USING PROTEINS TO STUDY THE STRUCTURE AND FUNCTION OF SOIL MICROBIAL COMMUNITIES

Robert Starke

Institute of Microbiology, Czech Academy of Sciences, Czech Republic

Soil provides ecosystem services in acting as a habitat for soil organisms, regulator of water quality, modifier of the atmospheric composition, medium for plant growth, and recycling system for nutrients and organic wastes [1]. A gram of soil can contain billions of organisms belonging to thousands of different bacterial species [2], which highlights the importance of understanding microbially-driven soil processes such as carbon cycling. The challenges in soil metaproteomics of efficient protein extraction, complete peptide measurement, and database construction for protein identification become even tougher since extracted proteins can belong to other taxonomic groups that are part of the soil food web. While we have a good understanding of the taxonomy and function from soil bacterial proteins, only little is known about the proteins of other taxonomic groups and their relative contribution to the metaproteome of the microbial community. Here, I will discuss (i) the soil metaproteomic workflow to investigate bacteria, how it can be modified (ii) for archaea, fungi, (iii) for a fair comparison between bacteria and fungi, and (iv) for more protein identifications.

References

1) E. Dominati, M. Patterson, A. Mackay, A framework for classifying and quantifying the natural capital and ecosystem services of soils, Ecol. Econ. (2010). https://doi.org/10.1016/j.ecolecon.2010.05.002.
2) V. Torsvik, L. Øvreås, Microbial diversity and function in soil: From genes to ecosystems, Curr. Opin. Microbiol. (2002). https://doi.org/10.1016/S1369-5274(02)00324-7.

CHALLENGES AND SOLUTIONS FOR EXPLORING TAXONOMICAL AND FUNCTIONAL CHANGES OF MICROBIOTA ALONG A SOIL CORE

Virginie Jouffret (1,2), Guylaine Miotello (1), Sophie Ayrault (2), Olivier Pible (1), Jean Armengaud (1)

1) *Laboratoire Innovations Technologiques pour la Détection et le Diagnostic (Li2D), Service de Pharmacologie etImmunoanalyse (SPI), CEA, INRA, F-30207 Bagnols-sur-Cèze, France*
2) *Laboratoire des Sciences du Climat et de l'Environnement (LSCE/IPSL), CEA-CNRS-UVSQ, Université Paris-Saclay,91191 Gif-sur-Yvette, France.*

Background and aims

Both metagenomics and metaproteomics are useful to get a comprehensive picture of the soil microbial communities, resulting in a taxonomical view and the functional repertoire for the former and the analysis of the biomass and the functional diversity for the latter. However, metaproteomics mostly relies on protein sequence databases for interpretation of MS/MS spectra acquired. Two main types of databases can be used: (i) sample-specific databases can be difficult to process because of a huge diversity of microorganisms in soil, (ii) public sequence databases where soil microbial organisms are currently largely under-represented. Today, the interpretation of soil metagenomics and metaproteomics data is a computational challenge and the identificationof proteins a key step of the workflow.

Methods

In order to identify the best strategies to analyze a sediment core from an anthropized floodplain along the SeineRiver downstream of Paris (France), we tested on a slice of the core several sample-specific metagenomics pipelines, public database construction strategies and cascade search methods to enhance the attribution yields of metaproteomics results (Jouffret et al (2021) Microbiome, in press). Once this exploration to maximize the results was achieved, we analyzed the entire sediment core that was sliced into 35 samples (every 3 centimeters) in triplicates, resulting in an impressive dataset corresponding to 105 nanoLC-MS/MS runs.

Results

The different strategies were evaluated by the number of peptide-spectrum matches as output. In one step strategies, the number was similar in cases of public or sample-specific metagenomics-derived databases. However, the combination of both types of databases in a two-step cascade search improves these numbers. Themicrobial community structure along the depth profile was characterized using 16S, metagenomics and metaproteomics datasets. Taxonomical and protein function results were compared with geochemical data, and the analysis of correlations between microorganisms and anthropogenic contaminants such as trace metals is under way.

Conclusions

We proposed a strategy to maximize the results of multi-omics for soil samples based on the combination of public and sample-specific metagenome databases. We analyzed a 1 m core of soil, corresponding to almost a century of sediment deposit. We correlated metaproteomics results with geochemical data acquired along 31 slices of this core.

EXPLORING PROTOZOAL FUNCTION AND THEIR GREATER METABOLIC INFLUENCE IN THE RUMEN MICROBIOME USING (META)-GENOME-RESOLVED METAPROTEOMICS

Thea Os Andersen (1), Ianina Altshuler (1), Arturo Vera Ponce de Leon (2), Laurence Bernard (3), Hélène Fougere (3), Diego P. Morgavi (3), Jeffrey F. Firkins (4), Zhongtang Yu (4), Torgeir Rhoden Hvidsten (2), Milka Popova (3), Magnus Øverlie Arntzen (2), Live Heldal Hagen (2), Phil B. Pope (1)

1) *Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Ås, Norway*
2) *Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway*
3) *Université Clermont Auvergne, INRAE, VetAgro Sup, UMR Herbivores, F-63122 Saint-Genes-Champanelle, France*
4) *Department of Animal Sciences, The Ohio State University, Columbus, OH, United States*

The rumen constitutes a specialized ecosystem composed of a dense and complex mixture of anaerobic bacteria, archaea, protozoa, fungi and phages, that interact closely in the degradation and fermentation of complex plant material into volatile fatty acids (VFAs), utilized for host energy metabolism, and methane gas. The metabolic functions carried out by the rumen microbiome are of scientific and industrial interest, as it contributes to feed efficiency and production of an important human food source (meat and dairy), but additionally contributes significantly to global methane emissions. Therefore, extensive efforts are needed to mitigate enteric methane emissions from ruminant animals without compromising livestock production. While the protozoal populations in the rumen microbiome can comprise up to 50% of the microbial biomass, their biological and metabolic features remain largely unsettled. Advances in culture-independent "meta-omics" approaches continue to increase our understanding of microbiomes, and by integrating multiple culture-independent meta-omics techniques, we can obtain a detailed real-time and in situ molecular portrait of which organisms occupy specific metabolic niches. By resolving (meta)genome-centric metaproteome datasets from rumen fluid samples originating from dairy cows and goats fed diets supplemented with different lipid sources, we explore protozoal function in the rumen microbiome. We specifically leveraged these data to investigate their active metabolic genes and pathways that are responsible for polysaccharide digestion, generation of hydrogen and production of VFAs. Our results illustrate the significant metabolic influence these under-explored eukaryotic populations have in the rumen, towards both fiber and hydrogen metabolism.

THE BACTERIAL MICROBIOME OF Lobaria pulmonaria L. HOFFM.: COMPOSITION AND VARIABILITY

Maria Grimm (1), Morten Streblow (1), Ulf Schiefelbein (2), Daniela Zühlke (1), Jörg Bernhardt (1), Mia Bengtsson (1)

1) *Center for Functional Genomics of Microbes, Institute of Microbiology, University of Greifswald, 17489 Greifswald,Germany*
2) *Botanical Garden, University of Rostock, Rostock, Germany*

Lichens represent self-supporting symbioses (composed of a myco- and photobiont), which occur in a wide range of habitats and which contribute significantly to mineral cycling and energy flow at a global scale. In comparison to other lichens, Lobaria pulmonaria L. Hoffm. is less stress-resistant. It only tolerates a quite narrow range of variation in ecological conditions. L. pulmonaria prefers undisturbed and relatively cool and humid habitats with minimal air pollution. Our ongoing project aims on deciphering phenotypic traits enabling this model lichen to adapt to changing environmental conditions. Bacterial communities are abundant, stable, specific, and structurally integrated elements of the classical lichen symbiosis. The diverse microbiota seems to contribute multiple aspects to the symbiotic system, including essential functions such as nutrient supply, resistance against abiotic factors or support of fungal and algal growth. The lichen microbiome is suggested to consist of a rather constant core microbiome complemented by a variable microbiome. Based on former studies, we hypothesize that the variable part of the microbiome is responsible to meet habitat-dependent challenges and may be acquired from the local surrounding environment. To address this hypothesis, we performed metaproteome analyses of samples, which have been collected at 14 different European sites. Comparisons of the microbiomes of all sampling sites confirmed that the microbiome of L. pulmonaria consists of a core and a variable part. By comparing the composition of the lichen-associated with the moss-associated bacterial microbiome, differences and similarities in taxonomical composition and functionality were illustrated. We suggest that mosses can act as a bacterial reservoir supporting the lichen during establishment. Our results show that the direct environment is able to shapes the variable part of the microbiome of L. pulmonaria and therefore itmay enable the holobiont to cope with environmental changes and, thus, can be considered as an important ecological trait mediating lichen stress tolerance.

PRELIMINARY FUNCTIONAL INSIGHTS INTO THE GUT MICROBIOTA OF COVID-19 PATIENTS PROBED BY METAPROTEOMICS

Lucia Grenga (1), Olivier Pible (1), Guylaine Miotello (1), Karen Culotta (1), Sylvie Ruat (1), Marie-Anne Roncato (1), Fabienne Gas (1), Laurent Bellanger (1), Pierre-Géraud Claret (2), Catherine Dunyach-Remy (3), Didier Laureillard (4), Albert Sotto (5), Jean-Philippe Lavigne (3), Jean Armengaud (1)

1) *Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), SPI, 30200Bagnols-sur-Cèze, France*
2) *Service d'Accueil des urgences, CHU de Nîmes, 30029 Nîmes, France*
3) *Virulence Bactérienne et Infections Chroniques, INSERM U1047, Université Montpellier, Service de Microbiologie etHygiène Hospitalière, CHU Nîmes, 30908 Nîmes, France.*
4) *Service des Maladies Infectieuses et Tropicales, CHU Nîmes, 30029 Nîmes, France.*
5) *Virulence Bactérienne et Infections Chroniques, INSERM U1047, Université de Montpellier, Service des MaladiesInfectieuses et Tropicales, CHU Nîmes, 30908 Nîmes, France.*

Since the beginning of the ongoing pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the gastrointestinal tract has emerged as an important organ influencing propensity to develop of the corresponding disease, COVID-19, and potentially its severity. In this context, the interplay between the gut microbiota and SARS-CoV2 and the contribution of intestinal SARS-CoV-2 infection to the pathogenesis of COVID-19 remains to be clarified. We employed a mass spectrometry-based approach to profile the gut microbiota in patients with intestinal COVID-19 infection. We showed that alterations to the composition of the gut microbiota are influenced by the levels of SARS-CoV-2 RNA in the gastrointestinal tract, but not by the presence of SARS-CoV-2 in the respiratory tract, COVID-19 severity, or GI symptoms. This difference in the microbiota profiles for faeces characterized by high SARS-CoV-2 RNA levels was further confirmed by a peptide-based functional metaproteomics analysis. These analyses revealed that mechanisms that potentially confer a competitive advantage in stressful environments were significantly more abundant in SARS-CoV-2-positive faeces. Similarly, host molecular functions linked to the enterocyte brush border, and supporting inflammation-induced enterocyte damage or increased intestinal permeability correlated positively with infection, and were associated with an increased abundance of several leaky-gut-related functions. These findings and the potential of profiling the microbiota via metaproteomics in the search for alternative therapies promoting viral clearance and/ or guide future clinical diagnosis will be discussed.

COMPREHENSIVE METAPROTEOMIC ANALYSIS OF NASOPHARYNGEAL SWABS TO INVESTIGATE CO-INFECTION STATUS IN COVID-19 PATIENTS

Surbhi Bihani (1), Subina Mehta (2), James Johnson (2), Saicharan Ghantasala (1), Arghya Banerjee (1), Kruthi Suvarna (1), Akanksha Salkar (1), Viswanthram Palanivel (1), Renuka Bankar (1), Aryan Gupta (1), Dr. Jayanthi Shashtri (3), Reid Wagner (2), Andrew Rajczewski (2), Timothy J. Griffin (2), Sanjeeva Srivastava (1), Pratik D. Jagtap (2)

1) *Indian Institute of Technology, Bombay, India*
2) *University of Minnesota, Minneapolis, MN*
3) *Kasturba Hospital for Infectious Diseases, Chinchpokli, Mumbai, India*

Respiratory infections disrupt the microbiota in the upper respiratory tract (URT), putting patients at risk for subsequent infections. The current COVID-19 pandemic, which is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has resulted in a slew of microbial co-infections because of impaired immunity caused by viral infections and medical interventions during hospitalization. These secondary infections engendered by opportunistic pathogens have greatly hampered therapy, prognosis, and overall disease management. For instance, several cases of a potentially fatal fungal infection - mucormycosis - were reported especially during the second wave of COVID-19 in India. The potential dangers and clinical difficulties of bacterial and fungal co-infections with SARS-CoV-2 necessitate the exploration of the URT microbiome from clinical samples.

In this study, mass spectrometry-based deep proteome data of nasopharyngeal swabs from COVID-19 patients was used for metaproteomic analysis. Acquired on an Orbitrap Fusion mass spectrometer in DDA mode, the MSdata from 68 individuals with varying degrees of COVID-19 severity were used to detect microbial peptides corresponding to potential pathogens using label-free quantification. The MS datasets were processed and analyzed in two batches, each with 44 (26 positives, 11 recovered, and 7 negatives) and 24 (11 severe and 13 non-severe) samples. Blank viral transport medium (VTM) - used to collect patient swabs - was used as a controlto detect organisms as a result of environmental contamination. The MS files were searched against a compact protein sequence database generated from a large database (approximately 33 million sequences belonging to common URT pathogens comprising 72 bacterial and fungal genera) using MetaNovo software. The reduced database was used to detect microbial peptides using Galaxy workflows that included software tools such as SearchGUI/Peptide Shaker (Identification), MaxQuant (Identification and Quantitation), PepQuery (peptide-spectrum match quality validation), and UniPept (taxonomy).

Preliminary analysis within the Galaxy platform detected around 400 peptides belonging to over 200 bacterial and fungal species. Peptides from bacteria such as Burkholderia, Bifidobacterium, Lactobacillus, Streptococcus, Acinetobacter, and Prevotella, and fungi like Fusarium, Rhizopus, Mucor, Candida, Aspergillus, Cryptococcus, and Syncephalastrum racemosum were detected exclusively in the clinical samples (both COVID-19 positive and negative samples) and not in blank VTM controls. Interestingly, we detected peptides from some opportunistic pathogens like Pseudomonas putida, Klebsiella, Clostridium botulinum, Acenitobacter baumannii, and Helicobacter pylori in a few individual samples. We also detected multiple peptides from fungi of the order Mucorales (Mucor, Rhizopus, and Syncephalastrum) that are known to cause mucormycosis, a rare but potentially fatal disease reported in severe and recovering COVID-19 patients during the recent pandemic wave in India. Moreover, several peptides from causative agents of candidiasis, aspergillosis, and fusariosis (opportunistic fungal infections that have been reported in COVID-19 patients) were also detected. We plan to validate the detection of these microbial peptides by using spectral visualization and parallel reaction monitoring. Based on this analysis, we aim to build a peptide panel that will help in the detection of pathogens from clinical samples so that it can aid in diagnosis and management of secondary infection from these and anyemerging pathogens.

FUNCTIONAL METAPROTEOMICS FROM BENCH TO CLINICS

Daniel Figeys

Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Canada

We are interested in better understanding the functional roles of the gut microbiome in diseases. Metaproteomics is well suited to reveal the functional changes that occur in microbiomes during diseases and therapeutics/nutritional interventions. Here we will present the application of metaproteomics to better understand the roles of the microbiome in pediatric inflammatory bowel disease(IBD). As well, we will discuss the effects that nutritional, and drugs have on individual gut microbiomes using the RapidAIM assay. Finally, we will discuss the application of metaproteomics in umbrella clinical trials to help select individual therapeutic interventions.

METAPROTEOMIC FEATURES RELATED TO INTESTINAL BOWEL DISEASES

Ariane Bassignani (1,2,3), Aaron Milan-Oropeza (1), Armelle Sejean (1), Olivier Langella (3), Harry Sokol (4), Céline Henry (1), Catherine Juste (1)

1) *Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, 78350, Jouy-en-Josas, France*
2) *Université Paris-Saclay, INRAE, MGP, 78350, Jouy-en-Josas,*
3) *Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE – Le Moulon, 91190, Gif-sur-Yvette, France*
4) *Assistance Publique Hôpitaux de Paris, Service de Gastroentérologie, Saint-Antoine Hospital, Paris, France*

Background and aims.
Diagnosis of the different phenotypes of intestinal bowel diseases (IBD) – CDC for Colic Crohn's disease, CDIC for Ileo-Colic Crohn's disease, and UC for Ulcerative Colitis) – is challenging for the clinician, and critical for the patient. As there are no phenotype-specific biomarkers, it is based on a body of clinical arguments (digestive and general disorders, other manifestations) and paraclinical exams (biology, endoscopy, etc.). It can sometimes take many years before clinicians and patients can put a name to the problem. The early correct diagnosis is however crucial since it allows for adapted treatment strategies as soon as possible, and lower risk of complications. In a discovery, without a priori metaproteomic approach of gut microbiota, we investigated the feasibility of identifying still unsuspected proteins that could assist in the early diagnosis of IBD phenotypes.

Methods.
Stool samples (n=40; 8 controls, 7 UC, 3 CDC and 2 CDIC) were collected and prepared for LC-MS/MS analysis on an Orbitrap Fusion™ Lumos™ Tribrid™. Focus was on the envelope-enriched fraction of the gut microbiota as the first line of interaction with the host mucosae. Considering the circumstances where the patients are far from the diagnosis centre, pairs of fresh and frozen samples were considered. For mass spectral interpretation we used X!Tandem in an iterative search strategy (ref) of the combined databases IGC 9.9 and Homo sapiens, and we used the grouping algorithm of X!TandemPipeline (https://doi.org/10.1021/acs.jproteome.6b00632) to provide a parsimonious list of proteins that explained all peptides identified in the samples. Proteins identified with a same set of peptides were grouped into 'metaproteins', which were each quantified by the sum of the spectral counts of their specific peptides. We used the R package 'metaprotr' (https://cran.r-project.org) to shape and gather the quantitative mass spectrometry data, and then display them graphically (metaproteins, taxonomic and functional entities).

Results and conclusions.
Based on this global approach and annotation of all metaproteins, we could provide a taxonomic and functional landscape of the different samples. More importantly, by implementing a highly stringent selection of metaproteins that were specifically either over- or underrepresented in the different IBD phenotypes, we could propose leads for earlier diagnosis of the different inflammatory bowel flares, illustrated by a gut metaproteomics-based decision tree.

CHARACTERIZATION OF TIME-DEPENDENT SIGNATURES OF ANTIBIOTIC ADMINISTRATION AND RESISTANCE-RELATED MICROBIAL MODULATION OF HOST IMMUNE RESPONSE IN PRETERM INFANTS

Samantha L. Peters (1,2), Robert L. Hettich (1,2)

1) Oak Ridge National Laboratory, Oak Ridge, TN, USA
2) University of Tennessee, Knoxville, TN, USA

The gut microbiome plays an essential role in host health by producing metabolic resources and the stimulation and training of the immune system. Due to the time-driven expansion of microbial complexity at early life, metaproteomics can track details about colonization and community functional dynamics based on environmental changes and capture host response information. This study was designed to examine the differences in the temporal relationships between microbial activities and the host proteome during the normal and abnormal establishment of the preterm infant gut microbiome. Using necrotizing enterocolitis (NEC) as a representative dysbiotic condition, 91 metaproteomic measurements were taken over the first 90 days of life from seventeen preterm infant fecal samples with various antibiotic treatment regimes and disease outcomes. Proteome Discoverer was used to compare spectra generated from a previous shotgun proteomics experiment (CT Brown et al., mBio, 2018) against individual-specific protein databases created from sample-matched metagenomes. The proteome databases were clustered based on sequence homology to reduce ambiguity related to shared proteins, and peptides were reassigned to protein groups based on the clustering. For protein quantification, peptide AUC intensities were used to assemble proteins by summing only peptide intensities uniquely mapping to one protein group.

On average, 3000 human and 9000 microbial protein groups were identified per infant, respectively. Among the quantified human proteins, proteins related to immune functions represent around 10-15% of human protein identifications, and 40-50% of quantified proteins are unique to specific immune pathways across all samples. Preliminary results show temporal patterns of host immune responses corresponding with gut microbiota function. Non-metric multidimensional scaling of Jaccard distances and PERMANOVA testing were performed to assess the functional β-diversity of samples. The ordinations revealed functional partitioning among infants based on antibiotic administration. Maternal intrapartum antibiotic administration significantly separates infants based on human proteins more predominantly than microbial proteins. Specifically, maternal antibiotic administration before delivery significantly impacts the infant's levels of expressed proteins related to immune processes, such as neutrophil degranulation and antimicrobial peptides early in life. In addition to monitoring host responses at critical windows of immune system development, metaproteomics allows the characterization of the microbial establishment. There are distinct longitudinal trends between healthy and diseased infants regarding antibiotic resistance proteins expressed by the microbiota. Tracking the longitudinal trajectories of protein expression levels related to microbial defense processes early in life may help predict host health outcomes and better inform treatment options. Overall, metaproteomics provides a unique approach to studying the developing gut environment. It allows simultaneous examination of host and microbial metabolic activities to elucidate potential immunomodulatory roles of the microbiota in gut-related dysbiotic conditions.

METAPROTEOMIC ANALYSIS OF COLONIC LUMINAL CONTENT MICROBIOTA FROM COLON CANCER PATIENTS

Alessandro Tanca (1), Marcello Abbondio (1), Giovanni Fiorito (1,2), Giovanna Pira (1), Rosangela Sau (1), Alessandra Manca (3), Maria Rosaria Muroni (4), Alberto Porcu (4), Antonio Mario Scanu (4), Paolo Cossu-Rocca (4,5), Maria Rosaria De Miglio (4), Sergio Uzzau (1)

1) *Department of Biomedical Sciences, University of Sassari, Sassari, Italy*
2) *MRC Centre for Environment and Health, Imperial College London, London, UK*
3) *Department of Pathology, AOU Sassari, Sassari, Italy*
4) *Department of Medical, Surgical and Experimental Sciences, University of Sassari, Sassari, Italy*
5) *Department of Diagnostic Services, Surgical Pathology Unit, "Giovanni Paolo II" Hospital, ASSL Olbia-ATS Sardegna,Olbia, Italy*

Recent studies have provided evidence of interactions among the gut microbiota, local host immune cells and intestinal tissues in colon carcinogenesis. However, little is known regarding the functions exerted by the intestinal microbiota in colon cancer (CC), particularly with respect to tumor clinical classification and lymphocyte infiltration. In addition, stool, usually employed as proxy of the gut microbiota, cannot fully represent the original complexity of CC microenvironment. This pilot study was aimed at characterizing the metaproteome of tumor-associated colonic luminal contents and evaluating its potential to identify associations between gut microbial protein functions and CC clinicopathological features, namely tumor stage, tumor grade and Tumor Infiltrating Lymphocytes (TILs).

Colonic luminal content samples were collected in operatory room immediately after surgery (colectomy) from 24 CC tissue specimens. After protein extraction and digestion, peptide mixtures were analyzed by high-resolution mass spectrometry. Bioinformatic analyses allowed peptide identification, label-free quantification and taxonomic/functional annotation. For each clinical variable, the sample set was divided intotwo groups (stage I-II vs III-IV; grade 1-2 vs 3; TILs positive/high vs negative/low) and the set of the most discriminating peptides was identified through a permutation-based sPLS regression approach. Significantly enriched taxa, functions and metabolic pathways were identified for each set of discriminating peptides, performing a sensitivity analysis considering covariate impact (age, sex, tumor site and the remaining clinical variables) and adjusting for multiple testing through a permutation-based approach.

A total of 57,102 peptide groups were quantified after mass spectrometry analysis, of which 34,862 were taxonomically classified as microbial and 9,338 as human. Peptide intensity data were used as quantitative measure and associated with taxonomic and functional annotations. Microbial peptides were assigned to 90 different microbial families (with Bacteroidaceae, Ruminococcaceae and Clostridiaceae being the most abundant) and over 1200 different functions, including carbohydrate transport and metabolism, response to stress, cell motility and translation. We identified 294, 94 and 568 microbial peptides discriminating for tumor stage, grade and TILs, respectively. Proteins produced by Bifidobacterium were found significantly enriched in high-stage tumors, whereas those expressed by Bacteroides spp. were over-represented in high-grade and TIL-negative tumor samples. Furthermore, microbial enzymes involved in tetrahydrofolate interconversion, glutamine biosynthesis and galactose catabolism were enriched in the colonic luminal metaproteome of high-stage/grade tumors. Discriminating peptide clusters correlated with host peptides enriched in specific functions, including prolactin-inducible protein, catalase and intelectin. Finally, power analysis indicated a minimum of 30 patients per group as the threshold to reach 80% probability to identify features with log foldchange 2.

In conclusion, this pilot study provided a detailed picture of the colonic luminal metaproteome. Moreover, promising correlations between the abundance of human and bacterial proteins and CC clinicopathological features were found. Future studies with higher numbers of samples, as suggested by power analysis results, are needed to extend the investigation and confirm the biological value of the findings, as well as to validate their potential to enhance our knowledge concerning CC progression.

QUANTITATIVE METAPROTEOMICS ANALYSIS OF A MOUSE MODEL FOR PARKINSON'S DISEASE

Jinqiu Xiao (1), Aletta Kraneveld (2), Christoph Turck (1), Paula Perez-Pardo (2)

1) *Max Planck Institute of Psychiatry, Proteomics and Biomarkers, 80804, Munich, Germany.*
2) *Division of Pharmacology, Faculty of Science, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht,The Netherlands.*

Parkinson's disease (PD) is a long-term neurodegenerative disease. Approximately 1% of the over 60-year-old population suffers from PD. A complete understanding of the pathobiology is still elusive. Many PD patients also develop gastrointestinal (GI) dysfunctions that precede the motor symptoms. This suggests that the microbiota-gut-brain-axis is involved in PD pathology. To address this, we have subjected fecal pellets collected from a cohort of control mice and rotenone-induced PD mice to quantitative metaproteomics analysis. We observed taxonomic differences and enriched taxa in rotenone-induced PD compared to control mice. The majority of the significantly changed microbial fecal pellet proteins correlated with rotenone-induced motor deficit symptoms. Our study is the first investigation of the PD mouse model's metaproteome and delineates taxaand molecular pathways that may be involved in PD pathobiology.

A HYBRID NANOPORE-METAPROTEOMICS APPROACH FOR THE CHARACTERIZATION OF RAW COW MILK RESISTOME

Alessio Soggiu (1), Gabriele Meroni (1), Valerio Sora (1), Vincenzo Cunsolo (2), Simona Arena (3), Tiziana Cabras (4), Piera Anna Martino (1), Paola Roncada (5), Raffaella Memmola (1), Alfonso Zecconi (1), Luigi Bonizzi (1)

1) *Department of Biomedical, Surgical and Dental Sciences, One-Health Section, University of Milan,Italy*
2) *Laboratory of Organic Mass Spectrometry (LOMS) Department of Chemical Sciences - University of Catania, Catania, Italy*
3) *ISPAAM-CNR, Portici (NA), Italy*
4) *Dipartimento di Scienze della Vita e dell'Ambiente Sez. Biomedica, Università di Cagliari, Monserrato, Italy*
5) *Department of Health Sciences University' Magna Græcia' of Catanzaro, Catanzaro, Italy*

Dairy cow mastitis (CM) is the cause of a great financial loss for all the animal production industry. The average cost of a single case of CM is around € 200. Its diagnosis increases the costs as well as the requirement of a more challenging treatment with antibiotics. Antibiotic resistance (AR) is a natural process that causes the selection and dissemination of antibiotic resistance genes (ARGs) and is mainly reinforced by the overuse or inappropriate use of antibiotics. In this context, the study of resistome, pangenome and metaproteome can be achieved using modern omics tools like whole genome sequencing, metagenomics and proteomics. The main aim of this work is to analyze through a multiomic approach both the bacterial and cellular compartment of raw milk to gain informations about the whole bacterial composition and the effects of antibiotic residues on the milk microbiota. Through the metagenomics approach (using third-generation sequencing technology), the microbiota of different milk samples has been analyzed to recover the taxonomical profile of each sample. An NGS-specific bioinformatics pipeline has been applied (EPI2ME) to obtain 16S taxonomic information between samples. The metaproteomic approach has been focused on the detection of key ARGs expressed in milk microbiota. Mass Spectrometry-based targeted analysis using specific databases (e.g., CARD) and pipelines allowed the monitoring of the relative abundance of ARG-related proteins among each sample analyzed and the bacterial species associated with the ARGs dissemination. To the best of our knowledge, this method allowed the characterization of AR in raw milk microbiota and provided a useful tool of practical importance for the detection of the indirect presence of antibiotic contamination in the environment.

Acknowledgements: work supported by PRIN 2017 "SAFE MILK: OMICS SCIENCE FOR MILK SAFETY AND QUALITY" (grant No. 2017MZ5KWM)

ACTIVITY OF MICROBIAL COMMUNITIES LIVING IN CARBONATE ROCKS AT SEAFLOOR METHANE SEEPS

Magdalena J. Mayr (1), Payal Chirania (2), Paul E. Abraham (2), Robert L. Hettich (2), Sergio Parra (1), Rafael L. Pérez (1), Antoine Crémière (1), Sean Mullin (1), Kriti Sharma (1), Victoria Orphan (1)

1) *Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CS, USA*
2) *Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA*

Cold seeps are seafloor ecosystems at the continental margins where methane from the subsurface supports microbial activity, resulting in the microbially mediated formation of carbonate rocks. These carbonate rocks were previously thought to be fossilized remnants of past activity but have recently been discovered to host viable methane-oxidizing communities, acting as globally significant methane sink. Carbonates pave about half of the seafloor around methane seeps and support unique animal communities of e.g., snails, tubeworms, mussels, and corals that show isotopic signatures suggestive of feeding on methane-derived carbon. Studies on the microbial communities thriving in and on carbonates are scarce, despite their important role in the ocean carbon cycle and for maintaining biodiversity associated with seep carbonates. To investigate the metabolic potential (energy and carbon acquisition) of microbes living on and in carbonate rocks we will apply metagenomics and metaproteomics. Future analyses will address which metabolic pathways are most expressedand how protein-based community composition compares to metagenomic results. Another question we plan toinvestigate is if it is possible to apply direct protein stable isotope fingerprinting – SIF metaproteomics to selected abundant carbonate microbes – analogous to animal analyses giving insights into trophic structure.

Optimization and successful protein extraction from the rocks and a first chromatogram have been performed and further analyses are underway. First microbial community analyses indicate that microbial taxa are shared between rocks from different seeps, with the dominating anaerobic methane oxidizing archaea and their symbiotic partners, the sulfate reducing bacteria. The roles of some other common groups, like Asgardachaeota and Caldatribacteriota are less clear. The surface biofilm of rocks, which is potentially subject to grazing by animals like snails showed a different community composition than the volumetrically much larger rock interior. We further hypothesize that rock methane oxidation activity influences the metaproteome of the microbial community. Future research directions include rock incubations with HPG (homopropargylglycine) a clickable methionine analog for detecting active protein synthesis, detectable with metaproteomics and incubations with isotopically labeled substrates, like ^{13}C-CH_4, ^{13}C-CO_2, CH_3D or D_2O, followed by SIP-metaproteomics. The aim is to link microbial taxonomic information with protein turnover and substrate uptake to illuminate trophic interactions and the timescales at which these slow growing rock communities operate.

INTEGRATIVE META-OMICS IN GALAXY

Valerie C. Schiml (1), Francesco Delogu (1), Praveen Kumar (2), Benoit Kunath (1), Subina Mehta (2), James E. Johnson (3), Bjoern Gruning (4), Phil B. Pope (1), Pratik Jagtap (2), Timothy J. Griffin (2), Magnus Ø. Arntzen (1)

1) *Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway;*
2) *Department of Biochemistry, Biophysics and Molecular Biology, University of Minnesota, Minneapolis, MN 55455*
3) *Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN 55455*
4) *Bioinformatics Group, Department of Computer Science, University of Freiburg, Freiburg, Germany*

The meta-omic technologies have opened new possibilities in analysing the complexity of a microbiome and its interaction with the environment on a scale not attainable before, offering a realistic picture of the natural habitat. Individually, metagenomics, metatranscriptomics and metaproteomics can provide great insight; while incombination, they offer a detailed understanding of which organisms occupy specific metabolic niches, how they interact, and how they utilize environmental nutrients. Metagenomics can provide access to the diversity and metabolic potential of microbial communities, while metatranscriptomics and metaproteomics can provide information on the actual proteins expressed, their abundancies and from which organisms they originated.
However, the complexity of informatics approaches required for multi-omics, limits their adoption by the widerresearch community. Here we have implemented and adapted a repertoire of commonly used omics tools spanning all three technologies into the Galaxy framework, in order to generate a user-accessible, scalable and robust analytical pipeline for integrated meta-omics analysis.

To showcase the pipeline, we applied it to a highly efficient cellulose-degrading minimal consortium isolatedand enriched from a biogas reactor in Fredrikstad, Norway, in order to analyse the key roles of uncultured microorganisms in complex biomass degradation processes. Metagenomic analysis recovered
metagenome-assembled genomes (MAGs) for several constituent populations including Hungateiclostridium thermocellum, Acetomicrobium mobile and multiple heterogenic strains affiliated to Coprothermobacter proteolyticus. Metatranscriptomic and metaproteomic analysis revealed co-expression of carbohydrate-active enzymes (CAZymes) from multiple populations, indicating deeper microbial interactions directed towards the
co-degradation of cellulose and hemicellulose. Due to the combination of meta-omics methods, the identificationand description of key roles played by specific uncultured microorganisms in complex degradation processes was possible.

To further enhance the data interpretation and exploration, we are currently developing an interactive R Shinyapplication tailored for use with the Galaxy outputs generated by this pipeline.

PEPGM: A PROBABILISTIC GRAPHICAL MODEL FOR TAXONOMIC PROFILING OF VIRAL PROTEOMES AND METAPROTEOMIC DATASETS

Tanja Holstein (1,2), Lennart Martens (2), Thilo Muth (1)

1) *Federal Institute for Materials Research and Testing, Berlin/Germany*
2) *VIB-UGent Center for Medical Biotechnology, VIB and Department of Biomolecular Medicine, Ghent University, Ghent,Belgium*

Taxonomic inference in mass spectrometry-based metaproteomics is a complex task. The presence of proteins and corresponding taxa must be inferred from a list of identified peptides which is often complicated by protein homology: many proteins do not only share peptides within a taxon but also between taxa. Correct taxonomic identification is crucial when identifying different viral strains with high sequence homology – considering, e.g., the different epidemiological characteristics of the various strains of SARS-CoV-2. Our work therefore aims for accurate viral strain identification.

For protein inference, the peptide-protein relationships can be represented as bipartite graphs. Probabilistic graphical models have been used successfully to propagate peptide scores to the protein level. However, similar methods are not yet available at the peptide-taxon level where uncertainty about the species present adds an additional level of complexity. Instead, current approaches rely on strategies such as peptide-spectrum-match counting or the use of unique peptides.

In our PepGM approach, we represent the peptide-taxon relationships as a bipartite graph where two types of nodes represent peptides and taxa, respectively. The resulting structure serves as scaffold for a factor graph, allowing for the computation of the marginal distributions of peptides and taxa. Propagation of peptide scores to taxa takes place through a message passing algorithm and results in taxonomic identifications with a corresponding statistically sound score. This graphical model will next be evaluated with viral and metaproteomic mass spectrometric data sets.

METAPROTEOMIC INVESTIGATION OF MARINE SEDIMENT TO EVALUATE THE METABOLIC POTENTIAL FOR HYDROCARBON DEGRADATION

Anne Ostrzinski (1), André R. Soares (2), Cedric C. Laczny (3), Jens Kallmeyer (4), Rolando di Primio(5), Alexander J. Probst(2), Anke Trautwein-Schult (1), Dörte Becher (1)

1) *Institute for Microbiology, University Greifswald*
2) *Environmental Microbiology and Biotechnology (EMB)/Aquatic Microbial Ecology (GAME), University ofDuisburg-Essen*
3) *Luxembourg Centre for Systems Biomedicine, University of Luxembourg*
4) *Geomicrobiology, Research Group Aquatic Geochemistry, Helmholtz Centre Potsdam*
5) *Lundin Energy Norway*

Hydrocarbons like oil and gas will still remain important resources for energy generation for several decades. At this stage geological and geophysical methods for hydrocarbon exploration in marine systems create massive environmental disturbances without the certainty to find hydrocarbon reservoirs. Therefore, the goal of the
EU-sponsored project PROSPECTOMICS is to develop a fundamentally new approach for oil and gas exploration based on molecular biological methods in order to minimize the environmental impact. Seepage from submarine oil and gas reservoirs does not always reach the sea floor but can induce a shift in microbial community composition and its metabolic activity profile in the subsurface sediment. The microbial footprint and biochemical properties of sediments above hydrocarbon deposits will be compared to sediments in areas without the occurrence of oil or gas, using several biogeochemical approaches as well as "omics" technologies like metagenomics, metatranscriptomics and metaproteomics.

The cell number and biological activity in marine sediments are expected to be low, whereas the diversity may be rather high. Currently, the different groups involved in the project are focusing in the extraction of biomolecules like DNA, RNA and protein from the marine sediment. Obstacles in protein extraction from these environmental samples are caused by the composition of the sediment. Sediment particles can bind proteins and trap them. Additionally, humic substances are coextracted with proteins and interfere with colorimetric assays for determination of protein concentration as well as disturb mass spectrometric measurements.
A standardized method for protein extraction from marine sediment is not available yet. Therefore, different published extraction procedures were tested and the number of recorded MS spectra as well as the number of identified proteins were used to compare the methods regarding their efficiency to extract proteins from this challenging environment. As starting material two types of sediments were used: shelf sediment from the Barents Sea and coastal sediment from the Baltic Sea. For protein extraction, several combinations of buffers and extraction procedures were tested. Finally, the extracted proteins were processed with an SDS-PAGE and digested with Trypsin. Resulting peptides were eluted from the gel and desalted by C18 ZipTips prior to MS measurement. Since no sample-specific database was available, a protein database containing marine sediment organisms was constructed.

In a first attempt, about 50 proteins could be identified ($\geq$2 peptides) by using a combination of heating and sonication of the samples prior to protein enrichment using StrataClean beads. Using the different extraction approaches, up to 150 proteins could be identified in both sediments with a minimum of one peptide. The low number of identified proteins indicates their low concentration in the extracts.

To conclude, the tested extraction methods can be used for protein extraction from marine sediments, but have tobe optimized and improved in order to increase the number of identifiable proteins.

CRITICAL ASSESSMENT OF METAPROTEOME INVESTIGATION (CAMPI): A MULTI-LAB COMPARISON OF ESTABLISHED WORKFLOWS

Tim Van Den Bossche (1,2), Benoit J. Kunath (3), Kay Schallert (4), Stephanie Schäpe (5), Lennart Martens(1,2), Thilo Muth (6), Metaproteomics Initiative (7)

1) *VIB - UGent Center For Medical Biotechnology, VIB, Ghent, Belgium*
2) *Department of Biomolecular Medicine, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium*
3) *Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg, Luxembourg*
4) *Bioprocess Engineering, Otto-von-Guericke University Magdeburg, Magdeburg, Germany*
5) *Helmholtz-Centre for Environmental Research - UFZ GmbH, Department of Molecular Systems Biology, Leipzig,Germany*
6) *Section eScience (S.3), Federal Institute for Materials Research and Testing, Berlin, Germany*
7) *www.metaproteomics.org*

Introduction

Metaproteomics, the study of the collective proteome within a microbial ecosystem, has matured into a powerful tool to assess functional interactions in microbial communities. This maturation has been driven by improved technologies and informatics approaches[1] and by the realization that metaproteomics can provide functional insights into microbial communities that go well beyond what can be studied with other methods such as metagenomics[2]. Although a variety of metaproteomic workflows has been developed, their impact on the resultsremains to be established.

Methods

To evaluate and compare existing metaproteomic workflows, we carried out the first community-driven, multi-lab comparison in metaproteomics: the Critical Assessment of MetaProteome Investigation (CAMPI) study[3]. Based on well-established workflows, we evaluated the effect of sample preparation, mass spectrometry, and bioinformatic analysis using two samples: a simplified, lab-assembled human intestinal model and a humanfecal sample.

Results

We found that meta-omics databases performed better than public reference databases across both samples. More importantly, even though larger differences were observed in identified spectra and unique peptide sequences, the different protein grouping strategies and the functional annotations provided similar results across the provided data sets from all laboratories. When minor differences could be observed, these were largely due to differences in wet-lab methods and partially to bioinformatic pipelines. Finally, for the taxonomic comparison, we found that overall profiles were similar between read-based methods and proteomics methods, with few exceptions.

Conclusion

To conclude, CAMPI demonstrates the robustness of present-day metaproteomics research, serves as a template for multi-lab studies in metaproteomics, and provides publicly available data sets for benchmarking future developments.

References

1) *Schiebenhoefer, H., et al. Expert Rev. Proteomics 2019, 16 (5), 375–390.*
2) *Lohmann, P.; et al. Expert Rev. Proteomics 2020, 17 (2), 163–173.*
3) *Van Den Bossche, T., et al. bioRxiv, 2021. https://doi.org/10.1101/2021.03.05.433915.*

METAPROTEOMICS PROFILING OF THE LUNG MICROBIOTA FROM CYSTIC FIBROSIS PATIENTS

Pauline Hardouin (1), Olivier Pible (1), Hélène Marchandin (2), Raphaël Chiron (3), Jean Armengaud (1), Lucia Grenga (1)

1) *Laboratory 'Innovative technologies for Detection and Diagnostics' CEA- Marcoule, Bagnols-sur-Cèze, France*
2) *Equipe PHySE, HydroSciences Montpellier, CNRS, IRD, Université de Montpellier,Montpellier, France*
3) *Centre de Ressources et de Compétences de la Mucoviscidose (CRCM), Centre Hospitalier Universitaire de Montpellier, Montpellier, France*

Background and aims
Cystic fibrosis (CF) is a hereditary disease caused by mutations in the CF transmembrane conductance regulator (CFTR) gene leading to the dysfunction of the anion channel in different organs such as the lungs and the gut. The main cause of morbidity and mortality is a chronic infection. The microbiota is now included among the additional factors that could contribute to the exacerbation of patient symptoms as well as to treatment outcome. Aiming at shedding new light on microbial communities associated with CF and host-microbiota interactions here we report the development of an innovative metaproteomic-based approach and its application to investigate the effect of the sputum sampling method on the composition of the CF microbiota.

Methods:
To profile the microbiota from sputum collected from CF patients, a metaproteomics approach was applied. Briefly, following a pre-treatment step aiming at disrupting the mucus and isolating the fraction containing the microbiota, proteins were extracted and submitted to in-gel proteolysis. The resulting peptides were analysed by shotgun tandem mass spectrometry. The taxonomical profile of the microbiota was obtained by interpreting the raw data via a three steps cascade search against a database whose complexity was reduced at each step according to the composition of the sample. Peptide-to-Spectrum Matches were filtered using an FDR < 1% and used to infer peptide and protein identifications. Proteins were then KEGG-annotated using the GhostKoala web service to retrieve functional information from the metaproteome.

Results:
The different components of the microbiota (bacteria, archaea and fungi), including the less explored or the uncultivable ones, were identified and quantified in three pairs of spontaneous and induced sputum samples. The comparative analysis of the metaproteomics data revealed the absence of significant differences both in terms of the microbial signal detected and taxonomical diversity between the two types of samples. Similarly, the sampling method doesn't impact the functional profile of both microbiota and host with functional terms related to the CF disease such as inflammation, lipid metabolism and infection identified both in spontaneous and induced samples.

Conclusions:
The metaproteomic-based approach allows quick and deep taxonomical and functional characterisation of the CF lung microbiota from sputum samples independently of the sampling method. The preliminary results obtained pave the way to numerous applications, e.g. to assess the impact of next-generation CFTR modulator therapies on the lung microbiota and the extent to which they might reduce susceptibility to chronic lung infection.

A METAPROTEOMICS APPROACH TO PROVE FUNCTIONAL STABILITY WITHIN THE INTESTINAL MICROBIOME OF PIGLETS AFTER CHALLENGE WITH CHLOROTONIL A

Madita Brauer (1,2), Stefanie Wagner (3), Arne Bublitz (4), Claudia Hirschfeld (1), Jennifer Herrmann (5), Rolf Müller (5), Dörte Becher (1), Susanne Sievers (1), Till Strowig (4), Thilo M. Fuchs (3), Katharina Riedel (1)

1) *Institute of Microbiology, University of Greifswald, Felix-Hausdorff-Str. 8, D-17489 Greifswald*
2) *Institute of Marine Biotechnology, Walther-Rathenau-Str. 49A, D-17489 Greifswald*
3) *Institute of Molecular Pathogenesis, Federal Research Institute for Animal Health, Naumburger Str. 96a, D-07743 Jena*
4) *Helmholtz Centre for Infection Research, Inhoffenstraße 7, D-38124 Braunschweig*
5) *Helmholtz Institute for Pharmaceutical Research Saarland, Department Microbial Natural Products, Helmholtz Centrefor Infection Research and Pharmaceutical Biotechnology, Saarland University, D-66123 Saarbrücken*

In healthy individuals an intact gastrointestinal microbiome protects against infections with the important anaerobic pathogen Clostridioides difficile. However, if colonization resistance is disturbed, for instance after antibiotic therapy, the pathogen can thrive inside the large intestine and cause disease. A sustainable therapy of
C. difficile infections therefore strongly relies on restoration of microbiome-mediated colonization resistance to prevent recurrence and reinfection. In consequence, antibiotics which selectively kill the pathogen while sparing the vast number of bacterial families within the intestinal tract are desirable. Data from in vitro experiments suggested that the macrocyclic natural compound Chlorotonils might meet these demands. To test this hypothesis, piglets were challenged with the lead compound Chlorotonil A and their feces were sampled before and after treatment to analyze the impact of Chlorotonil A treatment on the taxonomic and functional profile of the intestinal community by LC-MS/MS-based metaproteomics. To do so, proteins were extracted from feces using a phenol-based extraction method followed by suspension trapping-based protein digestion. Purified peptide samples were then analyzed in eight fractions by LC-MS/MS. Raw files were searched against a
sample-specific database using the search engine Mascot and data were further processed by using Scaffold and Prophane. Initial results of the analysis strongly suggest intestinal stability following Chlorotonil A treatment with only minor and very specific changes in the microbial community.

TACKLING CHALLENGES IN CLINICAL METAPROTEOMIC ANALYSIS OF BRONCHOALVEOLAR LAVAGE FLUID TO CHARACTERIZE MICROBIAL CONTRIBUTORS TO CYSTIC FIBROSIS

Monica E. Kruk (1), Pratik D. Jagtap (1), Subina Mehta (1), James E. Johnson (2), Reid Wagner (2), Tzu-Yi Yang (1), Emma Leith (1), Andrew McArdle (3), Chris H. Wendt (4,5), John B. O'Connor (6), Theresa Laguna (6,7), Timothy J. Griffin (1)

1) *Biochemistry, Mol. Biology and Biophysics, University of Minnesota, USA*
2) *Minnesota Supercomputing Institute, University of Minnesota, USA*
3) *Department of Infectious Disease, Imperial College London, London, UK*
4) *Pulmonary, Allergy, Critical Care and Sleep Medicine Section, Minneapolis Veterans Administration Health Care System,USA*
5) *Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, University of Minnesota, USA*
6) *Pulmonary and Sleep Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, USA*
7) *Pediatrics, Northwestern University Feinberg School of Medicine, USA*

The role of airway microbiota in the development and progression of cystic fibrosis (CF) lung disease remains unclear. MS-based metaproteomics of bronchoalveolar lavage fluid (BALF) can provide a unique look into the functional role of the lower airway microbiota in CF. To enhance detection of relatively low abundant microbial proteins, we used gas-phase fractionation (GPF) via FAIMS and high-resolution MS/MS, coupled with cutting-edge metaproteomic bioinformatic tools to characterize the metaproteome of BALF from CF patients. Our results demonstrate a more comprehensive portrayal of the CF microbiome afforded by this workflow for understanding microbial contributions in the disease, including insight into functional-taxonomic relationshipsmaking up the CF metaproteome.

BALF was pelleted through centrifugation, then lysed and trypsin digested before analysis with LC-MS/MS using an Orbitrap Eclipse equipped with high-field asymmetric ion mobility spectrometry (FAIMS). MetaNovo (https://doi.org/10.1101/605550) within the Galaxy for proteomics (Galaxy-P, galaxyp.org) platform was used for matching MS/MS to a large reference metaproteomics database (17M+ sequences) to generate a compact database of potential microbial constituents. The MSFragpipe tools suite was used for peptide spectrum match (PSM) analysis and protein inference. Galaxy-P tools such as Unipept (unipept.ugent.be/) and metaQuantome (10.1074/mcp.RA118.001240) were used for further characterization of taxonomy and function of identified peptides and proteins.

In initial studies, we generated a reference microbial metaproteomic database using MetaNovo. We then analyzed LC-FAIMS-MS/MS data from three CF samples using the MSFragpipe tool suite to generate PSMs against the microbial sequences appended to the human reference sequences. PSMs to microbial proteins from 446 spectra were generated, accounting for approximately 0.34% of total PSMs, with the vast majority matching to human proteins. Unipept analysis of the microbial PSMs identified 47 genera from 30 distinct bacterial species, although often represented by a single PSM, necessitating further validation. Among the confidently matched species, proteins from Stenotrophomonas maltophilia were identified, a multidrug resistant organism known to be a factor in CF, and Nocardia seriolae, an opportunistic pathogen linked to immunocompromised individuals, but with a lesser known role in CF. Compared to traditional LC-MS/MS, utilizing FAIMS suggested significant increases in PSMs of both human and microbial peptides and greater diversity of taxa. The proportion of microbial PSMs to the total acquired increased by 13% due to the enhanced selectivity of LC-FAIMS-MS/MS, indicating a potential enrichment for lower abundant peptide detection. Currently, we are constructing customized metaproteome sequence databases from a cohort of pooled CF and disease control patient samples, characterized by high or low microbial diversity. The pooled peptide mixtures will be fractionated by offline high pH RPLC, and concatenated fractions analyzed with LC-FAIMS-MS/MS. This deepcatalog of host and microbial proteins will be used to investigate protein signatures across 150 individual patient samples. Results obtained in this study through utilization of ultrasensitive MS-based analysis and informatics tools will be integrated with multi-omic data (16S rRNA, metabolomics, measurements of mucin integrity) fromthe same clinical BALF samples to gain a more complete understanding of microbial involvement in CF and assess taxonomic-functional relationships in disease progression.

SEWAGE TREATMENT PLANTS CAN INCREASE PREVALENCE OF ANTIMICROBIAL RESISTANCES IN AQUATIC ENVIRONMENTS

Tjorven Hinzke (1,2), Daniela Zühlke (1), Franziska Stressmann (3), Dominik Schneider (4), Katrin Premke (3), Rolf Daniel (4), Kathrin Riedel (1)

1) *University of Greifswald, Institute of Microbiology, Microbial Physiology & Molecular Biology*
2) *Institute of Marine Biotechnology e.V., Greifswald*
3) *Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin*
4) *Georg-August University of Göttingen, Institute of Microbiology and Genetics, Genomic and Applied Microbiology andGöttingen Genomics Laboratory*

Worldwide, antibiotics resistances are spreading at an alarming pace. Sewage treatment plants are potential hot spots for this spread of antibiotics resistances. To evaluate which antibiotic resistant bacteria are present in sewage treatment plants, which antibiotic resistance functions are actually expressed, and whether antibiotic-resistant bacteria can enter the environment via sewage treatment plants, we conducted a metaproteogenomic survey of two municipal treatment plants. We sampled the plants over a course of two years at several steps of the treatment process during all seasons, and also included the effluent-receiving aquatic environments. We integrated the metaproteomic data on abundances of classes of antibiotic resistance proteins with the taxonomy of the microbial communities at the different treatment steps, and with data on concentrations of antibiotics throughout the treatment plants. According to our data, sewage treatment plants can potentially increase and alter the prevalence of antimicrobial resistances in receiving environments: Several classes of antimicrobial resistance proteins had significantly higher relative abundances in the treatment plant effluents as compared to the receiving aquatic environments. The composition of microbial taxa producing these resistance proteins also differed between effluent and environment. In addition, several antibiotics were still detected in the effluent. Taken together, our study shows that sewage treatment plants contain a diverse microbial community producing a multitude of antibiotic resistance proteins, which can reach aquatic environments. This could lead to shifts in natural microbiota, and potentially to gene transfer events, with as yet to be determined ramifications forecosystem functions and human health.

METAPROTEOMICS INVESTIGATION ON THE ROLE OF THE GUT MICROBIOTA IN TYPE 1 DIABETES ONSET

Stefano Levi Mortera[a]*, Valeria Marzano[a], Pamela Vernocchi[a], Maria Cristina Matteoli[b], Alessandra Fierabracci[b], Valerio Guarrasi[c], Simone Gardini[c], Stefano Cianfarani[bd], Lorenza Putignani[e]

a) *Bambino Gesù Children's Hospital IRCCS – Human Microbiome Unit, Rome, Italy*
b) *Bambino Gesù Children's Hospital IRCCS – Diabetology and Growth Disorders Unit, Rome, Italy*
c) *GenomeUp S.R.L., Rome, Italy*
d) *Department of Systems Medicine, University of Rome 'Tor Vergata', Rome, Italy*
e) *Bambino Gesù Children's Hospital IRCCS –Microbiomics Unit, Human Microbiome Unit, Rome, Italy*

Type 1 diabetes (T1D) is the most frequent form of diabetes mellitus in childhood and adolescence and is characterized by a lack of insulin production caused by an autoimmune mediated progressive destruction of pancreatic β cells. The mechanism that induces the autoreactive activity of T lymphocytes is still not completely unveiled and recently some hypothesis have been raised, suggesting a link between the human gut microbiota (GM) profile and T1D onset.
In the last decade, 16S rRNA-based metagenomics has been extensively employed in investigations focused on GM ecology related to disease, focusing on its onset, progression, and response to treatments. Conversely, only few works have been published using metaproteomics or metabolomics approaches to interpret such relationships from a functional point of view.
We performed a comprehensive metaproteomic investigation on the GM of T1D children compared to a reference aged-matched healthy subjects control group (CTRLs) and to a smaller group of siblings, aiming at identifying microbial signatures associated with the triggering of the autoimmune response.
Combining the label free quantitation outcome with both the taxonomic assessment and the functional annotation of metaproteins, differential metaproteome patterns in the comparison of the T1D patients versus both siblings and CTRLs were observed. Particularly, up-regulated metaproteins, were found mainly associated with Bacteroides activities in T1Ds while in CTRLs up-regulated functions were mostly related to Bifidobacterium spp. These trends are in good accordance with what previously reported in a number of metagenomics studies, concerning taxa abundances. Interesting features were also observed by stratifying T1D children with respect to the insulin need threshold value, used as a proxy of the disease severity, showing a gradient of variation in both T1Ds and siblings when compared to CTRLs.
These results suggest that some specific GM features, particularly observed in "less severe" T1D children, might be shared in the familial milieu and be specifically associated with T1D risk.

LARGE-SCALE ANALYSES OF HUMAN MICROBIOMES REVEAL THOUSANDS OF SMALL, NOVEL GENES

Ami Bhatt

Department of Genetics, Stanford University, USA

Hundreds of the medications used in clinical medicine, such as penicillin, anthracyclines, and statins, are small molecule natural products of bacteria and fungi. Indeed, microbial products have transformed modern medicine. Despite these advances, many candidate drug targets, such as orphan G-protein coupled receptors (GPCRs) remain undruggable with available small molecules. Furthermore, many diseases, such as those caused by multidrug resistant (MDR) bacteria, have a high fatality rate given the lack of effective treatment options. Thus, we are in great need of additional breakthroughs in drug discovery. While scientists have aggressively prospected for novel small molecule drugs from microbes over the past decades, much less attention has been paid to microbial proteins, lipids, and carbohydrates. There is clear evidence that proteins are among the most prevalent molecules that mediate signaling at the cell surface throughout biology, which suggests great promise of these novel and largely under-mined drug-like proteins. However, it has been difficult to rationally identify small protein candidates for investigation. Recently, we overcame one of these major challenges in the protein space by identifying ≥4,500 new families of microbially encoded small proteins, of which a subset are predicted to be secreted and human microbiome-associated (Sberro et al, Cell 2019). We then developed and applied methods to validate the transcription/translation of these smORFs (Fremin et al, Nature Communications 2020; Fremin et al, Nature Protocols 2021) and automatically annotate these smORFs in metagenomic or genomic assemblies (Durrant and Bhatt, Cell Host and Microbe 2021). We postulate that some of these human microbe-associated small proteins mediate microbe-microbe or microbe-host communication; correspondingly, these small proteins have great promise as specific and potent potential drug-like molecules. By leveraging detailed information about the predicted structure, taxonomic restriction, and human microbe-associated nature of these molecules, we are embarking on an effort to identify and deeply mine a subset of these proteins for the potential role as drug-like molecules.

DISCOVERY OF NOVEL COMMUNITY-RELEVANT SMALL PROTEINS IN A SIMPLIFIED HUMAN INTESTINAL MICROBIOME

Hannes Petruschke (1), Christian Schori (2), Sebastian Canzler (1), Sarah Riesbeck (1), Anja Poehlein (3), Rolf Daniel (3), Christoph Kaleta (4), Nico Jehmlich (1), Christian Ahrens (2), Martin von Bergen (1,5)

1) *Department of Molecular Systems Biology, Helmholtz-Centre for Environmental Research – UFZ GmbH, Leipzig,Germany*
2) *Agroscope, Molecular Diagnostics, Genomics & Bioinformatics and SIB Swiss Institute of Bioinformatics, Wädenswil,Switzerland*
3) *Institute of Microbiology and Genetics, Department of Genomic and Applied Microbiology, Georg-August University ofGöttingen, Germany*
4) *Research Group Medical Systems Biology, Institute for Experimental Medicine, Christian-Albrechts-University Kiel,Germany.*
5) *Institute of Biochemistry, Faculty of Biosciences, Pharmacy and Psychology, University of Leipzig, Germany.*

The intestinal microbiota plays a crucial role in protecting the host from pathogenic microbes, modulating immunity and regulating metabolic processes. We studied the simplified human intestinal microbiota (SIHUMIx) consisting of eight bacterial species with a particular focus on the discovery of novel small proteinswith less than 100 amino acids (= sProteins), some of which may contribute to shape the simplified human intestinal microbiota. Although sProteins carry out a wide range of important functions, they are still often missed in genome annotations, and little is known about their structure and function in individual microbes andespecially in microbial communities.

We created a multi-species integrated proteogenomics search database (iPtgxDB) to enable a comprehensive identification of novel sProteins. Six of the eight SIHUMIx species, for which no complete genomes were available, were sequenced and de novo assembled. Several proteomics approaches including two earlier optimized sProtein enrichment strategies were applied to specifically increase the chances for novel sProtein discovery. The search of tandem mass spectrometry (MS/MS) data against the multi-species iPtgxDB enabled the identification of 31 novel sProteins, of which the expression of 30 was supported by metatranscriptomics data. Using synthetic peptides, we were able to validate the expression of 25 novel sProteins. The comparison ofsProtein expression in each single strain versus a multi-species community cultivation showed that six of these sProteins were only identified in the SIHUMIx community indicating a potentially important role of sProteins in the organization of microbial communities. Two of these novel sProteins have a potential antimicrobial function. Metabolic modelling revealed that a third sProtein is located in a genomic region encoding several enzymes relevant for the community metabolism within SIHUMIx.

We outline an integrated experimental and bioinformatics workflow for the discovery of novel sProteins in a simplified intestinal model system that can be generically applied to other microbial communities. The further analysis of novel sProteins uniquely expressed in the SIHUMIx multi-species community is expected to enablenew insights into the role of sProteins on the functionality of bacterial communities such as those of the humanintestinal tract.

MANTIS: FLEXIBLE AND CONSENSUS-DRIVEN GENOME ANNOTATION

Pedro Queiros (1), Francesco Delogu (1), Oskar Hickl (2), Patrick May (2), Paul Wilmes (1)

1) *Systems Ecology, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6 Avenue duSwing, 4367 Esch-sur-Alzette, Luxembourg*
2) *Bioinformatics Core, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6 Avenue du Swing, 4367 Esch-sur-Alzette, Luxembourg*

Background

The rapid development of the (meta-)omics fields has produced an unprecedented amount of high-resolution andhigh-fidelity data. Through the use of these datasets we can infer the role of previously functionally unannotatedproteins from single organisms and consortia. In this context, protein function annotation can be described as the identification of regions of interest (i.e., domains) in protein sequences and the assignment of biological functions. Despite the existence of numerous tools, challenges remain in terms of speed, flexibility, and reproducibility. In the big data era, it is also increasingly important to cease limiting our findings to a single reference, coalescing knowledge from different data sources, and thus overcoming some limitations in overly relying on computationally generated data from single sources.

Results

We implemented a protein annotation tool, Mantis, which uses database identifiers intersection and text mining to integrate knowledge from multiple reference data sources into a single consensus-driven output. Mantis is flexible, allowing for the customization of reference data and execution parameters, and is reproducible across different research goals and user environments. We implemented a depth-first search algorithm for
domain-specific annotation, which significantly improved annotation performance compared to sequence-wide annotation. The parallelized implementation of Mantis results in short runtimes while also outputting high coverage and high-quality protein function annotations.

Conclusions

Mantis is a protein function annotation tool that produces high-quality consensus-driven protein annotations. It is easy to set up, customize, and use, scaling from single genomes to large metagenomes. Mantis is available underthe MIT license at https://github.com/PedroMTQ/mantis.

CRITICAL ASSESSMENT OF METAGENOME INTERPRETATION - THE SECOND ROUND OF CHALLENGES

F. Meyer[1,2], A. Fritz[1,2,3] , A. C. McHardy[1,2,3,4,5]

1) Computational Biology of Infection Research, Helmholtz Centre for Infection Research, Braunschweig, Germany.
2) Braunschweig Integrated Centre of Systems Biology (BRICS), Braunschweig, Germany.
3) German Center for Infection Research (DZIF), Hannover-Braunschweig site.
4) Cluster of Excellence RESIST (EXC 2155), Hannover Medical School, Hannover, Germany
5) Technische Universität Braunschweig, Braunschweig, Germany

Evaluating metagenomic software is key for optimizing metagenome interpretation and focus of the community-driven initiative for the Critical Assessment of Metagenome Interpretation (CAMI). In its second challenge, CAMI engaged the community to assess their methods on realistic and complex metagenomic datasets with long and short reads, created from ~1,700 novel and known microbial genomes, as well as ~600 novel plasmids and viruses. Altogether 5,002 results by 76 program versions were analyzed, representing a 22x increase in results. Substantial improvements were seen in metagenome assembly, some due to using long-read data. The presence of related strains still was challenging for assembly and genome binning, as was assembly quality for the latter. Taxon profilers demonstrated a marked maturation, with taxon profilers and binners excelling at higher bacterial taxonomic ranks, but underperforming for viruses and archaea. Assessment of clinical pathogen detection techniques revealed a need to improve reproducibility. Analysis of program runtimes and memory usage identified highly efficient programs, including some top performers with other metrics. The CAMI II results identify current challenges, but also guide researchers in selecting methods for specific analyses.

CAN WE PREDICT A MICROBIAL ECOSYSTEM?

Francesco Delogu, Benoît J. Kunath, Patrick May, Paul Wilmes

Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg

Forecasting of microbial ecosystems allows to predict the state of the given microbial community via modellingof the present data and the environmental variables. Biological wastewater treatment plants (BWWTPs) can become a model systems to perfect such practice. BWWTPs have a constant influx of new populations whilst their parameters are partially controlled. Moreover, their microbial communities perform chemical transformations (e.g. nitrification and denitrification) and the lipid fraction in wastewater remains an untapped energy source. We characterized the community from a BWWTP foam to undercover its seasonal and physicochemical-dependent patterns via a temporal meta-omics time series over one-year period (weekly samples). Our analysis rendered the overarching biochemical key components of the systems and predicted the community patterns over the following 5 years, to be tested with 21 extra samples. Our findings suggest a cyclical and dynamic interaction of the taxa and genetic pool in the community maintains certain metabolic functions, highlighting critical nodes in the reaction network that should be considered when devising improvements and/or direct exploitation of these (or similar) processes for lipid harvesting in wastewater treatment plants. Moreover, the study demonstrates the power of microbial ecology forecasting as a tool to be deployed alongside meta-omics techniques.

TRANSKINGDOM NETWORK ANALYSIS ACROSS THE HOST-MICROBIOME NEXUS IN THE BOVINE RUMEN REVEALS ASSOCIATION TO HOST FITNESS

Ianina Altshuler (1), Arturo Vera Ponce De Leon (1), Rainer Roehe (2), Mick Watson (3), Carl Mathias Kobel (1), Aurélien Bernard (4), Milka Popova (4), Thea Os Andersen (1), Phillip B. Pope (1)

1) *Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Ås, Norway*
2) *Scotland's Rural College, Edinburgh, UK*
3) *The Roslin Institute and the Royal (Dick) School of Veterinary Studies, University of Edinburgh, UK*
4) *Université Clermont Auvergne, INRAE, VetAgro Sup, UMR Herbivores, F-63122 Saint-Genes-Champanelle, France*

Host-microbiome interactions are complex and difficult to pinpoint; yet are essential for the physiology and fitness of the host. Bovine rumen microbiota is particularly necessary for the survival of the host, as the microbe-mediated degradation of plant matter produces short-chain fatty acids required for the host's energetic requirements. In turn, the host is also thought to shape the structure of the microbiome. However, it is unclear how the molecular interactions of the microbiome with the host are linked to host fitness. Host-microbiome metaproteomic co-expression and co-exclusion networks can elucidate the active synergistic and antagonistic interactions of the bovine host and the symbiotic microbial population housed within the rumen. Our transkingdom metaproteomic analysis of microbial and bovine protein expression from the rumen epithelium tissue and the rumen interior of beef cattle demonstrate that unique microbial and bovine proteins are differentially expressed and spatially localized within the rumen. The majority of microbial proteins in the rumen interior were of bacterial, archaeal, and protozoan origin, while fungal proteins were recovered at a much lower rate. Microbial proteins were mostly recovered from the interior of the rumen compared to rumen epithelial tissue, however, several bacterial proteins belonging to members of Lachnospiraceae and Acutalibacteraceae were predominantly expressed within the rumen epithelium tissue compared to the rumen interior. These epithelium tissue-associated microbial proteins were involved in starch synthesis, peroxidases, and butyrate metabolism. Entodinium protozoa proteins were predominantly recovered from the interior of the rumen and included structural, housekeeping, and cilia proteins, as well as glycogen phosphorylases. Bovine proteins were mostly expressed within the rumen epithelial tissue; however, several bovine proteins were recovered at a higher rate from the rumen interior suggesting active secretion, these included gastric lipase and salivary proteins.

Network analysis of co-expression and co-exclusion of host-microbe proteins revealed links between the bovine host proteins with protozoan and bacterial proteins. Specifically, a modular network of proteins belonging to members of the Prevotella genus and Succinivibrionaceae, Ruminococcaceae, Acutalibacteraceae, Oscillospiraceae, and Lachnospiraceae families were linked to the expression of bovine and protozoan proteins. While a portion of these microbial network proteins was related to housekeeping functions, proteins involved in nitrogen metabolism, glucogenesis, pyruvate metabolism, and membrane binding were also part of this network. Coincidentally, expression of these microbial proteins was also positively and significantly correlated to the average daily feed intake (DFI) of the host and the expression of host bovine bacteriolytic lysozymes, lipid-binding proteins, and proteins involved in the vitamin B6 metabolism. Overall, our results suggest that host-microbiome interactions can potentially directly affect host fitness metrics.

DIET DRIVES HOST AND MICROBIAL CHANGES IN THE GASTROINTESTINAL TRACT OF RAINBOW TROUT (Oncorhynchus mykiss).

Arturo Vera-Ponce de León (1), Jacob Agerbo Rasmussen (2), Sabina Leanti La Rosa (1), Simen Rød Sandve (1), Morten Tømsberg Limborg (2), Phillip B. Pope (1)

1) Norwegian University of Life Sciences, 1432, Ås, Norway
2) University of Copenhagen, Copenhagen, Denmark

Aquaculture is one of the most expanding food-production sectors in the world and thus efficient and environmentally sustainable practices are required to ensure its long-term security. To solve these challenges, attractive solutions include developing new functional feed ingredients and better broodstock genetics to improve fish production and welfare. Intriguingly, it has been shown that both feed and host genetics can modulate the gut microbiome of animals and thus influence its integral connection to host phenotype. Within thiscontext, we aim to decipher the intimate functional coupling along the feed-microbiome-host axis in an applied context, with the emphasis on promising 'next generation' pro- and symbiotic feed supplements. Here we presentpreliminary data of our joint projects, which uses a (holo)genome-centric metaproteomic approach to gain insights into diet-driven changes produced in the gut of the rainbow trout (Oncorhynchus mykiss). Rainbow troutwere fed with different dietary formulations 1) a commercial fish diet without any additives (CTRL), 2) control feed plus the commercial probiotic BACTOCELL (PRO) and 3) control feed with a 'synbiotic' additive, i.e. a mix of BACTOCELL and galacto-oligosaccharides (SYN). For metaproteomics, gut content from the fish hindguts were sampled across biological replicates and total peptides were sequenced by nanoLC-MS/MS using a timsTOF mass spectrometer. Gut proteins were identified by mapping the peptides to a database containing O.mykiss genome as well as a Mycoplasma and a Lactiplantibacillus metagenome assembled genomes (MAGs), which, based on 16S rRNA gene analysis, are known to constitute greater than 90% of the O. mykiss gut microbiome. On average, a protein ID rate of ~30 % was obtained in CTRL, PRO and SYN samples, whereas a principal component analysis showed that replicates in CTRL and PRO cluster together within the ordination space, but not for SYN samples. Hierarchical clustering (HC) analysis were used to identify patterns caused by changes in the detection of both host or bacterial proteins, which resulted in two clear clusters for CTRL and PRO samples. Additionally, the HC analysis showed that there were changes in the protein level for both the fishand bacteria present in the gut. Interestingly, a higher protein abundance of the detected Mycoplasma was observed in CTRL samples in comparison to PRO and SYN. Functional annotation showed that all Mycoplasma proteins were classified as 'uncharacterized proteins', highlighting the unclear metabolic roles this highly prevalent yet enigmatic intracellular population has in the microbiome of fish. In contrast, a chitinase (GH18) from Lactiplantibacillus was recovered, suggesting a possible interaction with the dietary components. These initial datasets indicate that changes in the diet clearly affect the bacterial protein profiles in the rainbow trout and it is envisaged that deeper analysis will help us to identify exploitable interactions between specific feed components and microbiome functions that can be used to improve fish phenotype.

PROMOTING METAPROTEOMICS: EMPOWERING RESEARCHERS THROUGH ONLINE, ON-DEMAND METAPROTEOMICS EDUCATIONAL RESOURCES AND TRAINING MATERIALS VIA THE GALAXY PLATFORM

Pratik Jagtap (1), Subina Mehta (1), Saskia Hiltemann (2), Emma Leith (1), James Johnson (1), Berenice Batut (3), Helena Rasche (2), Bjoern Gruening (3), Timothy Griffin (1)

1) Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN (USA).
2) Clinical Bioinformatics, Department of Pathology, Erasmus Medical Center, Rotterdam (The Netherlands).
3) Bioinformatics Group, Department of Computer Science, University of Freiburg, Freiburg (Germany)

In the last decade, metaproteomics has emerged as an important tool to decipher the functions expressed by the microbiome. Furthermore, mass spectrometry-based metaproteomics research differs from traditional single organism proteomics and presents unique challenges. As an emerging field, many analytical tools and workflows are available for conducting metaproteomics research. However, there is a need for accessible educational resources to introduce new researchers to meta-omics methods for sample preparation, data acquisition, and data analysis for metaproteomics research.

In order to address this ongoing challenge for metaproteomic bioinformatics, the Galaxy for Proteomics (Galaxy-P) team has been conducting workshops at various annual research conferences for the last four years and via novel online and on-demand resources for the last 18 months (1). For example, via the Galaxy Training Network (2), we have authored tutorials that introduce researchers to metaproteomics, quantitative analysis, and metatranscriptomics methods. These online tutorials consist of a systematic walkthrough of the steps involved in data analysis, accompanied by an instructional video tutorial for the researchers to learn at their own pace. The metaproteomics introduction tutorial, for example, guides researchers through the bioinformatics analysis of an ocean metaproteomics dataset to detect taxonomic groups and functional composition of the microbiome.
Researchers can navigate through quantitative metaproteomics tutorials and learn how to generate inputs for a software suite named metaQuantome. metaQuantome software suite, available via Galaxy platform, enables statistical analysis and visualization of taxonomic and functional outputs from mass spectrometry data. The Galaxy Training Network also hosts a short and long version of a tutorial focused on metatranscriptomics analysis. The workflows described in this metatranscriptomics tutorial enable taxonomic and functional microbiome analysis using community RNA-Seq (metatranscriptomics) data.

Metaproteomics researchers can learn about the Galaxy-implemented software tools and analytical workflows using pilot datasets. Equipped with these tutorials, researchers can also use these advanced workflows on their own datasets. These software tools and workflows are made available via free and open Galaxy public instances, such as Galaxy Europe, Galaxy Australia, and Galaxy Main (U.S.). Many researchers from all across the world have accessed these tutorials. The Galaxy-P team seeks to continue integrating promising new software tools and workflows from leading laboratories into the Galaxy platform and encourages researchers to share their interests so that newer bioinformatics workflows can be made accessible via this resource.

In November this year, the Galaxy-P Team is also organizing an American Society of Microbiology-funded online workshop for Indian students that will focus on teaching the basics of microbiome analysis. Along with the Metaproteomics Initiative, we also see an opportunity in promoting talks by prominent metaproteomics researchers along with accessible educational resources (3). We anticipate that this website will act as a go-to hub for metaproteomics training resources. It will greatly benefit the metaproteomics research field as newer audio-visual and educational resources are shared via this portal. We believe that adding more tools, workflows, and educational resources via the Galaxy Training Network and Metaproteomics Initiative will educate and attract new talent with diverse expertise to advance the field of metaproteomics research.

References

1) http://galaxyp.org/workshops/
2) https://training.galaxyproject.org/
3) https://metaproteomics.org/education/

MAPPING METAPROTEOMICS DATA ON METABOLIC PATHWAYS AND PROTOTYPING METABOLIC MODELING

Robert Heyer (1,2), Daniel Walke (2), Emanuel Lange (2), Kay Schallert (2), Gunter Saake (1), Udo Reichl (2,3), Dirk Benndorf (2,3)

1) Otto von Guericke University Magdeburg, Database and Software Engineering Group
2) Otto von Guericke University Magdeburg, Chair of Bioprocess Engineering
3) Max Planck Institute for Dynamics of Complex Technical Systems Magdeburg, Bioprocess Engineering

Background and Aims:
Metaproteomics studies keep producing protein lists with increasing length thanks to rapid improvements in mass spectrometry, bioinformatics tools, and high-quality metagenomes for protein identification. The protein lists enable researchers to reconstruct the taxonomic and functional composition of diverse microbiomes and better understand their impact on the environment, biotechnical applications, and human health. Even more important than the exploration of the microbiome composition is to study the actual metabolic function. Therefore, identified proteins are assigned to metabolic networks, and flux balance and flux vector analysis may reveal the metabolic fluxes.

However, automated assignment of the single omics features to metabolic pathways of individual species is challenging due to the large quantity of data, taxonomic constraints, and the lack of matching metabolic pathways in standard repositories (e.g., KEGG). Furthermore, knowledge is required describing how well proteinabundances correlate to the actual metabolic flows. Therefore, the aim of our work was to develop a tool for the automatic assignment of protein lists to metabolic pathways and explore its application for the analysis of metabolic fluxes of (meta)-proteomics studies.

Methods:
We developed a stand-alone web application called “MPA Pathway Tool” (http://141.44.141.132:9001/home, https://github.com/danielwalke/MPA_Pathway_Tool) using Java for the backend and ReactJS for the frontend. It consists of one panel for the flexible construction of metabolic pathways and one for mapping protein lists. Furthermore, we explored the mapping of protein abundances tometabolic pathways using the Cobra Toolbox.

Results:
After successfully validating our tool using a metaproteomics dataset of biogas plants with manually assigned metabolic pathways, we used it for a combined metagenome and metaproteomics study of another biogas plant. Additionally, we evaluated our tool for clinical samples from the human gut. Thereby, we could show, that patients with inflammatory bowel disease possess in comparison with healthy controls, higher variations in the fermentation of short-chain fatty acids.

Subsequently, we prototyped the integration of flux balance and flux vector to our tool using a proteomics Escherichia coli data set from literature. It fitted the actual growth rates and showed the possible metabolic fluxes. The incorporation of protein abundances using the autopacman function further reduced the solutionspace of the metabolic flux vectors.

Conclusion:
Our tool facilitates and accelerates the data evaluation of (meta)-proteomics data. In the future, we will add the functions for metabolic modeling and integrate our tool into the new version of the MPA Cloud.

UNIPEPT DESKTOP: GETTING UNIPEPT READY FOR PROTEOGENOMICS

Pieter Verschaffelt, Bart Mesuere, Peter Dawyndt

Ghent University

Introduction

In recent years, proteomics has emerged as a novel technique that is situated at the interface of proteomics and genomics. It aims at improving the identification rate of proteomics data by first analysing metagenomes of a sample to better target the analysis of metaproteomes from the same sample. The genes predicted from a collection of reads are used to build a custom database for identifying measured spectra, instead of large general-purpose reference databases traditionally used in metaproteomics analyses. Limiting the protein searchspace improves sensitivity and specificity of metaproteomics analysis.

Methods

Unipept is a leading metaproteomics analysis tool that was initially developed as a web application. Its inherent web-based nature, however, limits the amount of data that can be analyzed. To overcome this limitation, we developed the Unipept Desktop application that is designed to drastically increase the throughput and capacity ofdata analysis.

Results and conclusions

The first stable version of the Unipept Desktop app was released in January 2021. It does not yet provide supportfor analyzing metaproteomics samples with custom databases, but focuses on improved analysis throughput.

Current development focuses on expanding the desktop app with support for custom databases. One way to reachthis goal is providing an automated pipeline to filter Uniprot proteins for a given list of taxa and build a custom database that can be queried locally. As a result, no internet connection is required to query custom databases and bottlenecks caused by limited network bandwidth are no longer an issue. In a later stage, we would like to support true proteogenomics analysis and allow users to construct databases directly from an annotated collection of DNA reads.

NovoLign: PEPTIDE ANNOTATION OF DE NOVO METAPROTEOMICS BY DIAMOND ALIGNMENT

Hugo Kleikamp, Ramon van Valderen, Martin Pabst

TU Delft

Background and aims
Due to recent technological advances, metaproteomic methods have found increasing success as a method of quantifying the contribution of organisms within microbial communities. Reduced sequencing costs have led to aaccelerated database growth, which necessitates the use of focused databases generated from metagenomes whenapplying database matching algorithms. De novo sequencing can use large public databases without significantlydropping in speed, with the added benefit of avoiding bias against out-of-database organisms.

NovoBridge, a recently established de novo metaproteomics pipeline, takes advantage of peptide de novo sequencing to enable fast taxonomic profiling (within a minute per sample) without requiring metagenomic database construction. Albeit taxonomic profiles of communities at e.g. the family level has been found comparable to database search approach outputs, the established pipeline depends on exact sequence matches (which depends on fragmentation spectra with near complete fragment ion coverage) with comprehensive protein sequence databases. This impacts on proteome coverage and taxonomic resolution, as well as limits the detection of poorly classified organisms. Other limitations that prevent a general application of de novo sequencing are its lower number of annotated peptides, and lower sequence specificity, due to its larger databasesize. The inclusion of homology alignment allows for non-exact matches, which can reduce the impact of de novo sequencing errors, as well as annotate sequences of novel taxa.

Methods
Metaproteomic data of pure cultures and synthetic mock communities was used to develop and validate new annotation strategies. To improve the number of matches different error mitigation strategies were applied. Anew peptide annotation pipeline was developed around peptide alignment with DIAMOND, dubbed “NovoLign”. To improve the taxonomic resolution of short peptides different lowest common ancestor (LCA)algorithms were applied. Weighted and focused LCA algorithms are applied with a range of parameters, and their effect on the number and specificity is compared, as well as quantification accuracy.

Results
The number of peptide matches in NovoBridge pipeline can be increased dramatically by including missed-cleavage handling and multiple candidate submission. The newly developed pipeline NovoLign enables great flexibility in database construction and output processing, and can annotate novel peptides. Weighted LCAsignificantly increases the amount and specificity of matches, but also skews quantification towards specific organisms.

Conclusions
Incorporating error mitigation strategies into NovoBridge dramatically improves the number of matches. TheNovoLign pipeline was developed, based around peptide alignment with DIAMOND, which in combination with different LCA algorithms improves peptide recall and specificity, but introduces biases in the taxonomicdistribution, which require further improvement.

References

1) Kleikamp, H. B. C., Pronk, M., Tugui, C., Lin, Y. M., Loosdrecht, M. C. M. Van, Kleikamp, H. B. C., … Lin, Y. M. (2021). Methods in Brief Database-independent de novo metaproteomics of complex microbial communities ll ll Methods in Brief Database-independent de novo metaproteomics of complex microbial communities, 375–383.
2) Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND, 12(1). https://doi.org/10.1038/nmeth.3176

MS2RESCORE: USING PREDICTED FRAGMENT ION INTENSITIES AND RETENTION TIMES TO INCREASE IDENTIFICATION RATES IN METAPROTEOMICS

Tim Van Den Bossche (1,2), Ralf Gabriels (1,2), Bart Mesuere (1), Thilo Muth (3), Lennart Martens (1,2)

1) VIB - UGent Center For Medical Biotechnology, Ghent, Belgium
2) Department of Biomolecular Medicine, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium
3) Section eScience (S.3), Federal Institute for Materials Research and Testing, Berlin, Germany

Introduction

The field of metaproteomics, the study of the collective proteome of whole (microbial) ecosystems, has seensubstantial growth over the past few years. Despite its high relevance, the field still suffers from low identification rates in comparison to single-species proteomics. The underlying challenge here is a lack of sequence resolution and statistical validation in the current identification algorithms, which are typically designed for single-species proteomics [1,2].

Methods

To solve this issue, we updated and applied the machine learning-based MS²ReScore algorithm on several multi-species, metaproteomics datasets. In the original version of MS²ReScore [3], the search engine-dependentfeatures of Percolator [4] were replaced with MS2 peak intensity features by comparing the PSM with the
corresponding MS²PIP-predicted spectrum [5]. Here, we further improved the method by combining both featuresets and adding additional features from DeepLC [6], a novel deep learning retention time predictor. By combining all features, we gain enough sensitivity to drastically lower the estimated false discovery rate (FDR) threshold, while still retaining a higher number of identified spectra.

Results and Conclusion

When the updated MS²ReScore algorithm is applied on metaproteomics datasets, our results show that MS²ReScore leads to an increased identification rate, ranging from the number of PSMs to the taxonomic level,while the false discovery rate (FDR) remains under full control as validated in an entrapment experiment [7].

References

1) Colaert, N.., et al. 2011. Journal of Proteome Research 10 (12): 5555–61.
2) Muth, T. et al. 2015. Proteomics 15 (20): 3439–53.
3) C. Silva, A.S., et al. 2019. Bioinformatics 35 (24): 5243–48.
4) Käll, L. et al. 2007. Nature Methods 4 (11): 923–25.
5) Gabriels, R., et al. 2019. Nucleic Acids Research 47 (W1): W295–99.
6) Bouwmeester, R., et al. 2020. bioRxiv. https://doi.org/10.1101/2020.03.28.013003.
7) Vaudel, M., et al. 2012. Journal of Proteome Research 11 (10): 5065–71.

METAPROTEOMICS BOOSTED UP BY DATA-INDEPENDENT ACQUISITION: A COMPUTATIONAL PERSPECTIVE

Laura Elo

Medical Bioinformatics Centre, University of Turku

A common approach in metaproteomics is to use data-dependent acquisition (DDA) mass spectrometry, which is known to have limited reproducibility when analyzing samples with complex composition. To overcome such limitations, my lab develops computational tools to facilitate data-independent acquisition (DIA) metaproteomics. Recently, we demonstrated the applicability of DIA mass spectrometry for the analysis of complex metaproteomic samples. We have now developed both a DDA-assisted as well as a DIA-only approach for metaproteomics and demonstrated their feasibility using laboratory-assembled microbial mixtures as well as human fecal samples. The tools are available in our open-source software package. Both a web-based graphical user interface and a command line interface suitable for high-performance computing clusters are being developed.

NOVEL MICROFLUIDIC DEVICE FOR METAPROTEOMICS SAMPLE PREPARATION

Ranjith Kumar Ravi Kumar, Massamba Mbake Ndiaye, Joelle Vinh, Yann Verdier

Biological Mass Spectrometry and Proteomics, PSL University, SMBP LPC, CNRS UMR8249, ESPCI Paris, France

Metaproteomics has developed into a decisive field with the developments in genomics, proteomics, and associated bioinformatics. The impact can be seen with the growth of knowledge in identifying new proteins/structural modifications, characterization of new species, environmental studies, host-microbe interactions, and pathogenesis. Further, development in the field requires overcoming analytical challenges associated with Liquid Chromatography-Mass Spectrometry (LC-MS), especially in the case of heterogeneous multi-taxonomic environmental samples for which the organisms of interest may be very few in number. This requires a universal sample preparation method that enables efficient protein isolation and minimal loss from different species, even at very low numbers. This research focuses on miniaturizing the sample preparation process for LC-MS to overcome certain analytical limitations associated with existing procedures.

The ChipFilter microfluidic device consists of two microchambers of 0.6 µl volume separated by an ultrafiltration membrane composed of regenerated cellulose (10kD cut-off). The reaction chamber is essential for performing chemical reactions, with the membrane acting as a gate to stop large biomolecules, such as polypeptides, and pass small molecules. The continuous flow of the reactants increases exchanges by reducing the surface-to-volume ratio for reactions. Cell lysis, protein concentration, and rapid chemical or enzymatic treatment are performed in the ChipFilter, and direct loading of the elute to the LC-MS is adequate for analysis[1]. To demonstrate the sensitivity and performance towards microbial samples, the device was used to process mixed whole cells of Escherichia coli, Bacillus subtilis and Saccharomyces cerevisiae. To assess the efficiency, we compared the performances of ChipFilter with several existing sample preparations methods (in-solution, in-gel and filter- assisted sample preparation). All samples have been analysed by LC-MS/MS on a high-resolution mass spectrometer and sequence database searching using the Mascot search engine.

Results indicate deeper protein identification with ChipFilters compared to other methods (1588 proteins by ChipFilter Proteolysis (CFP) compared to 1410 by in-solution, 1289 in-gel, and 966 proteins by FASP proteolysis). Around 20% of the proteins' identifications were specific to the CFP method, indicating promisingproficiency in recognizing new proteins. The distribution of CFP identified proteins between the three species was uniform. In contrast, the in-gel proteolysis suffered from a negative bias for Bacillus subtilis. In parallel tothe increase of proteome depth, the greater number of unique peptides identified per protein allowed higher confidence for protein and PTM identifications. Moreover, the ChipFilter requires reduced reaction times especially for tryptic reaction, where only 1 hour was sufficient as compared to overnight digestions for other methods with excellent yields. This is suggested by the ratio of missed cleavages, similar to other methods.

Finally, the sample preparation in the microfluidic device can be automated, which opens the way to work withpathogenic samples safely.

As a proof of concept, raw environmental samples were processed in the ChipFilter before LC-MS, allowing theidentification of hundreds of proteins from various species. To conclude, our results clearly show that our ChipFilter microfluidic device can be helpful for heterogeneous sample preparation for the identification of multiple species in low-concentration samples.

References

1) Ndiaye MM, Ta HP, Chiappetta G, Vinh J. On-Chip Sample Preparation Using a ChipFilter Coupled to NanoLC-MS/MS for Bottom-Up Proteomics. J Proteome Res. 2020 Jul 2;19(7):2654-2663. doi: 10.1021/acs.jproteome.9b00832. Epub 2020 May 8. PMID: 32343577.

IMPROVING METAPROTEOMIC IDENTIFICATION BY FOCUSING ON THE SPECTRA THAT MATTER

Benoit Kunath (1), Shengbo Wang (2), Satwant Kaur (2), Juan Antonio (2), Rob Finn (2), Patrick May (1), Paul Wilmes (1)

1) Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg
2) EMBL-European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, United Kingdom

The human gut microbiome has been extensively studied because of its profound influence on human physiology and involvement in a range of chronic diseases. As microbial proteins perform the majority of cellular functions, their analysis via metaproteomics has emerged as the most relevant approach to characterize the functional expression of a microbial community and elucidate its activity. However, a majority of the mass spectrometry (MS/MS) spectra obtained from shotgun metaproteomics cannot be assigned to a peptide. Indeed, typical studies only achieve an identification rate of ~30% spectra, which substantially limits the analysis of the datasets to their full potential. The low percentage of spectra assigned to peptides is notably due to the limited extent of sequence databases as well as spectra with unknown or unconsidered modifications.

We developed a workflow to integrate sample-specific metagenomic sequencing information with the recent Unified Human Gastrointestinal Genome (UHGG) catalogue to improve the identification rate in metaproteomics. The workflow utilizes a comprehensive database while retaining the sample specificity andleads to improved MS/MS spectra identification.

Furthermore, metaproteomic samples do not necessarily have associated metagenomic data. Spectral library searching is a powerful technique that could mitigate the need for a sequence database and has the potential to increase the efficiency of the searches. While spectral libraries remain a substantially underutilized resource, here we present spectral clustering and libraries applied for the first time to metaproteomic datasets. We showthat our method can extract information from relevant, yet unidentified spectra and allow focusing on those for subsequent identification, notably via de novo analyses. Ultimately, our workflow identifies more spectra and retrieves information even from unidentified spectra, thus improving our metaproteomic analyses.

PROPOSAL OF A DECOY-FREE FDR APPROACH SUITABLE FOR METAPROTEOMICS

Olivier Pible, Virginie Jouffret, Karen Culotta, Jean Armengaud

Laboratoire Innovations Technologiques pour la Détection et le Diagnostic (Li2D), Service de Pharmacologie et Immunoanalyse (SPI), CEA, INRA, F-30207 Bagnols-sur-Cèze, France

Accurate and fast evaluation of the False Discovery Rate (FDR) of spectra-to-peptide sequences inference is a difficult task in Metaproteomics, because the extra-large databases used are often largely incomplete and include numerous non-sample-specific sequences, particularly when using microbiota gene catalogs or generalist databases.

The traditional approach relying on combined target-decoy databases doubles search time, decreases sensitivity because of a larger search space, and is often biased because of dataset-database inadequate matching. We propose a target-only FDR estimate based on a mixture-model of four beta distributions.

We verified its efficiency on a set of 94 result datasets, including 26 metaproteomics searches, and a specificsearch with a controlled mismatching metaproteomics database.

Based on these extensive results, we found this method to be adequate for FDR estimation at the Peptide-Spectrum Matches level for proteomics, proteogenomics, and metaproteomics searches.

DeePSIVal: DECOY-FREE PEPTIDE SPECTRUM MATCH VALIDATION USING A DEEP LEARNING APPROACH

Kay Schallert (1), Daniel Micheel (2), Robert Heyer (2), Gunter Saake (2), Udo Reichl (1,3), Dirk Benndorf (1,3)

1) Otto von Guericke University Magdeburg, Chair of Bioprocess Engineering
2) Otto von Guericke University Magdeburg, Database and Software Engineering Group
3) Max Planck Institute for Dynamics of Complex Technical Systems Magdeburg, Bioprocess Engineering Group

Objectives:
Metaproteomics uses shotgun proteomics and subsequent protein database searching to produce peptide spectrum matches (PSMs) that link experimental mass spectrometry data to protein identifications. Validation of these PSMs is an important step of data processing to address the problem of false positives. To estimate and control the false discovery rate (FDR), the target-decoy strategy is state-of-the-art, but this approach comes with several drawbacks: a higher computation time due to the decoy search, the fact, that the validity of PSMs is determined by all other PSMs in an experiment and an often varying or even wrong estimation of the FDR depending on the protein database chosen. The latter is especially problematic for metaproteomics, since large metagenomes are commonly used as protein databases. Here, our goal was to create a system that can validate a PSM individually by utilizing a deep learning approach.

Results:
We developed the tool DeePSIVal that uses pre-classified PSM data to create a model, which can then be deployed to classify PSMs via a REST interface. DeePSIVal is based on a convolutional neural network architecture, which requires a PSM as input or specifically the peak list of the experimentally determined peptidespectrum and the amino acid sequence of the identified peptide. The training can be performed in a reasonable time and the time for classification is negligible. Four different datasets were used to test the system: a mass spectrometry standard for a human cell line (HeLa), an E. coli culture, a complex biogas sample and a simple protein standard measured on a TimsTOF mass spectrometer. Training data was labelled by setting an FDR of 1% using the target-decoy approach and the best model achieves an FDR between 1.8% and 9% depending on the combination of sample data used for training and separate testing. Computation time for classification is negligible and the tool was deployed as a web service that will be made publicly available in the future.

Conclusion:
The tool DeePSIVal allows the validation of PSMs as a simple REST interface and requires no additional information other than the PSM itself. The FDR of 4% achieved by the mixture model falls short of the expected 1% that implies comparable results to the target-decoy approach. Further refinement of this prototype can likely be achieved by selecting the training data more carefully and by using more sensible parameters for the model itself.

APPLYING METAPROTEOMICS TO OCEAN ENVIRONMENTS: PROGRESS AND CHALLENGES

Mak Saito

Woods Hole Oceanographic Institution, Massachusetts, USA

The oceans are a major reservoir of Earth's biodiversity and have a critical in the role in maintaining Earth's carbon cycle. Metaproteomics provides a useful means to directly assess the functional capabilities of the marine microbes that drive biogeochemical cycles. Notably functional metaproteomic analysis can contribute to detection of nutrient stressors to indicate nutrients that are controlling ecosystem productivity. Moreover, the absolute abundance of enzymes can be directly measured to estimate biogeochemical reaction rates. Finally, data sharing capabilities are being developed through the Ocean Protein Portal to further enable global ocean research. Examples of these each topics will be discussed, with a focus on the underlying methodologies.

METAPROTEOMIC CHARACTERIZATION OF THE PARTICLE-ASSOCIATED MICROBIAL COMMUNITY OF A NORTH SEA SPRING PHYTOPLANKTON BLOOM

Robin Siebers (1), Doreen Schultz (1), Daniela Zühlke (1), Pierre Alexander Mücke (1), T. Ben Francis (2), Dörte Becher (1), Mia Bengtsson (1), Katharina Riedel (1)

1) Institute for Microbiology, University of Greifswald, Greifswald, Germany
2) Max-Planck Institute for Marine Microbiology, Bremen, Germany

While there has been extensive research on the planktonic microbial fraction of algal blooms, due to their complex composition only little is known about the particle-associated (PA) microbial communities. Our group established a protocol to extract proteins from the particulate fractions for metaproteomic analysis (1), which was now applied to analyze the PA microbial communities of a 2018 North Sea spring phytoplankton bloom.

Samples were taken on three timepoints from April to May 2018 near Helgoland, Germany, and the PA-fractions were sequentially filtrated through polycarbonate filters with pore sizes of 10 µm and 3 µm. Proteins were extracted from the filters via bead-beating followed by aceton precipitation. Extracts were separated and fractionated by 1D SDS-PAGE and in-gel trypsin digested. Samples were measured with an Orbitrap Velos^{TM} mass spectrometer, and Mascot and Scaffold were used for peptide-spectrum-matching, protein identification and protein grouping based on a metagenome-based database. Finally, Prophane was used for taxonomic and functional annotation of the identified protein groups.

Combining data of three technical replicates for all three timepoints and both filter sizes, we could identify atotal number of ~8,000 individual protein groups, considering only protein groups with two or more peptide-matches. Of these, 1,561 protein groups were found in two of three technical replicates, with 179 and 896 protein groups found specifically on the 3 µm and 10 µm filter, respectively. On average, 1,300 protein groups could be identified for each technical replicate. Taxonomic distribution of prokaryotes confirmed protein activity of classes already known from the planktonic fraction, e.g. Alpha- and Gammaproteobacteria as well as Bacteroidetes. While Proteobacteria were most abundant in all filter fractions of all timepoints, high abundances of Bacteroidetes were only found in the > 10µm fractions for the two dates in May. However, eukaryotic proteins made up more than 70% of the identified protein groups, and further investigation will be needed to get insight into that part of the metaproteome.

References

1) Schultz et al., Environmental Microbiology Reports (2020), An optimized metaproteomics protocol for a holistic taxonomic and functional characterization of microbial communities from marine particles, doi: 10.1111/1758-2229.12842

COMBINED METAGENOMICS AND METAPROTEOMICS OF A TWO-STAGE BIO GAS PLANT QUESTIONS THE BENEFIT OF THE HYDROLYSIS FERMENTER FOR PROCESS PERFORMANCE

Robert Heyer (1,2), Patrick Hellwig (1), Irena Maus (3), Nicole Meinusch (1), Daniel Walke (1), Andreas Schlüter (3), Julia Hassa (3), Alexander Sczyrba (3), Michael Klocke (1,4), Torsten Mächtig (5), Gunter Saake (2), Udo Reichl (1,6), Dirk Benndorf (1,6,7)

1) *Otto von Guericke University, Bioprocess Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany*
2) *Otto von Guericke University, Database and Software Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany*
3) *Center for Biotechnology (CeBiTec), Bielefeld University, Genome Research of Industrial Microorganisms, Universitätsstraße 27, 33615 Bielefeld, Germany*
4) *Institute of Agricultural and Urban Ecological Projects affiliated to Berlin Humboldt University (IASP), Philippstraße 13,10115 Berlin, Germany*
5) *Christian-Albrechts-Universität Kiel, Institute of Agricultural Engineering, Olshausenstr. 40, 24098 Kiel, Germany*
6) *Bioprocess Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, 39106Magdeburg, Germany*
7) *Microbiology, Anhalt University of Applied Sciences, Bernburger Straße 55, 06354 Köthen, Germany*

Motivation and aims

Anaerobic digesters (ADs) are important renewable energy sources in which complex microbial communities degrade biomass into biogas (methane and carbon dioxide). Biogas is converted into electricity and heat by combined heat and power plants. A fundamental understanding of the taxonomic and functional composition of microbial communities in biogas plants is important to increase biogas productivity and yield. The biogas process consists of four main steps: hydrolysis, acidogenesis, acetogenesis, and methanogenesis. Due to differentenvironmental requirements for the microbes involved in the process steps, the biogas process can be spatially separated hydrolysis fermenter optimized for hydrolysis and acidogenesis and the main fermenter optimized for acetogenesis and methanogenesis. Metaproteomics of an agricultural two-stage biogas plant was applied to show, the adaptation of the microbial community in hydrolysis and main fermenters on a taxonomic and functional level and its consequences for degradation performance.

Methods

Hydrolysis and main fermenter were sampled once a month for one year. A metagenome was sequenced at a single time point. The proteins of all samples were extracted with phenol extraction and measured with liquid chromatography coupled to a tandem mass spectrometer (LC-MS/MS). The resulting spectra were searched withthe MetaProteomAnalyzer (MPA) against the annotated and binned metagenome. Unique peptides were used forthe taxonomic and functional analysis. Additionally, the MPA_Pathway_Tool was used to map the microbiome data to the top 50 common metabolic pathways. Furthermore, the results of the metagenomic data were compared with the metaproteomic data.

Results

The technical and chemical parameters indicate a stable operation of the main fermenter, but varying conditionswithin the hydrolysis fermenter over the sampling period. Inline, microbiome composition was more variable inthe hydrolysis fermenter than in the main fermenter.

Metaproteomics analysis resulted in all triplicates from the 12-time points in 2,374,049 identified spectra and90,606 identified proteins. 262 Metagenome-Assembled Genome (MAG) could be generated and account foraround 75 % of all identified spectra. Interestingly, methanogenic Methanothrixis with 8.14 % (based on metaproteins) one of the top abundant high-quality MAGs (>90% completeness,

Conclusions

The combined metagenomic and metaproteomic analysis of a two-stage biogas plant showed the differentiation of the microbial communities due to environmental conditions in hydrolysis and the main fermenter. Hydrolysisof complex carbohydrates which is the limiting step takes place in the main fermenter instead of in the hydrolysis fermenter. This result questions the benefit of the separated hydrolysis step in two-stage biogas plants.

ADAPTATION OF A METHANOGENIC COMMUNITY TO DEMAND-ORIENTED BIOLOGICAL METHANATION

Hoda Khesali[1], Dirk Benndorf[1,2,3], Alexander Sczyrba[4] Udo Reichl[1]

1. *Bioprocess Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, 39106 Magdeburg, Germany*
2. *Otto von Guericke University, Bioprocess Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany;*
3. *Microbiology, Anhalt University of Applied Sciences, Bernburger Straße 55, 06354 Köthen, Germany*
4. *Center for Biotechnology (CeBiTec), Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany,*

Motivation and aims

Power to methane is increasingly discussed in the last decade as a crucial process to store the surplus of renewable electricity while moving forward on mitigation of the climate change. The conversion of green H_2 with biogenic CO_2 from the biogas production can upgrade biogas to pure biomethane. Furthermore, it could buffer the inherent oscillations of renewable electricity resources by converting the surplus of power to CH_4 and allowing its storage in the gas network.

While discontinuous H_2 feeding may damage the chemical methanation catalysts, biological methanation handled by microbial methanogenic archaea is guessed to be less sensitive to H_2 starvation. Nevertheless, changes in H_2 feeding could result in variations in the microbial communities and cause process disturbances.

Methods

In this study, a biological methanation fermenter running for more than a year (biogas as source of CO_2) was exposed to repeated H_2 fluctuations. The effect on headspace gas composition was followed by online gas chromatography. A metagenome was sequenced as database for protein identification. High-resolution metaproteomics combining SDS-PAGE and LC-MS/MS measurements was applied to reveal alterations in microbial communities on taxonomic and functional level.

Results

During discontinuous feeding of H_2, the community shifted towards a higher abundancy of archaea. The abundance of archaea increased from 15 % to 30% at the end of experiments. Taxonomical investigations at the order level displayed the strongest increase in abundancy for Methansarcinales and a smaller increase for Methanobacteriales. The energy metabolism of Methanobacteriales is limited to hydrogenotrophic methanogenesis, whereas Methanosarcinales can gain energy from a wide range of substrates such as acetate, methyl-compounds and also H_2 and CO_2. In contrast, the abundance of all bacterial species was reduced. As no increase in homoacetogenic bacteria and no strong decrease in pH value were detected, the drastic and continuous increase in the Methanosarcinales cannot be explained by an increase in volatile fatty acids such as acetic acid and propionate. Possibly, the discontinuous feeding of H_2 favoured the growth of Methanosarcinales with a lower affinity for H_2 rather than strictly hydrogenotrophic methanogenic archaea due to transiently increased H_2 concentration in the recovery phases.

The presence of extracellular proteases produced by highly abundant bacteria, e.g. Petrimonas implicated that the bacterial community was mainly feeding on biomass primarily produced by methanogenic archaea. On the hand, killing methanogens by proteolytic bacteria could limit the methane production of the community, on the other hand, biomass turnover using inactive biomass as substrate could support growth of methanogens by recycling essential nutrients.

Conclusions

Taken together, the ability of the microbial community to adapt to a discontinuous feeding regime confirms the robustness of biological methanation processes. Furthermore, it seems that discontinuous feeding strategies could be used to enrich methanogenic archaea during establishment of a microbial community. The high abundance of proteolytic bacteria needs further consideration regarding its impact on stability of the microbial community.

METAPROTEOMICS REVEALS ALTERATION OF THE GUT MICROBIOME IN WEANERS BY THE INGESTION OF THE MYCOTOXINS DEOXYNIVALENOL AND ZEARALENONE

Johan Saenz (1,2), Alina Kurz (1,2), Ursula Ruczizka (3), Maximiliane Dippel (3), Veronika Nagl (4), Bertrand Grenier (4), Andrea Ladinig (3), Jana Seifert (1,2), Evelyne Selberherr (5)

1) Institute of Animal Science, University of Hohenheim, Emil-Wolff-Str. 6-10, 70567 Stuttgart.
2) HoLMiR - Hohenheim Center for Livestock Microbiome Research, University of Hohenheim, Leon-ore-Blosser-Reisen Weg 3, Stuttgart, Germany.
3) University Clinic for Swine, University of Veterinary Medicine Vienna, Veterinaerplatz 1, 1210 Vienna, Austria;
4) BIOMIN Research Center, Technopark 1, 3430 Tulln, Austria.
5) Institute of Food Safety, Food Technology and Veterinary Public Health, Unit of Food Microbiology, University of Veterinary Medicine Vienna, Veterinaerplatz 1, 1210 Vienna, Austria.

Mycotoxins can cause different carcinogenic, hepatotoxic, teratogenic, and mutagenic effects on humans and animals. These secondary metabolites are of great concern for food safety as long-term studies have found that 88 % of the feed in the world is contaminated with at least one mycotoxin. Even though mycotoxins exert a negative effect on the gastrointestinal barrier function and seem to alter the gut microbiome and digesta composition, their effect on the gut microbiome function is scarcely known. For that reason, we used metaproteomics to evaluate the effect of the mycotoxins deoxynivalenol (DON) and zearalenone (ZEN) on the gut microbiome of 15 female weaners. Briefly, animals were randomly separated into groups and exposed for 28 days to feed contaminated with different concentrations of DON (DONlow: 870 µg/kg DON, DONhigh: 2493 µg/kg DON) or ZEN (ZENlow: 679 27 µg/kg ZEN, ZENhigh: 1623 µg/kg ZEN). Animals of the control group received an uncontaminated feed. At the end of the trial, digesta and mucosal content were collected from both jejunum and ileum. After, proteins were extracted, and the peptide mixtures were measured using a Q-Exactive HF-X mass spectrometer faced with an EasyLC 1000 nano-UHPLC. Identification and quantification of peptides/protein groups and their taxonomic assignment and functional annotation was done using MetaLab. In general, the influence of the mycotoxins on the bodyweight of weaners and clinical signs of mycotoxicosis was not observed. However, ZEN caused signs of hypoestrogenism such as enlargement of the vulva. The gut metaproteome composition in the high toxin groups shifted compared to the control and low mycotoxin groups, and it was also more similar between high toxin groups. DONhigh and ZENhigh changed the relative microbiome composition. DONhigh and ZENhigh decreased the relative abundance of Firmicutes and Planctomycetes compared to the control group, while Actinobacteria increased. Also, DONhigh was the treatment with more differentially abundant proteins compared to the control group (107 proteins), followed by DONlow (8 proteins). DONhigh and ZENhigh increased the abundance of proteins associated with the ribosomes and pentose-phosphate pathways while decreasing glycolysis and other carbohydrate metabolism pathways. Proteins related to genetic processing, mostly ribosomes, were ~2-5 times more abundant in DONhigh and ~0-2.8 times more abundant in ZENhigh compared to the control group. Moreover, DONhigh and ZENhigh increased 4.8 and 2.2 times the abundance of thioredoxin-dependent peroxiredoxin, suggesting that the mycotoxins may induce oxidative stress in gut bacteria. In summary, the alteration of the gut microbiome structure and function due to the ingestion of mycotoxins can be associated with an increase in oxidative stress.

FUNCTIONAL RESPONSE OF A LIGNOCELLULOLYTIC MICROBIAL CONSORTIUM DERIVED FROM COW RUMEN TO SUBSTRATE MODIFICATIONS INDUCED BY PRETREATMENT

Emeline Flajollet (1), Nico Jehmlich (2), Vincent Lombard (3,4), Bernard Henrissat (3,4,5), Guillermina Hernandez Raquet (1)

1) *Toulouse Biotechnology Institute - TBI, Université de Toulouse, CNRS, INRAE, INSA, 135 Avenue de Rangueil, F-31077Toulouse, France.*
2) *Department of Molecular Systems Biology, Helmholtz-Centre for Environmental Research - UFZ, Permoserstrasse 15,04318 Leipzig, Germany.*
3) *CNRS UMR 7257, Aix-Marseille University, F-13288, Marseille, France.*
4) *USC 1408 AFMB, INRAE, F-13288, Marseille, France.*
5) *Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.*

Background:
Understanding lignocellulose deconstruction is a growing industrial and social concern to developthe bioeconomy. In current biorefineries, biomass pretreatment is an unavoidable step to facilitate the enzyme activity and reduce the lignocellulose recalcitrance. The mix of enzymes required for biomass deconstruction varies according to the physico-chemical properties of the substrate. Microbial consortia enriched on lignocellulosic feedstocks represent an interesting source of complex mixture of carbohydrate active enzymes (CAZymes) that could be tailored to the properties of the substrate. Nevertheless, little is known about the specific enzyme profile expressed by microbial consortia on specific pretreated substrates. Here, we have elucidated how a microbial consortium adapts to transform wheat straw that has been pretreated by mechanicaland dry-chemical methods. This new knowledge could help to develop optimal microbial consortia and enzymatic cocktails for a more efficient biomass conversion.

Methods:
A cow rumen-derived anaerobic microbial consortium, named RWS, was grown with wheat straw treated by dry mechanical grinding to 0.1 and 2 mm particle size and dry-alkaline impregnation. Using a combination of metaproteomics and multivariate analysis, we investigated the diversity of the active RWS community members, their specific CAZyme expression profile and their metabolic abilities to convert biomassinto volatile fatty acids (VFA) to decipher how substrate modifications affect community structure and functional expression.

Results:
Our results demonstrated that, regardless of particle size, dry-alkaline pretreatment, improved the deconstruction of wheat straw. Metaproteomics revealed that although the taxonomic profile of RWS growing onthese pretreated substrates was relatively similar, the abundance of certain genera, such as Bacteroides and Lachnoclostridium, increased, suggesting adaptations to the substrates properties. Bacteroides are particularly known for their high metabolic capability to hydrolyze holocellulosic biomass without substrate attachment.
Metaproteomics analyses identified that, among the 15 most abundant genera, the abundance of 4 of them was positively impacted by alkali treatment, including the most dominant genus, Bacteroides. Dry-alkali pretreatment had a significant impact on the CAZymes expression, with 9 families overexpressed. Bacteroides and Clostridium were the major producer of these CAZymes, mainly related to xylanase degradation. These taxashowed signs of active regulation of their CAZyme repertory, showing a higher allocation indexes for some CAzymes when grown on dry-alkali pretreated substrates. Proteins involved in VFA production were also affected by the substrate changes induced by pretreatment, but such differences resulted from changes in taxa abundances rather than a regulatory mechanism of expression.

Conclusions:
Metaproteomic analysis allowed us to identify the active populations impacted by substrate changes induced by dry-chemical pretreatment. We also identified interesting CAZy families which, in combination with chemical pretreatment, could be exploited for the formulation of effective enzyme cocktails toenhance biomass bioconversion.