Unraveling Microbial Effectors: Metaproteomics in the One Health Framework

Robert Heyer[1], Maximilian Wolf[2], Dirk Benndorf[3], Sergio Uzzau[4], Jana Seifert[5], Lucia Grenga[6], Martin Pabst[7], Heike Schmitt[8], Bart Mesuere[9], Tim Van Den Bossche[10], Sven-Bastiaan Haange[11], Nico Jehmlich[11], Mariagrazia Di Luca[12], Manuel Ferrer[13], Sergio Serrano-Villar[14], Jean Armengaud[15], Helge Bode[16], Patrick Hellwig[17], Catherine Robbe Masselot[18], Renaud Léonard[19], Paul Wilmes[20]

1. Multidimensional Omics Analyses group, Leibniz-Institut für Analytische Wissenschaften – ISAS – e.V., Bunsen-Kirchhoff-Straße 11, 44139 Dortmund, robert.heyer@ISAS.de, Multidimensional Omics Analyses group, Faculty of Technology, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, robert.heyer@uni-bielefeld.de
2. Multidimensional Omics Analyses group, Faculty of Technology, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, maximilian.wolf@uni-bielefeld.de
3. Chair of Bioprocess Engineering, Otto-von-Guericke University Magdeburg, Germany, Bioprocess Engineering, Max Planck Institute for Dynamics of Complex Technical Systems Magdeburg, Germany, Microbiology, Anhalt University of Applied Sciences, Germany, dirk.benndorf@hs-anhalt.de
4. Department of Biomedical Sciences, University of Sassari, 07100 Sassari, Italy, uzzau@uniss.it
5. Institute of Animal Science, University of Hohenheim, Emil-Wolff-Str., Stuttgart, Germany; HoLMiR - Hohenheim Center for Livestock Microbiome Research, University of Hohenheim, Leonore-Blosser-Reisen Weg, Stuttgart, Germany, seifert.jana@uni-hohenheim.de
6. Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), SPI, Bagnols-sur-Cèze, France, lucia.grenga@cea.fr
7. Prof. Martin Pabst, Department of Biotechnology, Delft University of Technology, Delft, The Netherlands, m.pabst@tudelft.nl
8. Prof. Heike Schmitt, Department of Biotechnology, Delft University of Technology, Delft, The Netherlands, and Institute for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands, h.schmitt@tudelft.nl
9. Prof. Bart Mesuere, Department of Applied Mathematics, Computer Science and Statistics, Ghent University, 9000 Ghent, Belgium, bart.mesuere@ugent.be

10. Dr. Tim Van Den Bossche, (i) VIB - UGent Center for Medical Biotechnology, VIB, 9052 Ghent, Belgium, (ii) Department of Biomolecular Medicine, Faculty of Medicine and Health Sciences, Ghent University, 9052 Ghent; Belgium, tim.vandenbossche@ugent.be
11. Helmholtz-Centre for Environmental Research - UFZ GmbH, Department of Molecular Toxicology, Permoserstrasse 15, 04318 Leipzig, Germany, sven.haange@ufz.de, nico.jehmlich@ufz.de
12. Department of Biology, University of Pisa, 56127 Pisa, Italy, mariagrazia.diluca@unipi.it, ORCID: 0000-0002-0688-034X
13. Instituto de Catalisis y Petroleoquimica (ICP), CSIC, 28049 Madrid, Spain, mferrer@icp.csic.es, ORCID: 0000-0003-4962-4714
14. Department of Infectious Diseases, Hospital Universitario Ramón y Cajal, IRYCIS and Universidad de Alcalá, 28034 Madrid, Spain, CIBERINFEC, Instituto de Salud Carlos III, 28029 Madrid, Spain, sergio.serrano@salud.madrid.org, ORCID: 0000-0002-5447-3554
15. Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), SPI, Bagnols-sur-Cèze, France, jean.armengaud@cea.fr
16. Department of Natural Products in Organismic Interactions, Max-Planck-Institut for terrestrial Microbiology, Karl-von-Frisch-Str. 10, 35043 Marburg, Germany, Molecular Biotechnology, Department of Biosciences, Goethe-University Frankfurt,60438 Frankfurt, Germany, Center for Synthetic Microbiology(SYNMIKRO), Phillips University Marburg, 35043 Marburg, Germany, Department of Chemistry, Phillips University Marburg, 35043 Marburg, German, LOEWE Centre for translational Biodiversity Genomics (LOEWE-TBG) &Senckenberg Gesellschaft für Naturforschung, 60325Frankfurt, Germany, helge.bode@mpi-marburg.mpg.de
17. Chair of Bioprocess Engineering, Otto-von-Guericke University Magdeburg, Germany, Bioprocess Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, 39106 Magdeburg, Germany, hellwig@mpi-magdeburg.mpg.de
18. Université de Lille, CNRS, UMR 8576 - UGSF, Lille, France, catherine.robbe-masselot@univ-lille.fr
19. Université de Lille, CNRS, UMR 8576 - UGSF, Lille, France, renaud.leonard@univ-lille.fr
20. Luxembourg Centre for Systems Biomedicine, University of Luxembourg, L-4362 Esch-sur-Alzette, Luxembourg, Department of Life Sciences and Medicine, University of Luxembourg, L-4362 Esch-sur-Alzette, Luxembourg paul.wilmes@uni.lu

Abstract

One Health seeks to integrate and balance the health of humans, animals, and environmental systems. These three spheres are intricately interconnected through microbiomes, which are universally present and exchange microbes and genes, influencing not only human and animal health but also key environmental, agricultural, and biotechnological processes. Preventing the emergence of pathogens as well as monitoring and controlling the composition of microbiomes through microbial effectors including virulence factors, toxins, antibiotics, non-ribosomal peptides, and viruses holds transformative potential. However, the mechanisms by which these microbial effectors shape microbiomes and their broader functional consequences in relation to host and ecosystem health remain poorly understood to date. Metaproteomics offers a novel methodological framework as it provides insights into microbial dynamics by quantifying microbial biomass composition, metabolic functions and detecting effectors like viruses, antimicrobial resistance proteins, and non-ribosomal peptides. Here, we document the potential of metaproteomics for elucidating microbial effectors and their impact on microbiomes, and discuss their potential for modulating microbiomes to foster desired functions.

Abbreviations

AHA: L-azidohomoalanine

AMP: Antimicrobial peptides

BGC: Biosynthetic gene cluster

BONCAT: Bioorthogonal non-canonical amino acid tagging

DDA: Data-dependent acquisition

DIA: Data-independent acquisition

LC: Liquid chromatography

LOD: Limit of detection

FDR: False discovery rate

HPG: L-homopropargylglycine

LOQ: Limit of quantification

MAG: Metagenome-assembled genome

MD: Molecular dynamics

NRP: Non-ribosomal peptide

NRPS: Non-ribosomal peptide synthetases

ompA: Outer membrane proteins

PETR: PEristaltic mixed Tubular bioReactor

PKS: Polyketide synthases

PRM: Parallel reaction monitoring

PTMs: post-translational modifications

RIPP: Ribosomally synthesized and post-translationally modified peptides

SHIME: Simulator of the Human Intestinal Microbial Ecosystem

SIP: Protein-stable isotope probing

SRM: Selected reaction monitoring

SynComs: Synthetic microbial communities

T: Thiolation

Keywords

- Microbiome
- Microbial community
- Metaproteomics
- Non-ribosomal peptides
- Bacteriophages
- Microbial effectors

1. Introduction

The One Health concept is based on the close interconnection of human health, animals, agricultural and environmental ecosystems. The interactions between these dimensions are evident through examples such as zoonotic diseases, and the emerging spread of antibiotic resistance. A key factor linking human and animal health with environmental, agricultural, and biotechnological systems is their microbiomes. The exchange of species (including viruses) and genes between different microbiomes leads to (i) alterations in microbiome composition and function within a receiving system, (ii) the introduction of pathogenic species and genes, and (iii) the transfer of antibiotic resistance genes. In relation to the latter, globally, in 2019 alone, deaths related to antimicrobial resistance were estimated at 4.95 million [1]. Conversely, controlling microbiome composition holds transformative potential for healthcare and biotechnological applications. Identifying microbial effectors, such as virulence factors, toxins, antibiotics, non-ribosomal peptides (NRP)s, and viruses from various environments, is crucial for precisely managing microbial communities [2, 3].

Although the clear potential exists to leverage microbiome-based effectors in the future, a more comprehensive understanding of the mechanisms by which these effectors influence microbiomes and their broader functional impacts is still needed. Metaproteomics [4] provides the toolbox to identify and monitor microbial effectors. Metaproteomics has proven valuable across diverse applications, including for characterizing the impact of antibiotic therapy on human gut microbiomes [5], antibiotic resistance in animals [6] and their manure [7], exploring alternative gene coding in human gut bacteriophages (phages) [8], and identifying the human gut virome [9] and bacteriophage populations within anaerobic digesters [10].

Recent advancements in high-resolution mass spectrometers and overall progress in the metaproteomics field have made it possible to identify microbial effectors at higher sensitivity

and specificity. Further improvements in the metaproteomic workflow promise meaningful contribution to study of microbial effectors and their impact on microbiomes.

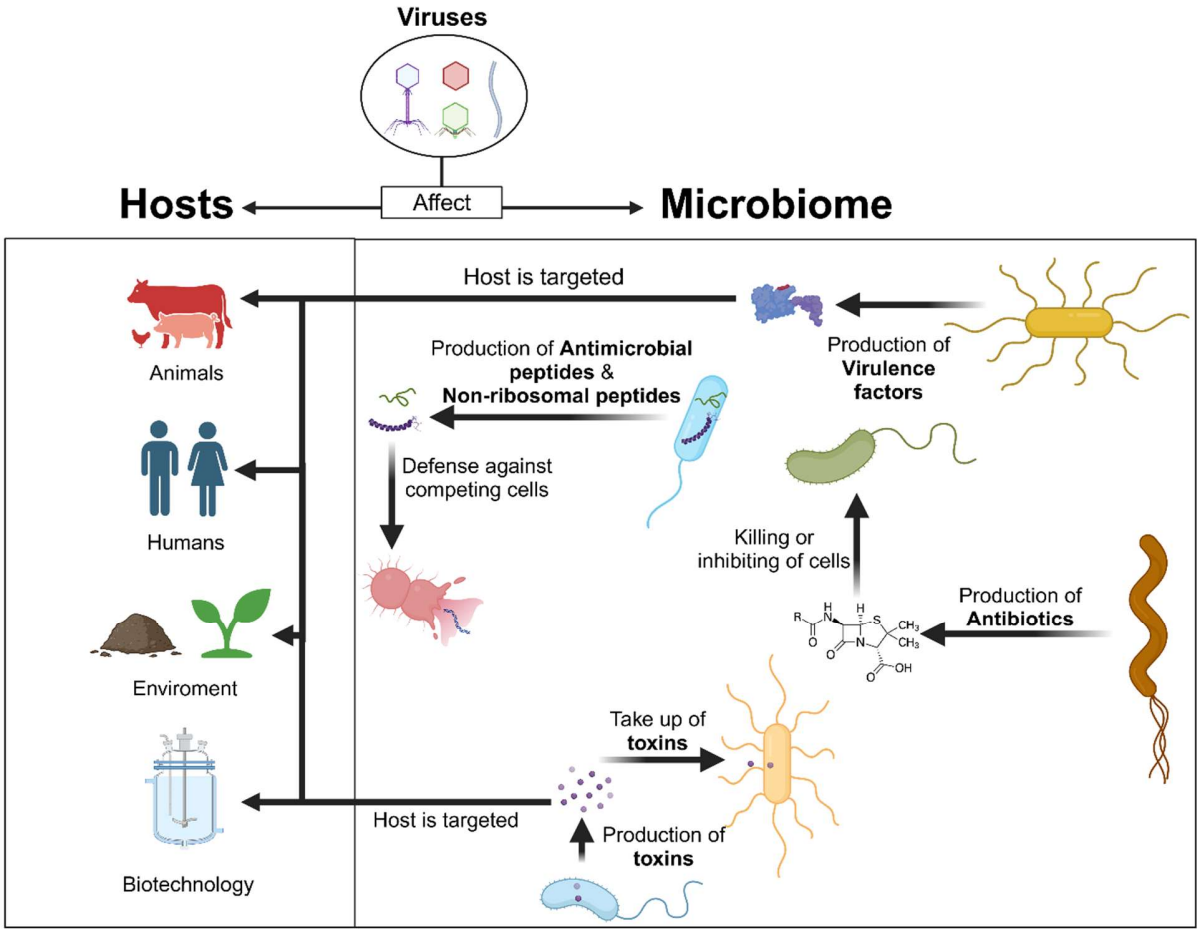
This current opinion **discusses how metaproteomics provides insights into the occurrence of microbial effectors and how these interact with microbiomes**. We focus on seven key challenges: **(i) identification and quantification of low-abundance microbial effectors through metaproteomics, (ii) identification of non-canonical peptides and proteins (e.g., NRPs), (iii) search databases for microbial effectors, (iv) taxonomic and functional annotation of microbial effectors, (v) mapping of microbial effectors to their hosts and targets, and (vi) identification strategies to explore their interactions.** To address these questions, we will first introduce the microbial effectors, followed by metaproteomics workflow and the required adaptations for microbial effector investigation, before focusing on their **application in microbiomes** and in the context of **One Health**.

2. Microbial effectors

Microbial effectors comprise diverse biomolecules used by microorganisms to combat with other species or modify their environment (Figure 1 and Table 1). For instance, microorganisms produce virulence factors (Chapter 2.1) to infect hosts, evade immune defenses, and cause disease. Among these virulence factors, toxins (Chapter 2.2) are particularly potent, as they can damage host cells, disrupt biological processes, or induce toxicity.

Additionally, several microbial species may synthesize antibiotics (Chapter 2.3) to inhibit the growth of competing species or eliminate them entirely. Furthermore, nearly all living organisms produce antimicrobial peptides (Chapter 2.4) as a defense mechanism against bacteria, viruses, fungi, and even tumor cells. These antimicrobial peptides can be classified into ribosomal peptides and NRPs.

Another class of microbial effectors are viruses, such as phages (Chapter 2.5), which function as mobile, self-replicating genetic elements. While metagenomics reveals the genetic potential of microbial communities to produce these effectors, metaproteomics offers a unique advantage by confirming their actual presence, expressed function, and the involvement of specific protein machinery in their synthesis.



158

159 **Figure 1:** Overview of the role of microbial effectors in microbiomes and their interaction with the

160 hosts.

Microbial effector	Definition	Structure	Producer	Target	Resistance
Virulence factor	Molecules or structures that enable pathogens to infect hosts, evade immune defenses, or cause disease.	Enzymes, toxins, adhesins	Bacteria, Archaea, Fungi, Viruses	Hosts including animals and plants	Usually slow, but fast adaptation by immune system
Toxins	Substances produced by organisms to damage host cells, disrupt biological functions, or cause toxicity.	Complex metabolites (may contain amino acids)	Bacteria, Archaea, Fungi, Viruses	Hosts including animals and plants	Usually slow, but fast adaptation by immune system
Antimicrobial peptide	Short peptides produced to defend against bacteria, viruses fungi and tumor cells	Canonical amino acids	All species	broad range against bacteria, virus, fungi or parasites	Rare
Non-ribosomal peptide	Peptides synthesized by non-ribosomal peptide synthetase (NRPS) enzymes, not ribosomes.	Peptide secondary metabolites	Bacteria, Archaea, Fungi	broad range against bacteria, virus, fungi or parasites	Rare
Antibiotics	Chemical compounds that specifically inhibit bacterial growth or kill bacteria.	Diverse small molecules (may contain amino acids)	Bacteria, Archaea, Fungi	Target microbial structures	Frequently
Viruses	Infectious agents consisting of nucleic acids	Nucleic acids (DNA/RNA) and proteins; sometimes lipids.	Self-reproducing in hosts	All domains	Usually slow, but fast adaptation by immune system

Table 1: Overview of microbial effectors

2.1 Virulence Factors

Virulence factors of microorganisms are described as the molecules that support the colonization of the host at the cellular level [11], and can be either secretory, associated with the membrane, or cytosolic. Virulence factors include, amongst others, proteins like toxins, adhesins, pili, proteases, hemolysins, or proteins from secretion systems, organized often as pathogenicity islands and regulatory pathways [12]. Their function is to allow nutrient acquisition, protect from the host's immune response, or enable adhesion, invasion, and intoxication of host cells [13]. The expression of these molecules is of outstanding importance for the health of humans, animals, and environmental systems. The transmission of virulence genes between or within species has the potential to create pathogen populations of enhanced virulence and favor the emergence of new diseases [14].

Transfer of DNA and thereby virulence factors is either accomplished by transduction through phages, conjugation through pili, or by uptake of naked, environmental DNA [15]. Environmental microbiomes from soils or glacial ice are reservoirs for clinically relevant virulence genes, and especially aquatic environments, such as wastewater treatment plants, are important drivers for the exchange and transmission of genes [16, 17]. Therefore, monitoring the presence of virulence factors in risky environments poses a crucial task in the One Health concept. The identification and quantification of virulence factors from microbial communities can be tentatively assessed using metagenomics and qPCR [18, 19], but the synthesis and secretion of virulence factors is prone to environmental influence, e.g., nutrient concentration or temperature [20, 21]. Therefore, proteomics and metaproteomics methods are more suitable to explore their synthesis and secretion. In practice, the production of several virulence factors was observed in different conditions with metaproteomics. For example, Graf et al. [22] identified several virulence factors, including staphylococcal leukocidin, in sputum samples of cystic fibrosis samples, and Messer et al. [23] observed protein virulence factors at marine plastic surfaces.

Thereby, metaproteomics is able to contribute to human health research and risk assessment, when dealing with environmental matrices. Furthermore, a number of studies used proteomics methods to quantify virulence factors, especially toxins [24].

2.2 Toxins

Toxins include a diverse set of bioactive compounds, including proteins, and are produced by microorganisms to establish a competitive advantage in their distinct environment and enhance survival [25]. They can be classified based on their biological effect on the target organism, e.g. into enterotoxins, cytotoxins, neurotoxins, leukotoxins, dermonecrotic toxins, and hemolytic toxins, and based on whether they are released to target cells (exotoxins) or cell-associated (endotoxins) [26]. Protein toxins are active in very low concentrations and can either act on the cell surface, by interfering with signal transduction, by damaging the membrane, or intracellularly, where they induce cell death, cytoskeleton alteration, or blockade of exocytosis [27].

Toxins produced by microorganisms can have severe or even life-threatening impacts on human and animal health, e.g., water- and food-borne intoxications, leading to diarrhea or septic shock syndrome, amongst others. Even probiotic strains are capable of producing hazardous toxins [28]. Transmission of toxins between environments is simple. For instance, toxins produced by microorganisms in an aquatic environment can accumulate in fish or seafood [29], or mycotoxins in contaminated food can accumulate in chicken tissue [30], which are then consumed by humans and animals impacting their health. Moreover, toxins produced by soil organisms can be washed into water and spread across a wide area.

Notably, xenobiotics derived from microbial toxins are currently being explored for their use in cancer treatment [31]. Furthermore, microbiomes can be used for the remediation of toxin-contaminated soils [32].

Therefore, the identification of microorganisms producing toxins and the detection of toxins in complex matrices (environmental samples or body fluids) is of importance for the One Health concept. Recently established databases condensing information about microbial toxins and antitoxins, like Toxinome [33], or tools for the prediction of toxin genes from (meta-) genomic data, like PathoFact [19], will facilitate the in-depth exploration of these microbial effectors. Metaproteomics was used to identify metabolic changes in the gut microbiome in response to toxins [34] and could help in the optimization of microbiome degradation capability for soil remediation purposes. Additionally, meta-(proteomics) can be used to verify the expression of protein toxins, or proteins involved in the biosynthesis and regulation of toxins [35].

2.3 Antimicrobials

Antimicrobials are capable of killing or inhibiting the growth of bacteria (antibiotics), or fungi (antifungals). The resistance of bacteria against antimicrobials represents a very significant public health concern [1]. Antibiotics are generally classified by their molecular targets, such as the bacterial cell wall (e.g., penicillins, cephalosporins), cell membranes, essential bacterial enzymes (e.g., rifamycins, quinolones, sulfonamides), or protein synthesis (e.g., macrolides, lincosamides, tetracyclines). Several databases are available for more detailed information on antimicrobials, including AntibioticDB [36], DrugBank [37], PubChem [38], and the ChEMBL [39] databases. Additionally, databases like CARD [40] and ResFinder [41] focus on collecting antimicrobial resistance genes.

A multitude of studies were performed to better understand the mode of action of antibiotics as well as the biological basis of drug resistance and to screen bacteria for the presence of antibiotic resistance mechanisms. Proteomics is key to elucidating mechanisms of actions of new antibiotics, but also to unravel cellular mechanisms of microbial adaptations to antibiotics – i.e. resistance to antibiotics [42–44].

Metaproteomics has been used to highlight changes in protein synthesis in specific bacteria resistant to antibiotics, i. a. to carbapenems [45]. While carbapenem resistance is genetically encoded in all the studied bacteria by genes encoding for the production of carbapenemase enzymes, more intricate proteome changes have been observed in the different genotypes ("New Delhi metallo-beta-lactamase", "Klebsiella pneumoniae *carbapenemase*", and "Imipenem-Hydrolyzing β-Lactamase"), showing that bacterial resistance can include more complex cellular responses than expected from the known enzymatic mechanisms of resistance. For example, an increase in the production of outer membrane proteins (ompA) under meropenem exposure was found mainly for *E. coli* of the "Imipenem-Hydrolyzing β-Lactamase" genotype. In contrast, the "New Delhi metallo-beta-lactamase" genotype rather showed an increase in the synthesis of DNA HU binding proteins and of the chaperonin protein complex GroEL/GroES alongside a higher increase in differential proteome abundance overall. A new application of proteomics addresses persister cells, i.e., subpopulations of cells in which antibiotics act at a slower rate. Recent proteomic studies showed that these persister cells have an overall reduced metabolic activity but also show adaptations that enable them to survive stress better, e.g., through SOS response [46].

2.4 Antimicrobial peptides/ non-ribosomal peptides

Antimicrobial peptides [47] are a subgroup of antimicrobials consisting of polypeptides of 12 to 50 amino acids, produced as part of the innate immune system response in all higher eukaryotes but also microorganisms. They play a key role in defending against other microbial species and may even target cancer cells or fungi [48]. Antimicrobial peptides are synthesized either through ribosomal pathways, utilizing canonical amino acids, often followed by extensive post-translational modification as in the case of Ribosomally synthesized and post-translationally modified peptides (RiPPs) [49, 50], or via nonribosomal peptide synthetases (NRPS), giant multifunctional enzymes found in bacteria, fungi but also higher eukaryotes [51].

While ribosomal antimicrobial peptides were mainly considered in the past to be linear, containing a few to several cationic amino acids, often showing membrane-interacting and pore-forming activity, the identification of several new classes of RiPPs in the past 15 years has changed this dogma. Today, several highly complex RiPPs are known, which are so heavily post-translationally modified including via additional ring systems, epimerizations, hydroxylations, acylations, and/or C- and N-methylations that they can hardly be recognized as being of ribosomal biosynthesis origin. However, the big difference to NRPS-derived peptides is that RiPPs initially rely only on the 20 canonical amino acids while in NRPS-derived peptides more than 400 different amino acids have been described. While most of these are incorporated into the peptide during the assembly-line-like mechanism, where they can also be modified including by C- and N-methylation, hydroxylation, oxidation, dehydratation, heterocyclization, acylation, or formylation. The resulting peptide can also be further modified post-NRPS by glycosylation, phosphorylation, sulfation, or deacylation. Furthermore, NRPS can occur in combination with polyketide synthases (PKS), forming so-called NRPS-PKS hybrid enzymes, based on the shared biochemical mechanism, whereby all biosynthesis intermediates are covalently bound to a peptidyl- or acyl-carrier protein or thiolation (T) domain, ensuring an efficient combination of amino acids with (further functionalized) malonyl- or acetyl-units. The resulting gamma- (elongation with one) or epsilon- (elongation with two PKS units) amino acids can add to the complexity of NRPS-derived peptides beyond what is possible through RiPPs.

The chemical diversity of NRPs and RiPPs with their several possible modifications, make their identification especially difficult, as they are highly stable against several proteases due to their D- or modified amino acids and cyclic structures. In most cases, where such complex peptides were identified, this was based on their good bioactivity (in the case of antibiotics) or after identification of the underlying biosynthetic gene cluster (BGC) encoding characteristic modifying enzymes (e.g., radical SAM-dependent enzymes), followed by heterologous production of the peptide. The identification of the classical linear AMPs is also very challenging, since although

287 they can be cleaved by proteases because of their linear structures, the presence of several
288 cationic amino acid residues often results in peptides too small for definitive identification. The
289 potential benefit of metaproteomics is that it enables quantifying even in communities the
290 abundance of the NRP-producing enzymes and thus predicting their presence and structure,
291 required for their targeted measurement.

292 2.5 Bacteriophages and archaeophages

293 Phages are viruses of bacteria (Figure 2) and archaea that selectively infect and rapidly kill cells
294 shaping microbial population dynamics [52].

295 Following the absorption on the cell surface, phages inject their nucleic acid into the bacterial
296 cytoplasm to immediately express early genes and manipulate DNA replication. They are divided
297 into two groups, temperate and virulent phages. While temperate phages can integrate their
298 genome into the genome of bacteria (prophage), chronic and virulent phages can only replicate
299 within the bacterium and lysate cells for the viral progeny release. Phage release from cells
300 occurs thanks to the production of holin and endolysin which target membranes and cell walls,
301 respectively [53]. Bacteriophages are also able to target biofilm-embedded bacteria, by
302 degrading extracellular matrix due to phage depolymerases, and to kill persister cells [54].

303 Phages are by their number the most abundant biological entities in the biosphere [55]. They
304 and their protein components, involved in the interaction with bacterial cells, have been identified
305 from different matrices (wastewater, soil, feces). Considering the small number of sequenced
306 phages, most phage proteins cannot be identified yet, due to the scarcity of primary sequenced
307 data. Thus, new phages should be isolated, and their genomes should be sequenced and
308 compared to the metaproteomic data to discover new molecules and new effectors.

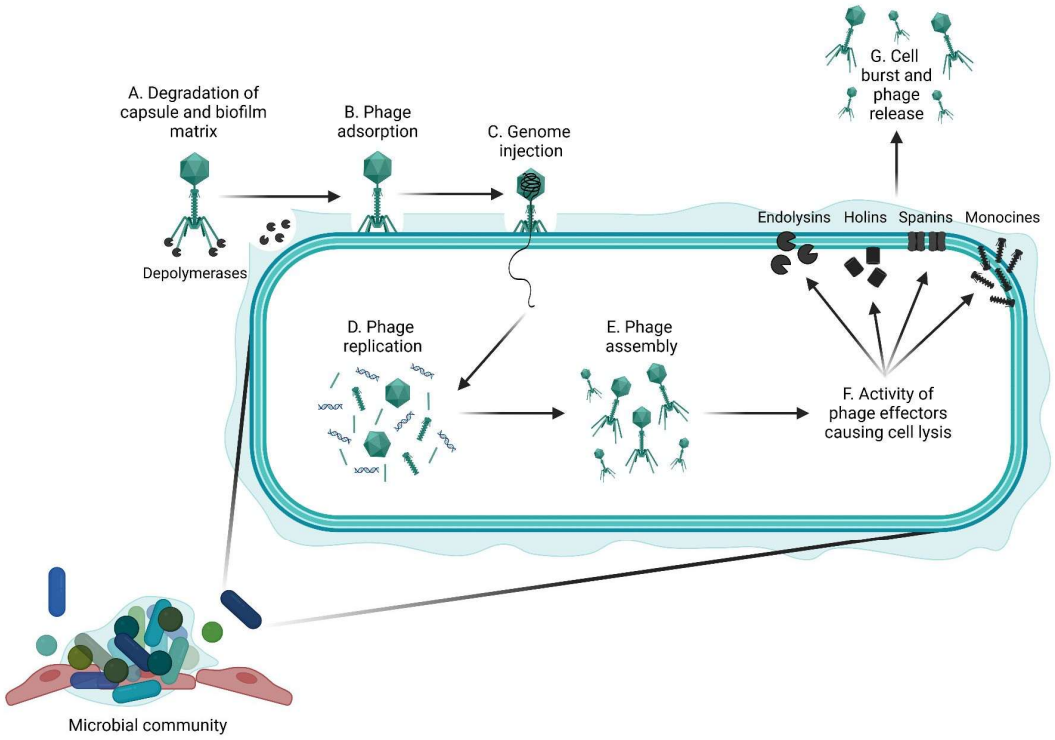
309 In the past, phages were very well studied for horizontal gene transfer and transduction
310 processes, especially for drug-resistant determinants and toxin genes. Now, they and their

proteins regain interest for their great antibacterial repertoire (including towards antibiotic-resistant strains), their narrow host range, and ease to engineer phages. All these features may potentially allow not only to use of phages as adjuvant of the antibiotics for the treatment of infections caused by drug-resistant bacteria [56], but also to modulate pathogenic, commensal, and pathobiont bacteria of the microbiome, impacting host physiology and immune system in a One Health approach.

One significant advantage of metaproteomics in phage research is its ability to confirm the expression of lysogenic phages and their associated proteins. For instance, *Corynebacterium diphtheriae* is a harmless bacterium unless it is infected by the prophage *corynephage β*, which encodes the diphtheria toxin responsible for causing diphtheria.

Metaproteomics also offers the opportunity to study host-phage interactions over time, including the role of host immune systems (e.g., CRISPR and other mechanisms [57] in combating phages [58].

However, a key challenge for analyzing the interactions in time rows lies in tracking mutations that alter protein sequences. These variations necessitate the incorporation of diverse protein isoforms into databases to ensure accurate protein identification and analysis.



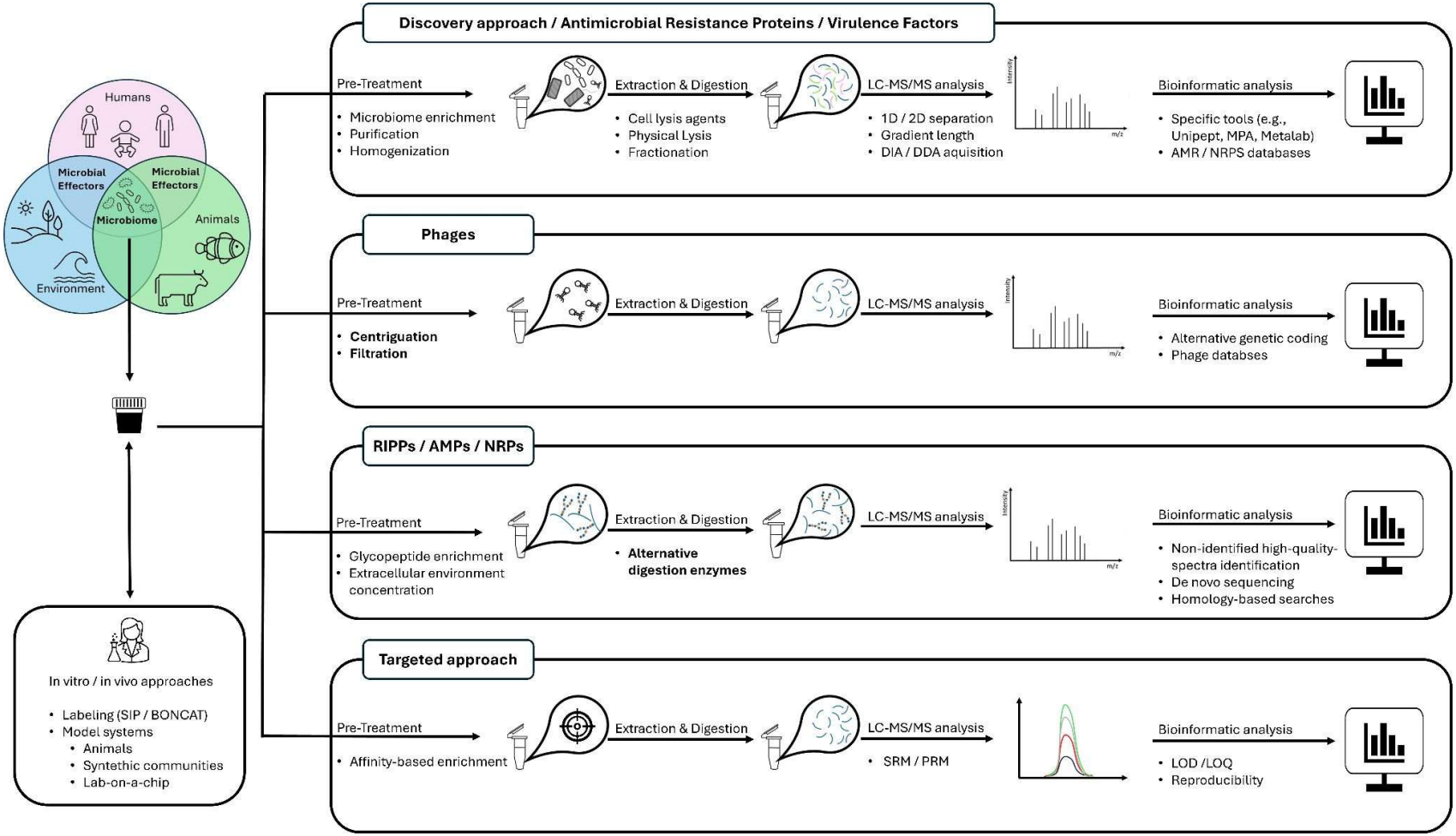
327

328 **Figure 2: Overview of phage infections.** The figure also shows the key molecules phages use

329 for cell membrane lysis.

3. Overview of metaproteomic workflows and requirements for the analysis of microbial effectors

Metaproteomics has achieved several major advances with respect to sample preparation (Chapter 3.10), mass spectrometry (Chapter 3.2), labeling (Chapter 3.3), bioinformatics (Chapter 3.4), multi-omics data integration (Chapter 3.5), and model systems (Chapter 3.6), facilitating the identification of microbial effectors and their impact on microbiomes. However, several challenges also have to be considered while studying microbial effectors using metaproteomics (Figure 3).



338

339 **Figure 3: Overview of metaproteomic workflow and key aspects that must be considered for studying microbial effectors**

340 **and their impact on microbiomes.** Abbreviations: RIPP: Ribosomally synthesized and post-translationally modified peptides, NRP:

341 non-ribosomal peptides, AMP: antimicrobial peptides. SRM: selected reaction monitoring, PRM: parallel reaction monitoring, LOD:

342 limit of detection, LOQ: limit of quantification, DIA: data-independent acquisition, DDA: data-dependent acquisition, SIP: stable

343 isotope probing, BONCAT: bioorthogonal non-canonical amino acid tagging.

3.1 Sample preparation for metaproteomics and considerations for the detection of microbial effectors

The first challenge for metaproteomic analysis is sampling, where biomass should be collected for metaproteome analysis and, if necessary, for metagenomic analysis simultaneously [59]. Furthermore, samples must represent the investigated microbiome, which can be challenging if the studied environment consists of heterogeneous biomass (e.g., feces, or soil). Microbial samples also require microbiome preservation if transportation before sample workup is required. Various strategies have been investigated recently to address this challenge. The subsequent metaproteomic sample preparation can be time-consuming because it often requires extensive extraction and purification steps, depending on the complexity of the sample matrix [60–66]. Common metaproteomics workflows begin with homogenization, cell lysis, protein extraction, and purification steps, followed by proteolytic digestion of proteins and subsequent mass spectrometry-based proteomic analysis. While numerous protocols are available in the literature describing metaproteomics, these have been commonly adapted to the specific microbial complexity, types of impurities, biomass content, and the scientific questions being addressed (e.g., deep proteome coverage, focus on central metabolic pathways, or identification of extracellular enzymes). For instance, microbiomes from water samples generally contain fewer impurities and instead require concentration of the microbial populations before cell lysis and protein extraction.

In contrast, soil or wastewater sludge samples contain high amounts of organic and phenolic compounds, polymeric substances, and inorganic compounds (including minerals), necessitating specialized purification and extraction protocols. Proteins can be adsorbed onto solid particles, such as clay minerals, which are often only partially reversible. The extracted proteome can be further fractionated to enable deeper proteome coverage and identification of low abundant microbial effectors, either before proteolytic digestion (e.g., via gel electrophoresis) or afterward

(e.g., ion exchange chromatography) [67]. However, direct analysis without fractionation and using 1D separation often provides sufficient coverage of the central metabolic pathways of the most abundant taxa [68]. This approach also saves significant time when analyzing large sample cohorts, such as in clinical studies. A particular challenge is posed by phages, hydrolytic enzymes or small peptides secreted into the extracellular environment. These require efficient extraction, concentration, and purification from often complex matrix components combined with enrichment of target molecules (e.g., antimicrobial glycopeptides) to achieve sensitive detection. Another challenge represents the detection of NRPs due to their complex and diverse structures. No general enrichment while sample preparation exists. However, possible NRPs can be predicted from genome information, and then targeted workflows for the detection of a subgroup of NRPs and their associated proteins can be developed [69–71].

3.2 Mass spectrometry for the detection of microbial effectors

High-resolution tandem mass spectrometry is a key technology for generating peptide tags in partial amino acid sequences, which confirm the presence of proteins and enable accurate quantification of their abundance. Peptides generated from proteins via trypsin proteolysis are typically separated by their hydrophobicity using reverse-phase chromatography and then introduced into the mass spectrometer through a nanospray interface directly connected to the outlet of the chromatography column. After their molecular weights are determined, peptides are either isolated (in Data-Dependent Acquisition mode, DDA) or pooled (in Data-Independent Acquisition mode, DIA) and subsequently fragmented. The molecular weights of the resulting fragments are then measured. The initial peptide mass data helps to narrow down potential sequence candidates, while the fragmentation patterns enable precise identification of the amino acid sequence.

Among a series of crucial parameters, those for the selection and fragmentation are paramount as they can significantly enhance the number of proteins identified and quantified. Compared to

traditional DDA setups, recent advances in DIA mode have shown increased sensitivity and broader protein coverage. Additionally, the recent introduction of a new generation of tandem mass spectrometers, specifically well adapted to address the complexity of metaproteomes, has significantly improved throughput and dynamic range [72, 73]. These instruments offer exceptional protein coverage, including of microbial effectors, and allow for deep characterization of dozens of samples or high-throughput monitoring of hundreds of samples in shorter time frames. Typically, 120,000 peptides can be identified and quantified within 30 minutes [73]. Interestingly, within this landscape 12 proteins were associated with toxins and 14 with phages. In contrast, a smaller number of proteins were linked to virulence (1 protein) and antibiotic-related function (2 proteins). The dynamic range observed in this dataset enables the identification and characterization of microorganisms comprising as little as 0.1% of the total biomass. Further fractionation of the samples can increase the dynamic range of this massive peptidome, allowing researchers to detect proteins from even low abundant organisms. Interestingly, NRPs can be characterized using the same experimental set-up, except that specific pre-enrichment should be carried out (e.g., for glycopeptides) [74]. Once the peptides are characterized, the monitoring of lowly abundant entities or cheap routine analysis of protein marker panels with targeted proteomics through selected reaction monitoring (SRM) mode can be straightforwardly developed for hundreds of samples. Fully harnessing the cutting-edge metaproteomics technology represents a major breakthrough for microbiome functional analysis, marking a transformative step forward in microbiome research [75].

3.3. Labeling approaches to study the impact of microbial effectors

Strategies to label proteins, such as protein-stable isotope probing (SIP), and click chemistry approaches, such as bioorthogonal non-canonical amino acid tagging (BONCAT), are additional tools to identify and quantify the impact of microbial effectors on the metabolic activities and nutrient fluxes of microbiomes. Newly synthesized proteins in actively growing cells are

detectable by incorporating labels, which cause a mass shift in the peptide spectra. Protein-SIP uses the incorporation of 2H, 13C, 15N, or 18O from respectively labeled substrates [76–78], whereas BONCAT is based on the incorporation of labeled amino acids tags such as l-homopropargylglycine (HPG) and l-azidohomoalanine (AHA) [79]. Although BONCAT is often combined with high-resolution microscopy and spectroscopy [80], the combination with mass spectrometric analyses was recently shown in studying the replication of phages during microbe-phage interactions [81], and the identification of effectors on bacterial pathogen infection [82, 83]. Thus, response mechanisms on various microbial effectors and the resulting physiological mechanisms can be identified with both methods. Bottlenecks such as restricted use of single labeled substrates in Protein-SIP and possible growth inhibitions by reactive substrate analogs must be considered and tested beforehand.

3.4 Bioinformatics

In (meta)proteomics, standard protein identification involves compiling a protein FASTA database from established repositories such as UniProtKB or sample-specific metagenomes. This step is followed by in-silico digestion and fragmentation to generate theoretical spectra, which are then compared to the experimentally measured spectra. Peptide and protein identifications are validated by calculating the false discovery rate (FDR). To enhance the number of valid peptide identifications in metaproteomics, applying *artificial error*-based rescoring of the FDR is particularly beneficial [84, 85]. This approach refines the accuracy of peptide detection and reduces false positives. Once valid peptides and their associated proteins are identified, taxonomic and functional annotations are assigned from the original databases or sequence-based comparison against other repositories.

Bioinformatic analysis for metaproteomics is challenging due to the usage of metagenomes for protein identification, the grouping of redundant protein identification from homologous proteins, and a comprehensive taxonomic and functional annotation. Therefore, several specific tools for

metaproteomics, like Unipept [86], MetaProteomeAnalyzer [87], and Metalab [88], were developed. These tools facilitate a better understanding of how microbial species contribute to resistance mechanisms across human, animal, and environmental health domains.

As outlined above, metagenomes derived from the same or similar environments, often supplemented with protein sequences from repositories (e.g., *Homo sapiens* entries in UniProt for human microbiome studies), are typically used as databases for protein identification. These metagenomes are usually assembled, genes are predicted (gene calling), and frequently, metagenome-assembled genomes (MAGs) are constructed to define sample-specific taxonomic units. For microbial effectors, tools such as PATHOFact (for virulence factors and antimicrobial resistance) [19], antiSMASH (for secondary metabolite biosynthesis) [89] or Macrel (for predicting antimicrobial peptides) [90] can be employed to annotate genes and their associated proteins (e.g., polyketide synthases or resistance genes). Furthermore, VirHostMatcher [91], phageAI [92], What the Phage [93] and PHASTEST [94] can be used to identify potential bacteriophage hosts and target structures.

A significant challenge in constructing metagenome databases for metaproteomics—particularly for cross-sample comparisons—is mapping genes and MAGs across different metagenomes or combined datasets. This task must accommodate the diversity of subspecies, mutations, and sequence variations while ensuring the database remains compact enough to support accurate FDR calculations. This issue is closely tied to proteogenomics, which combines genomic and proteomic data for deeper analysis [95]. In the context of metaproteomics, proteogenomics has demonstrated that phages can employ alternative genetic coding strategies [8]. Additionally, integrating a combined database or employing advanced tools enables a more detailed taxonomic and functional characterization of microbiomes from non-sequenced hosts [8].

Another major challenge in metaproteomics lies both in identifying non-tryptic peptides and in inferring homologous proteins across the vast diversity of microbial species. Due to the vast

diversity in microbial genomes and, thus, proteomes, traditional database-driven approaches often need help with incomplete or mismatched databases. Many proteins, especially those involved in resistance mechanisms, are poorly represented in existing databases. This challenge makes homologous protein identification a significant bottleneck, as the inability to match proteins accurately leads to gaps in understanding how microbial species express resistance.

RiPPs and NRPs can be chemically highly complex from various modifications and therefore often lack predictable structures, making them difficult to detect and identify using conventional proteomics workflows. As a result, they still need to be explored despite their crucial roles in AMR. A potential solution to identify RiPPs and NRPs is to screen for not-identified high-quality spectra (e.g., ranked by SPEQ [96]) and to leverage de novo sequencing [8] and homology-based searches [97]. These methods expand the search space beyond predefined peptide sequences and databases, allowing for the identification of non-tryptic peptides, homologous proteins, and NRPs. Moreover, integrating machine learning approaches and transfer learning can help refine peptide identification, making the detection of complex resistance mechanisms more accurate [98]. Another approach for NRPs or RiPPs for which the BGC has been identified is either heterologous production of the BGC or manipulation of the producer organism itself, followed by comparison of the respective expression, deletion or overproducing mutants with the wildtype strain. Although greatly facilitating the identification of compounds derived from these BGCs, this approach requires genetic tools for the desired peptide producer and knowledge about the biosynthesis pathway.

To better understand the effect of microbial effectors, identified proteins can be assigned to metabolic networks and used subsequently as input data for modeling microbiomes to study their effect on their taxonomic and functional composition [99].

3.5 Integration of further omics methods to enhance the identification of microbial effectors

Although metaproteomics is a powerful tool for studying microbiomes and microbial effectors, its potential is greatly enhanced by integrating additional omics approaches (Figure 4). Cytomics enables monitoring and sorting microbial subpopulations, providing insights into cell-specific behaviors. Furthermore, cytomics can measure cell viability [100], which is of great importance e.g., to assess the response of microbiomes to antimicrobials and could even provide the potential to evaluate community structure, diversity, and metabolite exchange in response to microbial effectors [101].

Metagenomics is essential for generating sample-specific databases used in protein identification, while predicted genes from metagenomic data serve as input for taxonomic and functional characterization (e.g., antibiotic production or resistance) using sequence alignment tools. Metagenomics typically offers higher resolution, allowing for more detailed insights into the taxonomic and functional composition of the microbiome. It also enables a unique opportunity to study operon structures and surrounding genes as well the encoding of microbial effectors in the genome or in mobile genetic elements [19]. Liquid chromatography (LC)-MS/MS-based metabolomics is another valuable method for studying microbiomes, allowing researchers to quantify metabolite pools and identify novel antibiotics and NRPs. Metabolomics provides an advantage in studying microbial effectors by enabling the screening of a broader range of chemical compounds through different LC and gas chromatography (GC) systems for prefractionation of the analytes [102, 103]. A key challenge in multi-omics research is integrating diverse data types, requiring standardized identifiers and ontologies, such as UniProt IDs [104] or KEGG ontologies [105]. An effective approach is aggregating all data into a graph-based structure or linking it to a knowledge graph. The benefit of knowledge graphs lies in their ability to integrate heterogeneous data, apply graph algorithms [106], and facilitate connections with

large language models, enabling improved data exploration and predictive analysis. Another strategy for multi-omics data integration is the calculation of correlation factors between the different omics features, which could be nicely visualized as co-occurrence networks [107].

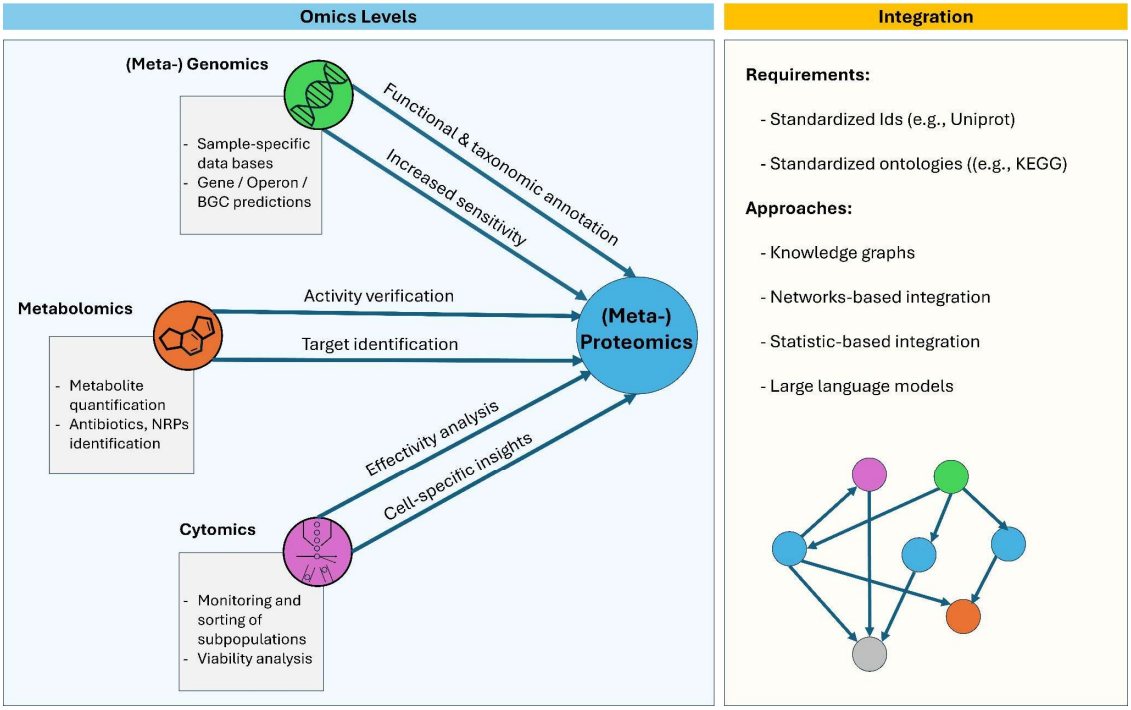


Figure 4: Strategies for combining metaproteomics with other omics tools.

3.6 Model systems to study microbial communities for the validation of microbial effectors

Researchers require controlled experimental models incorporating both in vitro and in vivo approaches to identify and validate novel microbial effectors and their impact on microbiomes. Synthetic microbial communities (SynComs), such as SIHUMIx, OligoMM, 14-SM, and other mock communities, represent gold-standard systems for studying microbial interactions and responses [108]. These models provide controlled environments that simulate natural microbial ecosystems, allowing for precise examination of effector molecules and their roles in community dynamics, signaling, and host interactions. SynComs offers a robust foundation for investigating

531 microbial functions within microbiomes by enabling direct observation of cause-and-effect
532 relationships while minimizing confounding variables. SynComs also enhances metaproteomic
533 research by using annotated genomes for each community member, which improves protein
534 identification accuracy. This genomic information enables detailed insights into strain-level
535 interactions, often unachievable in natural microbiomes due to similar protein sequences.
536 Additionally, SynComs allow for the study of low-abundance species, such as keystone taxa
537 [109, 110], which play essential roles within the microbiome but are often undetectable in
538 conventional metaproteomic analyses. *In vivo* mouse models replicate human physiological
539 responses for host-relevant microbial studies, providing insights into microbial-host interactions
540 and discovering microbial effectors pertinent to disease. Emerging "lab-on-a-chip" microfluidic
541 platforms complement these models by allowing precise control and monitoring microbial
542 communities in high-throughput formats, enhancing our understanding of microbial dynamics
543 under controlled conditions [111]. Fermentation systems, including the Simulator of the Human
544 Intestinal Microbial Ecosystem (SHIME) [112] and the PEristaltic mixed Tubular bioReactor
545 (PETR) [113], simulate gastrointestinal conditions, supporting long-term studies of microbial
546 fermentation and gut ecology. Of particular interest for the fermentation are systems that enable
547 co-cultures of microorganisms and human cells, such as HuMiX [114] or gut-on-chip [115], to
548 ascertain the effect of different microbiota-expressed effectors on the human cells to ascertain
549 the effect of different microbiota-expressed effectors on the human cells.

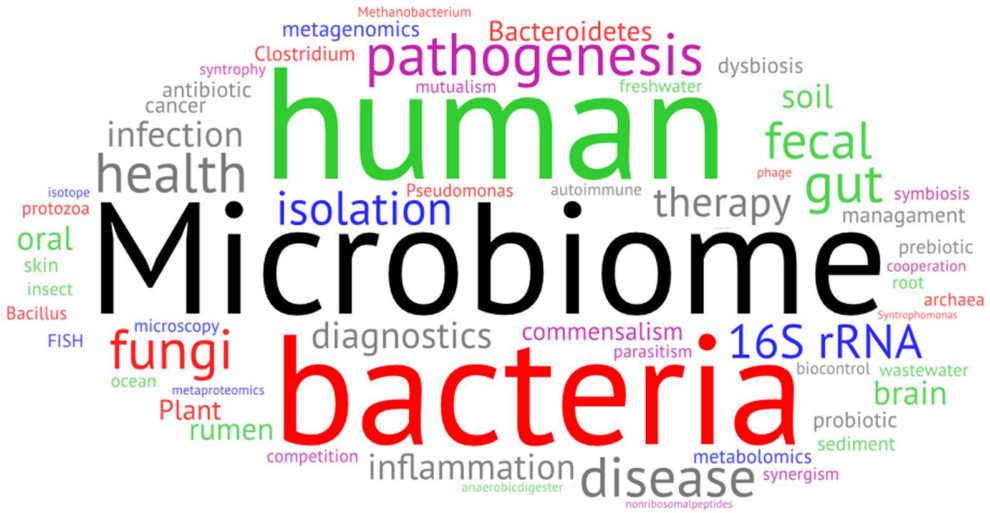
550 Together, these models offer a comprehensive toolkit for assessing microbial effects on host
551 health, advancing our understanding of microbial communities in health and disease.

4. Application of metaproteomics for studying microbial effectors and microbiomes in the One Health context

To evaluate the potential for identifying and studying microbial effectors in the One Health framework, we assessed the current status of the field of metaproteomics in this context (Chapter 4.1), followed by a detailed characterization of usage of microbiomes in the clinical (Chapter 4.2) and non-clinical context (Chapter 4.3).

4.1 Relevance of metaproteomics and microbial effectors in the context of microbiome research and One Health

As a summary of the current research status about microbiomes, metaproteomics, One Health, and microbial effectors, we created a word cloud based on the abundance of the keyword in combination with “microbiome” in Pubmed NCBI (Figure 5) presenting the relevance of the terms based on the text size.



564

565 **Figure 5: Word Cloud showing the abundance of keywords in combination with the**
566 **'Microbiome' in Pubmed NCBI.** As abundance values the rounded logarithm with the base of 2
567 of hits were used and submitted to https://wordart.com/create. For 'Microbiome' the number
568 without any combination was used for calculation.) The word cloud displays different aspects of
569 microbiome research: (i) sources of microbiomes (green), (ii) interactions (purple), (iii) involved
570 taxa (red), (iv) applied experimental approaches (blue), and (vi) societal effects and recent or
571 future applications (grey).

572 The word cloud highlights "microbiome" as a central term, primarily associated with human and
573 animal habitats like the gut or skin, while also representing ecosystems across the biosphere.
574 Advances in high-throughput sequencing have shifted microbiology from isolated studies to a
575 holistic view of microbial communities. While research often emphasizes pathogenic effects,
576 increasing evidence underscores the health-promoting roles of microorganisms and their
577 complex interactions, essential for microbiome stability.

578 Terms like "human," "gut," and "pathogen" reflect the focus on human disease, while
579 "commensalism," "mutualism," and "syntrophy" highlight cooperative interactions, such as
580 metabolite sharing, critical for microbiome functions. Conversely, terms like "phage," "antibiotic,"

and "nonribosomal peptides" point to regulatory mechanisms preventing dysbiosis and supporting microbial diversity.

Less frequent terms like "oceans" and "roots" indicate underexplored environmental microbiomes, aligning with the "One Health" approach. The prominence of "16S rRNA" and "cultivation" highlights a continued focus on community structure and species isolation, while the underrepresentation of "metaproteomics" suggests an experimental gap. Finally, terms like "probiotic," "prebiotic," and "biocontrol" reflect growing recognition of microbiomes' societal and environmental benefits.

4.2 Usage of microbiomes in the clinical context and potential for metaproteomics and microbial effectors

Currently, over 2,400 clinical studies are investigating the microbiome's relationship with various health factors (https://clinicaltrials.gov/). Of these, over 1,000 focus on the microbiome's role in 70 diseases, including autoimmune disorders, cancer, cardiovascular, digestive, and metabolic diseases (Figure 6). These efforts reflect the growing recognition of the microbiome as a key factor influencing disease diagnosis, prognosis, and treatment response.

The importance of the microbiome extends beyond human microbiome and health to encompass the interconnected animal and environmental dimensions of the One Health approach. Microbial communities in animals and the environment actively shape human microbiome composition and functionality. Through direct exposure, shared ecosystems, and environmental reservoirs, these interactions influence the microbiome's clinical impact. For instance, zoonotic pathogens or antimicrobial resistance genes are often mediated by microbial exchanges between humans, animals, and their habitats, demonstrating the profound interplay within these domains.

The high interest in the microbiome is particularly pertinent in diseases where immune and inflammatory mechanisms play a central role, as the microbiome may modulate both immune-

suppressive and stimulatory pathways [116]. Furthermore, microbiome-host interactions extend beyond individual organs that are vital for maintaining homeostasis and influencing disease progression [117]. Therefore, the interrelationship between microbial communities colonizing different human surfaces provides the advantage of collecting highly informative profiles from more accessible microbiome samples in clinical contexts where pathology mostly affects less reachable organs.

While different diseases are linked with different interactions between the microbiome and the host, key areas of investigation remain. These include understanding the microbiome's treatment response, metabolic consequences, underlying molecular pathways, and identifying microbiome components that enhance clinical status. To showcase the need of microbiome studies and the microbial effectors we showcase the following four clinical use cases:

(i.) Recent studies have shown that microbial proteins, which accumulate under specific conditions, as well as their sequence diversity, structural features, and post-translational modifications (PTMs) like acetylations, deaminations, hydroxylations, methylations, nitrosylations, oxidations, and phosphorylations, are critical for priming immune cells effectively [88]. Understanding these variations in proteins, whose nature can be revealed through metaproteome analysis, combined with advanced computational methods such as protein structure prediction [118] and all-atom molecular dynamics (MD) simulations [119], provides insights into the role of microbial proteins in immune regulation. This insight could potentially guide the development of targeted therapeutic strategies

(ii.) Cancer and infectious diseases, including HIV, underscore the importance of identifying specific microbiome-derived proteins that can boost immune function and mitigate inflammation while simultaneously managing antibiotic resistance in frequently hospitalized individuals [120]. Such patients often require repeated antibiotic cycles, which further complicates treatment by promoting resistance.

(iii.) Furthermore, the understanding of disease also requires a more holistic view. For example, *Helicobacter pylor*i was until recently considered the strongest risk factor for the development of gastric cancer, which is the fifth most common cancer worldwide. However, recent advances in metagenomics and metaproteomics techniques demonstrated changes in the complete microbiome during gastric carcinogenesis rather than that of single microbes. Hypochlorhydria, a state of low hydrochloric acid levels that affect the stomach's ability to digest and absorb proteins, induces changes in the complete microbiome (reducing diversity and abundance of commensal bacteria and promoting overgrowth of pathogenic and carcinogenic species) that might have a direct linked with gastric cancer [121]. This is further enhanced by the prolonged use of proton pump inhibitors which are widely used medications [122]. Thus, the risk of cancer could be identified by characterizing microbiome alterations in gastric juices and/or feces of patients. Monitoring microbial alterations could also help physicians and healthcare professionals assess the risks and benefits of using medications such as proton pump inhibitors, monitor medical care protocols, and optimize treatments for high-risk patients.

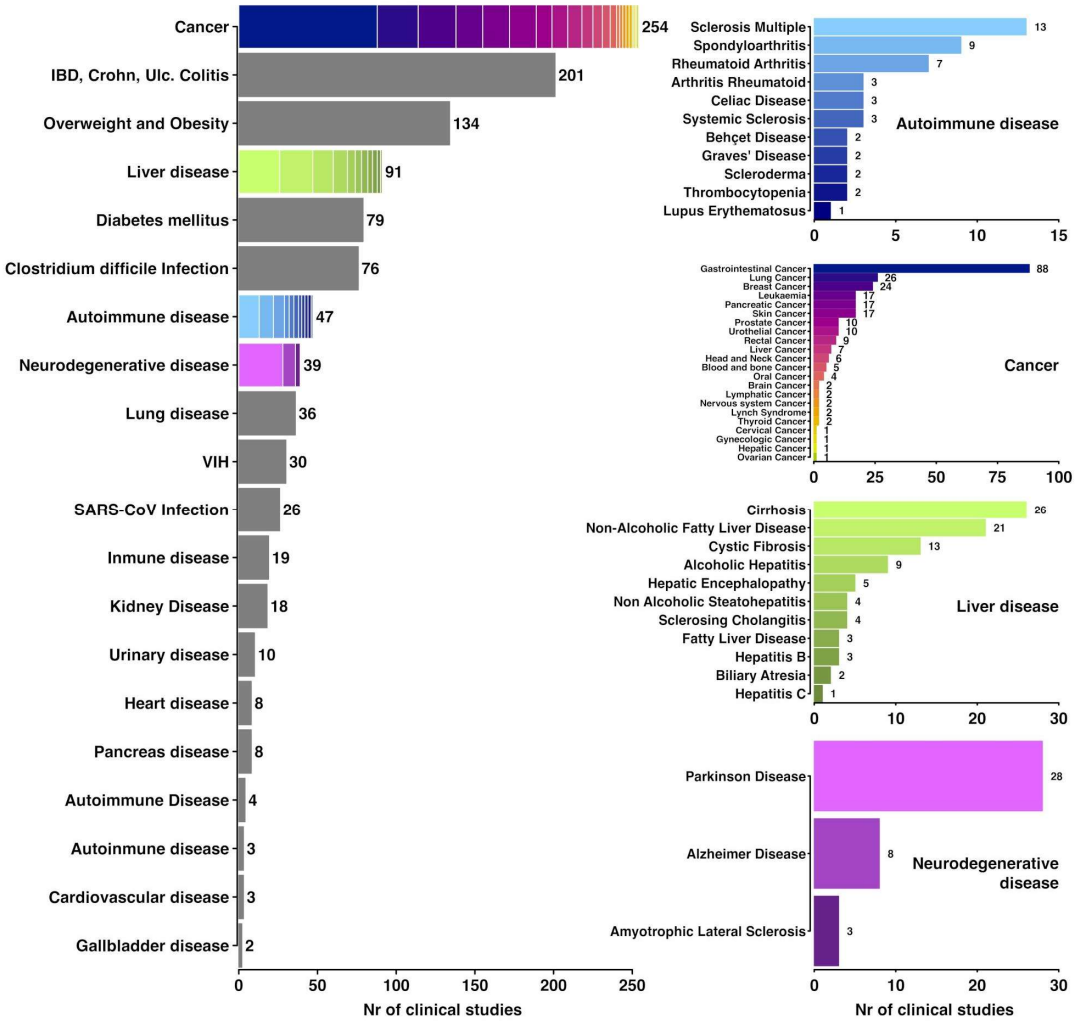
(iv.) Preclinical and clinical trials suggest that the alterations in the gut microbiome are also linked with toxicities induced by chemotherapies [123], and immunotherapies [124]. It has also been suggested that modulation of the gut microbiome before and during chemotherapy in cancer patients could reduce the occurrence of adverse events and improve the effectiveness of treatments [125]. Recent studies also suggested that the gut microbiome, available *via* fecal material, constitutes a promising source of biomarkers to predict and monitor treatment outcomes and potentially related adverse events [126]. Furthermore, tongue swab metaproteomics has for instance enhanced our understanding of the mechanism behind specific tongue coating formation and its potential role as an indicator of gastric cancer [127].

In sum to these examples, clinical needs for microbiome research include (i.) accurate and timely diagnosis of microbiome functional alterations; (ii.) monitoring the disturbances in

655 microbial communities and their components (genes, transcripts, proteins, metabolites) triggered
656 by clinical protocols; (iii.) evidence-based therapy to modulate the microbiome and regain its
657 homeostasis; and (iv.) the identification of novel microbial effectors for targeted microbiome
658 management. Metaproteomics can provide solutions to these critical domains, profiling how
659 clinical traits shape the microbiome, identifying microbial effectors involved in variations of the
660 microbial community structure and functions, and monitoring the outcome of experimental
661 clinical protocols based on microbial modulators, including antimicrobial peptides and phages.
662 For instance, the dynamics of the gut virome (i.e., phages) are strongly integrated with those of
663 the gut bacteriome [128] and may impact disease together. Host environmental stimuli can play
664 a role in the balance between inhibition or induction of viral replication by lysogenic phages.
665 Metaproteomic studies can highlight the relative abundance variations of viral and bacterial
666 proteins, establishing the possible correlation and leading to suggestions on the potential of
667 specific phages to modulate the abundance and the protein functions in specific bacterial taxa.
668 Further, metaproteomics can hold the promise to thoroughly investigate (i.) the potential use of
669 phages as antibacterial agents in several clinical conditions, (ii.) to monitor the functional
670 dynamics of their bactericidal effect and (iii.) to monitor therapy response due to the correlation
671 of the bacteria and host proteins. Thereby the unique dimension provided by metaproteomics is
672 that it monitors the functional products of gene expression and therefore reflects the functional
673 dimension of a microbiome.

674 Efforts to address these clinical needs, it is necessary to integrate metaproteomics-based
675 analysis, bioinformatics, machine learning, and computational approaches to fully realize the
676 potential of the microbiome in enhancing clinical outcomes. For example, machine learning
677 models have been used to predict patient responses to microbiome-targeted therapies by
678 analyzing complex microbial community data, offering insights into personalized treatment
679 strategies [129].

In conclusion, the active microbiome fraction that can be measured by state-of-the-art metaproteomics experiments holds immense potential in the clinical context, with significant strides being made in understanding its role across various diseases [130]. Despite promising results, unraveling the causal contributions of microbiome traits to host biology and translating microbiome-based biomarkers into clinical practice remains significantly challenging. Non-model gut commensals encode many traits, and the sample size in omics studies often lacks the power needed for functional monitoring of the microbiome in clinical contexts. Additionally, advancing the microbiome's clinical applications will require further well-designed clinical trials and the integration of multidisciplinary approaches. Metaproteomics, complemented with bioinformatics and machine learning, represents an essential toolbox for untangling the complex interactions between microbiome components and host health, thereby paving the way for novel therapeutic strategies. In this respect first assays for studying the drug response of microbiomes combined with metaproteomics were tested [131], highlighting the potential of pharmacomicrobiomics.



693

694 **Figure 6. Summary of over 2,400 clinical studies exploring the microbiome's role in**
695 **health, distributed by its relation to 70 diseases.** Cases where multiple types of cancers or
696 diseases are studied are detailed on the right. The figure was created using R programming
697 language v.4.4.1, with core-base functions and in-house scripts.

4.3 Non-clinical microbiomes, metaproteomics, and microbial effectors in the One Health context

The connection between microbiomes across environmental, agricultural, and biotechnological domains and the One Health framework extends beyond serving as a reservoir of novel microbial effectors for human therapeutics. Many microbial effectors developed for human use can also benefit pet and livestock health.

In plant agriculture, microbial antimicrobials have several promising applications. For instance, cyclic lipopeptides produced by *Pseudomonas* strains can act as natural insecticides, effectively targeting insect larvae. Phages may be employed to combat plant pathogens like *Pectobacterium atrosepticum*, which causes potato soft rot [132], while seed coatings with antimicrobial agents offer protective benefits [133]. Additionally, antimicrobials can enhance food safety by reducing microbial contamination in produce and food supply chains [132, 134].

In environmental management, cyanophages could be harnessed to mitigate harmful algal blooms, thus safeguarding aquatic ecosystems like oceans, seas, and lakes [135]. Meanwhile, antibiotics and other antimicrobials might stimulate the growth of contaminant-degrading microbes in nutrient-limited environments, such as certain groundwater systems, aiding in bioremediation [136].

Within biotechnological applications, phages offer a targeted approach to controlling filamentous bacteria, including *Microthrix parvicella* and *Nocardia* species, which cause foaming issues in wastewater treatment plants [137]. Phages also have emerging applications as structural components in nanomaterials, presenting exciting opportunities in materials science [138].

While microbial effectors offer significant potential, it is crucial to consider potential unintended impacts on microbiomes, such as effects on non-target species and the development of resistance mechanisms. Additionally, stressors—including those from human activities—can

accelerate the release of phages within microbiomes, leading to self-amplifying cycles and other stress responses. For instance, exposure to pesticides has been shown to increase bacterial antibiotic resistance by activating efflux pumps, reducing outer membrane permeability, and inducing gene mutations [139].

5. Future potential of microbial effectors and metaproteomics in the One Health context (under construction)

As the opposite of the spread of pathogenic species across different hosts (zoonoses), using microbial effectors from diverse environments holds transformative potential for treating diseases as well as monitoring and controlling microbiomes in biotechnological systems. Just as the "golden age" of antibiotic discovery opened new frontiers in medicine, broader screening of microbial effectors now offers the potential to treat pathogen-associated diseases. One advantage of many microbial effectors is their ability to target specific microorganisms, allowing for more precise treatment options and enhanced microbiome control. However, this specificity also increases the risk of resistance mechanisms, such as escape mutations, which require continuous adaptation of microbial effectors to maintain efficacy.

In this context, metaproteomics plays a crucial role in advancing microbial effector research through two key contributions: (i.): Metaproteomics aids in identifying new antimicrobial effectors by enabling the selection of microbiomes with a high abundance of proteins associated with microbial effectors. Therefore, it should be applied across various environments to maximize the number of novel identified microbial effectors. Environments under selective pressure—such as reptile saliva, amphibian skin, hospital wastewater, and livestock enclosures—are particularly interesting for discovering novel compounds. (ii.) Because metaproteomics allows for the analysis of expressed proteins and phages, it is an ideal tool for examining the effects of microbial effectors.

In addition to the focus on microbial effectors, metaproteomics is an important method for monitoring microbiomes and implementing strategies for controlling them. As shown in several studies metaproteome is very predictive of temporal variations within microbiomes [140] and hence a crucial indicator of how/when one can intervene for example by altering substrate availability, physico-chemical parameters or by CRISPR-based community-wide engineering/editing [141, 142]. In clinical settings, metaproteomics provides valuable insights into the current state of a patient's microbiome, facilitating the identification of dysbiosis and the prediction of drug-microbiome interactions. This approach also enables continuous monitoring of therapeutic progress.

Beyond microbiome monitoring, metaproteomics data can inform microbial abundance estimates, which are essential for developing control algorithms to manage microbiomes effectively [99]. Using these algorithms, control variables such as nutrient supply, process parameters (e.g., temperature), or the introduction of microbial effectors can be adjusted to achieve the desired microbiome functionality. This concept of microbiome management is indeed in analogy to animal gastrointestinal tracts regulating their microbiomes. Closing the gap to the One Health concept, the most comprehensively studied system for microbiome management is the human gut, whose control mechanism (e.g. antimicrobial peptides) could be applied also to manage microbiomes in environment and biotechnological applications.

Another application of metaproteomics lies in "pandemic preparedness" within a One Health framework. For example, tracking pathogen concentrations in wastewater treatment plants, as seen with COVID-19, enables early detection of disease outbreaks before hospitals or government agencies identify them. Sequencing methods with low detection limits (enabled by gene amplification) are primarily used for such monitoring. As demonstrated with the selective enrichment of COVID-19 peptides using advanced mass spectrometry (LC-MS/MS), proteomic methods could also play a valuable role in pandemic preparedness, offering timely and reliable

pathogen monitoring [143], enabling to monitoring the actual expression of the pathogens. Matching to this, tools such as PEPGM [144] have been developed to identify viruses from LC-MS/MS-based peptide identifications. As a result, the holistic tracking of wastewater through metaproteomics emerges as a powerful approach for detecting emerging pathogens and microbiome dysbioses associated with human diseases.

Acknowledgment

M. Ferrer thanks Rafael Bargiela for retrieving clinical studies exploring the microbiome's role in health from https://clinicaltrials.gov/.

Funding

Manuel Ferrer acknowledges funding from the Agencia Estatal de Investigación (AEI) (Digital Object Identifier MICIU/AEI/10.13039/501100011033), FEDER, the European Union, the NextGenerationEU/PRTR, and CSIC (COCREA program), under grants PID2023-153370OB-I00 and INMUNOBIOTA. Tim Van Den Bossche acknowledges funding from the Research Foundation Flanders (FWO) [1286824N]. Lucia Grenga and Jean Armengaud express their gratitude to the Région Occitanie (DeepMicro grant) and the Agence Nationale de la Recherche (ANR-20-CE34-0012 and ANR-24-INBS-0015).

AE

Grammarly and ChatGPT were used for spelling and language improvements

Conflicts of Interest:

Not applicable.

Author contributions:

Conceptualization: R.H., P.W.

Project administration: R.H.

Supervision: R.H., P.W

writing—original draft: ALL

writing—review and editing: ALL

All authors have read and agreed to the published version of the manuscript.

References

1. Naghavi M, Vollset SE, Ikuta KS, Swetschinski LR, Gray AP, Wool EE, et al. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. The Lancet. 2024;404:1199–226. doi:10.1016/S0140-6736(24)01867-1.
2. Cycoń M, Mrozik A, Piotrowska-Seget Z. Antibiotics in the Soil Environment-Degradation and Their Impact on Microbial Activity and Diversity. Front Microbiol. 2019;10:338. doi:10.3389/fmicb.2019.00338.
3. Joerger RD. Alternatives to antibiotics: bacteriocins, antimicrobial peptides and bacteriophages. Poult Sci. 2003;82:640–7. doi:10.1093/ps/82.4.640.
4. Wilmes P, Bond PL. The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. Environ Microbiol. 2004;6:911–20. doi:10.1111/j.1462-2920.2004.00687.x.
5. Heyer R, Schallert K, Siewert C, Kohrs F, Greve J, Maus I, et al. Metaproteome analysis reveals that syntrophy, competition, and phage-host interaction shape microbial communities in biogas plants. Microbiome. 2019;7:69. doi:10.1186/s40168-019-0673-y.
6. Buthasane P, Roytrakul S, Phaonakrop N, Tunsagool P, Buthasane W, Am-In N, Suriyaphol G. Metaproteomic Analysis of Gut Resistome in the Cecal Microbiota of Fattening Pigs Raised without Antibiotics. Microbiol Spectr. 2023;11:e0222323. doi:10.1128/spectrum.02223-23.
7. Zhang L, Li L, Sha G, Liu C, Wang Z, Wang L. Aerobic composting as an effective cow manure management strategy for reducing the dissemination of antibiotic resistance genes: An integrated meta-omics study. J Hazard Mater. 2020;386:121895. doi:10.1016/j.jhazmat.2019.121895.
8. Peters SL, Borges AL, Giannone RJ, Morowitz MJ, Banfield JF, Hettich RL. Experimental validation that human microbiome phages use alternative genetic coding. Nat Commun. 2022;13:5710. doi:10.1038/s41467-022-32979-6.

9. Lozano C, Pible O, Eschlimann M, Giraud M, Debroas S, Gaillard J-C, et al. Universal Identification of Pathogenic Viruses by Liquid Chromatography Coupled with Tandem Mass Spectrometry Proteotyping. Mol Cell Proteomics. 2024;23:100822. doi:10.1016/j.mcpro.2024.100822.
10. Pérez-Cobas AE, Gosalbes MJ, Friedrichs A, Knecht H, Artacho A, Eismann K, et al. Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. Gut. 2013;62:1591–601. doi:10.1136/gutjnl-2012-303184.
11. Sharma AK, Dhasmana N, Dubey N, Kumar N, Gangwal A, Gupta M, Singh Y. Bacterial Virulence Factors: Secreted for Survival. Indian J Microbiol. 2017;57:1–10. doi:10.1007/s12088-016-0625-1.
12. Lee YM, Almqvist F, Hultgren SJ. Targeting virulence for antimicrobial chemotherapy. Curr Opin Pharmacol. 2003;3:513–9. doi:10.1016/j.coph.2003.04.001.
13. Kreibich S, Hardt W-D. Experimental approaches to phenotypic diversity in infection. Curr Opin Microbiol. 2015;27:25–36. doi:10.1016/j.mib.2015.06.007.
14. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, et al. Emergence of a new disease as a result of interspecific virulence gene transfer. Nat Genet. 2006;38:953–6. doi:10.1038/ng1839.
15. Gyles C, Boerlin P. Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease. Vet Pathol. 2014;51:328–40. doi:10.1177/0300985813511131.
16. Pérez-Etayo L, González D, Vitas AI. The Aquatic Ecosystem, a Good Environment for the Horizontal Transfer of Antimicrobial Resistance and Virulence-Associated Factors Among Extended Spectrum β-lactamases Producing E. coli. Microorganisms 2020. doi:10.3390/microorganisms8040568.
17. Søborg DA, Hendriksen NB, Kilian M, Christensen JH, Kroer N. Bacterial Human Virulence Genes across Diverse Habitats As Assessed by In silico Analysis of Environmental Metagenomes. Front Microbiol. 2016;7:1712. doi:10.3389/fmicb.2016.01712.

18. Hoorzook KB, Barnard TG. Absolute quantification of E. coli virulence and housekeeping genes to determine pathogen loads in enumerated environmental samples. PLoS One. 2021;16:e0260082. doi:10.1371/journal.pone.0260082.
19. Nies L de, Lopes S, Busi SB, Galata V, Heintz-Buschart A, Laczny CC, et al. PathoFact: a pipeline for the prediction of virulence factors and antimicrobial resistance genes in metagenomic data. Microbiome. 2021;9:49. doi:10.1186/s40168-020-00993-9.
20. Hosseini H, Mahmoudi R, Pakbin B, Manafi L, Hosseini S, Pilevar Z, Brück WM. Effects of intrinsic and extrinsic growth factors on virulence gene expression of foodborne pathogens in vitro and in food model systems; a review. Food Sci Nutr. 2024;12:6093–107. doi:10.1002/fsn3.4281.
21. Penttinen R, Kinnula H, Lipponen A, Bamford JKH, Sundberg L-R. High Nutrient Concentration Can Induce Virulence Factor Expression and Cause Higher Virulence in an Environmentally Transmitted Pathogen. Microb Ecol. 2016;72:955–64. doi:10.1007/s00248-016-0781-1.
22. Graf AC, Striesow J, Pané-Farré J, Sura T, Wurster M, Lalk M, et al. An Innovative Protocol for Metaproteomic Analyses of Microbial Pathogens in Cystic Fibrosis Sputum. Front Cell Infect Microbiol. 2021;11:724569. doi:10.3389/fcimb.2021.724569.
23. Messer LF, Lee CE, Wattiez R, Matallana-Surget S. Novel functional insights into the microbiome inhabiting marine plastic debris: critical considerations to counteract the challenges of thin biofilms using multi-omics and comparative metaproteomics. Microbiome. 2024;12:36. doi:10.1186/s40168-024-01751-x.
24. Duracova M, Klimentova J, Fucikova A, Dresler J. Proteomic Methods of Detection and Quantification of Protein Toxins. Toxins (Basel) 2018. doi:10.3390/toxins10030099.
25. Nwaji AR, Arieri O, Anyang AS, Nguedia K, Abiade EB, Forcados GE, et al. Natural toxins and One Health: a review. Sci One Health. 2022;1:100013. doi:10.1016/j.soh.2023.100013.

26. Guerrant RL. Microbial toxins and diarrhoeal diseases: introduction and overview. Ciba Found Symp. 1985;112:1–13. doi:10.1002/9780470720936.ch1.
27. Popoff MR. Overview of Bacterial Protein Toxins from Pathogenic Bacteria: Mode of Action and Insights into Evolution. Toxins (Basel) 2024. doi:10.3390/toxins16040182.
28. Cui Y, Wang S, Ding S, Shen J, Zhu K. Toxins and mobile antimicrobial resistance genes in Bacillus probiotics constitute a potential risk for One Health. J Hazard Mater. 2020;382:121266. doi:10.1016/j.jhazmat.2019.121266.
29. Moreira C, Azevedo J, Antunes A, Vasconcelos V. Cylindrospermopsin: occurrence, methods of detection and toxicology. J Appl Microbiol. 2013;114:605–20. doi:10.1111/jam.12048.
30. Münke M, Harbers K, Jaenisch R, Francke U. Chromosomal mapping of four different integration sites of Moloney murine leukemia virus including the locus for alpha 1(I) collagen in mouse. Cytogenet Cell Genet. 1986;43:140–9. doi:10.1159/000132312.
31. Sharma PC, Sharma D, Sharma A, Bhagat M, Ola M, Thakur VK, et al. Recent advances in microbial toxin-related strategies to combat cancer. Semin Cancer Biol. 2022;86:753–68. doi:10.1016/j.semcancer.2021.07.007.
32. Zulfiqar U, Haider FU, Maqsood MF, Mohy-Ud-Din W, Shabaan M, Ahmad M, et al. Recent Advances in Microbial-Assisted Remediation of Cadmium-Contaminated Soil. Plants (Basel) 2023. doi:10.3390/plants12173147.
33. Danov A, Segev O, Bograd A, Ben Eliyahu Y, Dotan N, Kaplan T, Levy A. Toxinome-the bacterial protein toxin database. mBio. 2024;15:e0191123. doi:10.1128/mbio.01911-23.
34. Saenz JS, Kurz A, Ruczizka U, Bünger M, Dippel M, Nagl V, et al. Metaproteomics Reveals Alteration of the Gut Microbiome in Weaned Piglets Due to the Ingestion of the Mycotoxins Deoxynivalenol and Zearalenone. Toxins (Basel) 2021. doi:10.3390/toxins13080583.

35. Wang D-Z, Gao Y, Lin L, Hong H-S. Comparative proteomic analysis reveals proteins putatively involved in toxin biosynthesis in the marine dinoflagellate Alexandrium catenella. Mar Drugs. 2013;11:213–32. doi:10.3390/md11010213.
36. Farrell LJ, Lo R, Wanford JJ, Jenkins A, Maxwell A, Piddock LJV. Revitalizing the drug pipeline: AntibioticDB, an open access database to aid antibacterial research and development. J Antimicrob Chemother. 2018;73:2284–97. doi:10.1093/jac/dky208.
37. Knox C, Wilson M, Klinger CM, Franklin M, Oler E, Wilson A, et al. DrugBank 6.0: the DrugBank Knowledgebase for 2024. Nucleic Acids Res. 2024;52:D1265-D1275. doi:10.1093/nar/gkad976.
38. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2023 update. Nucleic Acids Res. 2022;51:D1373-80. doi:10.1093/nar/gkac956.
39. Zdrazil B, Felix E, Hunter F, Manners EJ, Blackshaw J, Corbett S, et al. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic Acids Res. 2024;52:D1180-D1192. doi:10.1093/nar/gkad1004.
40. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 2023;51:D690-D699. doi:10.1093/nar/gkac920.
41. Florensa AF, Kaas RS, Clausen PTLC, Aytan-Aktug D, Aarestrup FM. ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. Microb Genom 2022. doi:10.1099/mgen.0.000748.
42. Pulingam T, Parumasivam T, Gazzali AM, Sulaiman AM, Chee JY, Lakshmanan M, et al. Antimicrobial resistance: Prevalence, economic burden, mechanisms of resistance and strategies to overcome. Eur J Pharm Sci. 2022;170:106103. doi:10.1016/j.ejps.2021.106103.

926 43. Sulaiman SK, Musa MS, Tsiga-Ahmed FI, Dayyab FM, Sulaiman AK, Bako AT. A systematic
927 review and meta-analysis of the prevalence of caregiver acceptance of malaria vaccine for
928 under-five children in low-income and middle-income countries (LMICs). PLoS One.
929 2022;17:e0278224. doi:10.1371/journal.pone.0278224.
930 44. Tsakou F, Jersie-Christensen R, Jenssen H, Mojsoska B. The Role of Proteomics in
931 Bacterial Response to Antibiotics. Pharmaceuticals (Basel) 2020. doi:10.3390/ph13090214.
932 45. Sidjabat HE, Gien J, Kvaskoff D, Ashman K, Vaswani K, Reed S, et al. The use of SWATH
933 to analyse the dynamic changes of bacterial proteome of carbapanemase-producing
934 Escherichia coli under antibiotic pressure. Sci Rep. 2018;8:3871. doi:10.1038/s41598-018-
935 21984-9.
936 46. Barrett TC, Mok WWK, Murawski AM, Brynildsen MP. Enhanced antibiotic resistance
937 development from fluoroquinolone persisters after a single exposure to antibiotic. Nat
938 Commun. 2019;10:1177. doi:10.1038/s41467-019-09058-4.
939 47. Zhang Q-Y, Yan Z-B, Meng Y-M, Hong X-Y, Shao G, Ma J-J, et al. Antimicrobial peptides:
940 mechanism of action, activity and clinical potential. Mil Med Res. 2021;8:48.
941 doi:10.1186/s40779-021-00343-2.
942 48. Mohr KI, Volz C, Jansen R, Wray V, Hoffmann J, Bernecker S, et al. Pinensins: the first
943 antifungal lantibiotics. Angew Chem Int Ed Engl. 2015;54:11254–8.
944 doi:10.1002/anie.201500927.
945 49. Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, et al. Ribosomally
946 synthesized and post-translationally modified peptide natural products: overview and
947 recommendations for a universal nomenclature. Nat Prod Rep. 2013;30:108–60.
948 doi:10.1039/c2np20085f.
949 50. Montalbán-López M, Scott TA, Ramesh S, Rahman IR, van Heel AJ, Viel JH, et al. New
950 developments in RiPP discovery, enzymology and engineering. Nat Prod Rep.
951 2021;38:130–239. doi:10.1039/d0np00027b.

51. Süssmuth RD, Mainz A. Nonribosomal Peptide Synthesis-Principles and Prospects. Angew Chem Int Ed Engl. 2017;56:3770–821. doi:10.1002/anie.201609079.
52. Chevallereau A, Pons BJ, van Houte S, Westra ER. Interactions between bacterial and phage communities in natural environments. Nat Rev Microbiol. 2022;20:49–62. doi:10.1038/s41579-021-00602-y.
53. Cahill J, Young R. Phage Lysis: Multiple Genes for Multiple Barriers. Adv Virus Res. 2019;103:33–70. doi:10.1016/bs.aivir.2018.09.003.
54. Tkhilaishvili T, Lombardi L, Klatt A-B, Trampuz A, Di Luca M. Bacteriophage Sb-1 enhances antibiotic activity against biofilm, degrades exopolysaccharide matrix and targets persisters of Staphylococcus aureus. Int J Antimicrob Agents. 2018;52:842–53. doi:10.1016/j.ijantimicag.2018.09.006.
55. Clokie MR, Millard AD, Letarov AV, Heaphy S. Phages in nature. Bacteriophage. 2011;1:31–45. doi:10.4161/bact.1.1.14942.
56. Pirnay J-P, Djebara S, Steurs G, Griselain J, Cochez C, Soir S de, et al. Personalized bacteriophage therapy outcomes for 100 consecutive cases: a multicentre, multinational, retrospective observational study. Nat Microbiol. 2024;9:1434–53. doi:10.1038/s41564-024-01705-x.
57. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic discovery of anti-phage defense systems in the microbial pan-genome. Science 2018. doi:10.1126/science.aar4120.
58. Martínez Arbas S, Narayanasamy S, Herold M, Lebrun LA, Hoopmann MR, Li S, et al. Roles of bacteriophages, plasmids and CRISPR immunity in microbial community dynamics revealed using time-series integrated meta-omics. Nat Microbiol. 2021;6:123–35. doi:10.1038/s41564-020-00794-8.

59. Roume H, Heintz-Buschart A, Muller EEL, Wilmes P. Sequential isolation of metabolites, RNA, DNA, and proteins from the same unique sample. Methods Enzymol. 2013;531:219–36. doi:10.1016/B978-0-12-407863-5.00011-3.
60. Deusch S, Seifert J. Catching the tip of the iceberg - evaluation of sample preparation protocols for metaproteomic studies of the rumen microbiota. Proteomics. 2015;15:3590–5. doi:10.1002/pmic.201400556.
61. Heyer R, Kohrs F, Reichl U, Benndorf D. Metaproteomics of complex microbial communities in biogas plants. Microb Biotechnol. 2015;8:749–63. doi:10.1111/1751-7915.12276.
62. Keiblinger KM, Fuchs S, Zechmeister-Boltenstern S, Riedel K. Soil and leaf litter metaproteomics-a brief guideline from sampling to understanding. FEMS Microbiol Ecol 2016. doi:10.1093/femsec/fiw180.
63. Kleikamp HBC, Grouzdev D, Schaasberg P, van Valderen R, van der Zwaan R, van Wijgaart R de, et al. Metaproteomics, metagenomics and 16S rRNA sequencing provide different perspectives on the aerobic granular sludge microbiome. Water Res. 2023;246:120700. doi:10.1016/j.watres.2023.120700.
64. Kunath BJ, Minniti G, Skaugen M, Hagen LH, Vaaje-Kolstad G, Eijsink VGH, et al. Metaproteomics: Sample Preparation and Methodological Considerations. Adv Exp Med Biol. 2019;1073:187–215. doi:10.1007/978-3-030-12298-0_8.
65. Speda J, Johansson MA, Carlsson U, Karlsson M. Assessment of sample preparation methods for metaproteomics of extracellular proteins. Anal Biochem. 2017;516:23–36. doi:10.1016/j.ab.2016.10.008.
66. Zhang X, Li L, Mayne J, Ning Z, Stintzi A, Figeys D. Assessing the impact of protein extraction methods for human gut metaproteomics. J Proteomics. 2018;180:120–7. doi:10.1016/j.jprot.2017.07.001.
67. Kohrs F, Heyer R, Magnussen A, Benndorf D, Muth T, Behne A, et al. Sample prefractionation with liquid isoelectric focusing enables in depth microbial metaproteome

analysis of mesophilic and thermophilic biogas plants. Anaerobe. 2014;29:59–67. doi:10.1016/j.anaerobe.2013.11.009.

68. Hinzke T, Kouris A, Hughes R-A, Strous M, Kleiner M. More Is Not Always Better: Evaluation of 1D and 2D-LC-MS/MS Methods for Metaproteomics. Front Microbiol. 2019;10:238. doi:10.3389/fmicb.2019.00238.
69. Meier JL, Burkart MD. Proteomic analysis of polyketide and nonribosomal peptide biosynthesis. Curr Opin Chem Biol. 2011;15:48–56. doi:10.1016/j.cbpa.2010.10.021.
70. Chen Y, Ntai I, Ju K-S, Unger M, Zamdborg L, Robinson SJ, et al. A proteomic survey of nonribosomal peptide and polyketide biosynthesis in actinobacteria. J Proteome Res. 2012;11:85–94. doi:10.1021/pr2009115.
71. Bumpus SB, Evans BS, Thomas PM, Ntai I, Kelleher NL. A proteomics approach to discovering natural products and their biosynthetic pathways. Nat Biotechnol. 2009;27:951–6. doi:10.1038/nbt.1565.
72. Gómez-Varela D, Xian F, Grundtner S, Sondermann JR, Carta G, Schmidt M. Increasing taxonomic and functional characterization of host-microbiome interactions by DIA-PASEF metaproteomics. Front Microbiol. 2023;14:1258703. doi:10.3389/fmicb.2023.1258703.
73. Dumas T, Martinez Pinna R, Lozano C, Radau S, Pible O, Grenga L, Armengaud J. The astounding exhaustiveness and speed of the Astral mass analyzer for highly complex samples is a quantum leap in the functional analysis of microbiomes. Microbiome. 2024;12:46. doi:10.1186/s40168-024-01766-4.
74. Onigbinde S, Gutierrez Reyes CD, Sandilya V, Chukwubueze F, Oluokun O, Sahioun S, et al. Optimization of glycopeptide enrichment techniques for the identification of clinical biomarkers. Expert Rev Proteomics. 2024:1–32. doi:10.1080/14789450/2024.2418491.
75. Armengaud J. Metaproteomics to understand how microbiota function: The crystal ball predicts a promising future. Environ Microbiol. 2023;25:115–25. doi:10.1111/1462-2920.16238.

76. Starke R, Oliphant K, Jehmlich N, Schäpe SS, Sachsenberg T, Kohlbacher O, et al. Tracing incorporation of heavy water into proteins for species-specific metabolic activity in complex communities. J Proteomics. 2020;222:103791. doi:10.1016/j.jprot.2020.103791.
77. Kleiner M, Kouris A, Violette M, D'Angelo G, Liu Y, Korenek A, et al. Ultra-sensitive isotope probing to quantify activity and substrate assimilation in microbiomes. Microbiome. 2023;11:24. doi:10.1186/s40168-022-01454-1.
78. Bergen M von, Jehmlich N, Taubert M, Vogt C, Bastida F, Herbst F-A, et al. Insights from quantitative metaproteomics and protein-stable isotope probing into microbial ecology. ISME J. 2013;7:1877–85. doi:10.1038/ismej.2013.78.
79. van Kasteren S, Rozen DE. Using click chemistry to study microbial ecology and evolution. ISME Commun. 2023;3:9. doi:10.1038/s43705-022-00205-5.
80. Schaible GA, Jay ZJ, Cliff J, Schulz F, Gauvin C, Goudeau D, et al. Multicellular magnetotactic bacteria are genetically heterogeneous consortia with metabolically differentiated cells. PLoS Biol. 2024;22:e3002638. doi:10.1371/journal.pbio.3002638.
81. Hellwig P, Dittrich A, Heyer R, Reichl U, Benndorf D. Detection, isolation and characterization of phage-host complexes using BONCAT and click chemistry. Front Microbiol. 2024;15:1434301. doi:10.3389/fmicb.2024.1434301.
82. Sanderlin AG, Kurka Margolis H, Meyer AF, Lamason RL. Cell-selective proteomics reveal novel effectors secreted by an obligate intracellular bacterial pathogen. Nat Commun. 2024;15:6073. doi:10.1038/s41467-024-50493-9.
83. Hellwig P, Kautzner D, Heyer R, Dittrich A, Wibberg D, Busche T, et al. Tracing active members in microbial communities by BONCAT and click chemistry-based enrichment of newly synthesised proteins; 2024.
84. Picciani M, Gabriel W, Giurcoiu V-G, Shouman O, Hamood F, Lautenbacher L, et al. Oktoberfest: Open-source spectral library generation and rescoring pipeline based on Prosit. Proteomics. 2024;24:e2300112. doi:10.1002/pmic.202300112.

85. Buur LM, Declercq A, Strobl M, Bouwmeester R, Degroeve S, Martens L, et al. MS2Rescore 3.0 Is a Modular, Flexible, and User-Friendly Platform to Boost Peptide Identifications, as Showcased with MS Amanda 3.0. J Proteome Res. 2024;23:3200–7. doi:10.1021/acs.jproteome.3c00785.
86. Vande Moortele T, Devlaminck B, van de Vyver S, van den Bossche T, Martens L, Dawyndt P, et al. Unipept 6.0: Expanding Metaproteomics Analysis with Support for Missed Cleavages, Semi-Tryptic and Non-Tryptic Peptides; 2024.
87. Muth T, Behne A, Heyer R, Kohrs F, Benndorf D, Hoffmann M, et al. The MetaProteomeAnalyzer: a powerful open-source software suite for metaproteomics data analysis and interpretation. J Proteome Res. 2015;14:1557–65. doi:10.1021/pr501246w.
88. Cheng K, Ning Z, Zhang X, Li L, Liao B, Mayne J, Figeys D. MetaLab 2.0 Enables Accurate Post-Translational Modifications Profiling in Metaproteomics. J Am Soc Mass Spectrom. 2020;31:1473–82. doi:10.1021/jasms.0c00083.
89. Medema MH, Blin K, Cimermancic P, Jager V de, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011;39:W339-46. doi:10.1093/nar/gkr466.
90. Santos-Júnior CD, Pan S, Zhao X-M, Coelho LP. Macrel: antimicrobial peptide screening in genomes and metagenomes. PeerJ. 2020;8:e10555. doi:10.7717/peerj.10555.
91. Wang W, Ren J, Tang K, Dart E, Ignacio-Espinoza JC, Fuhrman JA, et al. A network-based integrated framework for predicting virus-prokaryote interactions. NAR Genom Bioinform. 2020;2:lqaa044. doi:10.1093/nargab/lqaa044.
92. Tynecki P, Guziński A, Kazimierczak J, Jadczuk M, Dastych J, Onisko A. PhageAI - Bacteriophage Life Cycle Recognition with Machine Learning and Natural Language Processing; 2020.

93. Marquet M, Hölzer M, Pletz MW, Viehweger A, Makarewicz O, Ehricht R, Brandt C. What the Phage: a scalable workflow for the identification and analysis of phage sequences. Gigascience 2022. doi:10.1093/gigascience/giac110.
94. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHASTEST: faster than PHASTER, better than PHAST. Nucleic Acids Res. 2023;51:W443-W450. doi:10.1093/nar/gkad382.
95. Nesvizhskii AI. Proteogenomics: concepts, applications and computational strategies. Nature Methods. 2014;11:1114–25. doi:10.1038/nmeth.3144.
96. Gholamizoj S, Ma B. SPEQ: quality assessment of peptide tandem mass spectra with deep learning. Bioinformatics. 2022;38:1568–74. doi:10.1093/bioinformatics/btab874.
97. van den Bossche T, Beslic D, van Puyenbroeck S, Suomi T, Holstein T, Martens L, et al. Metaproteomics beyond databases: addressing the challenges and potentials of de novo sequencing; 2024.
98. Bouwmeester R, Gabriels R, van den Bossche T, Martens L, Degroeve S. The Age of Data-Driven Proteomics: How Machine Learning Enables Novel Workflows. Proteomics. 2020;20:e1900351. doi:10.1002/pmic.201900351.
99. Lange E, Kranert L, Krüger J, Benndorf D, Heyer R. Microbiome modeling: a beginner's guide. Front Microbiol. 2024;15:1368377. doi:10.3389/fmicb.2024.1368377.
100. Priyadarsini M, Kushwaha J, Pandey KP, Rani J, Dhoble AS. Application of flow cytometry for rapid, high-throughput, multiparametric analysis of environmental microbiomes. J Microbiol Methods. 2023;214:106841. doi:10.1016/j.mimet.2023.106841.
101. Midani FS, David LA. Tracking defined microbial communities by multicolor flow cytometry reveals tradeoffs between productivity and diversity. Front Microbiol. 2022;13:910390. doi:10.3389/fmicb.2022.910390.
102. Phillips M, Cataneo RN, Chaturvedi A, Kaplan PD, Libardoni M, Mundada M, et al. Detection of an extended human volatome with comprehensive two-dimensional gas

chromatography time-of-flight mass spectrometry. PLoS One. 2013;8:e75274. doi:10.1371/journal.pone.0075274.

103. Anderson BG, Raskind A, Hissong R, Dougherty MK, McGill SK, Gulati AS, et al. Offline Two-Dimensional Liquid Chromatography-Mass Spectrometry for Deep Annotation of the Fecal Metabolome Following Fecal Microbiota Transplantation. J Proteome Res. 2024;23:2000–12. doi:10.1021/acs.jproteome.4c00022.

104. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51:D523-D531. doi:10.1093/nar/gkac1052.

105. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51:D587-D592. doi:10.1093/nar/gkac963.

106. Walke D, Micheel D, Schallert K, Muth T, Broneske D, Saake G, Heyer R. The importance of graph databases and graph learning for clinical applications. Database (Oxford) 2023. doi:10.1093/database/baad045.

107. Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front Microbiol. 2014;5:219. doi:10.3389/fmicb.2014.00219.

108. van Leeuwen PT, Brul S, Zhang J, Wortel MT. Synthetic microbial communities (SynComs) of the human gut: design, assembly, and applications. FEMS Microbiol Rev 2023. doi:10.1093/femsre/fuad012.

109. Herren CM, McMahon KD. Keystone taxa predict compositional change in microbial communities. Environ Microbiol. 2018;20:2207–17. doi:10.1111/1462-2920.14257.

110. Banerjee S, Schlaeppi K, van der Heijden MGA. Keystone taxa as drivers of microbiome structure and functioning. Nat Rev Microbiol. 2018;16:567–76. doi:10.1038/s41579-018-0024-1.

111. Li X-G, Chen M-X, Zhao S-Q, Wang X-Q. Intestinal Models for Personalized Medicine: from Conventional Models to Microfluidic Primary Intestine-on-a-chip. Stem Cell Rev Rep. 2022;18:2137–51. doi:10.1007/s12015-021-10205-y.

112. Molly K, Vande Woestyne M, Verstraete W. Development of a 5-step multi-chamber reactor as a simulation of the human intestinal microbial ecosystem. Appl Microbiol Biotechnol. 1993;39:254–8. doi:10.1007/BF00228615.

113. Vorländer D, Schultz G, Hoffmann K, Rasch D, Dohnt K. PETR: A novel peristaltic mixed tubular bioreactor simulating human colonic conditions. Biotechnol Bioeng. 2024;121:1118–43. doi:10.1002/bit.28636.

114. Shah P, Fritz JV, Glaab E, Desai MS, Greenhalgh K, Frachet A, et al. A microfluidics-based in vitro model of the gastrointestinal human–microbe interface. Nat Commun. 2016;7:11535. doi:10.1038/ncomms11535.

115. Kim HJ, Huh D, Hamilton G, Ingber DE. Human gut-on-a-chip inhabited by microbial flora that experiences intestinal peristalsis-like motions and flow. Lab Chip. 2012;12:2165–74. doi:10.1039/C2LC40074J.

116. Serrano-Villar S, Tincati C, Raju SC, Sáenz JS, Moreno E, Bargiela R, et al. Microbiome-derived cobalamin and succinyl-CoA as biomarkers for improved screening of anal cancer. Nat Med. 2023;29:1738–49. doi:10.1038/s41591-023-02407-3.

117. Kunath BJ, Rudder C de, Laczny CC, Letellier E, Wilmes P. The oral-gut microbiome axis in health and disease. Nat Rev Microbiol. 2024;22:791–805. doi:10.1038/s41579-024-01075-5.

118. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9. doi:10.1038/s41586-021-03819-2.

119. Taneishi K, Tsuchiya Y. Structure-based analyses of gut microbiome-related proteins by neural networks and molecular dynamics simulations. Curr Opin Struct Biol. 2022;73:102336. doi:10.1016/j.sbi.2022.102336.

120. Pu Y, Li L, Peng H, Liu L, Heymann D, Robert C, et al. Drug-tolerant persister cells in cancer: the cutting edges and future directions. Nat Rev Clin Oncol. 2023;20:799–813. doi:10.1038/s41571-023-00815-5.

121. Fossmark R, Olaisen M. Changes in the Gastrointestinal Microbiota Induced by Proton Pump Inhibitors-A Review of Findings from Experimental Trials. Microorganisms 2024. doi:10.3390/microorganisms12061110.

122. Sawaid IO, Samson AO. Proton Pump Inhibitors and Cancer Risk: A Comprehensive Review of Epidemiological and Mechanistic Evidence. J Clin Med 2024. doi:10.3390/jcm13071970.

123. Aarnoutse R, Ziemons J, Penders J, Rensen SS, Vos-Geelen J de, Smidt ML. The Clinical Link between Human Intestinal Microbiota and Systemic Cancer Therapy. Int J Mol Sci 2019. doi:10.3390/ijms20174145.

124. Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV, et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science. 2018;359:97–103. doi:10.1126/science.aan4236.

125. Villéger R, Lopès A, Carrier G, Veziant J, Billard E, Barnich N, et al. Intestinal Microbiota: A Novel Target to Improve Anti-Tumor Treatment? Int J Mol Sci 2019. doi:10.3390/ijms20184584.

126. Ervin SM, Ramanan SV, Bhatt AP. Relationship Between the Gut Microbiome and Systemic Chemotherapy. Dig Dis Sci. 2020;65:874–84. doi:10.1007/s10620-020-06119-3.

127. Chen J, Sun Y, Li J, Lyu M, Yuan L, Sun J, et al. In-depth metaproteomics analysis of tongue coating for gastric cancer: a multicenter diagnostic research study. Microbiome. 2024;12:6. doi:10.1186/s40168-023-01730-8.

128. Mills S, Shanahan F, Stanton C, Hill C, Coffey A, Ross RP. Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota. Gut Microbes. 2013;4:4–16. doi:10.4161/gmic.22371.

129. Li P, Luo H, Ji B, Nielsen J. Machine learning for data integration in human gut microbiome. Microb Cell Fact. 2022;21:241. doi:10.1186/s12934-022-01973-4.

130. Wolf M, Schallert K, Knipper L, Sickmann A, Sczyrba A, Benndorf D, Heyer R. Advances in the clinical use of metaproteomics. Expert Rev Proteomics. 2023;20:71–86. doi:10.1080/14789450.2023.2215440.

131. Li L, Ning Z, Zhang X, Mayne J, Cheng K, Stintzi A, Figeys D. RapidAIM: a culture- and metaproteomics-based Rapid Assay of Individual Microbiome responses to drugs. Microbiome. 2020;8:33. doi:10.1186/s40168-020-00806-z.

132. Cristobal-Cueto P, García-Quintanilla A, Esteban J, García-Quintanilla M. Phages in Food Industry Biocontrol and Bioremediation. Antibiotics (Basel) 2021. doi:10.3390/antibiotics10070786.

133. Rocha I, Ma Y, Souza-Alonso P, Vosátka M, Freitas H, Oliveira RS. Seed Coating: A Tool for Delivering Beneficial Microbes to Agricultural Crops. Front Plant Sci. 2019;10:1357. doi:10.3389/fpls.2019.01357.

134. Quinto EJ, Caro I, Villalobos-Delgado LH, Mateo J, De-Mateo-Silleras B, Redondo-Del-Río MP. Food Safety through Natural Antimicrobials. Antibiotics (Basel) 2019. doi:10.3390/antibiotics8040208.

135. Pankaj Bhatt, Bernard A. Engel, Mikael Reuhs, Halis Simsek. Cyanophage technology in removal of cyanobacteria mediated harmful algal blooms: A novel and eco-friendly method. Chemosphere. 2023;315:137769. doi:10.1016/j.chemosphere.2023.137769.

136. Kumar M, Jaiswal S, Sodhi KK, Shree P, Singh DK, Agrawal PK, Shukla P. Antibiotics bioremediation: Perspectives on its ecotoxicity and resistance. Environ Int. 2019;124:448–61. doi:10.1016/j.envint.2018.12.065.

137. Liu M, Gill JJ, Young R, Summer EJ. Bacteriophages of wastewater foaming-associated filamentous Gordonia reduce host levels in raw activated sludge. Sci Rep. 2015;5:13754. doi:10.1038/srep13754.

138. Kim S-M, Heo HR, Kim CS, Shin HH. Genetically engineered bacteriophages as novel nanomaterials: applications beyond antimicrobial agents. Front Bioeng Biotechnol. 2024;12:1319830. doi:10.3389/fbioe.2024.1319830.

139. Qiu D, Ke M, Zhang Q, Zhang F, Lu T, Sun L, Qian H. Response of microbial antibiotic resistance to pesticides: An emerging health threat. Science of The Total Environment. 2022;850:158057. doi:10.1016/j.scitotenv.2022.158057.

140. Delogu F, Kunath BJ, Queirós PM, Halder R, Lebrun LA, Pope PB, et al. Forecasting the dynamics of a complex microbial community using integrated meta-omics. Nature Ecology & Evolution. 2024;8:32–44. doi:10.1038/s41559-023-02241-3.

141. Rubin BE, Diamond S, Cress BF, Crits-Christoph A, Lou YC, Borges AL, et al. Species- and site-specific genome editing in complex bacterial communities. Nat Microbiol. 2022;7:34–47. doi:10.1038/s41564-021-01014-7.

142. Gelsinger DR, Vo PLH, Klompe SE, Ronda C, Wang HH, Sternberg SH. Bacterial genome engineering using CRISPR-associated transposases. Nature Protocols. 2024;19:752–90. doi:10.1038/s41596-023-00927-3.

143. van Puyvelde B, Dhaenens M. Add mass spectrometry to the pandemic toolbox. Elife 2021. doi:10.7554/eLife.75471.

144. Holstein T, Kistner F, Martens L, Muth T. PepGM: a probabilistic graphical model for taxonomic inference of viral proteome samples with associated confidence scores. Bioinformatics 2023. doi:10.1093/bioinformatics/btad289.