

DISSERTATION | DOCTORAL THESIS

Titel | Title

Prediction of Cross-feeding Interactions in Microbial
Communities using Metabolic Modelling

verfasst von | submitted by

Michael Predl BSc MSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Doctor of Philosophy (PhD)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 794 685 437

Dissertationsgebiet lt. Studienblatt | Field of
study as it appears on the student record
sheet:

Biologie

Betreut von | Supervisor:

Univ.-Prof. Mag. Dr. Thomas Rattei

Acknowledgements

I want to thank my supervisor Thomas Rattei for his valuable support and advice, which taught me a lot scientifically, as well as on planning and organizing research projects.

In addition to Thomas, I very much appreciated having Jürgen Zanghellini und Christine Moissl-Eichinger as members of my thesis advisory committee. Apart from the friendly and fruitful yearly meetings, their support eventually resulted in successful scientific collaborations.

Many thanks to all current and former CUBE members, with whom I had many interesting scientific and not so scientific discussions, and who provided invaluable feedback for this work. Special thanks also to the boardgame enthusiasts at CUBE and CeMESS, with whom I had the pleasure to share many joyful evenings.

I also want to thank the Vienna Doctoral School of Microbiology and Environmental Science for many interesting and high-quality courses that have tremendously improved my skills in scientific writing and figure design, the results of which can be seen in this work.

Finally, my deepest gratitude to Selina for her constant love and support, and for helping me keep my sanity during these four years. This journey would not have been possible without you.

1. Abstract

Life on earth takes place in communities of interacting organisms. These interactions add functionality and complexity on top of the members of a community. As such, they are key to understanding and predicting the behavior of communities, and ultimately applying these findings in medicine, engineering and environmental policies.

This work focuses on a certain type of interaction, called cross-feeding interaction. The term describes a relationship of one organism producing a metabolite, which is then consumed by another. Experimental study of cross-feeding interactions is resource intensive and often relies on prior knowledge of candidates for metabolites mediating said interactions.

Metabolic modelling is a bioinformatics technique for the simulation of metabolic states, which are combinations of reaction activities in a metabolic network of one or multiple organisms. As such, metabolic modelling could be used to explicitly model cross-feeding interactions and propose candidate metabolites for experimental investigation.

This work explores ways to harness the knowledge-base aspect of metabolic modelling for inferring cross-feeding interactions in microbial communities. As a starting point, the secondary metabolite production of a single organism, *Streptomyces venezuelae*, and its change after ethanol stress were investigated. The metabolic network of the organism was reconstructed, including a map from genes to enzymes to metabolic pathways. This allowed investigating gene expression changes in pathway enrichment analysis. Together with metabolomics data, changes in production of commercially desirable compounds, such as the antibiotic chloramphenicol, could be shown in response to ethanol shock.

In a follow up study, this experimental setup was repeated to analyze deletion mutants of regulatory genes that are related to ethanol shock, as identified in the first set of experiments. Massive gene expression changes could be shown in the deletion mutants compared to wild type, as well as over-expression of certain secondary metabolites.

Moving on to modelling of multiple organisms, several gaps and limitations were identified in the metabolic modelling toolbox. Two areas in particular, the generation of community metabolic models and the visualization of interactions, were prohibitive. To solve this, a python package for community metabolic modelling was developed (PyCoMo). PyCoMo generates community metabolic models from metabolic models of its members, implements a model structure allowing to switch between two types of linearization, fixed growth and fixed abundance, and allows efficient calculation of all potential cross-feeding interactions.

In addition, ScyNet, a Cytoscape app was developed, solving the problem of visualizing the complex metabolic networks of community metabolic models. By reducing the metabolic network to community members and exchanged metabolites, community metabolic models and their cross-feeding interactions can be informatively represented.

These methods were applied to a set of pairwise communities of gut archaea-bacteria co-cultures and predicted several possible cross-feeding interactions.

This work presents the potential of metabolic modelling for predicting and understanding cross-feeding interactions and closes gaps in the toolkit. With further improvements and validation, metabolic modelling could provide high quality candidates for metabolomics experiments to measure cross-feeding interactions and guide the engineering of microbial communities.

2. Kurzfassung

Leben auf der Erde besteht aus Lebensgemeinschaften von interagierenden Organismen. Diese Interaktionen erweitern Funktionalität und Komplexität von Lebensgemeinschaften über die Summe der Mitglieder hinaus. Daher sind diese Interaktionen der Schlüssel, um das Verhalten von Lebensgemeinschaften zu verstehen und vorherzusagen, sowie diese Erkenntnisse in Medizin, Technik und ökologischer Gesetzgebung anzuwenden.

Diese Arbeit beschäftigt sich mit einer bestimmten Art von Interaktion, den "cross-feeding" Interaktionen. Der Begriff beschreibt das Verhältnis zweier Organismen, bei dem ein Organismus einen bestimmten Metaboliten herstellt, welcher von einem anderen Organismus konsumiert wird. Um "cross-feeding" Interaktionen zu untersuchen oder nachzuweisen benötigt es teure und zeitaufwendige Experimente, und häufig müssen die zu testenden Kandidaten vorher bekannt sein.

Metabolische Modellierung ist eine bioinformatische Methode zur Simulation von metabolischen Zuständen, der Ansammlung von Aktivitäten chemischer Reaktionen im metabolischen Netzwerk eines oder mehrerer Organismen. Als solches erlaubt metabolische Modellierung die explizite Simulation von "cross-feeding" Interaktionen und dadurch potenziell die Vorhersage geeigneter Kandidaten für Experimente.

Diese Arbeit untersucht Methoden um den Wissensdatenbank Aspekt von metabolischen Modellen für die Vorhersage von "cross-feeding" Interaktionen in mikrobiellen Gemeinschaften zu nutzen. Den Startpunkt bildet die Untersuchung der Produktion von sekundären Metaboliten eines einzelnen Organismus, *Streptomyces venezuelae*, und der Änderungen nach Ethanol-Stress. Dazu wurde ein metabolisches Netzwerk des Organismus rekonstruiert, inklusive einer Verbindung von Genen, zu Enzymen, zu Stoffwechselwegen. Diese Verbindung erlaubte es, die Änderung in Genexpression mittels "pathway enrichment analysis" zu untersuchen. Zusammen mit Metabolom-Daten konnten Veränderungen in der Produktion kommerziell relevanter Stoffe, wie das Antibiotikum Chloramphenicol, in Zusammenhang mit Ethanol-Schock gefunden werden.

Eine nachfolgende Untersuchung wurde mit demselben experimentellen Aufbau die Deletionsmutanten von regulatorischen Genen untersucht, welche in den ersten Experimenten eine Verbindung zu Ethanol-Schock zeigten. Hier konnte eine massive Veränderung der Genexpression in den Mutanten im Vergleich zum Wildtyp beobachtet werden, sowie eine erhöhte Produktion einiger sekundärer Metabolite.

Im nächsten Schritt wird metabolische Modellierung auf mehrere Organismen erweitert, wobei sich mehrere Lücken und Limitationen im Repertoire der Methoden gezeigt haben. Zwei Aspekte waren besonders einschränkend: die Erstellung von metabolischen Modellen von Lebensgemeinschaften, sowie deren graphische Darstellung. Um das erste Problem zu lösen, wurde ein Python package namens PyCoMo für metabolische Modellierung von Lebensgemeinschaften entwickelt. PyCoMo erstellt metabolische Modelle von Lebensgemeinschaften basierend auf metabolischen Modellen der Mitglieder. Die Struktur der Modelle erlaubt einen Wechsel zwischen zwei Modi der Linearisierung, mit fixer Wachstumsrate oder mit fixen Abundanzen der Mitglieder. Weiters ermöglicht PyCoMo eine effiziente Berechnung aller möglichen "cross-feeding" Interaktionen.

Zusätzlich wurde ScyNet, eine Cytoscape App, entwickelt, welche das Problem der Visualisierung der komplexen metabolischen Netzwerke von Lebensgemeinschaften löst. Durch die Reduktion des metabolischen Netzwerks einer Lebensgemeinschaft auf ihre Mitglieder und deren ausgetauschte Metabolite können informative Grafiken erstellt werden.

Diese Methoden wurden auf einen Datensatz von paarweisen Kulturen von Archaeen und Bakterien angewandt und mehrere potentielle "cross-feeding" Interaktionen vorhergesagt.

Diese Arbeit stellt das Potential von metabolischer Modellierung für die Vorhersage der Untersuchung von "cross-feeding" Interaktionen vor, und schließt Lücken im Methoden-Repertoire. Durch weitere Verbesserungen und Validierung könnte metabolische Modellierung hochwertige Kandidaten für Metabolom-Experimente zur Untersuchung von "cross-feeding" Interaktionen liefern, sowie die Erstellung und Optimierung mikrobieller Lebensgemeinschaften leiten.

Contents

Acknowledgements	3
1. Abstract	5
2. Kurzfassung	7
3. Introduction	11
3.1. Communities	11
3.1.1. Analysis of microbial communities	11
3.1.2. Emergent properties	11
3.1.3. Interactions between community members	12
3.2. Cross-feeding interactions	12
3.2.1. Origin and evolution of cross-feeding interactions	12
3.2.2. Importance for cultivation	14
3.2.3. Investigation of cross-feeding interactions	15
3.3. Metabolic modelling	16
3.3.1. Content of metabolic models	16
3.3.2. Reconstruction of metabolic models	17
3.3.3. Modelling techniques	18
3.3.4. Modelling microbial communities	20
3.3.5. Known challenges and pitfalls	21
3.4. Outline	22
4. Effects of ethanol shock on <i>Streptomyces venezuelae</i> secondary metabolism	25
4.1. Context	25
4.2. Publication status	25
4.3. Author contributions	26
4.4. Other manuscripts	26
4.4.1. Three conserved extracytoplasmic function sigma factors in <i>Streptomyces</i> have a major impact on secondary metabolism	26
5. PyCoMo - generating and analyzing community metabolic models	27
5.1. Context	27
5.2. Publication status	27
5.3. Author contributions	28
5.4. Further experiments	28
5.4.1. Introduction	28
5.4.2. Methods and Results	29
5.4.3. Discussion	30
6. ScyNet - visualizing interactions within community metabolic models	31
6.1. Context	31
6.2. Publication status	31
6.3. Author contributions	32

7. Interactions in Archaea - Bacteria co-cultures	33
7.1. Context	33
7.2. Publication status	34
7.3. Author contributions	34
8. Discussion and Conclusion	35
Bibliography	37
Abbreviations	47
A. Appendix	49
A.1. Targeted Metabolomics and High-Throughput RNA Sequencing-Based Transcrip- tomics Reveal Massive Changes in the <i>Streptomyces venezuelae</i> NRRL B-65442 Metabolism Caused by Ethanol Shock	49
A.1.1. Publication status	49
A.1.2. Author contributions	49
A.1.3. Publication	49
A.1.4. Supplementary data	69
A.2. Three conserved extracytoplasmic function sigma factors in <i>Streptomyces</i> have a major impact on secondary metabolism	70
A.2.1. Publication status	70
A.2.2. Author contributions	70
A.2.3. Publication	70
A.3. PyCoMo: a python package for community metabolic model creation and analysis .	101
A.3.1. Publication status	101
A.3.2. Author contributions	101
A.3.3. Publication	101
A.3.4. Supplementary data	106
A.4. ScyNet: Visualising interactions in community metabolic models	113
A.4.1. Publication status	113
A.4.2. Author contributions	113
A.4.3. Publication	113
A.5. Expanding the cultivated human archaeome by targeted isolation of novel Methan- obrevibacter strains from fecal samples	123
A.5.1. Publication status	123
A.5.2. Author contributions	123
A.5.3. Publication	123
A.5.4. Supplementary data	159

3. Introduction

3.1. Communities

Communities are the fundamental organization of life. Other than colonies, which are groups of organisms of the same species or strain, communities are groups of organisms of different taxa. These organisms live in the same habitat and can interact with each other [1]. The term community is, however, not limited to cooperative relationships, it merely describes cohabitation of diverse organisms. Communities can include taxa from different kingdoms, such as having both unicellular and multicellular organisms, or plants and animals as members. As such, the term can also describe host-microbiome, or host-parasite relationships.

Defining the term makes clear, how ubiquitous and diverse communities are. This thesis focuses on microbial communities of bacteria and archaea. Microbial communities are present in almost every part on earth, from the deep-sea [2] to deserts [3], wastewater [4] or the human gut [5]. Given their prevalence, they are relevant to many scientific fields, but also agriculture, industry and medicine [5–7].

3.1.1. Analysis of microbial communities

A fundamental method for the analysis of natural microbial communities is metagenomics [8, 9]. It allows to assess the genomic and phylogenetic diversity without the need for cultivation. There are also bottom-up approaches using model communities of cultivable species [10–12]. These approaches are resource demanding, as there are many different parameters to be investigated. Thus, such model communities consist of only a few members, as the complexity and required work for experiments and analyses rises quickly with the number of community members. In both cases, many different experimental techniques can be applied to study specific aspects, such as metabolomics [13], metatranscriptomics [14], or spatial resolution with FISH and microscopy [15, 16]. These are restricted by the availability and amount of study material, which can be extremely limited for in vivo experiments, such as in ocean communities or medical samples.

3.1.2. Emergent properties

The reason communities are so widely prevalent is, that they are more than the sum of their members. In other words, communities have emergent properties [17–19]. For example, waste degradation is a complex metabolic process, involving numerous steps. Microbial communities are capable of such metabolic processes, despite each member only being able to carry out a subset of metabolic conversions. In addition, greater functional potential allows communities to better adapt to environmental changes, or to use a specific niche in the first place. Having redundancy in community members of specific functions can lead to higher stress tolerance. Furthermore, over the course of time, a community can differ in the abundance or even the presence of its members. Thus, communities are much more fluid than individual organisms, which have to mutate in order to evolve. As such, emergent properties are the reason for the success and prevalence of microbial communities.

3.1.3. Interactions between community members

A big driver of emergent properties are the interactions between community members. The term interaction describes a set of vastly different mechanisms and outcomes, which can also be used to categorize them [20, 21]. An interaction requires at least two different actors. A popular way of categorizing interactions is, by the outcome for the actors (positive, neutral or negative). The combinations of the interaction signs gives a total of six distinct interaction categories (Fig. 3.1). While this model can be intuitively understood, the sign of an interaction can be difficult to measure and also depends on the time frame in question (long-term vs. immediate).

Interactions can also be categorized by mechanism. For example, one actor can eat another (predation), or make nutrients available in the environment by secreting catalytic enzymes [21–23]. The scope of this work is a specific mechanistic type of interaction, the cross-feeding interaction, which will be explained in greater detail in section 3.2. Interactions also include complex traits such as quorum sensing, or biofilm generation [22, 24]. These two examples showcase interactions, where even the definition of the actors in the interactions is difficult. However, it is quite clear how these interactions can give rise to emergent properties in communities: Quorum sensing allows a cooperative reaction and regulation within the community, and biofilms can protect otherwise vulnerable microbial members. Thus, interactions are crucial for the functionality of communities. Consequently, understanding of interactions is important for ecological policies, designing and optimizing communities for industrial purposes and applications in human health.

3.2. Cross-feeding interactions

Cross-feeding interactions consist of one organism producing and secreting a metabolite, which is taken up and consumed by a second organism [25]. This process is, however, not limited to two interaction partners, but can involve several producers and consumers.

Cross-feeding interactions have a positive outcome for the recipient. The outcome for the producer is mostly neutral, however a positive outcome is possible if the cross-fed metabolite would be toxic or inhibiting for the producer [26]. Conversely, a negative outcome can be seen for the producer, if the release of the metabolite is involuntary, or the released metabolite was meant to stay in the environment or be taken up later [25].

3.2.1. Origin and evolution of cross-feeding interactions

In microbial communities, cross-feeding interactions are well known in trophic interactions, such as the human gut, or waste degradation [27]. However, the diversity and complexity of microbial community is far greater than the number of their trophic levels. Thus, community members are expected to be in competition, which has been seen as a disagreement with a fundamental principle in ecology, the competitive exclusion principle [25]. This states, that if organisms are in direct competition, under stable conditions and given enough time, only the organism with the highest fitness will survive. Multiple theories have formed, trying to explain this phenomenon: the existence of niches allows co-existence of competitors, natural environments being fluctuating and frequently perturbed, and cross-feeding interactions which provide a benefit greater than the cost of competition.

It has been shown that cross-feeding can be established beyond trophic interactions. The following paragraphs will give examples on mechanistic and evolutionary explanation for the origin and prevalence of cross-feeding interactions.

A key requirement for cross-feeding interactions is that the target metabolite needs to be available to the recipient. Apart from intended secretion of metabolites like waste products, this can also happen involuntarily, e.g. by leakage through the membrane [28–30]. The cell membrane is typically considered a clear separation between the inside and outside of the cell. However, metabolites can

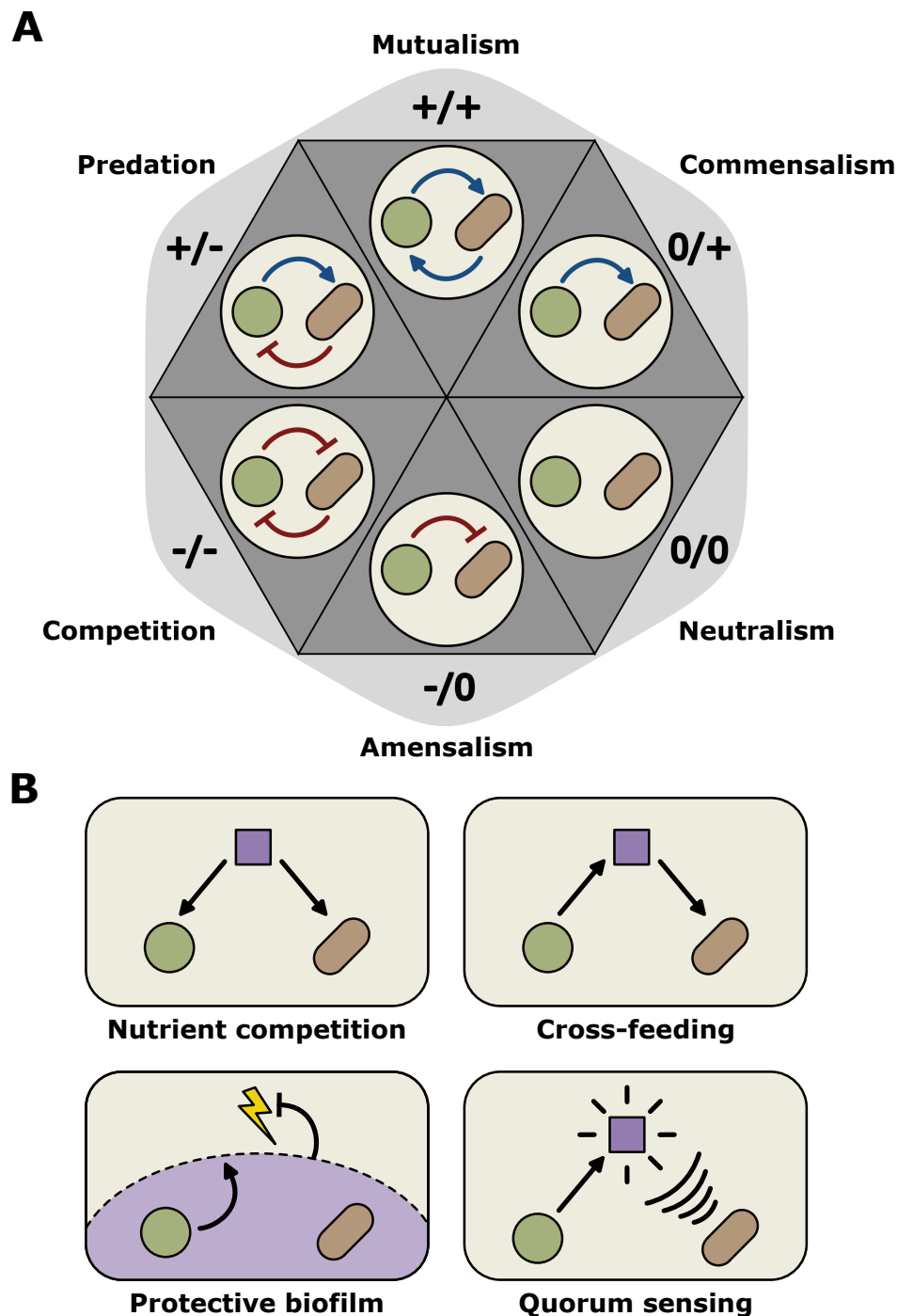


Figure 3.1.: Types of interactions in microbial communities. (A) Categorization by sign of the outcome for the interacting organisms. (B) Four examples of interaction mechanisms. Interacting organisms are depicted as circular and rod shapes, metabolites are symbolized by squares.

pass through the cell membrane, given they have specific characteristics [29]. Another mechanism for metabolite leakage are transporters [30]. If they are not coupled to an energy consuming mechanism, they are open in both directions.

A theoretical study has also predicted, that many organisms are able to produce and secrete metabolites without a decrease in fitness [31]. This costless secretion could then mediate cross-

feeding, which is beneficial if the interaction partner can in turn provide nutrients.

In cases of nutrient degradation involving multiple energy generating steps (like glucose to ethanol), it has been shown that organisms can decide to focus on a specific part of the degradation [25]. In cases of nutrient abundance, the specialization on degradation steps with high energy yield leads to a fitness advantage. As a consequence, intermediary metabolites are released, which can be taken up by other organisms to perform the remaining degradation. Such a pattern of cross-feeding has been shown in an evolutionary experiment of *E. coli* [32, 33].

This specialization of *E. coli* can also be seen as division of labor. A similar case has been observed in clonal yeast colonies, grown on agar plates with nutrients only available on specific locations of the plate [34]. Upon sufficient colony growth, the cells specialize into two subgroups, one with scouting behavior, advancing towards new nutrient pools over resource scarce spaces, and another which supplies the scouting group with nutrients. This cooperative behavior benefits the whole population, by allowing to access new resources.

Another origin of cross-feeding interactions is the black queen hypothesis [35]. It describes a community growing on a common nutritional source, that requires a specific function to be made available (e.g. polymer to monomer degradation), and the product of this function is available to all members. In this scenario, organisms who lose this function can gain a fitness advantage, by stealing the nutrients provided by others while not bearing the cost of their production (i.e. becoming cheaters). Of course, one last organism, the black queen, has to retain this function, otherwise the community would collapse. In fact, while the cheating organisms have higher fitness, there is negative frequency-dependent selection. This results in the cheaters' fitness being reduced, if the black queen's abundance decreases.

This hypothesis can also explain the establishment of mutual cross-feeding, if the black queen of one function becomes a cheater on another. This then creates a system of co-dependent auxotrophs, or a community with division of labor [25].

The full extent and prevalence of cross-feeding interactions is a matter of ongoing research. Some theoretical studies have predicted, that mutualistic cross-feeding interactions are rare, and the main mode of interaction is competition [31, 36]. However, the examples of the black queen hypothesis and costless secretion also show that establishing cross-feeding interactions leads to a fitness increase in recipient species. The outcome for the community can be difficult to define, but the concept of cross-feeding as a form of division of labor proves that an increase in fitness for the community is possible.

3.2.2. Importance for cultivation

Cross-feeding interactions shape the medium of microbial community members. This is important for the cultivation efforts of yet unculturable species [37, 38]. It is well established that only a small fraction of microbial diversity can be cultivated in laboratories (estimated ~1% of bacterial and archaeal species). This problem has been termed the great plate count anomaly. Unculturability is not a trait, rather it is the circumstance that the right conditions were not yet found for a certain species. These conditions include (1) physical parameters, such as pH and temperature, (2) the removal of toxic waste products and (3) the requirement of specific substrates (or certain concentrations of them) [37, 38].

The presence of cross-feeding interactions shows, that the space of available metabolites in a community differs from the input medium. An example is the external degradation of polymers such as chitin to accessible oligomers and monomers [23]. While, conceptually, all organisms in a chitin degrading community have chitin as their main source of energy, many community members cannot utilize it directly and require the presence of primary degraders. Further, studies by D'Souza et al. [39, 40] predict, that auxotrophies are prevalent in microbial organisms, which are likely complemented by cross-feeding interactions. Examples have also been found of co-dependent, syntrophic organisms, that could be cultivated together, but not individually [37]. This implies, that

knowledge of cross-feeding interactions can facilitate media design, by informing about the additives required for the growth medium.

3.2.3. Investigation of cross-feeding interactions

Experimental investigation

The identification of Cross-feeding interactions is challenging, as this would require proof of the production, the transmission and the consumption of the same metabolite in the two actors simultaneously. Stable isotope labelling methods allow to detect the presence of certain metabolites, but tracing the transmission can be difficult [41–43]. Interestingly, cross-feeding has been a source of noise for isotope labelling experiments in microbial communities, as these interactions obfuscate primary degraders [44]. Apart from these obstacles, labelling experiments rely on prior knowledge of candidate metabolites. Carrying out these experiments is also limited by the amount of sample material and accessibility for labelling, making them especially difficult for uncultured communities.

Apart from direct measurement, numerous methods can provide evidence for the presence of cross-feeding interactions, such as untargeted metabolomics, metatranscriptomics and meta-proteomics [45–47]. The latter two allow inferring active metabolic pathways, which can then be matched between the organisms. An overlap of pathways producing a metabolite in one organism and another consuming the metabolite in a second organism provides evidence for cross-feeding. However, the presence of transcripts and proteins does not prove that they are active, and the transmission of the metabolite can not be shown by these methods. Further, these methods are limited by correct functional prediction of genes, as well as the detection of transcripts and proteins of community members with lower abundance.

Untargeted metabolomics allows to measure the present metabolites. A big limitation of this method is the identification of measured metabolites, due to noisy data and ambiguous mapping of isomers [48–50]. In addition, metabolomics cannot determine the location of the metabolites. Thus, there is no distinction between metabolites being present in one organism or another. This problem can be tackled by comparing the metabolome of community members in single culture, with those in pair-wise co-cultures [23, 51]. The differences in presence and concentration of metabolites is then used to infer the consumption and production of metabolites, and ultimately the presence of cross-feeding interactions.

As an alternative to direct co-cultures, research is also being done on cultivation of supernatant co-cultures [51]. Here, one organism is grown in the base medium, changing it by consuming and producing metabolites. The cell-free supernatant is then harvested and used as growth medium for the second organism. The metabolomes of the base medium, the supernatant after growth of the first organism, and the final medium after growth of the second organism are measured and compared. This data can then be used to infer consumed and produced metabolites, as well as the presence of cross-feeding interactions.

While these co-culture experiments do not directly prove cross-feeding interactions, they provide strong evidence for them. However, they are also resource intensive and limited by the detection abilities of metabolomics methods. Unfortunately, these experiments can only be used on cultivable organisms and are hard to realize in vivo.

Prediction of cross-feeding interactions

Experimental methods for the investigation of cross-feeding interactions are resource intensive, and largely restricted to in vitro settings and cultured organisms. An alternative route is provided by bioinformatics methods. These allow to make predictions of likely cross-feeding interactions and can guide later experimental efforts validating such candidates.

Statistical methods are used to identify correlation between the presence and absence of organisms [52]. Paired with functional predictions (either genome based or aided by metatranscriptome

and metaproteome data) these correlations can be used to predict cross-feeding interactions [53, 54].

Machine learning methods have also been applied to a variety of topics related to the prediction of cross-feeding interactions, such as metabolite-organism interactions or metabolite profiles of communities [55, 56]. However, there are two factors limiting the success and applicability of machine learning on the prediction of cross-feeding. First of all, training machine learning models requires a large datasets, increasing with the complexity of the model. Due to the difficulty of proving or providing evidence for cross-feeding interactions, far too little experimental data are available. The second limitation is generalizability, or the applicability to new data. Even if enough data were available to train a cross-feeding prediction model on e.g. the human gut microbiome, it is highly unlikely that the model could provide any insights in other communities.

Lastly, metabolic modelling is a technique, based on the metabolic network of one or multiple organisms. As such, it can directly model cross-feeding interactions. This technique will be described in detail in the following section.

3.3. Metabolic modelling

Metabolic modelling is a bioinformatics method that investigates flux configurations in metabolic networks. It follows a systems biology approach, integrating information from genes to enzymes.

The aim of metabolic modelling is to provide mechanistic insights into the metabolic networks. Changes in the genotype or external conditions can be modeled and used to predict the phenotype under varying conditions.

While the beginnings of metabolic modelling are hard to define, the reconstruction of the first genome-scale metabolic model was achieved in 1995 by Fleischmann et al., of *Haemophilus influenzae* Rd. [57]. Since then, it has been applied to many different model and non-model organisms [58, 59]. Metabolic modelling has been a great success in designing microbial strains for metabolite production [60], but also found applications in many other areas, such as human health [61–63] and research on microbial communities [64, 65].

3.3.1. Content of metabolic models

A metabolic model, or also called metabolic network reconstruction, is a collection of chemical reactions and model parameters. The chemical reactions can be represented as a directed graph of reactions and metabolites, here called the metabolic network (see fig. 3.2). The metabolites in a metabolic network can be further organized in compartments, which represent physical spaces such as the medium, the inside of a cell, or organelles.

When used for mathematical modelling, the chemical reactions are typically represented by a stoichiometric matrix S . This matrix contains m rows of metabolites and r rows of reactions, resulting in a form of $S \in \mathbb{R}^{m \times r}$. The matrix S contains the stoichiometric coefficients of each metabolite i in reaction j at position $S_{i,j}$. A positive coefficient denotes the metabolite is a product, a negative value denotes a substrate, and 0 corresponds to the concentration of a metabolite not being changed by the respective reaction.

The activity of reactions is given by a flux vector $v \in \mathbb{R}^r$. A reaction flux corresponds to the turnover rate of the reaction, or simply put its activity. The unit of a flux is $\text{mmol} \cdot \text{h}^{-1} \cdot \text{gDW}^{-1}$ and as such, metabolic models are normalized to one gram of their dry mass. Multiplication of the stoichiometric matrix with the flux vector calculates the change in concentration of each metabolite in time t , resulting in a vector $c(t) \in \mathbb{R}^m$:

$$S * v = c(t) \tag{3.1a}$$

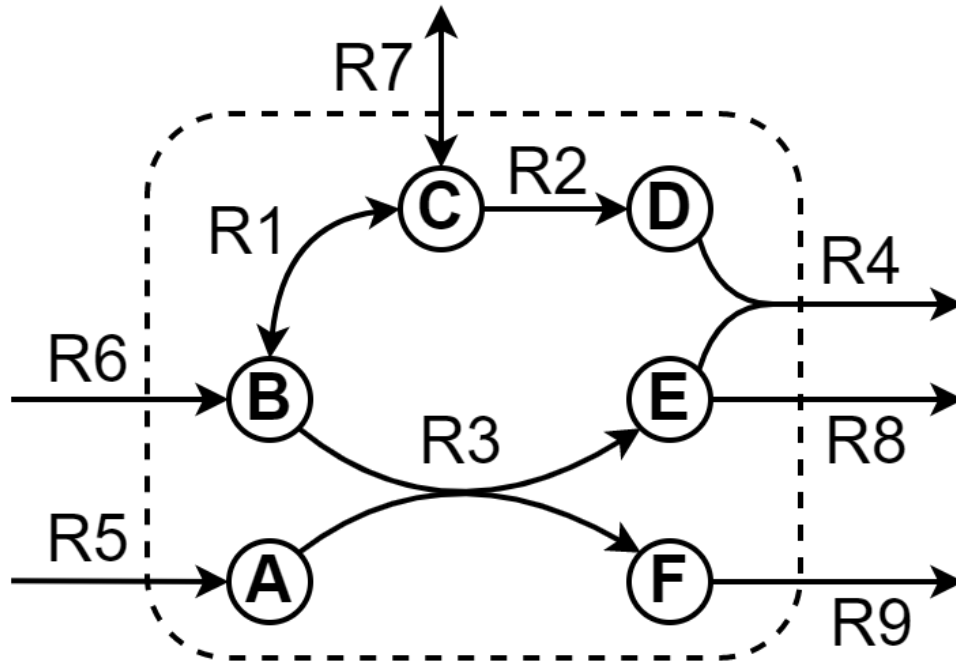


Figure 3.2.: A toy example of a metabolic network. Reactions are depicted as arrows and named R1-9. Metabolites are shown as circles and named A-F. The dotted line encasing the metabolites represents the boundary of the model. Reactions R5-9, which pass this boundary, are exchange reactions, allowing interaction with a hypothetical medium.

To allow influx and efflux of mass to and from the model, a special set of reactions exists, called boundary reactions. These reactions are not mass balanced and as such also no real chemical reactions. They typically take the form of a reaction producing a single metabolite from nothing, or consuming one without a product (see fig. 3.2 reactions R5-9).

Another special type of reaction is the biomass reaction (see fig. 3.2 reaction R4). It is an abstract reaction, symbolizing the growth of the modelled organism. A biomass reaction has the building blocks of a cell as substrate, such as nucleotides, amino acids, fatty acids and vitamins. The stoichiometric coefficients of these substrates are proportional to their abundance in the cell. By convention, the mass of substrates sums up to $1g$ (equal $1gDW$). This allows direct translation of the biomass flux to growth rate per hour.

Metabolic modelling can analyze the different metabolic states of the underlying network. These states are the combination of reaction activities, given by the flux vector v . However, the rate of chemical reactions is limited by a variety of factors, including physical conditions, thermodynamics or the availability of enzymes. Constraint-based modelling [66] includes such restrictions in the model, by providing a lower (lb) and upper bound (ub) for flux values of reactions:

$$lb_j \leq r_j \leq ub_j \quad (3.2a)$$

3.3.2. Reconstruction of metabolic models

Metabolic models can represent anything from specific metabolic aspects of single organisms [67], to complex communities [64], with different levels of detail and abstraction. In the past, metabolic models have often been built of select metabolic subnetworks, as the labor for manually constructing such models was immense. Developments in databases and automated reconstruction methods have shifted the field to use genome-scale metabolic models (sometimes called GEMs, or GSMMs), which aim to include the entire metabolic capability encoded in an organisms genome [68].

Thiele and Palsson have published a widely referenced protocol for the reconstruction of metabolic models [69]. They describe the reconstruction of a metabolic model as an iterative process, comprised of draft model reconstruction, refinement, simulation and evaluation.

The basis for the reconstruction of a metabolic model is the genome of the target organism. The next step is the annotation of the genome and functional prediction of the genes. The gene functions are then translated into metabolic capabilities, such as catalyzing chemical reactions, or the transport of metabolites. This is done with the help of one or several databases, such as KEGG, modelSEED, or the Pathway databases [70–72]. Based on this collection of proposed chemical reactions, a metabolic network is built.

This draft model is likely to contain gaps, e.g. due to missing functional predictions. As metabolic models require pathways to be connected in order for them to carry flux, these gaps need to be closed. This is done by the process of gap-filling [73], where candidate reactions are added to the draft model, until the gaps are closed. The selection of reactions can be guided by sequence similarity of enzymes to the target's genes, minimizing the number of added candidates, or phylogeny [73–75]. The identification of which gaps need to be filled is done by setting a medium, in which growth of the organism is expected, and then fixing gaps until flux through the biomass function is possible.

While the 96 steps of this protocol can still be followed manually, there has been effort to automate much of this process [68, 74]. There have also been some methods deviating from this single organism bottom-up approach. CARVEme constructs a single metabolic model, containing all possible chemical reactions and refining this complete model [76]. Then, organism specific metabolic models are created by carving away reactions without genomic evidence, while still ensuring that the biomass function can carry flux. The AGORA models are created by a reconstruction pipeline where manual curation in one organism is propagated to all other organisms with similar problems [77].

This iterative curation is a central principle for the development of genome-scale metabolic models, as can be seen by the long standing community efforts to reconstruct and improve metabolic models of model organisms such as *E. coli*, *S. cerevisiae* or *Homo sapiens* [63, 78, 79].

Metabolic models can be stored and transferred in a variety of file types. However, there is a community effort to create a standard, making tools and models compatible. This standard is called SBML (systems biology markup language) [80]. It is a text based file format, using xml structure to define the different model components and their attributes.

3.3.3. Modelling techniques

Metabolic models contain information on a set of chemical reactions. This set of reactions can be analyzed with a variety of methods, depending on the research question, available data and resources. Toolboxes for metabolic modelling have been implemented in Matlab, R and Python [81–83]. They provide many functionalities including import and export of metabolic models, their analysis and refinement. Importantly, they also provide an interface to linear programming solvers, which allow efficient calculations. In the following sections, an overview of common metabolic modelling techniques will be given.

Steady-state

As metabolic models can be very large and complex, efficient ways to calculate fluxes have been developed. One such way is to analyze the metabolism of an organism in steady-state. The term defines that the sum of mass and charge going into a system is equal to the amount of mass and charge leaving the system. Additionally, it is defined as a state, where the rates of production and consumption of each internal metabolite are equal. This results in a system, where the concentration

of its metabolites does not change, which is represented by the following equations.

$$c(t) = \frac{dc}{dt} = 0, \quad (3.3a)$$

$$S * v = 0 \quad (3.3b)$$

The steady-state assumption turns the system of equations from a set of differential equations into a set of linear equations, a much simpler problem to solve.

A cell can be assumed to be in steady-state when growing in a well mixed, continuous culture. Apart from this, quasi-steady-states can be assumed if organisms operate at temporary steady-states and the switch between such states is considered almost instantaneous [84]. A different justification for the applicability of steady-state is in oscillating cultures or communities, where a steady-state would describe the long-term average metabolic state [85].

FBA

A commonly used analysis technique of steady-state metabolic models is flux balance analysis (FBA). Here, a linear programming algorithm is used to find a flux vector which optimizes an objective function [86]. Such an objective function can be as simple as the flux through the biomass reaction (i.e. maximizing growth rate), but also more complicated with multiple objectives [87, 88]. FBA is a computationally efficient approach and thus frequently used. However, the applicability of optimality in natural systems has been a matter of debate [89]. A similar approach, called parsimonious flux balance analysis (pFBA), tries to deal with this optimality, by applying an additional objective of minimizing the sum of fluxes [87]. This represents the goal of organisms not to waste resources and has been shown to yield more accurate metabolic states in some models.

In both FBA and pFBA, only a single optimum is found. This makes these approaches computationally efficient, but often multiple metabolic states can reach the same optimum in the objective function. Flux variability analysis (FVA) is a technique used to calculate feasible flux ranges for all reactions within the constraints of a model [90].

Similar to the question of applicability of optimality in natural systems, the choice of objective functions is also a matter of debate [89]. While the goal of maximizing the growth rate is an obvious and suitable choice for microorganisms in logarithmic growth phase, it is unclear what metabolic goal, if any, can be phrased for dormant microorganisms or cells of multicellular organisms.

flux sampling

One approach to investigate the metabolic states without imposing an objective function is flux sampling [91, 92]. This technique uniformly samples feasible flux configurations from the solution space. Based on the gathered metabolic states, the range and probability distribution of all reactions' flux values can be calculated. The method's independence from cellular objectives has made it especially useful in metabolic models of multicellular organisms. For example, flux sampling has been applied to elucidate metabolic adaptations of *Arabidopsis thaliana* to colder environments [93].

Dynamic modelling

Some behaviors can only be described by including changes over time, such as adaptations to fluctuating environments or the growth of multicellular organisms. Simulations including these temporary aspects can be performed with dynamic metabolic modelling. One such method is kinetic modelling, which describes all chemical reactions in the model in a set of ordinary differential equations [94, 95]. As such, a big challenge in kinetic modelling is the increased computational complexity, compared to steady-state modelling. Further, this approach requires additional parameters for each reaction, such as Michaelis constants or enzyme concentrations. Such parameters are

only available for a small fraction of enzymes, presenting a limitation for the use of kinetic models [96]. Still, kinetic models have been applied for strain design in biotechnology or the prediction of adversary drug effects [61, 97].

There also exists a hybrid approach between steady-state and kinetic modelling, called dynamic flux balance analysis (dFBA) [98]. This method solves a steady-state problem at each time point and changes the concentrations of only the metabolites outside the model (i.e. the medium). The use of steady-state modelling is justified by the postulated presence of quasi-steady-states in dynamic systems [84, 85]. While more computational power is required, compared to standard steady-state methods, it is still vastly more efficient than kinetic modelling, which means it is also suited for larger, genome-scale metabolic models. DFBA has been applied to a multi-tissue metabolic model of *Arabidopsis thaliana* and elucidated the reprogramming of metabolism during growth [99]. Some methods have even extended the dFBA approach to also include spatial information, making it possible to predict the morphology of prokaryotic colonies [100, 101].

Knowledge-base aspect

Lastly, metabolic models can facilitate the analysis of an organism by serving as a knowledge-base [102]. The construction of a metabolic model consists of gathering information on the genes, gene functions, growth media and several other aspects [78, 103]. All these data types are connected via the metabolic network, providing additional benefit, such as linking genes to pathways. This also allows the investigation of metabolic network on a systems level, using network analysis approaches [104, 105].

3.3.4. Modelling microbial communities

Metabolic models can be built not only for individual organisms, but also for communities. While metabolic models of communities have been created since the early 2000s [106], the recent advances in automated model reconstruction and the increase in available computational power have let the field grow [107]. Metabolic models are well suited for the analysis of communities, as they can explicitly model the metabolic interplay between community members through resource competition and cross-feeding [20, 22, 107].

Compared to single organisms, the metabolic capabilities of communities quickly increase with the number of members. Thus, computational power can be a limiting factor for their analysis. To this end, metabolic models of communities have been constructed with different levels of detail [20].

The greatest level of abstraction is found in lumped models, which view communities as a "bag of genes" [22]. Such models do not consider cellular boundaries and instead build a metabolic network of the entirety of the communities' gene functions.

A different approach can be taken by modelling communities as a set of individual metabolic models. Here, each member's metabolic model is analyzed individually, and interaction takes place through the medium. The medium is iteratively updated by considering the analysis results of each of the member models, as done by Diener et al. [65].

A more fine-grained approach can be found in metabolic models of guilds [20, 108]. Here, a representative organism is chosen which carries out the function of its respective guild. Such a representative can be an actual organism, or an abstract combination of functions attributed to a guild. Each guild representative is, however, modelled as an individual organism, including cellular boundaries. Thus, there is a distinction between intracellular, private and extracellular, public metabolites and the exchange of metabolites requires transporters. This separation into different guild compartments allows explicit modelling of interactions and the identification of the facilitators of specific metabolic functions. On the other hand, many chemical reactions, such as the ones involved in the central carbon metabolism, will be present multiple times in the model, as each guild

needs to have their own copy. This increases the size of the metabolic network and thus requires more computational power [64].

Lastly, the highest level of detail is achieved in compartmentalized community metabolic models [20, 64, 109]. These are similar to models of the guild approach, but each community member is explicitly represented. Depending on the community, such a model would need to include hundreds of members, as in the case of the human gut microbiome.

For all the different types of community representations, the resulting metabolic models can be analyzed using the same techniques as for the single organisms described in section 3.3.3.

The construction of community metabolic models, however, requires some additional work [64]. The biomass reaction of community metabolic models needs to be constructed from the biomass reactions of all community members. Secondly, flux values now need to be normalized by the dry weight of the entire community, rather than individual organisms. Thus, the abundance (or mass fraction f of each community member needs to be taken into account when applying flux bounds to reactions. For steady-state conditions to apply, communities must also be in balanced growth. All community members need to exhibit equal growth rates, otherwise their relative abundance changes and the steady-state condition is violated. Thus, for a community with n members, the following equations need to be fulfilled:

$$\mu_c = \mu_1 = \dots = \mu_n \quad (3.4a)$$

$$\sum_{i=1}^n f_i = 1 \quad (3.4b)$$

$$\mu_i = r_{bio,i} / f_i \quad (3.4c)$$

$$lb_i * f_i \leq r_i \leq ub_i * f_i \quad (3.4d)$$

This results in a bilinear problem, which is computationally far more demanding to solve. The problem can be linearized by setting the growth rate or the community members' mass fractions to fixed values.

3.3.5. Known challenges and pitfalls

Metabolic models have been successfully applied to a variety of biological research questions. However, there are many known challenges and pitfalls at every step, from the experimental setup, the construction of metabolic models, to the final analysis. In this section, a few of the most relevant issues in regard to the prediction of cross-feeding interactions will be presented.

Starting with the reconstruction of metabolic networks, it is clear, that this step is limited by the capabilities of functional prediction [89, 110]. Many gene annotation and functional prediction methods rely on a few experimentally studied genes and enzymes, which mostly originate from model organisms. Thus, functional prediction is more incomplete and unreliable for organisms coming from less studied taxa.

There are, however, also restrictions coming from the translation of predicted functions to chemical reactions. The reactions of metabolic networks are precise, as they operate on specific metabolites and are required to maintain a mass and charge balance. Many functional predictions can only provide information on an abstract class of enzyme function or the involved metabolites [111, 112]. Most of these translations thus rely on biochemical databases [68, 69]. Each database has a different scope and coverage of chemical reactions. Combining information from multiple databases is hindered by the fact, that metabolites of the same name can have different mass, charge or level of detail [113, 114]. In an effort to create a mapping scheme of the metabolite namespaces, MetaNetX was created [115]. While it provides a valuable resource, the translation into a consistent metabolic network is still an open issue, which results in researchers choosing only a single database, or putting in extensive manual curation [74, 77].

A related problem is faced in the formulation of the biomass reaction [89]. The types and amount of metabolites present in an organism is often unknown. Therefore, biomass reactions are often estimated based on the genome (for nucleotide and amino acid utilization) and the taxonomy. However, the formulation of a biomass reaction can have a huge impact on the metabolic model [116]. Similarly, metabolic models can include a reaction representing energy spent for keeping up cellular functions. The formulation is usually the conversion of ATP to ADP at a fixed rate. Estimating the required energy is difficult, but can have a big impact on the analysis of the model [117].

Constraint based modelling aims to find realistic metabolic states by setting bounds of reactions to ranges feasible in the tested conditions. However, such flux ranges are difficult to measure and thus not much data are available [69]. Approximation via enzyme kinetics is possible, but also limited by available data. As such, setting reaction bounds for metabolic models of non-model organisms is unreliable.

Apart from the metabolic network of the target organisms, defining the growth medium can prove to be an obstacle as well. In vitro experiments can be done in well defined growth media, where the contained metabolites and their concentrations are known. However, many common growth media contain complex ingredients, such as peptone and yeast extract [118]. Such ingredients contain a variety of different and often unknown metabolites, and due to the process of their production, their content can change between suppliers and even batches of products. This problem is even more pronounced when modelling natural media, such as in the human gut or soil, where the contents are highly variable. These media need to be approximated with a selection of expected metabolites, based on literature or available experimental data [69].

The analysis of metabolic models faces limitations as well. One aspect is the required computational power. It increases non-linearly with the size of the metabolic network in most analysis methods [64, 119, 120]. This means, that the analysis toolbox is restricted by the size and complexity of the model.

Another aspect is the many choices in the analysis strategy. While the metabolic modelling techniques explained in 3.3.3 are only an overview, it already shows many assumptions and restrictions inherent to the different strategies: Whether to look into a single metabolic state, multiple states, or flux ranges, including the dynamics of a system or analyzing the steady-state, the choice of the objective function, or to simulate without an optimization principle [89].

These choices are also tied to a different issue, the differentiation between the potential and the realized metabolic state of an organism. Ideally, a genome-scale metabolic model will include the entirety of metabolic capabilities. However, it is expected that only a subset of the metabolic network is active at any given time. The goal of finding the realized metabolic state can be challenging, even with additional data such as transcriptomics provided [121].

Ensemble techniques have been proposed to acknowledge the uncertainty, but they also increase the solution space and are computationally more demanding [89]. Ultimately, iterative curation is the best way to deal with errors in metabolic models. Continuous refinement and evaluation on new data leads to models that can achieve more accurate predictions, as is the case for the several ongoing projects for model organisms. In addition to continuous efforts, validation of predictions and results of metabolic models would be very beneficial. However, data for such validation is scarce and follow up experiments are rarely done. This not only limits the usefulness of predictions done by metabolic models, but also hinders the evaluation of metabolic modelling methods.

3.4. Outline

The central aim of this work is to advance the field of metabolic modelling for the prediction of cross-feeding interactions in microbial communities. The ulterior goal is to utilize such predictions as candidates for experimental validation, for genome-guided cultivation, as well as design and optimization of microbial communities for metabolite production, waste degradation and human health.

The following chapters 4-7 present the scientific work, with the full publications and manuscripts being included in the appendix. The first work package explores the use of the knowledge-base aspect of metabolic models for the interpretation of transcriptome data in *Streptomyces venezuelae* (chapter 4). It is followed by two software packages for community metabolic modelling: PyCoMo, a Python package for generating and analyzing compartmentalized community metabolic models (chapter 5), and ScyNet, a network reduction and visualization algorithm for informative representation of interactions in community metabolic models (chapter 6). Finally, community metabolic modelling is applied to investigate potential cross-feeding interactions in pair-wise co-cultures of archaea and bacteria for the enrichment of *Methanobrevibacter* from human fecal samples. Chapter 8 discusses the impact of the presented work and examines further research questions and opportunities.

4. Effects of ethanol shock on *Streptomyces venezuelae* secondary metabolism

4.1. Context

As a first step towards using metabolic modelling in microbial communities and the prediction of cross-feeding interactions, we started with the analysis of a single organism. While metabolic modelling is an established technique, here it is applied on *Streptomyces venezuelae*, a non-model organism, at least relative to the field of metabolic modelling. Related to this circumstance, we wanted to test the quality of automated metabolic reconstructions, as manual curation is out of scope for the application in community metabolic modelling. Further, we aimed to test if the knowledge-base aspect of the reconstruction can be harnessed for the analysis of gene expression data.

This work package investigates the change in secondary metabolism of *Streptomyces venezuelae* in response to ethanol shock. Members of the *Streptomyces* genus are known for their rich repertoire of secondary metabolites, many of which are commercially or pharmaceutically relevant, such as the antibiotic chloramphenicol. As a result, they have also gained interest as production organisms in bioengineering. While it is known, that some of these secondary metabolites are produced in response to stressors, the regulatory mechanism and the extent of the capabilities of their secondary metabolism are not well understood.

To this end, the effects of ethanol shock on the transcriptome and secondary metabolome of *Streptomyces venezuelae* were investigated. For the analysis of the transcriptome, differential gene expression analysis was performed. To investigate the data also in the context of the metabolism, a metabolic network was reconstructed, which included a mapping of genes to enzymes to metabolic pathways. In combination with the differentially expressed genes, it enabled us to perform pathway enrichment analysis.

The combination of metabolome and transcriptome analysis, paired with the context of the metabolic reconstruction, showed a shift in secondary metabolism and parts of the primary metabolism of *Streptomyces venezuelae* upon ethanol shock. It also uncovered putative regulatory genes responsible for this shift, which are promising candidates for future research.

The analysis also revealed gaps in the reconstructed metabolic network, especially in the secondary metabolism of *S. venezuelae*. The presence of gaps was expected, both due to the nature of automated reconstructions, but also due to the known limited coverage of secondary metabolism in biochemical reaction databases [113]. Nevertheless, metabolic pathways of 9 secondary metabolite biosynthesis gene clusters were included in the reconstruction. Thus, despite being incomplete, the knowledge-base aspect of the metabolic reconstruction could be harnessed and proved useful for the analysis.

4.2. Publication status

Published work: Sekurova, O. N., Zehl, M., Predl, M., Hunyadi, P., Rattei, T., Zotchev, S. B. *Targeted Metabolomics and High-Throughput RNA Sequencing-Based Transcriptomics Reveal Massive Changes in the Streptomyces venezuelae NRRL B-65442 Metabolism Caused by Ethanol Shock*. Microbiol Spectr 10:e03672-22. DOI: <https://doi.org/10.1128/spectrum.03672-22>

This publication is included in the appendix, section A.1.

4.3. Author contributions

The author of this dissertation planned carried out the sequencing data analysis. He further planned and performed the construction of the metabolic network, as well as its integration and analysis with gene expression data. The author also produced figures and wrote parts of the manuscript.

4.4. Other manuscripts

4.4.1. Three conserved extracytoplasmic function sigma factors in *Streptomyces* have a major impact on secondary metabolism

Context

A follow up study of the effects of ethanol shock on the secondary metabolism of *Streptomyces venezuelae* was conducted. Here, the role of three regulatory genes was investigated, by analyzing their deletion mutants. The experimental setup largely resembles the first study, with the wildtype and mutant strains being subjected to ethanol shock or a mock treatment, then having transcriptome and secondary metabolome measured and analyzed. The transcriptome data is analyzed with differential gene expression analysis and pathway enrichment analysis, aided by the knowledge-base aspect of the metabolic reconstruction.

We found that the deletion of each of the three regulatory genes caused massive changes in gene expression. The mutants also had altered secondary metabolite profiles, with some of the compounds over expressed, compared to wildtype. These results are promising for the industrial production of secondary metabolites in *Streptomyces venezuelae*. As the three regulatory genes are conserved, similar results are expected in other members of the *Streptomyces* genus, as shown in the example of a marine sponge-derived *Streptomyces* species.

Publication status

Submitted for publication, under revision: Sekurova, O. N., Zehl, M., Predl, M., Hunyadi, P., Rattei, T., Zotchev, S. B., *Three conserved extracytoplasmic function sigma factors in Streptomyces have a major impact on secondary metabolism*. Submitted for publication in Microbial Cell Factories, confirmation of receipt: 30.03.2024

This publication is included in the appendix, section A.2.

Author contributions

The author of this dissertation planned carried out the sequencing data analysis. He further planned and performed the integration of the metabolic network with the gene expression data, as well as its analysis. The author also produced figures and wrote parts of the manuscript.

5. PyCoMo - generating and analyzing community metabolic models

5.1. Context

After the successful application of automatically generated metabolic networks, the next step was to model communities, instead of single organisms. As the focus of this work is the analysis of cross-feeding interactions, the compartmentalized community metabolic model structure is the best fit, since it explicitly includes these interactions. However, existing methods for generating such compartmentalized community metabolic models [7, 109] each face some of the following problems: (i) incomplete adherence to SBML standards [80], (ii) restrictions of metabolite and reaction namespaces in input models, (iii) availability of software (e.g. requirement of the proprietary software Matlab), or (iv) hard-coded implementation of linearization techniques. These issues restrict the construction and reuse of community metabolic models, and ultimately prohibit iterative curation.

To this end, we have developed PyCoMo, a Python package for generating compartmentalized community metabolic models, solving the mentioned limitations. (i) The generated models fully comply with the SBML standard. (ii) The input of single organism models is supported for a variety of file formats and tested for models of all major model generation pipelines and biochemical database namespaces. (iii) PyCoMo is freely available under an MIT license, does not require proprietary software, and is readily available on github, PyPI and bioconda [122]. (iv) PyCoMo uses a bound-free model structure to allow correct normalization of its members' reaction bounds. This allows for switching between linearization by fixed growth-rate, or fixed abundances.

PyCoMo also allows for efficient prediction of cross-feeding interactions. This task is usually accomplished by calculating feasible flux values (e.g. by FBA or FVA) and then extracting cross-fed metabolites. Since the linearization of the community metabolic model restricts this analysis to a specific growth-rate or abundance profile, the calculation needs to be repeated to search the entire parameter space. However, the bound-free model structure of PyCoMo allows to calculate all possible cross-feeding interactions in a single run of FVA. This is achieved by relaxing the constraint of balanced growth, which allows to search the entire solution space of the members' growth-rates and community compositions, but still correctly normalizes flux ranges to the community mass. Thus, an exhaustive search of cross-feeding potential is made readily available for community metabolic models.

Taken together, PyCoMo advances the methods toolbox for community metabolic modelling. The option to switch between linearization strategies, made possible by the bound-free model structure, allows for persistent community metabolic models, and thus enables an iterative curation process. PyCoMo also facilitates the exploration of interaction potential by efficient calculation of feasible cross-feeding interactions.

5.2. Publication status

Published work: Predl, M., Mießkes, M., Rattei, T., Zanghellini, J., *PyCoMo: a python package for community metabolic model creation and analysis*. Bioinformatics, Volume 40, Issue 4, April 2024, btae153, DOI: <https://doi.org/10.1093/bioinformatics/btae153>

This publication is included in the appendix, section A.3.

5.3. Author contributions

The author of this dissertation conceived of the study and software design, implemented the software, gathered and analyzed the test data, designed the figure and wrote the manuscript.

5.4. Further experiments

5.4.1. Introduction

PyCoMo is being actively maintained and new features have been included since its publication. A major problem in metabolic modelling is the presence of futile cycles [69, 123, 124]. These cycles are subnetworks that can carry flux without any input or output. While this flux does not influence the remaining network, it leads to inflated flux values in flux sampling and FVA. Typically, these cycles are deemed thermodynamically infeasible and can be broken up by corresponding methods.

In community metabolic models, there is, however, another reason for the presence of futile cycles. The conversion of two metabolites B and C can be reversible under natural conditions, and both B and C can be transported into and out of the cell. If two organisms in a community are able

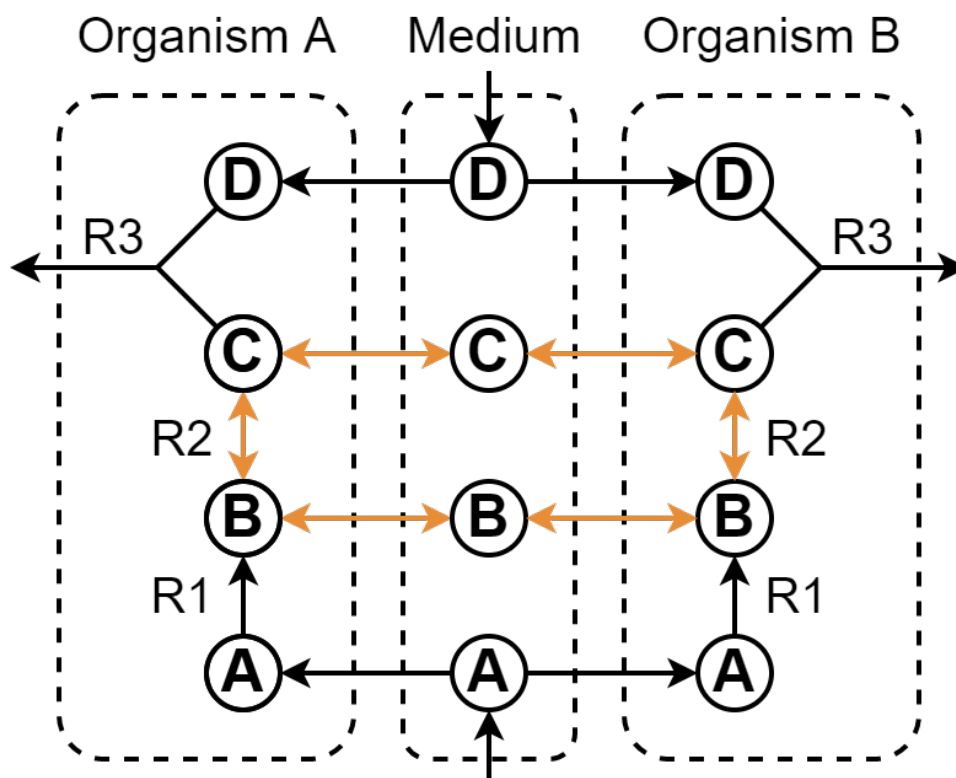


Figure 5.1.: A toy community metabolic modelling with a futile cycle. The model consists of two identical members. Each member requires metabolites C and D for biomass production. Metabolite C can be generated via B and A. Metabolites B and C can be taken up and secreted. Thus, the reversible conversion of B and C and their bidirectional transport forms a futile cycle (reactions shown in orange).

to transport and convert B and C, then a futile cycle is formed (see figure 5.1). While this circular metabolism is unlikely to be realized or biologically relevant, each of the reactions is part of valid metabolic capabilities. Thus, thermodynamic constraints cannot be used to deal with such futile cycles.

A different strategy, called the CycleFreeFlux algorithm, has been developed by Desouki et al. [124]. This method takes an existing flux solution (such as from FBA, or a step in FVA) and minimizes the flux through futile cycles in a second optimization step. To keep the relevant fluxes intact, the flux of all exchange reactions, the direction of the remaining reactions and the objective value are fixed. Then, the sum of reaction fluxes is used as the objective function and minimized with FBA. Due to the fixed metabolite exchange fluxes and reaction directions, the metabolic model can only reduce fluxes that are not relevant for these characteristics. The result is the removal of inflated flux values from futile cycles.

5.4.2. Methods and Results

A version of the CycleFreeFlux approach has been implemented in COBRApy, however, it is not compatible with the bound-free structure of models built by PyCoMo. Therefore, the algorithm was implemented directly in PyCoMo, with the following adaptations: First, the sink reactions of the lower and upper bound metabolites are not restricted, to keep the correct bound-free structure intact. Secondly, the original objective value is not fixed. This allows removing flux through futile cycles also in cases where reactions of such cycles are included in the objective function.

The size of community metabolic models results in many reactions to be analyzed with FVA. Ap-

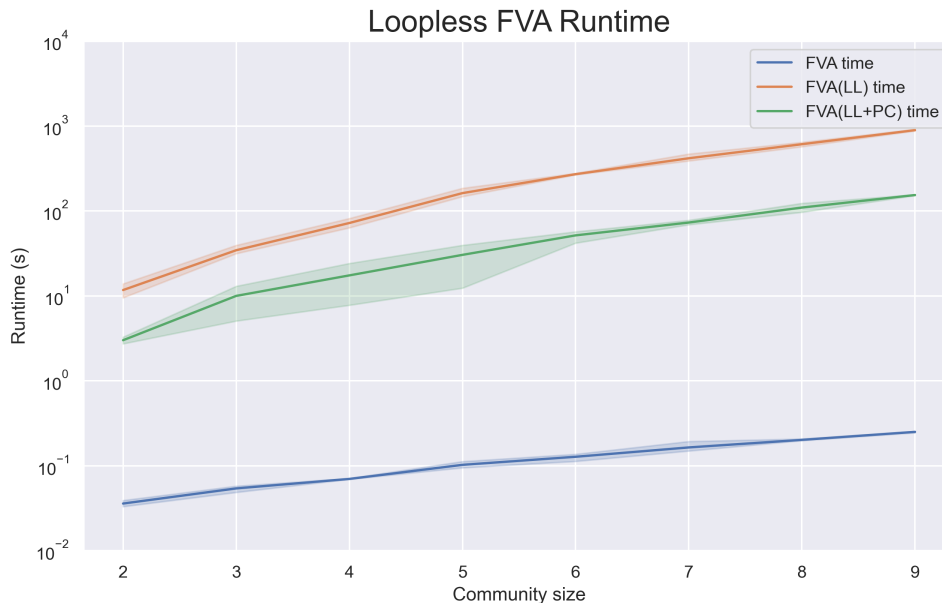


Figure 5.2.: Comparison of FVA runtimes. Abbreviations: FVA time - runtime of FVA without correction for futile cycles; FVA(LL) time - runtime of loopless FVA, correcting for futile cycles; FVA(LL+PC) time - runtime of loopless FVA, with precomputation of futile cycles, correcting for futile cycles only in reactions that are part of one. Communities of size two to nine members were generated by randomly selecting the respective number of models from the nine metabolic models of a biogas producing community [64] (with three replicates). The models are not genome-scale and as such only represent the trend in runtime, relative to the number of community members.

plying the CycleFreeFlux algorithm for each reaction leads to a large increase in runtime, compared to regular FVA. The PyCoMo implementation reduces this complexity by searching for cycle free solutions only in the reactions that are part of futile cycles. These reactions are identified in the solution of FVA without metabolite import, as reactions of futile cycles still carry flux. Despite the computational cost of the initial search step, this approach leads to drastic improvements in runtime, as can be seen in figure 5.2.

5.4.3. Discussion

Futile cycles are a major obstacle in the correct prediction of flux values and are likely to emerge in community metabolic models. The implementation of the CycleFreeFlux algorithm in PyCoMo solves this issue and implements a strategy to speed up the computation by precomputing reactions present in futile cycles. Ultimately, this achieves correct prediction of feasible and biologically relevant cross-feeding interactions.

6. ScyNet - visualizing interactions within community metabolic models

6.1. Context

The visualization of interactions of community members is another gap that was identified in currently available methods. Community metabolic models are highly complex and large, with thousands of reactions and metabolites. Nevertheless, the main mode of interacting with these models is via text, tables and programming interfaces. Of course, the network nature of the model is not captured by these representations. Visualizing the metabolic network would be favorable, and solutions have been developed for single organism metabolic models [125, 126].

In the case of smaller models it is possible to visualize the majority of the network in a manner that allows to gain understanding, as can be seen in the example of the *E. coli* core metabolism [125]. However, this is no longer possible for larger networks such as genome-scale metabolic models, where the visualized information needs to be filtered and reduced to be comprehensible. Common approaches are to select subnetworks by neighborhood [127] or pathways [128].

Community metabolic models face this problem on an increased level, as they are much larger and more complex than models of single organisms. While the existing reduction and abstraction approaches can be applied to them, none recover the community aspect, the interactions between the different community members. We hypothesized that a reduction of the community metabolic network to community members and exchanged metabolites would allow for exactly that.

This approach was implemented in ScyNet, an app for Cytoscape, a widely used network analysis framework [129]. After the import of community metabolic modes with Cy3sbml [130], ScyNet reduces the metabolic network to community members and external metabolites. Flux data can be used to contextualize the reduced network, showing the reaction directionality by arrow heads and color. ScyNet further includes a custom layout algorithm and allows for additional network reduction, showing only cross-feeding interactions.

The design of ScyNet visualizations was chosen under consideration of color-blindness. The color palette allows distinction of all colors in all three types of dichromacy (figure 6.1), and in the absence of coloring all elements can be clearly distinguished by shape.

ScyNet has been applied to several example communities of up to 17 members (results not shown). These examples revealed that with an increase in community size, even the extensive network reduction provided by ScyNet will result in visualizations, which are hard to interpret at a glance. For such communities, further reduction and abstraction will be needed. This could be achieved by clustering or abstraction of metabolites and organisms.

This work provides an important advancement for the visualization of cross-feeding interactions in community metabolic models.

6.2. Publication status

Submitted for publication: Predl, M., Gandolf, K., Hofer, M., Rattei, T., *ScyNet: Visualising interactions in community metabolic models*. Submitted for publication in Bioinformatics Advances, confirmation of receipt: 29.04.2024

This publication is included in the appendix, section A.4.

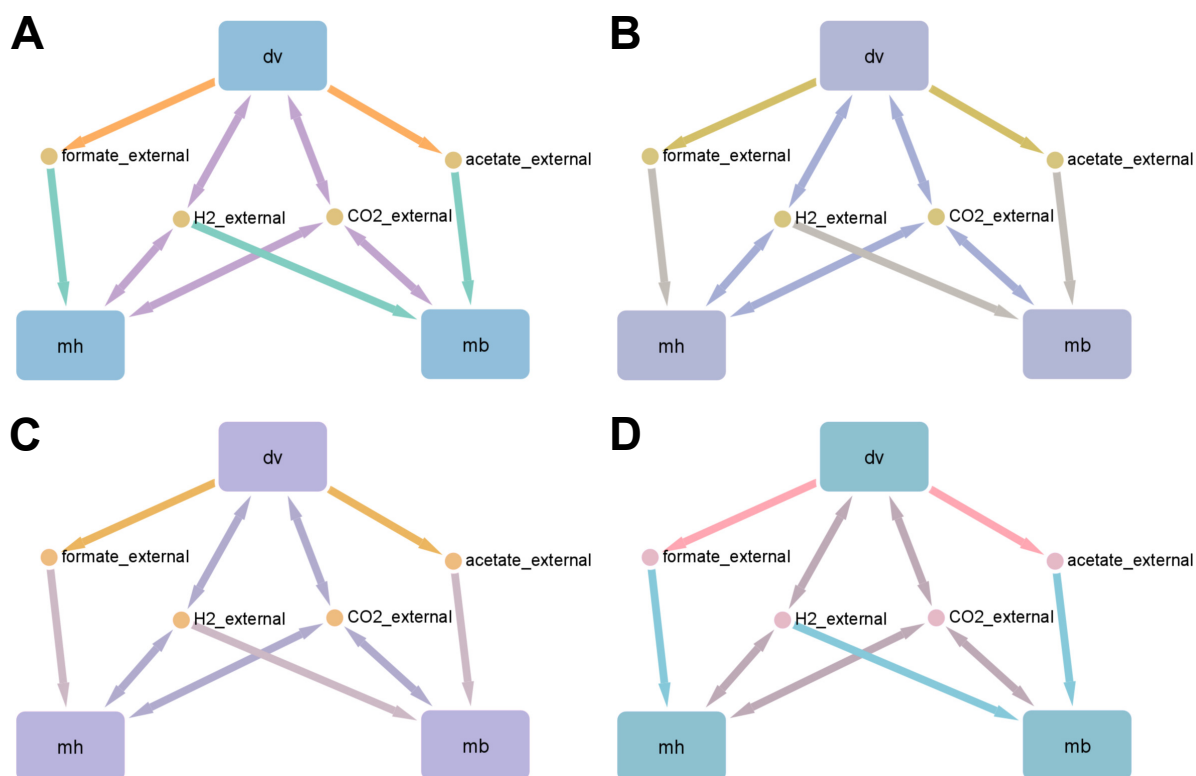


Figure 6.1.: Color-blind conscious design of ScyNet visualizations. The example shown is a three species biogas community, based on models of Koch et al. [64]. The community metabolic model was generated with PyCoMo. Flux ranges of an FVA result at optimal growth rate are visualized. (A) The figure as generated by ScyNet. The arrow direction and color indicates metabolite production (orange), consumption (green) or both (violet) by the community members. (B-D) Simulations of color-blindness using the Coblis webservice [131]. Three types of dichromacy are shown: Protanopia (B), Deuteranopia (C) and Tritanopia (D).

6.3. Author contributions

The author of this dissertation conceived of the study and software design, finished the implementation of the software based on a proof of concept version, gathered and analyzed the test data, designed the figure and wrote the manuscript.

7. Interactions in Archaea - Bacteria co-cultures

7.1. Context

With the methods for the generation of community metabolic models and the prediction of cross-feeding interactions in place, the next step was to apply them in biological research: Cultivation of archaea found in the human gastrointestinal tract (GIT).

The human GIT is colonized by diverse bacteria, archaea, viruses and fungi [132]. However, compared to bacteria, the archaea of the human gut microbiome are poorly studied, as most experimental procedures are designed for bacteria [133]. Consequentially, very few archaeal strains associated with human hosts are available as cultures, which in turn severely limits experimental research to understand these microbes. This project presents a cultivation strategy, which includes bacterial partners for the enrichment of archaea from human fecal samples.

Several examples have been found, where unculturable organisms could be cultured together with syntrophic partners [37]. In the case of *Methanobrevibacter*, a prominent archaeal genus in the human GIT, co-occurrence with specific bacteria, and increased (in vitro) growth rate with *Christensenella minuta* and *Bacteroides thetaiotamicron* could be shown [134–136]. These two bacterial species could be suitable syntrophic partners for the enrichment and culture of so far unculturable archaea.

To investigate potential cross-feeding interactions that could mediate the cultivation of *Methanobrevibacter*, community metabolic modelling was performed. The analysis started with the generation of metabolic models from the genomes of *Methanobrevibacter smithii*, *C. minuta* and *B. thetaiotamicron* using Gapseq [74]. Then, constituents of the complex growth media were estimated based on the protocol of Marinos et al. [118], and translated into a namespace matching the metabolic models. Gap-filling was used to allow growth of the models on the growth media. Finally, PyCoMo was used to generate pair-wise community metabolic models of *M. smithii* and a bacterial partner, and calculate potential cross-feeding interactions.

Analysis of the community metabolic models shows an increase in the predicted growth rate of *M. smithii* by up to 10-fold in co-culture, compared to single culture. These predictions are similar for both bacterial partners, who themselves are predicted to have no, or only a marginal increase in growth-rate. Further, several metabolites are predicted to be mediating cross-feeding interactions. Some of these interactions are known from literature, such as the supply of hydrogen and fermentation products like formate and acetate from *B. thetaiotamicron* to *M. smithii* [134, 136]. Other predicted cross-fed metabolites provided to *M. smithii* are sugars, amino acids and vitamin K2, for which some weaker evidence is present as well [136]. However, other proposed cross-fed metabolites are likely false positives, such as Coenzyme A and NAD being supplied by the bacteria, or hydrogen sulfide consumed by them. Based on these results, both bacteria seem equally capable partners for improving the growth of *Methanobrevibacter* species.

The enrichment protocol was then carried out to cultivate archaea from 16 human stool samples, providing either *B. thetaiotamicron*, *C. minuta*, or none as syntrophic partner. In 14 out of the 16 samples *Methanobrevibacter* could be enriched, which were identified as nine so far uncultivated strains. Interestingly, despite the evidence from co-culture experiments and the metabolic modelling predictions, *C. minuta* did not serve as a facilitator of *Methanobrevibacter* enrichment. Here, the reason can only be speculated, but it certainly shows that enrichment and co-cultivation are more complicated and cannot be reliably inferred by co-culture growth-rates alone.

This project shows, that metabolic modelling can lead to informative insights in cross-feeding interactions. The predicted metabolites likely contain false positives, but a number of literature validated and additional promising candidates were found. While experimental validation is needed, this points in the direction, that high-quality candidates for cross-fed metabolites can be found using metabolic models, even when they are automatically generated and of non-model organisms. It also suggests that genome-guided cultivation, the optimization of growth media, or finding potential partners for enrichment by co-cultivation are possible with the aid of metabolic modelling.

7.2. Publication status

Preprint publication: Duller, S., Vrbancic, S., Szydlowski, L., Mahnert, A., Blohs, A., Predl, M., Kumpitsch, C., Zrim, V., Hoegenauer, C., Kosciolk, T., Schmitz, R. A., Eberhard, A., Dragovan, M., Schmidberger, L., Zurabishvili, T., Weinberger, V., Moser, A. M., Kolb, D., Pernitsch, D., Mohammadzadeh, R., Kuehnast, T., Rattei, T., Moissl-Eichinger, C., *Expanding the cultivated human archaeome by targeted isolation of novel Methanobrevibacter strains from fecal samples*. bioRxiv 2024.04.10.588852; DOI: <https://doi.org/10.1101/2024.04.10.588852>

Submitted for publication: Duller, S., Vrbancic, S., Szydlowski, L., Mahnert, A., Blohs, A., Predl, M., Kumpitsch, C., Zrim, V., Hoegenauer, C., Kosciolk, T., Schmitz, R. A., Eberhard, A., Dragovan, M., Schmidberger, L., Zurabishvili, T., Weinberger, V., Moser, A. M., Kolb, D., Pernitsch, D., Mohammadzadeh, R., Kuehnast, T., Rattei, T., Moissl-Eichinger, C., *Expanding the cultivated human archaeome by targeted isolation of novel Methanobrevibacter strains from fecal samples*. Submitted for publication in Nature Communications, confirmation of receipt: 11.04.2024

This publication is included in the appendix, section A.5.

7.3. Author contributions

The author of this dissertation planned and carried out the metabolic modelling of pairwise co-cultures. He also wrote part of the manuscript.

8. Discussion and Conclusion

This work aims to advance the prediction of cross-feeding interactions in microbial communities by using metabolic modelling.

In chapter 4, the use of the knowledge-base aspect of metabolic models was investigated. A metabolic model was reconstructed for the non-model organism *S. venezuelae* and applied to the analysis of transcriptome data. This allowed putting the differentially expressed genes in the context of metabolic pathways, as well as analyzing them with pathway enrichment analysis. While the annotation of some secondary metabolic pathways was missing from the metabolic network, it added valuable information to the genes in the pathways that were included. As such, even in incomplete, automatically reconstructed metabolic models of a non-model organisms, the knowledge-base aspect can be utilized for analysis.

This aspect could still be further improved. An interesting approach would be further integration of differential gene expression analysis with metabolic networks, by searching for differentially expressed metabolic subnetworks.

Next, the scope is expanded from metabolic models of single organisms, to those of communities. Two gaps in the currently available methods were identified and tackled: The generation of community metabolic models and their visualization.

The PyCoMo package brings important advances to the field, by creating a model structure which allows switching between different linearization methods. The persistence of community metabolic models, paired with their reusability and compliance to the SBML community standards open the way for iterative curation in community metabolic models. Another important advancement is brought by the efficient calculation of all potential cross-feeding interactions, making this community property easily accessible to future research.

An important point for further improvements is additional validation. Given data are available, extending the functionality of PyCoMo and its model structure in the direction of OMICs data integration and dynamic flux balance analysis seems promising. Tools for such analyses already exist for single organism metabolic models, it could be sufficient to provide a suitable interface to models generated by PyCoMo.

There are also some interesting developments regarding model persistence and clearer compartmentalization of members, made available through the Hierarchical Model Composition package of SBML [137, 138]. Combining such an approach with the bound-free model structure would improve the ease of model curation, which is needed for the long term study and engineering of microbial communities.

The second software package, ScyNet, tackles the visualization of community metabolic models. Its network reduction algorithm allows for simple and quick visualization of interactions between community members. Such visualizations are crucial for understanding communities on a network level. This is much simpler by viewing a reduced network, than from studying spreadsheets or large complex hairball-like networks. Additionally, the figures created by ScyNet can be used for publications and presentations, facilitating the communication of results in community metabolic modelling research.

There are several future directions for improvement of ScyNet and the visualization of interactions within communities. To create informative figures for larger communities, further reduction is needed. This could be achieved by grouping or abstraction of community members and metabolites. Incorporation of additional data types, such as flux distributions, OMICs data, or a comparison

between metabolic states would open up ScyNet to new types of analyses. Also, including a selection of internal fluxes from community members would be useful for connecting intracellular metabolism and community phenotypes. Lastly, automation of ScyNet visualization via Python access of Cytoscape can be implemented and would allow the creation of bioinformatics pipelines from genomes to visualized community metabolic models.

In chapter 7, community metabolic modelling was applied to co-cultures of archaea and bacteria. The metabolic models showed intriguing predictions: a growth boost for archaea in co-culture with bacteria, as well as the prediction of cross-fed metabolites, some being known from literature, others proposed and or newly predicted. As such, the model provides a list of cross-feeding interaction candidates for further experiments. The predicted boost in growth-rate was also observed in previous experiments [134, 136], however, the enrichment of *Methanobrevibacter* was only successful with one of the bacterial partners. Further investigation of this phenomenon could be aided by metabolic modelling as well, for example by refining the community metabolic model with additional data such as metatranscriptome or metabolome.

The project showcases an important application community metabolic modelling: Genome-guided cultivation. Many such efforts have already been done [139]. However, this work shows the potential benefit of including the community aspect. Since many unculturable taxa naturally occur in communities, it is proposed that cultivation of these organisms requires substrates or growth conditions provided by their community members [37]. This project shows that PyCoMo can provide information on potential cross-feeding interactions, which are promising candidates for media design in cultivation efforts.

Metabolic modelling of communities still faces many challenges, such as model completeness, metabolite transport prediction, estimating reaction parameters or the incorporation of regulatory mechanisms. As these issues are under active research, future developments will likely improve all facets of model quality.

An important aspect for future research is the generation of validation datasets. They would allow for rigorous testing of community metabolic models and methods, making it possible to evaluate effective modelling strategies and make informed decisions. As establishing suitable datasets or community competitions has proven to be a successful way to assess the current status and push improvement of methods in other areas of bioinformatics [140, 141], metabolic modelling would benefit from such an effort as well.

Iterative curation has been a crucial strategy for the creation of high-quality metabolic models. Such a feat has not yet been achieved for the study of communities and would be promising. It is clear that generating experimental data is laborious. Including metabolic modelling in the scope of analysis early on in the study design can allow many of the experiments done for "biological research" to be used for modelling the target community as well [12]. On the other hand, iterative curation of community metabolic models was also restricted by their reusability, as changing the type of analysis would require creating a new model. This issue is now solved with the model structure of PyCoMo, taking a crucial step forward.

Taken together, the presented work has advanced the prediction of cross-feeding interactions in microbial communities by using metabolic modelling, and thus achieved its aim.

Multiple directions for further research have been proposed, but two areas seem especially beneficial: Launching long term refinement projects for metabolic models of select "model" communities, and gathering data for validating and evaluating the tools, strategies and choices in metabolic modelling.

Already in its current state, metabolic modelling shows promising applications in the study and design of communities and genome guided cultivation.

Bibliography

- [1] James T. Stroud et al. 'Is a community still a community? Reviewing definitions of key terms in community ecology'. In: *Ecol Evol* 5.21 (2015), pp. 4757–4765. doi: 10.1002/ece3.1651.
- [2] Teddy Namirimu et al. 'Microbial Community Structure and Functional Potential of Deep-Sea Sediments on Low Activity Hydrothermal Area in the Central Indian Ridge'. English. In: *Front. Mar. Sci.* 9 (2022). doi: 10.3389/fmars.2022.784807.
- [3] Stefanie Imminger et al. 'Survival and rapid resuscitation permit limited productivity in desert microbial communities'. en. In: *Nat Commun* 15.1 (2024), p. 3056. doi: 10.1038/s41467-024-46920-6.
- [4] Emily Lou LaMartina, Aurash A. Mohaimani and Ryan J. Newton. 'Urban wastewater bacterial communities assemble into seasonal steady states'. In: *Microbiome* 9.1 (2021), p. 116. doi: 10.1186/s40168-021-01038-5.
- [5] Yong Fan and Oluf Pedersen. 'Gut microbiota in human metabolic health and disease'. en. In: *Nat Rev Microbiol* 19.1 (2021), pp. 55–71. doi: 10.1038/s41579-020-0433-9.
- [6] Ambihai Shayanthan, Patricia Ann C. Ordoñez and Ivan John Oresnik. 'The Role of Synthetic Microbial Communities (SynCom) in Sustainable Agriculture'. English. In: *Front. Agron.* 4 (2022). doi: 10.3389/fagro.2022.896307.
- [7] Pavlos Stephanos Bekiaris and Steffen Klamt. 'Designing microbial communities to maximize the thermodynamic driving force for the production of chemicals'. en. In: *PLOS Computational Biology* 17.6 (2021), e1009093. doi: 10.1371/journal.pcbi.1009093.
- [8] Hiral G. Chaudhari et al. 'Decoding the microbial universe with metagenomics: a brief insight'. In: *Front. Genet.* 14 (24th Apr. 2023). doi: 10.3389/fgene.2023.1119740.
- [9] Ana Elena Pérez-Cobas, Laura Gomez-Valero and Carmen Buchrieser. 'Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses'. In: *Microbial Genomics* 6.8 (2020), e000409. doi: 10.1099/mgen.0.000409.
- [10] Natalie Becker et al. 'Human intestinal microbiota: Characterization of a simplified and stable gnotobiotic rat model'. In: *Gut Microbes* 2.1 (2011). doi: 10.4161/gmic.2.1.14651.
- [11] Sandrine Brugiroux et al. 'Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium'. In: *Nature Microbiology* 2.November 2016 (2016). doi: 10.1038/nmicrobiol.2016.215.
- [12] Kevin D'hoel et al. 'Integrated culturing, modeling and transcriptomics uncovers complex interactions and emergent behavior in a three-species synthetic gut community'. In: *eLife* 7 (2018). Ed. by Xochitl Morgan and Wendy S Garrett, e37090. doi: 10.7554/eLife.37090.
- [13] Anelize Bauermeister et al. 'Mass spectrometry-based metabolomics in microbiome investigations'. In: *Nat Rev Microbiol* 20.3 (Mar. 2022), pp. 143–160. doi: 10.1038/s41579-021-00621-9.
- [14] Migun Shakya, Chien Chi Lo and Patrick S.G. Chain. 'Advances and challenges in meta-transcriptomic analysis'. In: *Frontiers in Genetics* 10 (SEP 2019), pp. 1–10. doi: 10.3389/fgene.2019.00904.
- [15] Otto X Cordero and Manoshi S Datta. 'Microbial interactions and community assembly at microscales'. In: *Current Opinion in Microbiology*. Environmental microbiology * Special Section: Megaviromes 31 (1st June 2016), pp. 227–234. doi: 10.1016/j.mib.2016.03.015.

- [16] Ana Barbosa et al. 'Imaging biofilms using fluorescence in situ hybridization: seeing is believing'. In: *Front. Cell. Infect. Microbiol.* 13 (22nd May 2023). doi: 10.3389/fcimb.2023.1195803.
- [17] Allan Konopka. 'What is microbial community ecology?' In: *ISME J* 3.11 (Nov. 2009), pp. 1223–1230. doi: 10.1038/ismej.2009.88.
- [18] Philippe C. Baveye et al. 'Emergent Properties of Microbial Activity in Heterogeneous Soil Microenvironments: Different Research Approaches Are Slowly Converging, Yet Major Challenges Remain'. In: *Front. Microbiol.* 9 (27th Aug. 2018). doi: 10.3389/fmicb.2018.01929.
- [19] Naomi Iris van den Berg et al. 'Ecological modelling approaches for predicting emergent properties in microbial communities'. In: *Nat Ecol Evol* 6.7 (July 2022), pp. 855–865. doi: 10.1038/s41559-022-01746-7.
- [20] Octavio Perez-Garcia, Gavin Lear and Naresh Singhal. 'Metabolic Network Modeling of Microbial Interactions in Natural and Engineered Environmental Systems'. In: *Frontiers in Microbiology* 7 (2016).
- [21] Cristal Zuñiga, Livia Zaramela and Karsten Zengler. 'Elucidation of complexity and prediction of interactions in microbial communities'. In: *Microbial Biotechnology* 10.6 (2017), pp. 1500–1522. doi: 10.1111/1751-7915.12855.
- [22] Beatriz García-Jiménez, Jesús Torres-Bacete and Juan Nogales. 'Metabolic modelling approaches for describing and engineering microbial communities'. In: *Computational and Structural Biotechnology Journal* 19 (1st Jan. 2021), pp. 226–246. doi: 10.1016/j.csbj.2020.12.003.
- [23] Sammy Pontrelli et al. 'Metabolic cross-feeding structures the assembly of polysaccharide degrading communities'. In: *Science Advances* 8.8 (23rd Feb. 2022), eabk3076. doi: 10.1126/sciadv.abk3076.
- [24] Tiroyaone Shimane Tshikantwa et al. 'Current Trends and Potential Applications of Microbial Interactions for Human Welfare'. In: *Front. Microbiol.* 9 (1st June 2018). doi: 10.3389/fmicb.2018.01156.
- [25] Nick W. Smith et al. 'The Classification and Evolution of Bacterial Cross-Feeding'. In: *Front. Ecol. Evol.* 7 (14th May 2019). doi: 10.3389/fevo.2019.00153.
- [26] Tobias Großkopf and Orkun S Soyer. 'Synthetic microbial communities'. In: *Current Opinion in Microbiology*. Cell regulation 18 (1st Apr. 2014), pp. 72–77. doi: 10.1016/j.mib.2014.02.002.
- [27] Matti Gralka et al. 'Trophic Interactions and the Drivers of Microbial Community Assembly'. In: *Current Biology* 30.19 (5th Oct. 2020), R1176–R1188. doi: 10.1016/j.cub.2020.08.007.
- [28] J. Jeffrey Morris. 'Black Queen evolution: The role of leakiness in structuring microbial communities'. In: *Trends in Genetics* 31.8 (2015), pp. 475–482. doi: 10.1016/j.tig.2015.05.004.
- [29] Philipp Wendering and Zoran Nikoloski. 'COMMIT: Consideration of metabolite leakage and community composition improves microbial community reconstructions'. In: *PLoS Comput Biol* 18.3 (Mar. 2022), e1009906. doi: 10.1371/journal.pcbi.1009906.
- [30] James B. McKinlay. 'Are Bacteria Leaky? Mechanisms of Metabolite Externalization in Bacterial Cross-Feeding'. In: *Annual Review of Microbiology* 77 (Volume 77, 2023 15th Sept. 2023), pp. 277–297. doi: 10.1146/annurev-micro-032521-023815.
- [31] Alan R. Pacheco, Mauricio Moel and Daniel Segrè. 'Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems'. In: *Nature Communications* 10.1 (2019). doi: 10.1038/s41467-018-07946-9.

- [32] Tobias Großkopf et al. 'Metabolic Modelling in a Dynamic Evolutionary Framework Predicts Adaptive Diversification of Bacteria in a Long-Term Evolution Experiment'. In: *BMC Evolutionary Biology* 16.1 (20th Aug. 2016), p. 163. doi: 10.1186/s12862-016-0733-x.
- [33] Richard E. Lenski. 'Experimental Evolution and the Dynamics of Adaptation and Genome Evolution in Microbial Populations'. In: *The ISME Journal* 11.10 (Oct. 2017), pp. 2181–2194. doi: 10.1038/ismej.2017.69.
- [34] Sriram Varahan et al. 'Metabolite plasticity drives carbon-nitrogen resource budgeting to enable division of labor in a clonal community'. In: *bioRxiv* (2020), pp. 1–26. doi: 10.1101/2020.04.09.033571.
- [35] J. Jeffrey Morris, Richard E. Lenski and Erik R. Zinser. 'The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss'. In: *mBio* 3.2 (23rd Mar. 2012), 10.1128/mbio.00036–12. doi: 10.1128/mbio.00036-12.
- [36] Kevin R. Foster and Thomas Bell. 'Competition, Not Cooperation, Dominates Interactions among Culturable Microbial Species'. In: *Current Biology* 22.19 (9th Oct. 2012), pp. 1845–1850. doi: 10.1016/j.cub.2012.08.005.
- [37] Samay Pande and Christian Kost. 'Bacterial Unculturability and the Formation of Intercellular Metabolic Networks'. In: *Trends in Microbiology* 25.5 (1st May 2017), pp. 349–361. doi: 10.1016/j.tim.2017.02.015.
- [38] Eric J. Stewart. 'Growing Unculturable Bacteria'. In: *Journal of Bacteriology* 194.16 (27th July 2012), pp. 4151–4160. doi: 10.1128/jb.00345-12.
- [39] Glen D'Souza et al. 'LESS IS MORE: SELECTIVE ADVANTAGES CAN EXPLAIN THE PREVALENT LOSS OF BIOSYNTHETIC GENES IN BACTERIA'. In: *Evolution* 68.9 (1st Sept. 2014), pp. 2559–2570. doi: 10.1111/evo.12468.
- [40] Glen D'Souza et al. 'Plasticity and epistasis strongly affect bacterial fitness after losing multiple metabolic genes'. In: *Evolution* 69.5 (1st May 2015), pp. 1244–1254. doi: 10.1111/evo.12640.
- [41] Sebastian Behrens et al. 'Linking Microbial Phylogeny to Metabolic Activity at the Single-Cell Level by Using Enhanced Element Labeling-Catalyzed Reporter Deposition Fluorescence In Situ Hybridization (EL-FISH) and NanoSIMS'. In: *Applied and Environmental Microbiology* 74.10 (15th May 2008), pp. 3143–3150. doi: 10.1128/AEM.00191-08.
- [42] Roland Hatzenpichler et al. 'Next-Generation Physiology Approaches to Study Microbiome Function at Single Cell Level'. In: *Nature Reviews Microbiology* 18.4 (Apr. 2020), pp. 241–256. doi: 10.1038/s41579-020-0323-1.
- [43] Uria Alcolombri et al. 'Single-Cell Stable Isotope Probing in Microbial Ecology'. In: *ISME Communications* 2.1 (1st Dec. 2022), p. 55. doi: 10.1038/s43705-022-00142-3.
- [44] Maria Mooshammer et al. 'Flow-through stable isotope probing (Flow-SIP) minimizes cross-feeding in complex microbial communities'. In: *ISME J* 15.1 (Jan. 2021), pp. 348–353. doi: 10.1038/s41396-020-00761-5.
- [45] Seth Rakoff-Nahoum, Michael J. Coyne and Laurie E. Comstock. 'An Ecological Network of Polysaccharide Utilization among Human Intestinal Symbionts'. In: *Current Biology* 24.1 (6th Jan. 2014), pp. 40–49. doi: 10.1016/j.cub.2013.10.077. pmid: 24332541.
- [46] Alexandra L. McCully et al. 'An Escherichia Coli Nitrogen Starvation Response Is Important for Mutualistic Coexistence with Rhodospseudomonas Palustris'. In: *Applied and Environmental Microbiology* 84.14 (2nd July 2018), e00404–18. doi: 10.1128/AEM.00404-18.
- [47] Sudarshan A. Shetty et al. 'Inter-Species Metabolic Interactions in an In-vitro Minimal Human Gut Microbiome of Core Bacteria'. In: *npj Biofilms and Microbiomes* 8.1 (1 8th Apr. 2022), pp. 1–13. doi: 10.1038/s41522-022-00275-2.

- [48] Alexandra C. Schrimpe-Rutledge et al. 'Untargeted Metabolomics Strategies – Challenges and Emerging Directions'. In: *Journal of the American Society for Mass Spectrometry* 27.12 (Dec. 2016), pp. 1897–1905. doi: 10.1007/s13361-016-1469-y. pmid: 27624161.
- [49] Miriam Sindelar and Gary J. Patti. 'Chemical Discovery in the Era of Metabolomics'. In: *Journal of the American Chemical Society* 142.20 (20th May 2020), pp. 9097–9105. doi: 10.1021/jacs.9b13198. pmid: 32275430.
- [50] William Kumler, Bryna J. Hazelton and Anitra E. Ingalls. 'Picky with Peakpicking: Assessing Chromatographic Peak Quality with Simple Metrics in Metabolomics'. In: *BMC Bioinformatics* 24.1 (28th Oct. 2023), p. 404. doi: 10.1186/s12859-023-05533-4.
- [51] Pau Pérez Escriba, Tobias Fuhrer and Uwe Sauer. 'Distinct N and C Cross-Feeding Networks in a Synthetic Mouse Gut Consortium'. In: *mSystems* 7.2 (31st Mar. 2022), e01484–21. doi: 10.1128/msystems.01484-21.
- [52] Zachary D. Kurtz et al. 'Sparse and Compositionally Robust Inference of Microbial Ecological Networks'. In: *PLOS Computational Biology* 11.5 (7th May 2015), e1004226. doi: 10.1371/journal.pcbi.1004226.
- [53] Lindsey M. Solden et al. 'Interspecies Cross-Feeding Orchestrates Carbon Degradation in the Rumen Ecosystem'. In: *Nature Microbiology* 3.11 (Nov. 2018), pp. 1274–1284. doi: 10.1038/s41564-018-0225-4.
- [54] Him K. Shrestha et al. 'Metaproteomics Reveals Insights into Microbial Structure, Interactions, and Dynamic Regulation in Defined Communities as They Respond to Environmental Disturbance'. In: *BMC Microbiology* 21.1 (8th Nov. 2021), p. 308. doi: 10.1186/s12866-021-02370-4.
- [55] Akshit Goyal et al. 'Ecology-Guided Prediction of Cross-Feeding Interactions in the Human Gut Microbiome'. In: *Nature Communications* 12.1 (26th Feb. 2021), p. 1335. doi: 10.1038/s41467-021-21586-6.
- [56] Tong Wang et al. 'Predicting Metabolomic Profiles from Microbial Composition through Neural Ordinary Differential Equations'. In: *Nature machine intelligence* 5.3 (Mar. 2023), pp. 284–293. doi: 10.1038/s42256-023-00627-3. pmid: 38223254.
- [57] Robert D. Fleischmann et al. 'Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd'. In: *Science* 269.5223 (1995), pp. 496–512. doi: 10.1126/science.7542800.
- [58] Changdai Gu et al. 'Current status and applications of genome-scale metabolic models'. In: *Genome Biology* 20.1 (2019), pp. 1–18. doi: 10.1186/s13059-019-1730-3.
- [59] Won Jun Kim, Hyun Uk Kim and Sang Yup Lee. 'Current state and applications of microbial genome-scale metabolic models'. In: *Current Opinion in Systems Biology*. Regulatory and metabolic networks • Cancer and systemic diseases 2 (1st Apr. 2017), pp. 10–18. doi: 10.1016/j.coisb.2017.03.001.
- [60] Byoungjin Kim et al. 'Applications of genome-scale metabolic network model in metabolic engineering'. In: *Journal of Industrial Microbiology and Biotechnology* 42.3 (1st Mar. 2015), pp. 339–348. doi: 10.1007/s10295-014-1554-9.
- [61] Aarash Bordbar et al. 'Personalized Whole-Cell Kinetic Models of Metabolism for Discovery in Genomics and Pharmacodynamics'. In: *Cell Systems* 1.4 (28th Oct. 2015), pp. 283–292. doi: 10.1016/j.cels.2015.10.003.
- [62] Pouyan Ghaffari, Adil Mardinoglu and Jens Nielsen. 'Cancer Metabolism: A Modeling Perspective'. In: *Frontiers in Physiology* 6 (2015).
- [63] Elizabeth Brunk et al. 'Recon3D enables a three-dimensional view of gene variation in human metabolism'. In: *Nat Biotechnol* 36.3 (Mar. 2018), pp. 272–281. doi: 10.1038/nbt.4072.

- [64] Sabine Koch et al. 'RedCom: A strategy for reduced metabolic modeling of complex microbial communities and its application for analyzing experimental datasets from anaerobic digestion'. In: *PLOS Computational Biology* 15 (2019). doi: 10.1371/journal.pcbi.1006759.
- [65] Christian Diener, Sean M. Gibbons and Osbaldo Resendis-Antonio. 'MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota'. In: *mSystems* 5.1 (2020), pp. 1–17. doi: 10.1128/msystems.00606-19.
- [66] Aarash Bordbar et al. 'Constraint-based models predict metabolic and associated cellular functions'. In: *Nat Rev Genet* 15.2 (Feb. 2014), pp. 107–120. doi: 10.1038/nrg3643.
- [67] Oliver Hädicke and Steffen Klamt. 'EColiCore2: a reference network model of the central metabolism of Escherichia coli and relationships to its genome-scale parent model'. In: *Sci Rep* 7.1 (3rd Jan. 2017), p. 39647. doi: 10.1038/srep39647.
- [68] Sebastián N. Mendoza et al. 'A systematic assessment of current genome-scale metabolic reconstruction tools'. In: *bioRxiv* 20.1 (2019), pp. 1–20. doi: 10.1101/558411.
- [69] Ines Thiele and Bernhard Palsson. 'A protocol for generating a high-quality genome-scale metabolic reconstruction'. In: *Nature Protocols* 5.1 (2010), pp. 93–121. doi: 10.1038/nprot.2009.203.
- [70] Minoru Kanehisa et al. 'KEGG: New perspectives on genomes, pathways, diseases and drugs'. In: *Nucleic Acids Research* 45 (D1 2017), pp. D353–D361. doi: 10.1093/nar/gkw1092.
- [71] Samuel M.D. Seaver et al. 'The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes'. In: *Nucleic Acids Research* 49 (D1 2021), pp. D575–D588. doi: 10.1093/nar/gkaa746.
- [72] Ron Caspi et al. 'The MetaCyc database of metabolic pathways and enzymes - a 2019 update'. In: *Nucleic Acids Research* 48 (D1 8th Jan. 2020), pp. D445–D453. doi: 10.1093/nar/gkz862.
- [73] Ines Thiele, Nikos Vlassis and Ronan M.T. Fleming. 'FASTGAPFILL: Efficient gap filling in metabolic networks'. In: *Bioinformatics* 30.17 (2014), pp. 2529–2531. doi: 10.1093/bioinformatics/btu321.
- [74] Johannes Zimmermann, Christoph Kaleta and Silvio Waschina. 'gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models'. In: *Genome Biology* 22.1 (2021), p. 81. doi: 10.1186/s13059-021-02295-1.
- [75] Wai Kit Ong, Peter E Midford and Peter D Karp. 'Taxonomic weighting improves the accuracy of a gap-filling algorithm for metabolic models'. In: *Bioinformatics* 36.6 (15th Mar. 2020), pp. 1823–1830. doi: 10.1093/bioinformatics/btz813.
- [76] Daniel Machado et al. 'Fast automated reconstruction of genome-scale metabolic models for microbial species and communities'. In: *Nucleic Acids Research* 46.15 (2018), pp. 7542–7553. doi: 10.1093/nar/gky537.
- [77] Stefánía Magnúsdóttir et al. 'Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota'. In: *Nature Biotechnology* 35.1 (2017), pp. 81–89. doi: 10.1038/nbt.3703.
- [78] Jonathan M. Monk et al. 'iML1515, a knowledgebase that computes Escherichia coli traits'. In: *Nat Biotechnol* 35.10 (Oct. 2017), pp. 904–908. doi: 10.1038/nbt.3956.
- [79] Omid Oftadeh et al. 'A genome-scale metabolic model of Saccharomyces cerevisiae that integrates expression constraints and reaction thermodynamics'. In: *Nat Commun* 12.1 (9th Aug. 2021), p. 4790. doi: 10.1038/s41467-021-25158-6.

- [80] Sarah M Keating et al. 'SBML Level 3: an extensible format for the exchange and reuse of biological models'. In: *Molecular Systems Biology* 16.8 (Aug. 2020), e9110. doi: 10.15252/msb.20199110.
- [81] Laurent Heirendt et al. 'Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0'. In: *Nature Protocols* 14.3 (2019), pp. 639–702. doi: 10.1038/s41596-018-0098-2.
- [82] Gabriel Gelius-Dietrich et al. 'sybil – Efficient constraint-based modelling in R'. en. In: *BMC Syst Biol* 7.1 (2013), p. 125. doi: 10.1186/1752-0509-7-125.
- [83] Ali Ebrahim et al. 'COBRApy: CONstraints-Based Reconstruction and Analysis for Python'. In: *BMC Systems Biology* 7 (2013). doi: 10.1186/1752-0509-7-74.
- [84] Jose A. Gomez, Kai Höffner and Paul I. Barton. 'DFBALab: a fast and reliable MATLAB code for dynamic flux balance analysis'. In: *BMC Bioinformatics* 15.1 (18th Dec. 2014), p. 409. doi: 10.1186/s12859-014-0409-8.
- [85] Alexandra-M. Reimers and Arne C. Reimers. 'The steady-state assumption in oscillating and growing systems'. In: *Journal of Theoretical Biology* 406 (2016), pp. 176–186. doi: 10.1016/j.jtbi.2016.06.031.
- [86] Jeffrey D. Orth, Ines Thiele and Bernhard Ø Palsson. 'What is flux balance analysis?' en. In: *Nat Biotechnol* 28.3 (2010), pp. 245–248. doi: 10.1038/nbt.1614.
- [87] Nathan E Lewis et al. 'Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models'. In: *Molecular Systems Biology* 6.1 (Jan. 2010), p. 390. doi: 10.1038/msb.2010.47.
- [88] Seirana Hashemi, Zahra Razaghi-Moghadam and Zoran Nikoloski. 'Maximizing multi-reaction dependencies provides more accurate and precise predictions of intracellular fluxes than the principle of parsimony'. In: *PLOS Computational Biology* 19.9 (18th Sept. 2023), e1011489. doi: 10.1371/journal.pcbi.1011489.
- [89] David B. Bernstein et al. 'Addressing uncertainty in genome-scale metabolic model reconstruction and analysis'. In: *Genome Biology* 22.1 (2021), pp. 1–22. doi: 10.1186/s13059-021-02289-z.
- [90] Steinn Gudmundsson and Ines Thiele. 'Computationally efficient flux variability analysis'. In: *BMC Bioinformatics* 11.1 (29th Sept. 2010), p. 489. doi: 10.1186/1471-2105-11-489.
- [91] Hulda S Haraldsdóttir et al. 'CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models'. In: *Bioinformatics* 33.11 (1st June 2017), pp. 1741–1743. doi: 10.1093/bioinformatics/btx052.
- [92] Pedro A. Saa et al. 'LooplessFluxSampler: an efficient toolbox for sampling the loopless flux solution space of metabolic models'. In: *BMC Bioinformatics* 25.1 (2nd Jan. 2024), p. 3. doi: 10.1186/s12859-023-05616-2.
- [93] Helena A. Herrmann et al. 'Flux sampling is a powerful tool to study metabolism under changing environmental conditions'. In: *npj Syst Biol Appl* 5.1 (2nd Sept. 2019), pp. 1–8. doi: 10.1038/s41540-019-0109-0.
- [94] Mohammad Mazharul Islam, Wheaton Lane Schroeder and Rajib Saha. 'Kinetic modeling of metabolism: Present and future'. In: *Current Opinion in Systems Biology* 26 (1st June 2021), pp. 72–78. doi: 10.1016/j.coisb.2021.04.003.
- [95] Osvaldo D. Kim, Miguel Rocha and Paulo Maia. 'A Review of Dynamic Modeling Approaches and Their Application in Computational Strain Optimization for Metabolic Engineering'. In: *Front. Microbiol.* 9 (31st July 2018). doi: 10.3389/fmicb.2018.01690.

-
- [96] Alexander Kroll et al. 'Deep learning allows genome-scale prediction of Michaelis constants from structural features'. In: *PLOS Biology* 19.10 (19th Oct. 2021), e3001402. doi: 10.1371/journal.pbio.3001402.
 - [97] Milenko Tokic, Vassily Hatzimanikatis and Ljubisa Miskovic. 'Large-scale kinetic metabolic models of *Pseudomonas putida* KT2440 for consistent design of metabolic engineering strategies'. In: *Biotechnol Biofuels* 13.1 (28th Feb. 2020), p. 33. doi: 10.1186/s13068-020-1665-7.
 - [98] Radhakrishnan Mahadevan, Jeremy S. Edwards and Francis J. Doyle. 'Dynamic Flux Balance Analysis of Diauxic Growth in *Escherichia coli*'. In: *Biophysical Journal* 83.3 (Sept. 2002), pp. 1331–1340. doi: 10.1016/S0006-3495(02)73903-9.
 - [99] Rahul Shaw and C. Y. Maurice Cheung. 'A Dynamic Multi-Tissue Flux Balance Model Captures Carbon and Nitrogen Metabolism and Optimal Resource Partitioning During *Arabidopsis* Growth'. In: *Front. Plant Sci.* 9 (26th June 2018). doi: 10.3389/fpls.2018.00884.
 - [100] Eugen Bauer et al. 'BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities'. In: *PLoS Computational Biology* 13.5 (2017), pp. 1–22. doi: 10.1371/journal.pcbi.1005544.
 - [101] Ilija Dukovski et al. 'A metabolic modeling platform for the computation of microbial ecosystems in time and space (COMETS)'. In: *Nat Protoc* 16.11 (Nov. 2021), pp. 5030–5082. doi: 10.1038/s41596-021-00593-3.
 - [102] Edward J. O'Brien, Jonathan M. Monk and Bernhard O. Palsson. 'Using Genome-scale Models to Predict Biological Capabilities'. In: *Cell* 161.5 (21st May 2015), pp. 971–987. doi: 10.1016/j.cell.2015.05.019.
 - [103] Yara Seif et al. 'A computational knowledge-base elucidates the response of *Staphylococcus aureus* to different media types'. In: *PLOS Computational Biology* 15.1 (9th Jan. 2019), e1006644. doi: 10.1371/journal.pcbi.1006644.
 - [104] Adam Amara et al. 'Networks and Graphs Discovery in Metabolomics Data Analysis and Interpretation'. In: *Front. Mol. Biosci.* 9 (8th Mar. 2022). doi: 10.3389/fmolb.2022.841373.
 - [105] Varshit Dusad et al. 'Opportunities at the Interface of Network Science and Metabolic Modeling'. In: *Front. Bioeng. Biotechnol.* 8 (25th Jan. 2021). doi: 10.3389/fbioe.2020.591049.
 - [106] Sergey Stolyar et al. 'Metabolic modeling of a mutualistic microbial community'. In: *Mol Syst Biol* 3 (2007), p. 92. doi: 10.1038/msb4100131.
 - [107] Almut Heinken, Arianna Basile and Ines Thiele. 'Advances in constraint-based modelling of microbial communities'. In: *Current Opinion in Systems Biology* 27 (2021). doi: 10.1016/j.coisb.2021.05.007.
 - [108] Sabine Koch et al. 'Predicting compositions of microbial communities from stoichiometric models with applications for the biogas process'. In: *Biotechnology for Biofuels* 9.1 (2016), pp. 1–16. doi: 10.1186/s13068-016-0429-x.
 - [109] Siu Hung Joshua Chan, Margaret N. Simons and Costas D. Maranas. 'SteadyCom: Predicting microbial abundances while ensuring community stability'. In: *PLoS Computational Biology* 13.5 (2017), pp. 1–25. doi: 10.1371/journal.pcbi.1005539.
 - [110] Steven L. Salzberg. 'Next-generation genome annotation: we still struggle to get it right'. In: *Genome Biology* 20.1 (16th May 2019), p. 92. doi: 10.1186/s13059-019-1715-2.
 - [111] Munira Alballa, Faizah Aplop and Gregory Butler. 'TranceP: Predicting the substrate class of transmembrane transport proteins using compositional, evolutionary, and positional information'. In: *PLoS ONE* 15.1 (2020), pp. 1–23. doi: 10.1371/journal.pone.0227683.
-

- [112] Vladimir Gligorijević et al. ‘Structure-based protein function prediction using graph convolutional networks’. In: *Nat Commun* 12.1 (26th May 2021), p. 3168. doi: 10.1038/s41467-021-23303-9.
- [113] Marc Griesemer et al. ‘Combining multiple functional annotation tools increases coverage of metabolic annotation’. In: *bioRxiv* (2017), pp. 1–11. doi: 10.1101/160887.
- [114] Nhung Pham et al. ‘Consistency, inconsistency, and ambiguity of metabolite names in biochemical databases used for genome-scale metabolic modelling’. In: *Metabolites* 9.2 (2019). doi: 10.3390/metabo9020028.
- [115] Sébastien Moretti et al. ‘MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models’. In: *Nucleic Acids Research* 49 (D1 8th Jan. 2021), pp. D570–D574. doi: 10.1093/nar/gkaa992.
- [116] Duygu Dikicioglu, Betul Kirdar and Stephen G. Oliver. ‘Biomass composition: the “elephant in the room” of metabolic modelling’. In: *Metabolomics* 11.6 (1st Dec. 2015), pp. 1690–1701. doi: 10.1007/s11306-015-0819-2.
- [117] Hoang V. Dinh, Debolina Sarkar and Costas D. Maranas. ‘Quantifying the propagation of parametric uncertainty on flux balance analysis’. In: *Metabolic Engineering* 69 (1st Jan. 2022), pp. 26–39. doi: 10.1016/j.ymben.2021.10.012.
- [118] Georgios Marinos, Christoph Kaleta and Silvio Waschina. ‘Defining the nutritional input for genome-scale metabolic models: A roadmap’. In: *PLoS ONE* 15.8 (2020), pp. 1–17. doi: 10.1371/journal.pone.0236890.
- [119] David Gilbert et al. ‘Towards dynamic genome-scale models’. In: *Brief Bioinform* 20.4 (13th Oct. 2017), pp. 1167–1180. doi: 10.1093/bib/bbx096.
- [120] Daniel Machado. ‘A benchmark of optimization solvers for genome-scale metabolic modeling of organisms and communities’. In: *mSystems* 9.2 (22nd Jan. 2024), e00833–23. doi: 10.1128/msystems.00833-23.
- [121] Anne Richelle, Chintan Joshi and Nathan E. Lewis. ‘Assessing key decisions for transcriptomic data integration in biochemical networks’. In: *PLOS Computational Biology* 15.7 (19th July 2019). Ed. by Vassily Hatzimanikatis, e1007185. doi: 10.1371/journal.pcbi.1007185.
- [122] The Bioconda Team et al. ‘Bioconda: sustainable and comprehensive software distribution for the life sciences’. en. In: *Nat Methods* 15.7 (2018), pp. 475–476. doi: 10.1038/s41592-018-0046-7.
- [123] Jan Schellenberger, Nathan E. Lewis and Bernhard Ø Palsson. ‘Elimination of Thermodynamically Infeasible Loops in Steady-State Metabolic Models’. In: *Biophysical Journal* 100.3 (2nd Feb. 2011), pp. 544–553. doi: 10.1016/j.bpj.2010.12.3707. pmid: 21281568.
- [124] Abdelmoneim Amer Desouki et al. ‘CycleFreeFlux: Efficient Removal of Thermodynamically Infeasible Loops from Flux Distributions’. In: *Bioinformatics* 31.13 (1st July 2015), pp. 2159–2165. doi: 10.1093/bioinformatics/btv096.
- [125] Zachary A. King et al. ‘Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways’. en. In: *PLOS Computational Biology* 11.8 (2015), e1004321. doi: 10.1371/journal.pcbi.1004321.
- [126] Alberto Noronha et al. ‘ReconMap: an interactive visualization of human metabolism’. In: *Bioinformatics* 33.4 (2017), pp. 605–607. doi: 10.1093/bioinformatics/btw667.
- [127] Archana Hari and Daniel Lobo. ‘Fluxer: a web application to compute, analyze and visualize genome-scale metabolic flux networks’. In: *Nucleic Acids Research* 48.W1 (2020), W427–W435. doi: 10.1093/nar/gkaa409.

-
- [128] Thomas Pfau et al. 'IDARE2—Simultaneous Visualisation of Multiomics Data in Cytoscape'. en. In: *Metabolites* 11.5 (2021), p. 300. doi: 10.3390/metabo11050300.
- [129] Paul Shannon et al. 'Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks'. en. In: *Genome Res.* 13.11 (2003), pp. 2498–2504. doi: 10.1101/gr.1239303.
- [130] Matthias König, Andreas Dräger and Hermann-Georg Holzhütter. 'CySBML: a Cytoscape plugin for SBML'. In: *Bioinformatics* 28.18 (2012), pp. 2402–2403. doi: 10.1093/bioinformatics/bts432.
- [131] Colblindor. *Coblis — Color Blindness Simulator – Colblindor*. URL: <https://www.color-blindness.com/coblis-color-blindness-simulator/>.
- [132] Andrew B. Shreiner, John Y. Kao and Vincent B. Young. 'The Gut Microbiome in Health and in Disease:' in: *Current Opinion in Gastroenterology* 31.1 (Jan. 2015), pp. 69–75. doi: 10.1097/MOG.0000000000000139.
- [133] Alexander Mahnert et al. 'The Human Archaeome: Methodological Pitfalls and Knowledge Gaps'. In: *Emerging Topics in Life Sciences* 2.4 (14th Dec. 2018). Ed. by Nicholas P. Robinson, pp. 469–482. doi: 10.1042/ETLS20180037.
- [134] Albane Ruaud et al. 'Syntrophy via Interspecies H₂ Transfer between *Christensenella* and *Methanobrevibacter* Underlies Their Global Cooccurrence in the Human Gut'. In: *mBio* 11.1 (4th Feb. 2020), e03235–19. doi: 10.1128/mBio.03235-19. pmid: 32019803.
- [135] Christina Kumpitsch et al. 'Reduced B12 Uptake and Increased Gastrointestinal Formate Are Associated with Archaeome-Mediated Breath Methane Emission in Humans'. In: *Microbiome* 9.1 (Dec. 2021), p. 193. doi: 10.1186/s40168-021-01130-w.
- [136] Jennie L. Catlett et al. 'Metabolic Synergy between Human Symbionts *Bacteroides* and *Methanobrevibacter*'. In: *Microbiology Spectrum* 10.3 (29th June 2022). Ed. by Jan Claesen, e01067–22. doi: 10.1128/spectrum.01067-22.
- [137] Manuel Glöckler, Andreas Dräger and Reihaneh Mostolizadeh. 'Hierarchical modelling of microbial communities'. In: *Bioinformatics* 39.1 (2023), btad040. doi: 10.1093/bioinformatics/btad040.
- [138] Lucian P. Smith et al. 'SBML Level 3 package: Hierarchical Model Composition, Version 1 Release 3'. en. In: *Journal of Integrative Bioinformatics* 12.2 (2015), pp. 603–659. doi: 10.1515/jib-2015-268.
- [139] Sijia Liu et al. 'Opportunities and Challenges of Using Metagenomic Data to Bring Uncultured Microbes into Cultivation'. In: *Microbiome* 10.1 (12th May 2022), p. 76. doi: 10.1186/s40168-022-01272-5.
- [140] Fernando Meyer et al. 'Critical Assessment of Metagenome Interpretation: The Second Round of Challenges'. In: *Nature Methods* 19.4 (Apr. 2022), pp. 429–440. doi: 10.1038/s41592-022-01431-4.
- [141] Andriy Kryshtafovych et al. 'Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XIV'. In: *Proteins: Structure, Function, and Bioinformatics* 89.12 (2021), pp. 1607–1617. doi: 10.1002/prot.26237.

Abbreviations

dFBA	Dynamic flux balance analysis
FBA	Flux balance analysis
FVA	FLux variability analysis
GEM	Genome-scale metabolic model
GIT	Gastrointestinal tract
GSMM	Genome-scale metabolic model
pFBA	parsimonious flux balance analysis
SBML	Systems biology markup language

A. Appendix

A.1. Targeted Metabolomics and High-Throughput RNA Sequencing-Based Transcriptomics Reveal Massive Changes in the *Streptomyces venezuelae* NRRL B-65442 Metabolism Caused by Ethanol Shock

A.1.1. Publication status

Published work: Sekurova, O. N., Zehl, M., Predl, M., Hunyadi, P., Rattei, T., Zotchev, S. B. *Targeted Metabolomics and High-Throughput RNA Sequencing-Based Transcriptomics Reveal Massive Changes in the Streptomyces venezuelae* NRRL B-65442 Metabolism Caused by Ethanol Shock. Microbiol Spectr 10:e03672-22. DOI: <https://doi.org/10.1128/spectrum.03672-22>

A.1.2. Author contributions

The author of this dissertation planned carried out the sequencing data analysis. He further planned and performed the construction of the metabolic network, as well as its integration and analysis with gene expression data. The author also produced figures and wrote parts of the manuscript.

A.1.3. Publication



RESEARCH ARTICLE



Targeted Metabolomics and High-Throughput RNA Sequencing-Based Transcriptomics Reveal Massive Changes in the *Streptomyces venezuelae* NRRL B-65442 Metabolism Caused by Ethanol Shock

Olga N. Sekurova,^a Martin Zehl,^b Michael Predl,^{c,d} Peter Hunyadi,^c Thomas Rattei,^{c,d} Sergey B. Zotchev^a

^aDepartment of Pharmaceutical Sciences, Division of Pharmacognosy, University of Vienna, Vienna, Austria

^bDepartment of Analytical Chemistry, Faculty of Chemistry, University of Vienna, Vienna, Austria

^cUniversity of Vienna, Centre for Microbiology and Environmental Systems Science, Division of Computational System Biology, Vienna, Austria

^dUniversity of Vienna, Doctoral School in Microbiology and Environmental Science, Vienna, Austria

Olga N. Sekurova and Martin Zehl contributed equally to this work. The order of the first co-authors' names was decided based on the alphabetical order of their surnames.

ABSTRACT The species *Streptomyces venezuelae* is represented by several distinct strains with variable abilities to biosynthesize structurally diverse secondary metabolites. In this work, we examined the effect of ethanol shock on the transcriptome and metabolome of *Streptomyces venezuelae* NRRL B-65442 using high-throughput RNA sequencing (RNA-seq) and high-resolution liquid chromatography-tandem mass spectrometry (LC-MS/MS). Ethanol shock caused massive changes in the gene expression profile, differentially affecting genes for secondary metabolite biosynthesis and central metabolic pathways. Most of the data from the transcriptome analysis correlated well with the metabolome changes, including the overproduction of jadomycin congeners and a downshift in the production of desferrioxamines, legonoxamine, foroxymithin, and a small cryptic ribosomally synthesized peptide. Some of the metabolome changes, such as the overproduction of chloramphenicol, could not be explained by overexpression of the cognate biosynthetic genes but correlated with the expression profiles of genes for precursor biosynthesis. Changes in the transcriptome were also observed for several genes known to play a role in stress response in other bacteria and included at least 10 extracytoplasmic function σ factors. This study provides important new insights into the stress response in antibiotic-producing bacteria and will help to understand the complex mechanisms behind the environmental factor-induced regulation of secondary metabolite biosynthesis.

IMPORTANCE *Streptomyces* spp. are filamentous Gram-positive bacteria known as versatile producers of secondary metabolites, of which some have been developed into human medicines against infections and cancer. The genomes of these bacteria harbor dozens of gene clusters governing the biosynthesis of secondary metabolites (BGCs), of which most are not expressed under laboratory conditions. Detailed knowledge of the complex regulation of BGC expression is still lacking, although certain growth conditions are known to trigger the production of previously undetected secondary metabolites. In this work, we investigated the effect of ethanol shock on the production of secondary metabolites by *Streptomyces venezuelae* and correlated these findings with the expression of cognate BGCs and primary metabolic pathways involved in the generation of cofactors and precursors. The findings of this study set the stage for the rational manipulation of bacterial genomes aimed at enhanced production of industrially important bioactive natural products.

KEYWORDS *Streptomyces venezuelae*, ethanol shock, metabolomics, transcriptomics

Editor Paul A. Jensen, University of Michigan-Ann Arbor

Copyright © 2022 Sekurova et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Sergey B. Zotchev, sergey.zotchev@univie.ac.at.

The authors declare no conflict of interest.

Received 30 September 2022

Accepted 5 October 2022

Published 31 October 2022

Gram-positive filamentous bacteria belonging to the genus *Streptomyces* dwell in various complex environments, such as soils, marine sediments, plants, and animals, where they are often challenged with both biotic and abiotic stress factors (1). Hence, these bacteria must be able to adjust their gene expression patterns in order to meet these challenges and to change their metabolism accordingly. One of the well-known traits of streptomycetes is their ability to biosynthesize chemically diverse secondary metabolites in response to various stress factors, such as nutrient limitation, temperature shift, and the presence of other microorganisms nearby (2). Some of these secondary metabolites have antimicrobial activity, and quite a few have been developed into medically useful antibiotics, e.g., erythromycin, daptomycin, and tetracyclines. The roles that antibiotically active secondary metabolites play in natural environments are still being debated, but mounting evidence shows that besides their involvement in the “chemical warfare” to eliminate competitors for nutrient sources (1), they also act as signaling molecules at subinhibitory concentrations (3) that can shape microbial communities (4). Understanding of the environmental factor-dependent mechanisms that trigger the biosynthesis of secondary metabolites would provide important insights into the biology of streptomycetes and may also be applied to increase the industrial yields of medically important antibiotics.

In bacteria, one- and two-component systems, as well as extracytoplasmic function σ factors (ECFs), play pivotal roles in sensing the environment and transducing the signals inside the cell, leading to specific changes in the gene transcription patterns (5). The one-component systems consist of one protein with sensing and regulatory (DNA binding) domains (6), while the two-component systems are composed of a sensor kinase and a transcriptional regulator, subject to activation by the cognate kinase via phosphorylation (7). The activity of ECFs is usually regulated by transmembrane anti- σ factors, to which they are tightly bound until their release in response to an environmental stimulus. Notably, there also exist “hybrid” systems, where ECF sigma factors contain additional sensor domains that regulate their activity along with anti- σ factors (8). Both, two-component systems and ECFs are known to control secondary metabolism in *Streptomyces* (9–11) and thus appear to be crucial for the activation of the latter in response to environmental stimuli. However, molecular details of the mechanisms behind such signal transduction remain mostly obscure due to the extreme complexity of the regulatory networks that control secondary metabolism in streptomycetes. *Streptomyces venezuelae*, known as a producer of the antibiotic chloramphenicol, represents an interesting study subject in this respect since it is known to respond to phage infection, temperature shift, or ethanol shock by changing its secondary metabolome (12, 13). In *S. venezuelae* ATCC 10712, ethanol stress suppresses the production of chloramphenicol while simultaneously triggering the biosynthesis of angucycline polyketides, namely, jadomycins (12), and the regulatory mechanism behind such reciprocal regulation has been studied at the level of pathway-specific regulators (14, 15).

In this work, we investigated the effect of ethanol stress on the global transcriptome of *S. venezuelae* NRRL B-65442 by means of high-throughput RNA sequencing (RNA-seq) and correlated these data with the changes in the secondary metabolome. In addition, we examined the effect of ethanol shock on the transcription of genes governing major primary metabolic pathways and genes encoding ECFs as well as proteins known to be involved in stress response. The results obtained represent a significant step toward deciphering a complex process of stress response in *Streptomyces* bacteria and its connection to secondary metabolite biosynthesis.

RESULTS AND DISCUSSION

***Streptomyces venezuelae* NRRL B-65442 genome and its BGCs and secondary metabolites.** Various strains of *Streptomyces venezuelae* have been investigated by several research groups (16, 17). This species has remarkable properties for streptomycetes, in that it grows very fast and sporulates almost synchronously in liquid medium. The two very closely related strains *S. venezuelae* ATCC 10712 and NRRL B-65442 received considerable attention due to their ability to activate antibiotic biosynthesis upon ethanol shock

TABLE 1 Secondary metabolite biosynthesis gene clusters in the genome of *S. venezuelae* NRRL B-65442 and effects of ethanol shock on their expression and cognate metabolites

BGC	Type	Known or putative product ^a	Gene	Metabolome	Transcriptome	Pathway
1	Ectoine	Ectoine/betaine	RS01090–01095	ND ^b	Down	P101-PWY
2	Terpene	Geosmin	RS01290	ND	Down	PWY-5950
3	Pks/NRPS	Venemycin	RS02305–02380	ND	None	None
4	NRPS	Watasemycins/thiazostatins	RS02405–02475	Down	None	None
5	Lantipeptide	Core: ADhalDVPYDhbDhbGDhbIDhbVC	RS02610–02620	ND	None	None
6	Terpene	Terpenoid	RS02650–02655	ND	Down	None
7	Lantipeptide	Venezuelin	RS02970–02980	ND	None	None
8	Indole	Arcyriaflavin	RS03645–03670	ND	None	None
9	NRPS	Chloramphenicol	RS04425–04505	Up	Down/up	PWY-8032
10	CDPS	Cyclic dipeptide	RS09190–09195	None	None	PWY-7236
11	Siderophore	Desferrioxamine	RS12675–12690	Down	Down	PWY-6375
12	Lasso peptide	Core: GDAAELTQGQGGGQSEDKRRAVNC	RS15430–15450	ND	None	None
13	NRPS	Peptide	RS20250–20325	ND	None	None
	Other	Gaburedins	RS20785–20795	Down	Down	None
14	Butyrolactone	Butyrolactone	RS20800–20805	ND	None	None
15	Melanin	Melanin	RS23170–23175	ND	None	PWY-6481
16	Butyrolactone	Butyrolactone	RS25385	ND	None	None
17	Other	Pyrrolnitrin-like?	RS25415–25435	ND	None	PWY-6831
18	Thiopeptide	Core: SSGSIGTSSSSSTCSAC	RS25565–25620	ND	Down	None
19	Pks3	Flaviolin	RS26780	ND	None	None
20	Siderophore	Siderophore	RS27055–27060	ND	None	None
21	Siderophore	Legonoxamines	RS27330–27345	Down	Down	None
22	Bacteriocin	Bacteriocin	RS29140	ND	None	None
23	Pks2	Jadomycin	RS29740–29955	Up	Up	PWY-6679
24	NRPS	Peptide	RS30590–30630	ND	Up	None
25	Pks2-NRPS	Peptide-polyketide hybrid	RS30820–30995	ND	Down/up	None
26	Ladderane	Ladderane	RS31000–31150	ND	Down/up	None
27	NRPS	Terpapeptide (x-Phe-X-Gly?)	RS31155–31260	ND	Down	None
28	Terpene	Hopanoids	RS32065–32140	ND	Down	PWY-7072
29	Bacteriocin	Bacteriocin	RS32515	ND	None	None
30	Pks2	Spore pigment	RS33815–33860	ND	Down	None
31	Melanin	Melanin	RS34060–34065	ND	None	None
32	NRPS	Foroxymithines	RS35125–35150	Down	Down	None
33	Terpene	2-Methylisoborneol	RS35405–35410	ND	Down	None
34	Pks3	Alkylresorcinol	RS36050–36065	Down	Down	None
35	Terpene	Terpenoid	RS36960	ND	None	None
36	NRPS	Peptide	RS37045	ND	Down	None

^aKnown secondary metabolites for which cognate BGCs are known are shown in bold.^bND, not detected in conditions tested with the given workflow.

(12, 13). Gomez-Escribano et al. (16) recently compared the genomes of these strains, which differed in 46 nucleotide mismatches and 34 indels. These differences might account for the previously observed variation in the strains' response to ethanol shock, whereby chloramphenicol biosynthesis was suppressed by it in ATCC 10712 while being stimulated in NRRL B-65442 (mistaken for ATCC 10712 in reference 13). We once again examined the genome of NRRL B-65442 using antiSMASH6.0 (18), followed by manual curation, and amended the table of BGCs proposed by Gomez-Escribano et al. (16) as shown in Table 1. Since the effect of ethanol stress on both ATCC 10712 and NRRL B-65442 has been studied in one liquid medium, namely, Maltose Yeast Extract medium (MYM) (see Materials and Methods), we decided to use the same conditions in the follow-up experiments with strain NRRL B-65442, aimed at examining its secondary metabolome.

Previously, secondary metabolites corresponding to seven of the BGCs shown in Table 1 were reported from *S. venezuelae* ATCC 10712 and NRRL B-65442. Using a liquid chromatography-mass spectrometry (LC-MS)-based metabolomics workflow, the production of the following metabolites in MYM medium was confirmed by matching accurate mass, isotopic pattern, and tandem MS (MS/MS) data: gaburedins A to D (19); the 2-hydroxyphenylthiazolines watasemycin A/B, thiazostatin A/B, isopyochelin, pulicatin A to D, and aerugine (20); and chloramphenicol (21), as well as desferrioxamine B (22) and

numerous related siderophores, such as legonoxamine A (23). The *S. venezuelae* BGC-encompassing genes RS27330 to RS27345 (Table 1) could be linked to the biosynthesis of legonoxamines, as it encoded enzymes similar to those for the biosynthesis of desferrioxamines.

Although jadomycin B could not be detected, the related angucyclines L-digitoxosyl-phenanthroviridin described for *S. venezuelae* ISP5230 (24) and, in smaller amounts, phenanthroviridin aglycone (25) were tentatively identified. L-Digitoxosyl-phenanthroviridin was proposed to be the stable degradation product of the jadomycin B analogue with L-lysine incorporated instead of L-isoleucine (24). Our data support this hypothesis since a minor peak matching this transient jadomycin L-lysine was detected in some samples. In addition, the rare siderophore foroxymithine (26) was identified in *S. venezuelae* for the first time and tentatively matched to the BGC responsible for its production (see Table 1). Venemycin (27) and venezuelin (28) were not detected, but the production of both these compounds was described only in a mutant strain or upon heterologous expression and not in the wild-type *S. venezuelae*, where the corresponding BGCs appear to be silent under these cultivation conditions.

Besides the above-mentioned natural products, a large number of other metabolites were tentatively identified, such as amino acids and their derivatives (e.g., *N*-phenylacetylglutamine), diketopiperazines [e.g., cyclo(Phe-Pro)], nucleosides (e.g., various *N*- and *O*-methyladenosines), lipids (e.g., fatty acid amides and phosphatidylethanolamines), and cofactor-related compounds (e.g., biotin and pantothenic acid).

One of the main peaks in the chromatograms of the *S. venezuelae* NRRL B-65442 culture extract turned out to be a linear ribosomal peptide of 19 amino acids long with a formyl group on the *N*-terminal methionine (Met). With lower abundance, derivatives originating from the oxidation of the Met to the sulfoxide, C-terminal truncation down to the tetrapeptide, deamination of an Asn to Asp, and combinations of those were observed. The full sequence of the peptide, designated PepX, could be derived from the MS/MS spectra (with the exception of Leu and Ile, which cannot be differentiated easily in *de novo* sequencing). PepX was found to be encoded by a small gene provisionally named RS03776, located between RS03775 and RS03780 encoding a putative endolytic transglycosylase and an ABC transporter ATP-binding protein, respectively. No putative function could be assigned to the PepX peptide based on its amino acid sequence, which was translated as *N*-formyl-MNVITNLLAGVHFLGWLV in agreement with the MS/MS data. However, the *pepX* gene homologue could be identified in genomes of various streptomycetes and appears to be conserved. High-resolution MS and MS/MS spectra of the above-mentioned compounds and more details on their identification can be found in the Supplemental Material (Fig. S1 to S9).

Effect of the ethanol shock on the secondary metabolism of *S. venezuelae*. To investigate the impact of ethanol shock on the production of secondary metabolites and the relative concentration of other identified compounds in the culture, *S. venezuelae* NRRL B-65442 was cultivated in triplicates in MYM medium in the presence and absence of 6% ethanol added after 7 h of initial growth. The methanolic culture extracts were analyzed by LC-MS after 18 h, 24 h, and 48 h of incubation. The data set was compared using an untargeted data analysis workflow to identify the most significantly affected compounds. Ethanol shock was found to have a profound effect on the metabolite profile, and the amount of nearly all secondary metabolites detected was either increased or decreased (Fig. 1). The most obvious impact was on the hydroxamate siderophores because they are among the most abundant compounds already after 16 h of cultivation, and in the case of desferrioxamines/legonoxamines, they are represented by a large number of congeners and derivatives which all behaved identical qualitatively. The production of both foroxymithine and desferrioxamine B congeners were nearly abolished after the addition of ethanol to the medium. The same was true for the 2-hydroxyphenylthiazolines, such as thiazostatin A or B, which were however detected only at very low levels. Interestingly, the same very strong depletion was observed for PepX (Fig. 1), indicating a possible correlation in the regulation of the biosynthesis of this simple ribosomal peptide and the more complex multienzyme

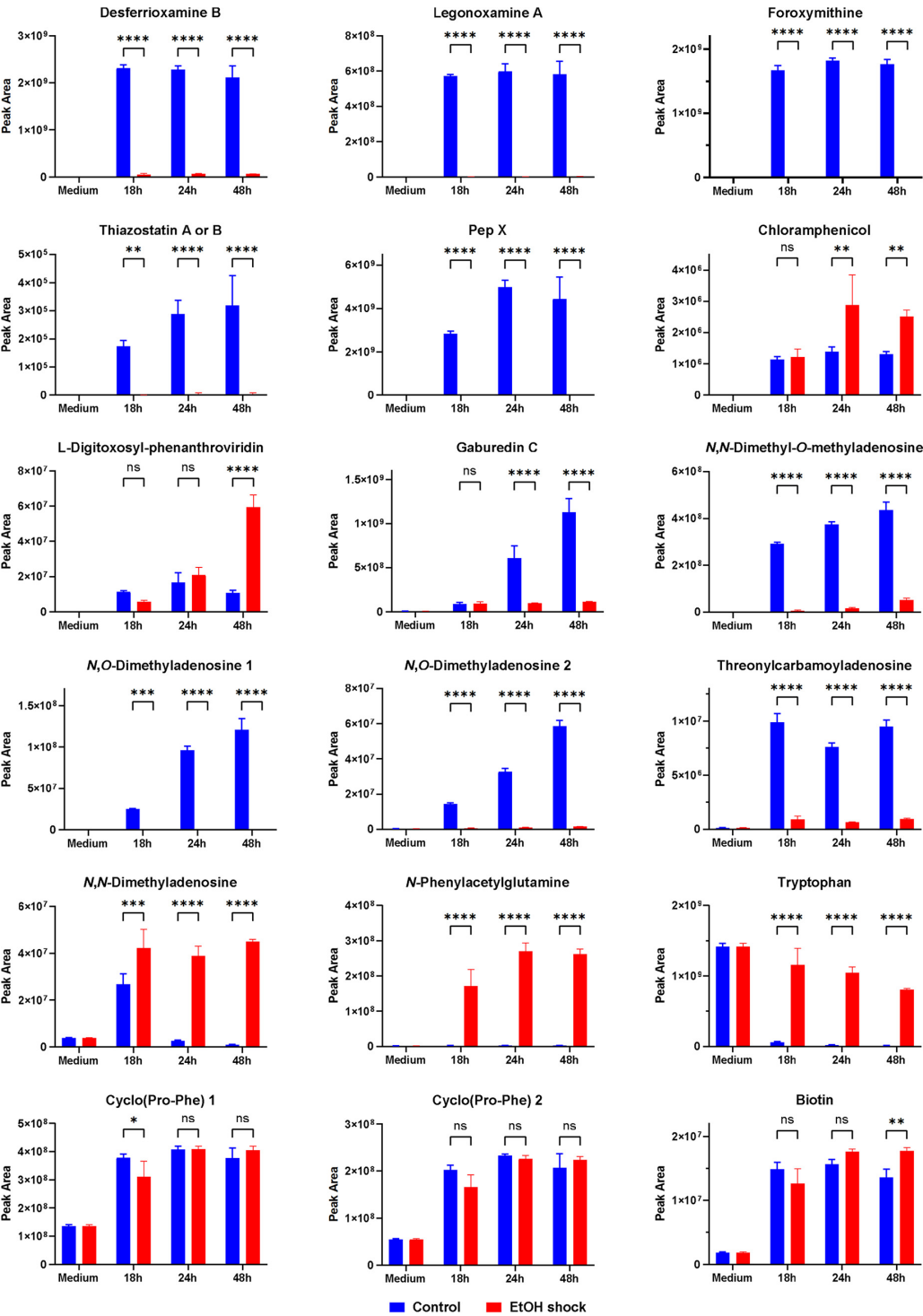


FIG 1 Effect of ethanol shock on the production of secondary metabolites and the relative concentration of other identified compounds in MYM cultures of *S. venezuelae* NRRL B-65442 after 18 h, 24 h, and 48 h of cultivation. After 7 h of cultivation, either 6% (vol/vol) of (Continued on next page)

factories responsible for the synthesis of foroxymithine, desferrioxamine B, legonoxamine A, and the 2-hydroxyphenylthiazolines. In agreement with our previous observations, the production of chloramphenicol (13) and of the jadomycin-related compounds L-digitoxosyl-phenanthroviridin and phenanthroviridin aglycone was increased by ethanol shock but only at later time points. Conversely, the level of the gaburedins seemed to be reduced by ethanol shock at the later time points (Fig. 1).

In the search for other compounds which were affected and could help to unravel the regulatory network, several modified nucleosides stood out. Most of the tentatively identified nucleosides, namely, *N,N*-dimethyl-*O*-methyladenosine, two *N,O*-dimethyladenosine isomers, and threonylcarbamoyladenosine, were downregulated strongly upon ethanol shock, while an *N,N*-dimethyladenosine was upregulated strongly (Fig. 1). Although the fragmentation pattern allows only the localization of the methyl groups on either the base or sugar part without further details, it is likely that the methylation of adenine takes place at the *N*⁶ position. *N*⁶,*N*⁶,2'-*O*-trimethyladenosine, *N*⁶,2'-*O*-dimethyladenosine, *N*⁶-threonylcarbamoyladenosine, and *N*⁶,*N*⁶-dimethyladenosine are all well-known products of RNA modification reactions that play an important role in transcription and translation (29). However, while the "RNA epigenetic" system of enzymes responsible for "writing, erasing and reading" the RNA *N*⁶-methyladenosine (*m*⁶A) modifications is reasonably well established for many eukaryotes, little is known about such a system in bacteria despite *m*⁶A being confirmed as a widespread modification (29, 30). Also, the relatively high levels of these free, modified nucleosides are puzzling considering that only <0.08% of the adenosine in mRNA from the two Gram-positive bacteria, *Staphylococcus aureus* and *Bacillus subtilis*, was found to be *N*⁶-methylated (31).

Another compound produced in larger amounts after ethanol shock was *N*-phenylacetylglutamine (Fig. 1), an amino acid derivative known as a gut microbiota-host metabolite that has been investigated as a potential biomarker for coronary heart disease (32). To the best of our knowledge, this compound has not yet been reported from streptomycetes and its biosynthetic origin is not clear. However, since the increase of *N*-phenylacetylglutamine coincides with the decrease in legonoxamine A content, it is possible that its production is not upregulated on a transcriptional or translational level but is influenced merely by the availability of larger amounts of phenylacetyl-coenzyme A (CoA) that is otherwise channeled into the biosynthesis of legonoxamine A.

We also observed that several amino acids and *N*-acetyl amino acids are more abundant upon ethanol shock. Since the amino acids are already present in the medium as nutrients in even larger amounts, their relatively higher concentration upon ethanol-induced stress might indicate lower uptake and consumption (Fig. 1). However, numerous other metabolites were not or only very little affected, as exemplified for the two cyclo(Phe-Pro) stereoisomers and biotin, demonstrating that the strong downregulation of several metabolites shown above cannot be simply explained by a general reduction in metabolism upon the ethanol shock.

In the course of a follow-up study, three additional batches of *S. venezuelae* NRRL B-65442 cultures in MYM medium in the presence and absence of 6% ethanol were analyzed. To test the robustness and reproducibility of the above-described findings, these data were examined by a targeted workflow, and the results are shown in Fig. S10 in the supplemental material. Despite strong variation in the peak areas between batches, all the trends reported above could be confirmed except for the gaburedins, for which no clear picture was obtained.

Ethanol shock causes massive changes in the transcriptome. A comparative RNA-seq-based transcriptome analysis was performed for the *S. venezuelae* cultures grown with and without ethanol shock. The samples for RNA isolation were taken from

FIG 1 Legend (Continued)

absolute ethanol (red bars) or distilled sterile water (blue bars) was added. Each bar represents the average of three independent experiments ($n = 3$), except for the medium control ($n = 2$). The error bars represent the standard deviation. The statistical differences between groups were evaluated using two-way ANOVA followed by the Šidák's multiple-comparison test (****, $P < 0.0001$; ***, $P < 0.0001$ to $P < 0.001$; **, $P < 0.001$ to $P < 0.01$; *, $P < 0.01$ to $P < 0.05$).

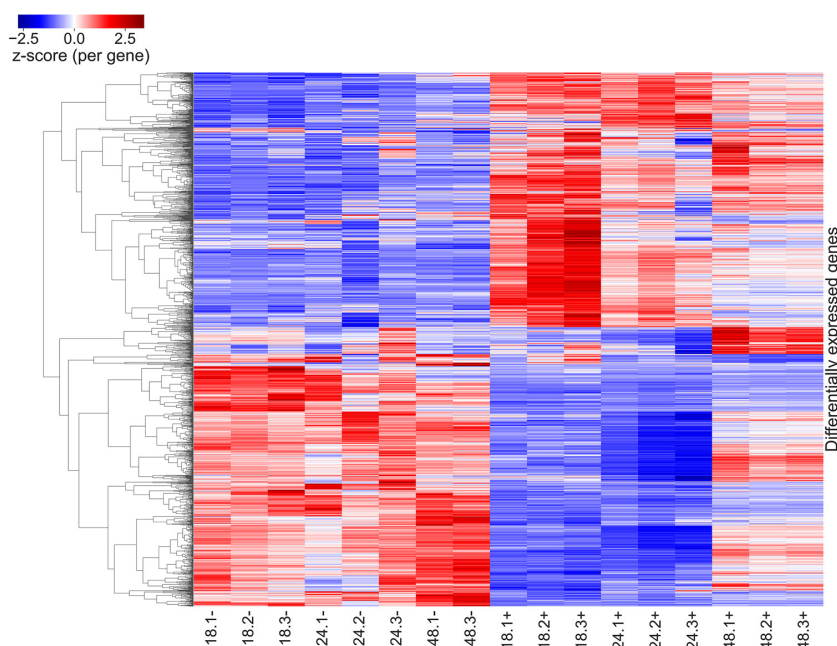


FIG 2 Comparison of differentially expressed genes across all samples and time points. Sample names on the bottom show the time point (18, 24, and 48 h) of collection, the ID of the biological replicate (1–3), and treatment with ethanol (+) or water (–). The individual cells each represent a differentially expressed gene in given sample. For each gene, a z-score is calculated based on mean and standard deviation across all samples. The Z-score is used to color the genes for a deviation from the mean with blue being less expressed, red being higher expressed, and white being expression close to the mean. Genes are clustered based on their correlation with each other, resulting in clear separation of ethanol-shocked and control samples.

the same cultures for which secondary metabolome analysis have been performed. After an 18-h post-ethanol shock, 1,246 out of the 7,204 annotated genes in the *S. venezuelae* genome (17.2%) were found to be differentially affected. At 24 and 48 h post-ethanol shock, 1,115 and 638 genes, respectively, were differentially expressed (see Table S1 in the supplemental material). A comparison of differentially expressed genes across all samples and time points is shown as a heat map in Fig. 2, where it is evident that the most significant changes in the transcriptome happen at 18 h and 24 h post-ethanol shock. It was also evident that at the 48-h time point that the expression of many genes is more similar to that of the control samples, as already indicated by the reduced number of differentially expressed genes. A comparison of all differentially expressed genes over all samples and time points reveals that their expression pattern clearly separates ethanol shock samples from control samples. The three different time points show distinct gene sets up- or downregulated consistently within each of the three ethanol shock replicates, with more similarity between 18 h and 24 h.

The effect of ethanol treatment on gene expression has been studied for bacteria used to produce alcohol-containing beverages, such as *Lactobacillus plantarum* and *Oenococcus oeni* (33, 34). The most frequently observed response to environmental stress, including ethanol shock, was reported to be changes in the bacterial membrane. Ethanol and other chemicals, such as hydroxylamine derivatives, increase membrane fluidity (35), and this process can be counteracted via several mechanisms. They include changes in the phospholipid composition (36) as well as overexpression of membrane-binding heat shock proteins, such as DnaK and GroEL (37). We specifically looked at the transcription profiles of *S. venezuelae* genes RS17015 and RS22115 encoding DnaK and GroEL heat shock proteins, respectively. Both genes were upregulated at 18 h and 24 h, with RS22115 also upregulated at 48 h after ethanol shock.

TABLE 2 Expression of *S. venezuelae* σ factors affected by the ethanol shock^a

Gene	Expression at:			σ domains
	18 h	24 h	48 h	
RS00060			Up	r2, r4
RS02270	Down	Down		r2, r3, r4
RS13895	Up	Up	Up	r2, r3, r4
RS15140	Up	Up		r1, r2, r3, r4
RS15760	Down	Down	Down	r2, r4
RS15950	Up	Up	Up	r2, r4
RS17255		Up		r2, r4, Snoal
RS18620	Up			r2, r4
RS18755	Down	Down	Down	r2, r3, r4
RS21030	Up			r2, r4
RS22475		Down		r2, r4
RS22785	Up	Up	Up	r2, r4
RS23885	Up			r2, r4
RS24280	Up	Up	Up	r2, r4

^aCells representing genes that encode ECFs are shaded in gray.

Besides being molecular chaperons that assist in correct protein folding, DnaK and GroEL are recruited to the bacterial membrane upon ethanol shock in *Bacillus subtilis* (37). At least for GroEL it was shown that such binding stabilizes the membrane (38). It is thus logical to assume that upregulation of DnaK and GroEL homologue expression in *S. venezuelae* upon ethanol shock serves the same purpose.

Some other proteins, such as Asp23 in *Staphylococcus aureus* overexpressed upon alkaline shock, were shown to be tethered to the membrane and implicated in stress response (39). Deletion of the *asp23* gene leads to upregulation of cell wall stress genes and increases resistance to the membrane-targeting antibiotic daptomycin (40). In *S. venezuelae*, expression of the gene RS07310 encoding an Asp23/Gls24 family envelope stress protein is downregulated at all three time points examined, which correlates well with the findings reported for its *S. aureus* counterpart.

From the data presented above, it appears plausible that the membrane of *S. venezuelae* undergoes significant alterations upon cell exposure to ethanol. One of the most important consequences of such perturbations could be changes in the state of membrane-bound proteins that may interact with ECFs and modulate their activity (41). With this in mind, we examined the effect of ethanol shock on the transcription of all 56 σ factors encoded in the *S. venezuelae* genome at 18 h, 24 h, and 48 h. The results of this analysis are presented in Table 2. In total, the expression of 14 σ factors, with 10 of them ECFs, was affected by the ethanol shock, with 2 being downregulated (1 of them ECF) and 4 upregulated (3 of them ECFs) at all three time points. We suggest that σ factors, whose expression was affected only at 24 h or/and 48 h and not at 18 h, are unlikely to be involved in the regulation of secondary metabolism since metabolome changes after ethanol shock were observed already at 18 h.

Correlation of the metabolome and transcriptome data does not always provide consistent results. We attempted to correlate the metabolome and transcriptome data, specifically looking at the genes known to be involved in the biosynthesis of detected secondary metabolites. Genes for the biosynthesis of desferrioxamine, legonoxamines, and foroxymithines, as well as the gene encoding peptide X, were clearly downregulated at the 18-h and/or 24-h time points, which correlated well with the drastically reduced accumulation of these compounds, while no such correlation was found for gaburedins. The finding that there was no correlation for gaburedins may be explained by the upregulation of the biosynthesis at the level of translation or via a decreased availability of the precursor γ -aminobutyric acid (19). The correlation between the transcriptome and metabolome was also poor for chloramphenicol and jadomycin BGCs, as none of the genes involved in the biosynthesis of corresponding scaffolds (e.g., nonribosomal peptide synthetase [NRPS] for chloramphenicol and ketosynthase for jadomycin) were affected by the ethanol shock. However, the expression of the regulatory genes *jadR1* (RS29820) and

jadR (RS29940), as well as some of the biosynthetic genes, namely, *jadJ* (RS29825, acyl-CoA carboxylase), *jadB* (RS29840, ketosynthase chain-length factor), *jadF1* (RS29860, FAD-dependent oxidoreductase), *jadH* (RS29870, FAD-dependent oxidoreductase), and *jadK* (RS29875, alpha/beta hydrolase) in the jadomycin BGC, were all upregulated at 18 h. It is possible that enzymes encoded by the upregulated *jad* genes represent “bottlenecks” in the jadomycin biosynthetic pathway, and hence, their overexpression circumvents this problem.

In the case of chloramphenicol, the situation appears to be more complex. The *cmIB* gene (RS04460) encoding aminodeoxychorismate (ADC) synthase is downregulated at the 24-h time point after the ethanol shock, which does not correlate with the apparent increase in chloramphenicol production. ADC synthase, besides chloramphenicol biosynthesis, is involved in the formation of *para*-aminobenzoic acid, a precursor for the formation of folates, the essential cofactors. Chang et al. (42) reported cloning of the *S. venezuelae* ISP5230 (ATCC 10712) PABA biosynthesis genes, *pabA* and *pabB*, which are located outside the chloramphenicol BGC and are responsible for the ADC synthesis. Disruption of *pabA* resulted in 97% reduction in the chloramphenicol biosynthesis, clearly suggesting that CmlB is not the main enzyme responsible for the generation of ADC starter. In *S. venezuelae* NRRL B-65442, homologous proteins are encoded by the genes RS31140 (*pabA*) and RS31145 (*pabB*), and their expression is not affected by the ethanol shock. Hence, downregulation of *cmIB* is unlikely to have any consequences for the chloramphenicol production. On the other hand, expression of the gene *cmIJ* (RS04490) encoding an SDR family oxidoreductase implicated into the last step of chloramphenicol biosynthesis (21) is upregulated at 24 h, which correlates well with the increase of this antibiotic yield after 24 h (Fig. 1). Apart from the regulation of protein expression or enzyme function, increased levels of chloramphenicol's main precursor, chorismate, could explain the observed stimulation of the biosynthesis. Our data show an increased expression of chorismate biosynthesis genes (see Fig. 5B), providing support for this hypothesis.

Notably, chorismate is the precursor for the biosynthesis of both chloramphenicol and thiazostatins/watasemycins (20). Hence, the suppression of the thiazostatins biosynthesis upon ethanol shock is likely due to the depletion of chorismate that may be channeled into the chloramphenicol biosynthesis pathway since no changes in the expression of the watasemycin biosynthetic genes could be detected.

Although the metabolome analysis did not detect other compounds which, according to the earlier reports could have been produced by *S. venezuelae*, we examined the gene expression patterns of their BGCs upon ethanol shock. According to such an analysis, the expression genes for ectoin, geosmin, and 2-methylisoborneol biosynthesis, as well as the PKS genes from the spore pigment BGC, were downregulated upon ethanol shock (Table S1). These compounds are likely produced in the culture but not detected by the applied metabolomics workflow due to their high polarity (ectoin), volatility (geosmin and 2-methylisoborneol), or low solubility (spore pigment).

Genome and omics-based analyses of selected *S. venezuelae* metabolic pathways. To better correlate the transcriptome and metabolome data for the secondary metabolite biosynthesis pathways affected by the ethanol shock in *S. venezuelae*, we performed their integration into the schemes presented in Fig. 3 to 5A and in Fig. S11 in the supplemental material. Figure 3A shows the desferrioxamine B biosynthesis pathway in MetaCyc (PWY-6376), overlaid with differential gene expression data for all four genes from the desferrioxamine BGC. The gene which product catalyzes the first reaction, conversion of L-lysine to cadaverine is downregulated at the 24-h time point, while the genes governing the next two reactions, up to *N*-hydroxy-*N*-succinylcadaverine, are also downregulated at 18 h. Only the last gene encoding the DesD enzyme responsible for the condensation of *N*-acylated *N*-hydroxycadaverine with two molecules of *N*-succinyl-*N*-hydroxycadaverine to form various desferrioxamines, including desferrioxamine B, was not differentially expressed.

Figure 3B shows the hopanoid biosynthesis pathway in MetaCyc (PWY-7072), overlaid with the differential gene expression data. The pathway from (2E,6E)-farnesyl diphosphate to squalene requires the products of three genes, but the exact reactions involved are unclear based on the predictions. The pathway assumes the conversion of

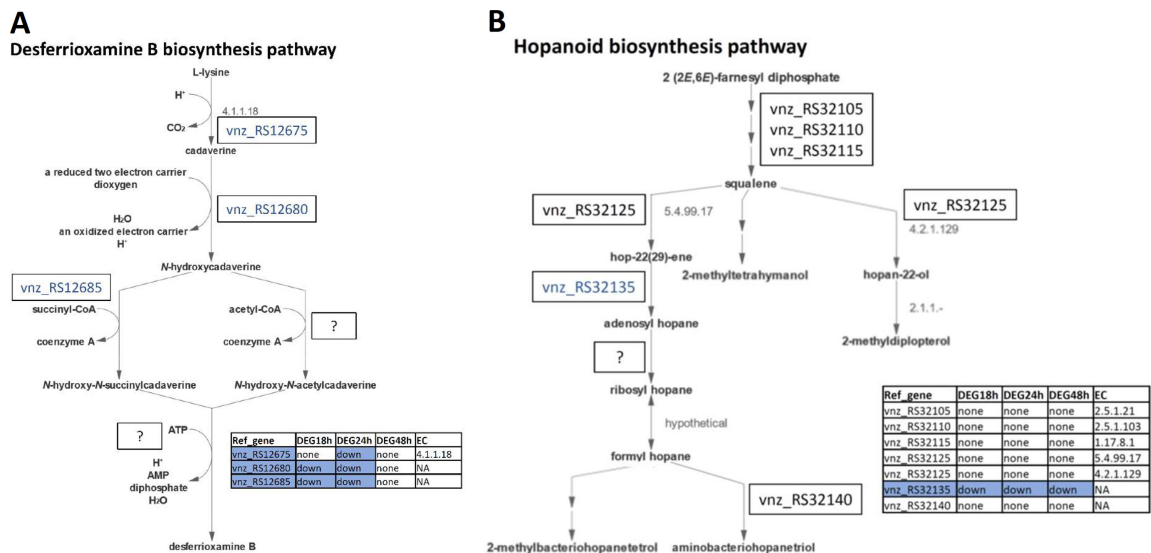


FIG 3 (A) Visualization of the desferrioxamine B biosynthesis pathway according to MetaCyc (PWY-6376). (B) Visualization of the hopanoid biosynthesis pathway according to MetaCyc (PWY-7072). *S. venezuelae* genes encoding enzymes involved in the pathways and their expression patterns at 3 time points after ethanol shock are shown. Downregulated genes are shown in blue.

presqualene diphosphate to squalene as two separate reactions via hydroxysqualene. Only for the second step, a gene coding for the enzyme responsible was predicted. However, a different gene was predicted to specify a one-step reaction from presqualene diphosphate to squalene, which is attributed to the epoxysqualene pathway. Furthermore, only some of the branches of the pathway are predicted to have genes encoding enzymes catalyzing these reactions, ending at hopan-22-ol and aminobacteriohopanetriol. The second branch contains the only reaction that is connected to a differentially expressed gene, RS32135, which is downregulated at all time points after the ethanol shock. This gene encodes adenosylhopane synthase HpnH that is responsible for the synthesis of C35 bacteriohopanepolyols, which represent an important part of bacterial membranes largely affecting their fluidity (43). Hence, downregulation of this gene likely results in a more rigid and less permeable membrane, thus adding to the stabilizing effect of DnaK/GroEL upregulation and Asp23 homologue downregulation.

The jadomycin biosynthesis pathway in MetaCyc (PWY-6679) is shown in Fig. 4, overlaid with differential gene expression data. All reactions in this pathway were matched with particular genes in the jadomycin BGC. Genes for enzymes catalyzing three different reactions were upregulated in this pathway; RS29860 is upregulated at 18 h and corresponds to the reaction from UWM6, an angucycline intermediate, to pre-jadomycin. The follow-up reaction is governed by RS29870, which is upregulated also at 24 h. Finally, the first reaction in the pathway, namely, the conversion of acetyl-CoA to malonyl-CoA, is predicted to be enabled by four different genes, of which only one (RS29825) belongs to the jadomycin BGC and is upregulated at 18 h. The other three genes are not located within jadomycin BGCs, but two of them are upregulated at 18 h and 24 h, with the last one not being differentially expressed.

Figure 5A shows the chloramphenicol biosynthesis pathway according to MetaCyc (PWY-8032), overlaid with differential gene expression data. This pathway consists of nine reactions catalyzed by the enzymes encoded by seven genes of the chloramphenicol BGC. One reaction [3-(4-aminophenyl)pyruvate to 4-amino-L-phenylalanine] does not have a gene prediction, as the MetaCyc database does not have a representative enzyme or sequence attributed to it. Out of the seven genes, only one is differentially

Jadomycin biosynthesis pathway

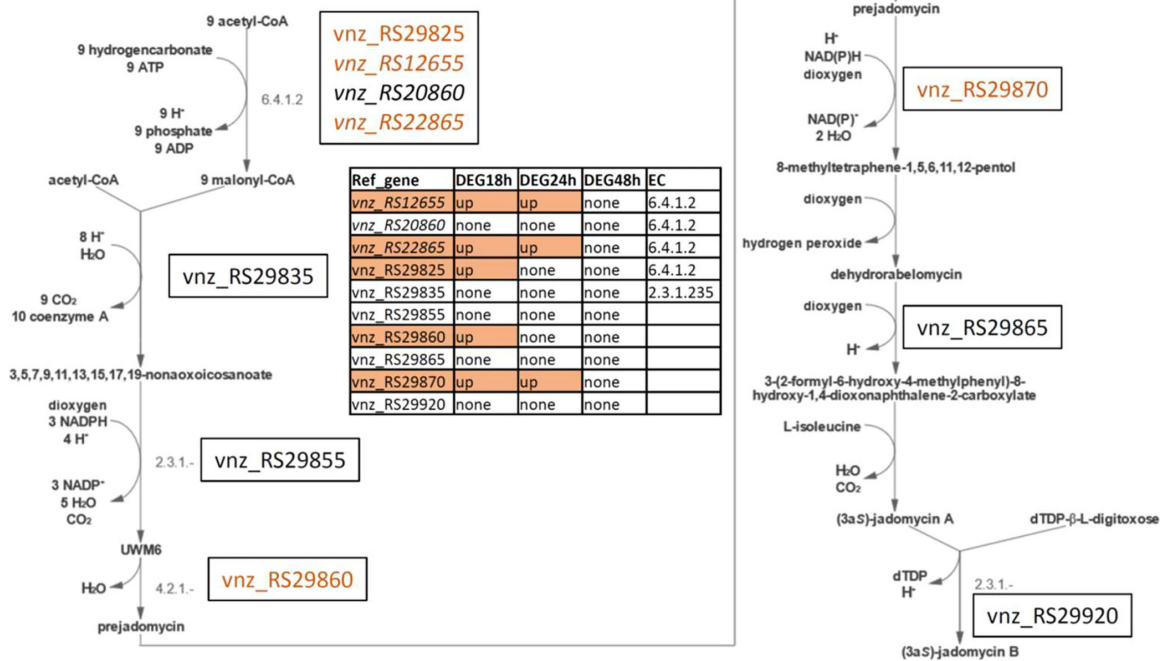


FIG 4 Visualization of the jadomycin biosynthesis pathway according to MetaCyc (PWY-6679). Genes encoding enzymes involved in the pathways and their expression patterns at 3 time points after ethanol shock are shown. Upregulated genes are shown in orange.

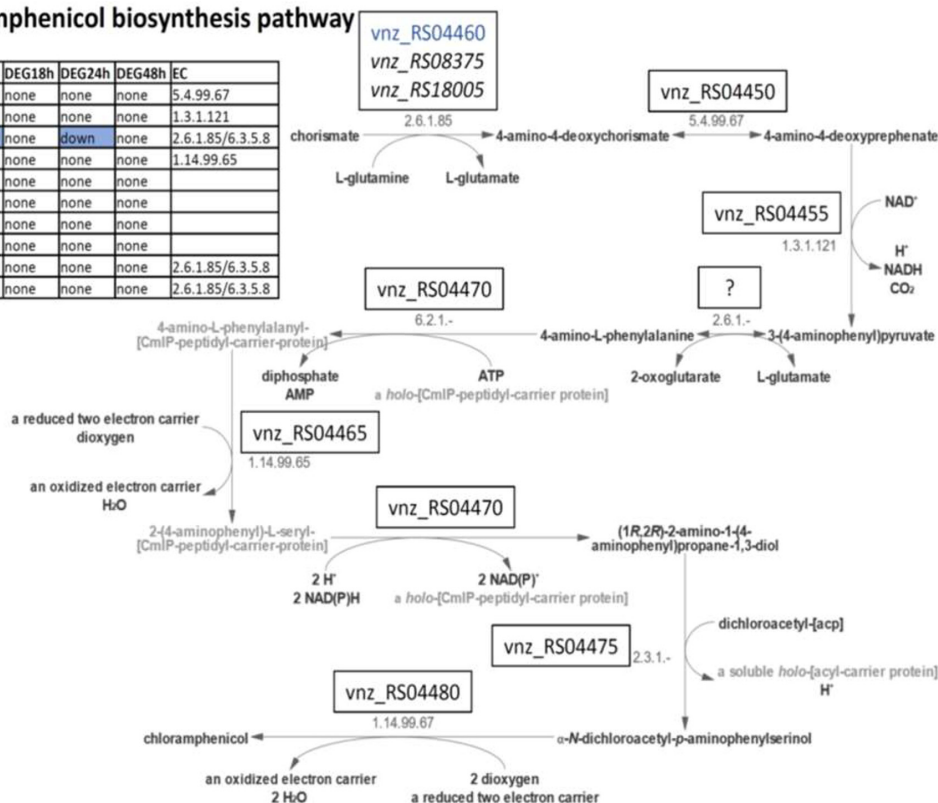
expressed; RS04460 is downregulated at 24 h and corresponds to the first step of the pathway.

Several central metabolic pathways are affected by ethanol shock. Besides the genes governing the biosynthesis of secondary metabolites, ethanol shock caused significant changes in the expression of genes involved in central metabolic pathways of *S. venezuelae* that are responsible for the generation of energy and primary metabolites. The tricarboxylic acid (TCA) cycle in *S. venezuelae*, schematically shown in Fig. 6 and, in more detail, in Fig. S11, includes a glyoxylate shunt. The glyoxylate shunt has been shown to be connected to oxidative and antibiotic stress response in various bacteria, but no clear understanding of its significance has yet been obtained (44). The gene expression profile shows an upregulation of the genes whose products are responsible for most parts of the TCA cycle at 18 h after the ethanol shock (genes RS12505, RS22220, RS22405, RS22565, RS22570, RS24745, RS26530, and RS30430). At 24 h, the expression profile changes drastically, as the genes for the reactions from succinate to malate are downregulated (genes RS01060 and RS34050) and most genes that were previously up-regulated are now not differentially expressed, leaving only RS12505, RS24745, and RS30430 expressed at a higher level. In addition, RS12505, whose product encodes citrate synthase responsible for the conversion of oxaloacetate to citrate, is upregulated at 18 and 24 h, while two other genes annotated as coding for “citrate synthase 2” (RS20910) and “bifunctional 2-methylcitrate synthase/citrate synthase” (RS27870) are not differentially expressed or downregulated, respectively. Most likely, the products of the genes RS20910 and RS27870 are not involved in the TCA. In the glyoxylate shunt, one of the genes involved in this pathway, namely, RS30430, is upregulated at 18 h and 24 h (Fig. 6, Table S1). At 48 h after ethanol shock, only a single gene involved in the TCA cycle is differentially expressed, RS24745, whose product is responsible for the conversion of 2-oxoglutarate to succinate semialdehyde. An upregulated TCA cycle, together with the glyoxylate shunt could thus provide increased amounts of NADH, NADPH, and

A

Chloramphenicol biosynthesis pathway

Ref_gene	DEG18h	DEG24h	DEG48h	EC
vnz_RS04450	none	none	none	5.4.99.67
vnz_RS04455	none	none	none	1.3.1.121
vnz_RS04460	none	down	none	2.6.1.85/6.3.5.8
vnz_RS04465	none	none	none	1.14.99.65
vnz_RS04470	none	none	none	
vnz_RS04475	none	none	none	
vnz_RS04480	none	none	none	
vnz_RS08375	none	none	none	2.6.1.85/6.3.5.8
vnz_RS18005	none	none	none	2.6.1.85/6.3.5.8



B

Chorismate biosynthesis

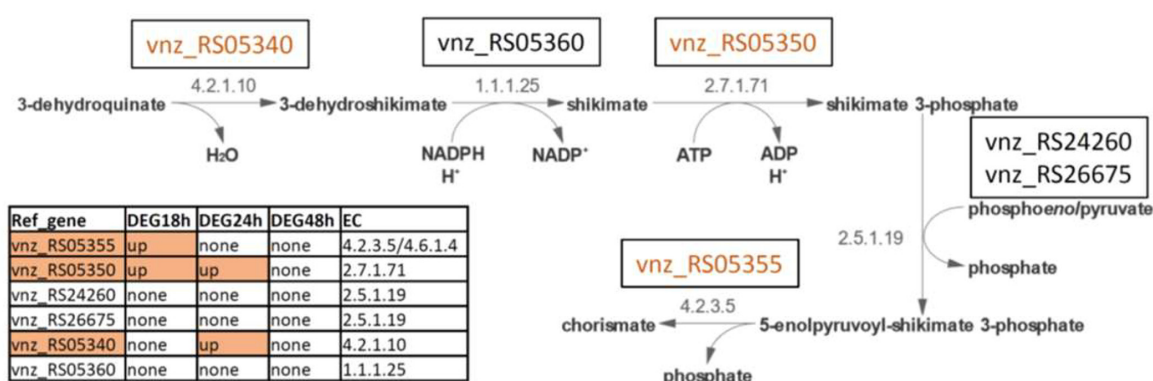


FIG 5 (A) Visualization of the chloramphenicol biosynthesis pathway according to MetaCyc (PWY-8032). (B) Visualization of the chorismate biosynthesis pathway according to MetaCyc (PWY-6163). Genes encoding enzymes involved in the pathways and their expression patterns at 3 time points after ethanol shock are shown. Downregulated genes are shown in blue and upregulated in orange.

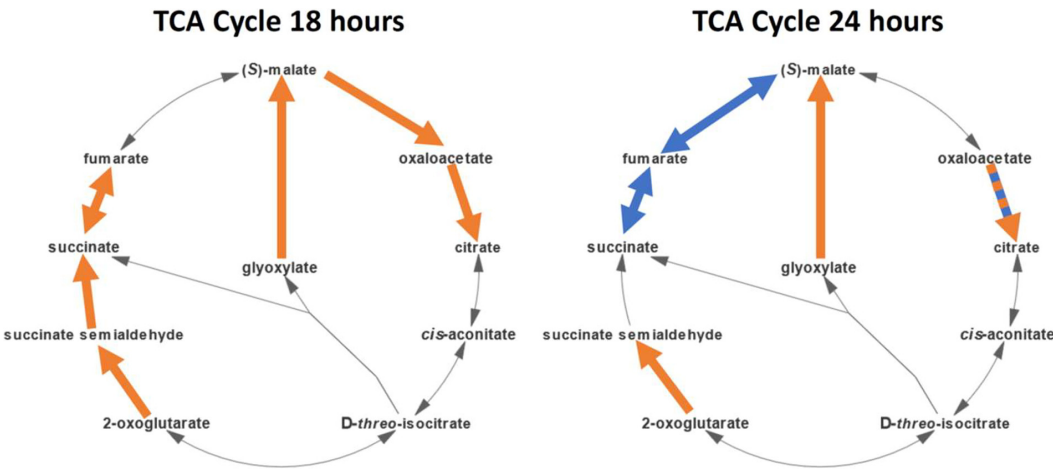


FIG 6 Reactions in the TCA cycle and glyoxylate shunt colored according to the gene expression state at 18 h and 24 h after ethanol shock. Reactions where at least one gene encoding an involved enzyme is upregulated are colored orange, and downregulated genes are colored blue. The reaction converting oxaloacetate to citrate can be catalyzed by the enzyme products of one up- and one downregulated gene and is colored both orange and blue.

coenzyme A, which are all known as cofactors of enzymes involved in the biosynthesis of secondary metabolites.

Several genes whose products catalyze reactions in gluconeogenesis (Fig. 7) were also found to be upregulated upon ethanol shock, with 9 genes (RS07775, RS07780, RS13425, RS14350, RS16930, RS22405, RS22865, RS29825, and RS36560) at 18 h after ethanol shock, dropping to 5 (RS07775, RS13425, RS14350, RS22865, and RS36560) at

Gluconeogenesis

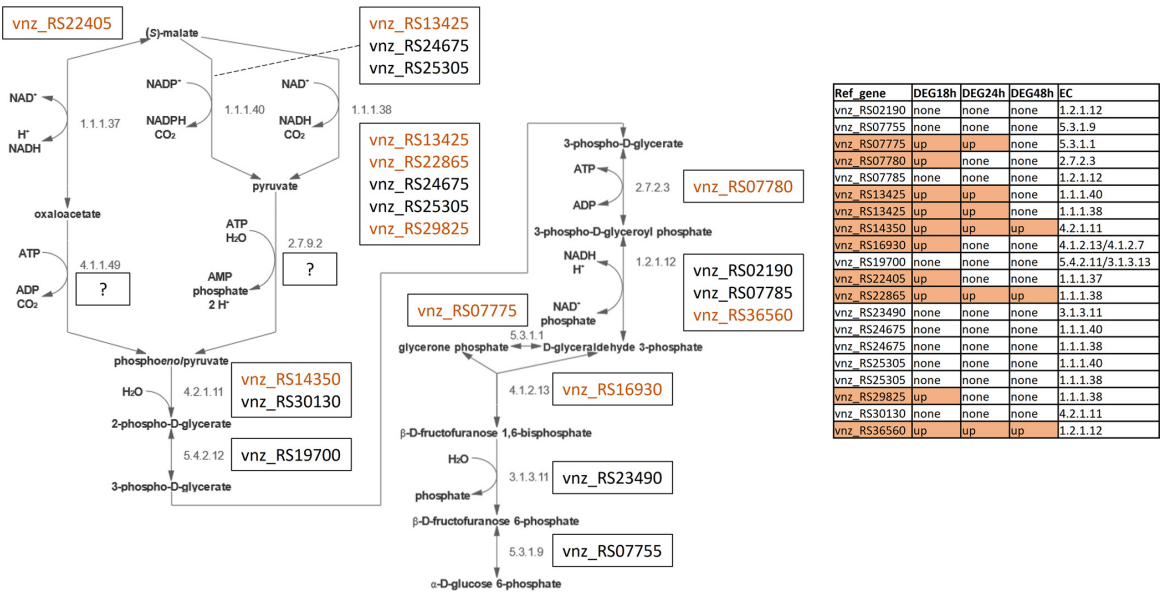


FIG 7 Visualization of the gluconeogenesis pathway according to MetaCyc (GLYCOLYSIS). *S. venezuelae* genes encoding enzymes involved in the pathways and their expression patterns at 3 time points after ethanol shock are shown. Genes are colored by their status of differential gene expression, namely, orange for higher expression and blue for lower expression.

Fatty Acid Oxidation

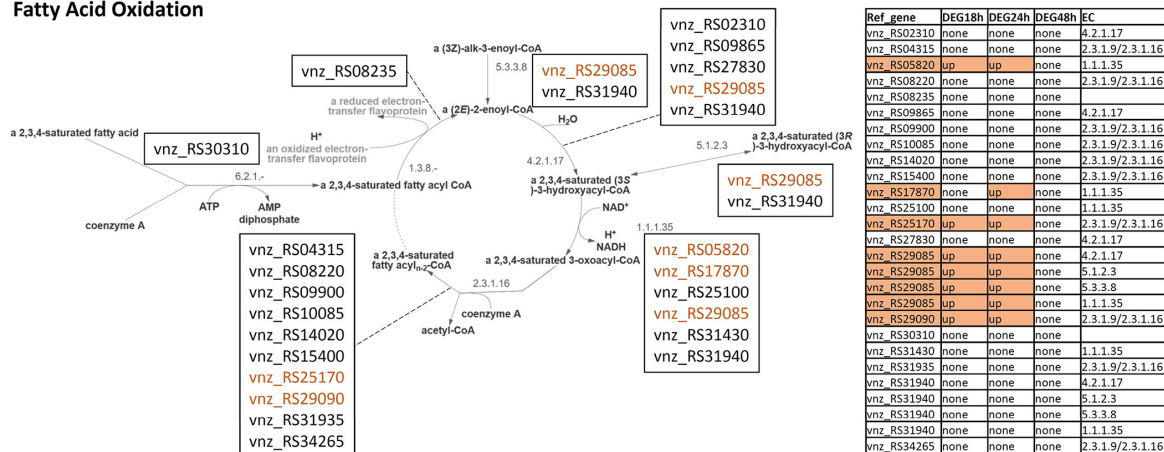


FIG 8 Visualization of the fatty acid oxidation pathway according to MetaCyc (FAO-PWY). *S. venezuelae* genes encoding enzymes involved in the pathways and their expression patterns at 3 time points after ethanol shock are shown. Genes are colored by their status of differential gene expression, namely, orange for higher expression and blue for lower expression.

24 h, and only 3 (RS14350, RS22865, and RS36560) at 48 h (Table S1). Gluconeogenesis is linked to the pentose phosphate pathway, which generates NADPH, an essential cofactor needed for anabolic processes, including the biosynthesis of secondary metabolites. Hence, upregulation of gluconeogenesis may boost secondary metabolism by supplying NADPH that can be utilized in pathways encoded by upregulated BGCs, such as the one for jadomycin biosynthesis (Fig. 4).

Acetyl-CoA is one of the central primary metabolites and is also used as a precursor or building block in the polyketide biosynthesis. The main catabolic pathway leading to acetyl-CoA is fatty acid oxidation, which according to MetaCyc involves up to 27 genes in *S. venezuelae* (Fig. 8). Four (RS05820, RS25170, RS29085, and RS29090)/five (RS17870) of these genes are upregulated at 18/24 h after the ethanol shock. The increased pool of acetyl-CoA may supply precursors for jadomycin biosynthesis, further boosting this process along with the overexpression of biosynthetic genes.

We also took a closer look at the expression profiles of the genes involved in the biosynthesis of chorismate, which is a precursor for both tryptophan and chloramphenicol, with chloramphenicol being overproduced upon ethanol shock (Fig. 1). Chorismate biosynthesis is accomplished via five reactions, starting from 3-dehydroquinate (Fig. 5B). Three of the five reactions are catalyzed by the enzymes encoded by genes whose expression is upregulated after ethanol shock. The upregulated genes are RS05355 at 18 h, assigned to the conversion of 5-enolpyruvyl-shikimate 3-phosphate to chorismate; RS05340 at 24 h, assigned to the reaction from 3-dehydroquinate to 3-dehydroshikimate; and RS05350 at 18 and 24 h, assigned to the reaction from shikimate to shikimate 3-phosphate. Considering this information, it is plausible to assume that the boost in chloramphenicol biosynthesis, in the absence of upregulation of its BGC, can be caused by the increased pool of chorismate. Accumulation of tryptophan upon ethanol shock can be caused by a reduced uptake of this amino acid from the medium, counteracted by an increased chorismate pool and overexpression of tryptophan biosynthesis genes to satisfy cellular demand for this amino acid. In the tryptophan pathway, 2 genes were found to be upregulated, namely, RS08410 (encoding phosphoribosyl anthranilate isomerase) at 18 h and RS08355 (encoding indole-3-glycerol phosphate synthase), at 24 h.

To have an indication of what the changes in the transcriptome might mean for the general cellular metabolism, we carried out a gene set enrichment analysis on several predicted pathways affected by ethanol shock (Fig. 9). The differentially expressed genes for each time point were used to generate a list of pathways up- or downregulated after the

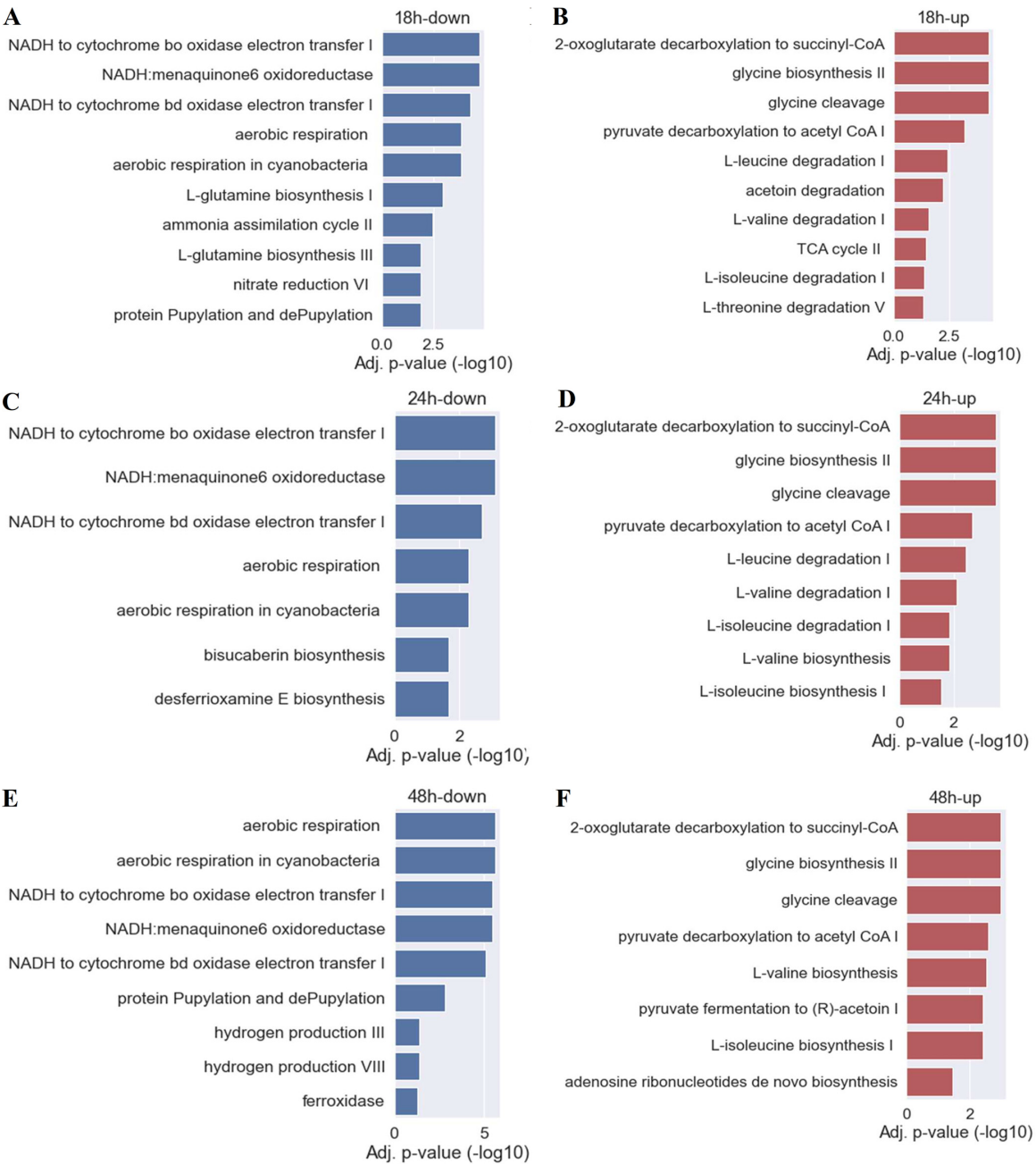


FIG 9 Enriched pathways in differentially expressed genes. Show pathways enriched in downregulated genes (A, C, E) and in upregulated genes (B, D, F), for 18 h, 24 h, and 48 h after ethanol shock. Only pathways with an adjusted *P* value (Benjamini-Hochberg correction) below 0.05 are shown.

ethanol shock. The number of enriched pathways ranges from 10 to 7. Across all time points, 2-oxo acid dehydrogenase complexes and amino acid metabolism pathways were upregulated (Fig. 9B, D, F). The only class of pathway enriched in downregulated genes across all samples was aerobic respiration (Fig. 9A, C, E). At 18 h, also the glutamine biosynthesis, ammonia assimilation, and nitrate reduction pathways were downregulated. The

related pathways for hydroxamate siderophores biosynthesis (bisucaberin and desferrioxamine) were downregulated at 24 h. Lastly, the protein Pupylation and dePupylation pathways were enriched at 18 h and 48 h, respectively. Notably, all genes involved in protein Pupylation and dePupylation (Pup: Prokaryotic Ubiquitin-like Protein) were statistically significantly downregulated at all three time points. That this pathway, containing only three genes, does not show up at 24 h is due to the higher number of genes being downregulated at this time point, leading to it not reaching the *P* value threshold for the enrichment. Protein Pupylation has been shown to be vital to oxidative stress response in *Streptomyces coelicolor*, while the proteasome was not (45, 46), and a *pup* deletion mutant was demonstrated to produce substantially smaller amounts of two pigmented antibiotics (46).

In conclusion, in this study, we demonstrated a complex response of *S. venezuelae* NRRL B-65442 to ethanol shock, including massive changes in the transcriptome which are reflected in the metabolome. We hypothesize that these changes are due to the alterations in the membrane via stress response mechanisms aimed at counteracting the membrane fluidization caused by ethanol. New insights in the stress response in *Streptomyces* bacteria, especially in connection with the secondary metabolite biosynthesis, will help to design new strategies for the metabolic engineering of these industrially important bacteria aimed at increasing the production of medically and agriculturally important natural products.

MATERIALS AND METHODS

Growth conditions, fermentation, and ethanol shock. *Streptomyces venezuelae* NRRL B-65442 was grown on ISP4 agar medium (Difco) at 28°C. Seeding cultures for fermentations were prepared by inoculation of 10 μ L of a spore suspension in 10 mL of liquid TSB medium (Ovoid) and incubation in 150-mL Erlenmeyer flasks for 16 h at 28°C with shaking (220 rpm/min). Next, 2.5 mL (10% vol/vol) of seed culture was used to start a 25-mL fermentation in MYM medium, containing 2.1 g/L MOPS (Sigma). Ethanol shock was performed after 7 h of cultivation by adding 1.5 mL (6% vol/vol) of absolute ethanol or 1.5 mL distilled sterile water (control) to the cultures. After ethanol shock, the fermentations continued for 18, 24, and 48 h. All the fermentations were performed in triplets.

Preparation of samples for RNA-seq. The fermentations were stopped after 18, 24, and 48 h. Next, 1.5-mL samples were immediately taken from each culture and transferred in sterile RNase-free tubes containing 3 mL of RNeasy Protect Bacterial reagent (Qiagen, Germany). The samples were vortexed for 10 sec, rested on a table for 5 min, transferred into 2.5-mL sterile RNase-free Eppendorf tubes and centrifuged for 12 min at 4,000 rpm and 4°C. Supernatants were removed and cells were frozen immediately at 80°C. One tube with the cells per sample was taken for RNA-seq analysis.

Preparation of extracts for LC-MS. After sampling for RNA-seq, the remaining cultures were freeze-dried and stored at -20°C until extraction. A total of 25 mL of methanol was added to the freeze-dried cultures and incubated for 60 min at a rotary shaker at 200 rpm. The mixtures were centrifuged for 10 min at 12,000 rpm, and the supernatants were transferred to new flasks. The methanol was removed using Rotavapor, and the dry residue was resuspended in 1 mL of methanol. A total of 100 μ L of the extracts was used for the LC-MS analyses.

LC-MS analyses and data interpretation. LC-MS analyses of the obtained extracts were performed on a Vanquish Horizon ultra-high-performance liquid chromatography (UHPLC) system (Thermo Fisher Scientific) coupled to the electrospray ionization (ESI) source of an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). Separation was carried out on an Acclaim 120 C₁₈, 2.1- by 150-mm, 3- μ m high-performance liquid chromatography (HPLC) column (Thermo Fisher Scientific) using water and acetonitrile, both modified with 0.1% formic acid, as mobile phase A and B, respectively. The sample components were separated and eluted with a linear gradient from 5% to 95% B for 45 min followed by an isocratic column cleaning (9.5 min at 95% B) and re-equilibration step (10 min at 5% B). The flow rate was 0.45 mL/min, and the column oven temperature was set to 25°C.

High-resolution ESI-MS spectra were recorded in positive ion mode with a Fourier-transform (FT) resolution of 60,000. High- or low-resolution ESI-MS/MS spectra of the three most intense precursor ions in each MS¹ spectrum were obtained in automated data-dependent acquisition mode using helium as the collision gas and the following settings: activation type, CID; isolation width, $\Delta m/z = 3$; normalized collision energy of 35.0; activation Q, 0.250; and activation time, 30 ms. For selected samples, negative ion mode data were recorded in addition to assist with the dereplication.

The sum formulas of the detected ions were determined using Thermo Xcalibur 4.1.31.9 Qual Browser based on the mass accuracy ($\Delta m/z, \leq 5$ ppm) and isotopic pattern. Dereplication was accomplished with the aid of GNPS Library Search (47), the Dictionary of Natural Products 30.2 (CRC Press, Taylor & Francis Group), and CAS SciFinder (American Chemical Society). MZmine 2 was used for untargeted and Skyline 21.1.0.146 for the more accurate targeted comparison of the peak intensities between different samples or groups of samples (48, 49). The statistical differences between groups were evaluated using the two-way analysis of variance (ANOVA) followed by the Sidák's multiple-comparison test (using Prism 9.4.1; GraphPad Software).

Differential gene expression analysis. Prinseq-lite (50) v0.20.4 is used to filter the raw reads (parameters: “trim_qual_right 28, min_len 75, min_qual_mean 25”). A public reference genome was retrieved from RefSeq under accession [NZ_CP018074.1](#) (*Streptomyces venezuelae* strain NRRL B-65442 genome), including the annotation information downloaded in a separate gff3 file. BWA-MEM (51) v0.7.16a with default parameters was used to align the reads to the reference. Aligned reads were separated from unaligned reads using SAMtools (52) sort (v1.15.1). From HTSeq (53) v2.0.1, the “htseq-count” function was used with the parameter “—nonunique=none” to count the aligned reads at each of the features in the gff3 reference file. Unaligned reads were analyzed with Kraken2 (54) v2.1.2, using the standard 8GB Kraken 2 Database built in 2019 from the NCBI REFSEQ bacteria, archaea, and viral libraries and the GRCh38 human genome (<https://benlangmead.github.io/aws-indexes/k2>) to explain their origin and check for contamination.

EdgeR (55) v3.36.0 and DESeq2 (56) v1.34.0 were used to predict differentially expressed genes. In EdgeR, the “TMM” method was used to normalize the expression values, and a quasi-likelihood F-test was performed to calculate *P* values. For DESeq2, the DESeq function was used with default parameters. All genes were analyzed by each of the two tools individually and considered differentially expressed with a fold change less than −0.6 or greater than 0.6 and a *P* value less than 0.05. The results were combined by considering only consensus predictions as differentially expressed, and conflicting results were deemed not differentially expressed.

Pathway analysis of biosynthetic gene clusters. Gapseq (57) was used to predict pathways and build a draft metabolic network from the reference genome. The genes from the RefSeq annotation were matched with genes predicted by gapseq via location. Genes without an exact location match were matched via overlap; further inspection ensured unique mapping and an overlap of >95% in both gapseq- and RefSeq-predicted genes.

Gene, reaction, and pathway associations were extracted from the generated draft metabolic network. Some reactions predicted by gapseq were not included in its metabolic network formulation database. Gene, reaction, and pathway associations for those genes were extracted from the reaction and pathway prediction output files of gapseq. Selection criteria were had a “good_blast” label and a blast bitscore of >200, similar to the parameters used in gapseq metabolic network creation.

The association of pathway identities (IDs) with biosynthetic gene clusters was performed by manual inspection of predicted reactions and pathways of genes located in biosynthetic gene clusters. Further reactions encoded by genes outside these clusters were found by searching for associations to the selected pathway IDs. Visualization of pathways was done using the pathway information and representation of the MetaCyc database and website (58).

Data availability. All RNA-seq reads generated and used in this study were deposited in the NCBI Sequence Read Archive under BioProject [PRJNA888945](#).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 2.7 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.6 MB.

ACKNOWLEDGMENTS

This study was supported by the Austrian Science Fund (FWF), grant no. 3098621.

We thank Anna Fabisikova for her assistance with LC-MS data acquisition and acknowledge support by the Mass Spectrometry Centre of the Faculty of Chemistry, University of Vienna.

REFERENCES

- van der Meij A, Worsley SF, Hutchings MI, van Wezel GP. 2017. Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiol Rev* 41:392–416. <https://doi.org/10.1093/femsre/fux005>.
- Urem M, Świątek-Polatyńska MA, Rigali S, van Wezel GP. 2016. Intertwining nutrient-sensory networks and the control of antibiotic production in *Streptomyces*. *Mol Microbiol* 102:183–195. <https://doi.org/10.1111/mmi.13464>.
- Tyc O, Song C, Dickschat JS, Vos M, Garbeva P. 2017. The ecological role of volatile and soluble secondary metabolites produced by soil bacteria. *Trends Microbiol* 25:280–292. <https://doi.org/10.1016/j.tim.2016.12.002>.
- Spagnolo F, Trujillo M, Dennehy JJ. 2021. Why do antibiotics exist? *mBio* 12:e0196621. <https://doi.org/10.1128/mBio.01966-21>.
- Mascher T. 2013. Signaling diversity and evolution of extracytoplasmic function (ECF) σ factors. *Curr Opin Microbiol* 16:148–155. <https://doi.org/10.1016/j.mib.2013.02.001>.
- Ulrich LE, Koonin EV, Zhulin IB. 2005. One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol* 13:52–56. <https://doi.org/10.1016/j.tim.2004.12.006>.
- Laub MT, Goulian M. 2007. Specificity in two-component signal transduction pathways. *Annu Rev Genet* 41:121–145. <https://doi.org/10.1146/annurev.genet.41.042007.170548>.
- Pinto D, Liu Q, Mascher T. 2019. ECF σ factors with regulatory extensions: the one-component systems of the σ universe. *Mol Microbiol* 112:399–409. <https://doi.org/10.1111/mmi.14323>.
- Honma S, Ito S, Yajima S, Sasaki Y. 2021. Nitric oxide signaling for actinorhodin production in *Streptomyces coelicolor* A3(2) via the DevS/R two-component system. *Appl Environ Microbiol* 87:e0048021. <https://doi.org/10.1128/AEM.00480-21>.
- McLean TC, Lo R, Tschowri N, Hoskisson PA, Al Bassam MM, Hutchings MI, Som NF. 2019. Sensing and responding to diverse extracellular signals: an updated analysis of the sensor kinases and response regulators of *Streptomyces* species. *Microbiology (Reading)* 165:929–952. <https://doi.org/10.1099/mic.0.000817>.
- López-García MT, Yagüe P, González-Quirón N, Rioseras B, Manteca A. 2018. The SCO4117 ECF sigma factor pleiotropically controls secondary metabolism and morphogenesis in *Streptomyces coelicolor*. *Front Microbiol* 9:312. <https://doi.org/10.3389/fmicb.2018.00312>.
- Doull JL, Singh AK, Hoare M, Ayer SW. 1994. Conditions for the production of jadomycin B by *Streptomyces venezuelae* ISP5230: effects of heat shock, ethanol treatment and phage infection. *J Ind Microbiol* 13:120–125. <https://doi.org/10.1007/BF01584109>.

13. Sekurova ON, Zhang J, Kristiansen KA, Zotchev SB. 2016. Activation of chloramphenicol biosynthesis in *Streptomyces venezuelae* ATCC 10712 by ethanol shock: insights from the promoter fusion studies. *Microb Cell Fact* 15:85. <https://doi.org/10.1186/s12934-016-0484-9>.
14. Xu G, Wang J, Wang L, Tian X, Yang H, Fan K, Yang K, Tan H. 2010. "Pseudo" gamma-butyrolactone receptors respond to antibiotic signals to coordinate antibiotic biosynthesis. *J Biol Chem* 285:27440–27448. <https://doi.org/10.1074/jbc.M110.143081>.
15. Robertson AW, Forget SM, Martinez-Farina CF, McCormick NE, Syvitski RT, Jakeman DL. 2016. JadX is a disparate natural product binding protein. *J Am Chem Soc* 138:2200–2208. <https://doi.org/10.1021/jacs.5b11286>.
16. Gomez-Escribano JP, Holmes NA, Schlimpert S, Bibb MJ, Chandra G, Wilkinson B, Buttner MJ, Bibb MJ. 2021. *Streptomyces venezuelae* NRRL B-65442: genome sequence of a model strain used to study morphological differentiation in filamentous actinobacteria. *J Ind Microbiol Biotechnol* 48:kuab035. <https://doi.org/10.1093/jimb/kuab035>.
17. Kim W, Lee N, Hwang S, Lee Y, Kim J, Cho S, Palsson B, Cho BK. 2020. Comparative genomics determines strain-dependent secondary metabolite production in *Streptomyces venezuelae* strains. *Biomolecules* 10:864. <https://doi.org/10.3390/biom10060864>.
18. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* 49:W29–W35. <https://doi.org/10.1093/nar/gkab335>.
19. Sidda D, Song L, Poon V, Al-Bassam M, Lazos O, Buttner MJ, Challis GL, Corre C. 2014. Discovery of a family of gamma-aminobutyrate ureas via rational derepression of a silent bacterial gene cluster. *Chem Sci* 5:86–89. <https://doi.org/10.1039/C3SC52536H>.
20. Inahashi Y, Zhou S, Bibb MJ, Song L, Al-Bassam MM, Bibb MJ, Challis GL. 2017. Watasemycin biosynthesis in *Streptomyces venezuelae*: thiazoline C-methylation by a type B radical-SAM methylase homologue. *Chem Sci* 8: 2823–2831. <https://doi.org/10.1039/c6sc03533g>.
21. Fernández-Martínez LT, Borsetto C, Gomez-Escribano JP, Bibb MJ, Al-Bassam MM, Chandra G, Bibb MJ. 2014. New insights into chloramphenicol biosynthesis in *Streptomyces venezuelae* ATCC 10712. *Antimicrob Agents Chemother* 58:7441–7450. <https://doi.org/10.1128/AAC.04272-14>.
22. Jones SE, Pham CA, Zambri MP, McKillip J, Carlson EE, Elliot MA. 2019. *Streptomyces* volatile compounds influence exploration and microbial community dynamics by altering iron availability. *mBio* 10:e00171-19. <https://doi.org/10.1128/mBio.00171-19>.
23. Hou Y, Braun DR, Michel CR, Klassen JL, Adnani N, Wyche TP, Bugni TS. 2012. Microbial strain prioritization using metabolomics tools for the discovery of natural products. *Anal Chem* 84:4277–4283. <https://doi.org/10.1021/ac202623g>.
24. Robertson AW, Martinez-Farina CF, Syvitski RT, Jakeman DL. 2015. Characterization of I-digitoxosyl-phenanthroviridin from *Streptomyces venezuelae* ISP5230. *J Nat Prod* 78:1942–1948. <https://doi.org/10.1021/acs.jnatprod.5b00277>.
25. de Koning CB, Ngwira KJ, Rousseau AL. 2020. Biosynthesis, synthetic studies, and biological activities of the jadomycin alkaloids and related analogues. *Alkaloids Chem Biol* 84:125–199. <https://doi.org/10.1016/bs.alkal.2020.02.001>.
26. Chen Y, Ntai I, Ju KS, Unger M, Zamdborg L, Robinson SJ, Doroghazi JR, Labeda DP, Metcalf WW, Kelleher NL. 2012. A proteomic survey of nonribosomal peptide and polyketide biosynthesis in actinobacteria. *J Proteome Res* 11:85–94. <https://doi.org/10.1021/pr2009115>.
27. Thanapipatsiri A, Gomez-Escribano JP, Song L, Bibb MJ, Al-Bassam M, Chandra G, Thamchaipenet A, Challis GL, Bibb MJ. 2016. Discovery of unusual biaryl polyketides by activation of a silent *Streptomyces venezuelae* biosynthetic gene cluster. *Chembiochem* 17:2189–2198. <https://doi.org/10.1002/cbic.201600396>.
28. Goto Y, Li B, Claesen J, Shi Y, Bibb MJ, van der Donk WA. 2010. Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. *PLoS Biol* 8:e1000339. <https://doi.org/10.1371/journal.pbio.1000339>.
29. Liu C, Cao J, Zhang H, Yin J. 2022. Evolutionary history of RNA modifications at N6-adenosine originating from the R-M system in eukaryotes and prokaryotes. *Biology (Basel)* 11:214. <https://doi.org/10.3390/biology11020214>.
30. Vargas-Blanco DA, Shell SS. 2020. Regulation of mRNA stability during bacterial stress responses. *Front Microbiol* 11:2111. <https://doi.org/10.3389/fmicb.2020.02111>.
31. Deng X, Chen K, Luo GZ, Weng X, Ji Q, Zhou T, He C. 2015. Widespread occurrence of N6-methyladenosine in bacterial mRNA. *Nucleic Acids Res* 43:6557–6567. <https://doi.org/10.1093/nar/gkv596>.
32. Nemet I, Saha PP, Gupta N, Zhu W, Romano KA, Skye SM, Cajka T, Mohan ML, Li L, Wu Y, Funabashi M, Ramer-Tait AE, Naga Prasad SV, Fiehn O, Rey FE, Tang WHW, Fischbach MA, DiDonato JA, Hazen SL. 2020. A cardiovascular disease-linked gut microbial metabolite acts via adrenergic receptors. *Cell* 180:862–877.e22. <https://doi.org/10.1016/j.cell.2020.02.016>.
33. van Bokhorst-van de Veen H, Abee T, Tempelaars M, Bron PA, Kleerebezem M, Marco ML. 2011. Short- and long-term adaptation to ethanol stress and its cross-protective consequences in *Lactobacillus plantarum*. *Appl Environ Microbiol* 77:5247–5256. <https://doi.org/10.1128/AEM.00515-11>.
34. Maitre M, Weidmann S, Dubois-Brissonnet F, David V, Covès J, Guzzo J. 2014. Adaptation of the wine bacterium *Oenococcus oeni* to ethanol stress: role of the small heat shock protein Lo18 in membrane integrity. *Appl Environ Microbiol* 80:2973–2980. <https://doi.org/10.1128/AEM.04178-13>.
35. Rubin E, Rottenberg H. 1982. Ethanol-induced injury and adaptation in biological membranes. *Fed Proc* 41:2465–2471.
36. Sajbidor J. 1997. Effect of some environmental factors on the content and composition of microbial membrane lipids. *Crit Rev Biotechnol* 17: 87–103. <https://doi.org/10.3109/07388559709146608>.
37. Seydlová G, Halada P, Fišer R, Toman O, Ulrych A, Svobodová J. 2012. DnaK and GroEL chaperones are recruited to the *Bacillus subtilis* membrane after short-term ethanol stress. *J Appl Microbiol* 112:765–774. <https://doi.org/10.1111/j.1365-2672.2012.05238.x>.
38. Török Z, Horváth I, Goloubinoff P, Kovács E, Glatz A, Balogh G, Vigh L. 1997. Evidence for a lipochaperonin: association of active protein-folding GroEL oligomers with lipids can stabilize membranes under heat shock conditions. *Proc Natl Acad Sci U S A* 94:2192–2197. <https://doi.org/10.1073/pnas.94.6.2192>.
39. Müller M, Reiß S, Schlüter R, Mäder U, Beyer A, Reiß W, Marles-Wright J, Lewis RJ, Pförtner H, Völker U, Riedel K, Hecker M, Engelmann S, Pané-Farré J. 2014. Deletion of membrane-associated Asp23 leads to upregulation of cell wall stress genes in *Staphylococcus aureus*. *Mol Microbiol* 93: 1259–1268. <https://doi.org/10.1111/mmi.12733>.
40. Barros EM, Martin MJ, Selleck EM, Lebreton F, Sampaio JLM, Gilmore MS. 2019. Daptomycin resistance and tolerance due to loss of function in *Staphylococcus aureus* *dsp1* and *asp23*. *Antimicrob Agents Chemother* 63: e01542-18. <https://doi.org/10.1128/AAC.01542-18>.
41. Asai K. 2018. Anti-sigma factor-mediated cell surface stress responses in *Bacillus subtilis*. *Genes Genet Syst* 92:223–234. <https://doi.org/10.1266/ggs.17-00046>.
42. Chang Z, Sun Y, He J, Vining LC. 2001. p-Aminobenzoic acid and chloramphenicol biosynthesis in *Streptomyces venezuelae*: gene sets for a key enzyme, 4-amino-4-deoxychorismate synthase. *Microbiology (Reading)* 147:2113–2126. <https://doi.org/10.1099/00221287-147-8-2113>.
43. Sato S, Kudo F, Rohmer M, Eguchi T. 2020. Characterization of radical SAM adenosylhopane synthase, HpnH, which catalyzes the 5'-deoxyadenosyl radical addition to diploptene in the biosynthesis of C35 bacteriohopanepolyols. *Angew Chem Int Ed Engl* 59:237–241. <https://doi.org/10.1002/anie.201911584>.
44. Dolan SK, Welch M. 2018. The glyoxylate shunt, 60 years on. *Annu Rev Microbiol* 72:309–330. <https://doi.org/10.1146/annurev-micro-090817-062257>.
45. Compton CL, Fernandopulle MS, Nagari RT, Sello JK. 2015. Genetic and proteomic analyses of pupylation in *Streptomyces coelicolor*. *J Bacteriol* 197:2747–2753. <https://doi.org/10.1128/JB.00302-15>.
46. Boubakri H, Seghezzi N, Duchateau M, Gominet M, Kofronová O, Benada O, Mazodier P, Pernodet JL. 2015. The absence of pupylation (prokaryotic ubiquitin-like protein modification) affects morphological and physiological differentiation in *Streptomyces coelicolor*. *J Bacteriol* 197:3388–3399. <https://doi.org/10.1128/JB.00591-15>.
47. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crusemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrew K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Siebottom AM, Sedio BE, et al. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 34:828–837. <https://doi.org/10.1038/nbt.3597>.
48. Pluskal T, Castillo S, Villar-Briones A, Oresic M. 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11:395. <https://doi.org/10.1186/1471-2105-11-395>.

49. Adams KJ, Pratt B, Bose N, Dubois LG, St John-Williams L, Perrott KM, Ky K, Kapahi P, Sharma V, MacCoss MJ, Moseley MA, Colton CA, MacLean BX, Schilling B, Thompson JW, Alzheimer's Disease Metabolomics Consortium. 2020. Skyline for small molecules: a unifying software package for quantitative metabolomics. *J Proteome Res* 19:1447–1458. <https://doi.org/10.1021/acs.jproteome.9b00640>.
50. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863–864. <https://doi.org/10.1093/bioinformatics/btr026>.
51. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv 1303.3997v2 [q-bio.GN]*. <https://arxiv.org/abs/1303.3997>.
52. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
53. Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169. <https://doi.org/10.1093/bioinformatics/btu638>.
54. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>.
55. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140. <https://doi.org/10.1093/bioinformatics/btp616>.
56. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
57. Zimmermann J, Kaleta C, Waschina S. 2021. gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biol* 22:81. <https://doi.org/10.1186/s13059-021-02295-1>.
58. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S, Subhraveti P, Karp PD. 2020. The MetaCyc database of metabolic pathways and enzymes—a 2019 update. *Nucleic Acids Res* 48:D445–D453. <https://doi.org/10.1093/nar/gkz862>.

A.1.4. Supplementary data

Supplementary data are available under:

https://journals.asm.org/doi/suppl/10.1128/spectrum.03672-22/suppl_file/spectrum.03672-22-s0001.pdf

https://journals.asm.org/doi/suppl/10.1128/spectrum.03672-22/suppl_file/spectrum.03672-22-s0002.xlsx

A.2. Three conserved extracytoplasmic function sigma factors in *Streptomyces* have a major impact on secondary metabolism

A.2.1. Publication status

Submitted for publication, under revision: Sekurova, O. N., Zehl, M., Predl, M., Hunyadi, P., Rattei, T., Zotchev, S. B., *Three conserved extracytoplasmic function sigma factors in *Streptomyces* have a major impact on secondary metabolism*. Submitted for publication in Microbial Cell Factories, confirmation of receipt: 30.03.2024

A.2.2. Author contributions

The author of this dissertation planned carried out the sequencing data analysis. He further planned and performed the integration of the metabolic network with the gene expression data, as well as its analysis. The author also produced figures and wrote parts of the manuscript.

A.2.3. Publication

Three conserved extracytoplasmic function sigma factors in *Streptomyces* have a major impact on secondary metabolism

Olga N. Sekurova¹, Martin Zehl², Michael Predl^{3,4}, Peter Hunyadi³, Thomas Rattei^{3,4}, Sergey B. Zotchev^{1*}

¹Department of Pharmaceutical Sciences, Division of Pharmacognosy, University of Vienna, 1090 Vienna, Austria

²Department of Analytical Chemistry, Faculty of Chemistry, University of Vienna, 1090 Vienna, Austria

³Centre for Microbiology and Environmental Systems Science, University of Vienna, 1030 Vienna, Austria

⁴Doctoral School in Microbiology and Environmental Science, University of Vienna, 1030 Vienna, Austria

Key words: *Streptomyces* bacteria, transcriptomics, extracytoplasmic function sigma factors, secondary metabolism

Corresponding author: Sergey B. Zotchev, Department of Pharmaceutical Sciences, Division of Pharmacognosy, University of Vienna, 1090 Vienna, Austria. Email: sergey.zotchev@univie.ac.at

Abstract

Background. Ethanol shock significantly affects expression of over 1200 genes in *Streptomyces venezuelae* NRRL B-65442, including those for secondary metabolite biosynthesis and a cryptic gene *pepX* encoding a 19-amino acid peptide with unknown function.

Results. To establish a possible correlation between PepX peptide and secondary metabolism in *S. venezuelae* its gene was deleted, followed by analyses of the transcriptome and secondary metabolome of the mutant. Although the secondary metabolome of the *pepX* mutant was not affected very strongly, *pepX* deletion, similarly to ethanol shock, mostly led to downregulated expression of secondary metabolite biosynthesis gene clusters (BGCs). At the same time, there was a reverse correlation between the expression of certain extracytoplasmic function sigma factors (ECFs) and several BGCs. Individual deletions of three selected ECF-coding genes conserved in *Streptomyces* that were upregulated upon both *pepX* deletion and ethanol shock had a profound positive effect on the expression of BGCs, which also correlated with overproduction of specific secondary metabolites. Deletion of one such ECF-coding gene in a marine sponge-derived *Streptomyces* sp. also significantly altered the secondary metabolite profile, thus confirming the role of this ECF in regulation of secondary metabolism.

Conclusions. These findings pave the way to activation or upregulation of BGCs in *Streptomyces* bacteria harboring genes for ECFs homologous to those identified in this study, hereby assisting in the discovery of novel bioactive secondary metabolites.

Background

Most bacteria live in complex environments such as soils, marine sediments, plant and animal hosts, where they constantly encounter multiple stressors in the form of living competitors for nutritional sources, predators, as well as abiotic factors such as shifts in temperature, salinity, pH etc. To cope with these challenges, bacteria have developed several regulatory systems that respond to particular environmental cues and adequately change gene expression patterns with the aim to quickly adapt to the new situation. One of such response systems is represented by an array of extracytoplasmic function sigma factors (ECFs) and their cognate anti-sigma factors whose genes have been identified in all sequenced bacterial genomes (1). Sigma factors represent the essential parts of the RNA polymerase holoenzyme that bind promoter regions and recruit the rest of the enzyme subunits to the transcription start site. Unlike housekeeping sigma factors, such as σ^{70} , which recognize promoters of genes having fundamental roles in functioning of the bacterial cells (e.g. DNA replication, protein and cell wall syntheses, etc.), ECFs represent the most diverse group of sigma factors that ensure expression of genes whose products are only needed in particular circumstances. The specificity of ECFs is supported by their ability to recognize only the -10 and -35 promoter regions with rather invariable spacing, while housekeeping sigma factors require in addition an upstream and extended -10 regions to efficiently bind the promoter (2). Typically, ECFs are bound to anti-ECFs, which are often tethered to the bacterial membrane, and thus are unable to bind promoters. This ECF sequestration is relieved upon proteolytic digestion or conformational change of anti-ECFs that may be caused by e.g. proteases produced by other bacteria or environmental stress affecting membrane integrity.

Bacteria of the genus *Streptomyces* are ubiquitous with regard to environments they dwell in, and have evolved a large number of ECF-anti-ECF pairs to deal with environmental cues. Genomes of *Streptomyces* bacteria usually encode over 50 sigma factors, most of them being ECFs, some of which are involved in regulation of secondary metabolism. Many secondary metabolites produced by *Streptomyces* are of medical importance, such as antibacterial, antifungal, antitumor and immunosuppressive agents (3). Insights into the correlation between environmental stress and secondary metabolite biosynthesis may yield new strategies for increasing the yields and production of novel compounds, whose cognate biosynthesis gene clusters (BGCs) are not expressed in standard conditions. ECFs can play different and complex roles in controlling secondary metabolite biosynthesis. Biosynthesis of 3-formamidosalicylate, the starter unit for antimycin biosynthesis, is positively regulated by ECF^{AntA} encoded within the antimycin BGC, which is antagonized by the ClpXP-protease system in *Streptomyces albus* S4 (4). On the contrary, Rebets et al. (5) have demonstrated that deletion of certain ECF-coding genes in *Streptomyces lividans* TK24 results in overproduction of the polyketide actinorhodin and the siderophore coelichelin (5). Although in the case of antimycin the dependence of gene expression on a specific sigma factor provides a straightforward explanation for the activator role of the ECF^{AntA}, the positive effect of ECF gene deletions on secondary metabolism may involve several possible scenarios. For example, such ECFs may control the expression of negative regulators suppressing particular BGCs, or

they themselves compete for promoter regions of BGC-situated genes with housekeeping sigma factors, thereby preventing the latter from binding (6).

In this work, we investigated the effects of several genes whose expression was affected by ethanol shock in *Streptomyces venezuelae* NRRL B-65442 by constructing deletion mutants and exploring their transcriptomes and secondary metabolomes. In particular, we identified three genes encoding ECF sigma factors conserved in *Streptomyces* bacteria, which differentially suppress expression of specific BGCs, and their individual deletions lead to significant increase in production of certain secondary metabolites. Similar results for at least one such ECF-coding gene were obtained for non-model marine sponge-derived *Streptomyces* sp. ADI96-15.

Materials and methods

Growth conditions, fermentations and ethanol shock

Growth conditions for *Streptomyces venezuelae* NRRL B-65442 and deletion mutants were as described in Sekurova et al. (7). All fermentations were performed in triplicates. Methanolic culture extracts were prepared according to Sekurova et al. (7)

Construction of plasmids and deletion mutants

Vector pSOK910. Suicide vector pSOK910 for introduction gene deletions into the streptomycetes chromosome via homologous recombination was constructed by ligation of synthetic part ermE*p-*gusA* (reporter gene for β -glucuronidase under control of a strong constitutive promoter ermE*p) with 3kb *HindIII*-*EcoRI* fragment from pSOK201 (8).

Vector pKNOCK. Suicide vector pKNOCK was used for deletion of gene Sig04314 from the chromosome of *Streptomyces* sp. ADI96-15. It was constructed by ligation of synthetic *sfp* gene (*Bacillus subtilis* phosphopantetheinyl transferase) with part of vector pPCV301 (Figure S1, Supplementary Information) cut with *HindIII* + *XbaI*, and containing reporter gene *bpsA* (encodes indigoidine synthetase and responsible for the production of blue pigment), *aac(3)-IV* gene for apramycin resistance and RP4 origin for ss DNA transfer.

pepX deletion mutant. The vector pPXD910 for *pepX* gene deletion was constructed by PCR amplification of flanking regions for *pepX* gene, using primers PEXDF1, PEXDR1 and PEXDF2, PEXDR2 (Table S1, Supplementary Information). Amplicon 1 was cut with *EcoRI*+*XbaI*, amplicon 2 - with *BamHI*+*XbaI*, and these fragments were ligated together with the vector pSOK910 cut with *EcoRI*+*XbaI*. The final construct pPXD910 was introduced into *S. venezuelae* via conjugation. The first crossover mutant went through 4 passages of growth in liquid TSB culture without antibiotic, and was plated for sporulation. The spore suspension and its serial dilutions then have been prepared. 100 μ l of -6 dilution was plated on 10 ISP4 +X-Gluc (5-bromo-4-chloro-3-indolyl- β -D-glucuronide cyclohexylammonium salt). After several days blue and white colonies appeared on the plates. White clones, which no longer harbor the reporter gene *gusA* (second crossover candidates), were selected. Genomic DNA from these clones were isolated and used as a template for PCR to confirm the deletion of *pepX* gene (primers PepXdel-F and PepXdel-R in the Table S1, Supplementary Information).

Construction of ECF sigma factors-encoding genes deletion mutants in *S. venezuelae*. The vectors pECF200, pECF300 and pECF501 for the deletion of ECF genes from *S. venezuelae* genome were constructed by PCR amplifications of flanking regions for corresponding genes, using primers ECF2delA-F and ECF2delA-R for flankA, and ECF2delB-F and ECF2delB-R for flank B for ECF200 deletion; primers ECF3delA-F, ECFdelA-R, ECF3delB-F and ECF3delB-R for ECF300 deletion, and ECF5delA-F, ECF5delA-R, ECF5delB-F and ECF5delB-R for ECF500 deletion (Table S1, Supplementary Information). All “A” amplicons were cut with *Pst*I+*Xba*I, all “B” amplicons -with *Pst*I+*Eco*RI. Cut fragments were ligated in pairs A+B into pSOK910 cut *Xba*I+*Eco*RI. The final constructs have been introduced into *S. venezueale* by conjugation. First crossover mutant as a result of homologous recombination went through several rounds of growth in liquid TSB culture without antibiotic, and were plated for sporulation. Serial dilutions of spore suspension were prepared, and -5 dilution plated on 10 ISP4+X-Gluc. After several days white and blue colonies appeared on the plates, and genomic DNA from selected white colonies, as the candidates for second crossover mutants, have been extracted. The following PCR with primers ECF2_Fw and ECF2_Rv, ECF3_Fw and ECF3Rv, ECF5_Fw and ECF5Rv, were used to confirm the second crossover event in the ECF2, ECF3 and ECF5 mutant strains, respectively.

Construction of deletion mutant for ECF gene Sig04314 in *Streptomyces* sp. ADI 96-15. The vector for deletion of gene 04314 was constructed by PCR amplifications of regions flanking the gene, using primers Sig04314A-F, Sig04314A-R for A flank, and Sig04314B-F, Sig04314B-R for the B (see Table S1, Supplementary Information). Flank A was then cut with *Age*I+*Pst*I, flank B- with *Pst*I+*Xba*I. Fragments A and B after enzyme restriction were ligated together with 7,3kb part from pKNOCK, digested by *Age*I+*Xba*I. The final construct pSig04314 was introduced into the *Streptomyces* sp. ADI 96-15 by conjugation. First crossover mutant passed several rounds of growth in liquid TSB cultures, and was plated for sporulation. The -5 dilution of spore suspension have been plated on 10 SFM plates. Colorless colonies were chosen as second crossover candidates. Their genomic DNA was extracted and used for PCR to confirm the deletion of the gene. Primers 04314delCheck-F and 04314delCheck-R (see Table S1, Supplementary Information) have been used in these reactions.

RNA sequencing

Cultures grown in MYM medium for 18, 24 and 48 h past ethanol shock, or without one, were used to prepare samples for RNA sequencing according to Sekurova et al. (7). RNA isolation, preparation of cDNA and sequencing were performed at LGC Genomics (Germany).

Differential gene expression and bioinformatics-based protein analyses

Raw reads were cleaned with prinseq-lite v0.20.4 (9) using the following parameters: trim_qual_right 28, min_len 75, min_qual_mean 25. Cleaned reads are mapped to a reference genome of *S. venezuelae* NRRL B-65442 (NZ_CP018074.1) using BWA-MEM v0.7.16a (10) with default parameters. Aligned reads were filtered with SAMtools sort v1.15.1 (11) and used to calculate feature counts using htseq-count of HTSeq v2.0.1 (12) with parameter nonunique=none. Contamination was estimated using unaligned reads (filtered with

SAMtools) and Kraken2 v2.1.2 (13) with the standard 8GB Kraken 2 Database (built in 2019). Calling of differential expression was performed by EdgeR v3.36.0 (14) (TMM normalization method and quasi-likelihood F-test) and DESeq2 v1.34.0 (15) (default parameters), considering only genes with absolute fold change greater than 0.6 and P-value less than 0.5 differentially expressed. The results of both methods were combined on consensus principle, with genes only being classified as differentially expressed, if both methods agree.

Analysis of proteins for the presence of specific domains and transmembrane helices was carried out using online tools MOTIF Search (<https://www.genome.jp/tools/motif/>) and DeepTMHMM (<https://dtu.biolib.com/DeepTMHMM>), respectively.

Pathway analysis for secondary and primary metabolism

Gene function and pathway presence were predicted with gapseq v1.1 (16), based on the reference genome of *S. venezuelae* strain NRRL B-65442. Genes predicted by gapseq were matched with reference genome loci by location under the conditions of more than 95% sequence overlap and unique matching. Associations of genes to reactions and pathways were selected based on the following criteria: Either gene, reaction, pathway associations were present in the draft metabolic model generated by gapseq, or such associations had a blast bitscore greater than 200 and were assigned a “good_blast” label in the functional prediction step of gapseq.

These gene-pathway associations were combined with their differential gene expression status to conduct gene set enrichment analysis. The analysis was performed with the GSEApv package and enrichr (17).

LC-MS analyses and data interpretation

The extracts from *S. venezuelae* NRRL B-65442 wild-type and mutants were analyzed by an untargeted LC-MS method with an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) exactly as described previously (7). The identification of the metabolites of interest is also reported in detail in Sekurova et al. (7). The targeted comparison of the peak intensities between different samples or groups of samples was performed with Skyline 21.1.0.146 (18). The statistical differences between groups were evaluated using either the two-way analysis of variance (ANOVA) followed by the Tukey's multiple comparisons test or the one-way ANOVA followed by the Dunnett's multiple comparisons test (using Prism 9.5.1; GraphPad Software).

LC-MS analyses of the MeOH extracts from *Streptomyces* sp. ADI96-15 wild-type and knock-out strain cultures were performed on a Vanquish Horizon UHPLC system (Thermo Fisher Scientific) coupled to the ESI source of a timsTOF fleX mass spectrometer (Bruker Daltonics). Separation was carried out on an Acquity Premier HSS T3 column, 2.1 x 150 mm, 1.8 μ m (Waters) using water and acetonitrile/water 9:1, both modified with 0.1% formic acid, as mobile phase A and B, respectively. The sample components were separated and eluted with a gradient starting with a linear increase from 0% to 20% B in 10 min, followed by a linear increase from 20% to 100% B in 25 min, and finally an isocratic column cleaning (4

min at 100% B) and re-equilibration step (6 min at 0% B). The flow rate was 0.5 mL/min and the column oven temperature was set to 40 °C.

High-resolution ESI-MS and MS/MS spectra were recorded in positive ion mode in the range of m/z 100-2500. CID spectra of the five most intense precursor ions in each MS¹ spectrum were obtained in automated data-dependent acquisition mode using nitrogen as collision gas. The sum formulas of the detected ions were determined using Bruker Compass DataAnalysis 5.3 based on the mass accuracy ($\Delta m/z \leq 5$ ppm) and isotopic pattern matching (SmartFormula algorithm). Dereplication was accomplished with the aid of The Natural Products Atlas (19), GNPS Library Search (20), and CAS SciFinder (American Chemical Society). MZmine 3 was used for comparing peak intensities between different samples or groups of samples (21). The peak areas of the dereplicated compounds were determined using Skyline 23.1.0.268 (18) and the statistical differences between the $\Delta sig04314$ mutant and wild-type of *Streptomyces* sp. ADI96-15 was evaluated using an unpaired t test assuming Gaussian data distribution using Prism 10.0.3 (GraphPad Software).

Results

Effect of the small ribosomally synthesized peptide PepX on the metabolome and transcriptome of *S. venezuelae*.

In our previous study focused on the effect that ethanol shock had on the transcriptome and secondary metabolome of *Streptomyces venezuelae* we identified a conserved 19-amino acid peptide PepX produced in copious amounts in normal conditions but drastically downregulated by ethanol shock (7). Since this effect correlated with both up- and downregulation of certain BGCs, we hypothesized that PepX might play a role in controlling secondary metabolism. To test this hypothesis, the *pepX* gene was deleted and the resulting mutant studied with respect to gene expression patterns and secondary metabolite biosynthesis with and without ethanol shock in comparison with the wild type strain. Only a few differences in the secondary metabolomes of the wild type strain and the $\Delta pepX$ mutant could be detected (Figure 1). In particular, the *pepX* deletion mutant showed a significantly lower production of the closely related hydroxamate siderophores desferrioxamine B and legonoxamine A, while levels of foroxymithine, another hydroxamate siderophore, were not altered. This effect was observed at all three time points (18, 24 and 48 h), but only in the absence of ethanol shock, which largely depletes these compounds in both the $\Delta pepX$ mutant and the wild type strains. Conversely, the concentrations of the 2-hydroxyphenylthiazoline congeners, such as thiazostatins A and B, were significantly higher after 48 h of cultivation of the $\Delta pepX$ mutant in the absence of ethanol shock. Biosyntheses of all other identified secondary metabolites were either not affected or the biological variation in their production levels were too high to enable detection of a statistically significant influence of the *pepX* deletion. The same holds true for all other metabolites previously shown to be up- or downregulated by the ethanol shock, such as threonylcarbamoyladenine or *N*-phenylacetylglutamine, as well as for selected metabolites that were not affected by ethanol shock, such as diketopiperazines and biotin (7). Clearly, on the level of secondary metabolites,

the response to ethanol shock was nearly identical for the wild type strain and the $\Delta pepX$ mutant, ruling out the possibility that PepX is a major mediator of this effect.

With respect to the wild type *S. venezuelae*, the $\Delta pepX$ mutant responded differently to the ethanol shock in terms of gene expression patterns. 417 genes were down- and 354 upregulated by the ethanol shock in the $\Delta pepX$ mutant at the 18 h time point post ethanol shock, while in the wild type strain these numbers were 431 and 679, respectively. It must be noted that the overlap between differentially expressed genes in the mutant and the wild type subjected to ethanol shock was significant, but not complete. The numbers of down/upregulated genes in the $\Delta pepX$ mutant were 349/308 and 327/213 at 24 h and 48 h post ethanol shock, respectively, while the corresponding numbers for the wild type strain were 496/579 and 308/305. It is worth noting that in the case of the $\Delta pepX$ mutant subjected to ethanol shock, the time-dependent decline in the number of downregulated genes was much steeper compared to the decline of upregulated genes. It appears thus that the effect of ethanol shock on gene expression in the $\Delta pepX$ mutant is milder relative to the wild type strain.

The RNAseq-based transcriptomics also revealed a significant effect of *pepX* deletion on gene expression patterns even in the absence of ethanol shock. Compared to the wild type strain, 224 genes were up- and 486 genes were downregulated in the $\Delta pepX$ mutant at 18 h time point, changing to 215 up/332 down at 24 h and 90 up/129 down at 48 h. According to the transcriptome data for 18 hours, which revealed the most profound changes, genes within 7 BGCs (including those for desferrioxamines, jadomycins and foroxymithins) were downregulated and genes in 3 BGCs (including those for watasemycins and chloramphenicol) were upregulated in the $\Delta pepX$ mutant as compared to the wild type strain (Table 1). This largely fits the LC-MS data on cognate secondary metabolites, although the observed differences in the levels of L-digitoxosyl-phenanthroviridin (jadomycin precursor) and chloramphenicol did not reach a confidence level > 95% (Figure 1). This picture to a certain extent resembled the response of the wild type strain to ethanol shock (15 BGCs down- and 3 upregulated). In particular, expression of genes for the biosynthesis of chloramphenicol (BGC9), desferrioxamine (BGC11), cryptic peptide-polyketide hybrid (BGC26), ladderane (BGC27) and foroxymithins (BGC33) was affected in the same way by the *pepX* deletion as by the ethanol shock in the wild type strain.

Apart from the BGCs, we also analyzed the impact of the *pepX* deletion on the primary metabolism-related transcriptome. While most genes coding for enzymes in primary metabolism did not exhibit any significant change in expression, two pathways were strongly affected. In the arginine biosynthesis pathway, 3 and 4 genes were upregulated at the 18h and 24h time points, respectively, under conditions without ethanol shock. Products of all these genes are involved in the conversion of glutamate to ornithine. After ethanol shock, 4, 5 and 4 genes for enzymes of the glutamine biosynthesis pathway were found to be upregulated at 18 h, 24 h and 48 h, respectively. In addition, gene set enrichment analysis identified two primary metabolism pathways as significantly enriched in the set of differentially expressed genes of the $\Delta pepX$ mutant. These pathways are the pyridoxal 5'-phosphate (vitamin B6) biosynthesis and threonine degradation at 24 h and 48 h time points, respectively.

***pepX* deletion and ethanol shock effects on the expression of ECF-coding genes.**

Considering the partial overlap of the effects on secondary metabolism by the *pepX* deletion and ethanol shock, we considered the possibility that these may be manifested via ECF sigma factors. The expression patterns of sigma factor-encoding genes in the $\Delta pepX$ mutant cultures at 18 h, 24 h and 48 h time points were examined. Compared to the wild type strain, which has 56 sigma factor-coding genes, 12 such genes (10 of those encoding ECFs) were differentially regulated in the *pepX* mutant (Table 2). Six of these genes, all of them encoding ECFs, were also affected in the wild type by the ethanol shock (7). At the 18 h time point, 5 ECF-encoding genes and 2 genes for the housekeeping sigma factors were downregulated in the $\Delta pepX$ mutant, while expression of only one ECF-coding gene was upregulated as compared to the wild type strain (Table 2). After 24 hour, 2 ECF-coding genes and one for the housekeeping sigma factor were downregulated, while 3 ECF-coding genes were upregulated in the $\Delta pepX$ mutant. The effect of *pepX* deletion was much less pronounced after 48 hours, with none of the sigma factors-coding genes differentially expressed relative to the wild type. The genes located immediately next to the sigma factor genes affected by the *pepX* deletion were analyzed for potential to encode anti-sigma factors (ASFs) or other proteins known to control ECFs (1). ECF-encoding RS00060 gene is co-transcribed with RS00065, which codes for a 102 aa protein with one TM domain predicted by DeepTMHMM software (22). It seems thus plausible that this protein represents a membrane-embedded ASF of an unknown family. Putative transmembrane proteins that may be involved in the control of cognate ECFs were also identified for RS00830 and RS16265, while proteins with C₂H₂ zinc-finger domains known to sense redox potential and sequester ECFs (23) were found to be encoded next to RS18175, RS21030 and RS23885.

Deletion of three ECF-encoding genes has profound differential effect on the primary metabolism and expression of *S. venezuelae* BGCs.

Our previous data on the effect of ethanol shock on gene expression in the wild type strain (7) revealed three ECF-encoding genes, whose expression was upregulated at all three time points, namely RS15950, RS22785 and RS24280. Expression of two of these genes, RS15950 and RS22785, was also upregulated in the $\Delta pepX$ mutant without ethanol shock. Keeping in mind that most of the secondary metabolite BGCs in *S. venezuelae* were downregulated by either ethanol shock or *pepX* deletion (Table 1), we reasoned that these particular ECFs might be involved in suppressing particular secondary metabolite biosynthesis pathways.

All three genes were individually deleted from the wild type strain, and the transcriptomes and targeted secondary metabolomes of the resulting mutants ECF200 (Δ RS15950), ECF300 (Δ RS22785) and ECF500 (Δ RS24280) were investigated. The heatmap showing gene expression patterns of the resulting mutants is shown in Figure 2. Deletion of the ECF-coding gene RS15950 caused downregulation of 1433 genes (including 12 genes for sigma factors) and upregulation of 1511 genes (including 13 genes for sigma factors). Deletion of the ECF-coding gene RS22785 resulted in downregulation of 1309 genes (including 12 genes for sigma factors) and upregulation of 1433 genes (including 10 genes for sigma factors). Deletion of the ECF-coding gene RS24280 caused downregulation of 2187 genes (including 16 genes for sigma factors) and upregulation of 2366 genes (including 14 genes for sigma factors). The

sigma factor-coding genes differentially affected in at least one of the three mutants (44 out of total 56) are shown in Table 4.

Effect of ECF gene deletions on the expression of genes involved in primary and secondary metabolism

The deletion of the three ECF genes also had a massive impact on the gene expression related to enzymes in primary metabolism. Pathways for the biosynthesis of amino acids, as well as glycolysis, gluconeogenesis, fatty acid oxidation and the TCA cycle were investigated in detail (see Supplementary Information, Table S3). On average, the fraction of differentially expressed genes in the pathways was close to the genome-wide prevalence. However, the distribution of up and downregulated genes in the pathways was not random (chi-square test, ECF300: p-value = 0.0289, ECF200 and ECF500: p-value << 0,01). This is mainly due to the expression differences for genes involved in amino acid biosynthesis. In the ECF200 mutant more than 50% of genes were found to be upregulated in the pathways for arginine, asparagine, lysine, proline and serine biosyntheses, with only a single gene being downregulated in arginine biosynthesis. Conversely, in the case of the ECF500 mutant at least half of the genes in biosynthesis pathways for histidine, isoleucine, leucine, serine, threonine, tryptophane and valine were downregulated, with only one gene being upregulated in the tryptophane biosynthesis pathway. In addition, half of the genes in the pathways for glycine biosynthesis were upregulated, with none downregulated. The ECF300 mutant, however, had the fewest pathways with clear changes in the patterns of gene expression – only the alanine biosynthesis pathway is notable, with 2 genes being upregulated.

Next, we examined the effect of the three ECF genes deletions on the expression of the *S. venezuelae* BGCs in comparison with the wild type strain (Table 4). BGC expression was considered upregulated when more than one biosynthetic gene from the cluster was upregulated, and none of the other genes were downregulated. A similar rule was applied for considering downregulation of BGCs. This type of analysis still left some ambiguity with regard to the functional effect of gene expression patterns changes since in several BGCs, especially in the ECF200 mutant, some genes were up- and some downregulated. More specifically, 11 out of 37 BGCs in the ECF200 mutant were up- and 6 downregulated, with no effect observed for 8 BGCs. In the ECF300 mutant 11 out of 37 BGCs were up- and 13 downregulated, with no effect observed for 5 BGCs. The most profound effect on secondary metabolite biosynthesis genes expression was shown for the ECF500 mutant, where 14 BGCs out of 37 were up- and 8 downregulated, with no effect observed for only 2 BGCs.

Expression of several BGCs was similarly affected by all three ECF gene deletions. In particular, in all three mutants expression of BGCs for geosmin, a cryptic lanthipeptide, a cryptic thiopeptide, flaviolin, 2-methylisoborneol and a cryptic non-ribosomally synthesized peptide was upregulated. Expression of four BGCs was downregulated in all three ECF mutants, namely those for desferrioxamine, a cryptic pyrrolnitrin-like compound, an alkylresorcinol and a cryptic terpenoid. Considering the eight sigma factor-encoding genes with similar changes in expression observed for all three mutants (Table 3), it is plausible that these regulate the expression of the abovementioned eleven BGCs either directly or via some kind of regulatory cascades.

The expression of the chloramphenicol BGC was a more complex case, as it includes both up- and down regulated genes in all three ECF gene deletion mutants. One of the genes downregulated in all mutants is the RS04425, which codes for a repressor protein. Interestingly, 6 and 7 genes encoding enzymes for the chloramphenicol biosynthesis are upregulated in ECF300 and ECF500 mutants, whereas only non-metabolic genes of this BGC are upregulated in the ECF200 mutant (Figure 3).

Among the many genes differentially expressed in the three ECF gene deletion mutants, gene set enrichment analysis found three pathways to be enriched: ectoine, UMP and meso-butanediol biosynthesis. All three pathways are enriched in both ECF200 and ECF300 mutants, but none in ECF500.

Secondary metabolite profiles of the three ECF-encoding gene deletion mutants.

The profound effect of the three ECF gene deletions on the expression of the *S. venezuelae* BGCs was also measurable on the metabolome level (Figure 4). In agreement with the transcriptome data, a strong downregulation in the production of desferrioxamine B and legonoxamine A was observed. Conversely, higher concentrations of foroxymithine and particularly the gaburedins were found in the cultures of the mutants. While all the above-mentioned secondary metabolites were detected in very similar levels in the ECF200 (Δ RS15950), ECF300 (Δ RS22785) and ECF500 (Δ RS24280) mutants, a more diverse response was seen for the 2-hydroxyphenylthiazoline congeners, jadomycin-related angucyclines and chloramphenicol. Thiazostatin A and B and their C-methylated congeners watasemycin A and B, which are only found in very low amounts in the wild type, are very strongly upregulated in all three mutants, but the data indicate relatively lower conversion of thiazostatins to watasemycons in ECF300, pointing towards a relatively lower activity of the type B radical-SAM methylase homologue of *sven0515* (24). However, the concentrations of these 2-hydroxyphenylthiazoline congeners show strong batch-to-batch variations, which limit the confidence in these observations. Even more interesting from a biotechnological perspective is the tremendous increase in levels of the antibiotic chloramphenicol (up to 1700 fold, compared to the wild type) and the jadomycin-related angucycline L-digitoxosyl-phenanthroviridin shown particularly upon RS22785 deletion in ECF300. Interestingly, the gene expression profile of the enzymes involved in jadomycin biosynthesis shows an opposing pattern, with the majority of genes significantly higher expressed in the ECF500 mutant, but only a fraction of genes differentially expressed in the other two ECF mutants. The relevance of the effect on the concentration of L-digitoxosyl-phenanthroviridin is more difficult to evaluate since the baseline level in the wild-type shows strong fluctuations between batches and similar levels as in the ECF300 mutant were previously seen upon ethanol shock (Figure 1 and (7)). However, the production of chloramphenicol in the ECF300 and ECF500 mutants was much higher than we have ever seen in the wild type, and in the case of ECF300 even surpasses the effect of ethanol shock. For comparison, constant levels of the diketopiperazine Cyclo(Pro-Phe) were measured, while a moderate but significant effect of all three ECF gene deletions on the concentrations of threonylcarbamoyladenosine and

biotin were observed, indicating some wider alterations of the metabolism in line with the large number of up- and downregulated genes described above.

Secondary metabolite profile of the sigma factor deletion mutant of *Streptomyces* sp. ADI96-15.

The facts that deletions of the three selected ECF sigma factor genes had a profound effect on secondary metabolome of *S. venezuelae*, and these ECFs are conserved in *Streptomyces* prompted us to test whether similar results can be obtained in other *Streptomyces*. As a target gene for deletion, the RS24280 homologue (deleted in the ECF500 mutant of *S. venezuelae*) was chosen, since this mutation, according to the RNAseq data, led to upregulation of the largest number of BGCs in *S. venezuelae*. For this experiment, we selected *Streptomyces* sp. ADI96-15, which was isolated from a marine sponge *Antho dichotoma* collected in the Trondheim fjord (Norway) (25). Since extracts from this isolate did not display antimicrobial activities, the secondary metabolome of this ADI96-15 has not been investigated previously. Hence, the extracts from ADI96-15 culture grown in MYM medium were analyzed for the presence of known secondary metabolites. De-replication with LC-MS assisted by antiSMASH genome analysis led to the tentative identification of compounds derived from eight different BGCs. The most abundant congeners of each group were the hydroxamate siderophore desferrioxamine B, the cyclic octapeptide surugamide A (26), the γ -aminobutyrate urea gaburedin B, the class III lanthipeptide SAL-2242 (27), the polyene macrolide candicidin D (28) and a compound with the sum formula $C_{13}H_{20}O_3$ that is most likely a γ -butyrolactone (GBL) as judged by the high similarity of its MS/MS spectrum to those of known GBLs (Table S1).

The homologue of the *S. venezuelae* gene encoding ECF500 (RS24280) was identified in the genome of ADI96-15 (GenBank accession number GCA_003846265.1) as str9615DRAFT_04314, later referred to as *sig04314*, its product showing 79% amino acid sequence identity to that of RS24280. *sig04314* was deleted from the chromosome of *Streptomyces* sp. ADI96-15 (see Materials and methods), and the secondary metabolome of the resulting mutant compared to that of the wild-type strain. As expected, the deletion of the *sig04314* gene in *Streptomyces* sp. ADI96-15 had a significant effect on the production of several secondary metabolites, thus confirming the role of this ECF in the regulation of natural product biosynthesis in this strain (Figure 5). In contrast to our observations with the three ECF deletion mutants of *S. venezuelae* NRRL B-65442, significantly higher amounts of desferrioxamine B were detected in the Δ *sig04314* strain, while no significant effect was seen on the gaburedin B level. The production of the lanthipeptide SAL-2242 and the presumed GBL were also upregulated, while the one of candicidin D was downregulated. However, overall, the impact of the *sig04314* deletion was less pronounced than for any of the three ECF-coding gene deletions in *S. venezuelae* NRRL B-65442, which can be due to the cultivation conditions used.

Discussion

Our previous study on the effect of ethanol shock on gene expression in *S. venezuelae* (7) revealed certain interesting facts that required further investigation. In particular, a correlation between downregulation of certain secondary metabolite BGCs and expression of a *pepX* gene encoding a small, hydrophobic cryptic peptide with the sequence *N*-formyl-MNVITNLLAGVVHFLGWLV seemed intriguing. Although deletion of this gene had only modest effect on secondary metabolism, the overall gene expression pattern was markedly changed in the $\Delta pepX$ mutant. Although deletion of this gene had only modest effect on secondary metabolism, overall gene expression pattern was markedly changed in the $\Delta pepX$ mutant. Hence, *pepX* deletion appears to mimic the effect of ethanol shock to a certain extent, at least when it concerns particular secondary metabolite biosynthesis genes. If we consider changes in the bacterial membrane caused by ethanol (29), this observation may indicate that PepX is membrane-bound and plays a certain role as its component. The fact that only traces of PepX could be identified in the culture supernatant of the wild type strain (data not shown) indirectly supports this suggestion. Since PepX does not appear to have any transmembrane helices, it may be associated with the membrane via an interacting protein partner, and its absence affects membrane topology and hence membrane-bound proteins that control certain ECFs. Such interacting partner(s) might be encoded by the neighboring genes RS03775 and RS03780 annotated as coding for an endolytic transglycosylase of the YceG family and an ABC transporter, respectively. Endolytic transglycosylase functions as a peptidoglycan terminase that cleaves nascent peptidoglycan strands to terminate their elongation (30). According to the prediction by DeepTMHMM and analysis by MOTIF Search online tool (<https://www.genome.jp/tools/motif/>), this protein has one N-terminal transmembrane domain with similarity to SLATT domain that has been implicated in cell signaling in bacteria (31). The RS03780-encoded ABC transporter, according to the prediction, has six transmembrane domains and thus must be associated with the membrane. Notably, a similar genetic context in terms of genes surrounding *pepX* homologues were identified in many *Streptomyces*, e.g. *Streptomyces laurentii* ATCC 31255, *Streptomyces fradiae* ATCC 10745 and *Streptomyces coelicolor* M1154. In the genome of the latter strain, the *pepX* homologue is not annotated, but was identified between genes RS14905 and RS14910 coding for an endolytic transglycosylase and an ABC transporter, respectively. We have found copious amounts of PepX homologue in the extracts of *S. coelicolor* M1154 cultures (data not shown), supporting the suggestion that this small peptide must have a particular basic function in streptomycetes. Inverse correlation between the expression of three particular ECFs and certain BGCs in *S. venezuelae* prompted investigation into their roles in regulation of secondary metabolism by generating and analyzing deletion mutants. The high number of genes affected in all three mutants, in particular ECF500, is most likely due to a cascade-like effect because of the differential expression of other sigma factor-encoding genes (Table 2). Hence, the true regulons of the three ECFs studied was not possible to establish in this study. The data presented in Table 3 suggests that RS15950- and RS22785-encoded ECFs have rather similar regulons, at least with respect to sigma factor genes affected, since deletions of these ECF genes cause identical changes in the expression patterns of 17 such genes (8 up- and 9

downregulated) while having no effect on other 14. This notion is in part supported by the gene expression pattern of all differentially expressed genes shown in Figure 2. The situation is different for RS24280, since the deletion of this ECF-encoding gene has the same effect (up- or downregulation) on the expression of only 9 sigma factor genes as the deletion of RS15950 and RS22785. It seems likely that these 9 genes represent part of regulons of all three abovementioned ECFs. Overall, the deletion of RS24280 had the most significant effect in *S. venezuelae*, changing expression patterns for 4553 genes (over 63% of all genes in the genome) in the conditions tested.

The activity of ECFs in bacteria is usually controlled by specific anti-sigma factors (ASFs), which are often tethered to the membrane and sequester ECFs via protein-protein interactions. Alternatively, ECF activity can be modulated by soluble ASFs. In the case of *S. coelicolor*, σ^R is controlled by the soluble ASF RsrA that binds σ^R and invokes a conformational change that prevents this ECF's binding to cognate promoters (32). This interaction is broken by a change in redox conditions induced by environmental stimuli that prompt formation of a disulfide bridge on RsrA, resulting in σ^R release. Out of three investigated ECFs of *S. venezuelae*, only RS24280 appears to have an ASF encoded by a downstream gene RS24281. This putative ASF is represented by a 105 amino acid protein predicted to contain a zinc-finger C₂H₂ domain. Such a domain is known to be involved in redox sensing (23), hence supporting the role of this ASF similar to that of RsrA in *S. coelicolor*.

How the ECF encoded by RS15950 is controlled is not that clear, since the latter gene is transcriptionally coupled to the downstream gene RS15955 specifying a 254-amino acid protein with no homology to known ASFs or any other characterized proteins. Immediately downstream RS15955 there are genes encoding a two-component system, which can be part of a control mechanism. A two-component response system encoded by genes immediately upstream of RS22785 and transcribed in opposite direction may also control the latter ECF. The ECF σ^E , a major regulator of cell envelope stress in *S. coelicolor*, is controlled by a two-component system (33), and it would be tempting to assume that ECFs responding to ethanol shock could play a similar role. The latter, however, remains an open question, since no appreciable homology (< 30%) could be found between the amino acid sequences of σ^E and the three ECFs characterized in this study.

For a long time it has been assumed that sigma factors in general play a strictly positive regulatory role, as they are required components of a transcriptional complex essential for promoter recognition, formation of an open complex and promoter clearance. Recent studies, however, challenge this paradigm, suggesting that ECFs can play both positive and negative roles in controlling gene expression, especially in *Streptomyces* bacteria. Rebets et al. (5) reported deletion of several sigma factor-encoding genes in *Streptomyces lividans* TK24, with some of the deletions apparently causing activation of silent BGCs for actinorhodin and undecylprodigiosin in this strain. New investigation on the regulatory role of the σ^E ECF in *S. coelicolor* firmly established its role as both positive and negative regulator (34), albeit whether the observed effects were direct or the results of a cascade-like regulation has not been studied. Our own data showing that deletion of the three ECF-encoding genes in *S. venezuelae* positively affects expression of particular BGCs strongly supports the differential negative regulatory roles of these ECFs in secondary metabolism.

The fact that deletion of RS24280 had a positive effect on the expression of the majority of BGCs prompted us to test whether this effect can be mimicked by a deletion of a homologous ECF-encoding gene in another *Streptomyces*. We have chosen the marine sponge-derived *Streptomyces* sp. ADI96-15 because it is a natural isolate and a non-model strain with no previous record on its secondary metabolome. Deletion of the RS24280 homologue in this strain indeed had a significant effect on secondary metabolism, affecting at least 4 biosynthetic pathways for molecules that could be identified using LC-MS-based analytics. Notably, the effect of this deletion on desferrioxamine biosynthesis was opposite to that observed in *S. venezuelae*, suggesting a difference in regulation of this pathway in two *Streptomyces* strains. Taking these data into account, it appears that deletions of RS24280 homologues in *Streptomyces* spp. will deregulate secondary metabolism, while it is not possible to predict whether the effect on certain pathways will be positive or negative. This ambivalent effect seems logical, considering that differences in regulation of secondary metabolite biosynthesis pathways reflect exposure of particular strains to specific environmental cues. The three ECFs, identified in this study as being heavily involved in regulation of secondary metabolism are conserved in hundreds of *Streptomyces* spp. The latter suggests that their deletions can be used as a tool to deregulate biosynthetic pathways, hereby eventually prompting discovery of new molecules with drug development potential or optimizing *Streptomyces* strains for improved secondary metabolite yield.

Conclusion

Streptomyces bacteria are a valuable source of bioactive secondary metabolites, but often the genes involved in their biosynthesis are transcriptionally silent in the laboratory conditions, preventing the discovery of novel molecules. Also, yields of known valuable secondary metabolites can be low, making their industrial production very challenging. In this work, we identified three sigma factors that are conserved in streptomycetes, which suppress secondary metabolism in these bacteria. Individual deletions of the genes encoding these sigma factors resulted in differential upregulation of secondary metabolite biosynthesis genes and overproduction (up to 1700 fold) of certain cognate molecules. Hence, deletion of homologous genes in various *Streptomyces* bacteria may provide a new effective tool for genome mining aimed at the discovery of new compounds, as well as for the increased production of known valuable metabolites.

Acknowledgements

We thank Anna Fabisikova from the Mass Spectrometry Centre of the Faculty of Chemistry, University of Vienna, for her assistance with LC-MS data acquisition.

Author contributions

SBZ and TR designed the experiments. ONS, MZ, PH and MP performed the experiments. SBZ, TR, MZ, ONS, MP and PH analyzed the data and concluded insights. SBZ, MZ, ONS and MP drafted the manuscript. All authors interpreted the results, read, and approved the final manuscript.

Funding

Open Access funding provided thanks to the University of Vienna agreement with Springer Nature. This research was funded by the Austrian Science Fund (FWF) [grant P30986]. M.P. was funded through FWF doc.funds project DOC 69 ``Microbial symbioses in dynamic environments: Metabolic interplay and novel interactions (MAINTAIN)".

Availability of data and materials

All data generated during this study are included in this article and the additional files. Raw data are available on reasonable request.

Competing interests

The authors declare that they have no competing interests.

References

1. Mascher T. Past, Present, and Future of Extracytoplasmic Function σ Factors: Distribution and Regulatory Diversity of the Third Pillar of Bacterial Signal Transduction. *Annu Rev Microbiol.* 2023;77:625-644.
2. Sineva E, Savkina M, Ades SE. Themes and variations in gene regulation by extracytoplasmic function (ECF) sigma factors. *Curr Opin Microbiol.* 2017;128-137.
3. Alam K, Mazumder A, Sikdar S, Zhao YM, Hao J, Song C, Wang Y, Sarkar R, Islam S, Zhang Y, Li A. *Streptomyces*: The biofactory of secondary metabolites. *Front Microbiol.* 2022; 13:968053.
4. Bilyk B, Kim S, Fazal A, Baker TA, Seipke RF. Regulation of Antimycin Biosynthesis Is Controlled by the ClpXP Protease. *mSphere.* 2020;5:e00144-20.
5. Rebets Y, Tsoilis KC, Guðmundsdóttir EE, Koepff J, Wawiernia B, Busche T, Bleidt A, Horbal L, Myronovskyi M, Ahmed Y, Wiechert W, Rückert C, Hamed MB, Bilyk B, Anné J, Friðjónsson Ó, Kalinowski J, Oldiges M, Economou A, Luzhetskyy A. Characterization of Sigma Factor Genes in *Streptomyces lividans* TK24 Using a Genomic Library-Based Approach for Multiple Gene Deletions. *Front Microbiol.* 2018; 9:3033.
6. Mauri M, Klumpp S. A model for sigma factor competition in bacterial cells. *PLoS Comput Biol.* 2014; 10:e1003845.
7. Sekurova ON, Zehl M, Predl M, Hunyadi P, Rattei T, Zotchev SB. Targeted Metabolomics and High-Throughput RNA Sequencing-Based Transcriptomics Reveal Massive Changes in the *Streptomyces venezuelae* NRRL B-65442 Metabolism Caused by Ethanol Shock. *Microbiol Spectr.* 2022;10:e0367222.
8. Zotchev S, Haugan K, Sekurova O, Sletta H, Ellingsen TE, Valla S. Identification of a gene cluster for antibacterial polyketide-derived antibiotic biosynthesis in the nystatin producer *Streptomyces noursei* ATCC 11455. *Microbiology.* 2000;146:611-619.
9. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics.* 2011;27:863-4.
10. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.* 2013;1303.3997v2 [q-bio.GN].
11. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078-9.
12. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31:166-9.
13. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019;20:257.
14. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139-40.
15. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.

16. Zimmermann J, Kaleta C, Waschina S. gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biol.* 2021;22:81.
17. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics.* 2013;14:128.
18. Adams KJ, Pratt B, Bose N, Dubois LG, St John-Williams L, Perrott KM, Ky K, Kapahi P, Sharma V, MacCoss MJ, Moseley MA, Colton CA, MacLean BX, Schilling B, Thompson JW; Alzheimer's Disease Metabolomics Consortium. Skyline for Small Molecules: A Unifying Software Package for Quantitative Metabolomics. *J Proteome Res.* 2020;19:1447-58.
19. van Santen JA, Poynton EF, Iskakova D, McMan E, Alsup TA, Clark TN, Fergusson CH, Fewer DP, Hughes AH, McCadden CA, Parra J, Soldatou S, Rudolf JD, Janssen EM, Duncan KR, Linington RG. The Natural Products Atlas 2.0: a database of microbially-derived natural products. *Nucleic Acids Res.* 2022;50(D1):D1317-D1323.
20. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapon CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrew K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol.* 2016;34:828-37.
21. Schmid R, Heuckeroth S, Korf A, Smirnov A, Myers O, Dyrland TS, Bushuiev R, Murray KJ, Hoffmann N, Lu M, Sarvepalli A, Zhang Z, Fleischauer M, Dührkop K, Wesner M, Hoogstra SJ, Rudt E, Mokshyna O, Brungs C, Ponomarov K, Mutabdzija L, Damiani T, Pudney CJ, Earll M, Helmer PO, Fallon TR, Schulze T, Rivas-Ubach A, Bilbao A, Richter H, Nothias LF, Wang M, Orešič M, Weng JK, Böcker S, Jeibmann A, Hayen H, Karst U, Dorrestein PC, Petras D, Du X, Pluskal T. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat Biotechnol.* 2023;41:447-9.

22. Hallgren J, Tsirigos KD, Pedersen MD, Armenteros JJA, Marcatili P, Nielsen H, Krogh A, Winther O. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv* 2022.04.08.487609.
23. Bregnard TA, Fairchild D, Erlandsen H, Semenova IV, Szczepaniak R, Ahmed A, Weller SK, Korzhnev DM, Bezsonova I. Conformational exchange at a C₂H₂ zinc-binding site facilitates redox sensing by the PML protein. *Structure*. 2023;31:1086-99.e6.
24. Inahashi Y, Zhou S, Bibb MJ, Song L, Al-Bassam MM, Bibb MJ, Challis GL. Watasemycin biosynthesis in *Streptomyces venezuelae*: thiazoline C-methylation by a type B radical-SAM methylase homologue. *Chem Sci*. 2017;8:2823-31.
25. Kersten RD, Yang YL, Xu Y, Cimermancic P, Nam SJ, Fenical W, Fischbach MA, Moore BS, Dorrestein PC. A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat Chem Biol*. 2011;7:794-802.
26. Guerrero-Garzón JF, Zehl M, Schneider O, Rückert C, Busche T, Kalinowski J, Bredholt H, Zotchev SB. *Streptomyces* spp. From the Marine Sponge *Antho dichotoma*: Analyses of Secondary Metabolite Biosynthesis Gene Clusters and Some of Their Products. *Front Microbiol*. 2020;11:437.
27. Matsuda K, Kuranaga T, Sano A, Ninomiya A, Takada K, Wakimoto T. The Revised Structure of the Cyclic Octapeptide Surugamide A. *Chem Pharm Bull*. 2019;67:476-80.
28. Haeder S, Wirth R, Herz H, Spiteller D. Candicidin-producing *Streptomyces* support leaf-cutting ants to protect their fungus garden against the pathogenic fungus *Escovopsis*. *Proc Natl Acad Sci U S A*. 2009;106:4742-6.
29. Maitre M, Weidmann S, Dubois-Brissonnet F, David V, Covès J, Guzzo J. Adaptation of the wine bacterium *Oenococcus oeni* to ethanol stress: role of the small heat shock protein Lo18 in membrane integrity. *Appl Environ Microbiol*. 2014;80:2973-80.
30. Yunck R, Cho H, Bernhardt TG. Identification of MltG as a potential terminase for peptidoglycan polymerization in bacteria. *Mol Microbiol*. 2016;99:700-18.
31. Burroughs AM, Zhang D, Schäffer DE, Iyer LM, Aravind L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res*. 2015;43:10633-54.
32. Park JH, Lee JH, Roe JH. SigR, a hub of multilayered regulation of redox and antibiotic stress responses. *Mol Microbiol*. 2019;112:420-431.
33. Tran NT, Huang X, Hong HJ, Bush MJ, Chandra G, Pinto D, Bibb MJ, Hutchings MI, Mascher T, Buttner MJ. Defining the regulon of genes controlled by σ^E , a key regulator of the cell envelope stress response in *Streptomyces coelicolor*. *Mol Microbiol*. 2019;112:461-81.
34. Pospíšil J, Schwarz M, Ziková A, Vítovská D, Hradilová M, Kolář M, Křenková A, Hubálek M, Krásný L, Vohradský J. σ^E of *Streptomyces coelicolor* can function both as a direct activator or repressor of transcription. *Commun Biol*. 2024;7:46.
35. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S, Subhraveti P, Karp PD. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res*. 2020;48:D445-D453.

Table 1. Comparison of the effects the *pepX* deletion and ethanol shock had on the BGC expression in *S. venezuelae*.

Type	Putative product	Genes	Transcriptome Δ pepX	Transcriptome WT+EtOH
ectoine	ectoine/betaine	RS01090-01095	none	down
terpene	geosmin	RS01290	none	down
pks/nrps	venemycin	RS02305-02380	none	none
nrps	watasemycins & congeners/shunt products	RS02405-02475	up	none
lantipeptide	core peptides: ADhaIDVPYDhbDhbGDhbIDhbVC	RS02610-02620	none	none
terpene	terpenoid	RS02650-02655	none	down
lantipeptide	venezuelin	RS02970-02980	none	none
indole	arcyriaflavin	RS03645-03670	none	none
nrps	chloramphenicol	RS04425-04505	up	up
cdps	cyclic dipeptide	RS09190-09195	none	none
siderophore	desferrioxamine	RS12675-12690	down	down
lasso peptide	core peptide: GDAAELTQGQGGGQSEDKRRAYNC	RS15430-15450	none	none
nrps	peptide	RS20250-20325	none	none
other	gaburedins	RS20785-20795	none	down
butyrolactone	butyrolactone	RS20800-20805	down	none
melanin	melanin?	RS23170-23175	none	none
butyrolactone	butyrolactone	RS25385	none	none
other	pyrrolnitrin-like	RS25415-25435	none	none
thiopeptide	core peptide: SSGSIGTSSSSSTCSAC	RS25565-25620	up	down
pks3	flaviolin	RS26780	none	none
siderophore	siderophore	RS27055-27060	none	none
siderophore	legonoxamines	RS27330-27345	none	down
bacteriocin	bacteriocin	RS29140	none	none
pks2	jadomycin & congeners/precursors	RS29740-29955	down	up
nrps	peptide	RS30590-30630	none	up
pks2-nrps	peptide-polyketide hybrid	RS30820-30995	down	down
ladderane	ladderane	RS31000-31150	down	down
nrps	terpapeptide (x-Phe-X-Gly?)	RS31155-31260	none	down
terpene	hopanoids	RS32065-32140	none	down
bacteriocin	bacteriocin	RS32515	none	none
pks2	spore pigment	RS33815-33860	none	down
melanin	melanin	RS34060-34065	down	none
nrps	foroxymithines	RS35125-35150	down	down
terpene	2-methylisoborneol	RS35405-35410	none	down
pks3	alkylresorcinol	RS36050-36065	none	down
terpene	terpenoid	RS36960	none	none
nrps	peptide	RS37045	none	down

Table 2. Expression of *S. venezuelae* σ factors affected by the deletion of the *pepX* gene relative to the wild type strain. Cells representing genes that encode ECFs are shaded in grey.

Gene	18 h	24 h	48 h	σ domains
RS00060	-	Up	-	r2, r4
RS00830	Down	-	-	r2, r4
RS01725	Down	-	-	r2, r3, r4
RS15950	-	Up	-	r2, r4
RS16265	Up	-	-	r2, r4
RS17230	-	Up	-	r2, r4, SnoaL
RS18175	Down	Down	-	r2, r3, r4
RS18620	Down	Down	-	r2, r4
RS21030	Down	Down	-	r2, r4
RS22785	-	Up	-	r2, r4
RS23885	Down	-	-	r2, r4
RS34660	Down	-	-	r2, r4

Table 3. Differential effect of ECF gene deletions on the expression of other sigma factor-encoding genes in *S. venezuelae*. Only genes affected in at least one of the mutants are shown.

Sigma factor-encoding gene	Sigma factor domains	Δ RS15950	Δ RS22785	Δ RS24280
RS00270	r2, r4	NA	NA	up
RS00630	r2, r4, SnoaL	down	down	NA
RS00830	r2, r4	NA	NA	up
RS01100	r2, r3, r4	up	up	down
RS01725	r2, r3, r4	up	up	down
RS01895	r2, r4	NA	NA	up
RS02075	r2, CAP	NA	NA	up
RS02270	r2, r3, r4	NA	NA	up
RS03595	r2, r4, DUF6596	NA	down	down
RS04145	r2, r4, SnoaL	down	down	NA
RS04775	r2, r4	down	down	NA
RS05705	r2, r4	NA	NA	down
RS11135	r1, r2, r3, r4	up	NA	NA
RS13895	r2, r3, r4	down	down	down
RS14430	r2, r4	down	NA	NA
RS15140	r1, r2, r3, r4	down	down	down
RS15760	r2, r4	up	up	up
RS15920	r4, BACON	up	up	up
RS15950	r2, r4	deleted	up	NA
RS16265	r2, r4	NA	up	down
RS16340	r2, r4	down	NA	NA
RS16715	r2, r4	up	NA	NA
RS17230	r2, r4, SnoaL	up	up	up

RS17255	r2, r4, SnoaL	down	down	down
RS18175	r2, r3, r4	down	down	down
RS18620	r2, r4	NA	down	up
RS18755	r2, r3, r4	up	NA	down
RS18935	r2, r4, SnoaL	NA	NA	up
RS19240	r2, r4, NPCBM	NA	NA	up
RS21030	r2, r4	up	NA	NA
RS22475	r2, r4	NA	down	NA
RS22610	r2, r4	up	up	NA
RS22720	r2, r4, SnoaL	up	NA	NA
RS22785	r2, r4	down	deleted	down
RS23885	r2, r4	down	down	down
RS24280	r2, r4	NA	NA	deleted
RS24470	r2, r3, r4	up	up	down
RS24785	r2, r4	up	up	up
RS26440	r2, r3, r4	NA	NA	down
RS27450	r1, r2, r3, r4	NA	NA	down
RS32930	r2, r4	NA	NA	down
RS33970	r2, r4	NA	NA	up
RS34660	r2, r4	down	down	NA
RS35505	r2, r4, phage int	NA	NA	up

NA – not affected.

Table 4. Effects of the three ECF-encoding gene deletions on the transcriptome and secondary metabolome of *S. venezuelae* as compared to the wild type strain.

Putative product	Genes	Metabolome ΔECF200	Transcriptome ΔECF200	Metabolome ΔECF300	Transcriptome ΔECF300	Metabolome ΔECF500	Transcriptome ΔECF500
ectoine/betaine	RS01090-01095		up		up		none
geosmin	RS01290		up		up		up
venemycin	RS02305-02380		mixed		down		mixed
wataemycins & congeners/shunt products	RS02405-02475		mixed		mixed		mixed
core peptides: ADhaIDVPYDhDbGDbhDhbVC;	RS02610-02620	none	up	up	down	up	mixed
terpenoid	RS02650-02655		up		up		down
venezuelin	RS02970-02980		up		none		up
aracylflavin	RS03645-03670		none		mixed		mixed
chloramphenicol	RS04425-04505	none	mixed	up	mixed	up	mixed
cyclic dipeptide	RS09190-09195		down		down		up
desferrioxamine	RS12675-12690		down		down		down
core peptide (?): GDAALITQGQGGQSEDKRRAYNC	RS15430-15450	down	mixed	down	up	down	mixed
peptide	RS20250-20325		none		up		up
gaburedins	RS20785-20795	up	none	up	up	up	up
butyrolactone	RS20800-20805		up		none		down
melanin?	RS23170-23175		up		down		down
butyrolactone	RS25385		none		none		none
pyrrolinrin-like?	RS25415-25435		down		down		down
core peptide: SSGSIGTSSSSSTCSAC	RS25565-25620		up		up		mixed
flaviolin	RS26780		none		none		up
siderophore	RS27055-27060		none		down		up
legonoxamines	RS27330-27345		mixed		mixed		mixed
bacteriocin	RS29140		none		down		up
jadomycin	RS29740-29955	up	mixed	up	mixed	none	mixed
peptide	RS30590-30630		mixed		mixed		mixed
peptide-polyketide hybrid	RS30820-30995		mixed		up		up
ladderane	RS31000-31150		mixed		mixed		mixed
terpapeptide (x-Phe-X-Gly?)	RS31155-31260		mixed		up		mixed
hopanoids	RS32065-32140		mixed		mixed		mixed
bacteriocin	RS32515		none		down		up
spore pigment	RS33815-33860		up		down		up
melanin	RS34060-34065		down		none		up
foroxymithines	RS35125-35150		mixed	up	down	up	down
2-methylisoborneol	RS35405-35410	up	up		up		up
alkylresorcinol	RS36050-36065		down		down		down
terpenoid	RS36960		down		down		down
peptide	RS37045		up		up		up

Figure legends

Figure 1. Effect of *pepX* deletion and ethanol shock on the production of secondary metabolites and the relative concentration of other identified compounds in MYM cultures of *S. venezuelae* NRRL B-65442 after 18 h, 24 h, and 48 h of cultivation. After 7 h of cultivation, either 6% (v/v) of absolute ethanol (red and yellow bars) or distilled sterile water (blue and green bars) was added. Each bar represents the average of three independent experiments ($n = 3$). The error bars represent the standard deviation. The statistical differences between groups were evaluated using two-way ANOVA followed by Tukey's multiple-comparison test (****, $P < 0.0001$; ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$).

Figure 2. Comparison of differentially expressed genes among ECF mutants. The sample names show the genotype (wild type, ECF200, ECF300, ECF500) and the biological replicate (1-3). Each cell in the heatmap corresponds to a gene in a given sample. The cells are colored with respect to gene expression, normalized within each sample to the overall expression and converted to a z-score based on mean and standard deviation across all samples. The order of the genes was determined by clustering, based on the z-score values.

Figure 3. The chloramphenicol biosynthesis pathway visualized according to MetaCyc (35) (PWY-8032) and its gene expression in ECF deletion mutants. For each gene encoding an enzyme of this pathway the differential expression status is shown for the ECF200, ECF300 and ECF500 deletion mutants. Higher and lower differential expression is denoted as up and down in the table and color-coded with orange and blue respectively. For one step of the pathway, the conversion of 3-(4-aminophenyl)pyruvate to 4-amino-L-phenylalanine, no gene could be identified, marked here with a question mark.

Figure 4. Effect of three ECF gene deletions on the production of secondary metabolites and the relative concentration of other identified compounds in MYM cultures of *S. venezuelae* NRRL B-65442 after 72 h of cultivation. Each bar represents the average of three independent experiments ($n = 3$). The error bars represent the standard deviation. The statistical differences to the wild type were evaluated using one-way ANOVA followed by Dunnett's multiple-comparison test (****, $P < 0.0001$; ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$).

Figure 5. Effect of the RS24280-homologue deletion (*Δsig04314*) on the production of secondary metabolites and the relative concentration of other identified compounds in MYM cultures of *Streptomyces* sp. ADI96-15 after 7 d of cultivation. Each bar represents the average of three independent experiments ($n = 3$). The error bars represent the standard deviation. The statistical differences to the wild type were evaluated using an unpaired t test (**, $P < 0.01$; *, $P < 0.05$).

Figure 1.

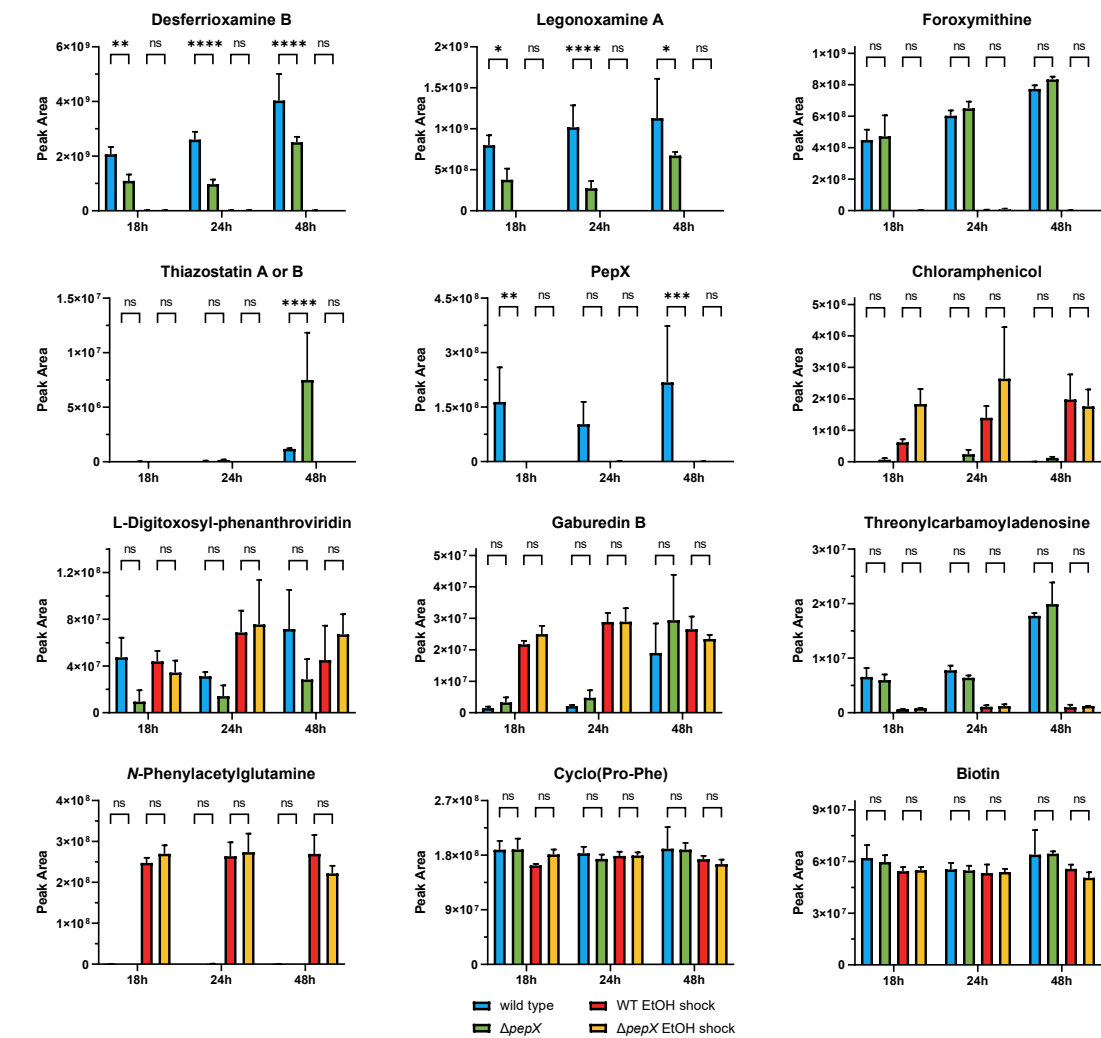


Figure 2.

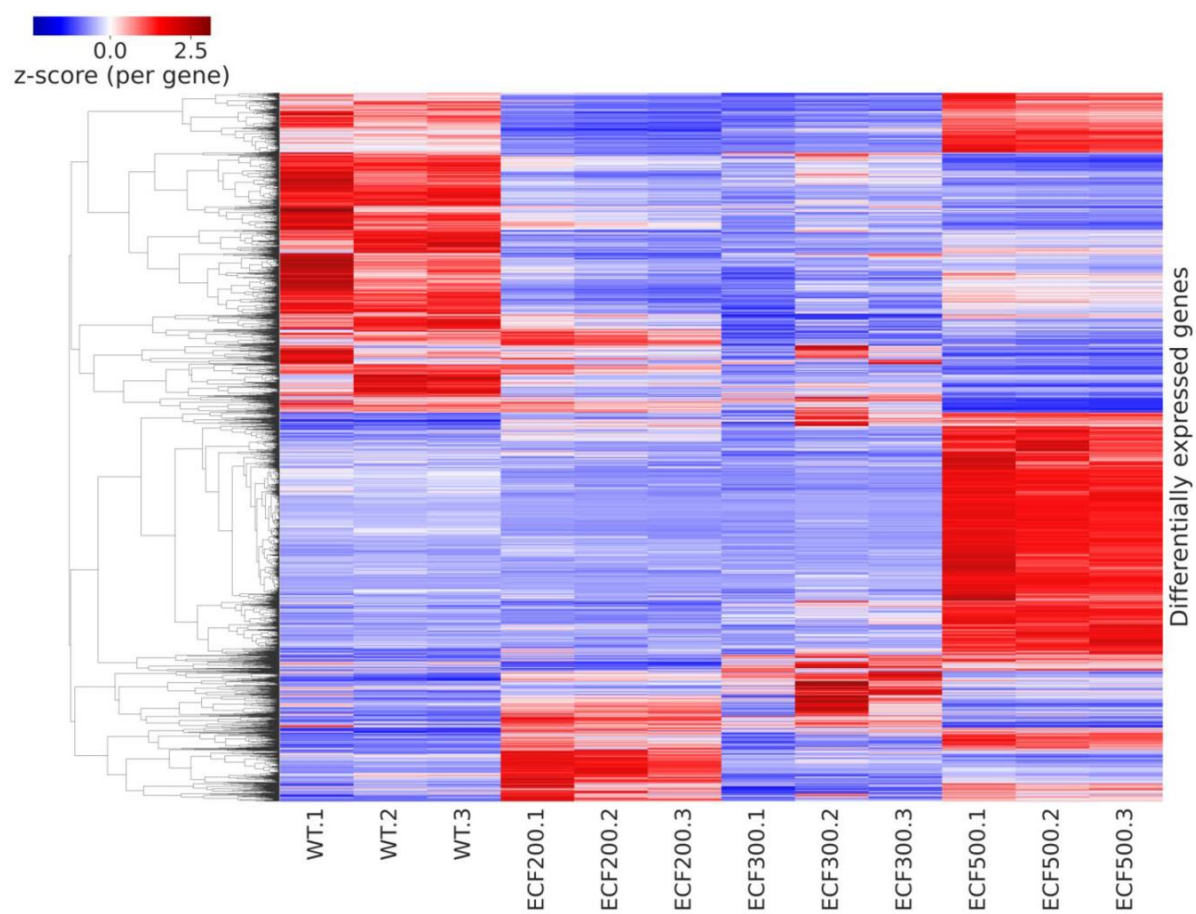


Figure 3.

Chloramphenicol biosynthesis pathway

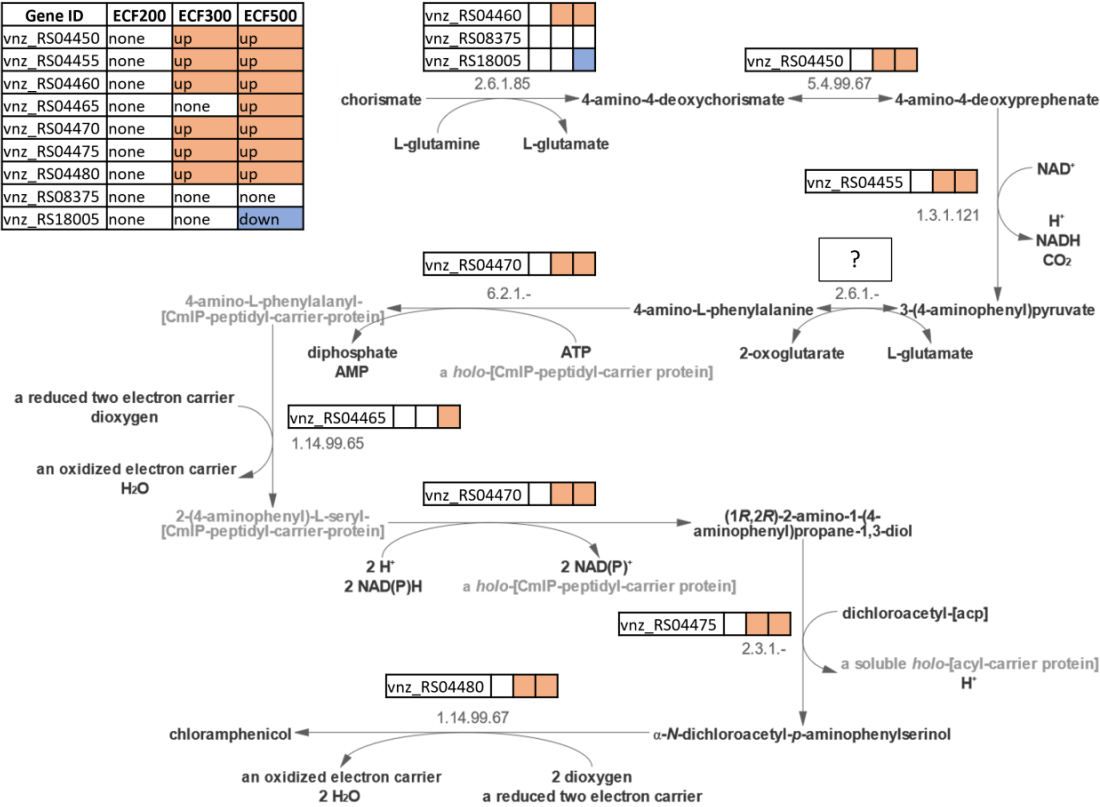


Figure 4.

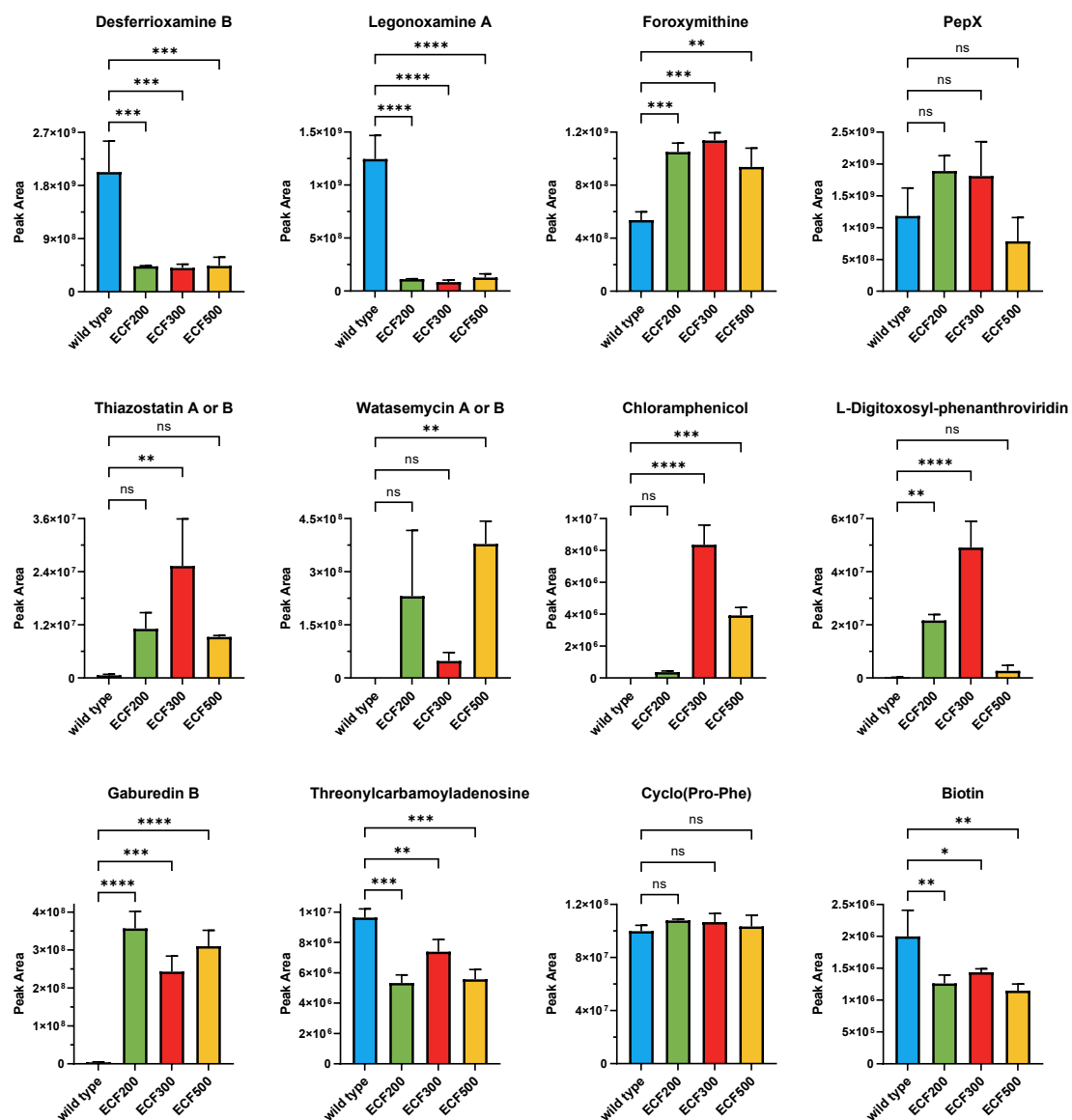
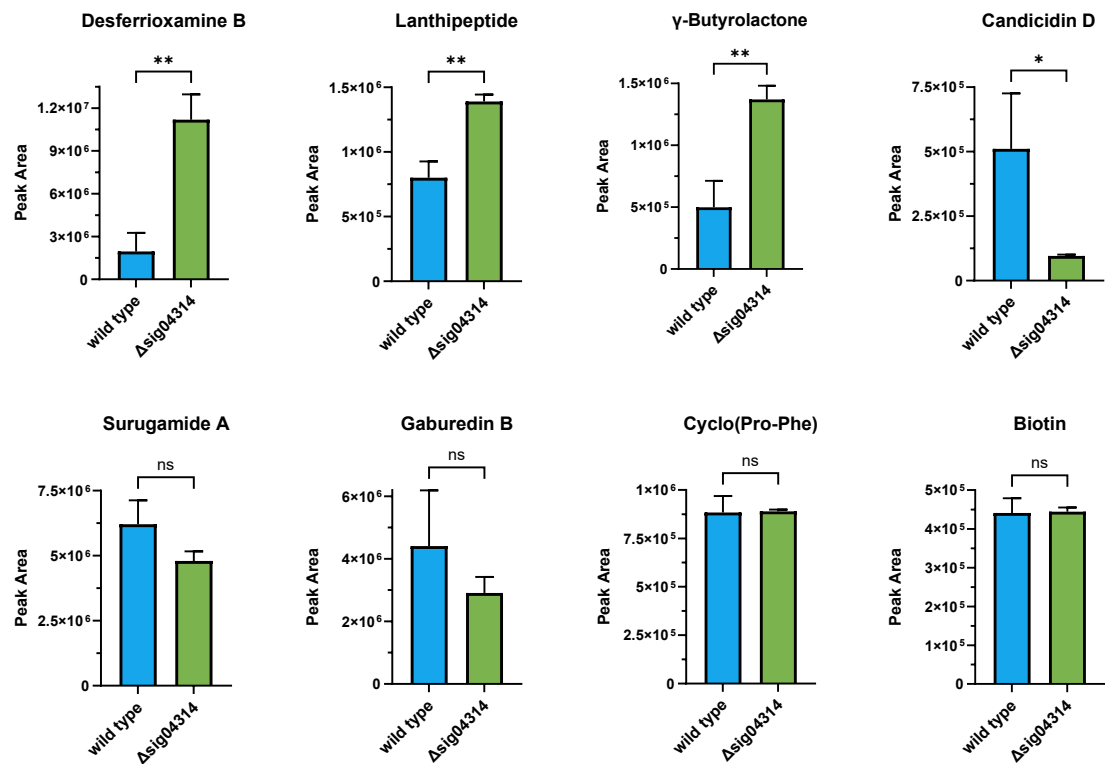


Figure 5.



A.3. PyCoMo: a python package for community metabolic model creation and analysis

A.3.1. Publication status

Published work: Predl, M., Mießkes, M., Rattei, T., Zanghellini, J., *PyCoMo: a python package for community metabolic model creation and analysis*. Bioinformatics, Volume 40, Issue 4, April 2024, btae153, DOI: <https://doi.org/10.1093/bioinformatics/btae153>

A.3.2. Author contributions

The author of this dissertation conceived of the study and software design, implemented the software, gathered and analyzed the test data, designed the figure and wrote the manuscript.

A.3.3. Publication

Bioinformatics, 2024, 40(4), btae153
<https://doi.org/10.1093/bioinformatics/btae153>
 Advance Access Publication Date: 26 March 2024
 Applications Note

OXFORD

Systems biology

PyCoMo: a python package for community metabolic model creation and analysis

Michael Predl ^{1,2}, Marianne Mießkes ^{3,4}, Thomas Rattei ^{1,2,*}, Jürgen Zanghellini ^{3,4,*}

¹Department of Microbiology and Ecosystem Science, Division of Computational Systems Biology, Centre for Microbiology and Environmental Systems Science, University of Vienna, 1030 Vienna, Austria

²Doctoral School in Microbiology and Environmental Science, University of Vienna, 1030 Vienna, Austria

³Department of Analytical Chemistry, Faculty of Chemistry, University of Vienna, 1090 Vienna, Austria

⁴Austrian Centre of Industrial Biotechnology, 1190 Vienna, Austria

*Corresponding authors. Department of Microbiology and Ecosystem Science, Division of Computational Systems Biology, Centre for Microbiology and Environmental Systems Science, University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria. E-mail: thomas.rattei@univie.ac.at (T.R.); Department of Analytical Chemistry, Faculty of Chemistry, University of Vienna, Sensengasse 8, 1090 Vienna, Austria. E-mail: juegen.zanghellini@univie.ac.at (J.Z.)

Associate Editor: Janet Kelson

Abstract

Summary: PyCoMo is a python package for quick and easy generation of genome-scale compartmentalized community metabolic models that are compliant with current openCOBRA file formats. The resulting models can be used to predict (i) the maximum growth rate at a given abundance profile, (ii) the feasible community compositions at a given growth rate, and (iii) all exchange metabolites and cross-feeding interactions in a community metabolic model independent of the abundance profile; we demonstrate PyCoMo's capability by analysing methane production in a previously published simplified biogas community metabolic model.

Availability and implementation: PyCoMo is freely available under an MIT licence at <http://github.com/univieCUBE/PyCoMo>, the Python Package Index, and Zenodo.

1 Introduction

Cross-feeding interactions are crucial for community formation and function. Predicting these interactions is an open challenge, with much development in the field of metabolic modelling (Heinken *et al.* 2021). It has expanded in the direction of microbial communities, with research being done on simple, cultured communities up to personalized models of the human gut microbiome (Diener *et al.* 2020, Ibrahim *et al.* 2021, Saa *et al.* 2022).

Metabolic modelling on the level of communities allows investigating questions that can only be stated in the context of communities, such as which interactions can occur within them. Different approaches have been developed to model communities of metabolic models: individual modelling of the members with iterative cross-talk via metabolite availability, lumped models representing the collective enzymatic capabilities of a community without cellular borders, and compartmentalized metabolic models, which place each member model in a separate compartment and add a shared exchange compartment for interaction (Perez-Garcia *et al.* 2016). While resulting in the largest and most complex models, the compartmentalized approach allows direct modelling of inter-species interactions such as cross-feeding.

The creation of compartmentalized community metabolic models includes many technical obstacles, such as matching the namespace of exchange metabolites, creating a community biomass function and preserving the attribution of metabolites and

reactions to the community members within the model. All these tasks can be automated when all necessary information is present, as shown in Commodelpy by Bekiaris and Klamt (2021). Yet, the models generated by the current version (0.0.3) of Commodelpy face several issues that hinder their reusability for further analyses and curation: the information which member a gene or compartment belongs to is not automatically retained, the flux bounds are changed due to normalization, and the abundance composition cannot be changed without re-running the merging procedure.

While standard formats for single organism metabolic models like SBML (Keating *et al.* 2020) have emerged and been widely adopted, no standard format for community metabolic models has been established yet. This not only poses a challenge for the compatibility and reuse of community metabolic models, but can also lead to a loss of information in the community model representation, as is the case for Commodelpy. Apart from retaining information, the abundances of community members and any changes to the original flux bounds need to be documented as well for the model to be reusable.

Here, we present PyCoMo, a python package for automated generation of compartmentalized community metabolic models. The generated models are suited for reuse and continuous improvement, as all information of the single organisms, as well as all community information, is stored in a format complying to existing standards. PyCoMo allows the quick calculation of the set of potential metabolite exchanges and resulting interactions

Received: 9 October 2023; Revised: 29 December 2023; Editorial Decision: 29 February 2024

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

across the entire community composition space, with a single flux variability analysis (FVA) run. We show an example of these functionalities by replicating the analysis of a previously published three species biogas community (Koch et al. 2019).

2 Description

PyCoMo is implemented in Python 3 and utilizes COBRApy (Ebrahim et al. 2013) for model input and output, as well as for fundamental structure and analysis methods. Three types of functionality for community metabolic modelling are implemented: (i) the preparation of the single organism models before merging into a community metabolic model, (ii) the generation of the community metabolic model in a structured format, and (iii) the contextualization and analysis of the community metabolic model (Fig. 1A). The resulting community model can then be used to predict the maximal community growth rate (at a given composition) or all feasible community compositions (at a given community growth rate).

The input for PyCoMo are the metabolic models of the community members, in any of the file formats supported by COBRApy (SBML, MATLAB, JSON, and YAML). The imported member models are prepared for merging, by prefixing all individual organism objects' (i.e. gene products, species, reactions, etc.) IDs with the name of the member they belong to. Further, a separate "medium compartment" is added, which

will allow metabolite exchange between organisms as well as the environment. Existing boundary reactions are converted into (internal) reactions transporting metabolites from the organisms' compartments to the medium compartment, and new boundary reactions are set for the medium compartment. Additionally, consistency checks and corrections for compliance with the SBML format (Keating et al. 2020) are performed.

Boundary metabolites of all members need to be matched in the medium compartment. PyCoMo resolves the matching either by ID, or by annotation. When matching by name, all metabolites with the same SBML ID are matched. Matching by annotation utilizes the information of database cross-referencing in the metabolite annotations. Such annotations are available to varying extent in all automatically generated metabolic models. The user can select a database to match all metabolites in the medium compartment with a reference to the same metabolite in the specified database. To avoid inconsistent matches, this process will only be followed through if all boundary metabolites map uniquely to the annotation database, otherwise PyCoMo stops and informs the user to use a different database or match the metabolites via names.

In a second step a community metabolic model object is created, using the prepared models as input. Here, all components of the models are merged. Metabolites, reactions, compartments and genes remain affiliated to their origin, due to the name change in the preparation.

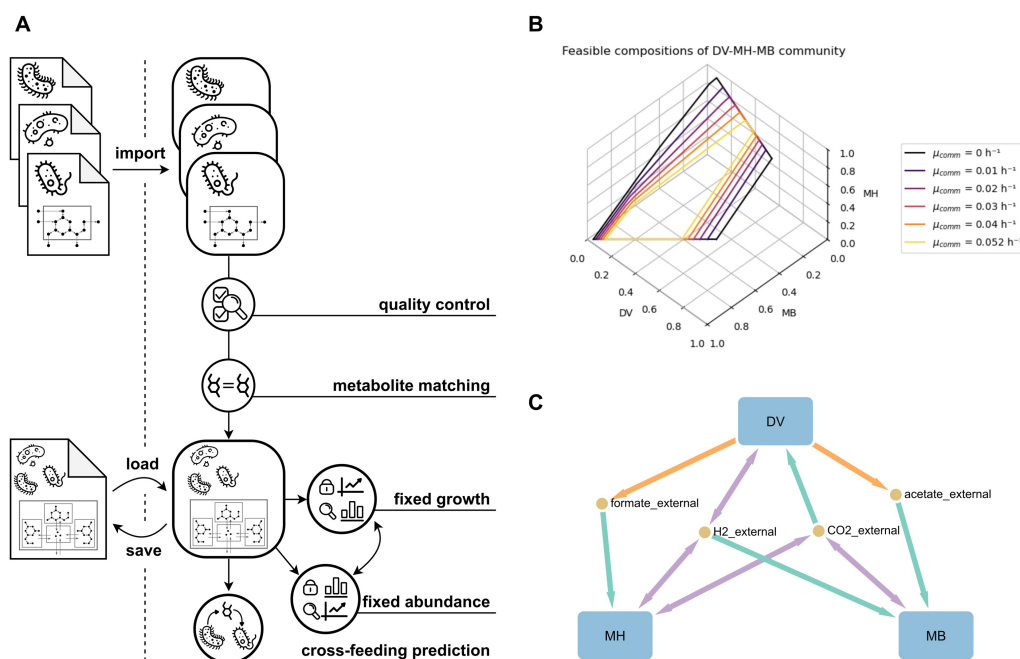


Figure 1. Summary of PyCoMo. (A) Overview of the automated steps in compartmentalized community metabolic model creation and the available analysis. All components of the final model can be traced back to their community member of origin. The community metabolic models can be saved and loaded into SBML format while retaining the original flux bounds, making the models reusable. Mass and charge balance of the entire model is checked, including all external metabolites matched between community members. The community metabolic model structure can be set to fixed growth or fixed abundance. All functionalities are available via a command-line interface. (B) Feasible compositions at different community growth rates in a three species biogas community (Koch et al. 2019). DV, *Desulfovibrio vulgaris*; MH, *Methanospirillum hungatei*; MB, *Methanosarcina barkeri*. The medium is specified to contain ethanol and CO₂, with ethanol being the substrate for *Desulfovibrio vulgaris*, which produces the substrates for the other community members. (C) Visualization of the community metabolic model and its cross-feeding interactions using ScyNet. The metabolite exchanges were calculated independently of the community composition and growth rate. Arrow colours denote feasible metabolite production (orange), uptake (green) or both (violet) by the communities' members.

Solving a community metabolic model leads to a nonlinear problem by default, due to the bilinear dependence of the community growth rate on the community abundance profile, and the growth rate of the individual members (Koch *et al.* 2019). To linearize the model, one can either hold the growth rate or the community member fraction constant by constraining the corresponding reactions (see [Supplementary Material](#)).

Typically, the fluxes in metabolic models are normalized to the (dry) mass of the system. In the context of microbial communities, this system mass represents the total dry mass of the community. Consequently, when consolidating multiple models, all cell-specific fluxes, including their corresponding bounds, need to be re-scaled to match the total community mass. This can be efficiently achieved by multiplying the cell-specific fluxes (alongside their bounds) by the mass fraction of the specific member in the community (Koch *et al.* 2019). However, when dealing with variable abundance profiles, flux bounds become dynamic, a characteristic not accommodated by the current SBML standard (level 3, Flux Balance Constraints Version 2 Release 2) (Olivier and Bergmann 2018). To address this, we convert each genome-scale metabolic model into an equivalent bound-free form, where flux bounds are represented as stoichiometric coefficients for newly introduced dummy metabolites. These metabolites are both produced and consumed by additional reactions, with their fluxes corresponding to the respective member's mass fraction within the community (refer to [Supplementary Material](#) for details). This approach aligns with current standards, ensuring seamless interchangeability among various openCOBRA tools.

The result of the merging and these additional steps is always a COBRApy model object, allowing the large environment of methods available to be utilized. On top of the COBRApy analysis methods, PyCoMo includes the functionality to calculate cross-feeding interactions in the community metabolic model. As community metabolic models have the additional parameter of community composition, typical steady-state, constraint-based modelling either only represents a single state of the community, or is computationally intensive, as it needs to account for the whole space of abundance combinations. By setting the model structure to variable community composition, but also removing the constraint of equal growth rate among the community members, FVA of the members' exchange reactions calculates the possible ranges of all metabolite exchanges and cross-feeding interactions. The resulting flux ranges may include cross-feeding interactions not feasible under balanced (community) growth if the production of metabolites is coupled to biomass production.

Community metabolic models generated by PyCoMo can easily be exported and transferred, as all the included export functionalities of COBRApy can be applied, and SBML format requirements are checked. PyCoMo was tested to work with input models of different origins [AGORA (Heinken *et al.* 2023), BIGG (King *et al.* 2016), CarveMe (Machado *et al.* 2018), PathwayTools (Karp *et al.* 2021), gapseq (Zimmermann *et al.* 2021)], although any model that can be imported using COBRApy will work. Further, community metabolic models generated by PyCoMo can be easily visualized using ScyNet (github.com/univieCUBE/ScyNet, manuscript submitted; see Fig. 1C), as all format requirements are fulfilled.

PyCoMo contains a command line interface, allowing to directly generate community metabolic models, reading all

member models from file and writing the resulting community metabolic model to file. All functionalities described above are available within the command line interface and thus facilitate an easy integration into existing workflows.

3 Results

As an example of a possible workflow, we have analysed a simplified biogas community of three species, aiming to construct a community metabolic model and compare all potential cross-feeding interactions to literature knowledge (Koch *et al.* 2019). The community consists of *Desulfovibrio vulgaris*, *Methanospirillum hungatei* and *Methanosarcina barkeri*, each representing a functional guild in the metabolic model. The members of the community are known to be actively cross-feeding, producing methane from ethanol or lactate, via the intermediary metabolites acetate, formate, H₂ and CO₂ (Koch *et al.* 2016). The models of the species members have been previously generated by Koch *et al.* (2019).

A community metabolic model is generated with PyCoMo following the workflow depicted in Fig. 1A. The input models are loaded from SBML files and the exchange metabolites are matched via identical IDs. The medium is set to contain ethanol and CO₂, as specified by Koch *et al.* (2019). The potential exchange metabolites and cross-feeding interactions were calculated. The result (Fig. 1B) shows that cross-feeding can occur via acetate, H₂, CO₂, and formate. For a clear representation of the interaction structure in the community, the community metabolic model and its cross-feeding interactions is visualized with ScyNet (Fig. 1C).

FVA of the generated community model shows that *D. vulgaris* can provide acetate and hydrogen to *M. barkeri*. These metabolites are known substrates for two different pathways for methanogenesis in *M. barkeri*, the acetoclastic pathway and the hydrogenotrophic pathway. Similarly, *D. vulgaris* can provide formate and hydrogen for *M. hungatei*, which are all processed via the hydrogenotrophic pathway.

The community metabolic model would also allow *D. vulgaris* to take up H₂ and CO₂. This uptake was not found in the analysis by Koch *et al.* (2019). While it is known that *D. vulgaris* can grow on acetate and hydrogen with sulphate as an electron acceptor, sulphate is not part of the growth medium, neither is acetate. PyCoMo could detect a thermodynamically infeasible cycle as the reason, allowing *D. vulgaris* to take up H₂ (see [Supplementary Material, Section 3](#)). In this cycle, *D. vulgaris* produces formate from H₂ and CO₂. Subsequently, *M. hungatei* converts the formate back into H₂ and CO₂, without the need for any other metabolites.

The performance of PyCoMo was tested using communities of different sizes (5–40 members). Genome-scale metabolic models from the AGORA collection were used as input models. Three tasks were timed on a system with 16 GB RAM and an AMD Ryzen 7 5700G using only a single core: the construction of the community metabolic model, flux balance analysis (FBA) at equal community member abundance, and the calculation of all potential exchange metabolites. The runtime scales approximately linear with community size for the construction of the community metabolic model and FBA, while the runtime of the calculation of all potential exchange metabolites scales quadratic, with a runtime of 11.3, 2.3, and 31.5 min respectively, at a community size of 40 members (average over three repetitions, see [Supplementary Material, Section 5](#)).

4 Conclusion

PyCoMo simplifies the generation and analysis of community metabolic models. This allows quickly addressing questions of which metabolites might be consumed, produced or exchanged in a community of interest, harnessing the knowledge base aspect of metabolic models on a community level. PyCoMo manages the construction of community models within a matter of seconds to minutes, yet the limitations of steady-state metabolic modelling methods apply. This means that simple optimization techniques like FBA will be feasible for larger communities, while the runtime for flux range and sampling approaches like FVA grows faster and can only be applied to smaller communities. Still, PyCoMo can find all exchange metabolites and cross-feeding interactions via FVA in a matter of 3 min for a 10-member community using genome-scale metabolic models.

Assembling community metabolic models with PyCoMo is reproducible and the resulting metabolic models can easily be used for further analysis with other metabolic modelling methods, including objectives other than biomass and objective free methods. All current openCOBRA file formats are supported for model input and export, with strict adherence to the SBML format. PyCoMo even facilitates (manual) curation on the level of communities by creating community metabolic models that fully preserve all available annotation and documentation in an interoperable, reusable manner adaptable to the necessary structures for different linearization approaches.

Finally, due to its command-line interface, PyCoMo can be easily integrated into (existing) genome analysis and phenotyping pipelines.

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest

None declared.

Funding

This work was supported by FWF doc.funds project DOC 69 “Microbial symbioses in dynamic environments: Metabolic interplay and novel interactions (MAINTAIN)” to M.P.; and the Austrian Climate and Energy Fund carried out under the program “Vorzeigeregion Energie” [project number: 887638 to M.M.].

Data availability

The data underlying this article are available in this article and its supplementary material, and at <https://pypi.org/project/pycomo/>.

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Bioinformatics, 2024, 40(4), btae153

<https://doi.org/10.1093/bioinformatics/btae153>

Advance Access Publication Date: 26 March 2024

Applications Note

References

- Bekiaris PS, Klamt S. Designing microbial communities to maximize the thermodynamic driving force for the production of chemicals. *PLoS Comput Biol* 2021;17:e1009093. <https://doi.org/10.1371/journal.pcbi.1009093>.
- Diener C, Gibbons SM, Resendis-Antonio O. MICOM: metagenome-scale modeling to infer metabolic interactions in the gut microbiota. *mSystems* 2020;5:1–17. <https://doi.org/10.1128/msystems.00606-19>.
- Ebrahim A, Lerman JA, Palsson BO *et al.* COBRApy: COntstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol* 2013;7:74. <https://doi.org/10.1186/1752-0509-7-74>.
- Heinken A, Basile A, Thiele I. Advances in constraint-based modelling of microbial communities. *Curr Opin Syst Biol* 2021;27:100346. <https://doi.org/10.1016/j.coisb.2021.05.007>.
- Heinken A, Hertel J, Acharya G *et al.* Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine. *Nat Biotechnol* 2023;41:1320–31. <https://doi.org/10.1038/s41587-022-01628-0>.
- Ibrahim M, Raajaraam L, Raman K. Modelling microbial communities: harnessing consortia for biotechnological applications. *Comput Struct Biotechnol J* 2021;19:3892–907. <https://doi.org/10.1016/j.csbj.2021.06.048>.
- Karp PD, Midford PE, Billington R *et al.* Pathway tools version 23.0 update: software for pathway/genome informatics and systems biology. *Brief Bioinform* 2021;22:109–26. <https://doi.org/10.1093/bib/bbz104>.
- Keating SM, Waltemath D, König M *et al.* SBML Level 3 Community Members. SBML level 3: an extensible format for the exchange and reuse of biological models. *Mol Syst Biol* 2020;16:e9110. <https://doi.org/10.15252/msb.20199110>.
- King ZA, Lu J, Dräger A *et al.* BiGG models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res* 2016;44:D515–22. <https://doi.org/10.1093/nar/gkv1049>.
- Koch S, Bendoric D, Fronk K *et al.* Predicting compositions of microbial communities from stoichiometric models with applications for the biogas process. *Biotechnol Biofuels* 2016;9:17–6. <https://doi.org/10.1186/s13068-016-0429-x>.
- Koch S, Kohrs F, Lahmann P *et al.* RedCom: a strategy for reduced metabolic modeling of complex microbial communities and its application for analyzing experimental datasets from anaerobic digestion. *PLoS Comput Biol* 2019;15:e1006759. <https://doi.org/10.1371/journal.pcbi.1006759>.
- Machado D, Andrejev S, Tramontano M *et al.* Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res* 2018;46:7542–53. <https://doi.org/10.1093/nar/gky537>.
- Olivier BG, Bergmann FT. SBML level 3 package: flux balance constraints version 2. *J Integr Bioinform* 2018;15:20170082. <https://doi.org/10.1515/jib-2017-0082>.
- Perez-Garcia O, Lear G, Singhal N. Metabolic network modeling of microbial interactions in natural and engineered environmental systems. *Front Microbiol* 2016;7:673.
- Saa P, Urrutia A, Silva-Andrade C *et al.* Modeling approaches for probing cross-feeding interactions in the human gut microbiome. *Comput Struct Biotechnol J* 2022;20:79–89. <https://doi.org/10.1016/j.csbj.2021.12.006>.
- Zimmermann J, Kaleta C, Waschina S. gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biol* 2021;22:81. <https://doi.org/10.1186/s13059-021-02295-1>.

A.3.4. Supplementary data

Supplementary material

PyCoMo: a python package for community metabolic model creation and analysis

Michael Predl^{1,2}, Marianne Mießkes^{3,4}, Thomas Rattei^{1,2,*}
and Jürgen Zanghellini^{3,4,*}

¹Division of Computational Systems Biology, Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria, ²Doctoral School in Microbiology and Environmental Science, University of Vienna, Vienna, Austria, ³Department of Analytical Chemistry, Faculty of Chemistry, University of Vienna, Vienna, Austria and ⁴Austrian Centre of Industrial Biotechnology, Vienna, Austria

*To whom correspondence should be addressed.

Availability and Implementation:

PyCoMo is freely available under an MIT licence at <http://github.com/univieCUBE/PyCoMo>, the Python Package Index and Zenodo (<https://doi.org/10.5281/zenodo.10431322>).

Contact: thomas.rattei@univie.ac.at, juergen.zanghellini@univie.ac.at

1. Equivalent, bound-free metabolic network

Consider a metabolic network represented by its stoichiometric matrix, $N \in \mathbb{R}^{m \times r}$ containing the net stoichiometric coefficients of m internal metabolites in r reactions. The vector $\mathbf{r} \in \mathbb{R}^r$ denotes a flux distribution (flux vector) through the network, and its components r_i with $i \in \{1, \dots, r\}$ are the respective reaction rates or fluxes.

In steady state, any feasible flux distribution through the network is a solution of the flux polyhedron given by

$$N\mathbf{r} = \mathbf{0}, \quad (1a)$$

$$\mathbf{l} \leq \mathbf{r} \leq \mathbf{u}, \quad (1b)$$

where the capacity constraints $\mathbf{l}$ and $\mathbf{u}$ denote the vectors of lower and upper bounds on the flux distribution. These limits also include irreversibility constraints, where $l_i = 0$ for irreversible reactions $i \in I_{\text{irr}}$.

By introducing the slack variable f and non-negative slack variable vectors $\mathbf{s}^l \in \mathbb{R}_{\geq 0}^r$ and $\mathbf{s}^u \in \mathbb{R}_{\geq 0}^r$ (associated with the lower and upper bounds of the reactions, respectively), the inhomogeneous set of equations in (1) can be transformed into a (higher-dimensional) set of homogeneous equations (Klamt et al.,

2017)

$$\begin{pmatrix} \mathbf{0} & N & \mathbf{0} & \mathbf{0} \\ -\mathbf{l} & I_r & -I_r & \mathbf{0} \\ \mathbf{u} & -I_r & \mathbf{0} & -I_r \end{pmatrix} \begin{pmatrix} f \\ \mathbf{r} \\ \mathbf{s}^l \\ \mathbf{s}^u \end{pmatrix} = \mathbf{0}, \quad (2a)$$

$$\mathbf{s}^l \geq \mathbf{0}, \quad (2b)$$

$$\mathbf{s}^u \geq \mathbf{0}. \quad (2c)$$

Here I_r denotes the identity matrix of size r .

By introducing the normalization condition

$$f = 1, \quad (3)$$

the solution spaces for $\mathbf{r}$ in (1) and (2) become identical. Thus, (2) can be interpreted as an equivalent, yet “bound-free” metabolic network, represented by the (transformed) stoichiometric matrix

$$\tilde{N} = \begin{pmatrix} \mathbf{0} & N & \mathbf{0} & \mathbf{0} \\ -\mathbf{l} & I_r & -I_r & \mathbf{0} \\ \mathbf{u} & -I_r & \mathbf{0} & -I_r \end{pmatrix},$$

and (transformed) flux vector

$$\tilde{\mathbf{r}} = \begin{pmatrix} f & \mathbf{r} & \mathbf{s}^l & \mathbf{s}^u \end{pmatrix}^T,$$

including the irreversible reactions $\mathbf{s}^l$, and $\mathbf{s}^u$. Note that for bound-free models of the community members the normalization

condition f introduced in equation 3 can also serve as a scaling factor, which enables one to rescale reaction fluxes and their bounds. This will prove useful in the context of the community metabolic models, where reaction fluxes and bounds are normalized relative to the community member's mass fraction, see (4a).

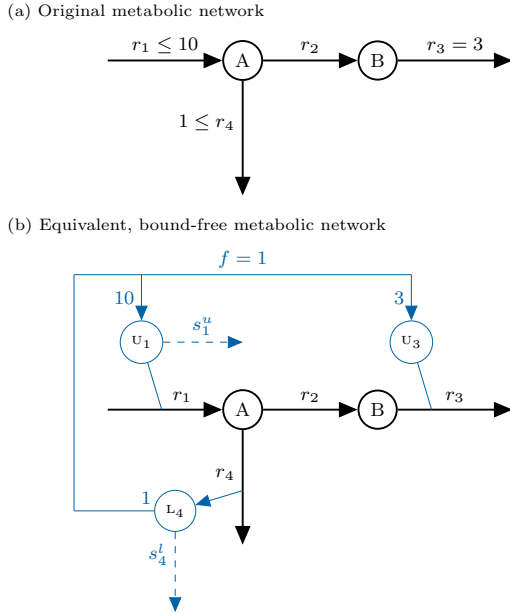


Fig. 1: Reconfiguring a constrained metabolic network (panel a) into an equivalent, bound-free metabolic network (panel b) via the set of Equations (2). Each non-standard bound (in panel a) is substituted with a dummy metabolite (blue circles in panel b). These dummy metabolites are either consumed (for upper bounds) or produced (for lower bounds) by their corresponding reactions. An additional reaction, labeled f (blue line), is introduced to balance these dummy metabolites. It both produces (for upper bounds) and consumes (for lower bounds) these dummy metabolites, with stoichiometric coefficients matched to the corresponding flux bound values. Finally, sink reactions (blue dashed arrows) are added to all dummy metabolites that do not correspond to equality constraints.

Example 1 Consider the simple constrained metabolic network depicted in Fig. 1a, composed of four irreversible reactions. Flux variability analysis (Mahadevan and Schilling, 2003) reveals the feasible ranges for reaction rates: r_1 spans (4 to 10) $\text{mmol g}^{-1} \text{h}^{-1}$, r_2 and r_3 are both $3 \text{ mmol g}^{-1} \text{h}^{-1}$, and r_4 ranges from (1 to 7) $\text{mmol g}^{-1} \text{h}^{-1}$.

The transformation of the toy network shown in Fig. 1a into its equivalent bound-free network using (2) yields the network depicted in Fig. 1b. Upon comparing Fig. 1a and Fig. 1b, it becomes apparent that the flux bounds in Fig. 1a translate into stoichiometric coefficients for newly introduced dummy metabolites, which are either produced or consumed by the new reaction f . These dummy metabolites function as co-factors,

and their availability constrains the respective original reactions r_1 to r_4 . Any excess can be removed through additional sink reactions, corresponding to the variables s^l and s^u . Note the simplification for the equality constraint $r_3 = 3 \text{ mmol g}^{-1} \text{h}^{-1}$: U_3 is fully consumed by r_3 , precisely determining the flux of r_3 and rendering both a sink for U_3 and a dummy metabolite for the lower bound unnecessary. For example, according to (3) reaction $f = 1 \text{ mmol g}^{-1} \text{h}^{-1}$, which produces $3 \text{ mmol g}^{-1} \text{h}^{-1} U_3$. U_3 can only be consumed by r_3 thus enforcing a flux of $3 \text{ mmol g}^{-1} \text{h}^{-1}$ through reaction r_3 . Consequently also $r_2 = 3 \text{ mmol g}^{-1} \text{h}^{-1}$. f also produces $10 \text{ mmol g}^{-1} \text{h}^{-1} U_1$, which can be consumed by r_1 or s_1^u , effectively imposing an upper limit of $10 \text{ mmol g}^{-1} \text{h}^{-1}$ on r_1 . Similarly, f drains $1 \text{ mmol g}^{-1} \text{h}^{-1} L_4$, which enforces a minimum flux of $1 \text{ mmol g}^{-1} \text{h}^{-1}$ through r_4 . Any surplus of L_4 is channeled through reaction s_4^l . We find a minimum flux of $r_1 = 4 \text{ mmol g}^{-1} \text{h}^{-1}$ and a maximum of $r_4 = 7 \text{ mmol g}^{-1} \text{h}^{-1}$ upon balancing these fluxes at metabolite A, ultimately fully recovering the limits obtained through flux variability analysis in the original network, as presented above.

(a) Equivalent, bound-free metabolic network of a constrained, reversible reaction

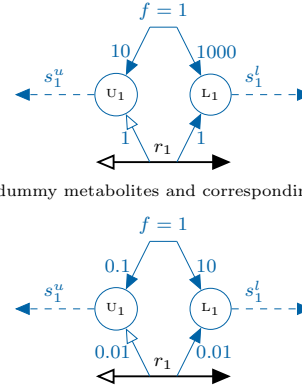


Fig. 2: Equivalent, bound-free metabolic network of a constrained, reversible reaction $-1000 \leq r_1 \leq 10$. Full and open arrows indicate forward and reverse directions of r_1 , dashed reactions represent sink reactions for the dummy metabolites L_1 , and U_1 . Panel a shows the reaction r_1 with a stoichiometry of 1 for the dummy metabolites U_1 and L_1 . Panel b shows the same reaction as implemented in PyCoMo, yet with stoichiometry coefficients for all dummy metabolites rescaled by a factor of 0.01. This scaling ensures reaction bounds equivalent to panel a, however, the maximum flux through the sink reactions is reduced from 1010 in panel a to 10.1 in panel b.

Example 2 In computational practice, the commonly adopted capacity constraints (l_i, u_i) are usually defined as $(-1000, 1000)$ for reversible reactions and $(0, 1000)$ for irreversible reactions. For instance, consider the reversible reaction $-1000 \leq r_1 \leq 10$, and its equivalent, bound-free network illustrated in Fig. 2a. Note that the negative lower flux bound produces the corresponding dummy metabolite L_1 by the reaction f .

Let's assume r_1 operates at its maximum forward flux. Consequently, r_1 produces $10 \text{ mmol g}^{-1} \text{ h}^{-1}$ of L_1 , while reaction f contributes $1000 \text{ mmol g}^{-1} \text{ h}^{-1}$ to the production of L_1 . Thus, s_1^l needs to drain off a total flux of $1010 \text{ mmol g}^{-1} \text{ h}^{-1}$. Such a scenario would be problematic if openCOBRA (Heirendt et al., 2019, 2017; Ebrahim et al., 2013; Jamshidi and Palsson, 2010) default constraints were applied to s_1^l .

To ensure that sink reactions s_i^l and s_j^u never encounter capacity constraints, we recommend employing considerably larger upper bounds for them, such as $10\,000 \text{ mmol g}^{-1} \text{ h}^{-1}$.

Increasing bounds for sink reactions can lead to sink reaction fluxes that are significantly larger than all other fluxes in the model, potentially causing numerical issues for the solver. PyCoMo addresses this by enabling the user to adjust the stoichiometric coefficients of associated auxiliary metabolites (Sun et al., 2013). For instance, in Figure 2b, the stoichiometric coefficients of the dummy metabolites have been reduced by a factor of 100 in all reactions producing those metabolites. Despite this adjustment, reaction bounds for r_1 are retained. The benefit of this scaling is that it avoids sink reaction fluxes higher than normal reaction fluxes, while still ensuring sufficient efflux for the dummy metabolites, avoiding potential numerical issues in the solver. The default scaling factor for dummy metabolite stoichiometry in PyCoMo is set to 10^{-2} as in the example above, but can be adjusted by users when generating a community metabolic model. The flux bounds for the sink reactions of dummy metabolites are scaled accordingly by PyCoMo, to ensuring that no capacity constraints are encountered.

2. Linearisation of community metabolic models

Fluxes in metabolic models are normalized by the dry mass of the model in question. Community metabolic models complicate this system, as the reaction constraints of all reactions of each community member need to be normalized by community mass, but also scaled by the mass fraction of the community member. Further, to comply with the steady-state assumption, the growth rate of all community members needs to be equal. The growth rate of each community member is the amount of this community member's newly produced biomass, divided by the existing mass. The concept of *balanced growth* states that a community can only be in steady state if all community members have equal growth rate on average (Koch et al., 2019). Otherwise, community members with higher growth rates will outgrow slower growing members, altering the community composition and thus violating the steady-state assumption. For a model with n community members m ; $m \in \{1, \dots, n\}$, the mass of the community is set as 1 and the mass of each community member m is equal to the mass fraction f_m . Each community member m has a biomass reaction $r_{biomass}^m$ and a growth rate μ_m . As such, in the community metabolic model setting the following constraints apply:

$$\sum_{m=1}^n f_m = 1, \quad (4a)$$

$$\mu_{com} = \mu_1 = \mu_2 = \dots = \mu_n, \quad (4b)$$

$$\mu_m = \frac{r_{biomass}^m}{f_m} \quad (4c)$$

Thus, the steady-state constraint leads to a bilinear problem in community metabolic models, as the relationship between

community member fraction, community member mass and growth rate is non-linear (see equation 4c).

2.1. Community and community member mass

To ensure efficient calculation, the model can be linearized by fixing either community member fractions or growth rate. Examples of the model structures can be seen in figures 3 - 5. First of all, the sum of abundance fractions of all community members needs to be 1. This is achieved by each community member's fraction reaction producing a fraction metabolite, which can be converted into a common fraction metabolite f_{com} . This common fraction metabolite is coupled to a demand reaction of flux 1, thus ensuring that the sum of fraction reaction fluxes equals 1 (see equations (4) and figure 3).

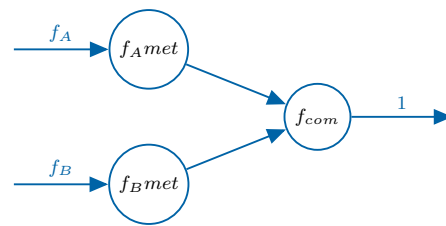


Fig. 3: Use of dummy metabolites to constrain the total mass of the community model. Fraction reactions f_m are introduced for each community member m . Fraction reactions have a flux corresponding to the mass fraction of the respective community member and are used to scale the bounds of each member's reactions. The sum of fraction reaction fluxes needs to be equal to the total mass of the community model, which is set to 1. To this end, each fraction reaction produces a dummy metabolite $f_m \text{ met}$ with a stoichiometric coefficient of 1. Each dummy metabolite $f_m \text{ met}$ is converted to a community mass dummy metabolite f_{com} , connected to a sink reaction of flux 1.

2.2. Fixed growth rate

Fixing the community growth rate is achieved by adding an upper bound metabolite to each of the community member's biomass reactions. These upper bound metabolites are not coupled to sink reactions, as they correspond to equality constraints. The production of these biomass bound metabolites is coupled with each community member's fraction reaction. The fraction reaction carries flux equal to the member's mass fraction in the community. The stoichiometric coefficient of the biomass bound metabolites is set equal to the community growth rate (see figure 4). This system leads to the production of biomass metabolites, scaled by the fraction of each member, in the end summing up to the total biomass required for the set growth rate.

$$\mu_{com} = \frac{r_{biomass}^{com}}{\sum_{m=1}^n f_m} = r_{biomass}^{com}, \quad (5a)$$

$$r_{biomass}^m = \mu_{com} * f_m, \quad (5b)$$

$$r_{biomass}^{com} = \sum_{m=1}^n r_{biomass}^m \quad (5c)$$

This further allows the fractions to be variable, enabling analysis across community compositions.

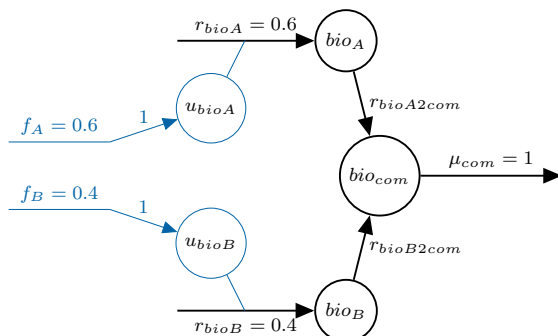


Fig. 4: Community metabolic model structure for fixed growth rate. An example with two community members, A and B, and a fixed community growth rate of 1 is shown. According to equations (5) the flux of the community member's biomass reactions, r_{bioM} is equal to $\mu_{com} * f_m$. In the example shown, solutions for the fraction reaction fluxes are 0.6 for A and 0.4 for B. This results in biomass reaction fluxes of 0.6 for A and 0.4 for B as well. The community biomass reaction flux is equal to the community growth rate μ_{com} , as well as the sum of the community members' biomass reaction fluxes.

2.3. Fixed community member mass fractions

Fixed community member fractions can be achieved by (i) constraining the fraction reactions to the respective fractions and (ii) introducing a reaction (r_{abd}) which produces upper bound metabolites for each of the community member's biomass reactions (see figure 5). The flux of reaction r_{abd} is equal to the community growth rate, but is variable in simulation. In r_{abd} the stoichiometric coefficients for the production of the biomass bound metabolites is equal to the fraction of the respective community member, fulfilling the constraints of equations (4) and (5). As in the fixed growth rate, these upper bound metabolites are not coupled to sink reactions, as they correspond to equality constraints.

3. Detection of thermodynamically infeasible cycles

Interactions in compartmentalized community metabolic models can create thermodynamically infeasible cycles, that cannot be detected in the member metabolic models alone. To accommodate for this circumstance, a detection of thermodynamically infeasible cycles is included in PyCoMo (accessible via the *get_loops* method of the *CommunityModel* class). The medium of the community metabolic model is set to be empty, allowing no influx of metabolites. Further, the constraints imposed by fixed growth or fixed abundance configurations on the fraction reactions and biomass reactions of the community members are lifted. Then, FVA is used to find reactions that can carry non-zero flux. An example output of the three species biogas community metabolic model (Koch et al., 2019), can be seen in table 1.

In total 31 reactions can carry flux while no medium is present. While all of them are part of thermodynamically infeasible cycles, not all of them are problematic. Many reactions are transport reactions, with a reversible transport being split into two irreversible reactions. As such, they create a cycle, transporting metabolites back and forth indefinitely (e.g. dv_Ac_ex

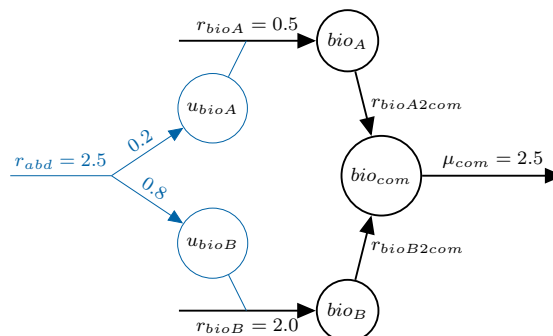


Fig. 5: Community metabolic model structure for fixed community member mass fractions. An example with two community members, A and B, and a fixed abundance profile of A : 0.2, B : 0.8 is shown. The stoichiometric coefficients for dummy metabolites in reaction r_{abd} are set to 0.2 for u_{bioA} and 0.8 for u_{bioB} . This leads to a balanced biomass production with variable growth rate, while following the constraints of equations (4) and (5). In this example, a community growth rate of 2.5 was chosen as solution. This results in a flux of 2.5 in r_{abd} , producing 0.5 of u_{bioA} and 2.0 of u_{bioB} . The production of dummy metabolites set the flux of the community biomass reactions accordingly. Ultimately, community biomass is produced in the form of 0.5 of bio_A and 2.0 of bio_B , summing up to a community growth rate μ_{com} of 2.5.

and dv_Ac_up). The same mechanism of reversible reactions being split into two irreversible counterparts can also be found in non-transport reactions, such as the conversion of fumarate to succinate and vice-versa (dv_Fum_Succ.dv_Internal.Species and dv_Succ_Fum.dv_Internal.Species). There is, however, one remaining thermodynamically infeasible cycle with a different pattern, spanning the community members *D. vulgaris* and *M. hungatei*: *D. vulgaris* converts formate into hydrogen and CO_2 , which are taken up by *M. hungatei* and turned into formate again. Upon detection of such cycles, users are strongly advised to curate the member and/or community metabolic models to resolve these cycles.

4. Matching of metabolite namespaces

Matching of boundary metabolites is a key step in allowing members of a community metabolic model to interact. PyCoMo resolves this by either matching via metabolite IDs, or matching via annotation. The latter will treat two boundary metabolites as equal if they have a matching identifier for a database selected by the user in the metabolite annotation. To avoid mapping metabolites to multiple different counterparts, the mapping fails if more than one identifier is present for the selected database in any boundary database. In addition, mass and charge balance is checked for all matched metabolites, and warnings are provided in case of violations.

The success rate of matching metabolite namespaces via annotations is dependent on the extent and similarity of metabolite annotation in the metabolic models about to be merged - not all databases (if any) are provided by all metabolic model generation pipelines, databases may change or provide new identifiers for existing compounds and the protonation state of

reaction	min_flux	max_flux
dv_TP_H2_TP_dv_External_Species	-10.0	0.0
dv_TP_CO2_TP_dv_External_Species	-10.0	0.0
dv_TP_Form_TP_dv_External_Species	0.0	10.0
dv_Fum__Succ_dv_Internal_Species	0.0	10.0
dv_Succ__Fum_dv_Internal_Species	0.0	10.0
dv_Ac_ex	0.0	0.5
dv_CO2_ex	0.0	10.0
dv_Eth_up	0.0	10.0
dv_Form_ex	0.0	10.0
dv_H2_up	0.0	10.0
dv_Ac_up	0.0	0.5
dv_H2_ex	0.0	10.0
dv_Eth_ex	0.0	10.0
dv_Fdred__H2_dv_Internal_Species	-10.0	0.0
dv_Fdred__Form_dv_Internal_Species	0.0	10.0
dv_CO2_up	0.0	10.0
mb_CO2_up	0.0	10.0
mb_CO2_ex	0.0	10.0
mh_TP_H2_TP_mh_External_Species	0.0	10.0
mh_TP_CO2_TP_mh_External_Species	0.0	10.0
mh_TP_Form_TP_mh_External_Species	-10.0	0.0
mh_Form__CO2_mh_Internal_Species	0.0	10.0
mh_Fum__Succ_mh_Internal_Species	0.0	10.0
mh_Succ__Fum_mh_Internal_Species	0.0	10.0
mh_NADP_Hydrogenase_mh_Internal_Species	0.0	10.0
mh_CO2__Form_mh_Internal_Species	0.0	10.0
mh_CO2_ex	0.0	10.0
mh_CO2_up	0.0	10.0
mh_Form_up	0.0	10.0
mh_H2_ex	0.0	10.0
mh_H2_up	0.0	10.0

Table 1. Reactions carrying flux while no medium is present. The results shown were generated using a three species biogas community metabolic model (Koch et al., 2019).

compounds can differ between databases, making exact matches impossible.

Thus, the matching of boundary metabolites via annotation provided by PyCoMo is intended for two purposes: (1) If metabolic models were generated by the same model generation pipeline, and annotations are present and unique in a given database for all boundary metabolites, this matching strategy should succeed and allow matching of metabolites regardless of their metabolite ID. (2) When creating a community metabolic model using metabolic models of different metabolic model generation pipelines, where most annotations are present and all are unique in a given database for all boundary metabolites, then matching via annotation can resolve at least the majority of metabolite matching without manual input. As mass and charge balance are checked, all metabolite matches found by PyCoMo should be correct. Users should then resolve all matches that were discovered to be not mass and charge balanced, as well as check all unmatched metabolites for potential missed matches due to any of the reasons provided before.

5. Runtime and complexity

The runtime complexity in relation to community size was estimated for three separate tasks: (i) The construction of the community metabolic model, (ii) flux balance analysis (FBA) at equal community member abundance, and (iii) the calculation of all potential exchange metabolites. The runtime was measured on a system running Microsoft Windows 11 with 16GB RAM and an AMD Ryzen 7 5700G using only a single core. To include the effects of diverse input model structures and the complexity of genome-scale metabolic models, the single organism metabolic models were randomly selected from the AGORA (Magnúsdóttir et al., 2017) collection. The three tasks were timed for communities of size 5 to 40 and repeated three times (each with randomly selected members). Figure 6 shows the results of the runtime measurements. The runtime complexity of the community metabolic model construction appears to be roughly linear, while FBA and the calculation of all potential exchange metabolites scale quadratically with the number of community members. It should be noted that fluctuations of an order of magnitude have been observed for the calculation of all potential exchange metabolites. This variability could stem from the differences in community members, but also the implementation of flux variability analysis of COBRApy, used for the calculation in PyCoMo. Further, the

runtime of flux variability analysis (FVA) can be decreased by making use of the parallel implementation in COBRApy. This parallel implementation is not as effective on Windows operating systems as on Linux or MacOS, due to operation system specific implementations of the underlying multiprocessing python library.

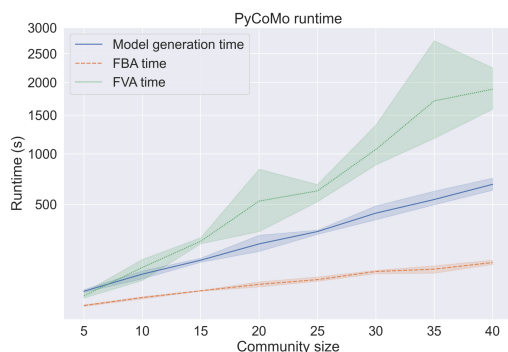


Fig. 6: Runtime measurements for PyCoMo. Single organism metabolic models were randomly sampled from the AGORA collection to generate communities of 5 to 40 members (each in triplicate). The runtime was measured for three tasks of PyCoMo: (i) The construction of the community metabolic model, (ii) flux balance analysis (FBA) at equal community member abundance, and (iii) the calculation of all potential exchange metabolites using flux variability analysis (FVA).

References

- A. Ebrahim, J. A. Lerman, B. O. Palsson, and D. R. Hyduke. Cobrapy: constraints-based reconstruction and analysis for python. *BMC systems biology*, 7:1–6, 2013.
- L. Heirendt, I. Thiele, and R. M. Fleming. Distributedfba.jl: high-level, high-performance flux balance analysis in julia. *Bioinformatics*, 33(9):1421–1423, 2017.
- L. Heirendt, S. Arreckx, T. Pfau, S. N. Mendoza, A. Richelle, A. Heinken, H. S. Haraldsdóttir, J. Wachowiak, S. M. Keating, V. Vlasov, et al. Creation and analysis of biochemical constraint-based models using the cobra toolbox v. 3.0. *Nature protocols*, 14(3):639–702, 2019.
- N. Jamshidi and B. Ø. Palsson. Mass action stoichiometric simulation models: incorporating kinetics and regulation into stoichiometric models. *Biophysical journal*, 98(2):175–185, 2010.
- S. Klamt, G. Regensburger, M. P. Gerstl, C. Jungreuthmayer, S. Schuster, R. Mahadevan, J. Zanghellini, and S. Müller. From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints. *PLoS computational biology*, 13(4):e1005409, 2017.
- S. Koch, F. Kohrs, P. Lahmann, T. Bissinger, S. Wendschuh, D. Benndorf, U. Reichl, and S. Klamt. RedCom: A strategy for reduced metabolic modeling of complex microbial communities and its application for analyzing experimental datasets from anaerobic digestion. *PLoS Computational Biology*, 15(2): e1006759, Feb. 2019. doi: 10.1371/journal.pcbi.1006759.
- S. Magnúsdóttir, A. Heinken, L. Kutt, D. A. Ravcheev, E. Bauer, A. Noronha, K. Greenhalgh, C. Jäger, J. Baginska, P. Wilmes, R. M. T. Fleming, and I. Thiele. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*, 35(1):81–89, 2017. doi: 10.1038/nbt.3703. Publisher: Nature Publishing Group.
- R. Mahadevan and C. H. Schilling. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic engineering*, 5(4):264–276, 2003.
- Y. Sun, R. M. Fleming, I. Thiele, and M. A. Saunders. Robust flux balance analysis of multiscale biochemical reaction networks. *BMC bioinformatics*, 14(1):1–6, 2013.

A.4. ScyNet: Visualising interactions in community metabolic models

A.4.1. Publication status

Submitted for publication: Predl, M., Gandolf, K., Hofer, M., Rattei, T., *ScyNet: Visualising interactions in community metabolic models*. Submitted for publication in Bioinformatics Advances, confirmation of receipt: 29.04.2024

A.4.2. Author contributions

The author of this dissertation conceived of the study and software design, finished the implementation of the software based on a proof of concept version, gathered and analyzed the test data, designed the figure and wrote the manuscript.

A.4.3. Publication

ScyNet: Visualising interactions in community metabolic models

Michael Predl^{1,2}, Kilian Gandolf¹, Michael Hofer¹ and Thomas Rattei^{1,2,*}

¹Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria

²Doctoral School in Microbiology and Environmental Science, University of Vienna, Vienna, Austria

*To whom correspondence should be addressed.

Abstract

Motivation: Genome-scale community metabolic models are used to gain mechanistic insights into interactions between community members. However, existing tools for visualising metabolic models only cater to the needs of single organism models.

Results: ScyNet is a Cytoscape app for visualising community metabolic models, generating networks with reduced complexity by focusing on interactions between community members. ScyNet can incorporate the state of a metabolic model via fluxes or flux ranges, which is shown in a previously published simplified cystic fibrosis airway community model.

Availability and Implementation: ScyNet is freely available under an MIT licence and can be retrieved via the Cytoscape App Store (<https://apps.cytoscape.org/apps/scynet>). The source code is available at Github (github.com/univieCUBE/ScyNet).

Contact: thomas.rattei@univie.ac.at

Keywords: Metabolic model, Microbial community, Visualization, Cytoscape

Introduction

Interactions play an important role in shaping the structure and properties of microbial communities (Mataigne et al., 2021). They can range from beneficial to detrimental for the community members and include various forms from direct interactions, like cross-feeding, to complex traits, like biofilm formation (Zuñiga et al., 2017). Due to these interactions,

communities of different taxa have emergent properties beyond the sum of their members' capabilities, and are actively researched in diverse areas such as microbial ecology, medicine and biotechnology (Heinken et al., 2021). In recent years the field of metabolic modelling has expanded from modelling single organisms to modelling communities, with the aim to gain mechanistic insights. Metabolic modelling has been applied to a variety of communities, including synthetic communities, human and environmental microbiomes, with community members spanning all domains of life (Heinken et al., 2021; Ibrahim et al., 2021; Saa et al., 2022).

Visualisation of metabolic models allows communicating concepts and results. However, metabolic models and especially genome-scale metabolic models are highly complex with thousands of metabolites and reactions. Community metabolic models only exacerbate this problem (Ibrahim et al., 2021). Many tools have been developed for the visualisation of single organism models trying to deal with this complexity (Hari & Lobo, 2020; König et al., 2012; Noronha et al., 2017; Pfau et al., 2021). One notable mention is cy3sbml (König et al., 2012), an app for the widely used network data visualisation platform Cytoscape (Shannon et al., 2003), which allows for importing and visualising metabolic models with all their information.

To avoid complex figures, a common technique is the reduction of the displayed network, e.g. highlighting a single pathway as done by IDARE2 (Pfau et al., 2021) or showing only the immediate neighbourhood of a reaction or metabolite as done by Fluxer (Hari & Lobo, 2020). These reduction approaches are not well suited for community metabolic models, as they focus only on a local area, missing the context of the community. Instead, a visualisation of the interactions would capture the essence of a community metabolic model while also drastically reducing the complexity of the figure. To our knowledge, no tool exists so far, which allows for automated visualisation of interactions between community members from metabolic models.

We present ScyNet, a Cytoscape app for visualising the exchange metabolite network of community metabolic models and the interactions between community members. ScyNet reduces the metabolic network of the whole community to community members and their consumed and produced metabolites, drastically decreasing the complexity. Flux data from flux balance analysis (FBA) and flux ranges from flux variability analysis (FVA) can be used to contextualise the visualisation. ScyNet is accompanied by PyCoMo (Predl et al., 2024), a python package for automated generation of a community metabolic model from single organism models, which can readily be used as input. Together they allow for quick

visualisation and analysis of microbial communities and their interactions starting with only the metabolic models of their members.

Description

ScyNet is a Java app developed for Cytoscape 3.9.0 or higher ([Shannon et al., 2003](#)). The input is a compartmentalised community metabolic model in SBML format, such as generated by PyCoMo. To distinguish different organisms in the community metabolic model, two requirements need to be met by the input SBML file: (1) The inclusion of the name of the community member as prefix in its metabolite and compartment IDs and (2) the presence of a shared external compartment between the community members called “medium”. The ID of the shared external compartment can be set to a name other than “medium”, by specifying the ID in the SBML parameter “shared_compartment_id”. Additionally, reactions transporting metabolites to and from the shared external compartment are assumed to carry only a single metabolite.

The ScyNet visualisation workflow is started with the import of the community metabolic model in Cytoscape, which is handled by Cy3sbml (König et al., 2012) (figure 1: step 1). ScyNet reduces the resulting network to its community members and their consumed and produced metabolites (figure 1: step 2). First, the nodes for each community member are generated by scanning the model for members in metabolite and compartment IDs. Next, reactions connected to metabolites in the shared external compartment are gathered. For each reaction, nodes are created for the external metabolites and connected to the corresponding community members node.

The simplified network can then be structured and stylized using a variety of methods provided by Cytoscape and its apps. ScyNet includes a tailored layout algorithm. It places the metabolites and organisms in four concentric circles: Going inwards, the first and largest circle contains all the external metabolites that are only connected to a single community member. The second circle are the community member nodes, evenly spaced. The third circle is reserved for metabolites connected to exactly two members. Such metabolites are instead placed between two member nodes, if they are adjacent in the layout, which is encouraged by the algorithm. The inner circle contains all remaining metabolites, with three or more interaction partners. This layout algorithm can be applied to all networks generated by ScyNet via the menu.

In addition, ScyNet allows styling of the network according to network state. For this, a flux vector can be imported and ScyNet will colour the edges accordingly. An example of such a contextualised network can be seen in figure 1, step 3. Networks can be stylized according

to reaction fluxes of a single state by providing a flux vector as a tab-separated file containing two columns: “reaction_id” and “flux”. Based on the reaction fluxes, all edge arrows and colours are set to represent the flux state. Reactions that do not carry flux are dropped from the visualisation. Similarly, flux ranges, as in the output of FVA can be visualised as well. Here, a tab separated file with three columns (“reaction_id”, “min_flux” and “max_flux”) is required.

In addition to the indication of reaction directionality by colour, ScyNet can visualise flux values by setting the edge width proportionally.

When flux data are present, the network can be further simplified to only visualise cross-feeding interactions, which are interactions of a community member producing a metabolite, which is taken up by a different community member. This definition allows more than the required two community members to be part of cross-feeding interactions of the same metabolite. Further, cross-feeding interactions are considered regardless of the presence of the target metabolite in the medium. A visualisation of the cross-feeding network can be achieved via the toggle cross-feeding function, hiding all reactions and metabolites not involved in cross-feeding. The same function can be used to reveal the hidden components of the network again.

Results

For testing ScyNet we chose to visualise a microbial community of cystic fibrosis airways, illustrating a typical visualisation use case of a small microbial community metabolic model. The microbial community was modelled using AGORA 1.02 models (Magnúsdóttir et al., 2017) as representatives for each genus, as done by Henson et al (Henson et al., 2019). The community consists of 17 member metabolic models, of which only the four key members (*Prevotella*, *Pseudomonas*, *Staphylococcus*, *Streptococcus*) identified by Henson et al. were visualised.

The single species models were merged into a community metabolic model using PyCoMo. All possible exchange reaction flux ranges were calculated and written to a tab separated file for input in ScyNet. The total community metabolic model of the 4 members has 4787 metabolites and 5636 reactions (figure 1: step 1). In the simplified network created by ScyNet these numbers are reduced to 234 metabolites and 568 reactions (figure 1: step 2).

After applying the possible exchange reaction flux ranges in the community identified by PyCoMo and removing infeasible reactions, the model contains only 194 metabolites and

470 reactions (figure 1: step 3). By reducing the visualisation to only cross-fed metabolites the network shrinks to 68 nodes and 222 edges.

Henson et al. analysed 16S rRNA data of 75 sputum samples of adult cystic fibrosis patients. This data showed that the community members can be present in varying compositions, which not only differ in flux amount, but also in activity and direction of exchange reactions, according to simulations of the metabolic models. We reproduced the visualisation in figure 8D, showing predicted cross-feeding fluxes in the context of communities with high *Pseudomonas* abundance (figure 1: step 4) as an example. An exact replication of the flux vector is not possible, as it is the average of the flux vectors across 500 simulations with randomly permuted reaction bounds. Instead, the flux vector was derived manually from the results of Henson et al., putting a flux of 1 for exchange reactions shown in figure 8D and omitting all others.

In the scenario of high abundance of *Pseudomonas*, alanine is predicted to be produced by *Staphylococcus*, *Prevotella* and *Streptococcus* and taken up by *Pseudomonas*. However at low abundance of *Pseudomonas* and increased abundance of *Streptococcus* the results of Henson et al. predict *Staphylococcus* to take up alanine produced solely by *Streptococcus*. This example reflects the flexibility of the direction of cross-feeding in this community. As shown in figure 1B, the majority of reactions that are part of cross-feeding interactions are feasible in both directions (77,5%). The reaction reversibility is not a general feature of external metabolites in the AGORA models. The fraction of reversible reactions is significantly lower in exchange reactions not involved in potential cross-feeding (8,1%), compared to exchange reactions of cross-fed metabolites (P-value << 0.05).

Conclusion

ScyNet enables comprehensible visualisation of community metabolic models and the interactions of their members. The reduction in complexity makes the visualisations suited for communication of results, as shown in the reproduced figure of the four member cystic fibrosis airways microbial community by Henson et al. (2019). The visualisation process takes less than 5 minutes on a desktop computer for the given example, including the time needed to generate the community metabolic model from the individual AGORA reconstructions. Thus, ScyNet can be readily applied during the construction, curation and analysis of community metabolic models.

ScyNet also allows to visualise different metabolic states of a community, e.g. for comparing communities under distinct conditions, different abundance profiles, growth on distinct media or showing FBA and FVA solutions. The contextualisation of the reduced network could be

extended by OMICs data integration as achieved by IDARE2 (Pfau et al., 2021), but focused on the community level.

Most community metabolic modelling applications are done in small to medium communities of less than 20 members (Heinken et al., 2021). Natural communities often consist of more than 100 different taxa. Few studies have already attempted to model such complex communities with metabolic modelling and it is to be expected that more work will follow.

While ScyNet drastically reduces the complexity of community metabolic networks, the amount of external metabolites and reactions grows with the size of the community (Ibrahim et al., 2021). The larger the network is, the less informative will be its visualisation.

To ensure that the represented network can be read at one glance, further reduction is necessary. An example is the selection of the five most frequent cross-feeding metabolites by Henson et al (2019). Alternatively, modularisation and abstraction can be employed to deal with larger communities. Such coarse-graining could be achieved by combining organisms to functional guilds or grouping metabolites in categories.

Despite the challenge of highly complex and information rich community metabolic models, ScyNet allows for quick visualisation, analysis and communication of interactions between community members.

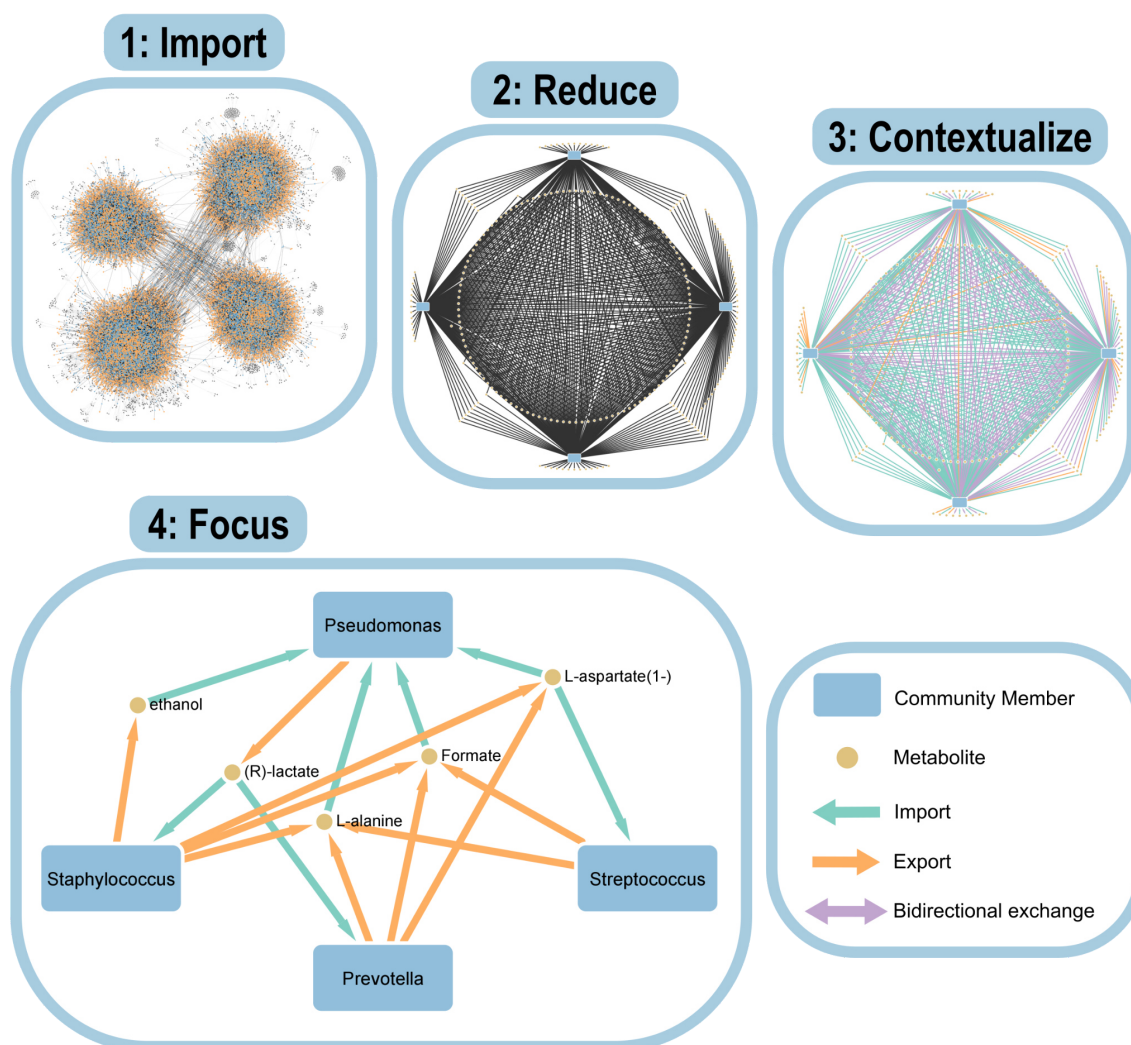


Figure 1: Workflow of visualising community metabolic networks with ScyNet. All networks are based on the same four AGORA models (Magnúsdóttir et al., 2017), converted into a community metabolic model using PyCoMo. The community members were chosen as representatives of their respective genus in a cystic fibrosis airways community model, by Henson et al (Henson et al., 2019). (Step 1) The complete network of the community metabolic model, as imported by cy3sbml (König et al., 2012). (Step 2) The network after reduction by ScyNet, containing only community members, consumed and produced metabolites as nodes, and exchange reactions as edges. (Step 3) Visualisation of flux ranges of potential exchange reactions, independent of the community composition and calculated with PyCoMo. All reactions with non-zero flux are shown. (Step 4) Replication of figure 8D by Henson et al.: Visualisation of the five metabolites identified by Henson et al. as most often cross-fed in simulations with randomly permuted exchange reaction flux bounds of samples with high abundance of *Pseudomonas*.

References

- Hari, A., & Lobo, D. (2020). Fluxer: A web application to compute, analyze and visualize genome-scale metabolic flux networks. *Nucleic Acids Research*, 48(W1), W427–W435. <https://doi.org/10.1093/nar/gkaa409>
- Heinken, A., Basile, A., & Thiele, I. (2021). Advances in constraint-based modelling of microbial communities. *Current Opinion in Systems Biology*, 27. <https://doi.org/10.1016/j.coisb.2021.05.007>
- Henson, M. A., Orazi, G., Phalak, P., & O'Toole, G. A. (2019). Metabolic Modeling of Cystic Fibrosis Airway Communities Predicts Mechanisms of Pathogen Dominance. *mSystems*, 4(2), e00026-19. <https://doi.org/10.1128/mSystems.00026-19>
- Ibrahim, M., Raajaraam, L., & Raman, K. (2021). Modelling microbial communities: Harnessing consortia for biotechnological applications. *Computational and Structural Biotechnology Journal*, 19, 3892–3907. <https://doi.org/10.1016/j.csbj.2021.06.048>
- König, M., Dräger, A., & Holzhütter, H.-G. (2012). CySBML: A Cytoscape plugin for SBML. *Bioinformatics*, 28(18), 2402–2403. <https://doi.org/10.1093/bioinformatics/bts432>
- Magnúsdóttir, S., Heinken, A., Kutt, L., Ravcheev, D. A., Bauer, E., Noronha, A., Greenhalgh, K., Jäger, C., Baginska, J., Wilmes, P., Fleming, R. M. T. T., & Thiele, I. (2017). Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*, 35(1), 81–89. <https://doi.org/10.1038/nbt.3703>
- Mataigne, V., Vannier, N., Vandenkoornhuysen, P., & Hacquard, S. (2021). Microbial Systems Ecology to Understand Cross-Feeding in Microbiomes. *Frontiers in Microbiology*, 12. <https://www.frontiersin.org/articles/10.3389/fmicb.2021.780469>
- Noronha, A., Daníelsdóttir, A. D., Gawron, P., Jóhannsson, F., Jónsdóttir, S., Jarlsson, S., Gunnarsson, J. P., Brynjólfsson, S., Schneider, R., Thiele, I., & Fleming, R. M. T. (2017). ReconMap: An interactive visualization of human metabolism. *Bioinformatics*, 33(4), 605–607. <https://doi.org/10.1093/bioinformatics/btw667>

- Pfau, T., Galhardo, M., Lin, J., & Sauter, T. (2021). IDARE2—Simultaneous Visualisation of Multiomics Data in Cytoscape. *Metabolites*, 11(5), Article 5.
<https://doi.org/10.3390/metabo11050300>
- Predl, M., Mießkes, M., Rattei, T., & Zanghellini, J. (2024). PyCoMo: A python package for community metabolic model creation and analysis. *Bioinformatics*, 40(4), btae153.
<https://doi.org/10.1093/bioinformatics/btae153>
- Saa, P., Urrutia, A., Silva-Andrade, C., Martín, A. J., & Garrido, D. (2022). Modeling approaches for probing cross-feeding interactions in the human gut microbiome. *Computational and Structural Biotechnology Journal*, 20, 79–89.
<https://doi.org/10.1016/j.csbj.2021.12.006>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11), 2498–2504. <https://doi.org/10.1101/gr.1239303>
- Zuñiga, C., Zaramela, L., & Zengler, K. (2017). Elucidation of complexity and prediction of interactions in microbial communities. *Microbial Biotechnology*, 10(6), 1500–1522.
<https://doi.org/10.1111/1751-7915.12855>

A.5. Expanding the cultivated human archaeome by targeted isolation of novel *Methanobrevibacter* strains from fecal samples

A.5.1. Publication status

Preprint publication: Duller, S., Vrbancic, S., Szydlowski, L., Mahnert, A., Blohs, A., Predl, M., Kumpitsch, C., Zrim, V., Hoegenauer, C., Kosciolk, T., Schmitz, R. A., Eberhard, A., Dragovan, M., Schmidberger, L., Zurabishvili, T., Weinberger, V., Moser, A. M., Kolb, D., Pernitsch, D., Mohammadzadeh, R., Kuehnast, T., Rattei, T., Moissl-Eichinger, C., *Expanding the cultivated human archaeome by targeted isolation of novel *Methanobrevibacter* strains from fecal samples*. bioRxiv 2024.04.10.588852; DOI: <https://doi.org/10.1101/2024.04.10.588852>

Submitted for publication: Duller, S., Vrbancic, S., Szydlowski, L., Mahnert, A., Blohs, A., Predl, M., Kumpitsch, C., Zrim, V., Hoegenauer, C., Kosciolk, T., Schmitz, R. A., Eberhard, A., Dragovan, M., Schmidberger, L., Zurabishvili, T., Weinberger, V., Moser, A. M., Kolb, D., Pernitsch, D., Mohammadzadeh, R., Kuehnast, T., Rattei, T., Moissl-Eichinger, C., *Expanding the cultivated human archaeome by targeted isolation of novel *Methanobrevibacter* strains from fecal samples*. Submitted for publication in Nature Communications, confirmation of receipt: 11.04.2024

A.5.2. Author contributions

The author of this dissertation planned and carried out the metabolic modelling of pairwise co-cultures. He also wrote part of the manuscript.

A.5.3. Publication

Expanding the cultivated human archaeome by targeted isolation of novel *Methanobrevibacter* strains from fecal samples

Stefanie Duller¹, Simone Vrbancic¹, Łukasz Szydlowski^{2,3}, Alexander Mahnert^{1,4}, Marcus Blohs¹, Michael Predl^{5,6}, Christina Kumpitsch^{1,2}, Verena Zrim⁷, Christoph Högenauer⁸, Tomasz Kosciolk^{2,3,9}, Ruth A. Schmitz¹⁰, Anna Eberhard¹, Melanie Dragovan¹, Laura Schmidberger¹, Tamara Zurabishvili¹, Viktoria Weinberger¹, Adrian Mathias Moser⁸, Dagmar Kolb¹¹, Dominique Pernitsch¹¹, Rokhsareh Mohammadzadeh¹, Torben Kühnast¹, Thomas Rattei⁵, Christine Moissl-Eichinger^{1,4*}

¹ D&R Institute of Hygiene, Microbiology and Environmental Medicine, Medical University of Graz, Graz, Austria;

² Malopolska Centre of Biotechnology, Jagiellonian University in Krakow, Poland;

³ Sano Centre for Computational Medicine, Krakow, Poland

⁴ BioTechMed Graz, Graz, Austria;

⁵ Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria;

⁶ Doctoral School Microbiology and Environmental Science, University of Vienna, Vienna, Austria

⁷ Center for Medical Research, Medical University of Graz, Graz, Austria;

⁸ Division of Gastroenterology and Hepatology, Department of Internal Medicine, Medical University of Graz, Graz, Austria;

⁹ Department of Data Science and Engineering, Silesian University of Technology, Gliwice, Poland;

¹⁰ Institute for General Microbiology, Christian Albrechts University, Kiel, Germany;

¹¹ Core Facility Ultrastructure Analysis, Medical University of Graz, Graz, Austria;

*Corresponding author

christine.moissl-eichinger@medunigraz.at

Abstract (250 words)

Archaea are integral components of the human microbiome but persist as understudied entities within the gastrointestinal tract (GIT), primarily due to the lack of cultured representatives for comprehensive mechanistic investigations. With only four *Methanobrevibacter smithii* isolates from humans available according to the Global Catalogue of Microorganisms (GCM), the existing cultures fail to adequately represent the observed diversity, as underscored by recent findings.

This study introduces a targeted cultivation method for enriching methanogenic archaea from human fecal samples. Applied to 16 stool samples from healthy and diseased donors, the method aimed to genomically characterize the archaeal cultures and establish correlations with gastrointestinal disorders. The procedure combines methane breath testing, *in silico* metabolic modelling, media optimization, FACS, dilution series, and genomic sequencing through Nanopore technology. Additional analyses include co-cultured bacteriome, comparative genomics of archaeal genomes, functional comparisons, and structure-based protein function prediction of unknown differential traits.

Successful establishment of stable archaeal cultures from 14 out of 16 fecal samples yielded nine previously uncultivated strains, eight of which were absent from a recent archaeome genome catalog. Comparative genomic and functional assessments of *Methanobrevibacter smithii* and *Candidatus* Methanobrevibacter intestini strains from diverse participant cohorts revealed features potentially associated with gastrointestinal diseases.

This work substantially broadens the scope of available archaeal representatives for functional and mechanistic studies in the human GIT. The established protocol facilitates the cultivation of methanogenic archaea from nearly every human fecal sample, offering insights into the adaptability of *Candidatus* Methanobrevibacter intestini genomes in critical microbiome situations.

Introduction

The human gut is a complex ecosystem harboring a diverse community of microbes, including bacteria, fungi, viruses, and archaea (1). While bacteria are well described, and their role in gut health is well studied, the function and role of archaea in the human gastrointestinal tract (GIT) is still not well understood. Archaea found in the GIT are mainly represented by the order Methanobacteriales, with the most abundant species *Methanobrevibacter smithii* and *Candidatus Methanobrevibacter intestini* (2–4).

These methanogenic archaea produce methane (CH₄) as a metabolic byproduct, while producing ATP via methanogenesis (5–9). The yielded CH₄ serves as an important indicator for the presence of methanogenic archaea (10).

CH₄ can be detected in humans' breath, as it is partially absorbed into the blood and excreted through the lungs (11). The impact of CH₄ on the human GIT has already been addressed and it has been reported that CH₄ leads to the reduced efficiency of intestinal smooth muscle activity, often associated with e.g. the constipation-predominant irritable bowel syndrome (12–14). Further, variations of CH₄ levels in the human breath were discussed as an indicator for enhanced oxidative stress and could therefore show a great potential for example in *in vivo* diagnostics (15–17). On the other hand, positive effects were also described, as CH₄ could also act as antioxidant, protecting cells from reactive oxygen species (ROS) and demonstrating potential anti-inflammatory properties (18).

Methanobrevibacter species show a prevalence of more than 90% in fecal samples with an average abundance of 0.56 % for *Methanobrevibacter smithii* and 0.13 % for *Cand. Methanobrevibacter intestini* (3,19). In high-CH₄ producing individuals (up to 40% of the middle European and Northern American population) (20), this average abundance is raised to 2 % for *Methanobrevibacter* genus in general (21).

Notably, *Methanobrevibacter* species are highly correlated with certain bacterial representatives of the gut microbiome, e.g. Christensenellaceae, and have been shown to be syntrophically interactive in co-culture experiments with e.g. *Christensenella minuta* or *Bacteroides thetaiotaomicron* (22,23).

Even though (methanogenic) archaea have been shown to be part of the human gut microbiome for more than 40 years (24), knowledge about their function and role within the human gut microbiome is still sparse. The reason for this lies in several

methodological issues as standard protocols are mainly bacteria-optimized, which are not optimally working for archaea due to their distinct cell structure, physiology, and metabolic activity. These distinct features and inappropriate protocols make it difficult to visualize, culture or analyze them with multi-omics approaches (25).

Furthermore, the lack of comprehensive reference databases poses a challenge in adequately evaluating methanogenic archaea in the datasets. Additionally, the presence of a strong bacterial or host background further complicates the matter (25). Therefore, it is of utmost importance to develop new protocols or modify existing ones to effectively study archaea in the complex microbiome environment.

Even though molecular biological techniques help to study the archaeome, isolating and culturing archaea is crucial for in-depth, mechanistic analysis of these microbes. Based on the Global Catalogue of Microorganisms (<https://gcm.wdcm.org/>) only few human host-associated archaeal isolates are available in culture (three *Methanobrevibacter oralis* strains from human oral cavity and subgingival plaque; and four *Methanobrevibacter smithii* isolated from human feces and large intestine), hindering detailed studies.

To tackle the lack of archaeal cultures for functional studies, we established herein a culturing procedure to efficiently enrich and isolate methanogenic archaea from human stool samples in a targeted manner. By providing optimal growth conditions (e.g. by providing a suitable bacterial partner), followed by archaea-targeting Fluorescence-Activated Cell Sorting (FACS), dilution series and quality control through Nanopore sequencing, we were able to establish stable methanoarchaeal cultures—from 14 out of 16 fecal samples obtained from both healthy and diseased subjects.

By this procedure, we obtained 9 novel methanoarchaeal strains, and were able to retrieve personalized representatives of *Methanobrevibacter smithii* and *Cand. Methanobrevibacter intestini* strains from stool samples of various participants with different health conditions, which were subsequently compared on genomic and functional levels.

As the role of archaea in health and disease is still unknown and it is still discussed whether they might have a potential beneficial or a pathogenic effect on human health, obtaining additional isolates from human samples will allow in-depth analysis, which might help to deepen our understanding of the role of methanogens in the human GIT.

Material and Methods

Ethics and participant recruitment

Sixteen participants with different health conditions (healthy or with any kind of gastrointestinal disorder) were recruited based on their baseline breath CH₄ levels (> 5 ppm / high CH₄ emitters). The recruitment was supported by medical doctors at the State Hospital Graz (Austria) and an independent physician also located in Graz, Austria. Participants were asked to undergo a breath test, provide a stool sample, and complete a questionnaire with health status information ([Supplementary Table S1](#)). Further, all participants signed an informed consent before participation. A list of exclusion and inclusion criteria can be found in [Supplementary Table S2](#). The study was approved by the local ethics committee at the Medical University of Graz, Austria (EK-Nr: 31-452 ex 18/19).

Sample collection, processing and storage

For the pre-selection of participants, a breath test was performed as previously described (21). For this purpose, participants performed the breath test either directly on a GastroCH₄ECK™ Gastrolyzer (Bedfont Scientific Ltd., UK) or with a GastroCH₄ECK™ breath bag (Bedfont Scientific Ltd., UK), which was used for a later measurement on the Gastrolyzer device. Measurements were performed according to the manufacturer's instructions. Participants, who exhaled more than 5 ppm CH₄, were considered as high CH₄ emitters based on (21) and were asked to provide a stool sample and personal data including their health status. For anaerobic cultivation the stool samples were collected with eSwabs™ (Copan, USA) and from the same source for metagenomics with a sterile stool sample container (Sarstedt, AUT). The stool sample container was pre-filled with 5 ml DNA/RNA free ethanol to stabilize the samples, as described earlier (26).

The participants were asked to return the samples within 24 hours after collection and store them at 4 °C in the meantime. After delivery, the samples for anaerobic cultivation were immediately transferred into sterile anoxic (N₂) serum bottles (Glasgerätebau Ochs Laborfachhandel e.K, GER) with sterile syringes (Injekt□-F Solo, B.Braun SE, GER) and the same volume of glycerol containing ascorbic acid (50 % glycerol (v/v), 50 % 1x PBS (v/v), and 1 g/l ascorbic acid) was added. Afterwards,

the samples were mixed by inverting, incubated for one hour at 4 °C and then stored at -80 °C until cultivation. Samples for DNA extraction were transferred in DNA-free 1.5 ml sterile reaction tubes (Eppendorf, GER) and frozen at -80°C until further processing.

Microbiome profiling of the original samples

The DNA of the original stool samples was isolated using the PureLink™ Microbiome DNA Purification Kit according to manufacturer's instructions (Invitrogen, Thermo-Fisher, USA).

Shotgun sequencing was conducted by the company Macrogen using the Illumina HiSeq device. Raw sequencing reads were quality filtered with ATLAS (version 2.16.2) which also included the removal of human reads (27). For taxonomic annotation the following tools were used: Kraken2 (version 2.1.2; with a conf. 0.1) and Bracken (version 2.7; read length 100) (28,29). To complete the hierarchical lineage information the script "kreport2mpa.py" was used that is publicly available at github (<https://github.com/jenniferlu717/KrakenTools/blob/master/kreport2mpa.py>) before merging the feature tables of all samples.

General workflow of enrichment and cultivation

Selection of bacterial syntrophy partners and metabolic modelling

Metabolic modeling was used to test the suitability of bacterial syntrophy partners on MS (30) and MpT1 (31,32) media for enrichment of *Methanobrevibacter smithii* (full recipes are available in (33)).

Draft metabolic models were generated using gapseq (v1.2) (34) on the genomes of the *B. thetaiotamicron*, *C. minuta* and *M. smithii* (GenBank IDs NC_004663.1, GCF_001652705.1 and ABYW000000000.1, respectively). The recipes for MS and MpT1 media were used to build a representation of media constituents in the ModelSEED (35) namespace. The compositions of the undefined media ingredients yeast extract and casein hydrolysate were approximated using the definitions of yeast extract and tryptone by Marinos et al. (36). The initial models were gap-filled with gapseq on each of the media separately, including the respective nutrient additives for

C. minuta and *B. thetaiotamicron*.

For the simulations of the co-cultures, PyCoMo (37) was used to build community metabolic models for each pair of organisms in each medium. The respective medium with bacterial additives were applied to the models. Growth of the communities was calculated at equal abundance of bacteria and *M. smithii*. Exchange metabolites and cross-feeding interactions were calculated from solutions of flux balance analysis with the same parameters.

Cross-feeding interactions profiles (metabolites produced by organism A and consumed by organism B in a given co-culture) were compared by calculating the fraction of the intersection of cross-fed metabolites divided by the union of cross-fed metabolites in two given profiles.

Cultivation

A general workflow of the enrichment process is shown in [Fig. 1](#) and detailed flowcharts for each sample can be found on our github repository (https://github.com/SDMUG/Methanobrevibacter_Enrichment). For the enrichment of methanogenic archaea, stool samples were gently thawed on ice and cultivated using following two approaches: i) First, the samples were inoculated in MS and MpT1 media supplemented with antibiotics (0.1 mg/ml penicillin G potassium salt (Carl Roth GmbH + Co. KG, GER) and 0.1 mg/ml streptomycin sulfate (Carl Roth GmbH + Co. KG, GER) to reduce bacterial growth. ii) For the second approach the samples were additionally inoculated in MS and MpT1 media with already growing bacterial cultures (purchased from Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, GER) either *Bacteroides thetaiotaomicron* (DSM 2079) or *Christensenella minuta* (DSM 22607); which were described to support the growth of the methanogenic archaea (22,23).

Inoculation for this approach was performed as described in (23). For the cultivation, both media were prepared under anaerobic conditions. In case of cultures with *B. thetaiotaomicron*, the following supplements were additionally added before inoculation to MS and MpT1 media, following the instructions of the DSMZ for the growth of *B. thetaiotaomicron* (medium 78 + haemine + Vit K1, <https://www.dsmz.de/collection/catalogue/details/culture/DSM-2079>): 0.2 ml hemin solution (50 mg hemin; Sigma-Aldrich, USA) in 1 ml 1 M NaOH (Carl Roth GmbH + Co. KG, GER), diluted to 100 ml with ddH₂O); 0.2 ml vitamin K1 solution (0.1 ml vitamin

K1 (Alfa Aesar GmbH & Co KG, GER) in 19 ml 95% (v/v) ethanol (Avantor Performance Material Deventer, NLD)) and 1 ml beef extract solution [100 µg/ml] (MP Biomedicals, USA). In case of *C. minuta*, the media were supplemented with 0.4 ml D-glucose solution [500 µg/ml] (Amresco Inc., USA). All supplements were also prepared under anaerobic conditions and pressurized with 1.0 bar N₂.

The cultures were subjected to incubation at a temperature of 37°C using an incubation shaker set at a shaking speed of 80 rpm. The growth of archaea was influenced by the medium and the presence of supporting bacteria. The incubation period ranged from 7 to 14 days.

The cultures from both approaches were screened for methane and F₄₂₀-positive cells. In the case of approach two, screening occurred prior to FACS sorting, as the sorting process was based on F420 autofluorescence of the methanogens.

bioRxiv preprint doi: <https://doi.org/10.1101/2024.04.10.588852>; this version posted April 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

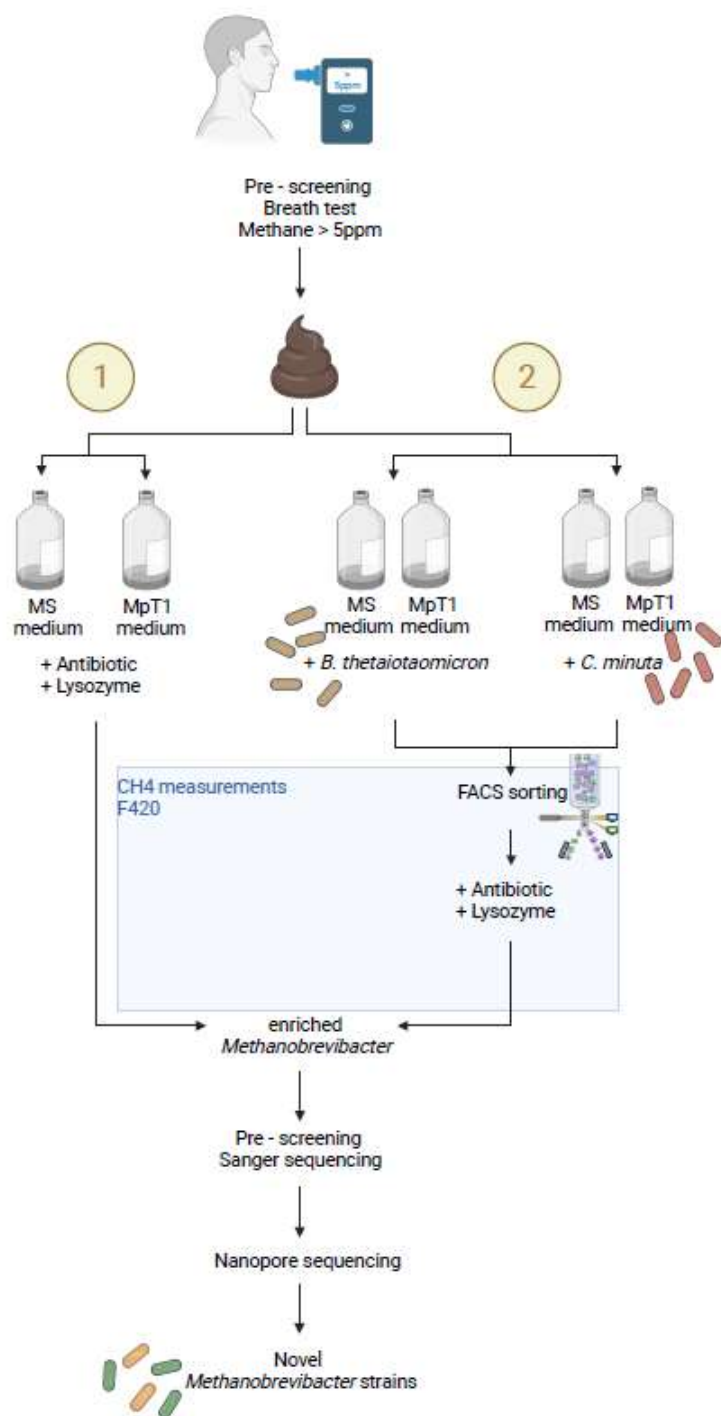


Figure 1: General workflow of the enrichment process.

CH₄ measurement and fluorescence microscopy

To detect the presence of methanogens, the CH₄ sensor BCP-CH₄ (BlueSens gas sensor GmbH, GER) was used to measure CH₄ in the gas phase of the culture bottle according to the manufacturer's manual. Any cultures with CH₄ levels above 2 % were classified as CH₄ positive.

The presence of methanogens was further checked via the typical fluorescence of coenzyme F₄₂₀. For this purpose, fluorescence microscopy (Zeiss microscope Axio Imager A1, Carl Zeiss AG, GER) with fluorescence upgrade was used with Zeiss filter set 05: BP 395-440 (excitation), FT 460 (beam splitter), LP 470 (emission) and 100x magnification (objective Plan-NEOFLUAR).

Fluorescence activated cell sorting (FACS)

To physically enrich the methanogens within the samples, the fluorescence activated cell sorting (FACS) was used on CH₄-positive cultures grown with *B. thetaiotaomicron* or *C. minuta*. Sorting was conducted via the Core Facility Flow Cytometry at the Center of Medical Research in Graz, Austria with FACSAria™ IIu system (BD Biosciences, USA) as described before (38). Prior to sorting, samples were prepared in an anaerobic chamber (Whitley A95 Anaerobic Workstation, Don Whitley Scientific Limited, UK) as follows: 1 ml of culture was centrifuged at 4000 x g for 5 min in a 1.5 ml reaction tube (Eppendorf AG, Hamburg, GER), and the resulting cell pellet was washed twice with 500 µl 1x PBS containing 1 g/L L(+)-ascorbic acid (AppliChem GmbH, GER). The cell pellet was then resuspended in 1 ml of 1x PBS + ascorbic acid and further diluted with an additional 1 ml of 1x PBS + ascorbic acid and transferred to a Falcon® 5 ml Polystyrene Round-Bottom Tube with 35 µm Cell-Strainer Cap (Corning Science Mexico S.A. de C.V., MEX) before sorting to ensure a single cell suspension.

Cells were sorted based on the autofluorescence of coenzyme F₄₂₀ (excitation wavelength at 420 nm) based on (39). F₄₂₀-negative and F₄₂₀-positive cells were distinguished using a 405 nm laser for excitation and a 450/50 band pass filter for detecting fluorescence emission. The presence of autofluorescent cells was observed in CH₄-positive cultures in contrast to its absence in CH₄-negative cultures. In addition, original stool samples, enrichment cultures, and pure cultures of various

methanogenic archaea (*Methanobrevibacter smithii* (DSM 2375), *Methanosphaera stadtmanae* (DSM 3091), and *Methanomassiliicoccus luminyensis* (DSM 25720)) were employed for identifying the sorting gate (Supplementary Figures S1- S4). The sort time was chosen for each culture individually and ranged between 3 and 7 minutes, depending on the archaeal cell count.

Sorted cells (min. 800 events, details for each sample can be found in the flowcharts on our github repository) were transferred into a 1.5 ml microcentrifuge tube containing 250 µl of 1x PBS + ascorbic acid and inoculated into fresh medium. Afterwards the cultures were incubated at 37°C with 80 rpm shaking speed.

Antibiotics, Lysozyme, Dilution series

To reduce bacterial growth and enrich methanogens within the cultures without the addition of *B. thetaiotaomicron* or *C. minuta* and cultures after FACS, antibiotics (penicillin G potassium salt (0.1 mg/ml) and streptomycin sulphate (0.1 mg/ml) were added to the media directly before sample inoculation. For further reduction of antibiotic resistant bacteria, lysozyme (0.5 mg/ml; all substances: Carl Roth GmbH + Co. KG, GER) was added after each culture transfer.

For further purification of the methanogens, dilution series from 10^{-1} - 10^{-13} were performed under anaerobic conditions. Growth was usually observed up to a dilution of 10^{-7} and 10^{-8} . After one week of incubation (37 °C, 80 rpm shaking speed) the CH₄ levels of positive growth cultures were measured and cultures were evaluated further via fluorescence microscopy.

Sequencing and sequence data analysis

Two sequencing approaches were followed to assess the composition of the cultures. Sanger sequencing was used to identify the (almost) pure archaeal and bacterial components of enrichment cultures, whereas Nanopore whole genome sequencing was used for a detailed assessment of metagenomic composition and genome reconstructions. Details are given in the Supplementary Information file.

Scanning electron microscopy (SEM)

The cell morphology of the enriched archaea and their remaining co-existing bacteria was visualized using the Zeiss Sigma 500 VP scanning electron microscope (Carl Zeiss AG, GER). Culture samples of 2 ml (with the supernatant not removed) were centrifuged at 4000 x g for 10 minutes, and the resulting pellets were handed over to the Core Facility Ultrastructure Analysis from Medical University of Graz (Austria) for SEM preparation. Briefly, the cells were mounted on coverslips, fixed with 2 % (wt/vol) paraformaldehyde in 0.1 M phosphate buffered saline, pH 7.4 and 2.5% glutaraldehyde in 0.1 M phosphate buffered saline, pH 7.4, and dehydrated stepwise in a graded ethanol series. Samples were post fixed with 1 % osmium tetroxide (Electron microscopy Sciences) for 1 h at room temperature and subsequently dehydrated in graded ethanol series (30-96 % and 100 % (vol/vol) EtOH). Further, HMDS (Merck | Sigma - Aldrich) was applied. Coverslips were placed on stubs covered with a conductive double coated carbon tape (40). SEM micrographs were taken with an electron voltage of 5 keV and a magnification of 3-21 kX, using the secondary electron detector (Zeiss Oberkochen).

Tree calculation, correlations and pangenome analysis

The ANI (average nucleotide identity) matrix of the genomes (all-vs-all) were calculated using the ANI Matrix calculator provided through (<http://enve-omics.ce.gatech.edu/g-matrix/>); the resulting phylogenetic tree was annotated using itol (41).

Correlation heatmaps were generated in R. All scripts are provided in our github repository (https://github.com/SDMUG/Methanobrevibacter_Enrichment). Statistical associations were based on spearman correlations, which are also included in the provided script.

Pangenome analyses was performed using kbase (42) based on RastK annotation, the Compute Pangenome and Pangenome Circle Plot (v1.2.0) function, including all available high-quality (> 90% completeness and < 10% contamination) *M. smithii* and *Cand. M. intestini* genomes (this study and (2)), with genomes of Gut_genome132205 and Gut_genome143185 serving as reference genomes.

Genomic, functional annotation and prediction using DeepFRI

Genome comparison was performed using drep v3.4.2 (43), and genome annotation was carried out using eggno-mapper v2.1.10 (default settings (e-value cut-off 0.001) (44–49). Prodigal was used for gene identification.

To examine the functional differences between *M. smithii* and *Cand. M. intestini* as well as to explore the differences between functions related to a specific health condition, PCoA were performed using a Bray-Curtis distance matrix. The visualization was done using the R package ggplot2 (50). PERMANOVA was performed with n=999 permutations to identify statistical significance in the functional profiles.

For identifying the functions of protein-coding genes in archaeal genomes, protein-coding genes were first predicted using Prodigal (<https://github.com/hyattpd/Prodigal>). Then, gene sequences translated to amino acids were annotated using a metagenomic pipeline (<https://github.com/bioinf-mcb/Metagenomic-DeepFRI/>) with embedded DeepFRI tool (51). The pipeline generated a query contact map using results from mmseqs2 target database (<https://github.com/soedinglab/MMseqs2>) searching for similar protein sequences with known structures. Next, the contact map alignment was performed to use it as input to DeepFRI Graph Convolutional Network (GCN) and annotations without matches to known structures were processed with Convolutional Neural Network (CNN). DeepFRI version 1.1 was used in this study; in version 1.1 we retrained the original DeepFRI architecture using high-quality models from the AlphaFold Database, thus increasing the number of functions the method can predict by up to 4x (52). For medium quality (MQ) and high quality (HQ) annotation, scores higher than or equal to 0.3 and 0.5 were used, respectively. For selected genes, a structure prediction was performed using AlphaFold database (52,53), followed by DeepFRI annotation (https://github.com/SDMUG/Methanobrevibacter_Enrichment). Wordclouds were generated using the pip package (<https://pypi.org/project/wordcloud/>).

Results

Set-up of a cultivation workflow and selection of bacterial syntrophic partners based on *in silico* predictions

For the cultivation of archaea from human fecal samples, we followed a combined strategy of: i) Providing *in silico*-customized media, partially including selected bacterial syntrophic partners to facilitate *Methanobrevibacter* growth, ii) pre-selection of sample donors for CH₄ exhalation and such a higher methanogen load, iii) rigorous and standardized screening of enrichments for CH₄ production and presence of (unique) methanoarchaeal key genes, and iv) FACS and antibiotic treatment for the mechanical and chemical enrichment of methanogens from the bacterial background of syntrophic partners or fecal samples. An overview of this strategy is provided in [Fig. 1](#).

C. minuta and *B. thetaiotaomicron* were selected as potential bacterial syntrophic partners, based on previous reports of successfully established co-cultures with *Methanobrevibacter* species (22,23), and the observed co-occurrence in metagenomic datasets (21). Both species, *C. minuta* and *B. thetaiotaomicron*, were reported to provide hydrogen, acetate and possibly other metabolites to the archaeal partners, with *C. minuta* reported to be more efficient than *B. thetaiotaomicron* (23). Nevertheless, we performed a more detailed prediction of the metabolic capacities of both bacteria to serve as a potential syntrophic partner for *Methanobrevibacter* species.

Metabolic models were generated for *M. smithii*, *C. minuta* and *B. thetaiotaomicron*, based on their genomes. At this early stage, *Cand. M. intestini* was not included in the simulations. Single and co-cultures with MS and Mpt1 media were simulated using flux balance analysis (FBA). Across all simulated co-cultures, the growth rate of *M. smithii* was increased compared to the single cultures on the same medium.

This growth boost is enabled by cross-feeding ([Supplementary Fig. S5](#)). Multiple metabolites could be identified as mediators of these interactions: The bacteria commonly provided H₂, amino acids (alanine, lysine, glutamine, cysteine), nucleobases and nucleosides (uracil, inosine, adenosine, deoxyguanosine, xanthosine), vitamins and coenzymes (menaquinone 7, coenzyme A, NAD), as well

as fructose, glycerol, formate, putrescine and ethanol (all output files are provided in our github repository). Based on the simulations, both *B. thetaiotamicron* and *C. minuta* were predicted to be fitting partners for *M. smithii*, without clear preference for either. As such, both bacteria and both media were selected for the enrichment cultures.

FACS was established based on the autofluorescence of coenzyme F₄₂₀ (39). Stool samples, and also methanogenic reference strains (see Materials and Methods) were used to identify the optimal sorting conditions, which were then applied to enrichment cultures.

Efficient enrichment of novel *Methanobrevibacter* from donors with detectable archaeal microbiome signatures

Fecal samples from all 16 study participants (Supplementary Table S1) were processed for enrichment of methanogenic archaea using suitable media with and without additionally added bacteria (see workflow overview Fig. 1). Overall, 8/8 samples of the healthy cohort and 6/8 of the diseased cohort yielded CH₄-positive cultures (88% of all samples). Some samples yielded several enrichment cultures by different approaches (e.g. P15 and P86).

After initial pre-selection for unique methanoarchaea (according to their *mcrA* gene sequence) of all CH₄-positive cultures, enrichment cultures of 14 participant samples were processed further (Table 1). FACS and visual monitoring for potential purity was performed, and after a minimum of 20 transfers they underwent Nanopore-based whole genome sequencing. An overview of the cultivation workflow and sequencing results (including taxonomic classification through kraken2/bracken) for each individual sample is provided on our github repository.

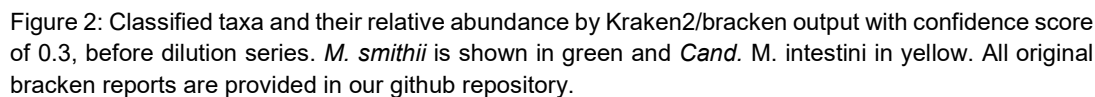
At this stage, 12 of 16 enrichment cultures were dominated by *Methanobrevibacter* representatives. In more detail, eight were predominated by *M. smithii*, and three by *Cand. M. intestini*. Further, two bacteria-dominated enrichment cultures (P51 and P46) contained a substantial amount of *Cand. M. intestini* (Fig. 2).

Notably, the dominance of either methanoarchaeon was evident in the original microbiome profiling. Samples where *Cand. M. intestini* was predominant in the profiling, exhibited similar dominance of this microbe in subsequent cultivation

attempts. Remarkably, both *Methanobrevibacter* species were present in all samples initially. Yet during the enrichment process, one of these archaea emerged as the predominant species, consistent with their original relative abundance (Table 1). However, in the case of sample P86, we achieved successful enrichment of both *Methanobrevibacter* species separately (Fig. 2).

Table 1: Overview of participants, their health status, the derived cultures and the outcome of the metagenomic profiling of the original fecal sample. CD: Crohn's Disease, IBS: Inflammatory Bowel Syndrome, Lactose I: Lactose intolerance, Fructose I: Fructose intolerance, SIBO: Small Intestinal Bacterial Overgrowth. CH₄ analysis indicated the presence of methanogenic archaea in the participant's microbiome. Derived cultures are labeled according to the participant number, the culture medium (MS: MS medium, MpT1: MpT1 medium), and the supplementation with *B. thetaiotaomicron* (B.t.). Additional genomes acquired through dilution series are annotated with +DS. The predominant archaeon as detected through Nanopore sequencing is given, and the outcome of the CH₄ measurement of the cultures 10 days of incubation at 37°C. The metagenomic information of the original fecal sample was analyzed with respect to the relative abundance of Bacteria, Archaea, *M. smithii* and *Cand. M. intestini*.

PARTICIPANTS			CULTIVATION			METAGENOMIC PROFILING			
#	Disorder	breath CH ₄ [ppm]	Cultures	Predominant archaeon	CH ₄ [%]	Bacteria [%]	Archaea [%]	<i>M. smithii</i> [% of archaea]	<i>Cand. M. intestini</i> [% of archaea]
P15	None	30	P15 (MpT1 / B. t.)	<i>M. smithii</i>	16.5	99.76	0.24	88.8	5.1
			P15 (MS / B. t.)	<i>M. smithii</i>	20.7				
P20	None	17	P20 (MS / B. t.)	<i>M. smithii</i>	22.2	99.82	0.18	76.5	3.5
P31	None	70	P31 (MS / B. t.)	<i>M. smithii</i>	23.6	n.a.			
P32	None	44	P32 (MS / B. t.)	<i>M. smithii</i> + DS	23.4	99.87	0.13	90.8	6.3
P45	None	16	P45 (MS / B. t.)	<i>Cand. M. intestini</i>	23.8	99.84	0.16	7.9	79.5
P46	None	21	P46 (MpT1 / B. t.)	<i>Cand. M. intestini</i>	24.2	99.49	0.51	8.6	89.9
P51	None	28	P51 (MS / B. t.)	<i>Cand. M. intestini</i>	5.5	99.96	0.04	6.5	83.1
P82	None	22	P82 (MpT1)	<i>Cand. M. intestini</i>	25.2	n.a.			
P34	CD, flatulence, abdominal pain	22	n.a.	n.a.	n.a.	n.a.			
P48	IBS	26	P48 (MpT1 / B. t.)	<i>M. smithii</i>	22.2	99.66	0.34	94.0	5.3
P61	IBS, Lactose I, SIBO	21	P61 (MpT1)	<i>M. smithii</i>	23	n.a.			
P66	Lactose I	55	P66 (MS / B. t.)	<i>M. smithii</i> <i>Cand. M. intestini</i>	21.8	99.16	0.84	94.6	4.8
P85	CD	6	n.a.	n.a.	n.a.	99.99	0.01	0	0
P86	Lactose intolerance	13	P86 (MS / B. t.)	<i>Cand. M. intestini</i> + DS	22.4	99.99	0.01	63.9	3.8
			P86 (MS)	<i>M. smithii</i> + DS	21.1				
P88	IBS, lactose I, fructose I	8	P88 (MS)	<i>M. smithii</i>	22.3	99.87	0.13	72.4	3.8
P89	IBS, lactose I, pancreatic inflam.	10	P89 (MS / B. t.)	<i>Cand. M. intestini</i>	21.2	99.78	0.22	5.6	88.6



Noteworthy is the observation that samples P34 and P85 failed to yield archaeal cultures, despite positive methane breath test results of the donors ([Table 1](#)). While microbiome profiles were accessible for P85 (P34 had to be excluded due to technical reasons), it was evident that this individual lacked archaeal signatures, suggesting an alternative source of methane production, such as undetected methanogens in the oral cavity or small intestine. It shall be noted that these two individuals were the only ones with Crohn's disease included in this study. It is indeed known that Crohn's disease patients have a reduced amount of archaea in their GIT (54).

To further purify the obtained enrichments and obtain monocultures of the methanogens, subsequent dilution series were and are being performed in combination with consequent antibiotics and lysozyme treatment (see Materials and Methods), leading to a substantial improvement of culture purity later on (e.g. P32: from 32.23 % *M. smithii* to 95.05 %; P 86 (MS + B.t.): from 80.48 % *Cand. M. intestini* to 98.4 %; [Supplementary Table S3](#)). Enrichment culture P45 (*Cand. M. intestini*) has meanwhile been obtained as a pure culture.

We further visually examined the enrichments with Scanning Electron (SEM) and fluorescence microscopy. The SEM micrographs revealed the presence of mainly short rods (single or in chains) in each of the enrichment cultures, representing the known morphology of *Methanobrevibacter* species. Some rods appeared very bloated, as shown in [Fig. 3](#) (arrows). All cultures were dense and revealed the presence of numerous, actively dividing cells. In agreement with the SEM examinations, we observed cells with *Methanobrevibacter*-typical morphology displaying the F₄₂₀ autofluorescence in all cultures ([Supplementary Figure S6](#)).

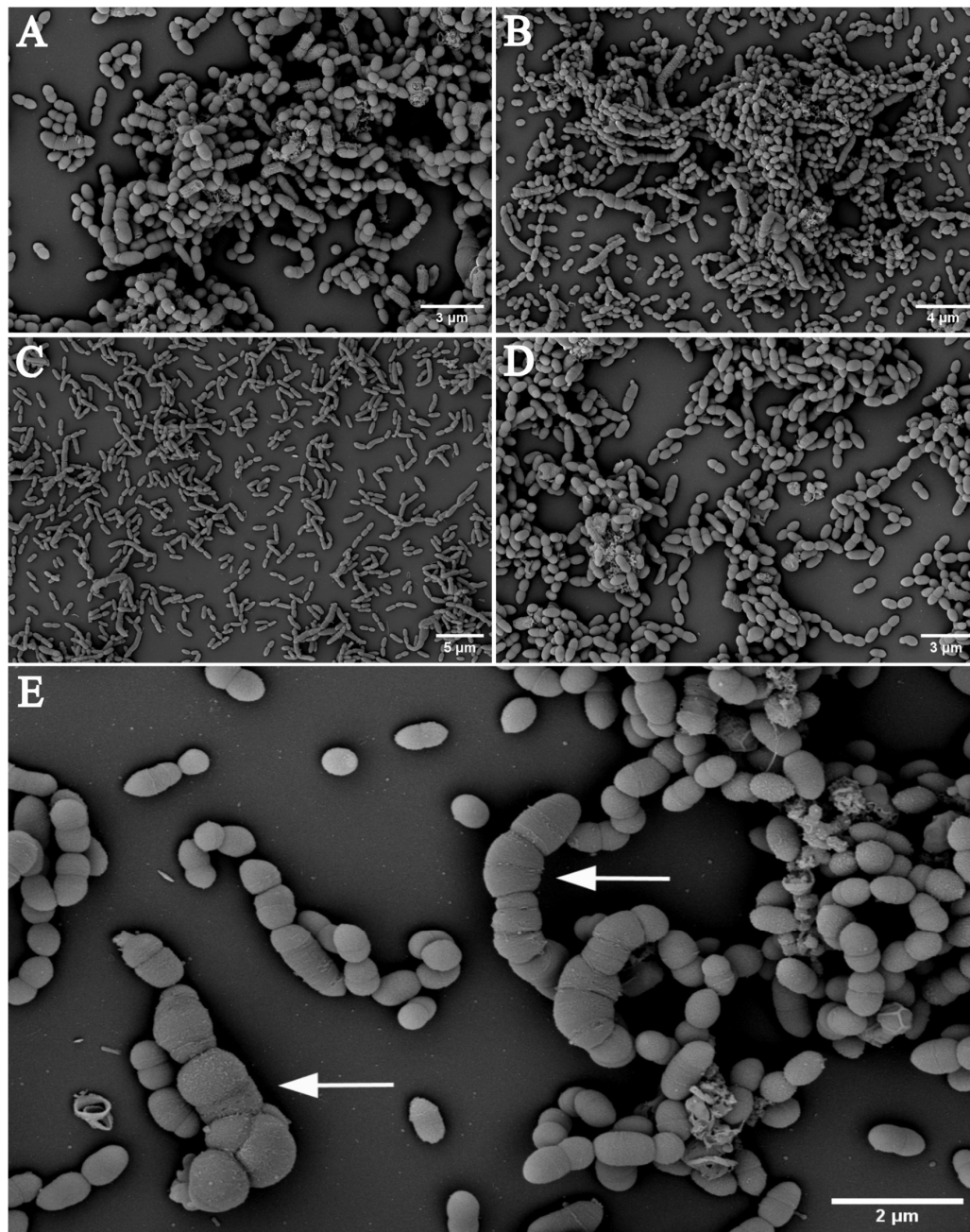


Figure 3: SEM micrographs visualizing the cell morphology of enriched archaeal cultures (*M. smithii*, *Cand. M. intestini*) with co-enriched background bacteria. A: enrichment culture P86, predominated by *M. smithii*. B: enrichment culture P89, predominated by *Cand. M. intestini*. C: enrichment culture P31, predominated by *M. smithii*. D: enrichment culture P51, presence of *Cand. M. intestini*. E: enrichment culture P86 (MS + B.t.), predominated by *Cand. M. intestini*, enlarged for better visibility of different cell phenotypes. Some rods appear bloated (arrows).

Bacteroides thetaiotaomicron, but not *Christensenella minuta*, is a potent syntrophic partner for selective archaeal enrichment

Methane, being uniquely produced by methanogens, serves as an indicator for their presence in cultures. We noted the presence of CH₄ in cultures grown in MS medium with and without supplementation with *B. thetaiotaomicron* after four days of incubation. Similarly, CH₄ was detected in cultures grown in MpT1 medium, with or without supplemental *B. thetaiotaomicron* albeit only after 10 days of incubation.

The addition of *B. thetaiotaomicron* demonstrated a beneficial impact on the efficiency of our enrichment approach. Its presence either had no discernible effect on methanogen enrichment (as evidenced by 12 out of 16 positive enrichment experiments with or without *B. thetaiotaomicron* concurrently) or resulted in an additional strain (P66) not obtained in the abiotic medium (Fig. 2).

However, cultures with additionally added *C. minuta* always remained CH₄ negative (up to 8 weeks of incubation). Also, no autofluorescence of coenzyme F₄₂₀ could be detected for these cultivation attempts. This finding was unexpected, as *C. minuta* had been described as an efficient syntrophic partner of *M. smithii* (23) and we had originally expected positive effects in supporting the growth of these methanogens by providing *C. minuta* in the medium.

Co-enriched bacteria give indications on bacterial partners within the gastrointestinal microbiome

Methanobrevibacter species were enriched with background bacteria from the fecal samples. Interestingly, enrichment cultures with a predominance of *M. smithii* (Supplementary Figure S7) showed only a few significant, weakly positive correlations with specific background bacteria, such as *Phocaeicola dorei* ($q=0.022052$; $r_s=0.3641$) and UMGS75 sp900538885, a so-far uncultivated Bacilli ($q=1.389 \times 10^{-9}$; $r_s=0.4242$; Spearman correlation and Benjamini Hochberg correction).

A number of negative associations of *M. smithii* were observed, including all *Bacteroides* (e.g. *B. cellulosilyticus*, $q=3.2267 \times 10^{-6}$; $r_s=-0.4112$) and *E. coli* species (e.g. *E. coli_D*, $q=7.8638 \times 10^{-8}$; $r_s=-0.4926$). These were, however, significantly

positive with *Cand. M. intestini* (*B. cellulosilyticus*, $q=0.002058$; $r_s=0.4828$; *E. coli_D*, $q=0.002042$; $r_s=0.2172$; [Supplementary Table S4](#); [Supplementary Fig. S7A](#))

Actually, in all found significant associations ([Supplementary Fig. S7B](#)), the *Methanobrevibacter* species showed opposed preference for bacterial partners, indicating a somewhat different niche preference in the gastrointestinal setting. Also, *M. smithii* showed numerous negative correlations, whereas *Cand. M. intestini*, showed mostly positive correlations to co-cultivated bacteria ([Supplementary Fig. S7B](#)).

Notably, *B. thetaiotaomicron*, although provided in the culture flasks to allow enrichment, was no longer predominant at later stages of the cultivation, indicating that the beneficial effect was rather on pre-priming the medium for methanogenic growth, which was also visible via the faster growth of *Methanobrevibacter* (increased CH₄ content after four days of incubation).

Although both methanoarchaeal species (*M. smithii* and *Cand. M. intestini*) co-exist in original gastrointestinal samples (3), their growth in the enrichment cultures was significantly negatively correlated (Spearman correlation, $r_s=-0.6559$, $q=1.967059 \cdot 10^{-7}$), indicating a competitive behavior, at least under the cultivation conditions applied.

Cultured archaea represent nine novel, thus far uncultivated *Methanobrevibacter* strains

From all enrichment cultures, the respective archaeal genomes could be recovered by Nanopore sequencing and genome assembly (see Material and Methods). A summary of the genome classification according to the GTDB database, as well as the checkM2 output representing genome completeness and contamination can be found in [Supplementary Table S5](#). The obtained genomes were placed in the context of available genomes from human-associated *Methanobrevibacter* strains for further analysis (2) (list of genomes provided in [Supplementary Table S6](#); all genome sequences are available through our github repository).

In general, we preserved genomes from nine *M. smithii* and eight *Cand. M. intestini* strains. From the *Cand. M. intestini* clade, we obtained overall four different strain-clusters (cutoff: 99% similarity; Figure 4; MIC-1 to MIC-4), while five clusters were obtained from the *M. smithii* clade (MS-1 to MS-5). All of these clades, except one (MIC-1) represent so far unknown genomes (below 99% identity compared to 1,167

genomes available; ANI values and matrix are given in Supplementary Table S7-S8; (2)). All strains had no cultivated representative thus far, indicating that our enrichment strategy is well suited to expand the cultivable *Methanobrevibacter* diversity.



Figure 4: Maximum Likelihood phylogenetic tree calculated based on ANI, retrieved by the comparison of all cultured *Methanobrevibacter* (red letters) and already available genomes from isolates and MAGs (Supplementary Table S6). The *M. smithii* clade is indicated with a green background color; the *Cand. M. intestini* is shown with yellow background color. Cultures are accompanied by additional information on their source (healthy/ diseased donor, and culture conditions). Clusters representing one strain are indicated by black boxes (Supplementary Table S8).

Differences of *Methanobrevibacter smithii* and *Candidatus* *Methanobrevibacter* *intestini* isolates revealed through sequence- and structure-based annotation

M. smithii and *Cand. M. intestini* show various differences in their genomic inventory (2). Based on the sequence-based annotation obtained through eggNOG profiling, our findings here corroborate this observation. Including only high-quality genomes (> 90% completeness and < 10% contamination), we identified a significantly different functional profile (PERMANOVA, $F_t = 2.2927$, $R^2 = 0.12533$, $\text{Pr}(> F) = 0.002$) among our enriched *Methanobrevibacter* representatives, clearly separating the two species (Supplementary Fig.S8). The analysis included reference genomes of *M. smithii* (Gut_genome132205; DSM2374) and *Cand. M. intestini* (Gut_genome143185; WWM1085; Supplementary Table S6-7).

The most discriminative genes of all obtained genomes (present in $\geq 90\%$ or $\leq 10\%$ of the genomes) are shown in Table 2.

Table 2: Summary of the most discriminative genes found in genomes of *M. smithii* and *Cand. M. intestini*. Table shows counts and percentages of genomes in which the respective function is present. Distinguishing characteristics derived from high-quality genomes are emphasized in bold. P: Inorganic ion transport and metabolism, I: Lipid metabolism, S: Function unknown, K: Transcription, F: Nucleotide transport and metabolism, G: Carbohydrate transport and metabolism, C: Energy production and conversion, H: Coenzyme transport and metabolism, Q: Secondary metabolites biosynthesis, transport and catabolism. N.d.: not done. **: High quality annotation (annotations with score ≥ 0.5); *: medium quality annotation (annotations with score ≥ 0.25)

eggNOG orthologous groups	Prevalence in <i>M. smithii</i> genomes		Prevalence in <i>Cand. M. intestini</i> genomes		COG category	Genomic annotation (EggNOG)	Functional annotation (DeepFri)
	count (total/present)	present in % of genomes	count (total/present)	present in % of genomes			
COG0725	12/12	100	0/10	0	P	ABC-type molybdate transport system, periplasmic component (<i>modA</i>)	ion binding (GO:0043167)**
COG4149	12/12	100	0/10	0	P	ABC-type molybdate transport system, permease component (<i>modB</i>)	regulation of metabolic process (GO:0080090)*/**; catalytic activity (GO:0003824)**; membrane-associated (GO:0016020)**
2EI95	12/12	100	1/10	10	F/G	unknown function	nucleotide-sugar transmembrane transporter activity (GO:0005338) **
COG4925	11/12	91.67	1/10	10	I	sulfurtransferase activity; cyclophilin-like domain-containing protein	transferase activity (GO:0016740)
COG0586	0/12	0	10/10	100	S	Uncharacterized membrane protein DedA, SNARE-associated domain	nd
COG2076	0/12	0	10/10	100	P	Membrane transporters of cations and cationic drugs	nd
arCOG00818	1/12	8.33	10/10	100	S	AbrB family transcriptional regulator	nd
2AVB5	0/12	0	9/10	90	F/G	unknown function	nucleotide-sugar transmembrane transporter activity (GO:0005338)
2C5WB	0/12	0	9/10	90	CHQ	unknown function	oxidoreductase (GO:0016491); oxalyl-CoA decarboxylase activity (GO:0008949)
2DDW5	0/12	0	9/10	90	G	unknown function	glycosyltransferase activity (GO:0016757); oxidoreductase (GO:0016616) involved in the steroid biosynthetic process (GO:0006694)
COG5015	0/12	0	9/10	90	S	Pyridoxamine 5'-phosphate oxidase family	nd
COG1733	0/12	0	9/10	90	K	HxIR-like helix-turn-helix	nd

As typical for archaeal genomes, a high number of proteins cannot be assigned to a function (54% across 1.8 million proteins from 1,167 genomes; (2)), due to missing databases and mechanistic information. For that reason, we performed a structure-based protein function prediction via DeepFRI which is reliant on Alpha-fold (51) for

the most discriminative functions. By doing so, three previously unknown functional genes (2AVB5, 2C5WB and 2DDW5) from *Cand. M. intestini* could be annotated (Table 2). Frequencies of the occurring GO terms in annotations are shown as word clouds in [Supplementary Fig. S9](#).

Most discriminative functions were involved in substrate/ ion transport, such as for molybdate, as all *M. smithii* genomes analyzed in this study carried the *modA* and *modB* subunits for the ABC (molybdate) transporter. *ModA* binds molybdate and tungstate with high affinity, and together with *ModB* transport permease protein is widely distributed across methanogens. Molybdopterin is an important cofactor for the formate dehydrogenase, indicating some difference in the formate metabolism of both species (55). Indeed, differences in the behaviour of formate metabolism in *Cand. M. intestini* was found, as it, in contrast to *M. smithii* strains, accumulated formate in its medium, although the formate dehydrogenase and formate transporters are coded in its genome (33). Further, unique nucleotide-sugar transmembrane transporters for each species were identified through DeepFri annotation.

Another finding was the potential presence of an oxalyl-CoA decarboxylase in the *Cand. M. intestini*; these enzymes are highly important to eliminate oxalates from human diet contained in e.g. coffee, tea and chocolate (56), and would allow for the release of CO₂ and formyl-CoA (<https://www.ebi.ac.uk/QuickGO/>).

Although *Cand. M. intestini* genomes were somewhat larger (1.8 Mbp) compared to *M. smithii* (1.7 Mbp; see also (2)), the overall gene count was similar (~1,900 genes per genome). The percentage of core genes was also very similar (66.39% for *M. smithii*; 66.61% for *Cand. M. intestini*), but variations were observed in the percentage of non-core functions (27.79% and 29.47, respectively) ([Supplementary Figure S10](#)). Notably, the representative *Cand. M. intestini* genome (Gut_genome132205) revealed double the amount (n=36) of genes involved in defense mechanisms than the representative genome of *M. smithii* (Gut_genome132205, DSM2374, n=18). The additional genes included mainly CRISPR-associated proteins or genes involved in Type I restriction modification ([Supplementary Table S9](#)), indicating a potentially more competitive environment for *Cand. M. intestini* ([Supplementary Fig. S11](#))

Functional annotations indicate potential differences of *Methanobrevibacter* species according donor health status

We explored the functional profiles of the archaeal cultures obtained from both healthy and diseased participants. This analysis excluded all reference genomes due to the unknown health status of the source individuals and thus only initial insights can be retrieved.

Overall, *Methanobrevibacter* genomes of healthy and diseased individuals did not differ significantly ([Supplementary Fig. S12](#), eggNOG annotation), but the *Cand. M. intestini* population revealed variances according to the health status of the respective donor. Some potential unique functions for representatives stemming from the diseased (n=4) and healthy cohort (n=4) could also be identified ([Supplementary Table S10](#)). Unique functions found in *Cand. M. intestini* genomes from subjects with gastrointestinal issues were mainly CRISPR-associated proteins (arCOG05124: CRISPR-associated protein TM1802 (cas_TM1802)); COG3649: (CRISPR-associated protein Cas7)). Three unresolved annotations (arCOG07602, 2DP7E, 2DQUR) were further tackled through structure-based function prediction and identified functions involved, amongst others, hydrolase activities ([Supplementary Fig. S13-14](#); [Supplementary Table S10](#)).

In general, the identified functions (DeepFri) for the unique COGs of *Cand. M. intestini* genomes derived from the diseased participants were mainly related to 5'-3' nuclease activity, RNA polymerase activity and DNA polymerase complex, suggesting the alteration in gene expression, with the metabolic functions being mainly organonitrogen compound biosynthesis and macromolecule metabolism (HQ annotations only; [Supplementary Fig. S15](#)).

However, the *Cand. M. intestini* genomes obtained from the healthy cohort revealed a higher number of unique functions, in particular when DeepFRI-based annotation was performed. Annotations exclusive to healthy cohorts included a wider range of functions, e.g. peptidase regulation, lipid metabolism, glycosyltransferase, methyltransferase and nucleotidyltransferase, as well as regulation of signal transduction and cell communication. Other annotations were associated with functions associated with membrane, cytoskeleton and chromatin ([Supplementary Fig. S15](#)).

In summary, genes unique to the *Cand. M. intestini* genomes preliminarily identified in the healthy cohort in this study, encode a broad range of functions (cellular processing, metabolism, information storage and processing), whereas genes unique to the diseased cohort were mainly involved in CRISPR activities, DNA and RNA processing ([Supplementary Fig. S15](#)), warranting a closer look into this matter with additional cultures and genomes.

Discussion

Despite acknowledging archaea in the human gut microbiome for over 40 years (14), our understanding of their functions and roles are still limited. The major bottleneck is the rare availability of cultivable representatives that would allow for deeper, mechanistic, comprehensive analyses, including multi-omics and interaction studies (57).

Development of an efficient cultivation procedure

In this study, we developed a targeted culturing procedure to efficiently enrich and isolate methanogenic archaea from human fecal samples. This approach enabled the stable cultivation of several *M. smithii* and *Cand. M. intestini* strains from participants with varying health conditions, facilitating genomic and functional comparisons. These cultures include nine previously uncultivated strains, with only one revealing overlap with one of more than 1,000 recently recovered metagenome assembled genomes (2).

From all donors, except Crohn's disease patients, stable methanogen cultures could be recovered, which is a proof of the efficient enrichment pipeline. Such high efficiency could even allow personalized archaeome studies based on individual microbiomes.

Growth inhibition by *C. minuta*

(Human) methanogens often rely on syntrophy with bacterial partners that provide small carbon compounds and hydrogen for methanogenesis (58). In particular *Christensenella* and *Bacteroides* species were considered as potent partners, as they generally correlate positively with *Methanobrevibacter* species in natural gut microbiomes, in previous experiments (23) and confirmation through *in silico* modeling (22) (and this study). In our targeted approach, we utilized this positive interaction and added both bacterial species to the culture media in order to induce the low levels of methanogens to grow, as it has been proposed earlier (57). Nevertheless, in contrast to cultures in which *B. thetaiotaomicron* was provided as syntrophic partner, all our enrichment attempts remained methane- and methanogen-negative when *C. minuta* was added, indicating a potential competition under the artificial culture-media conditions, despite their re-current co-existence in natural gastrointestinal microbiomes (3,21). Pre-priming of the media with *B. thetaiotaomicron*, however, was slightly advantageous, as one *Methanobrevibacter* strain could only be cultured with its addition.

Co-enriched bacteria from human GIT microbiome

Besides those known interaction partners, we co-enriched background bacteria from feces, which might give indications on potential, natural bacterial interaction partners

within the GIT microbiome. In contrast to *M. smithii* enrichments, *Cand. M. intestini* was revealing a strong co-occurrence with bacteria, such as *E. coli* and *Bacteroides* species. Both bacterial taxa produce a variety of substances from more or less complex sugar components, such as succinate, formate, acetate, lactate and ethanol (59,60). This indicates a potential dependency of *Cand. M. intestini* on these or other components that are not well provided in the medium, or an inhibition of *M. smithii* by those or other components.

Retrieval of pure cultures

The provision of syntrophic partners, as well as the co-enrichment of bacteria from the original samples results in a minor, but stable background contamination in the enrichment cultures. To obtain final pure cultures, this background contamination has to be removed, by e.g. using optical tweezers, micromanipulators, encapsulation methods or high-throughput single cell isolation (57). In our case, we used a combination of antibiotics/lysozyme treatment and dilution series, which substantially improved the archaeal predominance, and meanwhile led to the first pure isolate of a novel *Cand. M. intestini* strain (P45).

As all enrichment cultures are very stable and maintained for months and >50 transfers in our laboratory now, we also might have to consider that some co-cultures that resist clonal purification, are reliant on the archaeal-bacterial interaction, i.e. by syntrophy or any other mechanisms. Co-cultures reflect natural conditions to a greater extent, and creative solutions are needed to still obtain pure cultures (61,62), which are, for example, a prerequisite for the inclusion of these organisms in a culture collection. However, providing these strains to the scientific community is a declared goal, in order to extend the wealth of available *Methanobrevibacter* isolated from the human gut for future mechanistic studies.

Insights into the genomes and advantages of structure-based predictions

M. smithii and *Cand. M. intestini* stand out as distinct species, characterized by unique microbial networks, physiological roles, and genomic potentials (33,63). Like many other archaea, much of their genetic landscape remains unexplored, as a significant portion of their genes lack annotation, and genetic models are yet to be established. To delve deeper into their characteristics, we employed structure-based predictions to complement sequence-based tools for annotation (64). As previously observed (65), also archaeal COG groups when further annotated using DeepFRI (51), exhibit enhanced accuracy and coverage of annotated functions compared to orthodox orthology-based methods. As such, we were able to retrieve a higher resolution in the distinction of *M. smithii* and *Cand. M. intestini* genomes, and identified some novel functions (such as the putative oxalyl-CoA decarboxylase) that could be relevant for the gastrointestinal ecosystem.

The role of methanogenic archaea for health and disease are still a matter of ongoing studies (summarized in (3), and (18)). In particular, the role of *Cand. M. intestini* remains to be explored, as earlier studies could not resolve both species separately, due to the high sequence similarity in the 16S rRNA gene (33). In our study, the number of participants and the subsequent cultures and genomes was quite low, so that a relevance of either species for disease/health cannot be concluded at this stage. Nevertheless, structure-based analyses of the clusters of orthologous groups of the *Cand. M. intestini* genomes revealed some differences based on the health status of the donor, with a broader variety of functions associated with the healthy phenotype, and a number of functions associated with CRISPR, and DNA/RNA processing in the diseased phenotypes, lacking also a clear indication of their subcellular location, which might indicate a potential foreign origin (e.g. phage/virus). These results should be considered as preliminary evidence, showcasing the potential lying in the cultures and genomes, but also emphasizing the need for further investigation.

Conclusion and outlook

In this study, we have successfully extended the cultivable human archaeome through a targeted cultivation approach that relied on a strategic selection of samples and optimized culture conditions including syntrophic bacterial strains. This approach facilitated the growth of *Methanobrevibacter* species, resulting in the stable cultivation of nine novel strains belonging to the two major archaeal species in the human gastrointestinal ecosystem, namely *M. smithii* and *Cand. M. intestini*. The derived genomes can be used to obtain first insights into the physiology and functions that might be relevant for human health or disease.

Furthermore, our method has proven highly efficient, as evidenced by ongoing work in our laboratory, as we were able to enrich methanogens even from individuals with methane content below 5 ppm in their breath. This highlights the robustness of our approach and its potential for widespread application.

Our work opens the way for comprehensive analyses that promise to deepen our understanding of methanogen dynamics within the human gastrointestinal tract. Expanding the public isolate collection is essential for gaining further insights into the roles of methanogens in the human microbiome.

Data availability

All used and created data is made publically available on our github repository:
https://github.com/SDMUG/Methanobrevibacter_Enrichment

Acknowledgments

We want to acknowledge the support of Petra Jurše Kupper and Astrid Koller in the recruitment process, and we thank all donors that have contributed with their samples. This research was funded in whole or in part by the Austrian Science Fund (FWF) [grants P 32697, P 30796, SFB F-83, COE 7, doc.funds project DOC 69]. For open access purposes, the author has applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission. SD and MP were supported by their local doctoral programs MolMed (Medical University of Graz) and the Doctoral School in Microbiology and Environmental Science (University of Vienna).

Conflict of interests

The authors declare no competing interests.

Authors' contributions

The study was designed by SD and CME. Method establishment was done by SD, MB and AE. Recruitment was carried out by SD and supported by LSC, CH and AMM. Sample collection and sample preparation was performed by SD, SV, LSC, and TZ. The samples were cultivated by SD, SV and MD. FACS was done by SD, SV, MD and VZ. AM, CK, SV and MD performed Nanopore sequencing, and SD, AM, LS and CME did data analysis. SD, SV, MD, VW, and CME created the shown figures and micrographs. DK and DP prepared the samples for SEM. MP and TR performed metabolic modeling, LS and TK performed the DeepFRI analyses. SD and CME wrote the manuscript, and SV, TK, RM, RS, MP, and AM contributed to the writing of the manuscript, which was read and approved by all authors.

List of Supplementary Materials

Supplementary Tables (can be found in the github repository)

Supplementary Table S1: Overview of the 16 recruited participants and their metadata

Supplementary Table S2: Overview of inclusion and exclusion criteria for participant recruitment

Supplementary Table S3: Composition of enrichment cultures after antibiotics and lysozyme treatments

Supplementary Table S4: Spearman correlations of co-cultured archaea and bacteria

Supplementary Table S5: Recovered genomes, their classification according to GTDB, completeness and contamination.

Supplementary Table S6: List of *Methanobrevibacter* genomes

Supplementary Table S7: ANI Matrix

Supplementary Table S8: ANI distances from different clusters to themselves and *Candidatus M. intestini* (CMI, WWM1085) and *M. smithii* (MS, DSM2374).

Supplementary Table S9: Kbase annotation output on CRISPR Functions for *M. smithii* (left table) and *Cand. M. intestini* (right table)

Supplementary Table S10: Genomic and functional annotation of genes divergent in *Methanobrevibacter* genomes from diseased and healthy individuals

Supplementary Information and Figures

Supplementary Methods: Sequencing

Supplementary Figure S1: General gating strategy of FACS sorting.

Supplementary Figure S2: FACS sorting: Example of a (A) CH₄-negative and (B) CH₄-positive culture.

Supplementary Figure S3: FACS Sorting of pure cultures.

Supplementary Figure S4: FACS Sorting of mixed cultures.

Supplementary Figure S5: Comparison of cross-feeding interaction profiles in simulated co-cultures.

Supplementary Figure S6: Light- and corresponding fluorescence micrographs

Supplementary Figure S7A: Clustered heatmap of the relative abundances of archaeal and bacterial taxa in each enrichment culture.

Supplementary Figure S6B: Correlation matrix showing positive (blue dots) and negative correlations with co-enriched bacteria.

Supplementary Figure S7: PCoA, based on the presence/absence matrix obtained through functional annotation

Supplementary Fig. S8. Occurrences of DeepFRI annotations (as Go term annotations) of otherwise unknown functions of *Cand. M. intestini*.

Supplementary Fig. S9: Pangenome circle plot

Supplementary Fig. S10: Inventory of CRISPR-associated genes

Supplementary Fig. S11: Comparison of the genomic inventory (PcoA) of the enriched *Methanobrevibacter* representatives

Supplementary Fig. S12: Occurrences of DeepFRI annotations in diseased cohort

Supplementary Fig. S13: Occurrences of DeepFRI annotations in healthy cohort

Supplementary Fig. S14. Comparison of *Cand. M. intestini* genomes from healthy and diseased cohorts, DeepFRI results.

References

- Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M. W., Henderson, M. L., Riehl, W. J., Murphy-Olson, D., Chan, S. Y., Kamimura, R. T., Kumari, S., Drake, M. M., Brettin, T. S., ... Yu, D. (2018). KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology*, 36(7), 566–569. <https://doi.org/10.1038/nbt.4163>
- Balch, W. E., Fox, G. E., Magrum, L. J., Woese, C. R., & Wolfe, R. S. (1979). Methanogens: reevaluation of a unique biological group. *Microbiological Reviews*, 43(2), 260.
- Bang, C., & Schmitz, R. A. (2015). Archaea associated with human surfaces: not to be underestimated. *FEMS Microbiology Reviews*, 39(5), 631–648. <https://doi.org/10.1093/femsre/fuv010>
- Borrel, G., Brugère, J. F., Gribaldo, S., Schmitz, R. A., & Moissl-Eichinger, C. (2020). The host-associated archaeome. In *Nature Reviews Microbiology* (Vol. 18, Issue 11, pp. 622–636). Nature Research. <https://doi.org/10.1038/s41579-020-0407-y>
- Buchfink, B., Reuter, K., & Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, 18(4), 366–368. <https://doi.org/10.1038/s41592-021-01101-x>
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38(12), 5825–5829. <https://doi.org/10.1093/molbev/msab293>
- Catlett, J. L., Carr, S., Cashman, M., Smith, M. D., Walter, M., Sakka, Z., Kelley, C., Pierobon, M., Cohen, M. B., & Buan, N. R. (2022). Metabolic Synergy between Human Symbionts *Bacteroides* and *Methanobrevibacter*. *Microbiology Spectrum*, 10(3). https://doi.org/10.1128/SPECTRUM.01067-22/SUPPL_FILE/SPECTRUM.01067-22-S002.XLSX
- Chibani, C. M., Mahnert, A., Borrel, G., Almeida, A., Werner, A., Brugère, J.-F., Gribaldo, S., Finn, R. D., Schmitz, R. A., & Moissl-Eichinger, C. (2022). A catalogue of 1,167 genomes from the human gut archaeome. *Nature Microbiology*, 7(1), 48–61. <https://doi.org/10.1038/s41564-021-01020-9>
- Dridi, B., Henry, M., El Khechine, A., Raoult, D., & Drancourt, M. (2009). High prevalence of *Methanobrevibacter smithii* and *Methanosphaera stadtmanae* detected in the human gut using an improved DNA detection protocol. *PLoS One*, 4(9), e7063–e7063.
- Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Computational Biology*, 7(10), e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
- Fricke, W. F., Seedorf, H., Henne, A., Krüer, M., Liesegang, H., Hedderich, R., Gottschalk, G., & Thauer, R. K. (2006). The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H₂ for methane formation and ATP synthesis. *Journal of Bacteriology*, 188(2), 642–658.
- Gaci, N., Borrel, G., Tottey, W., O'Toole, P. W., & Brugère, J.-F. (2014). Archaea and the human gut: New beginning of an old story. *World Journal of Gastroenterology*, 20(43), 16062. <https://doi.org/10.3748/wjg.v20.i43.16062>
- Gasińska, A., & Gajewska, D. (2007). Tea and coffee as the main sources of oxalate in diets of patients with kidney oxalate stones. *Roczniki Panstwowego Zakladu Higieny*, 58(1), 61–67.
- Gligorijević, V., Renfrew, P. D., Kosciolk, T., Leman, J. K., Berenberg, D., Vatanen, T., Chandler, C., Taylor, B. C., Fisk, I. M., Vlamakis, H., Xavier, R. J., Knight, R., Cho, K., &

- Bonneau, R. (2021). Structure-based protein function prediction using graph convolutional networks. *Nature Communications*, 12(1), 3168. <https://doi.org/10.1038/s41467-021-23303-9>
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., & Jensen, L. J. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1), D309–D314.
- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1), 119. <https://doi.org/10.1186/1471-2105-11-119>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kaeberlein, T., Lewis, K., & Epstein, S. S. (2002). Isolating “Uncultivable” Microorganisms in Pure Culture in a Simulated Natural Environment. *Science*, 296(5570), 1127–1129. <https://doi.org/10.1126/science.1070633>
- Kieser, S., Brown, J., Zdobnov, E. M., Trajkovski, M., & McCue, L. A. (2020). ATLAS: A Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. *BMC Bioinformatics*, 21(1), 1–8. <https://doi.org/10.1186/S12859-020-03585-4/FIGURES/2>
- Koehler Leman, J., Szczerbiak, P., Renfrew, P. D., Gligorijevic, V., Berenberg, D., Vatanen, T., Taylor, B. C., Chandler, C., Janssen, S., Pataki, A., Carriero, N., Fisk, I., Xavier, R. J., Knight, R., Bonneau, R., & Kosciolk, T. (2023). Sequence-structure-function relationships in the microbial protein universe. *Nature Communications*, 14(1), 2351. <https://doi.org/10.1038/s41467-023-37896-w>
- Kuehnast, T., Kumpitsch, C., Mohammadzadeh, R., Weichhart, T., Moissl-Eichinger, C., & Heine, H. (2024). Exploring the Human Archaeome: Its Relevance for Health, Disease, and its Complex Interplay with the Human Immune System. *The FEBS Journal*, *accepted*.
- Kumpitsch, C., Fischmeister, F. P. S., Mahner, A., Lackner, S., Wilding, M., Sturm, C., Springer, A., Madl, T., Holasek, S., & Högenauer, C. (2021). Reduced B12 uptake and increased gastrointestinal formate are associated with archaeome-mediated breath methane emission in humans. *Microbiome*, 9(1), 1–18.
- Lambrecht, J., Cichocki, N., Hübschmann, T., Koch, C., Harms, H., & Müller, S. (2017). Flow cytometric quantification, sorting and sequencing of methanogenic archaea based on F420 autofluorescence. *Microbial Cell Factories*, 16(1), 180. <https://doi.org/10.1186/s12934-017-0793-7>
- Lee, S. J., Lee, D.-Y., Kim, T. Y., Kim, B. H., Lee, J., & Lee, S. Y. (2005). Metabolic Engineering of Escherichia coli for Enhanced Production of Succinic Acid, Based on Genome Comparison and In Silico Gene Knockout Simulation. *Applied and Environmental Microbiology*, 71(12), 7880–7887. <https://doi.org/10.1128/AEM.71.12.7880-7887.2005>
- Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293–W296. <https://doi.org/10.1093/nar/gkab301>
- Lu, J., Breitwieser, F. P., Thielen, P., & Salzberg, S. L. (2017). Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science*, 3, e104.

- Mahnert, A., Blohs, M., Pausan, M. R., & Moissl-Eichinger, C. (2018). The human archaeome: methodological pitfalls and knowledge gaps. *Emerging Topics in Life Sciences*, 2.4, 469–482.
- Maranga, M., Szczerbiak, P., Bezshapkin, V., Gligorijevic, V., Chandler, C., Bonneau, R., Xavier, R. J., Vatanen, T., & Kosciolk, T. (2023). Comprehensive Functional Annotation of Metagenomes and Microbial Genomes Using a Deep Learning-Based Method. *MSystems*, 8(2). <https://doi.org/10.1128/msystems.01178-22>
- Marinos, G., Kaleta, C., & Waschina, S. (2020). Defining the nutritional input for genome-scale metabolic models: A roadmap. *PLOS ONE*, 15(8), e0236890. <https://doi.org/10.1371/journal.pone.0236890>
- Mauerhofer, L.-M., Zwirtmayr, S., Pappenreiter, P., Bernacchi, S., Seifert, A. H., Reischl, B., Schmider, T., Taubner, R.-S., Paulik, C., & Simon, K.-M. R. (2021). Hyperthermophilic methanogenic archaea act as high-pressure CH₄ cell factories. *Communications Biology*, 4(1), 1–12.
- May, H. D., Patel, P. S., & Ferry, J. G. (1988). Effect of molybdenum and tungsten on synthesis and composition of formate dehydrogenase in *Methanobacterium formicicum*. *Journal of Bacteriology*, 170(8), 3384–3389. <https://doi.org/10.1128/jb.170.8.3384-3389.1988>
- Miller, T. L., Wolin, M. J., de Macario, E. C., & Macario, A. J. (1982). Isolation of *Methanobrevibacter smithii* from human feces. *Applied and Environmental Microbiology*, 43(1), 227–232.
- Mohammadzadeh, R., Mahnert, A., Duller, S., & Moissl-Eichinger, C. (2022). Archaeal key-residents within the human microbiome: characteristics, interactions and involvement in health and disease. *Current Opinion in Microbiology*, 67, 102146. <https://doi.org/10.1016/J.MIB.2022.102146>
- Mohammadzadeh, R., Mahnert, A., Shinde, T., Kumpitsch, C., Weinberger, V., Schmidt, H., & Moissl-Eichinger, C. (2024). Age-Related Dynamics of Methanogenic Archaea in the Human Gut Microbiome: Implications for Longevity and Health. *BioRxiv*, 2024.02.09.579604. <https://doi.org/doi:https://doi.org/10.1101/2024.02.09.579604>
- Morris, B. E. L., Henneberger, R., Huber, H., & Moissl-Eichinger, C. (2013). Microbial syntrophy: Interaction for the common good. In *FEMS Microbiology Reviews* (Vol. 37, Issue 3, pp. 384–406).
- Neal, M., Thiruppathy, D., & Zengler, K. (2023). Genome-scale metabolic modeling of the human gut bacterium *Bacteroides fragilis* strain 638R. *PLOS Computational Biology*, 19(10), e1011594. <https://doi.org/10.1371/journal.pcbi.1011594>
- Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME Journal*, 11(12), 2864–2868. <https://doi.org/10.1038/ismej.2017.126>
- Paul, K., Nonoh, J. O., Mikulski, L., & Brune, A. (2012). “Methanoplasmatales,” Thermoplasmatales-related archaea in termite guts and other environments, are the seventh order of methanogens. *Applied and Environmental Microbiology*, 78(23), 8245–8253.
- Plastira, I., Bernhart, E., Joshi, L., Koyani, C. N., Strohmaier, H., Reicher, H., Malle, E., & Sattler, W. (2020). MAPK signaling determines lysophosphatidic acid (LPA)-induced inflammation in microglia. *Journal of Neuroinflammation*, 17(1), 127. <https://doi.org/10.1186/s12974-020-01809-1>
- Polag, D., & Keppler, F. (2019). Global methane emissions from the human body: Past, present and future. *Atmospheric Environment*, 214, 116823.
- Predl, M., Mießkes, M., Rattei, T., & Zanghellini, J. (2024). PyCoMo: a python package for community metabolic model creation and analysis. *In Revision*.

- Ruaud, A., Esquivel-Elizondo, S., de la Cuesta-Zuluaga, J., Waters, J. L., Angenent, L. T., Youngblut, N. D., & Ley, R. E. (2020). Syntrophy via Interspecies H₂ Transfer between *Christensenella* and *Methanobrevibacter* Underlies Their Global Cooccurrence in the Human Gut. *MBio*, 11(1). <https://doi.org/10.1128/mBio.03235-19>
- Scanlan, P. D., Shanahan, F., & Marchesi, J. R. (2008). Human methanogen diversity and incidence in healthy and diseased colonic groups using mcrA gene analysis. *BMC Microbiology*, 8(1), 1–8. <https://doi.org/10.1186/1471-2180-8-79>
- Seaver, S. M. D., Liu, F., Zhang, Q., Jeffries, J., Faria, J. P., Edirisinghe, J. N., Mundy, M., Chia, N., Noor, E., Beber, M. E., Best, A. A., DeJongh, M., Kimbrel, J. A., D'haeseleer, P., McCorkle, S. R., Bolton, J. R., Pearson, E., Canon, S., Wood-Charlson, E. M., ... Henry, C. S. (2021). The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Research*, 49(D1), D575–D588. <https://doi.org/10.1093/nar/gkaa746>
- Shreiner, A. B., Kao, J. Y., & Young, V. B. (2015). The gut microbiome in health and in disease. *Current Opinion in Gastroenterology*, 31(1), 69–75. <https://doi.org/10.1097/MOG.000000000000139>
- Steinberger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11), 1026–1028. <https://doi.org/10.1038/nbt.3988>
- Sun, Y., Liu, Y., Pan, J., Wang, F., & Li, M. (2019). Perspectives on Cultivation Strategies of Archaea. *Microbial Ecology*. <https://doi.org/10.1007/s00248-019-01422-7>
- Thauer, R. K., & Shima, S. (2006). Methane and microbes. *Nature*, 440(7086), 878–879. <https://doi.org/10.1038/440878a>
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Židek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., ... Velankar, S. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1), D439–D444. <https://doi.org/10.1093/nar/gkab1061>
- Wang, Z., Zolnik, C. P., Qiu, Y., Usyk, M., Wang, T., Strickler, H. D., Isasi, C. R., Kaplan, R. C., Kurland, I. J., Qi, Q., & Burk, R. D. (2018). Comparison of Fecal Collection Methods for Microbiome and Metabolomics Studies. *Frontiers in Cellular and Infection Microbiology*, 8. <https://doi.org/10.3389/fcimb.2018.00301>
- Weinberger, V., Mohammadzadeh, R., Blohs, M., Kalt, K., Mahnert, A., Moser, S., Cecovini, M., Mertelj, P., Zurabishvili, T., Wolf, J., Shinde, T., Madl, T., Habisch, H., Kolb, D., Pernitsch, D., Hingerl, K., Metcalf, W., & Moissl-Eichinger, C. (2024). Expanding the cultivable human archaeome: *Methanobrevibacter intestini* sp. nov. and strain *Methanobrevibacter smithii*“GRAZ-2” from human feces. *Under Review*.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.
- Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, 20(1), 1–13.
- Zimmermann, J., Kaleta, C., & Waschina, S. (2021). gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biology* 2021 22:1, 22(1), 1–35. <https://doi.org/10.1186/S13059-021-02295-1>
- Zinder, S. H., & Salyers, A. A. (2001). Microbial Ecology—New Directions, New Importance. In *Bergey's Manual® of Systematic Bacteriology* (pp. 101–109). Springer New York. https://doi.org/10.1007/978-0-387-21609-6_12

A.5.4. Supplementary data

Supplementary data are available under:

<https://www.biorxiv.org/content/10.1101/2024.04.10.588852v1.supplementary-material>

https://github.com/SDMUG/Methanobrevibacter_Enrichment