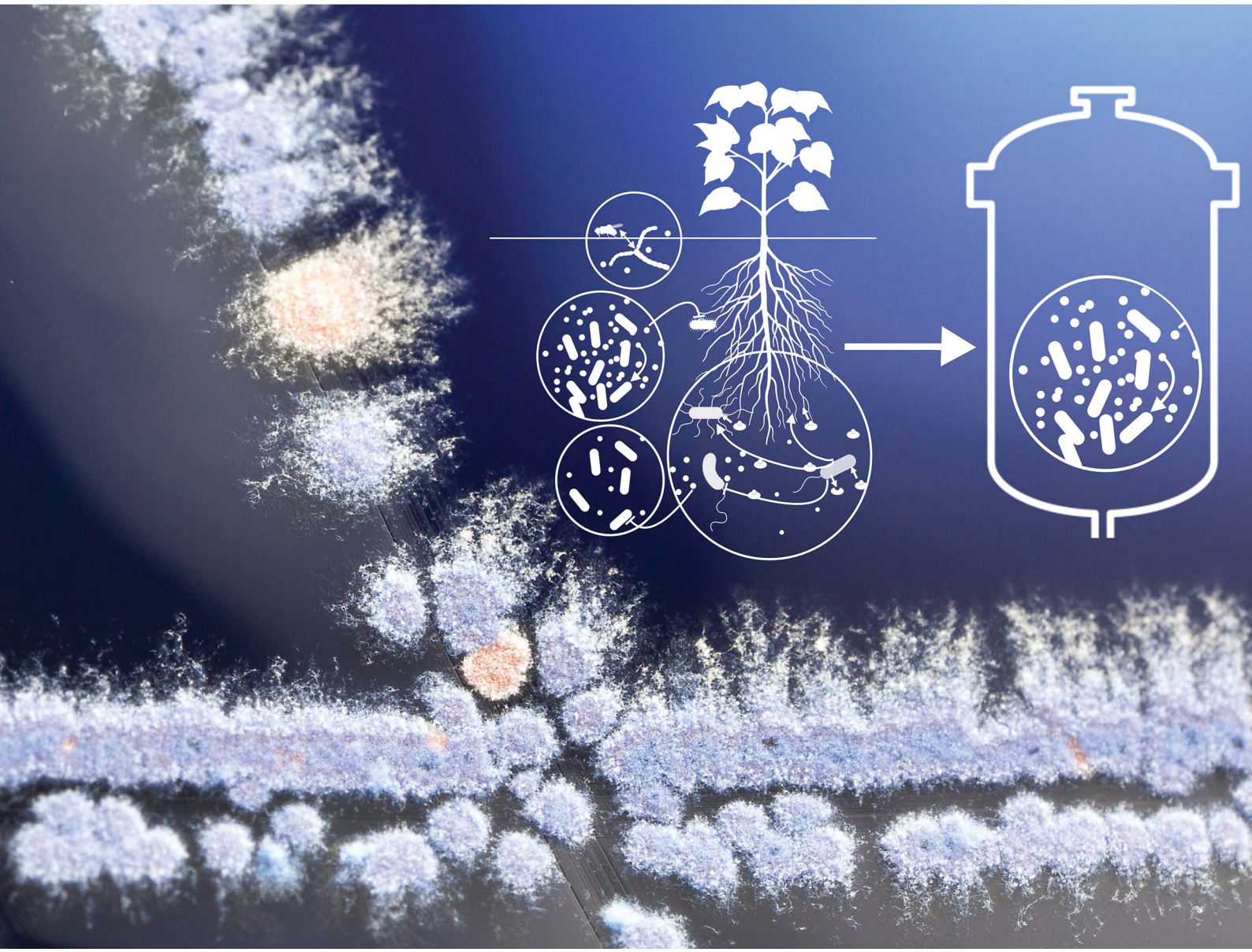
Volume 43
Number 2
February 2026
Pages 273–394

Natural Product Reports

rsc.li/npr



ISSN 0265-0568

ROYAL SOCIETY OF CHEMISTRY

HIGHLIGHT

Francesco Del Carratore and Rainer Breitling
Engineering microbiomes for natural product discovery and production

Natural Product Reports

ROYAL SOCIETY OF CHEMISTRY

HIGHLIGHT

View Article Online
View Journal | View Issue

Cite this: *Nat. Prod. Rep.*, 2026, **43**, 301

Engineering microbiomes for natural product discovery and production

Francesco Del Carratore*[a] and Rainer Breitling *[bc]

Covering: 2021 to 2025

Microbial communities represent a vast and largely untapped source of natural products with potential applications in various fields, including medicine, agriculture, and the biomanufacturing industry. Secondary metabolites play a crucial role in mediating interspecies interactions within these communities, influencing their structure and function. Recent advances in microbial genetic engineering and multi-omics technologies have enabled the harnessing of these interactions for enhanced natural product discovery and production. These techniques, coupled with systems biology and mathematical modelling, allow for the rational design and manipulation of microbial consortia to elicit the expression of cryptic biosynthetic gene clusters and to optimize the production of desired compounds. Additionally, direct mining of microbiomes using metagenomics, metatranscriptomics, and metabolomics has revealed a wealth of novel biosynthetic gene clusters and secondary metabolites with potential therapeutic and industrial value. Despite the challenges associated with cultivating and characterizing diverse microbial species, ongoing advancements in computational tools and data analysis are rapidly expanding our ability to explore and exploit the seemingly inexhaustible reservoir of natural products hidden within microbial communities.

Received 7th May 2025

DOI: 10.1039/d5np00038f

rsc.li/npr

1 Introduction

Bacterial secondary metabolism is a rich source of powerful bioactive compounds, which play a pivotal role in the interactions governing microbial communities,[1] in addition to being of great commercial and clinical importance. This means that complex microbial communities rather than monocultures of single strains could be the most obvious source to tap for natural product discovery and development. Based on this fundamental insight, there has been a long tradition of using microbial co-cultures to elicit the expression of cryptic and/or silent biosynthetic gene clusters (BGCs).[2] Similarly, the supplementation of elicitors that simulate naturally occurring interspecies interactions is routinely used for the discovery of natural products.[3–5] Recent technological advances in microbial genetic engineering[6] and microbiomics[7,8] have opened up an array of novel and exciting opportunities for harnessing microbial interspecies interactions for natural product discovery and production. For example, engineering small microbial communities into distributed natural products production systems is now a realistic prospect,[9] and dynamic division of labour (DDOL), mediated by horizontal gene transfer, can be employed to ensure the stable maintenance of the productive community.[10] Moreover, constant improvements of molecular profiling technologies allow the natural product community to access the seemingly inexhaustible reserve of secondary metabolites represented by microbiomes from a highly diverse range of environments, from the deep gut to the deep sea.[11–14] Advances in multi-omics data acquisition and analysis are increasingly freeing us from the constraints posed by the limited culturability of isolated microbiome constituents with regards to natural products discovery,[15] as they can be applied directly to the non-cultured microbiome *in situ*. At the same time, the insights emerging from this work are highlighting the extent to which we need to rethink current cultivation methods, as we are aiming for microbiome engineering and the exploitation of synthetic communities and co-cultivation approaches on a larger scale.

[a]*Department of Biochemistry, Cell and Systems Biology, Institute of Integrative, Systems and Molecular Biology, University of Liverpool, Liverpool L69 3BX, UK. E-mail: francesco.del-carratore@liverpool.ac.uk*

[b]*Manchester Institute of Biotechnology, Faculty of Science and Engineering, University of Manchester, Manchester M1 7DN, UK. E-mail: rainer.breitling@manchester.ac.uk*

[c]*Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore*

2 The role of secondary metabolites within microbial communities

Secondary metabolites exert a profound influence on the dynamics of microbial communities, shaping their structure, function, and stability.[16–18] These small extracellular molecules,

produced and released by microorganisms, often in response to external stimuli, act as crucial mediators of interspecies and intraspecies interactions, facilitating competition, cooperation, and communication within complex microbial networks[1,7,19] (Fig. 1). It is often the case that during the stationary phase of growth, when nutrients are limited, bacteria upregulate their secondary metabolism and secrete bioactive compounds in the extracellular environment.[20] For example, it has been shown that the soil-dwelling bacterium *Streptomyces coelicolor* switches on the production of different antibiotics in response to phosphate starvation.[21] Similarly, iron-starvation is a well-studied trigger for the production of siderophores in a wide range of bacteria.[22–24] In both examples, the production of secondary metabolites shapes microbial communities through competition. Compounds such as antibiotics and bacteriocins (bacterial produced antimicrobial peptides) cause direct inhibition of other bacteria competing for the same resources in the same spatial niche.[25,26] On the other hand, compounds like siderophores, which are high-affinity iron-chelating compounds, are often produced to scavenge iron from the environment, drastically limiting its availability to other microorganisms (competition) or by helping others which are capable of exploiting the same siderophore.[22] Similarly, biosurfactants can be produced to enhance nutrient availability or accessibility, such as rhamnolipids (a type of glycolipid) primarily produced by *Pseudomonas aeruginosa* to access hydrophobic substrates like hydrocarbons or to disperse biofilms of competing bacteria.[27,28] Furthermore, other secondary metabolites act as signalling molecules, facilitating communication within and between microbial species. These signals can elicit a range of responses in recipient cells, including alterations in gene expression, biofilm formation, virulence factor production, and the synthesis of secondary metabolites.[29] *N*-acyl homoserine lactones (AHLs) are a well-studied example of this kind of molecule. AHLs are secondary metabolites commonly used by Gram-negative bacteria in quorum sensing. Quorum sensing is a process that regulates several cellular activities (*e.g.*, antibiotic production, biofilm formation and virulence) depending on population density.[30,31] Moreover, AHLs have been implicated in interkingdom interactions, such as the communication between symbiotic bacteria and their host.[32–34] Geosmin represents another example of a secondary metabolite believed to be involved in interkingdom interactions. It is believed that this compound, usually produced by the members of the genus *Streptomyces*, acts as an olfactory signal that attracts certain insects such as the vinegar fly *Drosophila melanogaster* and the mosquito *Aedes aegypti*, helping with spore dispersion.[35–37] To sum up, bacterial secondary metabolites are often bioactive compounds secreted in the extracellular medium, driving microbial community dynamics by a complex web of competitive, cooperative and communicative interactions.

3 Microbial co-cultures can elicit the expression of cryptic and/or silent biosynthetic gene clusters (BGCs)

Thanks to computational tools such as antiSMASH,[38] we are able to identify a very large number of biosynthetic gene clusters (BGCs) responsible for the production of secondary metabolites,[39] thus dramatically reducing the number of cryptic BGCs. However, only a small fraction of these BGCs have been characterised and associated with their end compound.[40] This is, at least partially, due to the fact that a large number of bacterial species struggle to grow in lab conditions, while they thrive in their ecological niche.[41] Moreover, bacterial BGCs are often expressed in response to external stimuli which are unknown or not easily reproducible in controlled environments,[42] thus remaining silent in laboratory conditions. For example, secondary metabolites are often produced as a response to interspecies interactions,[43] and this phenomenon has been

Francesco Del Carratore

Dr Francesco Del Carratore is a Lecturer (Associate Professor) in Computational Biology at the University of Liverpool. He earned his BEng in Biomedical Engineering and MEng in Chemical Engineering from the University of Cagliari, before completing his PhD in Bioinformatics at the University of Manchester in 2018. Subsequently, he held postdoctoral and research fellow positions as a Computational Biologist at the University of Manchester. Since joining the University of Liverpool in 2023, Dr Del Carratore has established his research group in Computational Biology, directs the MSc Bioinformatics programme, and serves as an associate director of the Computational Biology Facility.

Rainer Breitling

*Professor Rainer Breitling studied biochemistry at the University of Hannover and received his PhD from the Technical University of Munich in 2001. He held postdoctoral positions as a bioinformatician at San Diego State University and the University of Glasgow, followed by professorial appointments at the Universities of Groningen and Glasgow. Since 2012 he has been Professor of Systems Biology at the Manchester Institute of Biotechnology, University of Manchester, currently in an honorary capacity. In January 2025 he joined the Bioinformatics Institute of A*STAR in Singapore, where he is responsible for the computational developments underpinning advanced synthetic biology applications.*

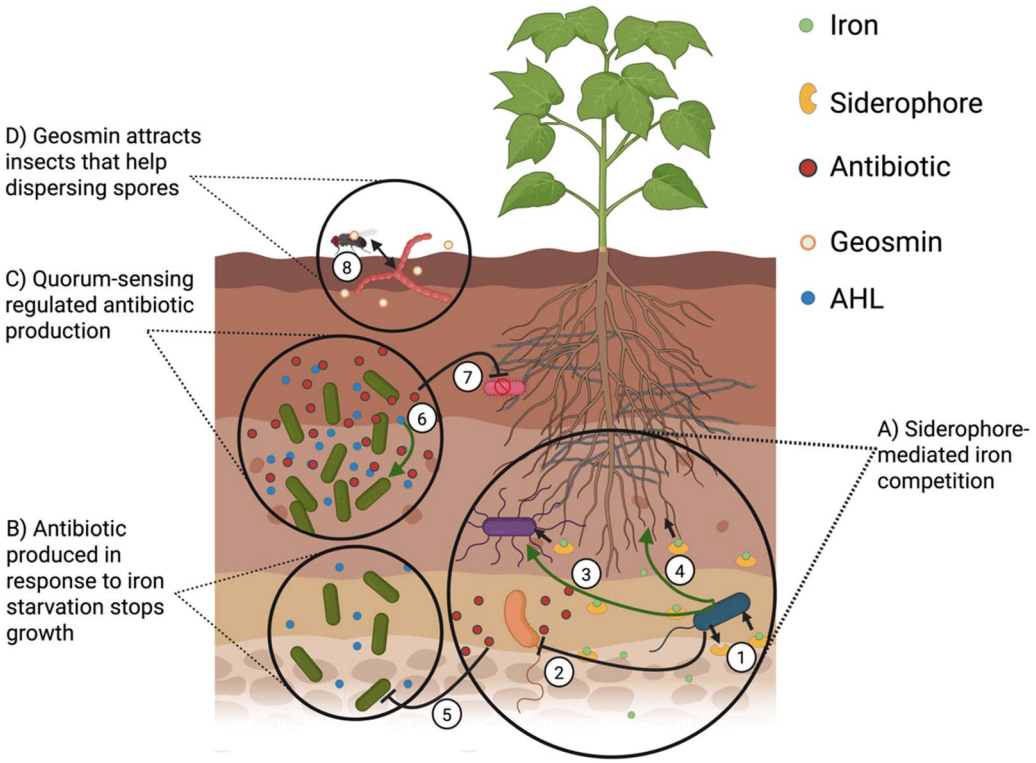


Fig. 1 Secondary metabolites as mediators of microbial community dynamics: this schematic illustrates how secondary metabolites released by microbes shape community structure. (A) Iron-competition: the blue bacterium releases a siderophore which strongly binds to iron. In turn the siderophore facilitate iron uptake for the producer (1) and the other organisms able to exploit it (3 and 4). The orange bacterium, not able to take up the siderophore, goes into iron starvation and starts producing an antibiotic as stress response (2). (B) The antibiotic released by the orange bacterium inhibits the growth of the green bacteria (5), which does not produce enough AHL (Quorum Sensing molecule) to trigger antibiotic production. (C) Far from the orange bacterium, the green bacteria can grow and produce enough AHL to self-induce antibiotic production (6), which kills the red plant pathogen (7). (D) Through the same regulatory system that controls sporulation, Streptomyces produces the volatile compound geosmin. Geosmin attracts insects, which in turn help dispersing the spores (8). Created in BioRender. Del Carratore, F. (2025) https://BioRender.com/ium6gv7.

exploited (Fig. 2). For example, supplementation of the growth medium with cell extracts or parts of the cell walls of other organisms has been successfully used to mimic interspecies interactions, leading to diversified secondary metabolite production.[44,45] Secondary metabolite production can also be elicited by signalling molecules secreted by plants and animals. For example, several catechol-containing molecules, such as epinephrine, have been shown to elicit the production of various secondary metabolites in *Streptomyces* sp. MBT84.[5] While proven successful, this approach only exploits unidirectional interactions. Conversely, microbial co-culturing can be used to successfully exploit bidirectional interactions for the enhancement of secondary metabolite production as well as the discovery of new natural products.[2,46] For example, nutrient competition might elicit the production of an extracellular signalling molecule by one organism that in turn elicits the production of an antibiotic from the other. It has been shown that four bacterial strains, *Lactobacillus helveticus* KLDS1.9207, *Enterococcus faecium* KLDS4.0352, *Lactobacillus reuteri* KLDS1.0737 and *Enterococcus faecalis* KLDS4.0313 induced the production of the bacteriocin plantaricin by *Lactobacillus plantarum* KLDS1.0391.[47] Such an effect was only seen when live cells of the partner organism were present in the culture.[47] Diffusible molecules are often responsible for such interactions. For example, when cultured together, *Wickerhamomyces anomalus* and *Monascus purpureus* led to a significant increase in production of 13*S*-hydroxyoctadecadienoic acid, a putative quorum-sensing molecule. In turn, this dramatically boosted the production of Monacolin K and natural pigments by *M. purpureus*.[48] For the most part, attempts to enhance production, or eliciting the production of novel natural products through co-culturing have been characterised by trial and error. However, multi-omics approaches, mathematical modelling and synthetic engineering will allow for the rational engineering of more complex and effective microbial communities.

4 How can we exploit this?

4.1 The mathematical modelling of these interactions is a necessary step

As in any complex biological system, it is important to understand that a reductionist approach focused on simply understanding each individual interaction is not enough to fully

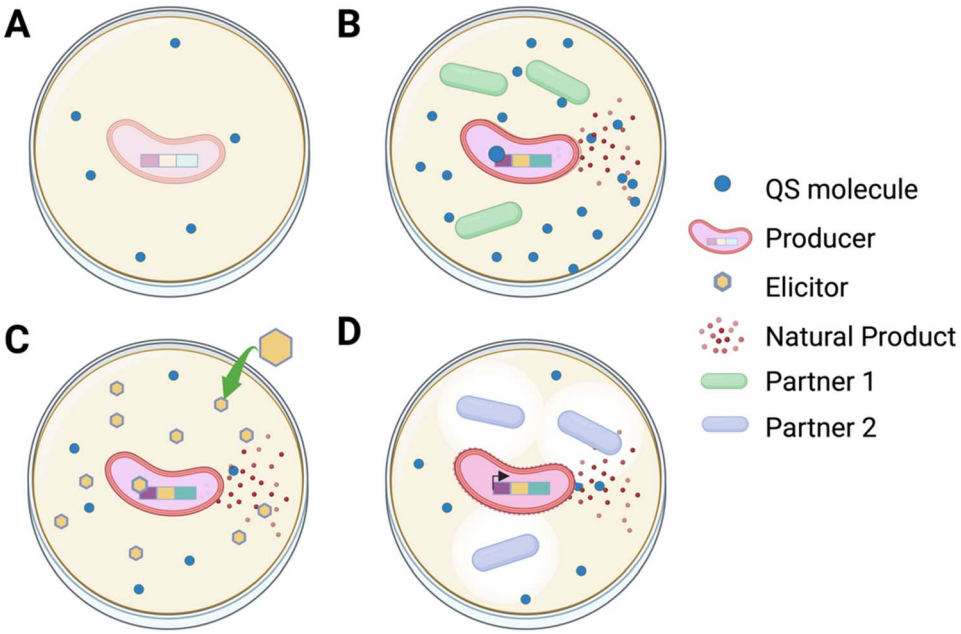


Fig. 2 Co-culturing strategies to stimulate natural product production: different strategies that can be used to trigger the production of natural products from microbial cultures: (A) monoculture: when grown alone in lab condition, this bacterium does not express the BGC responsible for the biosynthesis of a natural product of interest. This BGC is usually expressed as response to nutrient starvation, interspecies interaction, or it is induced when the QS molecule, produced by the bacterium itself, reaches a specific concentration. While able to grow, in this condition the bacterial population does not reach the necessary density that would lead to the required concentration of the QS molecule. (B) Co-culture with mutualistic partner: when growing with another partner that establishes a positive interaction, the bacterium can grow better and/or produce an higher amount of the QS molecule, which acts as an autoinducer for the BGC of interest. (C) Supplementation of elicitors: it is possible to supplement the culture medium with molecules extracted from another organism (elicitors). Such molecules can mimic the interspecies triggering the expression of the BGC and the production of the natural product. (D) Co-culture with competitor: when growing with a competitive partner able to more effectively use the nutrients available in the medium, the bacterium enters nutrient starvation which in turn triggers the production of the natural product. Created in BioRender. Del Carratore, F. (2025) https://BioRender.com/acnnm44.

capture the complexity governing even the simplest microbial community.[49] Using a systems biology-based approach, we can understand how individual secondary metabolite-mediated interactions have a broader effect on the entire community. Understanding such emergent properties of microbial networks requires multi-omics approaches coupled with advanced mathematical modelling of such networks. By integrating genomics, transcriptomics, proteomics, and metabolomics, multi-omics approaches facilitate the prediction of the metabolic potential of community members[50,51] (Fig. 3), the identification of secondary metabolites produced,[52,53] and assessing the transcriptional and metabolic responses of neighbouring organisms to such compounds.[54] The interpretation and contextualization of the huge number of biological insights generated by multi-omics data, requires complex computational biology and mathematical modelling methods. Genome-scale metabolic models for example, can be used for this purpose.[55,56] These models are built from the functional annotation of individual genomes or metagenome-assembled genomes (MAGs), and, at their best, can describe the complete metabolic potential of single organisms or microbial communities.[57–59] These models can be used to integrate and contextualise omics data[60] as well as estimate feasible metabolic flux distributions that support observed phenotypes, such as growth, and in turn predict nutrient uptake, byproduct secretion, and potential cross-feeding interactions between community members.[61,62] Recent advances in this field open the way for their use to describe how abiotic factors (*e.g.*, nutrient availability) or biotic perturbations (*e.g.*, addition or removal of a members) can alter microbial networks.[62,63] Ecological models, such as Lotka–Volterra models, can be used to simulate and predict how the composition of microbial communities' changes over time.[64–66] More advanced dynamic simulations, combining GSMMs and methods such as dynamic Flux Balance Analysis (dFBA), allow the description and prediction of the temporal changes in community members' abundance and the concentration of extracellular metabolites, allowing the understanding of the effect of initial conditions, growth rates and interspecies interactions.[67] Together with multi-omics approaches, mathematical modelling is a necessary tool for the understanding of the dynamics governing microbial communities, hence, enabling the rational design of synthetic microbial consortia possessing the desired function, such as the efficient production of natural products.

4.2 Engineering communities for enhanced NP production

Once a promising design for a novel functional microbial community has been identified through multi-omics and computational modelling, its implementation usually requires

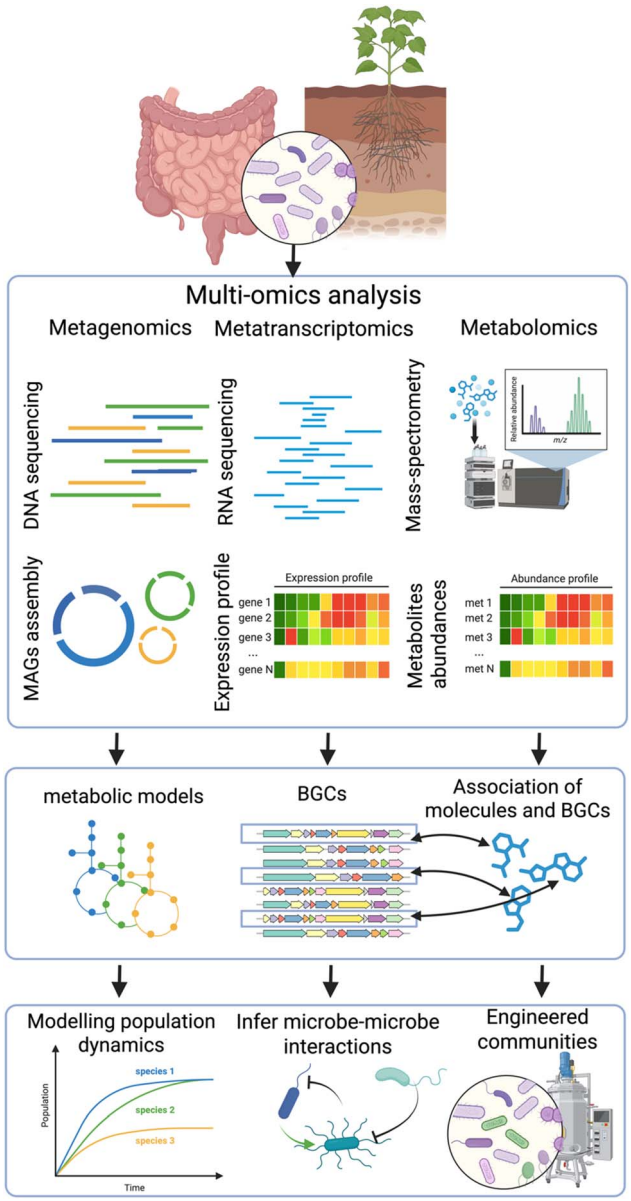


Fig. 3 Leveraging multi-omics for functional insights and predictive modelling: combining genomic, transcriptomic, and metabolomic data allows to connect microbial genetic potential to functional output. This is key for discovering novel natural products by associating metabolites with the BGCs responsible for their synthesis. The main strength of this approach is represented by the possibility of sampling directly from the metabolome natural environment. Metagenomics data allows the reconstruction of genome-scale models of the sequenced organisms, *in silico* identification of BGCs and modelling population dynamics (*e.g.* Lotka–Volterra model). Metatranscriptomics data can be used to identify the actively transcribed BGCs, improve the metabolic models, and detect changes in gene expression in response to different stimuli. Metabolomics can also be used to improve the metabolic models and measure the produced natural compounds, which can be putatively assigned to the detected BGCs. Metabolic models provide insights on the possible metabolic interactions between the different microorganisms (*e.g.* cross-feeding). The modelling of the population dynamics allows the identification of the interactions governing the microbiome (*e.g.*, parameters of the Lotka–Volterra model) which can be either mediated *via* metabolic exchanges or secondary metabolites. Created in BioRender. Del Carratore, F. (2025) https://BioRender.com/33ifd7c.

extensive microbial genetic engineering. Significant technological advances are providing the scientific community with unprecedented abilities to manipulate microbial genomes. Combined with the understanding of complex community dynamics, these technologies open the way to harness the vast biosynthetic potential of the microbial world. For instance, the development of CRISPR-Cas technology provided us with versatile, precise and efficient tools for the genetic engineering of microbial genomes.[68,69] This allows the manipulation of gene expression and the introduction of heterologous pathways, paving the way for the rewiring of metabolic networks in a greatly expanded range of microbial species, including the ones relevant in natural microbiomes.[70] It is now possible to build signal-responsive synthetic gene expression circuits able to sense specific molecules or changes in the environment and trigger the activation of different metabolic pathways, including the ones responsible for the production of natural products.[71,72] For example, a rhamnose-inducible regulatory system has been developed for the dynamic control of different metabolic pathways in several *Streptomyces* species.[73] CRISPR-Cas technology also provided us with the ability to simultaneously edit multiple genes at the same time with high efficiency. Such multiplex genome editing techniques are crucial for the engineering of complex metabolic pathways and for manipulating multiple members of the same community. For example, using the recently developed CRISPR-BEST system,[74] it has been possible to simultaneously edit 17 different target sites in *Streptomyces coelicolor*.[75] Thanks to these developments, the engineering of small microbial communities is now a realistic prospect.

Synthetic biology tools are routinely used to enable heterologous expression of BGCs in single strains;[76] however, the metabolic burden placed on the host cell by the introduction of a complex, multistep pathway can impede the accumulation of productive biomass and limit pathway efficiency.[77–79] A strategy based on the division of the labour (DOL) associated with such complex pathways across the different members of a microbial community, can be an effective solution for addressing this issue.[80,81] Within a small microbial community artificially put together in a Petri dish or in a liquid medium, it is possible to engineer some members for the production of specific nutrients or precursor compounds, that in turn enhance the production of specific secondary metabolites for another member of the community.[82] A similar mechanism was implemented in a synthetic three-member community consisting of *Corynebacterium glutamicum*, *Yarrowia lipolytica* and *Bacillus subtilis*.[83] In this study, *C. glutamicum* was engineered to produce high levels of precursor compounds (Threonine, Proline, Valine and Isoleucine), while the recombinant *Y. lipolytica* YL-21 was used to produce a higher level of C16 fatty acid. This resulted in an enhanced production of fengycin (an antifungal particularly effective against *Rhizoctonia solani*) by *B. subtilis* CGF26-IV.[83]

In a similar attempt, Mehta *et al.* designed a synthetic community consisting of 2 engineered *Escherichia coli* subpopulations each encoding a part of the biosynthetic pathway of violacein. This strategy led to a 2.5-fold increase in production when compared to the monoculture alternative.[84] Using

a different approach, Rafieenia and colleagues designed a synthetic community composed of mutant strains of *Yarrowia lipolytica* to upcycle fermentation byproducts, specifically citric acid (CA), to increase production yields.[85] A glucose-non-consuming "upcycler" strain was created through metabolic engineering and adaptive laboratory evolution to efficiently consume CA. When co-cultivated with a wild-type strain that produces CA, this community showed higher cell densities and improved β-carotene production compared to the monoculture. The presence of the "upcycler" strain improved growth by 2.35-fold and β-carotene production by 2-fold.[85] The successful implementation of the DOL strategies relies on the ability to obtain stable synergistic communities. Focusing on *Saccharomyces cerevisiae*, Aulakh *et al.* identified 49 pairs of auxotrophic mutants that spontaneously form syntrophic communities through high-throughput phenotypic screening. These communities were found to grow synergistically, with some pairs exchanging metabolic intermediates rather than just end products.[86] The authors demonstrated the biotechnological potential of these communities by splitting a malonic semi-aldehyde (MSA) biosynthesis pathway between different auxotrophs, which resulted in increased production of the compound. A specific pair (Δtrp2–Δtrp4) demonstrated an up to a six-fold increase in MSA production per unit biomass compared to a monoculture.[86] Similarly, Park *et al.* demonstrated that a combination of Δtrp2 and Δtrp4 mutants of *Y. lipolytica* and *S. cerevisiae* forms stable interspecies and intraspecies communities.[87] By splitting the biosynthetic pathway of 3-hydroxypropionic acid (3-HP) between the strains, they applied a DOL approach leading to a 19.3-fold increase in 3-HP production in the best-performing intraspecies community compared to the monoculture, and an 18.6-fold increase in the best-performing interspecies community.[87]

Peng *et al.* used a design-led approach by engineering 15 auxotrophic strains and overproduction strains for various essential metabolites.[88] Using these strains as modular "donors" and "receivers", they designed novel two- and three-member communities and explored how factors like initial population ratio and metabolite exchange rate influence community dynamics. The obtained toolkit was successfully used to employ a DOL strategy for producing resveratrol. By splitting the resveratrol biosynthetic pathway among the community members, the authors achieved a 3.16-fold higher production than the monoculture control.[88] While extremely promising, the major limitation of such a DOL strategy is represented by the fact that the maintenance of different engineered subpopulations is challenging due to competition and convoluted inter-strain population dynamics. In a recent study, Hamrick *et al.* (2024) showed through mathematical modelling, how dynamic division of labour (DDOL) can overcome these limitations.[10] DDOL is defined as a DOL where the different metabolic steps are not persistently associated with one single strain but are dynamically interchangeable and reversible. This can be achieved by encoding the different steps on plasmids capable of horizontal gene transfer. Through mathematical modelling the authors showed that, compared with a static DOL approach, DDOL enables persistence of multi-step pathways within a microbial community, as well as being able to generate microbial communities more resilient to asymmetric burden.[10] Alternatively, obligate commensalism can be designed within synthetic communities, which can allow the persistence of the community members as well as providing a tool for the containment of genetically engineered organisms. Forti *et al.* designed a 2-member synthetic community. The first member consisted of an engineered *E. coli* strain able to produce non-standard amino acids from simple carbon sources (producer), while the second consisted of another engineered *E. coli* depending on such non-standard amino acids for growth as a synthetic auxotroph (utilizer). The authors showed that the dependence of the utilizer on the producer persisted for up to 14 days *in vitro* and for 24 h within a simplified synthetic maize root-associated community.[89]

4.3 Direct mining of microbiomes for compound discovery

Recent technological advancements in the omics sciences have opened new routes of access to the vast natural products found within microbiomes, eliminating the need to isolate and cultivate individual strains in a laboratory setting. For instance, metagenomics allows for the sequencing of the total DNA from virtually any microbial community.[90,91] The data generated from these metagenomic studies can be used for the reconstruction of metagenome-assembled genomes (MAGs), providing the genomic context associated with any specific microbiome,[57,92] including the genetic code of the unculturable members, which often represent the dominant fraction of natural microbial communities. Novel computational pipelines for the automated analysis of such data are constantly being developed, greatly improving our ability to interpret MAGs.[57,93] The metaGEM pipeline is also able to automatically reconstruct genome-scale metabolic models from MAGs, shedding light on the metabolic networks governing microbial communities.[57] The assembled MAGs can also be easily mined for the *in silico* identification of BGCs with well-established computational tools such as antiSMASH and BAGEL,[38,94] the only limitation being the contig length, as successful BGC mining typically depends on all members of the cluster being located on a single contiguous sequence. Once high-quality assemblies are available, these tools allow for the identification and classification of BGCs directly from MAGs, as well as predicting the types of secondary metabolites likely synthesised by these clusters. New very promising computational tools exploiting the most recent advances in artificial intelligence and deep learning are being developed, both for BGCs and end-product predictions.[95–97] Additionally, it is possible to compare them with known and well-characterised BGCs, for example by comparing them to the ones present in the MIBIG database.[39] Moreover, it is possible to use advanced statistical methods for the computational identification of co-evolving multi-gene modules within BGCs found in the same microbial communities.[98] This approach can help with the identification of subunits responsible for the same function across different BGCs, enhance their annotation and provide insights on how these entities evolve within microbial communities. Metatranscriptomics is another powerful tool

that allows measuring the gene expression level of potentially all the genes present in microbial communities.[99,100] Importantly, it is possible to sample directly from the natural ecological niche where different microbiomes are found, hence allowing observations on which BGCs are actively transcribed in real life scenarios.[99] Therefore, such an approach provides insights on the potential function and regulation of all the identified BGCs. Metabolomics also self-evidently plays a very important role in the discovery of natural products directly from microbiomes. By comprehensively measuring all the small molecules (metabolites) in a sample, untargeted metabolomics represents a necessary tool for the identification of the large diversity of natural products produced by different microbiomes.[101] Mass spectrometry-based techniques are most commonly used for untargeted metabolomics experiments, and in some cases allow the detection of metabolites that can be directly linked to the BGCs observed in the same sample.[101,102] The association of the ions measured *via* mass spectrometry with the molecules that generated them (mass spectrum annotation) continues to be the major bottleneck that greatly limits the application of metabolomics to natural products discovery.[103] A number of new and exciting tools for the annotation of mass spectrometry data are constantly being developed,[104–107] helping with addressing this challenge. However, most of these tools rely on matching with the known compounds recorded in different databases. However, these tools do not help with the identification of novel compounds, which are not present in any database. There is a big effort put towards the development of computational tools that can predict chemical formulas and even structures for these novel compounds, such as SIRIUS[108] and GNPS.[109] Tay *et al.* (2023) demonstrated how it is possible to use a recurrent neural network trained on known natural products to generate over 67 million natural products-like molecular structures to complement existing records. This approach could help tackle the issue of incomplete natural products databases.[110] When we collect samples directly from microbiomes, we often end up with very complex mixtures, including all sorts of primary metabolites and natural products, which further complicates the use of untargeted metabolomics for natural products discovery. An approach that partially addresses these challenges is represented by the Small Molecule *In situ* Resin Capture (SMIRC) method. SMIRC is a new, culture-independent method that directly obtains natural products from their native environments by capturing compounds using adsorbent resins. This approach bypasses the need for laboratory cultivation and traditional microbe-first strategies, which often fail to access the full biosynthetic complement of microbiomes. SMIRC has successfully captured numerous novel compounds, including those with unprecedented structural features, demonstrating its potential for exploring previously unexplored regions of chemical space for natural product discovery.[111] More generally, recent advances in artificial intelligence and deep learning are boosting the development of novel computational methods that could help every aspect of novel natural products discovery.[112] These omics technologies, often used in combination (multi-omics), have significantly advanced our ability to explore the vast reservoir of BGCs encoded in diverse microbiomes, paving the way for the discovery of novel natural products.[113] In particular, multi-omics has confirmed the widely held assumption that a vast number of secondary metabolite biosynthetic gene clusters (BGCs) can be found in various microbiomes, such as the human gut,[14] freshwater sponges,[13] and mangrove swamps.[12] A very high proportion of the BGCs identified from metagenomics studies of natural microbiomes are uncharacterised and show little to no similarity to characterised BGCs, which suggests a high degree of chemical novelty among their products.

For example, Zhang *et al.* (2023) employed a multi-omics approach combining metagenomics, metatranscriptomics and untargeted metabolomics to explore the microbial biosynthetic potential of the mangrove swamp microbiome.[12] The analysis of the metagenomics data revealed a total of 3740 BGCs. A significant fraction of these, particularly those predicted to encode for polyketides and nonribosomal peptides, showed very little similarity to known clusters. Such novel clusters were mainly found in less-studied microbial groups like Desulfobacterota and Chloroflexota. Metatranscriptomics data provided evidence that many of these novel BGCs were actively expressed in the mangrove environment, suggesting they play an important role in the community ecology. Additionally, untargeted metabolomics data revealed that 98% of the detected mass spectra did not match to any known compounds, further supporting the novelty of the biosynthetic potential of this community.[12]

In another study, Graffius *et al.* (2023) investigated the potential of the microbial community within the freshwater sponge *Spongilla lacustris* to produce secondary metabolites, using both cultivation-based and cultivation-independent (metagenomic) methods.[13] The study identified a diverse range of bacteria, isolating representatives from 41 genera and recovering 20 metagenome-assembled genomes (MAGs). Analysis revealed a substantial capacity for secondary metabolite production, with a total of 306 biosynthetic gene clusters (BGCs) detected across the isolates' genomes and the MAGs. Genome mining experiments successfully activated the production of specific secondary metabolites, such as coprisidins, in selected *Streptomyces* isolates, confirming the predicted biosynthetic potential.[13]

The human microbiome is another rich and underexplored source of natural products. In fact, King *et al.* (2023) systematically mined 2229 human microbiome genomes, identifying numerous gene clusters for ribosomally synthesized and post-translationally modified peptides (RiPPs), specifically lanthipeptides and lasso peptides.[14] The authors were also able to express some of these in *E. coli* and functionally characterise 23 new BGCs, testing the activity of their end products against both healthy commensal bacteria and pathogens. Through this approach they were able to discover new antibiotics effective against bacteria implicated in skin, nasal, and vaginal dysbiosis, including some active against multidrug-resistant pathogens like MRSA and VRE.[14] This study demonstrated the potential of mining human-associated microbiomes for novel antimicrobial discovery. In a first important step towards

advancing these efforts, Zou *et al.* (2023) built the sBGC-hm database, containing 36 583 BGCs mined from human gut microbiomes, creating a valuable resource to facilitate the exploration of these communities for new bioactive natural products.[11]

5 Conclusions and future perspectives

The study of microbial communities and their secondary metabolites has unveiled a vast potential for natural product discovery and production. The intricate interactions within these communities, mediated by secondary metabolites, play a crucial role in shaping community dynamics and ecological functions. Advances in genetic engineering and multi-omics technologies have paved the way for harnessing these interactions, enabling the engineering of synthetic microbial consortia for enhanced production of natural products and the direct mining of microbiomes for novel compound discovery. Despite the challenges associated with culturing and characterizing diverse microbial species, ongoing advancements in computational tools and data analysis are rapidly expanding our ability to explore and exploit the seemingly inexhaustible reservoir of natural products represented within microbial communities. As the scientific community continues to unravel the complexities of microbial interactions and unlock the biosynthetic potential of unculturable microorganisms, there is a huge potential for the discovery of new natural products with possible applications in medicine, agriculture, and industrial biotechnology. The engineering of microbiomes represents a transformative frontier in biotechnology, shifting from single-strain approaches to harnessing complex microbial communities for natural product discovery and production. We predict that the next decade will see a convergence of systems biology, machine learning, and large-scale genome editing, enabling the precision engineering of microbiomes to unlock their vast reservoirs of bioactive compounds. A key goal the whole community should aim for is the development of rational microbiome design platforms, integrating multi-omics data, metabolic modelling, and ecological principles. Such platforms will enable the bottom-up engineering of consortia optimized for metabolite yield, stability, resilience, and scalability across diverse bioproduction settings. Decoding microbial interactions, both cooperative and competitive, will allow us to choreograph synthetic communities that mimic or surpass natural biosynthetic efficiency in a non-natural bioindustrial setting. Looking further ahead, the integration of closed-loop design–build–test–learn pipelines with lab automation and AI will be critical. Biofoundries will enable the parallel engineering and functional phenotyping of hundreds of newly assembled microbial consortia, drastically reducing the time from gene discovery to product validation. Treating the microbiome as a designable and evolvable unit will unlock a new era of distributed biosynthesis and ecological integration, leading to more robust and versatile bioproduction systems. Ultimately, we expect to see the emergence of a new-generation bioeconomy driven by programmable microbial communities, instead of relying on individual engineered strains.

6 Conflicts of interest

There are no conflicts to declare.

7 Data availability

There is no additional data associated with this article.

8 References

1 J. L. Chodkowski and A. Shade, *mSystems*, 2024, **9**(4), e00064.
2 D. M. Selegato and I. Castro-Gamboa, *Front. Microbiol.*, 2023, **14**, 1117559.
3 A. Y. Alwali and E. I. Parkinson, *J. Ind. Microbiol. Biotechnol.*, 2023, **50**(1), kuad019.
4 A. van der Meij, S. S. Elsayed, C. Du, J. Willemse, T. M. Wood, N. I. Martin, J. M. Raaijmakers and G. P. van Wezel, *Appl. Environ. Microbiol.*, 2023, **89**(11), e01239.
5 D. A. van Bergeijk, S. S. Elsayed, C. Du, I. N. Santiago, A. M. Roseboom, L. Zhang, V. J. Carrión, H. P. Spaink and G. P. van Wezel, *Commun. Chem.*, 2022, **5**, 14.
6 Y. H. Han, G. Kim and S. W. Seo, *Curr. Opin. Biotechnol.*, 2023, **79**, 102874.
7 P. Geesink, J. ter Horst and T. J. G. Ettema, *FEMS Microbiol. Ecol.*, 2024, **100**(4), fiae029.
8 S. D. Burz, S. Causevic, A. Dal Co, M. Dmitrijeva, P. Engel, D. Garrido-Sanz, G. Greub, S. Hapfelmeier, W.-D. Hardt, V. Hatzimanikatis, C. M. Heiman, M. K.-M. Herzog, A. Hockenberry, C. Keel, A. Keppler, S.-J. Lee, J. Luneau, L. Malfertheiner, S. Mitri, B. Ngyuen, O. Oftadeh, A. R. Pacheco, F. Peaudecerf, G. Resch, H.-J. Ruscheweyh, A. Sahin, I. R. Sanders, E. Slack, S. Sunagawa, J. Tackmann, R. Tecon, G. S. Ugolini, J. Vacheron, J. R. van der Meer, E. Vayena, P. Vonaesch and J. A. Vorholt, *Microbiol. Mol. Biol. Rev.*, 2023, **87**(4), e00063.
9 S. M. Brooks, C. Marsan, K. B. Reed, S.-F. Yuan, D.-D. Nguyen, A. Trivedi, G. Altin-Yavuzarslan, N. Ballinger, A. Nelson and H. S. Alper, *Nat. Commun.*, 2023, **14**, 4448.
10 G. S. Hamrick, R. Maddamsetti, H.-I. Son, M. L. Wilson, H. M. Davis and L. You, *ACS Synth. Biol.*, 2024, **13**, 1142–1151.
11 H. Zou, T. Sun, B. Jin and S. Wang, *Bioinformatics*, 2023, **39**(3), btad131.
12 J.-W. Zhang, R. Wang, X. Liang, P. Han, Y.-L. Zheng, X.-F. Li, D.-Z. Gao, M. Liu, L.-J. Hou and H.-P. Dong, *Appl. Environ. Microbiol.*, 2023, **89**(6), e00102.
13 S. Graffius, J. F. G. Garzón, M. Zehl, P. Pjevac, R. Kirkegaard, M. Flieder, A. Loy, T. Rattei, A. Ostrovsky and S. B. Zotchev, *Microbiol. Spectrum*, 2023, **11**(2), e04353.
14 A. M. King, Z. Zhang, E. Glassey, P. Siuti, J. Clardy and C. A. Voigt, *Nat. Microbiol.*, 2023, **8**, 2420–2434.

15 S. Wu, H. Zhou, D. Chen, Y. Lu, Y. Li and J. Qiao, *Briefings Bioinf.*, 2024, **25**(4), bbae264.
16 R. P. Jacoby, A. Koprivova and S. Kopriva, *J. Exp. Bot.*, 2021, **72**, 57–69.
17 M. G. Chevrette, C. S. Thomas, A. Hurley, N. Rosario-Meléndez, K. Sankaran, Y. Tu, A. Hall, S. Magesh and J. Handelsman, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**(42), e2212930119.
18 X. Dong, T. Zhang, W. Wu, Y. Peng, X. Liu, Y. Han, X. Chen, Z. Gao, J. Xia, Z. Shao and C. Greening, *Sci. Adv.*, 2024, **10**(17), eadl2281.
19 J. Michailidu, O. Maťátková, A. Čejková and J. Masák, *Molecules*, 2025, **30**, 431.
20 J. L. Chodkowski and A. Shade, *mSystems*, 2020, **5**(6), e00493.
21 K. Nieselt, F. Battke, A. Herbig, P. Bruheim, A. Wentzel, Ø. M. Jakobsen, H. Sletta, M. T. Alam, M. E. Merlo, J. Moore, W. A. Omara, E. R. Morrissey, M. A. Juarez-Hermosillo, A. Rodríguez-García, M. Nentwich, L. Thomas, M. Iqbal, R. Legaie, W. H. Gaze, G. L. Challis, R. C. Jansen, L. Dijkhuizen, D. A. Rand, D. L. Wild, M. Bonin, J. Reuther, W. Wohlleben, M. C. Smith, N. J. Burroughs, J. F. Martín, D. A. Hodgson, E. Takano, R. Breitling, T. E. Ellingsen and E. M. Wellington, *BMC Genomics*, 2010, **11**, 10.
22 J. Kramer, Ö. Özkaya and R. Kümmerli, *Nat. Rev. Microbiol.*, 2020, **18**, 152–163.
23 M. Ellermann and J. C. Arthur, *Free Radical Biol. Med.*, 2017, **105**, 68–78.
24 I. J. Schalk, *Nat. Rev. Microbiol.*, 2025, **23**, 24–40.
25 S. Heilbronner, B. Krismer, H. Brötz-Oesterhelt and A. Peschel, *Nat. Rev. Microbiol.*, 2021, **19**, 726–739.
26 J. D. Palmer and K. R. Foster, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**(38), e2205407119.
27 T. L. Wood, T. Gong, L. Zhu, J. Miller, D. S. Miller, B. Yin and T. K. Wood, *npj Biofilms Microbiomes*, 2018, **4**, 22.
28 S. Cerqueira dos Santos, C. Araújo Torquato, D. de Alexandria Santos, A. Orsato, K. Leite, J. M. Serpeloni, R. Losi-Guembarovski, E. Romão Pereira, A. L. Dyna, M. G. Lopes Barboza, M. H. Fernandes Arakawa, J. A. Pires Bitencourt, S. da Cruz Silva, G. C. da Silva Sá, P. Dias Rodrigues, C. M. Quintella and L. C. Faccin-Galhardi, *Sci. Rep.*, 2024, **14**, 4629.
29 B. K. Okada and M. R. Seyedsayamdost, *FEMS Microbiol. Rev.*, 2017, **41**, 19–33.
30 Y. Cai and X. Zhang, *J. Bacteriol.*, 2024, **206**(2), e00430.
31 A. C. Armes, J. L. Walton and A. Buchan, *Microbiol. Spectrum*, 2022, **10**(6), e02615.
32 Y. Su and T. Ding, *Gut Microbes*, 2023, **15**(2), 2252780.
33 A. Hartmann, T. Binder and M. Rothballer, *FEMS Microbiol. Ecol.*, 2024, **100**(6), fiae076.
34 R. Masteling, W. de Boer, J. M. Raaijmakers, P. Garbeva and F. Dini-Andreote, *Front. Ecol. Evol.*, 2022, **10**, 960198.
35 L. Zaroubi, I. Ozugergin, K. Mastronardi, A. Imfeld, C. Law, Y. Gélinas, A. Piekny and B. L. Findlay, *Appl. Environ. Microbiol.*, 2022, **88**(7), e00093.
36 N. Melo, G. H. Wolff, A. L. Costa-da-Silva, R. Arribas, M. F. Triana, M. Gugger, J. A. Riffell, M. DeGennaro and M. C. Stensmyr, *Curr. Biol.*, 2020, **30**, 127–134.
37 P. Garbeva, M. Avalos, D. Ulanova, G. P. van Wezel and J. S. Dickschat, *Environ. Microbiol.*, 2023, **25**, 1565–1574.
38 K. Blin, S. Shaw, H. E. Augustijn, Z. L. Reitz, F. Biermann, M. Alanjary, A. Fetter, B. R. Terlouw, W. W. Metcalf, E. J. N. Helfrich, G. P. van Wezel, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2023, **51**, W46–W50.
39 M. M. Zdouc, K. Blin, N. L. L. Louwen, J. Navarro, C. Loureiro, C. D. Bader, C. B. Bailey, L. Barra, T. J. Booth, K. A. J. Bozhüyük, J. D. D. Cediel-Becerra, Z. Charlop-Powers, M. G. Chevrette, Y. H. Chooi, P. M. D'Agostino, T. de Rond, E. Del Pup, K. R. Duncan, W. Gu, N. Hanif, E. J. N. Helfrich, M. Jenner, Y. Katsuyama, A. Korenskaia, D. Krug, V. Libis, G. A. Lund, S. Mantri, K. D. Morgan, C. Owen, C.-S. Phan, B. Philmus, Z. L. Reitz, S. L. Robinson, K. S. Singh, R. Teufel, Y. Tong, F. Tugizimana, D. Ulanova, J. M. Winter, C. Aguilar, D. Y. Akiyama, S. A. A. Al-Salihi, M. Alanjary, F. Alberti, G. Aleti, S. A. Alharthi, M. Y. A. Rojo, A. A. Arishi, H. E. Augustijn, N. E. Avalon, J. A. Avelar-Rivas, K. K. Axt, H. B. Barbieri, J. C. J. Barbosa, L. G. Barboza Segato, S. E. Barrett, M. Baunach, C. Beemelmanns, D. Beqaj, T. Berger, J. Bernaldo-Agüero, S. M. Bettenbühl, V. A. Bielinski, F. Biermann, R. M. Borges, R. Borriss, M. Breitenbach, K. M. Bretscher, M. W. Brigham, L. Buedenbender, B. W. Bulcock, C. Cano-Prieto, J. Capela, V. J. Carrion, R. S. Carter, R. Castelo-Branco, G. Castro-Falcón, F. O. Chagas, E. Charria-Girón, A. A. Chaudhri, V. Chaudhry, H. Choi, Y. Choi, R. Choupannejad, J. Chromy, M. S. C. Donahey, J. Collemare, J. A. Connolly, K. E. Creamer, M. Crüsemann, A. A. Cruz, A. Cumsille, J.-F. Dallery, L. C. Damas-Ramos, T. Damiani, M. de Kruijff, B. D. Martín, G. Della Sala, J. Dillen, D. T. Doering, S. R. Dommaraju, S. Durusu, S. Egbert, M. Ellerhorst, B. Faussurier, A. Fetter, M. Feuermann, D. P. Fewer, J. Foldi, A. Frediansyah, E. A. Garza, A. Gavriilidou, A. Gentile, J. Gerke, H. Gerstmans, J. P. Gomez-Escribano, L. A. González-Salazar, N. E. Grayson, C. Greco, J. E. G. Gomez, S. Guerra, S. G. Flores, A. Gurevich, K. Gutiérrez-García, L. Hart, K. Haslinger, B. He, T. Hebra, J. L. Hemmann, H. Hindra, L. Höing, D. C. Holland, J. E. Holme, T. Horch, P. Hrab, J. Hu, T.-H. Huynh, J.-Y. Hwang, R. Iacovelli, D. Iftime, M. Iorio, S. Jayachandran, E. Jeong, J. Jing, J. J. Jung, Y. Kakumu, E. Kalkreuter, K. Bin Kang, S. Kang, W. Kim, G. J. Kim, H. Kim, H. U. Kim, M. Klapper, R. A. Koetsier, C. Kollten, Á. T. Kovács, Y. Kriukova, N. Kubach, A. M. Kunjapur, A. K. Kushnareva, A. Kust, J. Lamber, M. Larralde, N. J. Larsen, A. P. Launay, N.-T.-H. Le, S. Lebeer, B. T. Lee, K. Lee, K. L. Lev, S.-M. Li, Y.-X. Li, C. Licona-Cassani, A. Lien, J. Liu, J. A. V. Lopez, N. V. Machushynets, M. I. Macias, T. Mahmud, M. Maleckis, A. M. Martinez-Martinez, Y. Mast, M. F. Maximo, C. M. McBride, R. M. McLellan, K. M. Bhatt, C. Melkonian, A. Merrild,

M. Metsä-Ketelä, D. A. Mitchell, A. V. Müller, G.-S. Nguyen, H. T. Nguyen, T. H. J. Niedermeyer, J. H. O'Hare, A. Ossowicki, B. O. Ostash, H. Otani, L. Padva, S. Paliyal, X. Pan, M. Panghal, D. S. Parade, J. Park, J. Parra, M. P. Rubio, H. T. Pham, S. J. Pidot, J. Piel, B. Pourmohsenin, M. Rakhmanov, S. Ramesh, M. H. Rasmussen, A. Rego, R. Reher, A. J. Rice, A. Rigolet, A. Romero-Otero, L. R. Rosas-Becerra, P. Y. Rosiles, A. Rutz, B. Ryu, L.-A. Sahadeo, M. Saldanha, L. Salvi, E. Sánchez-Carvajal, C. Santos-Medellin, N. Sbaraini, S. M. Schoellhorn, C. Schumm, L. Sehnal, N. Selem, A. D. Shah, T. K. Shishido, S. Sieber, V. Silviani, G. Singh, H. Singh, N. Sokolova, E. C. Sonnenschein, M. Sosio, S. T. Sowa, K. Steffen, E. Stegmann, A. B. Streiff, A. Strüder, F. Surup, T. Svenningsen, D. Sweeney, J. Szenei, A. Tagirdzhanov, B. Tan, M. J. Tarnowski, B. R. Terlouw, T. Rey, N. U. Thome, L. R. Torres Ortega, T. Tørring, M. Trindade, A. W. Truman, M. Tvilum, D. W. Udwary, C. Ulbricht, L. Vader, G. P. van Wezel, M. Walmsley, R. Warnasinghe, H. G. Weddeling, A. N. M. Weir, K. Williams, S. E. Williams, T. E. Witte, S. M. W. Rocca, K. Yamada, D. Yang, D. Yang, J. Yu, Z. Zhou, N. Ziemert, L. Zimmer, A. Zimmermann, C. Zimmermann, J. J. J. van der Hooft, R. G. Linington, T. Weber and M. H. Medema, *Nucleic Acids Res.*, 2025, **53**, D678–D690.

40 K. Blin, S. Shaw, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2024, **52**, D586–D589.

41 L. A. Pazos-Rojas, A. Cuellar-Sánchez, A. L. Romero-Cerón, A. Rivera-Urbalejo, P. Van Dillewijn, D. A. Luna-Vital, J. Muñoz-Rojas, Y. E. Morales-García and M. d. R. Bustillos-Cristales, *Microorganisms*, 2023, **12**, 39.

42 B. C. Covington, F. Xu and M. R. Seyedsayamdost, *Annu. Rev. Biochem.*, 2021, **90**, 763–788.

43 J. H. Kim, N. Lee, S. Hwang, W. Kim, Y. Lee, S. Cho, B. O. Palsson and B.-K. Cho, *J. Ind. Microbiol. Biotechnol.*, 2021, **48**(3–4), kuaa001.

44 Z. Song, Z. Ma, A. Bechthold and X. Yu, *Appl. Microbiol. Biotechnol.*, 2020, **104**, 4445–4455.

45 H. Shakeri Moghaddam, B. Shahnavaz, A. Makhdoumi and M. Iranshahy, *Biocontrol Sci. Technol.*, 2021, **31**, 1248–1266.

46 T. Boruta, A. Ścigaczewska and M. Bizukojć, *Front. Bioeng. Biotechnol.*, 2021, **9**, 713639.

47 L.-L. Man, X.-C. Meng and R.-H. Zhao, *Food Control*, 2012, **23**, 462–469.

48 H. Liu, M. Zhang, L. Xu, F. Xue, W. Chen and C. Wang, *Heliyon*, 2024, **10**, e31619.

49 N. I. van den Berg, D. Machado, S. Santos, I. Rocha, J. Chacón, W. Harcombe, S. Mitri and K. R. Patil, *Nat. Ecol. Evol.*, 2022, **6**, 855–865.

50 E. A. Eloe-Fadrosh, F. Ahmed, Anubhav, M. Babinski, J. Baumes, M. Borkum, L. Bramer, S. Canon, D. S. Christianson, Y. E. Corilo, K. W. Davenport, B. Davis, M. Drake, W. D. Duncan, M. C. Flynn, D. Hays, B. Hu, M. Huntemann, J. Kelliher, S. Lebedeva, P. E. Li, M. Lipton, C. C. Lo, S. Martin, D. Millard, K. Miller, M. A. Miller, P. Piehowski, E. P. Jackson, S. Purvine, T. B. K. Reddy, R. Richardson, M. Rudolph, S. Sarrafan, M. Shakya, M. Smith, K. Stratton, J. C. Sundaramurthi, P. Vangay, D. Winston, E. M. Wood-Charlson, Y. Xu, P. S. G. Chain, L. A. McCue, D. Mans, C. J. Mungall, N. J. Mouncey and K. Fagnan, *Nucleic Acids Res.*, 2022, **50**(D1), D828–D836.

51 Y. Zhang, J. P. Thomas, T. Korcsmaros and L. Gul, *Cell Rep. Med.*, 2024, **5**, 101738.

52 S. Wang, X. Li, W. Yang and R. Huang, *Microb. Biotechnol.*, 2024, **17**(8), e14533.

53 T. T. H. Dat, G. Steinert, N. T. K. Cuc, P. V. Cuong, H. Smidt and D. Sipkema, *Mar. Drugs*, 2022, **21**, 29.

54 Q. Li, W. Lin, X. Zhang, M. Wang, Y. Zheng, X. Wang, G. Gao, Y. Li, D. Zhao and C. Zhang, *Microbiol. Res.*, 2024, **280**, 127598.

55 P. Sen and M. Orešič, *Metabolites*, 2023, **13**, 855.

56 G. Zampieri, S. Campanaro, C. Angione and L. Treu, *Cells Rep. Methods*, 2023, **3**, 100383.

57 F. Zorrilla, F. Buric, K. R. Patil and A. Zelezniak, *Nucleic Acids Res.*, 2021, **49**, e126.

58 A. Heinken, J. Hertel, G. Acharya, D. A. Ravcheev, M. Nyga, O. E. Okpala, M. Hogan, S. Magnúsdóttir, F. Martinelli, B. Nap, G. Preciat, J. N. Edirisinghe, C. S. Henry, R. M. T. Fleming and I. Thiele, *Nat. Biotechnol.*, 2023, **41**, 1320–1331.

59 A. Mardinoglu and B. Ø. Palsson, *Nat. Rev. Genet.*, 2025, **26**, 123–140.

60 A. Passi, J. D. Tibocha-Bonilla, M. Kumar, D. Tec-Campos, K. Zengler and C. Zuniga, *Metabolites*, 2021, **12**, 14.

61 C. Gu, G. B. Kim, W. J. Kim, H. U. Kim and S. Y. Lee, *Genome Biol.*, 2019, **20**, 121.

62 N. Giordano, M. Gaudin, C. Trottier, E. Delage, C. Nef, C. Bowler and S. Chaffron, *Nat. Commun.*, 2024, **15**, 2721.

63 A. R. Pacheco and D. Segrè, *J. R. Soc. Interface*, 2021, **18**, 20210348.

64 M. Kumar, B. Ji, K. Zengler and J. Nielsen, *Nat. Microbiol.*, 2019, **4**, 1253–1267.

65 J. E. McDonald, J. R. Marchesi and B. Koskella, *Proc. R. Soc. B*, 2020, **287**, 20202886.

66 E. Lange, L. Kranert, J. Krüger, D. Benndorf and R. Heyer, *Front. Microbiol.*, 2024, **15**, 1368377.

67 I. Dukovski, D. Bajić, J. M. Chacón, M. Quintin, J. C. C. Vila, S. Sulheim, A. R. Pacheco, D. B. Bernstein, W. J. Riehl, K. S. Korolev, A. Sanchez, W. R. Harcombe and D. Segrè, *Nat. Protoc.*, 2021, **16**, 5030–5082.

68 N. Ali, C. Vora, A. Mathuria, N. Kataria and I. Mani, *Prog. Mol. Biol. Transl. Sci.*, 2024, **208**, 59–81.

69 Y. Lee, S. Hwang, W. Kim, J. H. Kim, B. O. Palsson and B.-K. Cho, *J. Ind. Microbiol. Biotechnol.*, 2024, **51**, kuae009.

70 M. Abavisani, N. Faraji, S. Faraji, N. Ebadpour, P. Kesharwani and A. Sahebkar, *Biochem. Eng. J.*, 2024, **211**, 109443.

71 M. M. Müller, K. M. Arndt and S. A. Hoffmann, *Front. Synth. Biol.*, 2025, **3**, 1548572.

72 J. Cornwall-Scoones, D. Benzinger, T. Yu, A. Pezzotta, A. Sagner, L. Gerontogianni, S. Bernadet, E. Finnie, G. L. M. Boezio, H. T. Stuart, M. Melchionda,

O. C. K. Inge, B. Dumitrascu, J. Briscoe and M. J. Delás, *bioRxiv*, 2025, DOI: 10.1101/2025.03.07.642002.
73 Q. Yang, M. Luan, M. Wang, Y. Zhang, G. Liu and G. Niu, *ACS Synth. Biol.*, 2024, **13**, 3461–3470.
74 Y. Tong, C. M. Whitford, H. L. Robertsen, K. Blin, T. S. Jørgensen, A. K. Klitgaard, T. Gren, X. Jiang, T. Weber and S. Y. Lee, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 20366–20375.
75 C. M. Whitford, T. Gren, E. Palazzotto, S. Y. Lee, Y. Tong and T. Weber, *ACS Synth. Biol.*, 2023, **12**, 2353–2366.
76 K. Seshadri, A. N. D. Abad, K. K. Nagasawa, K. M. Yost, C. W. Johnson, M. J. Dror and Y. Tang, *Chem. Rev.*, 2025, **125**, 3814–3931.
77 S. Snoeck, C. Guidi and M. De Mey, *Microb. Cell Fact.*, 2024, **23**, 96.
78 G. H. Rajacharya, A. Sharma and S. S. Yazdani, *Sci. Rep.*, 2024, **14**, 12271.
79 L. L. B. Kastberg, I. H. Jacobsen, E. Özdemir, C. T. Workman, M. K. Jensen and J. Förster, *FEMS Yeast Res.*, 2025, **25**, foaf007.
80 M. Wang, X. Chen, Y. Fang, X. Zheng, T. Huang, Y. Nie and X.-L. Wu, *Cell Syst.*, 2024, **15**, 63–74.
81 N. Roothans, M. C. M. van Loosdrecht and M. Laureni, *ISME J.*, 2025, **19**(1), wraf020.
82 A. Matuszyńska, O. Ebenhöh, M. D. Zurbriggen, D. C. Ducat and I. M. Axmann, *Synth. Biol.*, 2024, **9**(1), ysae011.
83 S.-Y. Wei, G.-R. Gao, M.-Z. Ding, C.-Y. Cao, Z.-J. Hou, J.-S. Cheng and Y.-J. Yuan, *J. Nat. Prod.*, 2024, **87**, 28–37.
84 H. Mehta, J. Jimenez, R. Ledesma-Amaro and G. B. Stan, *ACS Synth. Biol.*, 2025, **14**(7), 2703–2709.
85 R. Rafieenia, C. Klemm, P. Hapeta, J. Fu, M. G. García and R. Ledesma-Amaro, *Trends Biotechnol.*, 2025, **43**, 601–619.
86 S. K. Aulakh, L. Sellés Vidal, E. J. South, H. Peng, S. J. Varma, L. Herrera-Dominguez, M. Ralser and R. Ledesma-Amaro, *Nat. Chem. Biol.*, 2023, **19**, 951–961.
87 Y.-K. Park, H. Peng, P. Hapeta, L. Sellés Vidal and R. Ledesma-Amaro, *Nat. Commun.*, 2024, **15**, 8924.
88 H. Peng, A. P. S. Darlington, E. J. South, H.-H. Chen, W. Jiang and R. Ledesma-Amaro, *Nat. Microbiol.*, 2024, **9**, 848–863.
89 A. M. Forti, M. A. Jones, D. N. Elbeyli, N. D. Butler and A. M. Kunjapur, *Nat. Microbiol.*, 2025, **10**, 1404–1416.
90 N. Kim, J. Ma, W. Kim, J. Kim, P. Belenky and I. Lee, *Exp. Mol. Med.*, 2024, **56**, 1501–1512.
91 S. Yen and J. S. Johnson, *Mamm. Genome*, 2021, **32**, 282–296.
92 L. Liu, Y. Yang, Y. Deng and T. Zhang, *Microbiome*, 2022, **10**, 209.
93 N. Tadrent, F. Dedeine and V. Hervé, *F1000Research*, 2023, **11**, 1522.
94 A. J. van Heel, A. de Jong, C. Song, J. H. Viel, J. Kok and O. P. Kuipers, *Nucleic Acids Res.*, 2018, **46**, W278–W281.
95 C. Rios-Martinez, N. Bhattacharya, A. P. Amini, L. Crawford and K. K. Yang, *PLoS Comput. Biol.*, 2023, **19**(5), e1011162.
96 M. Liu, Y. Li and H. Li, *J. Mol. Biol.*, 2022, **434**, 167597.
97 Z. Yang, B. Liao, C. Hsieh, C. Han, L. Fang and S. Zhang, *bioRxiv*, 2021, DOI: 10/1101/2021.11.15.468547.
98 F. Del Carratore, K. Zych, M. Cummings, E. Takano, M. H. Medema and R. Breitling, *Commun. Biol.*, 2019, **2**(1), 83.
99 K. Ranjan, M. K. Bharti, R. A. Siddique and J. Singh, in *Microbial Metatranscriptomics Belowground*, Springer Singapore, Singapore, 2021, pp. 1–36.
100 L. Aitmanaitė, K. Širmonaitis and G. Russo, *Int. J. Mol. Sci.*, 2023, **24**, 13786.
101 A. Bauermeister, H. Mannochio-Russo, L. V. Costa-Lotufo, A. K. Jarmusch and P. C. Dorrestein, *Nat. Rev. Microbiol.*, 2022, **20**, 143–160.
102 L. K. Caesar, R. Montaser, N. P. Keller and N. L. Kelleher, *Nat. Prod. Rep.*, 2021, **38**, 2041–2065.
103 L. Chen, W. Lu, L. Wang, X. Xing, Z. Chen, X. Teng, X. Zeng, A. D. Muscarella, Y. Shen, A. Cowan, M. R. McReynolds, B. J. Kennedy, A. M. Lato, S. R. Campagna, M. Singh and J. D. Rabinowitz, *Nat. Methods*, 2021, **18**, 1377–1385.
104 F. Del Carratore, K. Schmidt, M. Vinaixa, K. A. Hollywood, C. Greenland-Bews, E. Takano, S. Rogers and R. Breitling, *Anal. Chem.*, 2019, **91**(20), 12799–12807.
105 F. Del Carratore, W. Eagles, J. Borka and R. Breitling, *Bioinformatics*, 2023, **39**(7), btad455.
106 F. Russo, F. Ottosson, J. J. J. van der Hooft and M. Ernst, Deep learning models for LC-MS untargeted metabolomics data analysis, in *From Computational Logic to Computational Biology: Essays Dedicated to Alfredo Ferro to Celebrate His Scientific Career*, Springer Nature Switzerland, Cham, 2024, pp. 128–144.
107 M. Ludwig, L.-F. Nothias, K. Dührkop, I. Koester, M. Fleischauer, M. A. Hoffmann, D. Petras, F. Vargas, M. Morsy, L. Aluwihare, P. C. Dorrestein and S. Böcker, *Nat. Mach. Intell.*, 2020, **2**, 629–641.
108 M. Ludwig, M. Fleischauer, K. Dührkop, M. A. Hoffmann and S. Böcker, *Computational Methods and Data Analysis for Metabolomics*, 2020, pp. 185–207.
109 A. T. Aron, E. C. Gentry, K. L. McPhail, L.-F. Nothias, M. Nothias-Esposito, A. Bouslimani, D. Petras, J. M. Gauglitz, N. Sikora, F. Vargas, J. J. J. van der Hooft, M. Ernst, K. Bin Kang, C. M. Aceves, A. M. Caraballo-Rodríguez, I. Koester, K. C. Weldon, S. Bertrand, C. Roullier, K. Sun, R. M. Tehan, C. A. P. Boya, M. H. Christian, M. Gutiérrez, A. M. Ulloa, J. A. Tejeda Mora, R. Mojica-Flores, J. Lakey-Beitia, V. Vásquez-Chaves, Y. Zhang, A. I. Calderón, N. Tayler, R. A. Keyzers, F. Tugizimana, N. Ndlovu, A. A. Aksenov, A. K. Jarmusch, R. Schmid, A. W. Truman, N. Bandeira, M. Wang and P. C. Dorrestein, *Nat. Protoc.*, 2020, **15**, 1954–1991.
110 D. W. P. Tay, N. Z. X. Yeo, K. Adaikkappan, Y. H. Lim and S. J. Ang, *Sci. Data*, 2023, **10**, 296.
111 A. Bogdanov, M. N. Salib, A. B. Chase, H. Hammerlindl, M. N. Muskat, S. Luedtke, E. B. da Silva, A. J. O'Donoghue, L. F. Wu, S. J. Altschuler, T. F. Molinski and P. R. Jensen, *Nat. Commun.*, 2024, **15**, 5230.
112 M. W. Mullowney, K. R. Duncan, S. S. Elsayed, N. Garg, J. J. J. van der Hooft, N. I. Martin, D. Meijer, B. R. Terlouw, F. Biermann, K. Blin, J. Durairaj, M. Gorostiola González, E. J. N. Helfrich, F. Huber,

S. Leopold-Messer, K. Rajan, T. de Rond, J. A. van Santen, M. Sorokina, M. J. Balunas, M. A. Beniddir, D. A. van Bergeijk, L. M. Carroll, C. M. Clark, D.-A. Clevert, C. A. Dejong, C. Du, S. Ferrinho, F. Grisoni, A. Hofstetter, W. Jespers, O. V. Kalinina, S. A. Kautsar, H. Kim, T. F. Leao, J. Masschelein, E. R. Rees, R. Reher, D. Reker, P. Schwaller, M. Segler, M. A. Skinnider, A. S. Walker, E. L. Willighagen, B. Zdrazil, N. Ziemert, R. J. M. Goss, P. Guyomard, A. Volkamer, W. H. Gerwick, H. U. Kim, R. Müller, G. P. van Wezel, G. J. P. van Westen, A. K. H. Hirsch, R. G. Linington, S. L. Robinson and M. H. Medema, *Nat. Rev. Drug Discovery*, 2023, **22**, 895–916.

113 J. A. van Santen, S. A. Kautsar, M. H. Medema and R. G. Linington, *Nat. Prod. Rep.*, 2021, **38**, 264–278.