November 19th
Microbiota session

Mathew Ryan : Microbial Resource Centres and the Microbiome Challenge
CABI
Living microbial collections play an important role in underpinning scientific research and development. The World Data Centre for Microorganisms provides a global view of microorganisms held in *ex situ* collections, and there are almost 3.2 million strains available for reference and research, of which 849,724 are fungal strains. However, culture collections almost exclusively maintain strains in axenic format and rarely store mixed cultures or consortia. A microbiome is all the microbes present in any one ecosystem. Recently much attention has been focused on the need for collections to meet the needs of the microbiome research community.
Infrastructure to underpin microbiome research is fragmented at both a national and international level. For example, the requirements for microbiome often falls outside the remit existing ESFRI research infrastructures. The needs for biobanks and collections was reviewed as an output from the EU microbiome support project in early 2021[1].There are very few microbiome focussed resources. Projects such as the UK Crop Microbiome Cryobank (UK CMCB www.agmicrobiomebase.org) are seeking to bridge the gap between collections/biobanks and the microbiome research community. However, more is required and the microbial resource community now has to lead the discussion in association with research, policy and industry stakeholders. Further this should incorporate a one health approach covering human, animal and plant domains.

[1] Ryan MJ, Schloter M, Berg G, Kostic T, Kinkel LL, Eversole K, Macklin JA,Schelkle, B; Kazou M, Sarand I, Lima N et al. (2021) Development of microbiomebiobanks- challenges and opportunities. *Trendsin Microbiology* 29: 89-92 https://doi.org/10.1016/j.tim.2020.06.009

Thomas Riedel: The DZIF Pathogen Repository
Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany
The Pathogen Repository is an infrastructure unit of the German Center for Infection Research (DZIF), which is located at the Leibniz Institute DSMZ. Currently, the DZIF Pathogen Repository comprises approximately 3,000 authenticated, clinically relevant microorganisms from >200 bacterial genera. This collection of isolates is increasing continuously. Tailored to the individual needs of DZIF researchers, various deposit options have been developed (public collection, closed collection with exclusive access, security deposit and backup storage), which are free of charge. In order to support the selection of relevant strains by the depositors, a novel key strain concept has been developed, which includes relevant selection criteria (e.g. new pathogens, reference strains, strains of multicenter studies, important outbreak strains, microbiome strains). Thus, an infrastructure adapted to the needs of the DZIF TTUs was established at the Leibniz Institute DSMZ, which ensures the availability of quality-controlled, standardized and well-documented microorganisms and the associated data.

Perrine PORTIER: The Microstore project: Effects of preservation on taxonomic composition and functional diversity of microbial communities
Missimahou OUSSOU, Géraldine TAGHOUTI, Steven JAGLINE,Thomas LERENARD, Audrey LATHUS, Cécile DUTRIEUX, **Perrine PORTIER**

UnivAngers, Institut Agro, INRAE, IRHS, SFR QUASAV, CIRM-CFBP, F-49000 Angers,France

Plant-associated microbiotas play an important role in plant health and development. The use of native or synthetic microbiotas is developing with the hope to understand causality between the composition of microbiotas and plant resistance or to propose perspectives of application. If synthetic communities (SynCom) are fairly easy to manipulate, they do not represent the whole complexity of microbiotas. However,native plant microbiotas are composed of numerous microorganisms, which only about 60% are cultivable. Thus, the preservation of these communities will have an impact on their taxonomic composition, and in turn can have an impact on their functional diversity.
We studied the impact of preservation on plant-associated communities. Radish seed and leaves-associated microbiotas were extracted and preserved by three different methods (deep freezing at -80°C or at -196°C in liquid nitrogen, and lyophilization). Before and after preservation, the taxonomic composition(metabarcoding 16S, *gyrB* and ITS1)and the functional diversity (BIOLOG Ecoplates) were assessed with or without an amplification stage (culture on solid or in liquid medium).
This permitted us to assess the bias induced by preservation on the taxonomic composition and the functional diversity of native plant-associated communities.
This work permit to better understand how preservation and amplification impact the taxonomic composition and functional diversity of plant-associated microbiotas. These results will permit scientists to better choose how to manipulate the microbial communities.

Hauke Smidt: UNLOCK - integrated biodiscovery-, bioreactor- and FAIR data facilities to unlock microbial diversity for society

Personal Chair Complex Microbial Ecosystems
Head Molecular Microbial Ecology Group
Scientific Director UNLOCK Infrastructure for Microbiome Research
Laboratory of Microbiology
Wageningen University & Research

UNLOCK is an experimental and data platform, enabling breakthrough research and knowledge sharing on natural and synthetic mixtures of microorganisms. Such microbial communities are of key importance at different scales in our society, ranging from individual-based health issues related to microbial communities inhabiting the human body, to global greenhouse gas (i.e. CH_4 and N_2O) emissions related to microbial activity. Three major limitations exist in our current experimental procedures in research on microbial communities: (i) the lack of (medium to) high-throughput cultivation facilities for comparative analysis of microbial ecosystem development, particularly for anaerobic microbiomes, (ii) the effective integration of these cultivation studies with molecular systems characterization, and (iii) the transparent and uniform storage and processing of the generated data. It is exactly these three limitations that *UNLOCK* will address, through enabling research on mixed microbial communities at an unprecedented scale and efficiency. With *UNLOCK*, Wageningen University and Delft University of Technology have joined forces to integrate the expertise of the research groups involved in four complementary platforms. The first platform is the Biodiscovery platform for high-throughput discovery and characterization

of yet-uncultured microbes, specifically focusing on fastidious anaerobes. The Modular Bioreactor platform is specifically suitable for investigating sustainable solutions for environmental challenges, such as degradation of (micro)pollutants, sustainable energy generation, and recovery of resources from complex waste streams. The Parallel Bioreactor platform facilitates users to conduct dozens of high-resolution cultivation experiments in bioreactors in parallel for comparative analysis of how process variables affect system development. Finally, the FAIR-Data platform allows for data storage, data extraction and analysis of high-throughput data in a cloud-based infrastructureThe data generated will be FAIR by design, enabling transparent procedures.
*UNLOCK*represents a unique infrastructure for exploring new horizons for research onmicrobial communities. Owing to the smooth access procedure, the experimentalhardware available, and the expertise of involved researchers, *UNLOCK* enables research that to date isconsidered too complex and too expensive for one researcher (or one researchgroup) to conduct. Combining and exploring the fields of expertise of the *UNLOCK* research groups will boost theresearch community and strengthen the position of Dutch research in the fieldsof microbiology, microbial ecology, and bioprocess engineering. *UNLOCK*is open to excellence-driven users from universities, knowledge institutes andindustries, placing them in the unique position to conduct research at unmetspeed and resolution.

Vincent Thomas: Flow Cytometry for targeted culturomics of complex ecosystems species Bioaster, Paris , France

Sequencing studies have highlighted the association of deficiencies in a variety of gut commensal species with various pathological conditions. It is therefore of great interest to use well-characterized strains to complement dysbiotic microbiota. However, this approach is still hampered by the fact that there are usually only few or even no strains available for many species of interest, due to specific nutritive requirements, extreme oxygen sensitivity (EOS) or under-representation in the gut ecosystem. In an attempt to circumvent these limitations, we developed flow cytometry and cell sorting under anaerobic conditions using a modified BD Influx® cell sorter. Viability- and Gram staining as well as specific polyclonal antibodies were investigated as characterization tools. We demonstrated that viability of the EOS species *Faecalibacterium prausnitzii* was preserved during the anaerobic sorting process, while complete loss of viability was observed in normal sorting conditions. Staining procedures had only marginal effects on cultivability. Using antibodies directed against strains that belong to two different phylogroups, we established a collection of 15 strains of the EOS species *F. prausnitzii*. We then focused on the species *Christensenella minuta* that is usually found in very low amount in the human fecal microbiota and were able to establish a collection of 7 strains that belong to this health-related species. These developed tools allow rapid fingerprint of microbiota composition and accurate isolation of EOS bacteria from complex microbial communities. New developments are underway to apply reverse genomics strategies to identify sequences encoding immunogenic proteins in order to produce antibodies even in the absence of already cultivated strains.

MALDI-TOF session:

Anneleen Wieme: MALDI-TOF MS, an identification tool used in QC and culturomics projects @BCCM/LMG, Gent, Belgium.

Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry (MALDI-TOF MS) has been generally accepted as a rapid, high-throughput and low-cost tool for genus/species-level identification of pure bacterial cultures in medical, environmental and food-related projects. At BCCM/LMG, identification of bacterial cultures is offered as a high-throughput and fast customer service using the MALDI-TOF MS MicroflexTM LT/SH (smart) systems of Bruker Daltonics (Bremen, Germany). These commercial systems include an extensive database comprising numerous mass spectra generated mostly from clinically relevant bacteria. To meet the dynamic identification needs of the clients, BCCM/LMG continuously works on the development of a proprietary MALDI-TOF MS identification database with sufficient coverage of the intraspecies diversity.
Next to basic MALDI-TOF MS identification services, MALDI-TOF MS is also applied in the routine quality control of BCCM/LMG reference strains. In brief, mass spectra generated before and after preservation are compared to monitor the authenticity of the reference strains in a fast and straightforward manner.
More recently BCCM/LMG also started using MALDI-TOF MS for bacterial diversity studies in a cultivation-dependent approach, meaning cultivating the bacterial fraction present in a sample using a broad range of cultivation conditions as step one. After incubation, numerous colonies are isolated and processed via MALDI-TOF MS. A high-throughput dereplication screening including up to thousands of mass spectra generated can be performed using the in-house developed script SPeDE (Dumolin et al., 2019). This dereplication step enables us to group isolates that represent the same taxon. Dereplication is thus the assessment of novelty and aims to reduce a large number of isolates to a smaller, non-redundant set for further identification.

Sylvain Brisse : *Klebsiella* MALDI TypeR: a web-based tool for *Klebsiella* identification based on MALDI-TOF mass spectrometry

Biological Resources Center of Institut Pasteur (CRBIP), Paris, France
Research Unit Biodiversity and Epidemiology of Bacterial Pathogens,Institut Pasteur, Paris, France
Klebsiella pathogens affect human and animal health and are widely distributed in the environment. Among these, the *Klebsiella pneumoniae* species complex, which includes seven phylogroups, is an important cause of community and hospital infections. The *Klebsiella oxytoca*species complex also causes hospital infections and antibiotic-associated haemorrhagic colitis. The unsuitability of currently used clinical microbiology methods to distinguish species within each of these species complexes leads to high rates of misidentifications that are masking the true clinical significance and potential epidemiological specificities of individual species.MALDI-TOF reference spectra databases from commercial providers may be lagging behind recent taxonomic updates.
We developed a web-based tool, *Klebsiella* MALDITypeR, a platform-independent and user-friendly application that enables uploading MALDI-TOF mass spectrometry data in order to identify *Klebsiella*isolates at the species complex and phylogroup levels. The tool, available at https://malditypeR.pasteur.fr, leverages a database of previously identified biomarkers that are specific for species complexes, individual phylogroups, or related phylogroups. We obtained 84% to100% identification accuracy depending on phylogroup. Identification results are obtained in a few seconds from batches of uploaded spectral data.*Klebsiella* MALDI TypeR enables fast and reliable identification of *Klebsiella*s trains that are often misidentified with standard microbiological

methods. This web-based identification tool may be extended in the future to other human bacterial pathogens and might represent a useful tool for microbial biological resources centers.

Vincent Robert (KNAW, Bioware) : Towards a universal and community-based MALDI-TOF database and identification system

Vincent Robert [1],Teun Boekhout [1], Karl-Otto Kraeuter [2]
WesterdijkFungal Biodiversity Institute [1], The Netherlands; Bruker [2],Germany

MALDI-TOF technology was launched during the eighties and has been in use for bacterial identification since 1996. With more than 4000 MALDI Biotyper sold until 2020, Bruker is the leader in the market while BioMerieux is another serious player in the field.
The technique is now strongly established as a fast and reliable identification method for Bacteria and is becoming important for yeasts and fungi as well. 300 million identifications have been done with MALDI-TOF in 2020 alone, mainly in the hospital settings.
MALDI-TOF builders are also proposing reference databases that are well established and controlled but they certainly do not cover the whole range of microorganisms that are known today. Therefore, Bruker, the Center for Disease Control and Prevention(CDC) and the Westerdijk Institute (WI) have decided to join forces to create a virtually unified reference database that will allow users of Bruker MALDI Biotyper to identify unknown strains by comparing their profiles against their local database (including Bruker profiles), the CDC and/or the WI. The pilot project will start with the yeasts database (www.theyeasts.org).
If the project is successful, the concept will be opened to more data providers that would be interested to join and share their data with any or a selection of clients.

Genomic session:
Antonis Rokas : Evolutionary Genomics as a tool for studying fungal diversity and revising taxonomy

Antonis Rokas, Ph.D., Cornelius Vanderbilt Chair in Biological Sciences, Professor of Biological Sciences and Biomedical Informatics, and Director of the Vanderbilt Evolutionary Studies Initiative

1. *Department of BiologicalSciences, Vanderbilt University, Nashville, TN, 37235, USA*
2. rokas@vanderbilt.edu

Genome sequencing has revolutionized the study of fungal evolution and diversity. One of the major foci of work focuses on yeasts, unicellular fungi that do not form fruiting bodies. Although the yeast lifestyle has evolved multiple times, most known species belong to the subphylum Saccharomycotina (hereafter yeasts). This diverse group includes the premier eukaryotic model system, *Saccharomyces cerevisiae*; the common human commensal and opportunistic pathogen, *Candida albicans*; and over 1,000 other known species. Yeasts are found in every biome and continent and are more genetically diverse than either plants or bilaterian animals. Ease of culture, simple life cycles, and small genomes have made yeasts exceptional models for molecular genetics, biotechnology, and evolutionary genomics. Funded by the US National Science Foundation's Dimensions of Biodiversity and Rules of Life programs, we and our collaborators have undertaken a large-scale examination to sequence and study the evolution, diversity, and taxonomy of the type

strains of all ~1,200 known species in the subphylum. In my talk, I will describe how we use evolutionary genomic approaches to study the diversity of budding yeasts and revise their higher-level taxonomy.

Sarah Leeber : Implementation of genomics for the characterization of Lactobacillus diversity and taxonomy

Dept of Bioscience Engineering, University of Antwerp

Lactobacilli are among the most widely used bacteria for biotechnology and health applications. You can find them in food as natural fermentation microbes or deliberately added starter cultures and probiotics. They are also the most dominant bacteria in the female reproductive tract under healthy conditions. Similarly, lactobacilli can also dominate body sites of particular animals such as insects (honey bee gut) and chickens (crop). However, the species and genera that dominate these different habitats and environments generally differ quite a lot. We have recently updated the taxonomy of the old Lactobacillus genus complex to better grasp and show this diversity. This taxonomy update was largely driven by our comparative genomics pipeline. We used a combination of phylogenetic tree building based on core genes and the assignment of genera based on shared signature genes. Results from our Citizen Science projects on vaginal lactobacilli (https://isala.be/en) and fermented vegetables ('Ferme Pekes') will be presented.

Federica Palma : Genomic libraries to catalogue and analyse strain biodiversity within bacterial species: perspectives for mBRCs.

Federica PALMA and Sylvain BRISSE CRBIP, Institut Pasteur, Paris, France

Background

Genomic information is essential for taxonomic, phylogenetic, and evolutionary research,and to comprehensively decipher the characteristics of bacterial strains (*e.g.*,antimicrobial resistance or virulence), to explore their genotypic-phenotypic associations, and to answer a variety of biological questions. Strain's identification, classification, and nomenclature have also progressed considerably by harnessing of genome sequence information, allowing global laboratories and public health institutions to effectively communicate on strain subtypes using a unified language. This enables to better track emerging bacterial strains in One Health and Global Health contexts. To achieve this, advances in bioinformatics platforms and the development of genomic libraries are essential. BIGSdb-Pasteur, the genomic taxonomy web platform of Institut Pasteur's mBRC (CRBIP), is a successful example of integrated resource of genomic sequences and isolate's information (*i.e.*, genomic libraries) and bioinformatics tools enabling the storing, sharing and comparison of data.

Objectives

BIGSdb-Pasteur,powered by the Bacterial Isolate Genome Sequence database software tool (developedat Oxford University), integrates a set of genome libraries that provide genome-based identification and strain genotyping systems. A nomenclature ofvariants and attached Linnaean taxonomy is used for identification (ribosomal MLST), while nomenclatures of sublineages and strains are maintained for specific bacterial pathogen species (*Klebsiella*, *Listeria*, etc)and curated by expert communities, serving as a one-stop platform for bacterial strain classifications and taxonomy.

Results

BIGSdb-Pasteur hosts a large collection of curated typing schemes (multilocus sequence typing(MLST), core-genome MLST, virulence, antimicrobials and heavy metals resistance,among others) along with several thousand of isolates and sequence data from bacterial species representing the diversity of Pasteur's mBRC and species of public health importance (e.g., *Klebsiella pneumoniae* and *Listeria monocytogenes*) or industrial interest (e.g., *Lactobacillus casei*). All data in BIGSdb-Pasteur can be easily uploaded, accessed and queried through the web platform with different levels of permissions (*i.e.*, private or public libraries). Although BIGSdb-Pasteur follows the philosophy of open data, creation of dedicated private databases and projects is supported to ensure the privacy of sensitive data, pending original publications and patent deposits. Graphical visualisation tools are also implemented to support users in the interpretation and valorisation of data. Genomic libraries in BIGSdb-Pasteur are dynamic resources that are constantly expanded, and several expert groups are dedicated to database curation and maintenance. Currently, thousands of users contribute by submitting data from known or newly emerging bacterial genotypes on a daily basis. The BIGSdb-Pasteur web platform is freely accessible via a web interface at https://bigsdb.pasteur.fr and automatically via a RESTful web service and application programming interface.

Conclusions

Collaborative projects are being defined between mBRCs and the genomic libraries platform, with the aim of genome-based identification, strain taxonomy and comparative genomics. Accessibility of genomic data from mBRC resources will contribute to valorise collections of bacterial strains. Where large collections of strains of a single bacterial species are maintained, the genomic libraries can help organise and characterise the vast amounts of biodiversity within species, making it easier to locate relevant information and pick representative strains for preservation and distribution by mBRCs. Vice-versa, strain identifiers from mBRCs made available within the genomic libraries will facilitate detection of resources of interest based on genomic data, to eventually define their phenotypic properties for clinical or industrial purposes.