4. Microbial Ecosystems

Background

Microbes are the most ancient, abundant, and diverse form of life on Earth. They have co-evolved with and shaped our planet, and without them life as we know it would not exist. Although often considered as simple or primitive, these unicellular organisms colonise, proliferate on, and impact every biotic and abiotic surface and subsurface of our planet, even in its most inhospitable corners. Microbes can be found literally everywhere: in and on animals, on plants, in the soil, in aquatic environments, in food chains, in everything that humans can and cannot touch (from household and medical appliances to the bottom of the ocean, and from microbial mats in hot springs to the permafrost). Microbes do this mostly in the form of complex communities, in which they interact with each other and with their surroundings, forming complex microbial ecosystems. In the past two decades, DNA sequencing technologies have exposed the breadth, complexity, diversity, and ecogeography of such microbial communities, also referred to as **microbiomes**. It is currently estimated that more than 10^{12} microbial species reside in these communities, yet only around 10^{4} of them have been isolated or have available reference genomes. Their interactions and functional capacities, as well as the effects they have on their surroundings, are even less well characterised.

Mapping the microbial diversity within an ecosystem is the first step towards understanding the functional role in the ecosystem of individual microbes and their communities. Knowledge of *who* is there paves the way for answering questions on *what* actions they perform, *how*, and *why*. The biological principles, molecules, interactions, and functional outputs will help answer questions on community stability and niche specificity. At present, the only microbiome for which there is a good understanding of its diversity is the human microbiome – the collection of microbes that live on and inside humans. The human gut microbiome in particular has emerged in the past decade as a prototypical microbial ecosystem, due to its accessibility, tractability, richness, and direct relevance to human health. As a 'human organ' with roughly 100 times more genes than the human genome, the gut microbiota plays an essential role in host health: helping in host metabolism, immunity, brain function, and response to medication. Numerous studies associate changes in microbiota composition to susceptibility to infections and diseases such as diabetes, colon cancer, cardiovascular and neurological conditions, as well as many others. The development and progress of these diseases is thought to be linked to altered functional outputs of the microbiota as a result of changes in its composition. However, many of these links, as well as the overall view of the gut microbiota and its interplay with the host and the environment, remain largely descriptive, relying on associations between phenotypes of interest and microbial community composition. The next challenge lies in mapping the causal effects and understanding the underlying molecular mechanisms of gut microbiome–host–environment interactions. The knowledge derived can lay the groundwork for understanding other complex microbial ecosystems as they are better characterised.

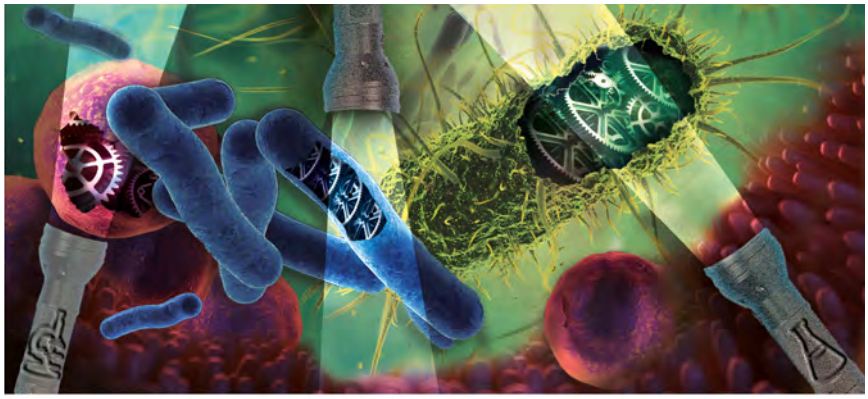


Figure ME1 | Shedding light on the dark genetic matter of the gut microbiome to help establish *de novo* gut model organisms.

The Opportunity

To better understand microbial ecosystems, their functional capacities, and their molecular interplay with the environment, EMBL researchers first need to study the individual players within microbiomes: the microbes themselves. Characterising microbial communities and the interactions within them and with their natural environment, will create an understanding of microbial communities' roles in their habitat, and of the underlying molecular players (genes, proteins, metabolites) and mechanisms. This knowledge will lay the foundations for intervention and rational modulation of microbiomes from dysbiotic states, and/or to help to rebalance or restore the ecosystems they live in. In the case of the human microbiota, it will enable personalised, microbiome-related therapies. In the case of soil and marine microbiotas, it will open unprecedented opportunities for reviving and rebalancing natural ecosystems.

The first step in this process is to collect the available information on microbiomes in one place, allowing for systematic curation and dissemination to the scientific community. This will set standards for future investigations, provide a framework to report and query study results, and facilitate the identification of knowledge gaps to tackle further research. The second step is to develop experimental and computational resources and tools to functionally study isolated microbes and microbial communities in and out of their natural context. These resources encompass strain collections, pipelines for systematic genetic manipulation of the microbes they contain, model communities, experimental platforms to study these communities (e.g. approaches to link genotypes to phenotypes, and to assess the impact of controlled conditions and perturbations), and frameworks to integrate multi-omics and multiscale data. Building from its pioneering work and growing expertise on the human microbiome, EMBL is in a unique position to start dissecting and understanding the underlying principles and molecular mechanisms of the assembly, dynamics, and properties of microbial ecosystems.

Research Aims

Over the past decade, EMBL has been instrumental in developing computational and experimental tools and approaches that have propelled research on the human microbiome, facilitating our current understanding of the microbiome's diversity and its role in human health. These technological advances (Tech Dev Boxes TD1-3_ME) have provided novel insights into human gut microbiome composition across populations, the impact on it of perturbations and age, its links to disease, its interfaces with drugs, its strain resolution, and its encoded functional diversity. In the new Programme, EMBL aims to become a leading hub of microbiome research and resources in Europe by strengthening existing tools and developing new ones, in close collaboration with other leading research organisations in the field, in member states and elsewhere. This will involve the assimilation and curation of microbiome sequencing data, followed by integration with systematic experimental strategies to gain insights into the fundamental principles that shape microbial ecosystems and the functional capacities they encode. **Microbes that colonise the human gut will initially be used as models**, with the focus gradually expanding towards more fastidious and understudied environmental microbiomes related to, for example, soil, plants, marine waters, or other natural environments, with roles in plant growth, carbon cycling, antibiotic biogeography, or pollution degradation (Chapter 7: Planetary Biology). EMBL is in a unique position to address this ambitious initiative, which will **promote the transition of microbiome research from descriptive and correlative to molecular and causal**, revealing the underlying mechanisms and interactions supporting complex microbial communities and their interactions with the environment.

This level of understanding will ultimately enable the rational modulation of these communities, when required. EMBL specifically aims to:

I. **Understand the functional diversity of individual microbial species and strains**. Current knowledge of bacterial gene functions, pathways, and cellular architecture stems from very few model bacteria, which fail to capture the phylogenetic and genetic diversity of the gut microbiota. As a consequence, the vast majority of genes in the gut microbiome remain 'dark matter' with respect to function: that is, of elusive or completely unknown function. EMBL will be the hub for community efforts to **systematically tackle the vast genetic dark matter in the human gut microbiome and to establish new relevant model microbes**. These efforts will build upon existing resources to develop the next tier of microbiome-related computational tools and databases, and the development of automated high-throughput experimental pipelines. This will unravel novel pathways and protein machines, illuminating how microbes produce bioactive molecules, communicate with each other, survive stress, resist or modify xenobiotics, and metabolise nutrients. Importantly, it will also provide a roadmap for generating functional knowledge about key microbes in any microbiome or ecological habitat.

II. **Dissect the interactions and properties of microbial communities**. Microbes within complex communities, such as the gut microbiota, compete for resources but also cooperate to break down complex food sources, communicate with each other, fend off intruders, and deal with stress or fluctuating environments. Using high-throughput experimental setups at different levels of complexity and tractability (from monocultures and microbial communities to human donor cohorts), EMBL researchers aim to understand the underlying principles driving the organisation, stability, and characteristics of gut microbial communities, and to establish model communities. A specific focus will be on the role of the gut microbiota in containing and combating pathogens, and on the development and spread of antimicrobial resistance.

III. **Place microbial communities in their ecological context**. Microbes and microbial communities are shaped largely by interactions with their environments. They also have functionalities that only become relevant in their natural ecological context. To study microbial organisms and communities in their natural context, which in the case of the human gut microbiota is the human host, EMBL aims to study the functional outputs and characteristics of such communities in co-culture with their host cells, in organoids, or in gnotobiotic animal models. Within this endeavour lies a unique opportunity to understand the principles and molecular mechanisms by which intrinsic factors (e.g. bacterial genomes, metabolic pathways, and protein assemblies), extrinsic factors (e.g. nutrients, xenobiotics, and host immune response), and evolutionary processes determine the composition and functioning of microbial ecosystems.

IV. **Modulate microbial communities and their interactions**. The tools and knowledge generated on single species functionalities and on community organisation, interactions, and properties, will serve as a basis for moving towards strategies for rational modulation of the microbiome. An iterative approach, combining high-throughput experimentation with modelling and machine learning, will pinpoint abiotic (pharmaceuticals, xenobiotics, food, prebiotics) and biotic (phages, probiotics, microbial species) strategies to shift or change microbial community compositions in a targeted manner. These modulations will range from precise removal or exchange of a single strain (e.g. a pathogen or a microbe carrying easily transmittable antibiotic resistance) to more radical partial or whole community transplantations. The ability to rationally modulate microbial communities will pave the way for new therapeutics and biotechnological applications.

V. **Expand and translate knowledge to other microbiomes**. The human gut microbiome will be used as an exemplary microbial ecosystem to chart interactions, understand gene functions, dissect the underlying mechanisms, and probe their impact and role in their ecological context. The accumulated knowledge of microbial function and biological principles, as well as the developed bioinformatics and experimental pipelines, will be used as a springboard to ask similar questions and start work on other relevant microbiomes. Studying other microbiomes in the human body (skin or respiratory or reproductive systems) or ones that humans are exposed to daily in their lives (those found in food and in the natural and built environments in which we live) would be the natural extension to these efforts. In conjunction with plans for field work and *in natura* measurements (Chapter 7: Planetary Biology), EMBL will seek to develop appropriate experimental platforms for cultivating, probing, and characterising microbes and microbial communities from such environmental settings.

EMBL's Approach

Understanding the Functional Diversity of Microbial Species and Strains

Current knowledge of microbial gene functions stems from a few model organisms (e.g. *Escherichia coli*, *Bacillus subtilis*) and pathogens, which bear little resemblance to most species found in the gut microbiota (Figure ME2A-B). Diversity in microbes is high, due to their long evolutionary history and fast reproduction. Hence the phylogenetic distances between bacterial species are much larger than those between eukaryotic organisms, which limits the utility of inferences about gene function and physiological context based on homology to current microbial model organisms. Not surprisingly, more than 60% of the genes in the human gut microbiome remain of unknown function, a **dark genetic matter** that impedes our understanding of their encoded functional capacities and contributes significantly to the outstanding body of proteins of unknown function found across all sequenced organisms (Figure ME2C). To bridge this gap, **EMBL proposes to generate the functional knowledge and resources required to establish new model organisms spanning the diversity of the human gut microbiome**.

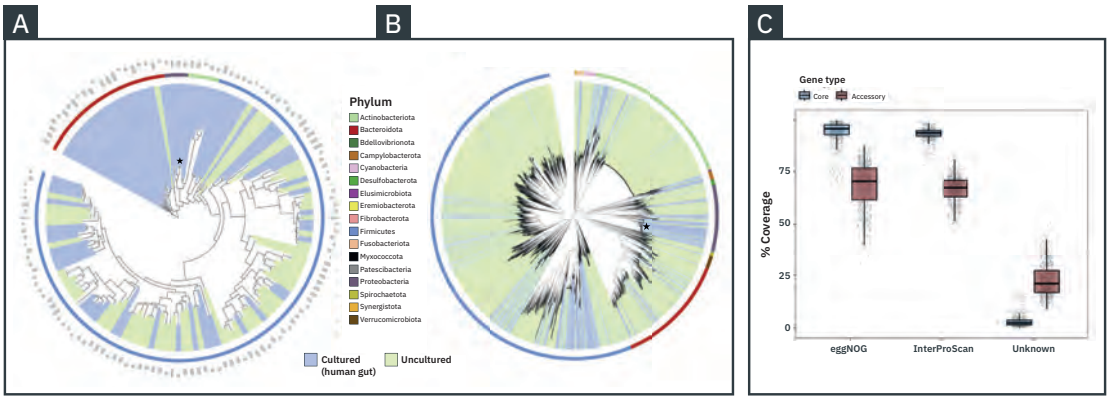


Figure ME2 | The gut microbial species diversity highlights current challenges – the lack of model species and many functionally uncharacterised proteins.
(A) Phylogenetic tree of the most prevalent and abundant bacterial species in the human gut. Species are shown in which prevalence >10% and relative abundance >10^{-5} in 2,803 healthy individuals across the world. *E. coli* (black star) is the only model organism in the list of species. The phylum of these species is indicated in the outer coloured ring. Most of the species that are cultured (light blue, inner colouring) are part of EMBL's culture collection. **(B)** Phylogenetic tree representing the 4,644 bacterial species found in the human gut, based on meta-analysis of >12,000 human gut metagenomes (Almeida *et al. Nature* 2019). Over 70% of these species are yet to be cultured (green, inner colouring). The location of *E. coli* is indicated as before. **(C)** Pangenome (the entire gene set of all strains of a species) analysis of the complete set of 205,000 novel genomes produced, following functional annotation by EMBL resources (eggNOG and InterProScan) reveals that, while most core proteins can be functionally annotated, our knowledge of accessory proteins is far more limited. Typically, 20% of the accessory proteins lack any functional annotation, highlighting the current microbial dark genetic matter (Almeida *et al. Nature* 2019).

Establishing Comprehensive Databases and Computational Resources

Metagenomics has enabled the vast functional diversity encoded in the human gut microbiome to be understood. With microbiome sequencing information being released at an ever-increasing pace, and with any two human individuals differing by more than 90% in terms of their microbiome strain content, while being more than 99% genetically identical, there is a need for a single comprehensive catalogue of human gut microbiota composition and functions. Propelled by work led by EMBL groups on sequencing and defining bacterial pangenomes, mapping strain diversity in microbiomes, and assembling genomes from metagenomics data, it has become evident that less than 30% of the approximately 4,500 bacterial species in the human gut currently have culture isolates (Figure ME2B). This makes harmonisation of data-driven efforts crucial to put future metagenomics studies in context, and to guide experiments that aim at dissecting these microbiomes at the molecular level.

EMBL will build on its experience as a leading provider of diverse computational resources in the life sciences, many of them, such as SpecI, mOTUs, MGnify, proGenomes, GenomeProperties, eggNOG, InterPro, STRING, iTOL, and iPATH, being intimately related to analysing microbiome data and/or representing their phylogenetic and functional diversity (Tech Dev Box TD1_ME). These databases will continue to be expanded, and will also serve as the basis for a new integrated resource, which systematically captures and catalogues all bacteria, viruses, and eukaryotes in the human gut microbiome. Information on genomes from isolated species and metagenome assembled genomes (MAGs) will be a particular focus, as they provide comprehensive and uniform taxonomic and functional annotation of genes, genomes, and pangenomes to drive functional hypothesis generation at the level of single microbial species and strains. They also dramatically increase the power to identify functional associations based on genomic context information. Key host metadata such as geographical distribution, age, and health status of associated samples will be curated and linked to these functional resources, to enable associations between functional capacities and environmental factors. The data coordination efforts to build this resource would unite various fragmented efforts in the field, remove redundancy, and ensure consistent quality controls. The large amount of data collected, curated, and managed will provide an unprecedented resource for investigating pangenome information and strain-level variation across microbiomes, not only those living in or on humans but also those found in the ocean or soil.

Developing Experimental Resources and Automated High-throughput Pipelines

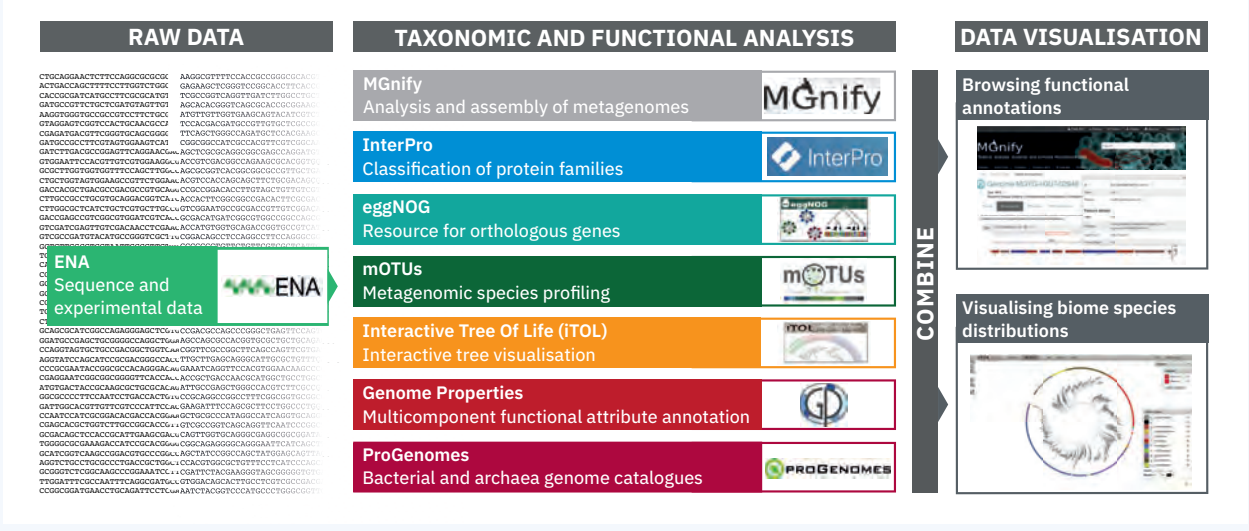
Automated experimental setups to cultivate and study gut microbes have recently started to emerge, with EMBL at the forefront of such developments (Tech Dev Box TD2_ME). Such setups enable the study of these still largely unknown microorganisms at the molecular level in controlled settings. EMBL aims to continue and expand these efforts. First, EMBL will isolate, biobank, and sequence thousands of strains from human individuals to build a repository for functional studies and for understanding the dynamics and properties of different microbial ecosystems. Working in close collaboration with national strain collections, other microbiome cultivation initiatives, and expert labs, EMBL will focus on assembling large strain collections of a selected set of abundant and prevalent gut bacterial species. These collections will be invaluable in efforts to link genes to functions and to other genes, e.g. genes co-occurring or co-evolving and required for a given function or phenotype. Second, EMBL will expand its present capacities to systematically cultivate, perturb, and monitor gut microbes in controlled environments and automated settings. This will involve building new tailored quantitative assays (fitness-, morphological-, genomics-, proteomics-, and metabolomics-based readouts) and pipelines in which microbes can be exposed in parallel to hundreds of ecologically relevant perturbations while their phenotypes are monitored (Tech Dev Box TD2_ME). Third, EMBL will establish genetic tools for synthetic biology approaches and will construct genome-wide mutant libraries, including single-gene knockout, knockdown, and overexpression libraries, for systematic functional studies. These libraries will enable systematic studies of gene function, as well as tailored mechanistic work on human gut microbes and their communities.

Technology Development Box TD1_ME | Microbiome computational tools and databases.

EMBL has pioneered the development of computational tools, databases, and web resources for microbiome research. These include ways to delineate prokaryotic lineages (proGenomes2 – Mende *et al. NAR* 2020), analyse phylogenetic trees (iTOL – Letunic and Bork, *NAR* 2019), profile shotgun metagenomes (mOTUs2 – Milanese *et al. Nature Communications* 2019), and assess eukaryotic microbial genome quality (EukCC – Saary *et al. bioRxiv* 2020). Further integrated computational workflows will increase utility to cater to various research areas and applications, from clinical microbiome biomarkers to diverse environmental microbial communities.

Functional analysis of the genetic diversity encoded in metagenomes is still challenging. Researchers at EMBL were among the first to catalogue the genetic and functional diversity of the human gut microbiome. Databases maintained at EMBL (Pfam, SMART, InterPro, UniProt, eggNOG) are key for functional annotations. EMBL scientists will ensure that these continue to incorporate the latest information, and will transfer annotations to all sequenced organisms.

Recently, over 200,000 bacterial genomes were assembled from human gut metagenomes, and over a billion distinct microbial proteins from diverse habitats were catalogued in EMBL databases (Almeida *et al. bioRxiv* 2020). To increase data accessibility and enable discoveries, integrated infrastructures (e.g. MGnify, Ensembl) and new concepts for data mining will be developed.

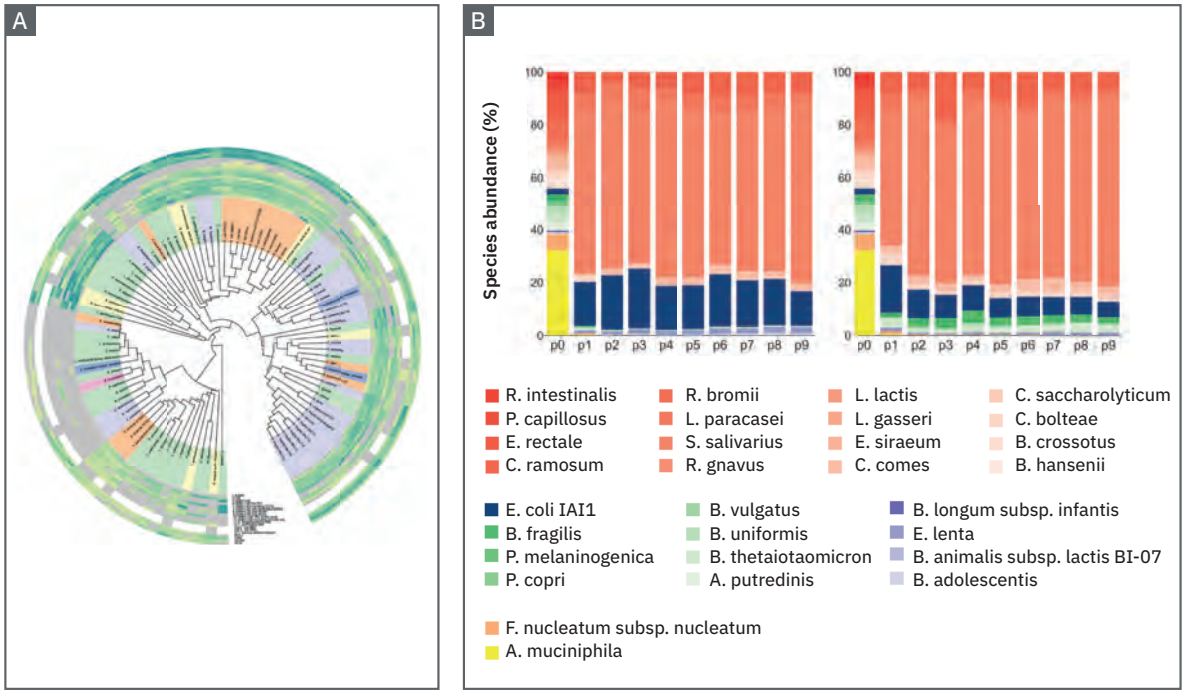


Technology Development Box TD2_ME | Automated microbiomics platforms.

Human microbiome research has been propelled by data-driven science, largely directed by genomics analyses of samples without cultivation. Although these approaches provide unique *in natura* insights into communities and permit associations to health and environmental parameters, they are not enough to understand the underlying causal relationships and mechanisms. EMBL has been at the forefront of establishing experimental setups to investigate the human gut microbiome.

(A) EMBL researchers from the Bork, Patil, and Typas groups have assembled a collection of prevalent and abundant human gut species (now containing >100 species), developed robust and automated cultivation pipelines, and mapped the metabolic capacities of these species (Tramontano *et al. Nature Microbiology* 2018). EMBL researchers have used this collection to systematically profile the interactions of gut microbes with environmental perturbations (Maier *et al. Nature* 2018; *bioRxiv* 2020) and with each other. **(B)** The same pipelines were used to assemble large numbers of complex communities. Here, two stable communities are shown over time, after mixing 32 species in different media, to profile emergent phenotypic traits of communities.

As a next step, EMBL is moving its suite of unique high-throughput reverse genetics (Nichols *et al. Cell* 2011; Kritikos *et al. Nature Microbiology* 2017; Galardini *et al. eLife* 2017; Brochado *et al. Nature* 2018; Zimmermann *et al. Nature* 2019) and MS-based approaches (Tech Dev Box TD_ME_3) from model microbes to less well-understood gut microbes. EMBL researchers are also developing more complex, high-throughput microbiomics setups (Figure ME3), with a special focus on interrogating personalised microbiome communities to empower precision medicine solutions.



Mapping Gene Function, Protein Complexes, and Cellular Pathways

Gut microbes have an enormous capacity to degrade and utilise nutrient sources; produce bioactive molecules, including essential vitamins for the host; sense and transduce signals; respond to and protect themselves from stress; interact with each other; train the host immune system; and fight off pathogenic intruders. To gain insights into this functional diversity and create the foundational functional knowledge required for new model organisms, EMBL will combine its unique experimental resources and automated pipelines, with cutting-edge technologies spanning omics to structures, and with computational approaches for data analysis and integration. For example, using strain collections and genome-wide mutant libraries in novel metabolomics- and proteomics-based read-outs will allow EMBL scientists to link genes to substrates and products, map enzymes and transporters to pathways and to their ligands, and chart the organisation of the metabolic network of these microbes (Tech Dev Box TD3_ME). Similarly, combining hundreds of genetic, chemical, or environmental perturbations with quantitative read-outs (e.g. fitness-based omics) will unravel genotype-to-phenotype relationships, and uncover gene function and organisation en masse (Tech Dev Box TD2_ME). Both the establishment of innovative and diverse quantitative omics readouts and the integration of multi-omics data lie within the core expertise and areas of excellence of EMBL. The data generated from these approaches will be used to chart the main functional units of a plethora of evolutionarily distant gut microbes, as well as provide insights into their function, regulation, and interconnections. The data will also shed light on the microbial dark genetic matter, making it possible to infer the functions of thousands of orphan proteins, and providing leads for further molecular characterisation through biochemical or structural investigations. Here, EMBL will capitalise on its expertise in structural biology (cryo-EM, X-ray crystallography, and NMR spectroscopy) to answer mechanistic questions about protein complexes of any level of complexity or size by determining high-resolution structures. These techniques are optimally suited to map diversity at the level of sequence, function, and physiological relevance onto the atomic level.

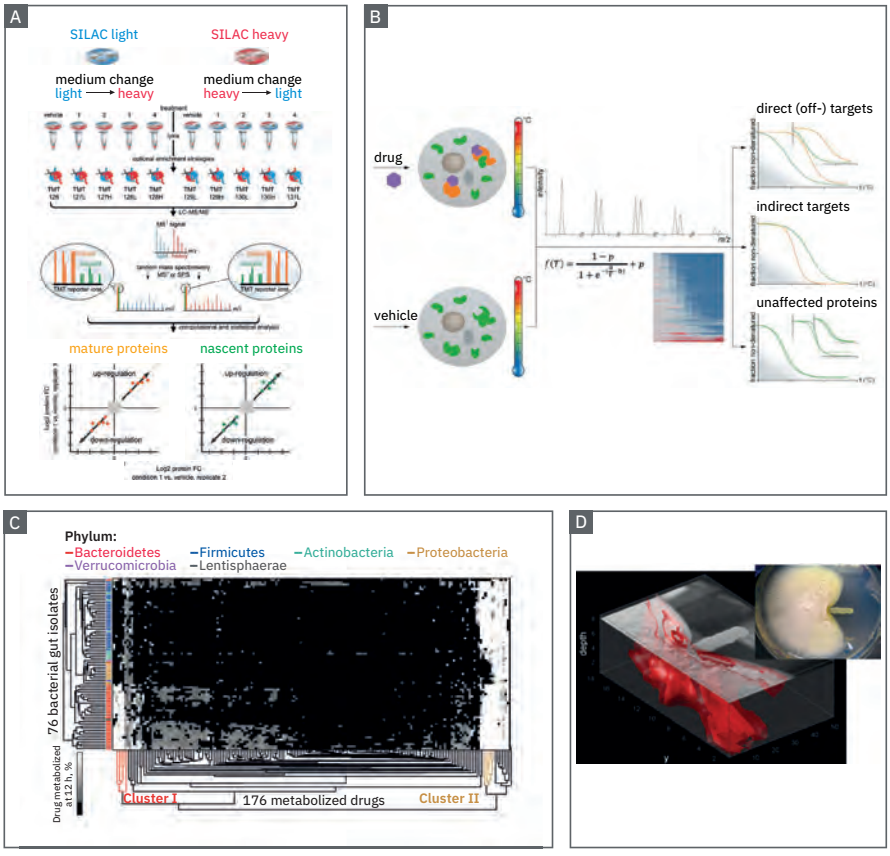
To get this ambitious plan off the ground, a workshop entitled ‘Unlocking the Gut Functional Diversity’ (Figure ME1) has been organised to bring together key stakeholders, including scientists, journal editors, and funders, from EMBL member states and beyond. The goal is to chart the framework of tools, approaches, and strategies required for generating comprehensive functional knowledge about prevalent and abundant gut species. This knowledge and the available tools will act as a foundation for establishing representative model organisms for this microbial ecosystem.

Technology Development Box TD3_ME | Mass spectrometry-based functional analysis.

Recent technical developments in mass spectrometry (MS) have enabled the characterisation of cells and their environment at the molecular level. Proteomics, metabolomics, and lipidomics measure the composition, interaction, and modification of macromolecules and chemicals to provide direct insights into molecular mechanisms of organismal functions.

(A) The Savitski Team have developed cutting-edge methods for unbiased determination of protein state *in vivo* in microbial and host cells (Thermal Proteome Profiling; Savitski *et al. Science* 2014; Becher *et al. Cell* 2018), and **(B)** for disentangling global protein degradation and synthesis through multiplexed proteome dynamics profiling (Savitski *et al. Cell* 2018). **(C)** The Zimmermann Group uses a combination of high-throughput metabolomics measurements, massive parallelised microbial culturing, and bacterial genetics to process up to 10,000 samples at a time. This unique setting allows the systematic identification of novel functional units of microbial strains and their communities, such as, high-throughput metabolomics analyses of clusters of human gut microbes, based on their xenobiotic-converting activity (Zimmermann *et al. Nature* 2019). **(D)** The Alexandrov Team is world-leading in spatial metabolomics approaches. These approaches can be used to quantify and visualise the chemical environment shaped by microbial colonies and the molecular interactions both between microbes, and between microbes and the host (Watrous *et al. ISME J.* 2013).

EMBL has diverse expertise in MS-based technologies, enabling world-class research to understand life at the molecular level. These technologies are key to moving microbiome research from associations to identifying the underlying molecular interactions in these microbial ecosystems.



Dissecting Interactions and Properties of Microbial Communities

As microbiotas are more than the sum of their individual members, they manifest community properties that the individual species do not exhibit or possess. These collective behaviours are relevant to the stability of the community, and to the way it interacts with its environment – both in terms of how microbiotas perceive and process signals, and how they respond to them. For the gut microbiota, these behaviours are relevant in their responses to the environment (e.g. xenobiotics) and to pathogenic intruders, their ability to degrade food, and their interaction with the host. Yet their prevalence and the molecular underpinnings that lead to these responses are largely unknown.

For an initial assessment, several EMBL groups performed a pilot experiment by comparing the sensitivity to 30 drugs of approximately 30 gut species in isolation or in a community setting. Communal phenotypes were frequent, with cross-protection (drug-sensitive species becoming resistant in the community setting) being more common than cross-sensitisation (Figure ME3B). Interestingly, these communal traits became less relevant at higher drug concentrations, indicating that communities have a robustness threshold against perturbations. As a proof of principle, EMBL identified the strains and enzymes providing protection against the drug niclosamide. In the future, this setup will be used to systematically probe different perturbations and communities.

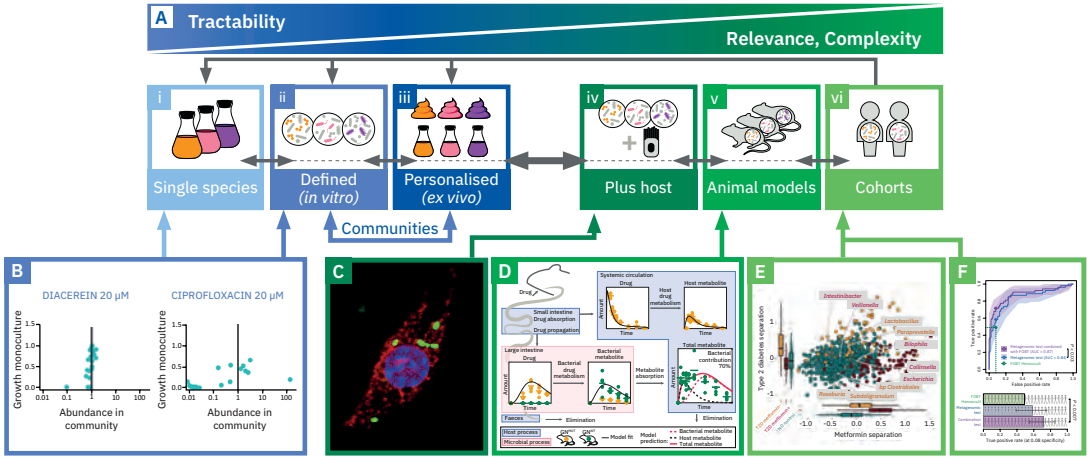


Figure ME3 | Experimental setups to study microbial ecosystems, from monocultures to microbiomes in clinical cohorts.
(A) Experimental approaches to study microbial communities, at various levels of complexity and tractability, that are employed by EMBL researchers **(i–vi)**. **(B)** Single species behave the same in isolation as in a community (ciprofloxacin), or the community exhibits collective behaviours (diacerein – all strains become resistant) in the presence of a drug. **(C)** Intracellularly growing *Salmonella enterica* Typhimurium (green) leads to re-trafficking of active cathepsins (red) to the nucleus (blue) of macrophage cells (Selkrig *et al. Nature Microbiology* 2020). **(D)** Quantifying microbial contributions to drug metabolism *in vivo* using bacterial genetics, gnotobiotic mice, and pharmacokinetic models, which include drug absorption, GI tract propagation, host and microbial drug conversion, and systemic drug elimination (Zimmermann *et al. Science* 2019). **(E)** Human cohort study disentangles the effect of type 2 diabetes and metformin medication on patients' gut microbiota (Forslund *et al. Nature* 2015). **(F)** The use of gut microbiome composition signatures as a biomarker to diagnose human colorectal cancer (Zeller *et al. Molecular Systems Biology* 2014; Wirbel *et al. Nature Medicine* 2019).

Understanding the Stability and Dynamics of Microbial Communities

Community-specific traits emerge through microbial interactions and are challenging to identify and prove solely through bioinformatics analyses of metagenomics data or experiments with bacterial isolates. Therefore, EMBL has recently been developing strategies to study microbial communities at various degrees

of complexity, from synthetic assemblies to *ex vivo* communities (Figure ME3; Tech Dev Box TD 2_ME). The former are artificially assembled communities, derived from mixing together individual strains, and the latter come from cultivating complex individualised communities directly from human stool. EMBL will amplify these efforts in the next Programme, with the goal being to **understand the impact of such interactions, their molecular nature, and their underlying general principles**. This will include studies of how interactions evolve over time and across dynamic environments; what their inter-individual variability is; and whether environments can foster or break communal behaviours, such as their intrinsic stability.

EMBL will focus on investigating the role of specific genetic variants, strains, and species in microbial interactions and community dynamics. Building on its established pipelines to systematically probe the gut microbe–medication interface, EMBL will expand to assessing the effect of nutrition, environmental changes, and xenobiotics such as drugs, excipients, pollutants, and food additives, on the stability and dynamics of different gut microbiome community assemblies. To do this, hundreds of communities in parallel with distinct, defined, and stable compositions will be combined with defined nutrients, environments, and perturbations in a high-throughput manner. These approaches will be coupled with high-content quantitative readouts to provide mechanistic insights into the underlying phenotypes. The results will provide a better understanding of how microbial and environmental factors shape community composition, and how emergent or collective phenotypic traits of these communities feed back to their environments; in this case, the human host. Ultimately this information will advance the ability to interpret microbiome data from clinical cohorts, and to understand the underlying associations to health and disease. It will also enable the development of predictive models of community dynamics and impact.

Dissecting the Role of the Gut Microbiota in Infection and Antibiotic Resistance

A comprehensive understanding of the molecular processes by which benign bacteria within the human microbiome become pathogenic, and the role of these dense communities in the development of antimicrobial resistance and the emergence of difficult-to-treat to pathogens, are topics of high interest. The human gut has an intrinsic ability to fend off intruders (colonisation resistance), but at the same time harbours numerous pathobionts (non-harming symbionts that under specific conditions can become pathological) and opportunistic pathogens (pathogens that normally do not cause disease in a healthy host, but can do when opportunity arises) as regular inhabitants. Examples of the latter are enterotoxigenic strains of *Bacteroides fragilis* or *Clostridioides difficile*, a leading cause of diarrheal illness. The interface between the human microbiota in different parts of the human body (gut, skin, mouth, nasopharynx, lung, urinary tract, and vagina) and pathogenic subpopulations is complex and has been largely elusive to date.

A fundamental question is: what are the signals that trigger or prevent the transition of a silent resident pathogen to an active one, causing dysbiosis or even disease? Combining state-of-the-art bioinformatics and molecular biology, systems biology, and structural biology expertise, EMBL is in a unique position to reshape molecular infection biology and place it in the context of the holobiont (microbiome and host) and the environment. Research on the molecular players and mechanisms used by pathogens during infection lies at the core of infection biology and is widespread across sites at EMBL (Chapter 5: Infection Biology). These research themes range from characterising secretion machines and the molecules that bacterial pathogens use to hijack different parts of host biology, to monitoring pathogen physiology and signalling during infection. EMBL will take advantage of its ability to systematically investigate and perturb microbial communities to identify what opens up those windows of opportunity for pathogens, to understand if there are ways to better contain the enemies within, and to discover the effects of removing these species from communities, or exchanging them with avirulent variants. EMBL will also focus on understanding the molecular entities and machines that microbes use to fight pathogenic intruders and that pathogens use to counterattack – from small molecules to secreted proteins and from phages to large protein machines that act as weapons (Figure ME4).

The technologies available at EMBL enable these studies to be carried out across scales, from atomic resolution of isolated molecules to the visualisation of these molecules within whole cells.

Moreover, the high density and diversity of the gut microbiota (10^{13}–10^{14} regular microbial inhabitants representing 3,000–4,000 species and many more strains, being attacked by 10^{14}–10^{15} phages) facilitates the exchange of genetic material, i.e. horizontal gene transfer (HGT). This high density and the strong environmental exposures (such as new microbes coming with food or selective pressure from drug treatments) make the gut microbiome the perfect reservoir **for the development and spread of antimicrobial resistance (AMR).** AMR genes are frequently carried on mobile genetic elements that can move within and between bacteria, promoting the development of difficult-to-treat superbugs in such hotspots of microbial interaction. Understanding AMR transmission is critical for human health and is a core part of research at EMBL (Chapter 5: Infection Biology). In the case of the gut microbiome, EMBL researchers are interested in using their expertise in bioinformatics, microbiology, genetics, and structural biology to chart how AMR spreads in these communities, mapping transmission routes, underlying mechanisms, and possible Achilles' heels (i.e. the fitness cost it confers to the community). This understanding is essential to identify strategies to limit or even prevent the development and spread of AMR to pathogens.

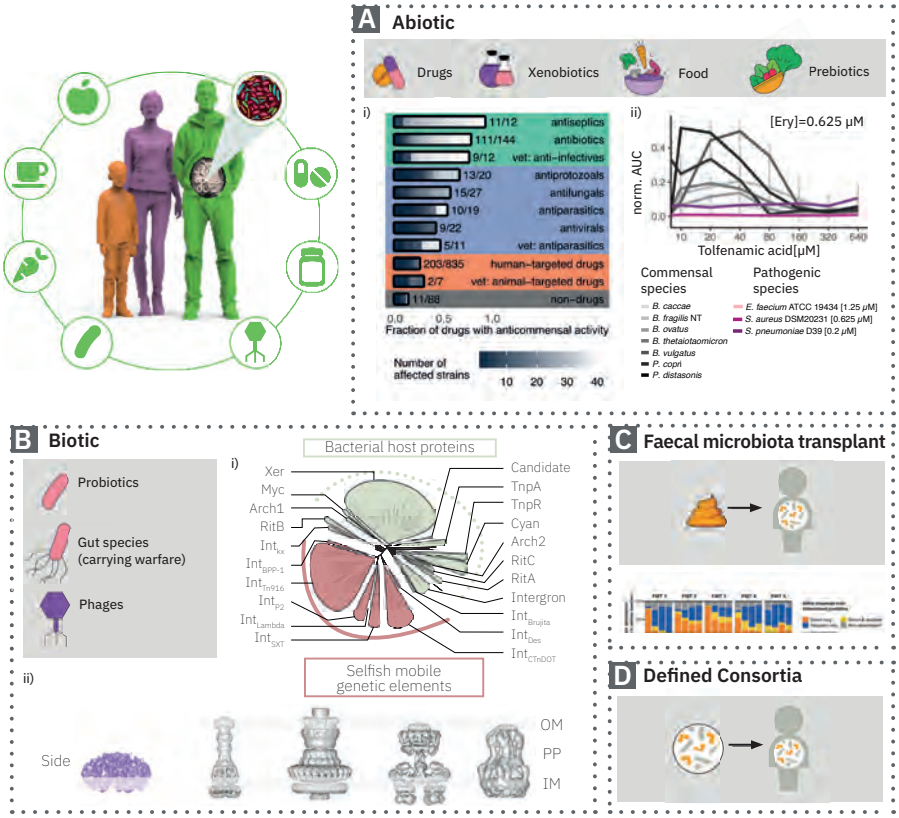


Figure ME4 | Towards a rational modulation of microbial ecosystems such as the human gut.
(A) Abiotics, such as drugs, food, and prebiotics, and combinations thereof, can be used to shift microbiota compositions: **(i)** many non-antibiotic drugs can directly inhibit specific gut microbes (Maier *et al. Nature* 2018); **(ii)** species-specific outcomes of drug combinations (Brochado *et al. Nature* 2018) can be used to rescue the collateral damage caused by antibiotics on gut microbes. Tolfenamic acid, a painkiller, alleviates the effect of the antibiotic erythromycin in several commensal species (black traces), but allows erythromycin to retain activity against pathogens (purple traces) (Maier *et al. bioRxiv* 2020). **(B)** Biotics, such as probiotics, phages, and species carrying secreted proteins targeting other bacteria or the host can be employed to shift bacterial community composition: **(i)** a systematic survey of genetic elements that can autonomously move between bacterial species reveals their diversity and evolution, and provides new tools for mapping HGT routes and traits (Smyshlyaev *et al. Biorxiv*, 2019); **(ii)** the first electron microscopy structure of a Type VII secretion system and its comparison with other known secretion systems (Beckham *et al. Nature Microbiology* 2017). **(C)** FMT can change the microbiota of the recipient: mapping success across individuals (Li *et al. Science* 2016). **(D)** Defined consortia can be used to enable targeted rational modulation of the gut microbiota.

Placing Microbial Communities in their Ecological Context

Microbial communities are largely shaped by interactions with their environment, with which they form a complex and dynamic ecosystem. To investigate these complex interactions and to separate causes from consequences, EMBL will use experimental settings that mimic the natural environment and probe the effects of the environment to the community and vice versa via multi-level molecular profiling. The goal is to gain mechanistic insights into the functional interplay between microbes and their environment, which in case of the human gut microbiome is the host. The generated results will expand the interpretation of available sequencing data and will eventually enable targeted manipulations of microbial communities. Furthermore, the general framework proposed will provide a roadmap to study other microbiomes, and how molecular signals and pressures within these microbial communities shape their composition and interactions within the ecosystem.

Impact of the Functional Microbiome Output on the Host

To understand how changes in microbial community composition and physiology affect the host, EMBL will use *ex vivo* host models, such as human- and animal-derived cell lines, primary cells, and organoids (Figure ME3). These simplified host systems will be exposed to microbial extracts or communities. Tailored host readouts (reporter assays, high-throughput microscopy, multi-omics, and FACS with immune markers) will be used to assess the impact of synthetic microbial or *ex vivo* communities. This line of experimentation will enable the identification of causal effects of the microbiota on specific host responses, such as neurotransmitter secretion by enteroendocrine cells, stimulation of immune cells and drug responses of different cell types. In addition, it will point to the perpetrator species and strains, and with the help of bacterial genome-wide mutant libraries and multi-omics readouts will reveal the underlying molecules, genes, pathways, and mechanisms of these interactions. The experimental setups developed here will link to complementary approaches used for mapping the effect of the microbiota on drug responses of cancer cells (Chapter 6: Human Ecosystems).

Reciprocal Microbiome–Host Interactions in Complex Models

The host plays a crucial role in shaping its colonising microbial communities and the multifactorial impact of the host on these microbes is hard to fully reproduce *in vitro*. EMBL will therefore employ gnotobiotic animals colonised with communities at various complexity levels (including genetically engineered marker strains or genetically barcoded libraries of a specific species) to carry out spatially resolved assays of compositional adaptation to dietary, chemical, or infectious perturbations (Figure ME3). These data will be paired with omics and physiological measurements, allowing direct comparisons with results from the more controlled *in vitro* settings described above.

The microbiota also influences the host physiology at various levels, which are invisible without the use of whole organisms to determine behaviour, reproduction, and organ function. EMBL will exploit its resources and expertise in mouse biology to re-derive germ-free and/or genetically modified mouse strains. This can include humanised mice (mice with transplanted human gut microbiota), and specific target knockout mice such as those lacking liver CYP450 enzymes, which results in a compromised drug metabolism. Mouse models and readouts will be selected to investigate microbial community influence on drug responses, cancer development, the germline, the liver, and the gut–brain axis. The knowledge from the *in vitro* and *ex vivo* data, especially on causal genes, strains, and communities, will facilitate the prioritisation of experiments in genetically modified mice. The developed models and the data generated will be instrumental to both reproducing interactions seen in *ex vivo* data, and testing associations drawn in large human microbiome

cohort studies. Ultimately, mirroring the effect of microbiota on health and disease in appropriate animal models will not only help scientists to identify causal effects and mechanisms, but will also provide a solid basis for testing intervention strategies for rational modulation of the composition of the gut microbiota.

Modulating Microbial Communities and their Interactions

The microbiome has become a primary therapeutic target, with thousands of clinical trials currently taking place globally. Most of these studies aim to modulate the microbiome towards a health-promoting state for its human host. The means to do so vary immensely, from antibiotics or phages to eliminate community members, to prebiotics to promote certain strains of beneficial bacteria, and from probiotic strain cocktails to complement the community with selected members or traits, to faecal microbiota transplantations (FMT), which involve transplanting a sample of the microbiota from one person to another (Figure ME4). However, these interventions are mostly based on empirical observations of particular microbiome properties, lacking systematic insight into the repertoire of underlying factors and molecular mechanisms that impact the outcome of the intervention and dictate its success. Hence most of these trials are bound to fail or have reproducibility issues when the size of the study increases and the inter-individual complexity and variability of these communities increases. Therefore **the rational design of microbiome modulation strategies remains the pinnacle in translational microbiome research**, which aims to develop therapeutic strategies to treat infections, allergies, inflammatory conditions, metabolic syndromes, cancer, and many more dysbiosis-associated human diseases. A sufficient understanding of microbial processes, in their interactions with each other and the host, are the basis for moving towards more controlled strategies for modulating the gut microbiome. EMBL proposes to combine bottom-up and top-down approaches to achieve this. Bottom-up approaches will be based on mapping and predicting interactions between molecules and microbes within the community context, such as those that occur via mobile elements or phages, and interactions between microbes and the host. Top-down approaches will rely on computational analysis of complex data from clinical microbiome intervention studies to identify patterns associated with successful interventions or to validate findings from defined community settings.

Precision Modulation of Communities

EMBL will utilise its automated high-throughput pipelines to systematically assess the effects of simple **abiotic** modulators, such as chemicals, prebiotics, and food compounds or additives, or combinations of these, on single microbes and communities (Figure ME4A). This will provide foundational knowledge about how to specifically suppress or promote the growth of certain species or to alter community compositions. Such modulations could be minimal, with the overarching goal being to remove specific strains while retaining the overall community composition intact. For instance, replacing an AMR pathobiont by its cured derivative would increase treatment options if this becomes pathogenic for the human host. The use of *ex vivo* communities, strain collections, and genome-wide mutant libraries, together with the ability to rapidly run millions of such experiments, will allow EMBL researchers to systematically assess the role of strain diversity and genetic traits on the outcome of such perturbations. Using machine learning approaches on such rich experimental data will make it possible to predict modulators and responding bacterial species and strains, as a first step towards rationalising more complex modulations.

Investigation of single-agent live microbiome (**biotic**) modulators, such as probiotic bacterial strains or phages, will require first pinpointing the microbial genes and metabolites involved in their interactions with microbial community members or the host. Bioinformatic analyses based on protein homology and genomic organisation can expedite the identification of secondary metabolites, toxins, secretion, and adhesion machines. The increase in genomic sequence information fuelled by large-scale metagenomics assemblies

will further facilitate such strategies, especially ones that use genomic context information to fish out novel microbial weapons that bear no resemblance to previous systems. EMBL has been pioneering this area of research with multiple databases and tools (InterPro, eggNOG, Pfam, STRING) that map protein families or genomic context information and use it to catalogue or identify new functions. These resources and tools will continue to develop in the next EMBL Programme, and aim to take advantage of and cater to the continuing increase of microbiome metagenomics data. *In silico* predictions will directly interface with experimental approaches, which aim to systematically map the effects of bacterial strains, species, and communities on the composition of personalised microbial communities (*ex vivo* communities from different individuals). As bigger parts of the microbial armoury are unravelled, selected molecules, such as proteins or metabolites, will provide candidate agents for microbiome modulation. Experimental follow-up combining omics and analytical technologies will provide mechanistic insights into interesting modulating agents.

Another potential area of study involves **mobile genetic elements** and lytic **phages**, which have been largely unexplored as microbiome modulators. Due to the spread of AMR, the interest in using phages to treat pathogenic infections has increased, with some clinical trials currently ongoing. There are many potential advantages of phage treatment over antibiotics, such as high specificity, low inherent toxicity, and minimal disruption of surrounding tissues or normal flora. However, there are also many issues that remain to be resolved, such as delivery to the infection site or resistance development. What EMBL can contribute to the field at this stage are improved bioinformatics tools for mapping and characterising both mobile genetic elements (Figure ME4B) and the vast phage arsenal present in the human microbiome and in environmental microbiomes. This information will improve understanding of the resistance reservoirs carried within the microbiotas of individuals and the specificity of phage treatments as microbiome modulators.

Whole Community Transplantation

Faecal microbiota transplantations (FMT) are confounded by various factors that affect the success rate of this therapy. EMBL will employ data-driven approaches that involve the analysis of microbiome data from ongoing and past clinical interventions across multiple indications. Previous such work by EMBL (Figure ME4C) of strain dynamics following FMT will be complemented with data from interventions involving complex dietary shifts or phage cocktails. The goal will be to deduce the principles by which gut bacterial taxa or communities are resilient to such perturbations and associate the characteristics of modulations (such as the FMT donor) with success rates. This effort will be empowered by our increased understanding of microbe–microbe and microbe–host interactions. Overall, these approaches will provide a framework for testing any general characteristics that underlie success rates, both in *in vitro* or *ex vivo* communities or animal models.

Understanding how individual human host differences factor into the outcome of FMTs, together with a molecular understanding of microbial communities and ecosystems, will enable predictive modelling of microbiome modulations using defined microbial consortia. These interventions will bestow desired traits and expected molecular impacts on the targeted gut community, such as the elimination of pathogens or microbes carrying AMR. This modelling attempt will not only elucidate what might have to be supplemented to the recipient's microbiome or which pathogens are to be displaced, but will also identify the microbial consortia required to achieve the desired effect. The resulting predictions can be tested using synthetic communities and faecal *ex vivo* cultivation in animal intervention experiments, yielding results that can be used to further improve the models.

Expanding and Translating to Other Microbial Ecosystems

Although EMBL's track record of studying microbial ecosystems is focused primarily on the gut microbiota, diverse microbial communities are found in other parts of the human body, including the entire digestive and respiratory systems, the skin, reproductive organs, and several internal organs. In addition to bacteria, there is a plethora of fungi, phages, and protists within these communities. Over the past few years, advances in sampling and bioinformatics analyses have led to an increased appreciation of the diversity and roles of these less-studied organisms within microbial communities. The computational and experimental approaches pioneered for gut bacteria will act as a roadmap to facilitate the cataloguing and functional dissection of other microbiota members and microbiomes. EMBL will foster research in these areas, aiming to gain a better understanding of the collective role of microbes in human health, the degree of interplay between the different human microbiotas, and the role of the environment in shaping microbiome composition and phenotypic traits.

Humans are exposed to microbes and microbiotas at every step of their lives – through the food chain, social interactions, and via their natural and built environments. Scientists are only now starting to understand the microbial exposome in a quantitative and molecular way, including what it is, its influence on humans and life on the planet, and how much human activities change microbiome composition and behaviours. The computational tools for mapping this microbial diversity and its functional underpinning will be key for moving towards a better understanding.

As EMBL's experience in cultivating microbial species increases, and in collaboration with European institutes that have decades of expertise in specific microbial ecosystems, EMBL will create automated experimental setups to map species and functional diversity for defined microbial communities outside of the human microbiome. This includes building strain collections, frameworks for systematically monitoring perturbations and microbial phenotypes, and tailoring these advanced molecular technologies to be applicable to the study of these systems. Employing these approaches, EMBL and collaborators can start assessing the principles (e.g. stability, communication, competition versus cooperation, emergent behaviours, gene transfer) that define the organisation of these microbial communities and facilitate niche specificity. Integration of this information with geography and ecological measures will be the next challenge to conquer for big data and machine learning-based approaches. The ultimate goal will be to understand the flow of information and the overarching principles that govern stability of microbial ecosystems and (evolution of) life on Earth.

Impact

Of all microbial ecosystems, the human gut microbiome is currently by far the most studied. However, despite intense research there are still fundamental gaps in knowledge of microbial genes, community behaviours, and microbial interactions. EMBL proposes a plan to systematically fill these gaps through a coordinated effort, combining cutting-edge computational and experimental approaches. Many aspects of human physiology are linked to the gut microbiome, including obesity, immune system function, and even mental health. EMBL's unique technology platforms will form essential and foundational resources for the scientific community, in member states and beyond.

The molecular insights gained from the approaches EMBL will undertake in the next Programme will ultimately help scientists understand how genetic and environmental factors shape microbial community composition, the collective phenotypic traits of these communities, and their impact on the environment. Ultimately, this information will enable scientists to rationally design therapeutic interventions targeting the microbiome to provide health benefits for individuals and society. The developed approaches, tools, resources, and data integration frameworks will pave the way for future studies of other microbial ecosystems, such as ones found in the natural or built environment, the composition and diversity of which we are only now beginning to map.