Ann Kraut**. Conservation of Wood-Inhabiting Biodiversity – Semi-Natural Forests as an Opportunity. Tartu, 2016, 141 p.
288. **Tiit Örd**. Functions and regulation of the mammalian pseudokinase TRIB3. Tartu, 2016, 182. p.
289. **Kairi Käiro**. Biological Quality According to Macroinvertebrates in Streams of Estonia (Baltic Ecoregion of Europe): Effects of Human-induced Hydromorphological Changes. Tartu, 2016, 126 p.
290. **Leidi Laurimaa**. *Echinococcus multilocularis* and other zoonotic parasites in Estonian canids. Tartu, 2016, 144 p.
291. **Helerin Margus**. Characterization of cell-penetrating peptide/nucleic acid nanocomplexes and their cell-entry mechanisms. Tartu, 2016, 173 p.
292. **Kadri Runnel**. Fungal targets and tools for forest conservation. Tartu, 2016, 157 p.
293. **Urmo Võsa**. MicroRNAs in disease and health: aberrant regulation in lung cancer and association with genomic variation. Tartu, 2016, 163 p.
294. **Kristina Mäemets-Allas**. Studies on cell growth promoting AKT signaling pathway – a promising anti-cancer drug target. Tartu, 2016, 146 p.
295. **Janeli Viil**. Studies on cellular and molecular mechanisms that drive normal and regenerative processes in the liver and pathological processes in Dupuytren's contracture. Tartu, 2016, 175 p.
296. **Ene Kook**. Genetic diversity and evolution of *Pulmonaria angustifolia* L. and *Myosotis laxa sensu lato* (Boraginaceae). Tartu, 2016, 106 p.
297. **Kadri Peil**. RNA polymerase II-dependent transcription elongation in *Saccharomyces cerevisiae*. Tartu, 2016, 113 p.
298. **Katrin Ruisu**. The role of RIC8A in mouse development and its function in cell-matrix adhesion and actin cytoskeletal organisation. Tartu, 2016, 129 p.
299. **Janely Pae**. Translocation of cell-penetrating peptides across biological membranes and interactions with plasma membrane constituents. Tartu, 2016, 126 p.
300. **Argo Ronk**. Plant diversity patterns across Europe: observed and dark diversity. Tartu, 2016, 153 p.
301. **Kristiina Mark**. Diversification and species delimitation of lichenized fungi in selected groups of the family Parmeliaceae (Ascomycota). Tartu, 2016, 181 p.
302. **Jaak-Albert Metsoja**. Vegetation dynamics in floodplain meadows: influence of mowing and sediment application. Tartu, 2016, 140 p.

303. **Hedvig Tamman**. The GraTA toxin-antitoxin system of *Pseudomonas putida*: regulation and role in stress tolerance. Tartu, 2016, 154 p.
304. **Kadri Pärtel**. Application of ultrastructural and molecular data in the taxonomy of helotialean fungi. Tartu, 2016, 183 p.
305. **Maris Hindrikson**. Grey wolf (*Canis lupus*) populations in Estonia and Europe: genetic diversity, population structure and -processes, and hybridization between wolves and dogs. Tartu, 2016, 121 p.
306. **Polina Degtjarenko**. Impacts of alkaline dust pollution on biodiversity of plants and lichens: from communities to genetic diversity. Tartu, 2016, 126 p.
307. **Liina Pajusalu**. The effect of CO_2 enrichment on net photosynthesis of macrophytes in a brackish water environment. Tartu, 2016, 126 p.
308. **Stoyan Tankov**. Random walks in the stringent response. Tartu, 2016, 94 p.
309. **Liis Leitsalu**. Communicating genomic research results to population-based biobank participants. Tartu, 2016, 158 p.
310. **Richard Meitern**. Redox physiology of wild birds: validation and application of techniques for detecting oxidative stress. Tartu, 2016, 134 p.
311. **Kaie Lokk**. Comparative genome-wide DNA methylation studies of healthy human tissues and non-small cell lung cancer tissue. Tartu, 2016, 127 p.
312. **Mihhail Kurašin**. Processivity of cellulases and chitinases. Tartu, 2017, 132 p.
313. **Carmen Tali**. Scavenger receptors as a target for nucleic acid delivery with peptide vectors. Tartu, 2017, 155 p.
314. **Katarina Oganjan**. Distribution, feeding and habitat of benthic suspension feeders in a shallow coastal sea. Tartu, 2017, 132 p.
315. **Taavi Paal**. Immigration limitation of forest plants into wooded landscape corridors. Tartu, 2017, 145 p.
316. **Kadri Õunap**. The Williams-Beuren syndrome chromosome region protein WBSCR22 is a ribosome biogenesis factor. Tartu, 2017, 135 p.
317. **Riin Tamm**. In-depth analysis of factors affecting variability in thiopurine methyltransferase activity. Tartu, 2017, 170 p.
318. **Keiu Kask**. The role of RIC8A in the development and regulation of mouse nervous system. Tartu, 2017, 184 p.
319. **Tiia Möller**. Mapping and modelling of the spatial distribution of benthic macrovegetation in the NE Baltic Sea with a special focus on the eelgrass *Zostera marina* Linnaeus, 1753. Tartu, 2017, 162 p.
320. **Silva Kasela**. Genetic regulation of gene expression: detection of tissue- and cell type-specific effects. Tartu, 2017, 150 p.
321. **Karmen Süld**. Food habits, parasites and space use of the raccoon dog *Nyctereutes procyonoides*: the role of an alien species as a predator and vector of zoonotic diseases in Estonia. Tartu, 2017, p.
322. **Ragne Oja**. Consequences of supplementary feeding of wild boar – concern for ground-nesting birds and endoparasite infection. Tartu, 2017, 141 p.
323. **Riin Kont**. The acquisition of cellulose chain by a processive cellobiohydrolase. Tartu, 2017, 117 p.

324. **Liis Kasari**. Plant diversity of semi-natural grasslands: drivers, current status and conservation challenges. Tartu, 2017, 141 p.
325. **Sirgi Saar**. Belowground interactions: the roles of plant genetic relatedness, root exudation and soil legacies. Tartu, 2017, 113 p.
326. **Sten Anslan**. Molecular identification of Collembola and their fungal associates. Tartu, 2017, 125 p.
327. **Imre Taal**. Causes of variation in littoral fish communities of the Eastern Baltic Sea: from community structure to individual life histories. Tartu, 2017, 118 p.
328. **Jürgen Jalak**. Dissecting the Mechanism of Enzymatic Degradation of Cellulose Using Low Molecular Weight Model Substrates. Tartu, 2017, 137 p.
329. **Kairi Kiik**. Reproduction and behaviour of the endangered European mink (*Mustela lutreola*) in captivity. Tartu, 2018, 112 p.
330. **Ivan Kuprijanov**. Habitat use and trophic interactions of native and invasive predatory macroinvertebrates in the northern Baltic Sea. Tartu, 2018, 117 p.
331. **Hendrik Meister**. Evolutionary ecology of insect growth: from geographic patterns to biochemical trade-offs. Tartu, 2018, 147 p.
332. **Ilja Gaidutšik**. Irc3 is a mitochondrial branch migration enzyme in *Saccharomyces cerevisiae.* Tartu, 2018, 161 p.
333. **Lena Neuenkamp**. The dynamics of plant and arbuscular mycorrhizal fungal communities in grasslands under changing land use. Tartu, 2018, 241 p.
334. **Laura Kasak**. Genome structural variation modulating the placenta and pregnancy maintenance. Tartu, 2018, 181 p.
335. **Kersti Riibak**. Importance of dispersal limitation in determining dark diversity of plants across spatial scales. Tartu, 2018, 133 p.
336. **Liina Saar**. Dynamics of grassland plant diversity in changing landscapes. Tartu, 2018, 206 p.
337. **Hanna Ainelo**. Fis regulates *Pseudomonas putida* biofilm formation by controlling the expression of *lapA*. Tartu, 2018, 143 p.
338. **Natalia Pervjakova**. Genomic imprinting in complex traits. Tartu, 2018, 176 p.
339. **Andrio Lahesaare**. The role of global regulator Fis in regulating the expression of *lapF* and the hydrophobicity of soil bacterium *Pseudomonas putida*. Tartu, 2018, 124 p.
340. **Märt Roosaare**. *K*-mer based methods for the identification of bacteria and plasmids. Tartu, 2018, 117 p.
341. **Maria Abakumova**. The relationship between competitive behaviour and the frequency and identity of neighbours in temperate grassland plants. Tartu, 2018, 104 p.
342. **Margus Vilbas**. Biotic interactions affecting habitat use of myrmecophilous butterflies in Northern Europe. Tartu, 2018, 142 p.

343. **Liina Kinkar**. Global patterns of genetic diversity and phylogeography of *Echinococcus granulosus* sensu stricto – a tapeworm species of significant public health concern. Tartu, 2018, 147 p.
344. **Teivi Laurimäe**. Taxonomy and genetic diversity of zoonotic tapeworms in the species complex of *Echinococcus granulosus* sensu lato. Tartu, 2018, 143 p.
345. **Tatjana Jatsenko**. Role of translesion DNA polymerases in mutagenesis and DNA damage tolerance in Pseudomonads. Tartu, 2018, 216 p.
346. **Katrin Viigand**. Utilization of α-glucosidic sugars by *Ogataea* (*Hansenula*) *polymorpha.* Tartu, 2018, 148 p.
347. **Andres Ainelo**. Physiological effects of the *Pseudomonas putida* toxin grat. Tartu, 2018, 146 p.
348. **Killu Timm**. Effects of two genes (DRD4 and SERT) on great tit (*Parus major*) behaviour and reproductive traits. Tartu, 2018, 117 p.
349. **Petr Kohout**. Ecology of ericoid mycorrhizal fungi. Tartu, 2018, 184 p.
350. **Gristin Rohula-Okunev**. Effects of endogenous and environmental factors on night-time water flux in deciduous woody tree species. Tartu, 2018, 184 p.
351. **Jane Oja**. Temporal and spatial patterns of orchid mycorrhizal fungi in forest and grassland ecosystems. Tartu, 2018, 102 p.
352. **Janek Urvik**. Multidimensionality of aging in a long-lived seabird. Tartu, 2018, 135 p.
353. **Lisanna Schmidt**. Phenotypic and genetic differentiation in the hybridizing species pair *Carex flava* and *C. viridula* in geographically different regions. Tartu, 2018, 133 p.
354. **Monika Karmin**. Perspectives from human Y chromosome – phylogeny, population dynamics and founder events. Tartu, 2018, 168 p.
355. **Maris Alver**. Value of genomics for atherosclerotic cardiovascular disease risk prediction. Tartu, 2019, 148 p.
356. **Lehti Saag**. The prehistory of Estonia from a genetic perspective: new insights from ancient DNA. Tartu, 2019, 171 p.
357. **Mari-Liis Viljur**. Local and landscape effects on butterfly assemblages in managed forests. Tartu, 2019, 115 p.
358. **Ivan Kisly**. The pleiotropic functions of ribosomal proteins eL19 and eL24 in the budding yeast ribosome. Tartu, 2019, 170 p.
359. **Mikk Puustusmaa**. On the origin of papillomavirus proteins. Tartu, 2019, 152 p.
360. **Anneliis Peterson**. Benthic biodiversity in the north-eastern Baltic Sea: mapping methods, spatial patterns, and relations to environmental gradients. Tartu, 2019, 159 p.
361. **Erwan Pennarun**. Meandering along the mtDNA phylogeny; causerie and digression about what it can tell us about human migrations. Tartu, 2019, 162 p.

362. **Karin Ernits**. Levansucrase Lsc3 and endo-levanase BT1760: characterization and application for the synthesis of novel prebiotics. Tartu, 2019, 217 p.
363. **Sille Holm**. Comparative ecology of geometrid moths: in search of contrasts between a temperate and a tropical forest. Tartu, 2019, 135 p.
364. **Anne-Mai Ilumäe**. Genetic history of the Uralic-speaking peoples as seen through the paternal haplogroup N and autosomal variation of northern Eurasians. Tartu, 2019, 172 p.
365. **Anu Lepik**. Plant competitive behaviour: relationships with functional traits and soil processes. Tartu, 2019, 152 p.
366. **Kunter Tätte**. Towards an integrated view of escape decisions in birds under variable levels of predation risk. Tartu, 2020, 172 p.
367. **Kaarin Parts**. The impact of climate change on fine roots and root-associated microbial communities in birch and spruce forests. Tartu, 2020, 143 p.
368. **Viktorija Kukuškina**. Understanding the mechanisms of endometrial receptivity through integration of 'omics' data layers. Tartu, 2020, 169 p.
369. **Martti Vasar**. Developing a bioinformatics pipeline gDAT to analyse arbuscular mycorrhizal fungal communities using sequence data from different marker regions. Tartu, 2020, 193 p.
370. **Ott Kangur**. Nocturnal water relations and predawn water potential disequilibrium in temperate deciduous tree species. Tartu, 2020, 126 p.
371. **Helen Post**. Overview of the phylogeny and phylogeography of the Y-chromosomal haplogroup N in northern Eurasia and case studies of two linguistically exceptional populations of Europe – Hungarians and Kalmyks. Tartu, 2020, 143 p.
372. **Kristi Krebs**. Exploring the genetics of adverse events in pharmacotherapy using Biobanks and Electronic Health Records. Tartu, 2020, 151 p.
373. **Kärt Ukkivi**. Mutagenic effect of transcription and transcription-coupled repair factors in *Pseudomonas putida.* Tartu, 2020, 154 p.
374. **Elin Soomets**. Focal species in wetland restoration. Tartu, 2020, 137 p.
375. **Kadi Tilk**. Signals and responses of ColRS two-component system in *Pseudomonas putida.* Tartu, 2020, 133 p.
376. **Indrek Teino**. Studies on aryl hydrocarbon receptor in the mouse granulosa cell model. Tartu, 2020, 139 p.
377. **Maarja Vaikre**. The impact of forest drainage on macroinvertebrates and amphibians in small waterbodies and opportunities for cost-effective mitigation. Tartu, 2020, 132 p.
378. **Siim-Kaarel Sepp**. Soil eukaryotic community responses to land use and host identity. Tartu, 2020, 222 p.
379. **Eveli Otsing**. Tree species effects on fungal richness and community structure. Tartu, 2020, 152 p.
380. **Mari Pent**. Bacterial communities associated with fungal fruitbodies. Tartu, 2020, 144 p.

381. **Einar Kärgenberg**. Movement patterns of lithophilous migratory fish in free-flowing and fragmented rivers. Tartu, 2020, 167 p.
382. **Antti Matvere**. The studies on aryl hydrocarbon receptor in murine granulosa cells and human embryonic stem cells. Tartu, 2021, 163 p.
383. **Jhonny Capichoni Massante**. Phylogenetic structure of plant communities along environmental gradients: a macroecological and evolutionary approach. Tartu, 2021, 144 p.
384. **Ajai Kumar Pathak**. Delineating genetic ancestries of people of the Indus Valley, Parsis, Indian Jews and Tharu tribe. Tartu, 2021, 197 p.
385. **Tanel Vahter**. Arbuscular mycorrhizal fungal biodiversity for sustainable agroecosystems. Tartu, 2021, 191 p.
386. **Burak Yelmen**. Characterization of ancient Eurasian influences within modern human genomes. Tartu, 2021, 134 p.
387. **Linda Ongaro**. A genomic portrait of American populations. Tartu, 2021, 182 p.
388. **Kairi Raime**. The identification of plant DNA in metagenomic samples. Tartu, 2021, 108 p.
389. **Heli Einberg**. Non-linear and non-stationary relationships in the pelagic ecosystem of the Gulf of Riga (Baltic Sea). Tartu, 2021, 119 p.
390. **Mickaël Mathieu Pihain**. The evolutionary effect of phylogenetic neighbourhoods of trees on their resistance to herbivores and climatic stress. Tartu, 2022, 145 p.
391. **Annika Joy Meitern**. Impact of potassium ion content of xylem sap and of light conditions on the hydraulic properties of trees. Tartu, 2022, 132 p.
392. **Elise Joonas**. Evaluation of metal contaminant hazard on microalgae with environmentally relevant testing strategies. Tartu, 2022, 118 p.
393. **Kreete Lüll**. Investigating the relationships between human microbiome, host factors and female health. Tartu, 2022, 141 p.
394. **Triin Kaasiku**. A wader perspective to Boreal Baltic coastal grasslands: from habitat availability to breeding site selection and nest survival. Tartu, 2022, 141 p.
395. **Meeli Alber**. Impact of elevated atmospheric humidity on the structure of the water transport pathway in deciduous trees. Tartu, 2022, 170 p.
396. **Ludovica Molinaro**. Ancestry deconvolution of Estonian, European and Worldwide genomic layers: a human population genomics excavation. Tartu, 2022, 138 p.
397. **Tina Saupe**. The genetic history of the Mediterranean before the common era: a focus on the Italian Peninsula. Tartu, 2022, 165 p.
398. **Mari-Ann Lind**. Internal constraints on energy processing and their consequences: an integrative study of behaviour, ornaments and digestive health in greenfinches. Tartu, 2022, 137 p.
399. **Markus Valge**. Testing the predictions of life history theory on anthropometric data. Tartu, 2022, 171 p.
400. **Ants Tull**. Domesticated and wild mammals as reservoirs for zoonotic helminth parasites in Estonia. Tartu, 2022, 152 p.

401. **Saleh Rahimlouye Barabi**. Investigation of diazotrophic bacteria association with plants. Tartu, 2022, 137 p.
402. **Farzad Aslani**. Towards revealing the biogeography of belowground diversity. Tartu, 2022, 124 p.
403. **Nele Taba**. Diet, blood metabolites, and health. Tartu, 2022, 163 p.
404. **Katri Pärna**. Improving the personalized prediction of complex traits and diseases: application to type 2 diabetes. Tartu, 2022, 190 p.
405. **Silva Lilleorg**. Bacterial ribosome heterogeneity on the example of bL31 paralogs in *Escherichia coli.* Tartu, 2022, 189 p.