From Concept to Perspective: Digital Twins of Microbial Systems

Simon Labarthe, Nicolas Creusot, Clémence Frioux, Guillaume Gautreau, Elie Desmond-Le Quéméner, Lorenzo Sala, Agustín G. Yabo, Paul Dou, Gabriel Capson-Tojo, Eric Dugat-Bony, et al.

From Concept to Perspective: Digital Twins of Microbial Systems

Simon Labarthe[1,2], Nicolas Creusot[3], Clémence Frioux[2,4], Guillaume Gautreau[5], Elie Desmond-Le Quemener[6], Lorenzo Sala[5], Agustín G. Yabo[7], Paul Dou[8], Gabriel Capson-Tojo[6], Eric Dugat-Bony[9], Andrea Fernandez Diaz[10,11], Mathieu Mezache[5], Coralie Muller[2], Sthyve Tatho[1,2], Sahak Yeghiazaryan[6], Jerome Harmand[6], Jean-Roch Mouret[12], Thibault Nidelet[12], Sylvain Prigent[13], Ludovic Cottret[14,15], Anne Goelzer[16], Isabelle Lesur-Kupin[2,1], Lionel Rigottier-Gois[17], Franck Salin[1,2], Valentina Baldazzi[18,19], Celine Casenave[7], Helene Falentin[12], Mahendra Mariadassou[5], Rafael Muñoz-Tamayo[20], Milka Popova[8], Alain Rapaport[7], David Sherman[2], Jean-Philippe Steyer[6], Corinne Vacher[21], and the Artemis Consortium

Abstract

Digital twins (DTs) are increasingly recognized across diverse sectors for their capacity to enhance the control, efficiency, and comprehension of the physical or biological systems they represent. For microbial systems, DTs could allow model-guided improvements of the services provided by the microbial communities in the agrifood chain. While DTs definitions are generally built on the same core idea of bi-directional exchanges between digital and physical counterparts, where real-time data feeds digital models and model-driven insights guide the real system, a wide variety of definitions of what is a DT still co-exist across domains. This variability underscores the need for a clear, system-specific definition of DTs for microbial ecosystems.

In this perspective paper, we propose a conceptual framework for microbial system digital twins (MSDTs), defined as a collection of models dynamically linked to the microbiological system through in-line, at-line or off-line data and control flows. We illustrate this framework with examples spanning environmental, bioprocess, plant, animal, food, and human microbial systems, in a One Health perspective. For each ecosystem, we explore the potential applications of MSDTs. We also identify the scientific challenges that remain in experiments, bioinformatics, data science, modeling, control and microbial ecosystem engineering to build accurate MSDTs.

We advocate for the development of MSDT in laboratory settings, as a catalyst for interdisciplinary sciences, and we stress practical and ethical issues preventing the generalization of MSDT for large-scale applications. However, high-tech MSDTs in laboratory environments may pave the way for low-tech, generalizable microbial solutions for improved ecosystemic microbial services.

Keywords: Digital twins, Microbial ecology, microbial community models

[1]Univ. Bordeaux, INRAE, BIOGECO, 33610, Cestas, France, [2]Inria, Univ. Bordeaux, INRAE, F-33400 Talence, France, [3]INRAE, UR EABX, 33600, Cestas, France, [4]Université Paris-Saclay, INRAE, MGP, 78350, Jouy-en-Josas, France, [5]Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-enJosas, France, [6]INRAE, Univ Montpellier, LBE, 11100, Narbonne, France, [7]MISTEA, Univ. Montpellier, INRAE, Institut Agro, 34060 Montpellier, France, [8]Université Clermont Auvergne, INRAE, VetAgro Sup, UMR 1213 Herbivores, Saint-Gènes-Champanelle, France, [9]UMR 0782 SAYFOOD, Université Paris-Saclay, INRAE, AgroParisTech, 22 place de l'agronomie, 91120, Palaiseau, France, [10]RA Landscape Functioning, Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374, Müncheberg, Germany, [11]Thaer Institute, Faculty of Life Sciences, Humboldt University of Berlin, Unter den Linden 6, 10099, Berlin, Germany, [12]SPO, Univ Montpellier, INRAE, Institut Agro, Montpellier, France, [13]Univ. Bordeaux, INRAE, UMR1332 BFP, 33140 Villenave d'Ornon, France, [14]Toxalim, Université de Toulouse, INRAE, ENVT, EI-Purpan, Toulouse, France, [15]MetaToul-MetaboHUB, National Infrastructure of Metabolomics and Fluxomics, Toulouse, France, [16]University, INRAE, UR MIAT, Castanet-Tolosan, France, [17]Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, France, [18]Université Côte d'Azur, INRAE, CNRS, ISA, France, [19]Université Côte d'Azur, Inria, Inrae, CNRS, MACBES Project-Team, France, [20]Université Paris-Saclay, INRAE, AgroParisTech, UMR Modélisation Systémique Appliquée aux Ruminants, 91120, Palaiseau, France, [21]INRAE, Bordeaux Sciences Agro, ISVV, SAVE, 33882 Villenave-d'Ornon, France

Correspondence
simon.labarthe@inrae.fr

1. Introduction

Digital twins (DT), formally introduced in the early 2000s in an industrial context (Grieves, 2023), are attracting growing interest across a wide range of sectors such as mechanics (Ritto and Rochinha, 2021), urbanism (Lehner and Dorffner, 2020), healthcare systems (Vallée, 2023), agrifood systems (Tzachor et al., 2022a), agriculture (Metcalfe et al., 2023; Purcell and Neubauer, 2023; Pylianidis et al., 2021), environmental technologies (Torfs et al., 2024), and medicine (Björnsson et al., 2020; Laubenbacher et al., 2024). Progress in data acquisition (Marjani et al., 2017) and integration of artificial intelligence (AI) brings perspectives for the development and adoption of DTs, by improving simulation times, predictive maintenance, autonomous decision-making and hypothesis-based scenarii (De Domenico et al., 2025). For example, advances in omics, i.e. high-throughput measurements of biological molecule sets for comprehensive system-wide analysis, combined with AI progresses, offer opportunities to create an AI virtual cell that could be personalized using specific patients data to create a DT and suggest suitable interventions (Bunne et al., 2024). This scientific focus is accompanied by massive public investments on DTs, such as the EU initiatives Destination Earth for a DT of the planet Earth (Bauer et al., 2021) and EDITH for a European Virtual Human Twin (Viceconti et al., 2024), together with the Society for In silico Medicine VPH [1].

All of these DTs are built on the same core idea: a digital object made of code and models, the DT, is created to mirror a real world (biological) system, the biological twin, functioning as its virtual counterpart. Unlike digital shadows, which involve one-way data flow from the real system to its virtual entity but lack feedback, a digital twin adds feedback through bidirectional flows with its biological (or physical) counterpart: data flows in, while control commands flow back to influence the real-world system (Fuller et al., 2020). Despite this general consensus, definitions of DT vary widely in the literature, with significant application-dependent differences in terms of model specifications or data flows between digital and biological twins (Wright and Davidson, 2020). These discrepancies in DT definitions result in two main issues: (i) they can obscure the differences between DT and classical mathematical models, thus making it difficult to determine whether DT is a genuine breakthrough for model implementation or just a scientific hype; (ii) they make it more difficult to identify a clear and uniform definition of DT, forcing each research field involved in their development or application to adapt the concept to its own context (Wright and Davidson, 2020). The goal of this opinion paper is to clearly define and investigate the potential of DTs within the context of microbial systems.

Microbial systems underpin key processes in health (e.g., barrier effect against pathogens, microbiota-based products and probiotics), food production (e.g., fermented food, food preservation), agriculture (e.g. for plant protection or animal health), industry (e.g., biological wastewater treatment and anaerobic digestion) and the environment (e.g., biogeochemical cycles, primary production), placing them as key players for sustainable development within the One-Health concept (Affagard et al., 2026; Banerjee and van der Heijden, 2023; Winkler et al., 2025). Microbial communities are particularly well suited for DT approaches because they combine fast and measurable process dynamics with a high degree of experimental control and numerous observation options, enabling flexible experimental designs and data availability. In controlled environmental settings, many multi-modal data streams can now be routinely acquired with readily available equipments. This combination of observability, controllability, replicability and societal relevance

[1] VPH website: https://vph-society.org/who-we-are/

makes microbial systems an especially attractive domain for developing DTs. Conversely, DTs of microbial systems could bring breakthrough insights in the functions and interactions driving the microbial communities, while providing a practical tool to pilot the outcome of the microbial processes. After a brief overview of digital models of microbial systems in various areas of microbiology (sec. 2 and boxes 1 to 6), the paper will focus on clearly defining what a Microbial System DT (MSDT) is (sec. 3). Next, the article explores the promises MSDTs hold for fundamental and applied research on microbial systems management (sec. 4) and the key scientific topics to be addressed for the development of efficient MSDT (sec. 5). Finally, the technical, theoretical, and ethical issues associated to MSDTs are discussed (sec. 6).

2. Current models of microbial systems

2.1. Modeling microbial communities across ecosystems

In microbiology, mathematical modeling of microbial systems has been a major focus as early as the 1940s, with the introduction of the Monod equation (Monod, 1949) standing as a foundational example for modeling the growth of micro-organisms. Since then, a substantial body of literature on microbial systems (MS) modeling has emerged, including process-based kinetic modeling, partial-differential equation (PDE) modeling, genome-scale metabolic modeling (GSMM), machine-learning (ML) or ecological modeling (Cerk et al., 2024; Wade et al., 2016; Widder et al., 2016). In this paper, MS refer to both complex natural communities and synthetic communities (SynComs) that are isolated, designed and assembled for a specific application, associated with a specific host or free-living in their environment. MS exploration thus encompasses various topics in microbial ecology, such as biodiversity and interactions within natural ecosystems (as holobiont or free-living), as well as microbiota engineering issues, here understood as ecological management strategies such as screening, controlling and optimizing microbial interactions and functions in designed or selected synthetic consortia. To outline the current landscape of key topics in both microbial ecology and modeling, a selection of emblematic MS across a diverse range of applied fields within a One Health perspective is addressed, including environmental, biotechnological, plant, animal, food, and human MS (see Fig. 1 for an overview and boxes 1 to 6 for a brief review of microbial ecology and modeling issues in each ecosystem). Reviewing these MS reveals that the objectives of MS models are ecosystem-specific, as they are directly tied to the primary ecosystem services delivered by the microbial community. In contrast, modeling methodologies are often transferable across ecosystems and are more closely associated with the complexity of the microbial community itself. For SynCom models, genome-scale approaches can be employed, leveraging microbial genomes, metabolic modeling, inference of microbial interactions, and dynamical models calibrated with omics data. For more complex communities, functional and/or taxonomic simplifications are conducted using knowledge-based, ML-based or deep-learning approaches, leading to the development of process-based models at the community level (Widder et al., 2016).

2.2. Microbial systems have intrinsic advantages for system biology and ecology

The large diversity of microbial community models illustrated in the Box 1 to 6 reflects key intrinsic advantages of MS over macroscopic ecosystems (i.e. those involving macroscopic individuals such as animals or plants) for system biology approaches and articulation with mathematical models.

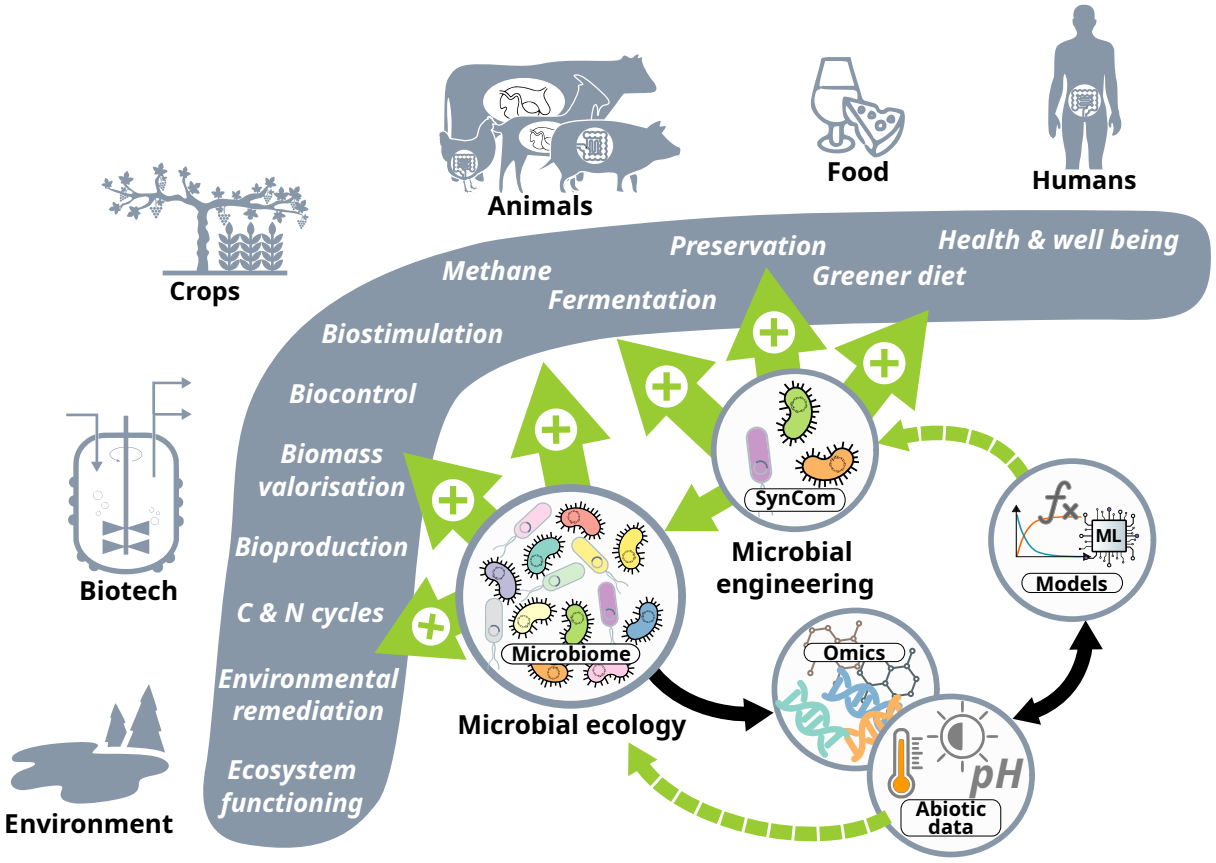

Figure 1 – Microbial systems and applications. MS are found within various environments, whether macroscopic hosts (e.g. plants, animals or humans), manufactured systems (e.g. bioreactors or food) or open environments (e.g. seawater or soil). MS perform complex metabolic functions, that extend the metabolic landscape of their environment, thus providing a number of valuable services such as ecosystem function regulation and environmental remediation, biomass valorisation and bioproduction of high value-added compounds, crop protection and plant health promotion, livestock production efficiency and environmental impact, food preservation and human health enhancement. To better understand microbial dynamics, meta-omics (like metabarcoding, metagenomics, metatranscriptomics or metametabolomics) or abiotic data (like temperature, pH, light, etc) give insight into the ecology (functioning and taxonomic assembly) of the MS. Mathematical models help decipher the omics data, and give a framework for efficiently build microbiota engineering strategies, such as SynComs design, in order to steer the microbial communities and improve the microbial services. For these ecosystems, MSDTs could effectively close this bi-directional interaction loop by automatizing data production and analysis, dynamical modeling and model-guided microbial system steering.

Controlled and reproducible experiments of reduced ecosystems. In microbiology, assembling and cultivating defined consortia of few microbial species is technically feasible. These SynComs can be cultivated on chemically defined media, allowing precise control over the nutritional environment, or in interaction with host models (e.g. foliar discs as proxies for the phyllosphere, or organoids and gut-on-a-chip devices to simulate colonic environment) for simplified host-microbiota interactions. These reductionist approaches significantly improve experimental control and reproducibility.

Towards molecular ecology. Since microbial genomes are compact, efficient methods now exist to assemble (meta-) genome-scale metabolic models that capture both individual species metabolic activity and interspecies trophic chains at the community level (Belcour et al., 2020;

Box 1: Modeling MS to explore microbial regulations in environmental microbiology. Environmental microbiology explores both free-living and host-associated microbial communities in natural environments, such as rivers, lakes, wetlands, grasslands, or forests, and their role in essential ecosystem services, such as carbon fixation, detoxification, biogeochemical cycling, organic matter decomposition, and primary production (Falkowski et al., 2008; Morin and Artigas, 2023). These complex multi-kingdom communities form dynamic, interconnected networks that underpin the resilience and functionality of the ecosystem (Pascual-García et al., 2020; Van Bruggen et al., 2019): understanding their adaptive mechanisms is vital for predicting how ecosystems will cope wil multi-stress scenarios, including climate change, chemical and biological pollution, or land anthropization (Hellal et al., 2023; Philippot et al., 2021).

Modeling environmental scales. A key challenge for accurate MS models in natural environments is bridging scales: while macro-scale cycles are central in ecosystem services, key microbial functions often depend on a small number of keystone species involved in complex ecological interactions, such as cross-feeding and antibiosis (Mataigne et al., 2021). Reconciling these scales requires the development of advanced modeling methods that articulate functional omics data integration, metabolic modeling (Creusot et al., 2025; Muller et al., 2018) and spatial models accounting for spatially-structuring features such as quorum sensing or multi-species biofilms (Lami et al., 2023). By integrating metabolic, physiological, and environmental data at the community level, these models can predict ecosystem responses to stressors and identify critical tipping points (Barbosa et al., 2025; Iqbal et al., 2025).

Box 2: Modeling industrial bioprocesses: the example of anaerobic digesters. Anaerobic digesters are bioreactors that convert organic feedstocks such as agricultural waste and energy crops into valuable outputs like biogas (a renewable energy source) and digestate (a soil fertilizer). These systems host diverse microbial communities, primarily bacteria and archaea, continuously shaped by non-sterile inputs and fluctuating operational conditions. At first order approximation, the process involves four interconnected biochemical stages: hydrolysis, acidogenesis, acetogenesis, and methanogenesis (Batstone et al., 2002). These steps are governed by intricate mechanisms, including syntrophic relationships (Pan et al., 2021), which are particularly prone to instability due to narrow thermodynamic windows. Additionally, stable digester operation requires microbial adaptation to rapid shifts in operating conditions, a challenge for slow-growing microbes like syntrophs and certain archaea.

Bioprocess-based models for anaerobic digestion Efforts to model anaerobic digestion led to the development of the Anaerobic Digestion Model 1 (ADM1) (Batstone et al., 2002), a foundational bioprocess-based compartment dynamic model that has inspired numerous adaptations (Batstone et al., 2006; Capson-Tojo et al., 2021; Hassam et al., 2015; Ramirez et al., 2009). Robust models incorporate dozens of state variables to capture the complexity of microbial ecosystems, multiphasic interactions (gas, liquid, solid), chemical equilibria, and ion-pairing reactions (Flores-Alsina et al., 2016). However, variability in poorly characterized inputs and spatial heterogeneity pose challenges. Data scarcity, particularly for microbial populations and feedstock composition, further complicates model calibration. Mathematical modeling optimizes anaerobic digestion by ensuring continuous, robust treatment of all inputs while balancing stability and performance. ML can integrate large industrial datasets, and when combined with mechanistic models in hybrid approaches, it enhances interpretability (Vanrolleghem et al., 2025). In wastewater treatment, facility-wide DTs are emerging, though microbial modeling remains simplified to limit computational complexity (Wang et al., 2024).

Cerk et al., 2024). These models also serve as powerful tools for integrating and deciphering omics data. By contrast, macroscopic ecosystems lack comparable ecosystem-wide models capable of resolving molecular-scale interactions and metabolic dynamics between macroscopic individuals (plants, animals) with the same accuracy and reproducibility.

Large microbiome databases. Large databases of annotated microbial genomes, pathways and proteins acquired on microbiome samples are now available (Richardson et al., 2023). These

Box 3: Model-guided enhancement of food fermentation. Fermenting food with microorganisms offers a "3-in-1" benefit: improving safety and conservation, enhancing flavor and texture, and promoting health through probiotics and prebiotics. This process also supports low-carbon footprint diets by using plant-based ingredients and reducing energy-intensive steps like cooling in alcoholic fermentation (Harlé et al., 2025; Rul et al., 2022; Teng et al., 2021). Fermentation typically occurs in closed systems with varying pH (3–9), anaerobic conditions, and minimal external interactions. The microorganisms involved—bacteria, fungi, or both—depend on the food type (vegetables, dairy, meat), its structure (raw, liquid, sliced), and processing factors (temperature, salt, hydration). Starter cultures ensure safety and consistency, while microbial activity—through alcoholic (e.g., yeast in beer), lactic (e.g., bacteria in yogurt), or propionic fermentation—produces preservatives (alcohol, organic acids) and beneficial compounds (aroma, vitamins, health-promoting metabolites).

Optimizing Fermentation with Models. Predictive models simulate microbial dynamics and key parameters like metabolite levels, pH, and flavor compounds (Li et al., 2025b). These models use metabolic or process-based frameworks (Karimi et al., 2026), enriched with multi-omic data (metabolomics, transcriptomics) and environmental factors (temperature, pH) (Lecomte et al., 2024; Somerville et al., 2022). They can also be combined with machine learning in hybrid models. Predictive fermentation models help informed decision and optimization of production stability, sensory quality, and resource efficiency while maintening safety (Yabo and Casenave, 2023). In precision fermentation—where engineered microbes produce specific compounds (proteins, enzymes)—DTs could address biosynthesis bottlenecks (Helmy et al., 2024; Melkonian et al., 2023), with potential applications to modernize and streamline processes.

Box 4: Crop Microbiome Models for Disease Biocontrol. Plants host complex microbial communities that enhance nutrient uptake, growth, stress tolerance, immune priming, and pathogen resistance (Chialva et al., 2022). These microbial services are increasingly harnessed for sustainable agriculture through microbe-based biocontrol products, biofertilizers, and biostimulants (Compant et al., 2025). While a first generation of single-strain biocontrol solutions showed variable field efficacy against diseases, research now focuses on multi-strain SynComs for greater resilience to environmental changes (Compant et al., 2025). Other research avenues include breeding plants with an enhanced capacity to associate with beneficial microorganisms (French et al., 2021), rewilding the microbiome with ancestral microbes from wild environments (Raaijmakers and Kiers, 2022) and managing plant diversity to steer the microbiome toward a more protective state (Meyer et al., 2022b; Moreau et al., 2025; Wyckhuys et al., 2025).

Modeling simplified communities for plant disease biocontrol. To develop microbiome-based strategies for disease biocontrol, integrative frameworks combine metagenomics to identify key micro-organisms and functions (Choi et al., 2026; Pacheco and Vorholt, 2023), culturomics to isolate them (Sarhan et al., 2019), and SynComs assembly based on top-down and/or bottom-up approaches (Mehlferber et al., 2024) for lab, greenhouse and field testing against pathogens (Choi et al., 2026; Pacheco and Vorholt, 2023). SynComs represent a necessary step for elucidating the molecular mechanisms underlying biocontrol effects. GSMM and multi-omics data can help elucidate microbial interactions and dynamics (Feierabend and Töpfer, 2025; Schäfer et al., 2023). Microorganisms exhibiting similar dynamics can be identified and grouped with dedicated ML models, revealing shared functions, and high-throughput numerical exploration of possible interactions can be conducted to select efficient SynComs for targeted function. Iterating this process could lead to the development of a digital twin for model-guided design of biocontrol consortia.

data can be mined and arranged to build and explore metagenome-assembled genomes and genome-scale metabolic network reconstructions (Cerk et al., 2024) or to build foundation AI models able to explore and predict sequences and functions (Brixi et al., 2025; Wiatrak et al., 2025). Large datasets of microbiome samples (Carlino et al., 2024; Kuhn et al., 2026; Pasolli et al., 2017; Rodrigues et al., 2025) and their associated metadata can enable data-driven studies

Box 5: Modeling the rumen microbiome to mitigate livestock methane emissions. The rumen microbiome comprises bacteria, archaea, protozoa, fungi, and viruses, and is vital for the nutrition and health of ruminants. It influences key outcomes such as feed efficiency and ruminal acidity through microbial fermentation. It is also involved in ruminant methane emissions, which is a major contributor to the environmental footprint of livestock. Despite age-related changes, the mature rumen is a resilient ecosystem. Rumen microbes degrade plant fibers, producing essential compounds such as volatile fatty acids (VFAs) and hydrogen, which methanogens convert into methane. (Huws et al., 2018)

Dynamical models of rumen fermentation use ordinary differential equations (ODEs) to represent microbial metabolism and transport phenomena. Some rumen models (Muñoz-Tamayo et al., 2021, 2016) share structural similarities with anaerobic reactor models (e.g., ADM1 - Box 2). Rumen specific features include differential degradation rates of feed particles (Pressman and Kebreab, 2024), and host factors such as VFA absorption, saliva secretion and eating behavior patterns (Vivares et al., 2025). Incorporating microbial genomic information into rumen models should improve their predictive power and enhance model capabilities for investigating microbial levers to manage animal health and ruminal livestock methane production (Muñoz-Tamayo et al., 2023). Recent advances include the use of state observers to link microbial data to VFA dynamics (Davoudkhani et al., 2024), and the use of metabolic network reconstruction to model the metabolism of a cellulolytic bacterium (Fakih et al., 2023).

Box 6: Modeling the human gut microbiota. Humans maintain a continuous symbiotic relationship with diverse microbial communities, particularly within the colon, where hundreds to thousands of species contribute to host physiology. The gut microbiota ferments indigestible dietary residues, such as fibers, supplying a portion of the host's energy requirements (Arnoldini et al., 2025). It also plays a critical role in immune system maturation, epithelial cell turnover, and barrier effect against pathogens (Fan and Pedersen, 2021). Host-microbiota selection is driven by factors like colonic hypoxia, immunity-mediated beneficial species sorting (Bandyopadhyay et al., 2025; Crouch et al., 2024), with diet significantly shaping microbial diversity (Schmidt et al., 2018). This symbiosis is being challenged by environmental changes (unhealthy diets, processed foods, antibiotics, and modern sanitary practices) leading to a loss of microbial diversity and dysbiosis (Carding et al., 2015). Microbiota dysbiosis is linked to chronic diseases (e.g., obesity, diabetes, inflammation, cancer, neurodegenerative disorders or ocular conditions (Rocks et al., 2025)) and infectious diseases (Fan and Pedersen, 2021). The gut microbiota exhibits spatial structures, varying longitudinally (proximal vs. distal colon) and radially (mucus-associated vs. lumen communities), with strong interactions at the epithelial interface (McCallum and Tropini, 2024).

Spatial models of the gut microbiota and intestinal physiology. Spatial models must capture these features, either by using compartmental dynamical systems (Kettle et al., 2018; Muñoz-Tamayo et al., 2010) or partial differential equations to represent intestinal regions (Cremer et al., 2017; Labarthe et al., 2019; Moorthy et al., 2015). Multi-scale integration, from microscopic interactions (e.g., microbial-metabolic exchanges, stem cell regulation (Darrigade et al., 2022; Haghebaert et al., 2024)) to organ-level processes (e.g., peristalsis, nutrient absorption, mucus turnover (Labarthe et al., 2019)), remains a key challenge. For biomedical applications, model personalization is essential, incorporating diet, host physiology, and patient-dependent microbial profiles, leveraging machine learning analyses of metagenomic datasets to define microbial structures and functions (Frioux et al., 2023a; Labarthe et al., 2023; MetaHIT Consortium et al., 2011).

through ML or AI approaches (Frioux et al., 2023a; MetaHIT Consortium et al., 2011; Tap et al., 2023).

Leveraging known bioprocesses. Complex microbial communities often exhibit activities that can be distilled into fundamental metabolic processes such as hydrolysis, proteolysis, fermentation, respiration or catabolism, centered around biomass formation and metabolism of key compounds like fibers, sugars, proteins or short chain fatty acids. These community-scale functions

can be supplemented by specialized metabolism when it is key for the community dynamics, like biocide production in pathosystems. This metabolic simplification lends itself to implementation of process-based kinetic models, enabling macroscopic-scale predictions of community dynamics. Beyond modeling, it also provides a structured approach to formulate testable scientific hypotheses in microbiology.

Scaling up experiments and models. The small size of microorganisms and their rapid replication rates simplify experimental scale-up, which can be further amplified by robotics and miniaturization, such as microfluidic devices. This enables high-throughput parallel cultivation of diverse microcosms at minimal operational cost and time. On the computational side, parallelization, high-performance computing (HPC), and model simplification techniques (e.g., surrogate modeling) facilitate comparable scaling up of community-level models. In addition, MS experiments impose fewer ethical and regulatory constraints than animal or plant subjects for laboratory experiments.

Steering MS and microbial ecosystem engineering. Experimentalists have at hand several levers to finely manage the dynamics of MS. This control can be exerted through environmental tuning (changing culture media, modulating feeding rates in bioreactors, or applying stressors such as pH, temperature or light), as well as consortium engineering by adding or removing targeted species, e.g. using phages. These experimental options can be efficiently integrated with control engineering and mathematical framework for model-guided steering of MS.

These inherent advantages position MS as a unique platform for advancing the synergy between models and experimental systems, paving the way for MSDTs.

3. Digital twin of microbial systems

Classical definitions of DTs emphasize bidirectional interconnection between the biological system and its digital twin, enabling continuous data flow and model-guided feedbacks (Fuller et al., 2020). While model-guided control of microbial systems has a long history, in particular in bioprocesses and food fermentation (see boxes 2 and 3), DTs introduce additional layers of complexity beyond traditional control. Some definitions highlight DT personalization to a specific individual (Viceconti et al., 2024), but this concept does not easily translate to complex microbial ecosystems. Others stress real-time observation and control of the system (Fuller et al., 2020), which is rarely feasible in microbiology. To adapt DT definition to MS, we propose a precise definition of a MSDT, structured around three core components: 1) a rigorous characterization of the biological twin, hence clarifying how the concept of personalization adapts to microbial ecosystems, 2) the conceptualization of the MSDT as an ensemble of computational models rather than a unique model, in order to extend classical modeling and control approaches, and, 3) the distinction of three types of interactions (in-line, at-line and off-line interactions) between the biological and digital twins, aligning MSDT with existing data production and control methods.

3.1. Characterizing the biological twin

Specifying a MS. The first step in constructing a MSDT is the explicit delineation of its real-world biological counterpart. While this may appear somewhat obvious, it actually deserves clarification, particularly for systems as versatile, diverse and heterogeneous as microbial communities. Traditionally, a digital twin represents a given individual: for example in an industrial context, a DT of an aircraft represents a specific plane within a production series, resulting in as many DTs

as aircrafts. However, the concept of an “individual” becomes ambiguous when applied to microbial ecosystems or communities, depending on how MS are defined. MS can be specified in two contrasting ways. They can follow a **constructive explicit definition**: the system is precisely specified by providing the precise composition of the microbial community in term of proportions of defined isolates from determined microbial collections within a biochemically controlled environment. This definition is related to defined consortia in microbial ecology. MS can also be defined with a **non-constructive implicit definition**: the system is described using generic terms such as "gut microbiome" or "phyllosphere microbiome" which, in practice, refer to samples obtained through standardized protocols without controlling the community composition nor the environment. In such a case, the “individual” may correspond either to a given sampling realization (i.e., a specific instance of the system) or to an averaged proxy derived from multiple samples. This definition is related to undefined microbiota in microbial ecology.

MSDT for personalization or system simplification. These definitions lead to distinct MSDT paradigms. For constructive definition or single sampling realizations, building a MSDT is a **personalization task**: the MSDT must adapt to the unique characteristics of the biological twin. For averaged proxies of different samples, building a MSDT is a **simplification task**: the goal is to identify a parsimonious set of features representative of the different samples of the biological twin. We note that this definition broadens the usual scope of digital twins, usually tightly linked to a proper individual. Such an extension is necessary given the inherent complexity and dynamics of microbial ecosystems interacting with a complex environment and sometimes with a host.

Illustrative examples. To illustrate this expanded framework, we present diverse examples of MS biological twins, each exemplifying distinct definitions of MS and applying to the microbial systems described above.

(1) In environmental microbiology, a biological twin can be a given lake together with its microbiota defined as the pooling of different spatially distributed samples. The MS is then an averaged proxy of spatialized populations of a specific lake.
(2) In a context of bioprocesses, a MS can be specified with different samples of the same sludge inoculated in different chemostats fed with controlled medium. The MS is then this collection of chemostat devices hosting different samples of the implicitly defined microbial population. The MSDT will be built upon a simplified common description of the sludge microbial population.
(3) In a context of food fermentation, a MS can be a given bioreactor with a defined starter and plant-based liquid substrate. In this fully explicit example, different MSDTs should be built for every bioreactor.
(4) In a context of plant microbiology, a MS can be defined as a given biobank of hundreds of microbial isolates and a given high-throughput cultivation platform able to assemble and grow a large number of simplified communities. The MS is then fully explicit and constructive.
(5) In the context of the ruminant holobiont, a given animal is modeled with its complex rumen microbiota characterized with time-resolved rumen samples. The MS is a proxy of the unique microbial community of this specified animal, in interaction with its host. In this case, the MS is implicitly characterized.

(6) In human microbiology, a MS can be set by choosing a human individual, while specifying their diet habits and microbiota with dedicated surveys and proxies that usually determine the taxonomic or genetic composition of the community. The microbial compartment is defined implicitly but the MSDT can be personalized to the given individual.

3.2. Conceptualization of the MSDT with modular architecture

State-dependant model structure. Rather than a monolithic model, we define a MSDT in a modular design paradigm as a collection of inter-connectable models. This architecture enables context-dependent instantiation, where specific modules are selectively activated or deactivated based on the current state of the biological twin. When a first version of the MSDT is instantiated, the current model configuration is fitted using incoming data streams. However, if the system behavior deviates beyond the validity domain of the current model, due to perturbations or external stressors, the MSDT dynamically reconfigures by assembling a new combination of modules from its repository. These new modules are not only a recalibration of the same models: they can have different model structures, with e.g. the addition of a new term in the equations, or even a new modeling paradigm, with e.g. a switch between process-based model and GSMM. This adaptive framework ensures that the MSDT remains both accurate and computationally efficient across varying conditions. A classical modeling approach using only one model can qualify as a MSDT, but extending its range of validity in future versions of the MSDT may necessitate a modular architecture to accommodate increased complexity and adaptability.

Example of a modular MSDT of the gut microbiota. A gut microbiota MSDT can exemplify this modularity. Here, distinct modeling bricks may include an accurate model of intestinal crypt dynamics, a detailed representation of the mucus layers, different versions of microbiota models with various proportions of pathogens or short-chain fatty acid producers, and an immune response module. During inflammatory episodes, the immune and crypt turnover modules are activated to simulate pro-inflammatory signaling and epithelial remodeling, providing mechanistic insights into disease progression. Conversely, when homeostatic conditions are recovered, these modules can be deactivated as their contributions become negligible, in order to avoid unnecessary computational load and to speed up computation. This selective activation ensures that the MSDT adapts its complexity and its biological relevance to the current state of the biological twin.

Example of a modular MSDT for anaerobic digestion models. Similarly, in a wastewater treatment plant MSDT, modularity enables robust responses to microbial unbalance. For instance, if an abrupt microbial input disrupts the microbial consortium and inhibits the sludge digestion, the microbial model can be reconfigured by redefining the structure of the population dynamics model to take into account the new microbial populations and metabolic shifts (Tartakovsky et al., 2002). This model can also describe new corrective monitoring (e.g., substrate adjustments or bioaugmentation) to steer the community until reactivation of the sludge.

3.3. Three types of interactions between biological and digital twins.

The data flow and interactions between the biological or ecological system and the MSDT can be decomposed into four basic actions, all forming a cyclical framework (Fig. 2).

(1) **Observe**: the first step, MS observation, encompasses systematic generation, preprocessing and analysis of experimental data, thus involving experiments, bioinformatics and data science.

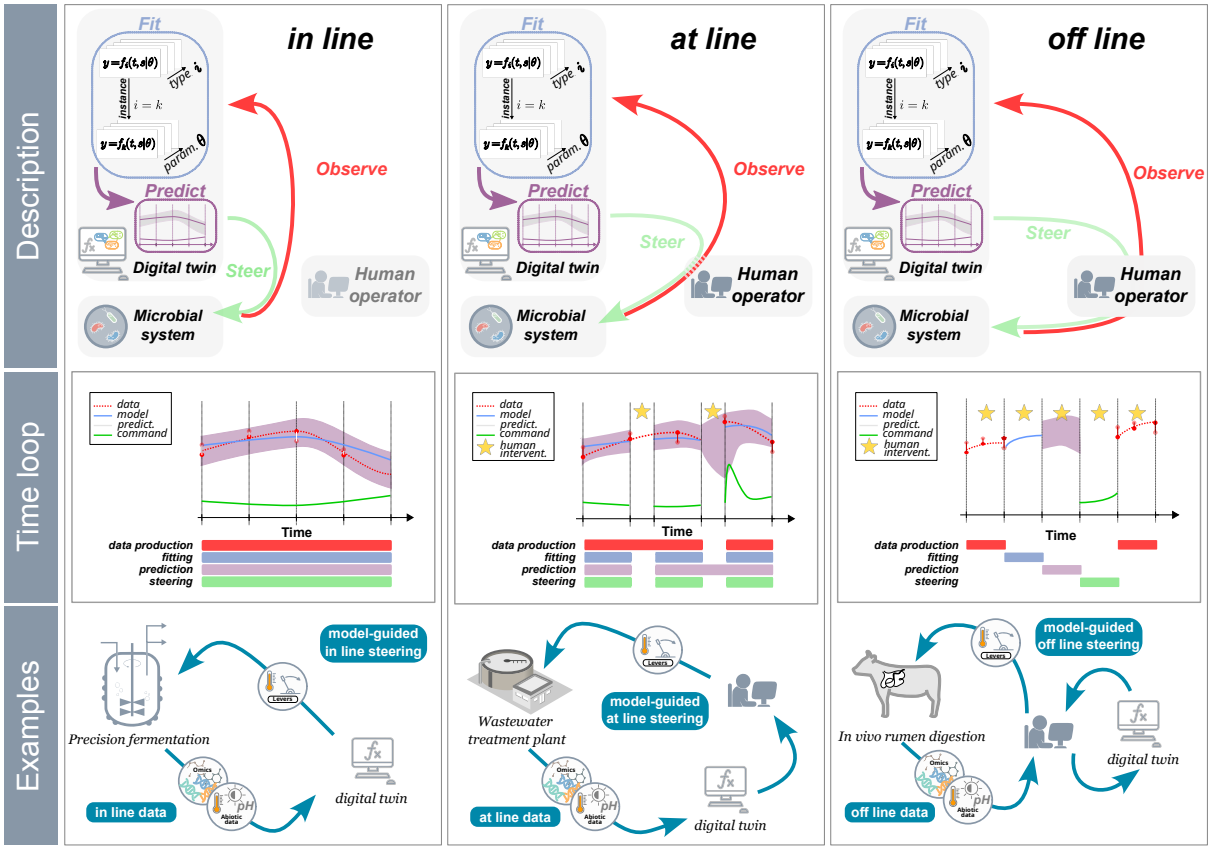

Figure 2 – **In-line, at-line and off-line loops.** Three types of interactions between MS and MSDT can be identified: in-line, at-line and off-line loops. In each loop type, from a precise delineation of the system, a collection of models (different model types) can be built, the specific instantiation and fitting (parameter θ inference) of which depend on current observations of the MS. Then, four actions (observe (red), fit (blue), predict (purple) and steer (green)) are articulated in an automated pipeline (upper pannel). The dynamic articulation of these actions is detailed for each loop (middle panel: dynamics of data (red), fitted model (blue), forecast (purple) and command (green), together with activation of data production, fitting, prediction and steering steps (colored bars)). Finally, emblematic examples are provided (bottom panels). **In-line (left)**. In in-line loops, data production, model fit and prediction and MS steering are automatic, real-time and simultaneous, keeping humans out of the loop, as an external observer (upper figure). An archetypal example is a MSDT designed as a controller of the MS for real-time control of precision fermentation (bottom-left figure). **At-line (center)**. The at-line loop integrates rapid data collection or command with slower semi-automatic procedures involving human operators (upper figure). In at-line loops, time-demanding bottlenecks with human interventions (central figure, yellow star) pause the process, either for data production or model computation. A canonical situation could be a MSDT designed for data assimilation of semi-automatic data for water treatment plant management, where the human operator is kept inside the loop for model reconfiguration in case of extreme events (bottom figure). **Off-line (right)**. The off-line loop is performed on a larger time-scale, with required human interventions (upper figure). In off-line loops, experiments, model fitting, model exploration and steering are made sequentially, with systematic human interventions (central figure). A typical example could be a MSDT used for digital exploration and model-guided decision and experimental design in animal experiments.

(2) **Fit**: next, MSDT fitting involves the instantiation of a context-appropriate model architecture followed by parameter calibration using the observed data. It is therefore at the crossroads of data science and modeling.

(3) **Predict**: the fitted model then enables MSDT prediction, for forecasting future system states and computing optimal interventions, thus bridging analytical insights with actionable strategies. It involves modeling and control engineering.
(4) **Steer**: finally, MS steering implements these model-informed interventions in the real-world biological system, using control theory and microbiota engineering strategies, completing the feedback loop.

This integrated workflow transforms raw biological data into actionable knowledge while maintaining continuous synchronization between the biological twin and the MSDT. Depending on the time scale and degree of automation of these actions, three distinct interaction paradigms can be defined for the coupling between microbial systems and their MSDT: in-line, at-line and off-line loops, illustrated in Fig. 2.

Level 1: In-line loops. In-line loops feature fully automated, real-time execution of the four actions: observe, fit, predict and steer. No direct human intervention is required, appart external observations, enabling autonomous operation. A canonical example is a model-guided control of fermentation, where the digital twin continuously optimizes process parameters (e.g., substrate feeding, pH, temperature) to maintain optimal (or viable) performance under dynamic conditions (Zhao et al., 2025).

Level 2: At-line loops. In at-line loops, the four actions are operated under semi-automated execution, with intermittent human intervention required to address persistent bottlenecks, e.g. time-demanding production of additional data when in-line sensors are insufficient, manual model reconfiguration, computationally intensive optimization tasks, or implementation delays for control actions (e.g., substrate preparation). In this mode, the operator is temporarily integrated into the loop to resolve situations that exceed the autonomous system capabilities. A telling example is a MSDT of a wastewater treatment plant: while the MSDT operates fully automatically in nominal functioning, an interface allows the testing of what-if scenario by human operators to search for more efficient operating conditions. Based on these additional insights, efficient operations are defined and implemented on the real system, allowing automatic steering to resume with improved accuracy (Daneshgar et al., 2024).

Level 3: Off-line loops. In off-line loops, the four actions are executed sequentially and asynchronously on distinct timescales with continuous human oversight. The operator remains fully integrated into the loop, coordinating manual interventions at every stage. An emblematic example of off-line loops could be a MSDT of the rumen or gut microbiome. In such system, real-time data acquisition and processing are infeasible, and microbiome modulation via pre- or probiotics operates on large timescales constrained by digestive physiology. Off-line loops have recently been used in the *design-built-test-learn* paradigm as introduced for a food microbiology MSDT, where iterative refinement relies on discrete, human-guided experimentation and analysis (Helmy et al., 2024).

These feedback loops, whether in-line, at-line, off-line or a mix of it, are fundamental to the definition of digital twins. Consequently, certain models previously classified as MSDT (Sizemore et al., 2024) would not fall under this proposed definition, by lacking a feedback loop from the MSDT towards the biological twin. We review in Figure 3 and Tables 1, 2 and 4 different data acquisition and command methodologies for microbial populations and microbial functions available for the different loop types.

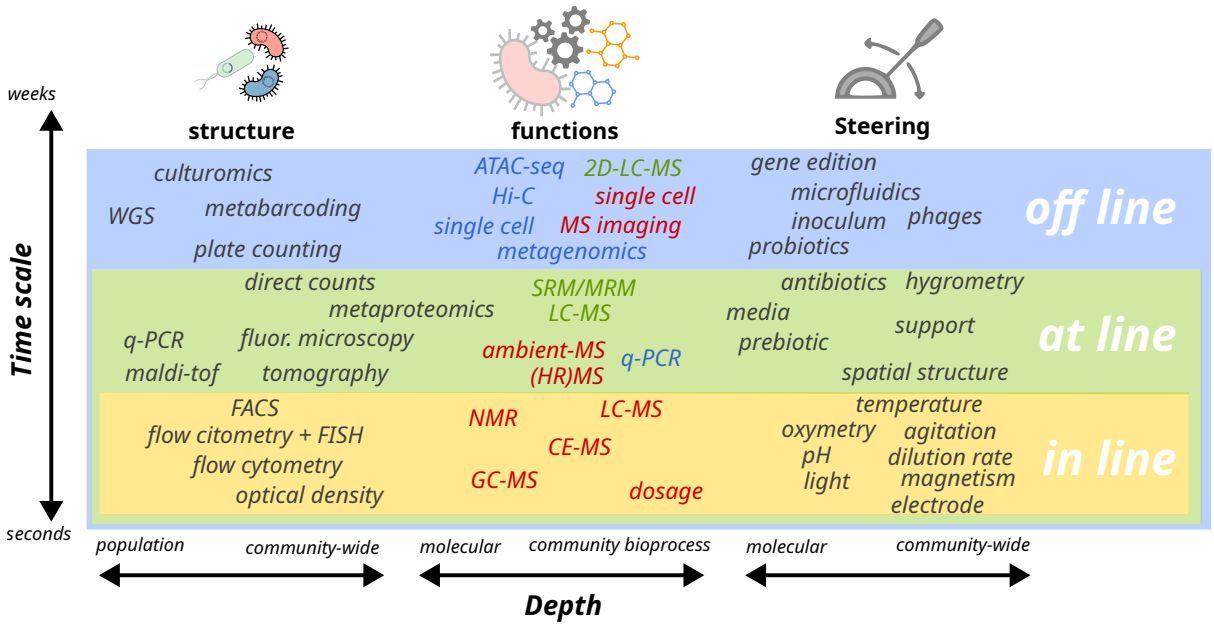

Figure 3 – Data, commands and feedback loops. Different biodiversity (left panel) and functional data (central panel) are reviewed, together with ecological levers that can be activated to steer microbial communities (right panel). Data types and commands are placed along two gradients: production time scales and depth, understood as the functional or taxonomic resolution of the data and the command, providing at a glance available inputs and outputs for in-line, at-line and off-line loops, together with expected precision of the action. For functional data, we indicate in blue metagenomics and metatranscriptomics data, in green metaproteomic data and in red metabolomics data. This representation supplements the information provided in tables 1, 2 and 4, that also contain abbreviation definitions.

4. Promises of digital twins for microbial ecology

4.1. MSDT: a research object

Accelerating research and reducing costs. By putting the emphasis in accelerating data production and analysis, as well as model computation and microbial steering, MSDT can significantly speed up the iterative interactions between experiments and models. Furthermore, the development of MSDTs may facilitate critical transformations in research workflows, transitioning from off-line interactions toward at-line and in-line processes, with real-time integration as the ultimate goal. With model-guided experimental design and experimental steering, MSDTs will reduce experimental times, costs and environmental footprints. In the animal science domain, MSDT can contribute to the implementation of the 3Rs (Replace, Reduce, Refine) principles for the use of animals in scientific research.

Helping to address and investigate scientific questions. When defining a MSDT to answer a scientific question, experimentalists, bioinformaticians, modellers and control engineers have to set up together the experimental design at the same time as formalizing the analysis pipeline, the model characteristics and the microbial control strategy. This dialogue makes it possible to identify new scientific questions at the conjunction of their respective disciplines, extending the original question in an interdisciplinary dialogue. An example of this can be determining microbial interactions in a community. A dedicated experimental design, such as a microfluidic set up interconnecting isolated wells with individual microbial populations, can be enhanced with

an adapted microbial community model focused on interactions: the model can identify metabolites mediating metabolic interactions, hence giving specifications for the measurements and the control strategy. This multidisciplinary crosstalk facilitates the anticipation of extreme situations: when the biological twin shows an unexpected behaviour, an off-line loop of the MSDT can reconfigure the model to detect dysfunctions, understand their underlying reasons, and bring back the system towards normal functioning. Furthermore, the system biology or ecology approach underlying MSDT development allows extending the MSDT in a modular way to integrate the community into a wider environment, like host-microbiota interactions. Finally, as the MSDT is clearly directed towards functional management of the community, the link between microbial taxonomic diversity and community-wide functions is central in the MSDT outcome, hence bringing insights into the system's microbial ecology. In particular, it can help deciphering the role of low abundant populations or under represented keystone species.

4.2. A large panel of potential applications

MSDT can be envisioned in all microbiology fields.

Model-guided management of environmental water quality. In environmental microbiology, a MSDT could be designed to mirror the microbial taxonomic composition and/or metabolites (e.g. toxins) production dynamics in a lake, with the final goal of anticipating and mitigating the effect of the bloom of pathogenic or toxic micro-organisms. A MSDT could take into account large scale levers like effluent follow-up or water management of the rivers, considering various physico-chemical parameters and ecological processes.

Waste and wastewater treatment plants. A MSDT of a water treatment plant could involve a microbial compartment providing integrated information to the human operator, including model predictions, links towards databases and memory of the process operations. MSDT outputs must handle discrete data such as human observations or previous decisions, and should be able to propose appropriate actions to avoid failure or recover when the process is stopped. The ultimate goal of the MSDT is robustness, by maintaining the treatment of the whole influx without failure, rather than optimality.

Controlled fermented food. DTs offer a solution for improving the functioning of fermentation systems. By providing real-time feedback and control, in-line digital twins can predict anomalies in the fermentation process and help restore equilibrium before product quality is compromised by abiotic (e.g. salt) or biotic (e.g. inoculation of a new strain) perturbations. This predictive capability could radically change the way fermentation processes are managed, leading to more consistent and reliable outcomes in food production. DT could also offer a solution for the optimal selection of microbial consortia adapted to the fermentation of a given raw material (e.g. novel plant-based fermented product) (Karimi et al., 2026).

SynCom selection in plant microbiome. The functional augmentation of plant microbiomes using SynComs is emerging as a promising alternative to synthetic pesticides for enhancing cultivated plant health. However, the effective selection of SynComs—those possessing both targeted functions and robust engraftment capabilities—remains a significant challenge. Traditional approaches relying on knowledge-based selection of strains from microbial collections are limited by the existing knowledge and the combinatorial explosion when combining them into SynComs, which far exceeds experimental feasibility. A MSDT of a SynCom-testing platform offers a transformative solution. By enabling model-guided SynCom selection, a massively parallel MSDT

can perform *in silico* screening of potential SynComs, identifying putative consortia for subsequent *in vitro* high-throughput testing in microfluidic experiments. Initially, this MSDT could be off-line, while progresses in robotization could enable at-line or in-line SynCom-testing platform MSDT in the future. The experimental data generated can then refine the digital exploration conducted by the MSDT, creating an iterative, funnel-like process that progressively converges towards SynComs of interest.

Fostering animal health. Modeling microbial proxies of key physiological variables within a MSDT of ruminants (and other livestock species) enables individualized management in a precision livestock farming context. Given the pivotal role of the rumen microbiome in digestion, animal health, and greenhouse gas emissions, the MSDT can be engineered to optimize precision feeding strategies that improve rumen function while mitigating methane emissions. However, microbial phenotypes often involve trade-offs such as those between feed efficiency and methane reduction, or between immune resilience and productivity. By accounting for these functional trade-offs, the MSDT can support balanced decision-making tailored to specific management goals. Furthermore, by integrating microbial biomarkers associated with health disorders, the MSDT can trigger an off-line diagnostic loop for early detection and individualized preventive or curative interventions, ultimately enhancing both animal well-being and operational efficiency.

Personalized gut microbiota management. In humans, a MSDT of the gut microbiota holds significant potential for personalized microbiome management, thereby enhancing health and well-being. The MSDT could integrate modular components, including an accurate description of human diet in terms of microbial substrates, intestinal epithelium turnover, inflammatory responses, pathogenic infection and metabolic capabilities of microbial populations at different functional granularity. Tailored for specific contexts such as dysbiosis, inflammation or pathogenic colonization, the MSDT could enable personalized modulation of the gut microbiota. This could be achieved through targeted interventions, including precision nutrition, prebiotics, probiotics or antibiotic therapy.

5. Digital twins as a catalyst of interdisciplinarity

By design, digital twins reside at the crossroads of multiple scientific fields: experimental science, bioinformatics, data science, modeling, control engineering and microbiota engineering. Therefore, developing a MSDT requires advancements in methods and techniques at the interface of these fields, with two main objectives: speed and depth (see Fig. 4). Increasing speed entails accelerating each step of the interaction loops between both twins, with the ultimate goal of achieving real-time in-line loops between the MSDT and its biological twin. Increasing depth involves developing experimental, sampling and analytical tools that yield deeper insights into the functioning of the microbial system. This aims to transition from phenomenological to mechanistic understanding, from community-level descriptions to molecular-scale characterization, from macroscale to microscale resolution in both time and space. Reaching speed and depth are often at odds: the priority between these objectives depends on the applicative context. The key scientific fronts that MSDTs can advance to reconcile speed and depth are discussed below.

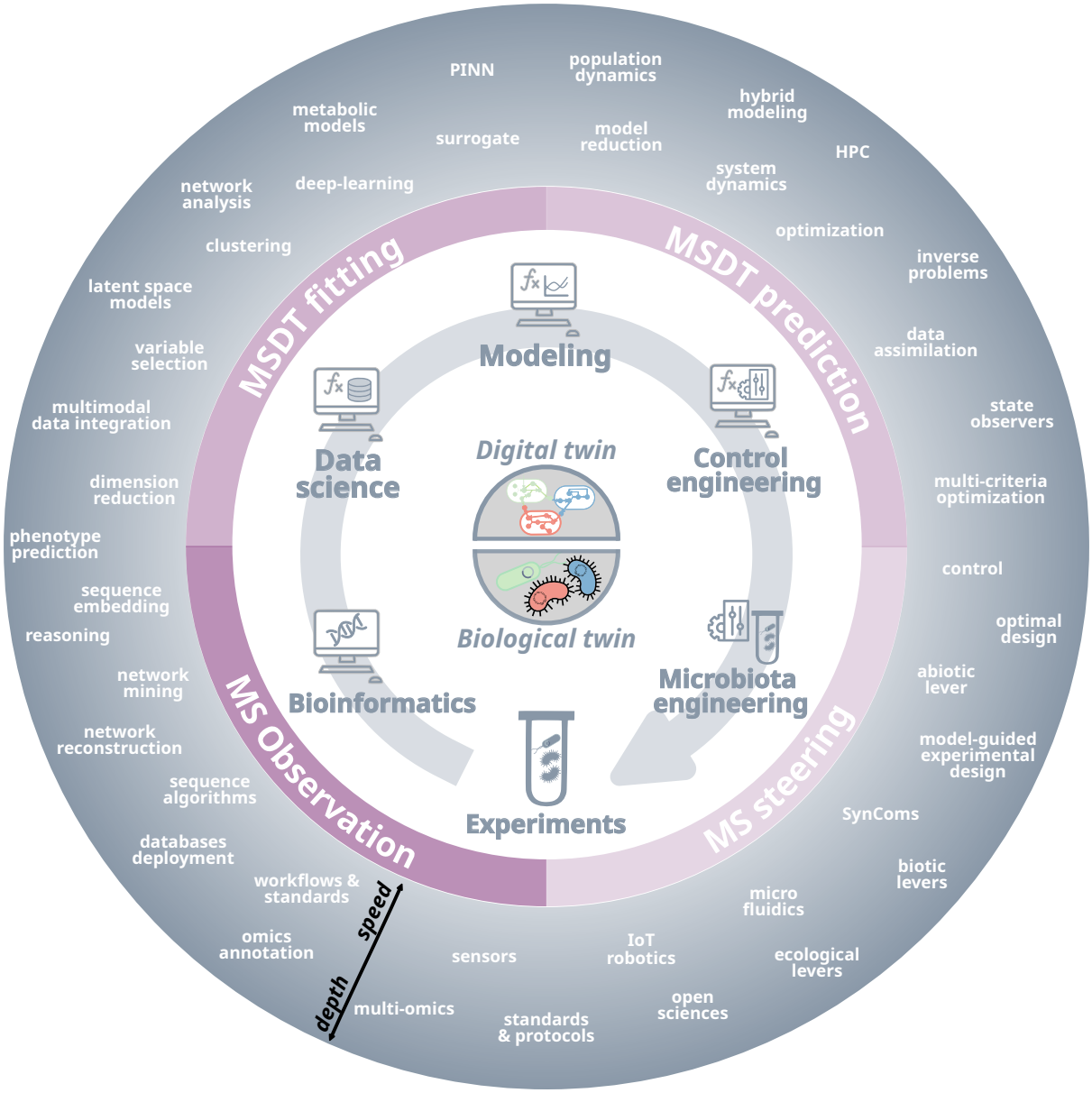

Figure 4 – MSDTs foster interdisciplinarity. By design, MSDTs sit at the intersection of wetlab experimentation, bioinformatics, data science, modeling, control, and microbiota engineering. Constructing a MSDT involves integrating a continuum of methods across these six disciplines, each presenting methodological challenges, directed towards enhancing speed or depth of data, analysis, models or commands. This interdisciplinary articulation not only advances technical capabilities but also spurs new biological questions, particularly in microbial ecology. Abbreviations: PINN (Physic-Informed Neural Network), HPC (high-performance computing), IoT (internet of things)

5.1. Forefront of scientific research in experimental and data acquisition sciences

Observing poorly accessible systems. Some microbial systems, such as the human gut or the rumen ecosystem, are inherently inaccessible for data acquisition through direct sampling. While stools might serve as practical proxies for the gut microbiota, they offer limited spatio-temporal resolution. Even if the recent development of ingestible intelligent capsules paves the way of *in vivo* intestinal microbome direct sampling (Ding et al., 2021), capturing high-throughput data, especially during transient microbial dynamic phases, remains a major challenge, particularly for

targeted metabolomics or exo-metabolome analysis. Advances in automation could allow for higher frequency measurements, which are crucial for monitoring and maintaining optimal fermentation conditions. Both in human and ruminants, microsensor technologies offer a promising practical solution to monitor in situ key variables such as pH, temperature and volatile fatty acids (Han et al., 2022; So et al., 2023). Technologies for the analysis of exhaled breath (coupled to systems such as the GreenFeed (Hammond et al., 2016) for monitoring methane produced by ruminants) offer noninvasive approaches for assessing rumen function (Islam et al., 2024; Jorge-Smeding et al., 2025). These devices allow for 3-4 measurements per day. Increasing both the time rate and the range of detectable metabolites would provide richer input for a MSDT.

Miniaturization for improved measurements. The microscopic nature of microorganisms poses technical challenges for sensor design in microbial systems. Miniaturization addresses these challenges by improving spatial and temporal resolution. For instance, mass spectrometry imaging and ambiant MS (e.g., DESI, MALDI, or SIMS; see Table 2) enables microscale metabolomics, while millifluidic devices allow high-throughput sampling of microbial communities grown in droplets and monitored via optical systems (Boitard et al., 2015). When coupled with robotics, milli or microfluidic systems can scale up experiments by running numerous experimental conditions in parallel. Finally, microfluidics can decipher microbial interactions: isolated microorganisms can be cultivated in interconnected microwells with controllable metabolite exchange via micropipes mimicking metabolic interactions between populations (Ugolini et al., 2024). The field of culturomics focuses on advancing these techniques.

Faster and deeper omics data generation. Over the past decade, sequencing technologies have seen dramatic improvements in sequencing depth, read length, throughput, error rates and correction, and cost reduction. However, sample conditioning, DNA extraction, and post-processing, including metagenomic assembly, remain bottlenecks, limiting DNA sequencing to off-line loops. A similar trend applies to chromatography-based technologies for metametabolomics and metaproteomics: while sample preparation, chromatographic rates, processing times, and costs continue to improve using robotics, real-time, in-line measurements have just started to be explored in both MS (Cortada-Garcia et al., 2024) and NMR-based analysis (Bouillaud et al., 2019). Robotization and real-time sequencing devices could accelerate these steps, gradually enabling the integration of sequencing into at-line and in-line MSDT. Real-time sequencing of microbial communities could provide invaluable insights into ongoing fermentation process, facilitating timely adjustments. However, expectations regarding the outcome of such data generation must account for the computational cost and the potential for automation of the downstream data processing: fast analyses such as reference-based mapping of sequenced reads may for instance be preferred over *de novo* assembly (Meyer et al., 2022a). We reviewed available omics facilities to capture microbial biodiversity and functions, focusing on processing times and depth (see Tables 1, 2, 3 and 4 and Fig. 3).

5.2. Challenges in bioinformatics

Managing omics data avalanche. The accumulation of omics data is only valuable when paired with efficient analytical tools. Bioinformatic pipelines for sequence binning, genome assembly, and pangenomic graph construction are essential for effectively exploiting DNA and RNA-seq data (Meyer et al., 2022a). A core challenge lies in accelerating data processing to scale analyses in line with increasing sequencing depth and read length, while preserving accuracy. High-performance computing and parallelization can partially address scalability, but AI

techniques offer further potential. By learning the underlying structure of genomic information and by providing universal representations of sequences using the embedding induced by foundational models, AI could streamline data analysis by projecting exponentially growing volume of data into dense latent spaces with a large but fixed dimension and a simpler geometry (Vaswani et al., 2017a). In the way, the emergence of microbiome large cohort studies in (meta)metabolomics requires to speed up data processing and annotation as well as improving data fusion and normalization to streamline the huge amount of complex data (Hajjar et al., 2023). Foundational models pre-trained or fine-tuned on microbiome datasets are beginning to emerge (Medearis et al., 2026; Pope et al., 2025b; Zhang et al., 2026). While their current applications primarily rely on transfer learning for classification tasks, these models are also expected to enable generative AI capabilities for MSDT steering in the near future.

Improving omics data contextualization. Beyond technological breakthroughs for deeper and faster measurements, bioinformatic advances are needed to improve the contextualization of raw omics data. Building open databases and efficient querying methods enables the integration of biological knowledge into observations. Marker genes, genomes or metagenomic-assembled genomes databases enrich sequencing data with taxonomic or functional insights, while gene, metabolite, and pathway annotation databases further enhance data interpretation. Ontologies help organize this information for improved knowledge extraction. A continuous effort must be maintained to improve databases size and accuracy, as well as querying speed, in order to improve automatic reconstruction pipelines and overall data signals in MSDT (Ma et al., 2023; Pop et al., 2024).

Metabolic models for deciphering microbial community functions. Metabolic network reconstructions organize genomic data into the metabolic pathways that microorganisms are genetically equipped with and can potentially activate in favourable contexts. Scalable computational pipelines can be implemented to automatically generate draft network reconstruction for well-characterized microbiomes, such as the human gut microbiota (Heinken et al., 2023). However, achieving high-accuracy network reconstruction remains a significant challenge, particularly for microbial ecosystems with limited or sparse database resources. By linking putative metabolites to actionable functions, these networks enable the integration of genomic, metatranscriptomic, metaproteomic and metametabolomic data. AI models can be used to enhance network reconstruction (Boer et al., 2024). GSMM considering those networks and environmental conditions can be leveraged to simulate cellular responses to their environment and predict the involved cellular biochemical mechanisms (Dillard et al., 2021), even for complex metabolism such as proteolysis (Paulay et al., 2024). At the community level, articulating multiple metabolic networks reveals metabolic exchanges that underpin trophic interactions among microbial populations (Cerk et al., 2024). The ability to mine large datasets of metabolic networks and scale this process to match observed microbial diversity is critical for deciphering community-wide metabolic functioning. Improving the accuracy and scalability of metabolic predictions therefore remains a central challenge for developing effective MSDTs of complex communities (Li et al., 2025a).

5.3. Advancements for data analysis

Omic data management. After bioinformatic processing, omic data acquired from microbial communities (metagenomics, metabarcoding, metranscriptomics or metametabolomics) result in data tables that pose significant challenges for statistical analyses. The tables are compositional (i.e. involving frequencies rather than absolute counts), high dimensional, zero-inflated (i.e. most

omic features are unobserved in a significant proportion of the samples) and span several orders of magnitude. They require special handling and developments for classical tasks such as differential analysis, variable selection in predictive framework or variance decomposition (Hernández Medina et al., 2022). Moreover, the joint analyses of multi-omics data requires the use of data integration techniques, whether early (table concatenation, latent space and factor analysis, etc) or late (kernel based integration) (Rohart et al., 2017). In time-varying MSDTs, managing time-series is critical, requiring dedicated methods such as time smoothing, parametric curve fitting, vector autoregression models or curve clustering to extract meaningful temporal relationships between features and outcomes of interest (growth rate, acetate production, etc) (Sherwani et al., 2025).

Microbiota simplification and pattern recognition. Identifying universal patterns or delineating discrete microbial types is a major research axis in microbial ecology. It aims to decipher the variability of omics features and simplify microbiome data into distinct groups characterized by simple community-scale biomarkers. Classical machine learning such as soft and hard clustering or dimensionality reduction (PCA, t-SNE, UMAP, NMF [2], etc) have long been used to reveal these patterns. These analysis can be conducted on taxonomic or functional data (Frioux et al., 2023a; Labarthe et al., 2023; MetaHIT Consortium et al., 2011; Tap et al., 2023).

Deep learning for microbiome omics data. Deep learning, particularly through the use of pre-trained models, presents a promising avenue for advancing pattern recognition in microbial systems data (Hernández Medina et al., 2022). Convolutional neural networks (CNNs), which have demonstrated exceptional efficiency in image-based pattern recognition, or transformers, the architecture grounding natural language models (Vaswani et al., 2017b), can be adapted to identify characteristic structures within omics data matrices (Pope et al., 2025a; Sharma et al., 2020). Autoencoder-based deep learning models enable the extraction of robust, low-dimensional representations from high-dimensional biological datasets (Oh and Zhang, 2020). Recurrent neural networks (RNNs) further facilitate the analysis of time-series data, allowing for the identification of microbial interactions and trophic networks (Baranwal et al., 2022).

However, these methods still need further development to handle multi-omics integration, better incorporation of metadata to correct for unbalanced representation of conditions, and scalability in the face of dimensionality increase, computational cost and interpretability issues, to come up with robust and simple biomarkers that are highly predictive of the microbial state. In MSDT, projection on reduced lattent spaces are essential for simplified screening of the microbial system and effective trade-offs between speed, computational burden and accuracy for microbiota representations.

5.4. Advancements for mechanistic modeling

Navigating across scales. Mathematical modeling of microbial systems requires integration across multiple scales: spatial, temporal, structural (taxonomic composition and biodiversity) and functional. Spatial scales range from cellular-level phenomena (e.g., active swimming, cell-cell interactions) to millimeter-scale processes (e.g. biofilm formation, flocculation or quorum sensing), or even larger structures (e.g., microbial populations in soil, organs, or environments). Temporal scales span millisecond metabolic reactions, hourly growth dynamics, and seasonal variations like in plant microbiota. Structural scales cover several species in SynComs to thousands of taxa in

[2] PCA: principal component analysis, t-SNE: t-Distributed Stochastic Neighbor Embedding, UMAP: Uniform Manifold Approximation and Projection, NMF: non-negative factorization

complex communities. Functional scales include genome-scale metabolic descriptions of microbial physiology describing each enzyme function, supplemented by molecular insights provided by metatranscriptomes, metaproteomes and metametabolomes, and community-level functions with coarse-grained description of large processes. Effectively navigating across these scales is essential for modeling open microbial systems such as those in plants, lakes, or long-term engineered environments like wastewater treatment plants. A diverse array of mathematical formalisms, including individual-based models, constraint-based metabolic modeling, partial differential equations, ordinary differential equations, stochastic models, and kinetic models, can be employed to address this multiscale complexity (Widder et al., 2016). The process of model fitting (parameter estimation) can be challenging, with key issues to address including structure selection, parameter identifiability and uncertainty quantification in a context of multiple time scales. A variety of methods are available (Banga and Villaverde, 2025; Heinrich et al., 2025) and could be used for model building in the MSDT context.

Managing model modularity. Building a MSDT necessitates efficient procedures to adapt or refresh the model on the fly depending on the system state. First, it should be feasible to easily change the structure of the microbial community model depending on the community composition, by either selecting corresponding GSMM in a bank of pre-built metabolic models or by adapting the ODE state space (biomass and metabolites) in a process-based model. Second, this microbial community model should be easily coupled with various modeling blocks each representing other compartments of the microbial system, such as host tissues and organs (epithelium, mucus, rumen, rhizome, phyllosphere…) or direct environment (river, lake, cheese, wine tank, substrate, waste, stressors…). A hybrid hierarchical multi-model approach could be a practical way to implement such a feature: different models of the same system can be implemented with increasing complexity and accuracy (but decreasing computational efficiency) and an upper modeling layer could dynamically activate the suitable model according to the current needs, seeking for parcimony.

Speeding up computations. For a given model structure, reaching real-time and MSDT inline feedback loops necessitates constant efforts in speeding up computations, since mechanistic models are often computationally intensive and incompatible with real-time constraints in their original form. Computation speed-up can first be analytic. Using theoretical analysis methods such as asymptotic analysis, singular perturbation, scale decomposition or homogeneization, model approximations can be analytically derived providing fast-to-compute proxis of the initial model (Labarthe et al., 2019). Surrogate modeling is a data-driven approach presenting acceleration potential (Frioux et al., 2023b). Based on a learning database constructed with the original model or real data, the relationship between model input and output is learned using dedicated statistical tools (reproducing kernel Hilbert space, polynomial chaos, spline smoothing, proper orthogonal decomposition…) or deep learning. Physic-informed neural networks (PINN) incorporates to the learning database additional conservation laws derived from physics to strengthen approximation accuracy (Hossie et al., 2025). Finally, high performance computing (HPC), which includes both dedicated algorithmic for scientific computing and parallelization methods to scale up computations on large clusters or GPUs, can be used to further speed up computations.

5.5. Key challenges in control engineering

Adapting automatic control theories to modular MSDT. Automatic control is part of the broader field of “Control Science”, a discipline developed since the 1960s across several industries including aeronautics, automobile, chemical processes or bioprocesses (Bastin and Dochain, 2013; d’Andréa-Novel and De Lara, 2013). It introduces a rigorous framework for describing “input-output” systems and the role of feedback mechanisms, including notions such as “real-time control", “output feedback" or “adaptive regulation” (Ariyur and Krstić, 2004). These concepts find a direct transposition in the definition of DTs. However, since they have been defined for a fixed structure of the model under study, further developments are required to extend them for modular MSDT, when there is a heterogeneity of the mathematical representations of the various entities involved in the digital twin and on-the-fly modifications of the model structure. As such, in a hybrid framework an automaton may supervise the activation of modules depending on the system state. An applied example could be a gut microbiota MSDT, with modules describing inflammatory response, epithelial turn-over or the mucus layer that could be enabled or disabled by a supervisor automaton depending on the pathogenic status of the microbiota. Distinct definitions and methods would apply to controllers and feedback on the upper pilot automaton and on the sub-modules, in a context where the model structure is dynamic. Given the complexity of the models underlying DT (hybrid systems, large number of variables, non-linearity), piecewise linear representations (or approximations) of the dynamics, combined with simple controllers in each region, could be used by AI techniques to obtain hybrid controllers (however, the challenge in controlling hybrid systems lies in the fact that switching between two stable models can lead to instability...)

Defining effective observers for microbial system control. To effectively control a microbial system from the available information provided by the sensors, it is often necessary to reconstruct internal state variables of the model required to control the system, and thus to identify an appropriate set of key observables from which this is possible. Indeed, a practical trade-off exists between observation dimensionality (the number of measured variables) and system controllability (the theoretical capacity to guide the system toward a desired state). As the number of variables required to fully characterize the system increases, the challenge of designing an effective control strategy grows. Thus, a crucial objective is to identify a limited number of sensors or state observers that sufficiently capture system dynamics with the precision needed to define and implement efficient control feedbacks.

Multi-objective control under complex constraints. MSDTs are characterized by numerous state variables of diverse nature, making it theoretically and practically challenging to control all of them simultaneously. Meanwhile, experimentalists have limited actuators, i.e. practical control inputs that can be activated to steer the microbial system, such as abiotic parameters (e.g. pH modulation to inhibit specific microbial populations or inflow rate), or biotic levers (e.g. microbial populations or metabolic compounds). Therefore, a key challenge is to identify achievable objectives with the available actuators and designing multi-objective control strategies that navigate trade-offs while coping with the system constraints. For example, consider a MSDT for wine processing. The organoleptic properties of wine are defined by a vast array of compounds, making fine-tuned control of all variables impractical. However, focusing on a subset of key volatile compounds together with the microbial levers managing them could enable targeted trade-offs, ensuring the wine maintains a desired organoleptic profile. In order to adapt to the small numbers

of inputs and outputs, model reduction may be relevant to obtain minimal representations (i.e. controllable and observable models). To this end, variables aggregation could be explored using AI-type techniques. An additional layer of complexity arises from the diverse nature of variables and objectives. Following up with the example of wine processing, one might aim to minimize the energetic cost of cooling wine-making tanks while optimizing fermentation outcomes, or to accelerate fermentation. Controlling simultaneously energetic costs, microbial processes and fermentation duration represents a significant methodological challenge, due to the different—and potentially contradictory—nature of these objectives. Multi-criteria optimization, when criteria are prioritized, could benefit from being tackled through optimization under constraints satisfaction levels. The viability sets associated with these levels could be explored by combining geometric conditions and AI tools (Aubin et al., 2011).

5.6. Key challenges in microbiota engineering

Defining new levers for microbial management. Various levers can be activated for microbial control, including *prebiotics* (nutrients designed for targeted promotion of micro-organisms), *probiotics* (micro-organisms selected for their ability of shaping the community), *postbiotics* (nutritional environment obtained after inoculation of targeted micro-organism), *antibiotics* (compound targeting microbial populations), *phages* (bacterial viruses), *abiotic factors* (acidity, light, heat, salinity, oxygen…) or *nutritional control* (designed substrate to mitigate or enhance populations). More systemic actions can be defined such as ecological levers (use of promoting or competing micro-organisms, management of microbial diversity) or spatial control (shaping of the 3D environment, biofilm promotion or repression) (Delgado-Baquerizo et al., 2025) (see Table 4 for an extended review). Expending the catalog of actionable levers for microbiota engineering is key for developing new MSDT. These levers can be selected in the MSDT control strategy. Their impact on the system must be modeled to determine their optimal use: timing, intensity, duration, and modulation. While multiple levers can enhance control precision, they also increase computational demands for defining an effective command strategy.

Automatization, sensor development and remote control. Automatization of control processes both increase speed and accuracy of microbial manipulation. In microfluidic setups, automatic inoculation or flow modulation is a way to finely manage the microbial system (Boitard et al., 2015). Pipetting robots and miniaturization can also scale up substrate preparation, microbial inoculation and culturing. Developing new sensors (such as in line sequencing or HPLC, MS or optic devices) or improvement of existing ones by miniaturization (e.g. microfluidics-based sample preparation and LC), increasing measurement rates and precision or data collection, will enhance observations. Furthermore, remote screening and control of the MSDT with remote applications (Internet of Objects, IoT) will facilitate human actions on the system (Zhao et al., 2025).

Simplified communities for targeted microbiota engineering. The selection of SynComs is key for targeted ecological engineering. A bottom-up approach involves assembling defined consortia from isolates available in culture collections, focusing on micro-organisms that carry a targeted function and their symbionts, which enhance SynCom stability and resilience (Mehlferber et al., 2024). This ensures the SynCom can express the desired function across diverse contexts. A top-down approach begins with a natural complex community exhibiting the target phenotype. The community is simplified, e.g. through serial dilutions or selective stress, while preserving the

phenotype, followed by sequencing to define the SynCom (Jacquiod et al., 2022). In both strategies, deep characterization of SynCom members is critical for understanding community dynamics and adapting management to specific contexts. Additionally, studying the engraftment of the SynCom into natural communities is crucial to assess its ability to enhance the target function within complex ecosystems. These ecological studies can be complemented by high-throughput phenotyping—such as parallel cultivation in diverse media or random gene knockouts—to further elucidate key organism functions and interactions.

6. Conclusive discussion

6.1. MSDT as a new modeling paradigm

MSDTs represent a novel paradigm moving beyond traditional modeling and control engineering frameworks. It begins with a clear delimitation of the microbiological system, identifying the relevant level of complexity to capture the critical phenomena driving system dynamics. This systemic conceptualization of the biological twin enables the rigorous definition of its digital counterpart. Unlike conventional modeling approaches, the MSDT is not built on a single model but rather on a collection of modular models that can be coupled on the fly as needed. The MSDT is tightly linked to its biological twin through one or more interaction loops, facilitating a bidirectional data and command flow: experimental data feeds the MSDT, while model-based insights guide the steering of the microbial system. These loops can operate in-line (real-time data and command flows), at-line (real-time with occasional human-dependent delays), or off-line (all steps involve time-consuming human-operated processes). These interaction loops are essential for adapting the MSDT to the system specificities, enabling model personalization or microbial ecosystem simplification. We advocate for an application-oriented definition of MSDTs: whenever iterative cycles of data-driven model refinement and model-guided system updates enhance microbial system outcomes, the model qualifies as an MSDT. Such iterations have already been implemented in food microbiology (Helmy et al., 2024; Zhao et al., 2025). However, MSDT research should prioritize accelerating these iterations, with the goal of transitioning off-line loops to at-line, and ultimately to in-line loops in future digital twin versions, together with improving data depth and model accuracy for enhanced MSDT outcome. This evolution, that follows DT maturation schemes (Metcalfe et al., 2023), is key to unlocking the full potential of MSDTs.

6.2. MSDT: a fruitful research object

Originally introduced in product lifecycle management (Grieves, 2023), digital twins have become integral to the manufacturing sector, where they have been widely adopted to enhance the design, production, and maintenance of complex systems (Liu et al., 2021). Their main objective was fostering innovation to achieve operating real-time control of a process. In this paper, we argue for importing MSDT into the microbiology laboratory for fundamental research. A research-oriented MSDT can integrate multiple disciplines, enhancing data flow, modeling, and control to yield deeper insights into microbial systems. The first key objective is leveraging omics and models to better personalize the MSDT by adapting it to system specificities. Second, the MSDT can help identifying patterns to simplify microbial ecosystem description, enabling efficient modeling and control. Finally, accelerating interaction loops is central in MSDT research since faster data, model and control treatment enhances MSDT throughput. While deeper omics

typically conflict with faster pipelines, MSDT research must strike a balance between these competing priorities to achieve both depth and operating efficiency. Microbial systems provide an ideal context for demonstrating the value of digital twins in research: microbiology offers highly controllable, miniaturizable and replicable systems, with rapid and cost-effective experiments compared to other biology fields.

6.3. Boosting interdisciplinary science

MSDTs are intrinsically at the interface of diverse disciplines, spanning experimental microbiology, bioinformatics, data science, mathematical modeling, control and microbiota engineering. Building a MSDT requires establishing a common system biology and ecology framework that federate experimentalists and modelers. This involves defining collections of observables of interest, specifying the model features and designing control strategies to steer the system. This formal definition demands contributions from every discipline along the MSDT pipeline, dismantling traditional silos and fostering interdisciplinary collaborations. It can be the basis of new scientific questions, both driving methodological developments, like deeper and faster analytical methods, and advancing microbial ecology by unravelling the ecological dynamics of the microbial system. From these fundamental scientific advancements, real-world applications emerge, enabling innovative management of microbial systems in practical settings. The acceleration and consolidation of MSDTs will be boosted by the adoption of Open Science practices such as data and code sharing. Interdisciplinary approaches should also include human sciences like economics or sociology to facilitate MSDT transfer out of the laboratory or investigate societal implications of MSDT use.

6.4. Epistemological risks

Beyond their scientific relevance, MSDTs also bring forward epistemological questions about how we define microbial reality. A model explicitly assumes it simplifies reality, whereas the terminology of “twin” suggests the digital construct is a faithful, or even augmented, copy of the microbial system, obscuring the modelling assumptions that a term like “model” already signals. This framing can lead to the erroneous idea that the real system could be replaced by its digital twin, substituting experiments with digital exploration without consequence for knowledge construction. It shifts the phenomenological focus from the “Microbes-in-themselves” to the model as the primary object of microbiological inquiry. Such confusion arises more easily than one might think, particularly when researchers are immersed in data rather than in hands-on experimentation, a tendency consistent with the documented overconfidence of modellers in their own constructs (Puy et al., 2025; Saltelli et al., 2020).

A further risk lies in the way MSDTs implicitly frame microbial ecosystems as systems whose dynamics could be steered and controlled, a mindset inherited from engineering disciplines whose direct transfer to living systems remains questionable. By concentrating on what can be encoded in models, we risk overlooking the inherent complexity of life, and in particular the “messiness” of microbial life that characterise living systems. For these reasons, the simplifications involved in the development of MSDTs, which depend on the questions asked and the objectives pursued, must be made explicit.

6.5. Ethics of digital twins: high tech science for low tech solutions

Digital twins are often presented as high-impact research investments, promising productivity gains and progress towards sustainable development goals (Tzachor et al., 2022b). While they may yield high returns on investment in industrial settings, their large-scale deployment in the agrifood sector beyond controlled environment like greenhouses (Ariesen-Verschuur et al., 2022) raises questions, and profitable use cases in agriculture could stay limited. However, MSDTs hold potential in areas such as fermentation optimization, biomass valorization, water and soil remediation, or pest management. Yet, as modeling increasingly influences policy, there is a risk of overestimating digital twins' potential to accelerate breakthroughs (Saltelli et al., 2025). This places a responsibility on modelers to actively and explicitly guide stakeholders to distinguish between realistic visions and speculative fictions about digital twins (Saltelli et al., 2025). Without careful stewardship, digital twins could fuel unrealistic expectations, echoing past critiques of technosolutionism, a concern already noted in the context of Earth digital twins (Saltelli et al., 2024). This risk is specifically pronouced for complex microbial ecosystems. Since MSDT are explicitely goal-oriented towards anthropocentric ends, they may induce problem framing in which ecological disturbances appear as microbial malfunctions that require optimisation, overlooking underlying ecological drivers: MSDT could thereby perpetuate scenarios where technology is deployed to treat symptoms rather than address root causes.

Nonetheless, laboratory-oriented MSDT research remains valuable for boosting fundamental research and deepening our understanding of microbial communities. MSDTs could be directed towards the identification of simple, scalable management strategies for field applications: by introducing operational constraints to the MSDT development framework, researchers could intentionally prioritize microbial solutions that require minimal technological intervention, bridging the gap between lab innovation and field practicality. In this way, high-tech science in the lab may lead to low-tech, widely accessible solutions in the field. High-tech science for low-tech solutions: this could be a more credible path forward for MSDTs.

Acknowledgements

We thank the Artemis consortium members. All the authors of this paper belong to the Artemis consortium, along with Beatrice Laroche, Romain Briandet, Suzanne Touzeau, Ludovic Mailleret, Frédéric Grognard, Jean-Luc Gouzé, Chabname Ghassemi Nedjad and Laurent Tournier to whom is addressed a special thank for their insightfull participation during the Artemis workshops. We also thank the INRAE Digit-bio metaprogram steering commity members for the fruitful discussions we shared.

Use of AI tools

During the preparation of this manuscript, the Mistral-AI agent has been used for language improvement and chatGPT-4 for a draft version of Table 2, metabolomics. All outputs from these tools were critically reviewed, edited, and approved by the authors to ensure accuracy, integrity, and originality.

CRediTs

The contribution of each author of the Artemis consortium is listed, according to the Contributor Role Taxonomy (CRediT):

S.Labarthe: Funding acquisition, Supervision, Conceptualization, Visualization, Validation, Writing – original draft, Writing – review & editing. N.Creusot: Validation, Writing – original draft, Writing – review & editing. C.Frioux: Conceptualization, Visualization, Validation, Writing – original draft, Writing – review & editing. G.Gautreau: Conceptualization, Visualization, Validation, Writing – original draft, Writing – review & editing. E.Desmond-Le Quemener: Conceptualization, Visualization, Validation, Writing – original draft, Writing – review & editing. L.Sala: Conceptualization, Visualization, Validation, Writing – original draft, Writing – review & editing. A.Yabo: Conceptualization, Validation, Writing – original draft, Writing – review & editing. P.Dou: Visualization, Writing – review & editing. G.Capson-Tojo: Writing – original draft, Writing – review & editing. E.Dugat-Bony: Writing – original draft, Writing – review & editing. A.Fernandez Diaz: Conceptualization, Writing – review & editing. M.Mezache: Conceptualization, Writing – review & editing. C.Muller: Writing – review & editing. S.Tatho: Writing – review & editing. S.Yeghiazarian: Writing – review & editing. J.Harmand: Conceptualization, Validation, Writing – original draft, Writing – review & editing. J.-R.Mouret: Writing – original draft, Writing – review & editing. T.Nidelet: Conceptualization, Writing – review & editing. S.Prigent: Conceptualization, Writing – review & editing. L.Cottret: Conceptualization, Visualization, Validation, Writing – original draft, Writing – review & editing. A.Goelzer: Conceptualization, Visualization, Writing – review & editing. I.Lesur-Kupin: Visualization, Writing – review & editing. L.Rigottier-Gois: Conceptualization, Visualization, Writing – original draft. F.Salin: Visualization, Writing – review & editing. V.Baldazzi: Conceptualization, Writing – review & editing. C.Casenave: Conceptualization, Validation, Writing – review & editing. H.Falentin: Writing – original draft, Writing – review & editing. M.Mariadassou: Validation, Writing – original draft, Writing – review & editing. R.Muñoz-Tamayo: Conceptualization, Writing – original draft, Writing – review & editing. M.Popova: Conceptualization, Writing – original draft, Writing – review & editing. A.Rapaport: Conceptualization, Validation, Writing – original draft, Writing – review & editing. D.Sherman: Conceptualization, Writing – review & editing. J.Steyer: Conceptualization, Validation, Writing – original draft, Writing – review & editing. C.Vacher: Conceptualization, Visualization, Validation, Writing – original draft, Writing – review & editing.

Fundings

All the authors receive fundings from the Digit-bio metaprogram of INRAE institute through the Artemis consortium.

This work has been supported by the French National Research Agency (ANR) France 2030 PEPR Agroécologie et Numérique MISTIC (ANR-22-PEAE-0011), the Cultissimo project (PEPR Systèmes Alimentaires, Microbiome et Santé, CULTISSIMO ANR-24-PESA-0002), the H2Rumen project (ANR-24-CE20-7802), the VITAE project (ANR-20-PCPA-0010), the Digitwine project (ANR-24-CE10-4479) and the MetaboHUB infrastructure (MetaboHUB (ANR-11-INBS-0010). The Conseil Régional d'Aquitaine supported this work through the MicroMod project.

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

The authors declare the following non-financial conflict of interest: Simon Labarthe is recommender of PCI Microbiology and PCI Computational Biology. Rafael Muñoz-Tamayo is member of the managing board of PCI Animal Science and recommender of PCI Microbiology.

Data, script, code, and supplementary information availability

This opinion paper has not necessitated new data. All needed information is included in the main text.

References

Affagard H, Benamouzig R, Bork P, Bull CT, Clement K, De Montera B, De Vos WM, Dominguez-Bello MG, Dore J, Fontaine F, Gao GF, Henn M, Holmes E, Karciauskate I, Knight R, Koren O, Louis P, Maguin E, McAllister TA, Ohno H, et al. (2026). Integrating Microbiomes into One Health: Insights from the 2025 One Health World Microbiome Partnership Summit. *The Lancet Microbe*, 101319. https://doi.org/10.1016/j.lanmic.2025.101319.

Amstalden Van Hove ER, Smith DF, Heeren RM (2010). A Concise Review of Mass Spectrometry Imaging. *Journal of Chromatography A* **1217**, 3946–3954. https://doi.org/10.1016/j.chroma.2010.01.033.

Ariesen-Verschuur N, Verdouw C, Tekinerdogan B (2022). Digital Twins in Greenhouse Horticulture: A Review. *Computers and Electronics in Agriculture* **199**, 107183. https://doi.org/10.1016/j.compag.2022.107183.

Ariyur KB, Krstić M (2004). Slope Seeking: A Generalization of Extremum Seeking. *International Journal of Adaptive Control and Signal Processing* **18**, 1–22. https://doi.org/10.1002/acs.777.

Arnoldini M, Sharma R, Moresi C, Chure G, Chabbey J, Slack E, Cremer J (2025). Quantifying the Varying Harvest of Fermentation Products from the Human Gut Microbiota. *Cell* **188**, 5332–5342.e16. https://doi.org/10.1016/j.cell.2025.07.005.

Aubin JP, Bayen AM, Saint-Pierre P (2011). *Viability theory: new directions*. Springer Science & Business Media.

Bandyopadhyay A, Sarkar D, Das A, Das A (2025). Intersections of ABO Blood Group, Secretor Status, and the Gut Microbiome: Implications for Disease Susceptibility and Therapeutics. *Archives of Microbiology* **207**, 296. https://doi.org/10.1007/s00203-025-04515-9.

Banerjee S, van der Heijden MGA (2023). Soil Microbiomes and One Health. *Nature Reviews Microbiology* **21**, 6–20. https://doi.org/10.1038/s41579-022-00779-w.

Banga JR, Villaverde AF (2025). Mechanistic Dynamic Modelling of Biological Systems: The Road Ahead. *Current Opinion in Systems Biology* **42**, 100553. https://doi.org/10.1016/j.coisb.2025.100553.

Baranwal M, Clark RL, Thompson J, Sun Z, Hero AO, Venturelli OS (2022). Recurrent Neural Networks Enable Design of Multifunctional Synthetic Human Gut Microbiome Dynamics. *eLife* **11**, e73870. https://doi.org/10.7554/eLife.73870.

Barbosa MI, Silva G, Ribeiro P, Vieira E, Perrotta A, Moreira P, Rodrigues PM (2025). Unraveling the Microbiome–Environmental Change Nexus to Contribute to a More Sustainable World: A Comprehensive Review of Artificial Intelligence Approaches. *Sustainability* **17**, 7209. https://doi.org/10.3390/su17167209.

Bastin G, Dochain D (2013). *On-Line Estimation and Adaptive Control of Bioreactors*.

Batstone D, Keller J, Steyer J (2006). A Review of ADM1 Extensions, Applications, and Analysis: 2002–2005. *Water Science and Technology* **54**, 1–10. https://doi.org/10.2166/wst.2006.520.

Batstone DJ, Keller J, Angelidaki I, Kalyuzhnyi SV, Pavlostathis SG, Rozzi A, Sanders WTM, Siegrist HA, Vavilin VA (2002). The IWA Anaerobic Digestion Model No 1 (ADM1). *Water Science and technology* **45**, 65–73.

Bauer P, Stevens B, Hazeleger W (2021). A Digital Twin of Earth for the Green Transition. *Nature Climate Change* **11**, 80–83. https://doi.org/10.1038/s41558-021-00986-y.

Belcour A, Frioux C, Aite M, Bretaudeau A, Hildebrand F, Siegel A (2020). Metage2Metabo, Microbiota-Scale Metabolic Complementarity for the Identification of Key Species. *eLife* **9**, e61968. https://doi.org/10.7554/eLife.61968.

Björnsson B, Borrebaeck C, Elander N, Gasslander T, Gawel DR, Gustafsson M, Jörnsten R, Lee EJ, Li X, Lilja S, Martínez-Enguita D, Matussek A, Sandström P, Schäfer S, Stenmarker M, Sun XF, Sysoev O, Zhang H, Benson M, on behalf of the Swedish Digital Twin Consortium (2020).

Digital Twins to Personalize Medicine. *Genome Medicine* **12**, 4. https://doi.org/10.1186/s13073-019-0701-3.

Boer MD, Melkonian C, Zafeiropoulos H, Haas AF, Garza DR, Dutilh BE (2024). Improving Genome-Scale Metabolic Models of Incomplete Genomes with Deep Learning. *iScience* **27**, 111349. https://doi.org/10.1016/j.isci.2024.111349.

Boitard L, Cottinet D, Bremond N, Baudry J, Bibette J (2015). Growing Microbes in Millifluidic Droplets. *Engineering in Life Sciences* **15**, 318–326. https://doi.org/10.1002/elsc.201400089.

Bouillaud D, Farjon J, Gonçalves O, Giraudeau P (2019). Benchtop NMR for the Monitoring of Bioprocesses. *Magnetic Resonance in Chemistry* **57**, 794–804. https://doi.org/10.1002/mrc.4821.

Brixi G, Durrant MG, Ku J, Poli M, Brockman G, Chang D, Gonzalez GA, King SH, Li DB, Merchant AT, Naghipourfar M, Nguyen E, Ricci-Tam C, Romero DW, Sun G, Taghibakshi A, Vorontsov A, Yang B, Deng M, Gorton L, et al. (2025). *Genome Modeling and Design across All Domains of Life with Evo 2*. https://doi.org/10.1101/2025.02.18.638918. Pre-published.

Bunne C, Roohani Y, Rosen Y, Gupta A, Zhang X, Roed M, Alexandrov T, AlQuraishi M, Brennan P, Burkhardt DB, Califano A, Cool J, Dernburg AF, Ewing K, Fox EB, Haury M, Herr AE, Horvitz E, Hsu PD, Jain V, et al. (2024). How to build the virtual cell with artificial intelligence: Priorities and opportunities. *Cell* **187**, 7045–7063. https://doi.org/10.1016/j.cell.2024.11.015.

Cai J, Henion J (1995). Capillary Electrophoresis-Mass Spectrometry. *Journal of Chromatography A* **703**, 667–692. https://doi.org/10.1016/0021-9673(94)01178-H.

Capson-Tojo G, Astals S, Robles Á (2021). Considering Syntrophic Acetate Oxidation and Ionic Strength Improves the Performance of Models for Food Waste Anaerobic Digestion. *Bioresource Technology* **341**, 125802. https://doi.org/10.1016/j.biortech.2021.125802.

Carding S, Verbeke K, Vipond DT, Corfe BM, Owen LJ (2015). Dysbiosis of the Gut Microbiota in Disease. *Microbial Ecology in Health & Disease* **26**. https://doi.org/10.3402/mehd.v26.26191.

Carlino N, Blanco-Míguez A, Punčochář M, Mengoni C, Pinto F, Tatti A, Manghi P, Armanini F, Avagliano M, Barcenilla C, Breselge S, Cabrera-Rubio R, Calvete-Torre I, Coakley M, Cobo-Díaz JF, De Filippis F, Dey H, Leech J, Klaassens ES, Knobloch S, et al. (2024). Unexplored microbial diversity from 2, 500 food metagenomes and links with the human microbiome. *Cell* **187**, 5775–5795.e15. https://doi.org/10.1016/j.cell.2024.07.039.

Cerk K, Ugalde-Salas P, Nedjad CG, Lecomte M, Muller C, Sherman DJ, Hildebrand F, Labarthe S, Frioux C (2024). Community-scale Models of Microbiomes: Articulating Metabolic Modelling and Metagenome Sequencing. *Microbial Biotechnology* **17**, e14396. https://doi.org/10.1111/1751-7915.14396.

Chialva M, Lanfranco L, Bonfante P (2022). The Plant Microbiota: Composition, Functions, and Engineering. *Current Opinion in Biotechnology* **73**, 135–142. https://doi.org/10.1016/j.copbio.2021.07.003.

Choi S, Hwang HS, Lee HS (2026). Designing Field-Ready Biocontrol: Discovery–Syncom–Validation. *Plant Biotechnology Reports* **20**, 1. https://doi.org/10.1007/s11816-025-01029-0.

Colangelo CM, Chung L, Bruce C, Cheung KH (2013). Review of Software Tools for Design and Analysis of Large Scale MRM Proteomic Datasets. *Methods* **61**, 287–298. https://doi.org/10.1016/j.ymeth.2013.05.004.

Compant S, Cassan F, Kostić T, Johnson L, Brader G, Trognitz F, Sessitsch A (2025). Harnessing the Plant Microbiome for Sustainable Crop Production. *Nature Reviews Microbiology* **23**, 9–23. https://doi.org/10.1038/s41579-024-01079-1.

Cooks RG, Ouyang Z, Takats Z, Wiseman JM (2006). Ambient Mass Spectrometry. *Science* **311**, 1566–1570. https://doi.org/10.1126/science.1119426.

Cortada-Garcia J, Haggarty J, Weidt S, Daly R, Arnold SA, Burgess K (2024). On-line Targeted Metabolomics for Real-time Monitoring of Relevant Compounds in Fermentation Processes. *Biotechnology and Bioengineering* **121**, 683–695. https://doi.org/10.1002/bit.28599.

Cremer J, Arnoldini M, Hwa T (2017). Effect of Water Flow and Chemical Environment on Microbiota Growth and Composition in the Human Colon. *Proceedings of the National Academy of Sciences* **114**, 6438–6443. https://doi.org/10.1073/pnas.1619598114.

Creusot N, Bellvert F, Bertrand S, Bertrand C, Colas S, Cottret L, Diémé B, Eon M, Hubas C, Goddard ML, Jousse CC, Marie B, Marti G, Millard P, Pétriacq P, Valls Fonayet J, Salvia MV, Schmitt-Jansen M, Zi L (2025). Unraveling the Microbiome-Metabolome Nexus for Innovative One-Health Solution.

Crouch LI, Rodrigues CS, Bakshani CR, Tavares-Gomes L, Gaifem J, Pinho SS (2024). The Role of Glycans in Health and Disease: Regulators of the Interaction between Gut Microbiota and Host Immune System. *Seminars in Immunology* **73**, 101891. https://doi.org/10.1016/j.smim.2024.101891.

d'Andréa-Novel B, De Lara M (2013). *Control Theory for Engineers: A Primer*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-34324-7.

Daims H, Wagner M (2007). Quantification of Uncultured Microorganisms by Fluorescence Microscopy and Digital Image Analysis. *Applied Microbiology and Biotechnology* **75**, 237–248. https://doi.org/10.1007/s00253-007-0886-z.

Daneshgar S, Polesel F, Borzooei S, Sørensen HR, Peeters R, Weijers S, Nopens I, Torfs E (2024). A Full-scale Operational Digital Twin for a Water Resource Recovery Facility—A Case Study of Eindhoven Water Resource Recovery Facility. *Water Environment Research* **96**, e11016. https://doi.org/10.1002/wer.11016.

Darrigade L, Haghebaert M, Cherbuy C, Labarthe S, Laroche B (2022). A PDMP Model of the Epithelial Cell Turn-over in the Intestinal Crypt Including Microbiota-Derived Regulations. *Journal of Mathematical Biology* **84**, 60. https://doi.org/10.1007/s00285-022-01766-8.

Davoudkhani M, Rubino F, Creevey CJ, Ahvenjärvi S, Bayat AR, Tapio I, Belanche A, Muñoz-Tamayo R (2024). Integrating Microbial Abundance Time Series with Fermentation Dynamics of the Rumen Microbiome via Mathematical Modelling. *PLOS ONE* **19**. Ed. by Adham A. Al-Sagheer, e0298930. https://doi.org/10.1371/journal.pone.0298930.

De Domenico M, Allegri L, Caldarelli G, d'Andrea V, Di Camillo B, Rocha LM, Rozum J, Sbarbati R, Zambelli F (2025). Challenges and Opportunities for Digital Twins in Precision Medicine from a Complex Systems Perspective. *npj Digital Medicine* **8**, 37. https://doi.org/10.1038/s41746-024-01402-3.

Delgado-Baquerizo M, Singh BK, Liu YR, Sáez-Sandino T, Coleine C, Muñoz-Rojas M, Bastida F, Trivedi P (2025). Integrating Ecological and Evolutionary Frameworks for SynCom Success. *New Phytologist* **246**, 1922–1933. https://doi.org/10.1111/nph.70112.

Dillard LR, Payne DD, Papin JA (2021). Mechanistic Models of Microbial Community Metabolism. *Molecular Omics* **17**, 365–375. https://doi.org/10.1039/D0MO00154F.

Ding Z, Wang W, Zhang K, Ming F, Yangdai T, Xu T, Shi H, Bao Y, Yao H, Peng H, Han C, Jiang W, Liu J, Hou X, Lin R (2021). Novel Scheme for Non-Invasive Gut Bioinformation Acquisition with a Magnetically Controlled Sampling Capsule Endoscope. *Gut* **70**, 2297–2306. https://doi.org/10.1136/gutjnl-2020-322465.

Fakih I, Got J, Robles-Rodriguez CE, Siegel A, Forano E, Muñoz-Tamayo R (2023). **Dynamic Genome-Based Metabolic Modeling of the Predominant Cellulolytic Rumen Bacterium *Fibrobacter Succinogenes* S85**. *mSystems* **8**. Ed. by Vanni Bucci, e01027–22. https://doi.org/10.1128/msystems.01027-22.

Falkowski PG, Fenchel T, Delong EF (2008). The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **320**, 1034–1039. https://doi.org/10.1126/science.1153213.

Fan Y, Pedersen O (2021). Gut Microbiota in Human Metabolic Health and Disease. *Nature Reviews Microbiology* **19**, 55–71. https://doi.org/10.1038/s41579-020-0433-9.

Feierabend M, Töpfer N (2025). *In Silico* Encounters: Harnessing Metabolic Modelling to Understand Plant–Microbe Interactions. *FEMS Microbiology Reviews* **49**, fuaf030. https://doi.org/10.1093/femsre/fuaf030.

Fiehn O (2016). Metabolomics by Gas Chromatography–Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Current Protocols in Molecular Biology* **114**. https://doi.org/10.1002/0471142727.mb3004s114.

Flores-Alsina X, Solon K, Kazadi Mbamba C, Tait S, Gernaey KV, Jeppsson U, Batstone DJ (2016). Modelling Phosphorus (P), Sulfur (S) and Iron (Fe) Interactions for Dynamic Simulations of Anaerobic Digestion Processes. *Water Research* **95**, 370–382. https://doi.org/10.1016/j.watres.2016.03.012.

French E, Kaplan I, Iyer-Pascuzzi A, Nakatsu CH, Enders L (2021). Emerging Strategies for Precision Microbiome Management in Diverse Agroecosystems. *Nature Plants* **7**, 256–267. https://doi.org/10.1038/s41477-020-00830-9.

Frioux C, Ansorge R, Özkurt E, Ghassemi Nedjad C, Fritscher J, Quince C, Waszak SM, Hildebrand F (2023a). Enterosignatures Define Common Bacterial Guilds in the Human Gut Microbiome. *Cell Host & Microbe* **31**, 1111–1125.e6. https://doi.org/10.1016/j.chom.2023.05.024.

Frioux C, Huet S, Labarthe S, Martinelli J, Malou T, Sherman D, Taupin ML, Ugalde-Salas P (2023b). Accelerating Metabolic Models Evaluation with Statistical Metamodels: Application to *Salmonella* Infection Models. *ESAIM: Proceedings and Surveys* **73**. Ed. by Virginie Ehrlacher, Damiano Lombardi, Olga Mula, Fabio Nobile, and Tommaso Taddei, 187–217. https://doi.org/10.1051/proc/202373187.

Fuller A, Fan Z, Day C, Barlow C (2020). Digital Twin: Enabling Technologies, Challenges and Open Research. *IEEE Access* **8**, 108952–108971. https://doi.org/10.1109/ACCESS.2020.2998358.

Gao W, Zhang W, Meldrum DR (2011). RT-qPCR Based Quantitative Analysis of Gene Expression in Single Bacterial Cells. *Journal of Microbiological Methods* **85**, 221–227. https://doi.org/10.1016/j.mimet.2011.03.008.

Gitai Z (2009). New Fluorescence Microscopy Methods for Microbiology: Sharper, Faster, and Quantitative. *Current Opinion in Microbiology* **12**, 341–346. https://doi.org/10.1016/j.mib.2009.03.001.

Grandi FC, Modi H, Kampman L, Corces MR (2022). Chromatin Accessibility Profiling by ATAC-seq. *Nature Protocols* **17**, 1518–1552. https://doi.org/10.1038/s41596-022-00692-9.

Grieves MW (2023). Digital Twins: Past, Present, and Future. In: *The Digital Twin*. Ed. by Noel Crespi, Adam T. Drobot, and Roberto Minerva. Cham: Springer International Publishing, pp. 97–121. https://doi.org/10.1007/978-3-031-21343-4_4.

Haghebaert M, Laroche B, Sala L, Mondot S, Doré J (2024). A Mechanistic Modelling Approach of the Host–Microbiota Interactions to Investigate Beneficial Symbiotic Resilience in the Human Gut. *Journal of The Royal Society Interface* **21**, 20230756. https://doi.org/10.1098/rsif.2023.0756.

Hajjar G, Barros Santos MC, Bertrand-Michel J, Canlet C, Castelli F, Creusot N, Dechaumet S, Diémé B, Giacomoni F, Giraudeau P, Guitton Y, Thévenot E, Tremblay-Franco M, Junot C, Jourdan F, Fenaille F, Comte B, Pétriacq P, Pujos-Guillot E (2023). Scaling-up Metabolomics: Current State and Perspectives. *TrAC Trends in Analytical Chemistry* **167**, 117225. https://doi.org/10.1016/j.trac.2023.117225.

Hammond KJ, Waghorn GC, Hegarty RS (2016). The GreenFeed System for Measurement of Enteric Methane Emission from Cattle. *Animal Production Science* **56**, 181–189. https://doi.org/10.1071/AN15631.

Han CS, Kaur U, Bai H, Roqueto Dos Reis B, White R, Nawrocki RA, Voyles RM, Kang MG, Priya S (2022). Invited Review: Sensor Technologies for Real-Time Monitoring of the Rumen Environment. *Journal of Dairy Science* **105**, 6379–6404. https://doi.org/10.3168/jds.2021-20576.

Harlé O, Parayre S, Maillard MB, Henry G, Guédon É, Thierry A, Niay J, Deutsch SM, Falentin H (2025). Fermentation of Soy Juice by *Lactiplantibacillus Plantarum* CIRM-BIA777 Produces Flavor-Related and Health-Promoting Metabolites. *Food Frontiers* **6**, 3137–3153. https://doi.org/10.1002/fft2.70118.

Hassam S, Ficara E, Leva A, Harmand J (2015). A Generic and Systematic Procedure to Derive a Simplified Model from the Anaerobic Digestion Model No. 1 (ADM1). *Biochemical Engineering Journal* **99**, 193–203. https://doi.org/10.1016/j.bej.2015.03.007.

Heinken A, Hertel J, Acharya G, Ravcheev DA, Nyga M, Okpala OE, Hogan M, Magnúsdóttir S, Martinelli F, Nap B, Preciat G, Edirisinghe JN, Henry CS, Fleming RMT, Thiele I (2023). Genome-Scale Metabolic Reconstruction of 7,302 Human Microorganisms for Personalized Medicine. *Nature Biotechnology*, 1–12. https://doi.org/10.1038/s41587-022-01628-0. PMID: 36658342.

Heinrich M, Arutjunjan R, Timmer J (2025). On the Different Flavours of Practical Identifiability. *Current Opinion in Systems Biology* **42**, 100556. https://doi.org/10.1016/j.coisb.2025.100556.

Hellal J, Barthelmebs L, Bérard A, Cébron A, Cheloni G, Colas S, Cravo-Laureau C, De Clerck C, Gallois N, Hery M, Martin-Laurent F, Martins J, Morin S, Palacios C, Pesce S, Richaume A, Vuilleumier S (2023). Unlocking Secrets of Microbial Ecotoxicology: Recent Achievements and Future Challenges. *FEMS Microbiology Ecology* **99**, fiad102. https://doi.org/10.1093/femsec/fiad102.

Helmy M, Elhalis H, Rashid MM, Selvarajoo K (2024). Can Digital Twin Efforts Shape Microorganism-Based Alternative Food? *Current Opinion in Biotechnology* **87**, 103115. https://doi.org/10.1016/j.copbio.2024.103115.

Hernández Medina R, Kutuzova S, Nielsen KN, Johansen J, Hansen LH, Nielsen M, Rasmussen S (2022). Machine Learning and Deep Learning Applications in Microbiome Research. *ISME Communications* **2**, 98. https://doi.org/10.1038/s43705-022-00182-9.

Hossie PJ, Laroche B, Malou T, Perrin L, Saigre T, Sala L (2025). Surrogate Modeling of Interactions in Microbial Communities through Physics-Informed Neural Networks. *ESAIM: Proceedings and Surveys* **81**. Ed. by Didier Auroux, Martin Campos Pinto, Bruno Després, Victorita Dolean, Stéphane Lanteri, and Victor Michel-Dansac, 104–122. https://doi.org/10.1051/proc/202581104.

Hou J, Wang C, Rozenbaum RT, Gusnaniar N, De Jong ED, Woudstra W, Geertsema-Doornbusch GI, Atema-Smit J, Sjollema J, Ren Y, Busscher HJ, Van Der Mei HC (2019). Bacterial Density and Biofilm Structure Determined by Optical Coherence Tomography. *Scientific Reports* **9**, 9794. https://doi.org/10.1038/s41598-019-46196-7.

Huang Y, Sheth RU, Zhao S, Cohen LA, Dabaghi K, Moody T, Sun Y, Ricaurte D, Richardson M, Velez-Cortes F, Blazejewski T, Kaufman A, Ronda C, Wang HH (2023). High-Throughput Microbial Culturomics Using Automation and Machine Learning. *Nature Biotechnology* **41**, 1424–1433. https://doi.org/10.1038/s41587-023-01674-2.

Huws SA, Creevey CJ, Oyama LB, Mizrahi I, Denman SE, Popova M, Muñoz-Tamayo R, Forano E, Waters SM, Hess M, Tapio I, Smidt H, Krizsan SJ, Yáñez-Ruiz DR, Belanche A, Guan L, Gruninger RJ, McAllister TA, Newbold CJ, Roehe R, et al. (2018). Addressing Global Ruminant Agricultural Challenges Through Understanding the Rumen Microbiome: Past, Present, and Future. *Frontiers in Microbiology* **9**, 2161. https://doi.org/10.3389/fmicb.2018.02161.

Iqbal S, Begum F, Nguchu BA, Claver UP, Shaw P (2025). The Invisible Architects: Microbial Communities and Their Transformative Role in Soil Health and Global Climate Changes. *Environmental Microbiome* **20**, 36. https://doi.org/10.1186/s40793-025-00694-6.

Islam M, Räisänen S, Schudel A, Wang K, He T, Kunz C, Li Y, Ma X, Serviento A, Zeng Z, Wahl F, Zenobi R, Giannoukos S, Niu M (2024). Exhalomics as a Noninvasive Method for Assessing Rumen Fermentation in Dairy Cows: Can Exhaled-Breath Metabolomics Replace Rumen Sampling? *Journal of Dairy Science* **107**, 2099–2110. https://doi.org/10.3168/jds.2023-24124.

Jacquiod S, Spor A, Wei S, Munkager V, Bru D, Sørensen SJ, Salon C, Philippot L, Blouin M (2022). Artificial Selection of Stable Rhizosphere Microbiota Leads to Heritable Plant Phenotype Changes. *Ecology Letters* **25**. Ed. by Marc-André Selosse, 189–201. https://doi.org/10.1111/ele.13916.

Jorge-Smeding E, Martin C, Volmerange L, Violleau F, Salah N, Silberberg M (2025). Volatolomics of Peripheral Matrices as a Potential Tool to Assess Rumen Function and Host Biology in Lactating Dairy Cows Fed Diets Contrasting in Fiber and Starch Content. *Journal of Dairy Science* **108**, 6917–6933. https://doi.org/10.3168/jds.2024-25977.

Karimi E, Tap J, Champomier-Vergès MC, Chaillou S (2026). Microbiome Metabolic Modeling as a Tool for Innovation in Fermented Foods. *Current Opinion in Food Science* **67**, 101368. https://doi.org/10.1016/j.cofs.2025.101368.

Kettle H, Holtrop G, Louis P, Flint HJ (2018). microPop: Modelling Microbial Populations and Communities in R. *Methods in Ecology and Evolution* **9**. Ed. by Nick Golding, 399–409. https://doi.org/10.1111/2041-210X.12873.

Kleiner M (2019). Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities. *mSystems* **4**, e00115–19. https://doi.org/10.1128/mSystems.00115-19.

Kuhn M, Schmidt TSB, Ferretti P, Głazek A, Robbani SM, Akanni W, Fullam A, Schudoma C, Cetin E, Hassan M, Noack K, Schwarz A, Thielemann R, Thomas L, von Stetten M, Alves R, Iyappan A, Kartal E, Kel I, Keller MI, et al. (2026). Metalog: curated and harmonised contextual data for global metagenomics samples. *Nucleic Acids Res.* **54**, D826–D834.

Labarthe S, Plancade S, Raguideau S, Plaza Oñate F, Le Chatelier E, Leclerc M, Laroche B (2023). Four Functional Profiles for Fibre and Mucin Metabolism in the Human Gut Microbiome. *Microbiome* **11**, 231. https://doi.org/10.1186/s40168-023-01667-y.

Labarthe S, Polizzi B, Phan T, Goudon T, Ribot M, Laroche B (2019). A Mathematical Model to Investigate the Key Drivers of the Biogeography of the Colon Microbiota. *Journal of Theoretical Biology* **462**, 552–581. https://doi.org/10.1016/j.jtbi.2018.12.009.

Lami R, Urios L, Molmeret M, Grimaud R (2023). Quorum Sensing in Biofilms: A Key Mechanism to Target in Ecotoxicological Studies. *Critical Reviews in Microbiology* **49**, 786–804. https://doi.org/10.1080/1040841X.2022.2142089.

Laubenbacher R, Mehrad B, Shmulevich I, Trayanova N (2024). Digital Twins in Medicine. *Nature Computational Science* **4**, 184–191. https://doi.org/10.1038/s43588-024-00607-6.

Lecomte M, Cao W, Aubert J, Sherman DJ, Falentin H, Frioux C, Labarthe S (2024). Revealing the Dynamics and Mechanisms of Bacterial Interactions in Cheese Production with Metabolic Modelling. *Metabolic Engineering*, S1096717624000302. https://doi.org/10.1016/j.ymben.2024.02.014.

Lehner H, Dorffner L (2020). Digital geoTwin Vienna: Towards a Digital Twin City as Geodata Hub. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science* **88**, 63–75. https://doi.org/10.1007/s41064-020-00101-4.

Li L, Nielsen J, Chen Y (2025a). Personalized Gut Microbial Community Modeling by Leveraging Genome-Scale Metabolic Models and Metagenomics. *Current Opinion in Biotechnology* **91**, 103248. https://doi.org/10.1016/j.copbio.2024.103248.

Li Y, Gu Y, Cheng W, Li Z, Zhang X, Zhao Y, Ko K, Liu W, Liu X, Li H (2025b). Fermentation Modeling and Machine Learning for Flavor Prediction in Low-Sodium Radish Paocai with Potassium Chloride Substitution. *npj Science of Food* **9**, 156. https://doi.org/10.1038/s41538-025-00528-2.

Liu M, Fang S, Dong H, Xu C (2021). Review of Digital Twin about Concepts, Technologies, and Industrial Applications. *Journal of Manufacturing Systems* **58**, 346–361. https://doi.org/10.1016/j.jmsy.2020.06.017.

Lloréns-Rico V, Simcock JA, Huys GR, Raes J (2022). Single-Cell Approaches in Human Microbiome Research. *Cell* **185**, 2725–2738. https://doi.org/10.1016/j.cell.2022.06.040.

Ma L, Zou D, Liu L, Shireen H, Abbasi AA, Bateman A, Xiao J, Zhao W, Bao Y, Zhang Z (2023). Database Commons: A Catalog of Worldwide Biological Databases. *Genomics, Proteomics & Bioinformatics* **21**, 1054–1058. https://doi.org/10.1016/j.gpb.2022.12.004.

Marjani M, Nasaruddin F, Gani A, Karim A, Hashem IAT, Siddiqa A, Yaqoob I (2017). Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges. *IEEE access : practical innovations, open solutions* **5**, 5247–5261. https://doi.org/10.1109/ACCESS.2017.2689040.

Mataigne V, Vannier N, Vandenkoornhuyse P, Hacquard S (2021). Microbial Systems Ecology to Understand Cross-Feeding in Microbiomes. *Frontiers in Microbiology* **12**, 780469. https://doi.org/10.3389/fmicb.2021.780469.

McCallum G, Tropini C (2024). The Gut Microbiota and Its Biogeography. *Nature Reviews Microbiology* **22**, 105–118. https://doi.org/10.1038/s41579-023-00969-0.

Medearis NA, Zhu S, Zomorrodi AR (2026). BiomeGPT: A foundation model for the human gut microbiome.

Mehlferber EC, Arnault G, Joshi B, Partida-Martinez LP, Patras KA, Simonin M, Koskella B (2024). A Cross-Systems Primer for Synthetic Microbial Communities. *Nature Microbiology* **9**, 2765–2773. https://doi.org/10.1038/s41564-024-01827-2.

Melkonian C, Zorrilla F, Kjærbølling I, Blasche S, Machado D, Junge M, Sørensen KI, Andersen LT, Patil KR, Zeidan AA (2023). Microbial Interactions Shape Cheese Flavour Formation. *Nature Communications* **14**, 8348. https://doi.org/10.1038/s41467-023-41059-2.

MetaHIT Consortium, Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, et al. (2011). Enterotypes of the Human Gut Microbiome. *Nature* **473**, 174–180. https://doi.org/10.1038/nature09944.

Metcalfe B, Boshuizen HC, Bulens J, Koehorst JJ (2023). Digital Twin Maturity Levels: A Theoretical Framework for Defining Capabilities and Goals in the Life and Environmental Sciences. *F1000Research* **12**, 961. https://doi.org/10.12688/f1000research.137262.1.

Meyer F, Fritz A, Deng ZL, Koslicki D, Lesker TR, Gurevich A, Robertson G, Alser M, Antipov D, Beghini F, Bertrand D, Brito JJ, Brown CT, Buchmann J, Buluç A, Chen B, Chikhi R, Clausen PTLC, Cristian A, Dabrowski PW, et al. (2022a). Critical Assessment of Metagenome Interpretation: The Second Round of Challenges. *Nature Methods* **19**, 429–440. https://doi.org/10.1038/s41592-022-01431-4.

Meyer KM, Porch R, Muscettola IE, Vasconcelos ALS, Sherman JK, Metcalf CJE, Lindow SE, Koskella B (2022b). Plant Neighborhood Shapes Diversity and Reduces Interspecific Variation of the Phyllosphere Microbiome. *The ISME Journal* **16**, 1376–1387. https://doi.org/10.1038/s41396-021-01184-6.

Monod J (1949). THE GROWTH OF BACTERIAL CULTURES. *Annual Review of Microbiology* **3**, 371–394. https://doi.org/10.1146/annurev.mi.03.100149.002103.

Moorthy AS, Brooks SPJ, Kalmokoff M, Eberl HJ (2015). A Spatially Continuous Model of Carbohydrate Digestion and Transport Processes in the Colon. *PLOS ONE* **10**. Ed. by Peter E. Larsen, e0145309. https://doi.org/10.1371/journal.pone.0145309.

Moreau D, Ballini E, Chave M, Cordeau S, Djian-Caporalino C, Lavoir AV, Suffert F, Cortesero AM (2025). Potential of Service Plants for Regulating Multiple Pests While Limiting Disservices in Agroecosystems. A Review. *Agronomy for Sustainable Development* **45**, 38. https://doi.org/10.1007/s13593-025-01031-4.

Morin S, Artigas J (2023). Twenty Years of Research in Ecosystem Functions in Aquatic Microbial Ecotoxicology. *Environmental Toxicology and Chemistry* **42**, 1867–1888. https://doi.org/10.1002/etc.5708.

Muller EE, Faust K, Widder S, Herold M, Martínez Arbas S, Wilmes P (2018). Using Metabolic Networks to Resolve Ecological Properties of Microbiomes. *Current Opinion in Systems Biology* **8**, 73–80. https://doi.org/10.1016/j.coisb.2017.12.004.

Müller S, Nebe-von-Caron G (2010). Functional Single-Cell Analyses: Flow Cytometry and Cell Sorting of Microbial Populations and Communities. *FEMS Microbiology Reviews* **34**, 554–587. https://doi.org/10.1111/j.1574-6976.2010.00214.x.

Muñoz-Tamayo R, Davoudkhani M, Fakih I, Robles-Rodriguez C, Rubino F, Creevey C, Forano E (2023). Review: Towards the next-Generation Models of the Rumen Microbiome for Enhancing Predictive Power and Guiding Sustainable Production Strategies. *animal* **17**, 100984. https://doi.org/10.1016/j.animal.2023.100984.

Muñoz-Tamayo R, Chagas JC, Ramin M, Krizsan SJ (2021). Modelling the Impact of the Macroalgae Asparagopsis Taxiformis on Rumen Microbial Fermentation and Methane Production. *Peer Community Journal* **1**, e7. https://doi.org/10.24072/pcjournal.11.

Muñoz-Tamayo R, Giger-Reverdin S, Sauvant D (2016). Mechanistic Modelling of in Vitro Fermentation and Methane Production by Rumen Microbiota. *Animal Feed Science and Technology* **220**, 1–21. https://doi.org/10.1016/j.anifeedsci.2016.07.005.

Muñoz-Tamayo R, Laroche B, Walter É, Doré J, Leclerc M (2010). Mathematical Modelling of Carbohydrate Degradation by Human Colonic Microbiota. *Journal of Theoretical Biology* **266**, 189–201. https://doi.org/10.1016/j.jtbi.2010.05.040.

Mytilinaios I, Salih M, Schofield H, Lambert R (2012). Growth Curve Prediction from Optical Density Data. *International Journal of Food Microbiology* **154**, 169–176. https://doi.org/10.1016/j.ijfoodmicro.2011.12.035.

Nagana Gowda GA, Raftery D (2021). NMR-Based Metabolomics. In: *Cancer Metabolomics*. Ed. by Shen Hu. Vol. 1280. Cham: Springer International Publishing, pp. 19–37. https://doi.org/10.1007/978-3-030-51652-9_2.

Oh M, Zhang L (2020). DeepMicro: Deep Representation Learning for Disease Prediction Based on Microbiome Data. *Scientific Reports* **10**, 6026. https://doi.org/10.1038/s41598-020-63159-5.

Olsen RA, Bakken LR (1987). Viability of Soil Bacteria: Optimization of Plate-Counting Technique and Comparison between Total Counts and Plate Counts within Different Size Groups. *Microbial Ecology* **13**, 59–74. https://doi.org/10.1007/BF02014963.

Pacheco AR, Vorholt JA (2023). Resolving Metabolic Interaction Mechanisms in Plant Microbiomes, 9 p. https://doi.org/10.3929/ETHZ-B-000608898. HDL: 20.500.11850/608898.

Pan X, Zhao L, Li C, Angelidaki I, Lv N, Ning J, Cai G, Zhu G (2021). Deep Insights into the Network of Acetate Metabolism in Anaerobic Digestion: Focusing on Syntrophic Acetate Oxidation and Homoacetogenesis. *Water Research* **190**, 116774. https://doi.org/10.1016/j.watres.2020.116774.

Pascual-García A, Bonhoeffer S, Bell T (2020). Metabolically Cohesive Microbial Consortia and Ecosystem Functioning. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**, 20190245. https://doi.org/10.1098/rstb.2019.0245.

Pasolli E, Schiffer L, Manghi P, Renson A, Obenchain V, Truong DT, Beghini F, Malik F, Ramos M, Dowd JB, Huttenhower C, Morgan M, Segata N, Waldron L (2017). Accessible, curated metagenomic data through ExperimentHub. *Nat. Methods* **14**, 1023–1024.

Paulay A, Grimaud GM, Caballero R, Laroche B, Leclerc M, Labarthe S, Maguin E (2024). Design of a Proteolytic Module for Improved Metabolic Modeling of *Bacteroides Caccae*. *mSystems* **9**. Ed. by Nicholas Chia, e00153–24. https://doi.org/10.1128/msystems.00153-24.

Pauvert C, Buée M, Laval V, Edel-Hermann V, Fauchery L, Gautier A, Lesur I, Vallance J, Vacher C (2019). Bioinformatics Matters: The Accuracy of Plant and Soil Fungal Community Data Is Highly Dependent on the Metabarcoding Pipeline. *Fungal Ecology* **41**, 23–33. https://doi.org/10.1016/j.funeco.2019.03.005.

Philippot L, Griffiths BS, Langenheder S (2021). Microbial Community Resilience across Ecosystems and Multiple Disturbances. *Microbiology and Molecular Biology Reviews* **85**, e00026–20. https://doi.org/10.1128/MMBR.00026-20.

Pop M, Attwood TK, Blake JA, Bourne PE, Conesa A, Gaasterland T, Hunter L, Kingsford C, Kohlbacher O, Lengauer T, Markel S, Moreau Y, Noble WS, Orengo C, Ouellette BFF, Parida L, Przulj N, Przytycka TM, Ranganathan S, Schwartz R, et al. (2024). Biological Databases in the Age of Generative Artificial Intelligence. *Bioinformatics Advances* **5**. Ed. by Alex Bateman, vbaf044. https://doi.org/10.1093/bioadv/vbaf044.

Pope Q, Varma R, Tataru C, David MM, Fern X (2025a). Learning a Deep Language Model for Microbiomes: The Power of Large Scale Unlabeled Microbiome Data. *PLOS Computational Biology* **21**. Ed. by Stacey D. Finley, e1011353. https://doi.org/10.1371/journal.pcbi.1011353.

Pope Q, Varma R, Tataru C, David MM, Fern X (2025b). Learning a deep language model for microbiomes: The power of large scale unlabeled microbiome data. *PLoS Comput. Biol.* **21**, e1011353.

Pressman EM, Kebreab E (2024). A Review of Key Microbial and Nutritional Elements for Mechanistic Modeling of Rumen Fermentation in Cattle under Methane-Inhibition. *Frontiers in Microbiology* **15**, 1488370. https://doi.org/10.3389/fmicb.2024.1488370.

Props R, Monsieurs P, Mysara M, Clement L, Boon N (2016). Measuring the Biodiversity of Microbial Communities by Flow Cytometry. *Methods in Ecology and Evolution* **7**. Ed. by David Hodgson, 1376–1385. https://doi.org/10.1111/2041-210X.12607.

Purcell W, Neubauer T (2023). Digital Twins in Agriculture: A State-of-the-art Review. *Smart Agricultural Technology* **3**, 100094. https://doi.org/10.1016/j.atech.2022.100094.

Puy A, Bacon E, Carmona A, Flinders S, Gefen D, Khanjani M, Larsen KR, Lachi A, Linga SN, Lo Piano S, Melsen LA, Murray E, Sheikholeslami R, Sobhani A, Wei N, Saltelli A (2025). Socio-environmental modeling shows physics-like confidence with water modeling surpassing it in numerical claims. *iScience* **28**, 112184. https://doi.org/10.1016/j.isci.2025.112184.

Pylianidis C, Osinga S, Athanasiadis IN (2021). Introducing Digital Twins to Agriculture. *Computers and Electronics in Agriculture* **184**, 105942. https://doi.org/10.1016/j.compag.2020.105942.

Quince C, Walker AW, Simpson JT, Loman NJ, Segata N (2017). Shotgun Metagenomics, from Sampling to Analysis. *Nature Biotechnology* **35**, 833–844. https://doi.org/10.1038/nbt.3935.

Raaijmakers JM, Kiers ET (2022). Rewilding Plant Microbiomes. *Science* **378**, 599–600. https://doi.org/10.1126/science.abn6350.

Ramirez I, Volcke EI, Rajinikanth R, Steyer JP (2009). Modeling Microbial Diversity in Anaerobic Digestion through an Extended ADM1 Model. *Water Research* **43**, 2787–2800. https://doi.org/10.1016/j.watres.2009.03.034.

Richardson L, Allen B, Baldi G, Beracochea M, Bileschi ML, Burdett T, Burgin J, Caballero-Pérez J, Cochrane G, Colwell LJ, Curtis T, Escobar-Zepeda A, Gurbich TA, Kale V, Korobeynikov A, Raj

S, Rogers AB, Sakharova E, Sanchez S, Wilkinson DJ, et al. (2023). MGnify: The Microbiome Sequence Data Analysis Resource in 2023. *Nucleic Acids Research* **51**, D753–D759. https://doi.org/10.1093/nar/gkac1080.

Ritto T, Rochinha F (2021). Digital Twin, Physics-Based Model, and Machine Learning Applied to Damage Detection in Structures. *Mechanical Systems and Signal Processing* **155**, 107614. https://doi.org/10.1016/j.ymssp.2021.107614.

Rocks MC, Bhatnagar P, Verticchio Vercellin A, Sala L, Siesky B, Antman G, Wood K, Sacco R, Harris A (2025). Mathematical Modeling and Artificial Intelligence to Explore Connections Between Glaucoma and the Gut Microbiome. *Medicina (Kaunas, Lithuania)* **61**, 343. https://doi.org/10.3390/medicina61020343. PMID: 40005459.

Rodrigues JFM, Tackmann J, Malfertheiner L, Patsch D, Perez-Molphe-Montoya E, Näpflin N, Gaio D, Rot G, Danaila M, Peluso ME, Dmitrijeva M, Schmidt TSB, von Mering C (2025). The MicrobeAtlas database: Global trends and insights into Earth's microbial ecosystems.

Rohart F, Gautier B, Singh A, Lê Cao KA (2017). mixOmics: An R Package for 'omics Feature Selection and Multiple Data Integration. *PLOS Computational Biology* **13**. Ed. by Dina Schneidman, e1005752. https://doi.org/10.1371/journal.pcbi.1005752.

Rul F, Béra-Maillet C, Champomier-Vergès MC, El-Mecherfi KE, Foligné B, Michalski MC, Milenkovic D, Savary-Auzeloux I (2022). Underlying Evidence for the Health Benefits of Fermented Foods in Humans. *Food & Function* **13**, 4804–4824. https://doi.org/10.1039/D1FO03989J.

Saltelli A, Bammer G, Bruno I, Charters E, Di Fiore M, Didier E, Nelson Espeland W, Kay J, Lo Piano S, Mayo D, Pielke Jr R, Portaluri T, Porter TM, Puy A, Rafols I, Ravetz JR, Reinert E, Sarewitz D, Stark PB, Stirling A, et al. (2020). Five ways to ensure that models serve society: a manifesto. *Nature* **582**, 482–484. https://doi.org/10.1038/d41586-020-01812-9.

Saltelli A, Gigerenzer G, Hulme M, Katsikopoulos KV, Melsen LA, Peters GP, Pielke R, Robertson S, Stirling A, Tavoni M, Puy A (2024). Bring Digital Twins Back to Earth. *WIREs Climate Change* **15**, e915. https://doi.org/10.1002/wcc.915.

Saltelli A, Melsen LA, Puy A (2025). Digital Twins of the Earth Between Vision and Fiction. *Minerva*. https://doi.org/10.1007/s11024-025-09581-3.

Sarhan MS, Hamza MA, Youssef HH, Patz S, Becker M, ElSawey H, Nemr R, Daanaa HSA, Mourad EF, Morsi AT, Abdelfadeel MR, Abbas MT, Fayez M, Ruppel S, Hegazi NA (2019). Culturomics of the Plant Prokaryotic Microbiome and the Dawn of Plant-Based Culture Media – A Review. *Journal of Advanced Research* **19**, 15–27. https://doi.org/10.1016/j.jare.2019.04.002.

Schäfer M, Pacheco AR, Künzler R, Bortfeld-Miller M, Field CM, Vayena E, Hatzimanikatis V, Vorholt JA (2023). Metabolic Interaction Models Recapitulate Leaf Microbiota Ecology. *Science* **381**, eadf5121. https://doi.org/10.1126/science.adf5121.

Schmidt TS, Raes J, Bork P (2018). The Human Gut Microbiome: From Association to Modulation. *Cell* **172**, 1198–1215. https://doi.org/10.1016/j.cell.2018.02.044.

Sharma D, Paterson AD, Xu W (2020). TaxoNN: Ensemble of Neural Networks on Stratified Microbiome Data for Disease Prediction. *Bioinformatics* **36**. Ed. by Pier Luigi Martelli, 4544–4550. https://doi.org/10.1093/bioinformatics/btaa542.

Sherwani MK, Ruuskanen MO, Feldner-Busztin D, Nisantzis Firbas P, Boza G, Móréh Á, Borman T, Putu Erawijantari P, Scheuring I, Gopalakrishnan S, Lahti L (2025). Multi-Omics Time-Series Analysis in Microbiome Research: A Systematic Review. *Briefings in Bioinformatics* **26**, bbaf502. https://doi.org/10.1093/bib/bbaf502.

Sizemore N, Oliphant K, Zheng R, Martin CR, Claud EC, Chattopadhyay I (2024). A Digital Twin of the Infant Microbiome to Predict Neurodevelopmental Deficits. *Science Advances* **10**, eadj0400. https://doi.org/10.1126/sciadv.adj0400.

Smith CJ, Osborn AM (2009). Advantages and Limitations of Quantitative PCR (Q-PCR)-Based Approaches in Microbial Ecology: Application of Q-PCR in Microbial Ecology. *FEMS Microbiology Ecology* **67**, 6–20. https://doi.org/10.1111/j.1574-6941.2008.00629.x.

So D, Yao CK, Gill PA, Thwaites PA, Ardalan ZS, McSweeney CS, Denman SE, Chrimes AF, Muir JG, Berean KJ, Kalantar-Zadeh K, Gibson PR (2023). Detection of Changes in Regional Colonic Fermentation in Response to Supplementing a Low FODMAP Diet with Dietary Fibres by Hydrogen Concentrations, but Not by Luminal pH. *Alimentary Pharmacology & Therapeutics* **58**, 417–428. https://doi.org/10.1111/apt.17629.

Somerville V, Grigaitis P, Battjes J, Moro F, Teusink B (2022). Use and Limitations of Genome-Scale Metabolic Models in Food Microbiology. *Current Opinion in Food Science* **43**, 225–231. https://doi.org/10.1016/j.cofs.2021.12.010.

Souza AL, Patti GJ (2021). A Protocol for Untargeted Metabolomic Analysis: From Sample Preparation to Data Processing. In: *Mitochondrial Medicine*. Ed. by Volkmar Weissig and Marvin Edeas. Vol. 2276. New York, NY: Springer US, pp. 357–382. https://doi.org/10.1007/978-1-0716-1266-8_27.

Tap J, Lejzerowicz F, Cotillard A, Pichaud M, McDonald D, Song SJ, Knight R, Veiga P, Derrien M (2023). Global Branches and Local States of the Human Gut Microbiome Define Associations with Environmental and Intrinsic Factors. *Nature Communications* **14**, 3310. https://doi.org/10.1038/s41467-023-38558-7.

Tartakovsky B, Morel E, Steyer JP, Guiot SR (2002). Application of a Variable Structure Model in Observation and Control of an Anaerobic Digestor. *Biotechnology Progress* **18**, 898–903. https://doi.org/10.1021/bp010142c.

Teng TS, Chin YL, Chai KF, Chen WN (2021). Fermentation for Future Food Systems: Precision Fermentation Can Complement the Scope and Applications of Traditional Fermentation. *EMBO reports* **22**, e52680. https://doi.org/10.15252/embr.202152680.

Topić Popović N, Kazazić SP, Bojanić K, Strunjak-Perović I, Čož-Rakovac R (2023). Sample Preparation and Culture Condition Effects on MALDI-TOF MS Identification of Bacteria: A Review. *Mass Spectrometry Reviews* **42**, 1589–1603. https://doi.org/10.1002/mas.21739.

Torfs E, Nicolaï N, Daneshgar S, Copp JB, Haimi H, Ikumi D, Johnson B, Plosz BB, Snowling S, Townley LR, Valverde-Pérez B, Vanrolleghem PA, Vezzaro L, Nopens I (2024). The Transition of WRRF Models to Digital Twin Applications. In: *Modelling for Water Resource Recovery*. Ed. by Nicolas Derlon, Kris Villez, Heather Stewart, and Arifur Rahman. IWA Publishing. https://doi.org/10.2166/wst.2022.107.

Tzachor A, Richards CE, Jeen S (2022a). Transforming Agrifood Production Systems and Supply Chains with Digital Twins. *npj Science of Food* **6**, 47. https://doi.org/10.1038/s41538-022-00162-2.

Tzachor A, Sabri S, Richards CE, Rajabifard A, Acuto M (2022b). Potential and Limitations of Digital Twins to Achieve the Sustainable Development Goals. *Nature Sustainability* **5**, 822–829. https://doi.org/10.1038/s41893-022-00923-7.

Ugolini GS, Wang M, Secchi E, Pioli R, Ackermann M, Stocker R (2024). Microfluidic Approaches in Microbial Ecology. *Lab on a Chip* **24**, 1394–1418. https://doi.org/10.1039/D3LC00784G.

Ünlü M, Morgan ME, Minden JS (1997). Difference Gel Electrophoresis. A Single Gel Method for Detecting Changes in Protein Extracts. *ELECTROPHORESIS* **18**, 2071–2077. https://doi.org/10.1002/elps.1150181133.

Vallée A (2023). Digital Twin for Healthcare Systems. *Frontiers in Digital Health* **5**, 1253050. https://doi.org/10.3389/fdgth.2023.1253050.

Van Bruggen AH, Goss EM, Havelaar A, Van Diepeningen AD, Finckh MR, Morris JG (2019). One Health - Cycling of Diverse Microbial Communities as a Connecting Force for Soil, Plant, Animal, Human and Ecosystem Health. *Science of The Total Environment* **664**, 927–937. https://doi.org/10.1016/j.scitotenv.2019.02.091.

Van Den Bossche T, Armengaud J, Benndorf D, Blakeley-Ruiz JA, Brauer M, Cheng K, Creskey M, Figeys D, Grenga L, Griffin TJ, Henry C, Hettich RL, Holstein T, Jagtap PD, Jehmlich N, Kim J, Kleiner M, Kunath BJ, Malliet X, Martens L, et al. (2025). The Microbiologist's Guide to Metaproteomics. *iMeta* **4**, e70031. https://doi.org/10.1002/imt2.70031.

Van Gulik WM (2010). Fast Sampling for Quantitative Microbial Metabolomics. *Current Opinion in Biotechnology* **21**, 27–34. https://doi.org/10.1016/j.copbio.2010.01.008.

Vanrolleghem PA, Khalil M, Serrao M, Sparks J, Therrien JD (2025). Machine Learning in Wastewater: Opportunities and Challenges — "Not Everything Is a Nail!". *Current Opinion in Biotechnology* **93**, 103271. https://doi.org/10.1016/j.copbio.2025.103271.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017a). Attention is all you need. eprint: 1706.03762 (cs.CL).

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017b). Attention Is All You Need. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Vol. 30. Curran Associates, Inc.

Viceconti M, De Vos M, Mellone S, Geris L (2024). Position Paper From the Digital Twins in Healthcare to the Virtual Human Twin: A Moon-Shot Project for Digital Health Research. *IEEE Journal of Biomedical and Health Informatics* **28**, 491–501. https://doi.org/10.1109/JBHI.2023.3323688.

Vivares G, Dijkstra J, Bannink A (2025). Modeling Diurnal Rumen Metabolism Dynamics in Dairy Cattle: An Update to a Mechanistic Model Representing Eating Behavior, Rumen Content, Rumination, and Acid-Base Balance. *Journal of Dairy Science* **108**, 6934–6957. https://doi.org/10.3168/jds.2024-26121.

Wade M, Harmand J, Benyahia B, Bouchez T, Chaillou S, Cloez B, Godon JJ, Moussa Boudjemaa B, Rapaport A, Sari T, Arditi R, Lobry C (2016). Perspectives in Mathematical Modelling for Microbial Ecology. *Ecological Modelling* **321**, 64–74. https://doi.org/10.1016/j.ecolmodel.2015.11.002.

Wang AJ, Li H, He Z, Tao Y, Wang H, Yang M, Savic D, Daigger GT, Ren N (2024). Digital Twins for Wastewater Treatment: A Technical Review. *Engineering* **36**, 21–35. https://doi.org/10.1016/j.eng.2024.04.012.

Wang Y, Hammes F, De Roy K, Verstraete W, Boon N (2010). Past, Present and Future Applications of Flow Cytometry in Aquatic Microbiology. *Trends in Biotechnology* **28**, 416–424. https://doi.org/10.1016/j.tibtech.2010.04.006.

Webb KJ, Xu T, Park SK, Yates JR (2013). Modified MuDPIT Separation Identified 4488 Proteins in a System-wide Analysis of Quiescence in Yeast. *Journal of Proteome Research* **12**, 2177–2184. https://doi.org/10.1021/pr400027m.

Wiatrak M, Viñas Torné R, Ntemourtsidou M, Dinan A, Abelson DC, Arora D, Brbić M, Weimann A, Floto RA (2025). A contextualised protein language model reveals the functional syntax of bacterial evolution.

Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, Cordero OX, Brown SP, Momeni B, Shou W, Kettle H, Flint HJ, Haas AF, Laroche B, Kreft JU, Rainey PB, Freilich S, Schuster S, Milferstedt K, van der Meer JR, et al. (2016). Challenges in Microbial Ecology: Building Predictive Understanding of Community Function and Dynamics. *The ISME journal*. https://doi.org/10.1038/ismej.2016.45.

Winkler AS, Brux CM, Carabin H, Das Neves CG, Häsler B, Zinsstag J, Fèvre EM, Okello A, Laing G, Harrison WE, Pöntinen AK, Huber A, Ruckert A, Natterson-Horowitz B, Abela B, Aenishaenslin C, Heymann DL, Rødland EK, Berthe FCJ, Capua I, et al. (2025). The Lancet One Health Commission: Harnessing Our Interconnectedness for Equitable, Sustainable, and Healthy Socioecological Systems. *The Lancet* **406**, 501–570. https://doi.org/10.1016/S0140-6736(25)00627-0.

Wright L, Davidson S (2020). How to Tell the Difference between a Model and a Digital Twin. *Advanced Modeling and Simulation in Engineering Sciences* **7**, 13. https://doi.org/10.1186/s40323-020-00147-4.

Wyckhuys KA, Bushley K, Gratton C, Gurr GM, Pozsgai G, Tscharntke T, Wanger TC, Lu Y, Elkahky M (2025). Restoring Functional Farmland Biodiversity for Biological Pest Control. *Trends in Plant Science* **30**, 1097–1110. https://doi.org/10.1016/j.tplants.2025.03.012.

Yabo AG, Casenave C (2023). Aroma Synthesis and Energy Consumption in Wine Fermentation: A Multiobjective Optimization Approach. *IFAC-PapersOnLine* **56**, 6211–6216. https://doi.org/10.1016/j.ifacol.2023.10.743.

Yaffe E, Relman DA (2019). Tracking Microbial Evolution in the Human Gut Using Hi-C Reveals Extensive Horizontal Gene Transfer, Persistence and Adaptation. *Nature Microbiology* **5**, 343–353. https://doi.org/10.1038/s41564-019-0625-0.

Zhang H, Zhang Y, Kang Z, Xiong J, Yang R, Ning K (2026). MGM as a large-scale pretrained foundation model for microbiome analyses in diverse contexts. *Adv. Sci. (Weinh.)*, e13333.

Zhang N, Kandalai S, Zhou X, Hossain F, Zheng Q (2023). Applying Multi-omics toward Tumor Microbiome Research. *iMeta* **2**, e73. https://doi.org/10.1002/imt2.73.

Zhao S, Jiao T, Adade SYSS, Wang Z, Ouyang Q, Chen Q (2025). Digital Twin for Predicting and Controlling Food Fermentation: A Case Study of Kombucha Fermentation. *Journal of Food Engineering* **393**, 112467. https://doi.org/10.1016/j.jfoodeng.2025.112467.

	Data type	Description	Used for	Limitations	Quantif.	Depth	Time scale	Citation
Microbial identification & quantification	optical density	estimation of microbial density based on turbidity measurements with a spectophotometer	Non selective total population counts	total population counts only	yes	Community-wide	seconds	(Mytilinaios et al., 2012)
	flow cytometry	microfluidic cell numbering	cell numbering	Cells must be tagged or differenciable for population counts in community	yes	Population-specific	minutes	(Props et al., 2016)
	flow cytometry + FISH	microfluidic and fluorescent discrimination	cell numbering and discrimination	higher price, few populations	yes	Population-specific	minutes	(Wang et al., 2010)
	FACS (fluorescence activated cell sorting)	Flow cytometry based sorting of cells	cell sorting for accurate population numbering or characterization.	Cells must be tagged. High price.	yes	Population-specific	minutes	(Müller and Nebe-von-Caron, 2010)
	MALDI-TOF	micro-organism-specific mass spectrum	Identification of micro-organisms	Strains must be isolated	no	Species composition	minutes	(Topić Popović et al., 2023)
	metaproteomics	Micro-organism-specific peptide profiles	Identification of micro-organisms	Strains must be isolated	no	Species composition	minutes	(Kleiner, 2019)
	Fluorescence microscopy	Imaging cell fluorescence by microscopy	Specific population densities	GMO needed, few populations	yes	Population-specific	hours	(Gitai, 2009)
	Optical coherence tomography	Biofilm imaging technique	Microbial density and structures in biofilms	No species differentiation	yes	Community-wide	hours	(Hou et al., 2019)
	q-PCR	q-PCR amplification of marker gene	targeted absolute count of specific populations	Specific marker gene must be identified for each target	yes	Population-specific	hours	(Smith and Osborn, 2009)
	direct counts	Microscopy-based direct numbering	Total population counts	Time consuming, high operative cost	yes	Community-wide	hours	(Daims and Wagner, 2007)
	WGS	whole genome sequencing	untargeted relative counts + MAGs	Price, Evolved bioinformatics	relative	MAGs composition	days	(Quince et al., 2017)
	Plate-counting	growth on selective media	numbering of targeted species	existance of selective media, high errors	yes	Population-specific	days	(Olsen and Bakken, 1987)
	metabarcoding	Sequencing of amplicons	untargeted relative counts	Relative counts. Lot of wet lab.	relative	OTU/ASV composition	days	(Pauvert et al., 2019)
	culturomics	High-throughput isolation and cultivation	After characterization (e.g. by Maldi-TOF), isolated micro-organisms give an insight on biodiversity	Only cultivable micro-organisms, must be coupled with other technique.	no	Strain level	weeks	(Huang et al., 2023)

Table 1 – Microbial diversity data. The data available to decipher community composition and structure are reviewed. Their use and limitations are discussed, together with their quantification type (abosolute or compositional), their depth (here understood as their resolution, from strain to whole-community data) and their production time-scale.

	Data type	Description	Used For	Limitations	Quantif.	Depth	Time Scale	Citation
Metagenomics & Metatranscript.	q-PCR	q-PCR amplification of targeted genes	targeted absolute count of specific genes or transcripts	Specific probes must be designed for each gene	absolute	Gene-specific	hours	(Gao et al., 2011)
	shotgun metagenomics	untergeted sequencing of all (microbial) DNA in a sample	Taxonomic and functional characterisation of samples, reconstruction of genomes	Requires assembly of the sequences.	Relative counts	Whole-genome	days	(Quince et al., 2017)
	ATAC-seq	Identification of accessible DNA in the genome	Study the regulatory landscape of a genome	Cost, interpretation	Relative counts	Whole-genome	days	(Grandi et al., 2022)
	Hi-C	Fixation-based sequencing technique	Estimation of spatial proximity between DNA fragments. Can be used to facilitate assembly	Cost, interpretation	Relative counts	Whole-genome	days	(Yaffe and Relman, 2019)
	Single-cell	DNA sequencing of a single DNA molecule from an isolated cell	Functions characterization through single-amplified genomes	Cost, interpretation, operational difficulties	Relative counts	Whole-genome	days	(Lloréns-Rico et al., 2022)
	Note:	*Sequencing technologies are now strongly improving the length of sequenced fragments (from short to long reads) hence facilitating binning. Cell sorting by FACS, or microfluidic techniques can help to pre-process the samples in order to facilitate post processing tasks such as assembly. Same technologies can be used for metagenomics of metatranscriptomics.*						
Metaproteomics	LC-(HR)MS	Separates peptides by liquid chromatography (LC) before high-resolution (HR) mass spectrometry (MS).	Broad metabolite profiling	Complex sample preparation; matrix effects; may miss non-ionizable or extreme-polar compounds	Yes (with internal standards).	Whole-proteome	minutes	(Souza and Patti, 2021)
	tagged LC-MS	LC-MS samples are multiplexed by reagents (ICAT, iTRAQ, SILAC)	quantitative comparison between samples	Same limitations as LC-MS ; limited multiplexing possibilities	Yes	Whole-proteome	hours	(Van Den Bossche et al., 2025)
	SRM/MRM	Selective or Multiple Reaction monitoring (SRM/MRM) coupled to LC-MS	Targeted proteomics : SRM or MRM allow for peptide selection	Limited number of targeted peptides	Yes (with internal standards)	Protein-specific	hours	(Colangelo et al., 2013)
	2D LC-MS	Multidimensional LC-MS coupled with data-based identification (MudPIT)	Discovery proteomics ; protein identification	Not quantitative ; Complex sample preparation; matrix effects	No	Whole-proteome	hours	(Webb et al., 2013)
	Gel-based	Gel electrophoresis (GE), possibly multiplexed (DIGE), coupled to LC-MS	Seperate proteins, compare protein levels in two samples (DIGE)	Very sensitive to experimental conditions : lack of reproducibility.	No (but comparison)	Protein-specific	days	(Ünlü et al., 1997)
	Note:	*MS = mass spectrometry. After enrichment of phosphorylated proteins, LC-MS can be applied (giving metaphosphoproteomics). Data-independant (DIA) fractionation can be applied (giving DIA-MS). Spectra can be compared to data-bases or computed de novo.*						

Table 2 – Microbial functions data. 1/2: Metagenomics, metatranscriptomics and metaproteomics. The data available to decipher microbial functions are reviewed. Their use and limitations are discussed, together with their quantification type (abosolute or compositional), their depth (here understood as their resolution, from molecular to community-scale bioprocesses data) and their production time-scale.

	Data type	Description	Used For	Limitations	Quantif.	Depth	Time Scale	Citation
Metabolomics	dosage	Quantitative dosage of targeted metabolites	Rapid quantification of targeted metabolites	Low-throughput	Yes	Metabolite-specific	minutes	(Van Gulik, 2010)
	SRM/MRM	SRM/MRM in GC-MS and LC-MS	Targeted metabolomics	Limited number of targeted metabolites	Yes (with internal standards)	Metabolite-specific	minutes	(Zhang et al., 2023)
	GC–(HR)MS	Separates volatile small metabolites by gaz chromatography (GC) before MS	Profiling and quantification of small organic metabolites	Requires chemical derivatization; poor for non-volatile or larger compounds; moderate throughput	Yes (with internal standards)	Whole-metabolome	minutes	(Fiehn, 2016)
	LC–(HR)MS	Separates compounds by LC before (HR)MS.	Broad metabolite profiling	Complex sample preparation; matrix effects; may miss non-ionizable or extreme-polar compounds	Yes (with internal standards).	Whole-metabolome	minutes	(Souza and Patti, 2021)
	CE–(HR)MS	Capillary Electrophoresis (CE) for high-resolution separation of metabolites	Profiling ionic metabolites in low-volume samples	Low sensitivity due to low volumes; specialized instrumentation needed.	Yes (semi-quantitative)	Whole-metabolome	minutes	(Cai and Henion, 1995)
	NMR	Nuclear Magnetic Resonance (NMR) non-destructive analysis of metabolites	Untargeted profiling and absolute quantitation of abundant metabolites	Low sensitivity; overlapping peaks in complex mixtures; high instrument cost.	Yes	Whole-metabolome	minutes	(Nagana Gowda and Raftery, 2021)
	Ambient MS	Ambient analysis from surfaces or gases (e.g. DESI, DART…)	Rapid in situ metabolite profiling (e.g. microbial samples)	Generally qualitative; matrix effect.	Limited	Whole-metabolome	minutes	(Cooks et al., 2006)
	MS Imaging	Spatial distribution of metabolites on surfaces (e.g. MALDI, DESI, SIMS)	Spatialized metabolomics within microbial colonies or tissues.	Complex sample preparation; limited identification capabilities; slow image acquisition	Yes (semi-quantitative)	Whole-metabolome	hours	(Amstalden Van Hove et al., 2010)
	Note:	*Different metabolomic technics can be applied on extracellular media (exometabolomics), on volatile organic compounds (volatilomics) or to follow-up labeled metabolites to decipher fluxes (fluxomics).*						

Table 3 – Microbial functions data. 2/2: metabolomics. The data available to decipher microbial functions are reviewed. Their use and limitations are discussed, together with their quantification type (abosolute or compositional), their depth (here understood as their resolution, from molecular to community-scale bioprocesses data) and their production time-scale.

43

	Ecological lever	Description	Used for	Limitation	Depth	Time scale
Abiotic	Dilution rate	Management of bioreactor inflow and outflow rates	Modulation of nutritional environment and harshness through dilution	Same modulation for the whole community	Community-wide	seconds
	Temperature	Controlled temperature of the culture.	Influence growth rate, enzyme activity, metabolic pathways.	Same modulation for the whole community. Not targeted	Community-wide	seconds
	pH	Acidic/basic conditions of the culture.	Influence enzyme activity, nutrient availability, or stress response.	Buffering requirements, pH drift, not targeted.	Community-wide	seconds
	Light	Light exposure (intensity, wavelength, photoperiod).	Drive phototrophic growth, optogenetic control, or circadian rhythms.	Turbidity may disturb light access, complexity of optogenetic systems	Targeted functions or populations	seconds
	Oxymetry	Oxygen availability (aerobic/anaerobic/microaerophilic).	Control respiration, fermentation, or redox-sensitive pathways.	Equipment cost, operational complexity	Targeted pathways	minutes
	Magnetism	Exposure to magnetic fields.	Magnetic fields can modulate microbial metabolism and select receptive communities	It is unclear by which mechanism reactions or populations are enhanced. Not targeted.	Community-wide	minutes
	Electrodes	Application of electrical potential or current.	Drive electrogenic/electrotrophic activity, redox control, or bioelectrochemical systems.	Limited to systems where redoxBalance is the main constraint	targeted pathways	minutes
	Agitation	Mechanical mixing (stirring, shaking, sparging).	Improve mass transfer, homogeneity, or shear stress.	Limited to systems where mechanical environment or spatialStructures are key	Community-wide	minutes
	Spatial structure	Physical compartmentalization (e.g., beads, membranes, microfluidics, 3D-printed structures...).	Mimic natural niches, control interactions, or spatial gradients.	Limited to system where spatial structures are key, no temporal modulation ("batch experiment")	Community-wide	hours
	Hygrometry	Humidity or water activity control.	Optimize water availability, osmotic stress, or desiccation tolerance.	Untargeted, modulate a global Stress	Community-wide	hours
	Culture media	Composition of growth medium	Designed media for targeted community outputs (growth, metabolite production, trophic interaction…)	Complexity to select the media (reverse ecology).	Roughly targeted function	hours
	Prebiotic (in host)	Non-digestible compounds promoting beneficial microbes in a host.	Enhance growth of specific taxa, modulate community composition.	Specificity, potential off-target effects.	Roughly targeted function	hours
	Support (biofilm induction)	Different textures for substrate and support can be used	Enhance biomass retention, spatial organization, or stress resistance by biofilm formation.	No temporal modulation ("batch experiment")	Biofilm-forming population-wide	hours
	Antibiotics, biocides	Chemical agents targeting specific microbes.	Selective inhibition of targeted populations	Resistance development, non-specific toxicity	Targeted populations	hours
Biotic	Probiotic, bioaugmentation	Introduction of beneficial microbes.	Enhance function, stability, or resilience of the community.	Engraftment challenges due to competition.	Roughly targeted function or populations	hours
	Phages	Viruses targeting specific bacteria.	Population control or horizontal gene transfer.	Host range limitations, resistance.	Targeted populations or functions	hours
	Interactions (microfluidics)	Controlled micro-scale fluidic environments.	Study small-population dynamics, gradients, or spatial and metabolic interactions.	Technical complexity, scalability issues.	Targeted functions	hours
	Inocculum selection	Choice of initial microbial community.	Steer succession, community structure and functions.	Difficulty for community selection. No temporal modulation.	Targeted populations or functions	hours
	Gene editing	Genetic editing of community members	Add or knock-out functions to modulate community dynamics	Biosafety hazard, no temporal modulation ("batch experiment")	Targeted function	days

Table 4 – Microbial ecology engineering levers. The available levers to engineer microbial ecosystems are reviewed. Their use and limitations are discussed, together with their depth (here understood as their resolution, from strain to whole-community, or from molecular mechanisms to community-wide functions) and their activation time-scale.