BIOSYSMO

Deliverable D2.1

Initial report on model-driven biosystems design

Deliverable information	
Responsible Partner:	IDENER
Work Package	2
Author(s):	IDENER
Contributing Partner(s):	UBFC/UFC, JSI, ICL
Dissemination Level:	Public
Type:	R-Document, report
Due Date:	31.08.2024
Submission Date:	28.08.2024
Version:	1.0

Funded by the European Union under the GA no **101060211**.
Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or REA. Neither the European Union nor the granting authority can be held responsible for them.

Project Profile

Programme	Horizon Europe
Call	HORIZON-CL6-2021-ZEROPOLLUTION-01
Topic	HORIZON-CL6-2021-ZEROPOLLUTION-01-10
Number	101060211
Acronym	BIOSYSMO
Title	BIOremediation systems exploiting SYnergieS for improved removal of Mixed pOllutants
Start Date	1 September 2022
Duration	48 months

Document History

Version	Date	Author	Remarks
0.1	01.07.2024	IDENER	Original Draft
0.2	16.07.2024	IDENER	Inclusion of the atrazine degradation pathway in Section 2.1.1
0.3	09.07.2024	IDENER	Inclusion of revisions from LEITAT and other contributing partners
0.4	26.08.2024	IDENER	Minor revisions and inclusion of text on flux balance analysis
0.5	26.08.2024	IDENER	Modification of figures to replace the strain numbers mentioned for UBFC
1.0	28.08.2024	IDENER	Final Draft

Executive Summary

Deliverable D2.1 - *Initial report on model-driven biosystems* design is the first deliverable of WP2 - *Computational tools for biosystems design and implementation* and is devoted to address the work initiated in *Task 2.2 - Constraint-based genome-models* since M6. The primary objective behind this task is to develop and subject genomic-scale metabolic models (GEMs) to in-silico simulations of metabolic interactions pertaining to pollutant biotransformation or degradation. By exploiting the GEMs, IDENER supports the experimental partners by (1) guiding metabolic engineering activities, (2) selecting ideal candidates, and (3) identifying the key players or interactions for the proposed biosystems. Within the scope of this task, IDENER has classified the GEM construction method into two approaches: a bottom-up approach, which provides high-quality refined GEMs to assess performance of key strains; and a top-down approach, where semi-automatically generated GEMs from genomic or metagenomic datasets facilitates the study of existing microbial communities.

In this Deliverable, IDENER describes the status, and the results obtained from the initiation of the bottom-up approach for GEM reconstruction. This task was initiated by introducing the lindane (γ-HCH) degradation pathway into an existing GEM of *Pseudomonas putida* KT2440 to represent the mutant strain *P. putida* KT2440_pSEVA-lindane generated by ICL in WP3. Similarly, atrazine degradation pathway was included to further assist *P. putida* engineering strategies for atrazine consumption and degradation. This was further followed by the construction and exploratory analysis of the GEMs generated from six different fungal strains that were screened and sequenced by UBFC/UFC from specific contaminated sites within the BIOSYSMO project. This analysis included the identification of biosynthetic pathways for plant-growth promoting factors, such as indole acetic acid, jasmonic acid, and gibberellins. Other traits relevant to plant-microbe bioremediation systems such as prospective phosphate solubilization mechanisms and metal interaction genes were identified with the help of the PLaBase database as well as BIOSYSMO's integrated database platform.

Table of Contents

List of Figures

List of Tables

Table of Abbreviations

Abbreviation	Definition
GEM	Genomic-scale metabolic models
γ-HCH	γ-hexachlorocyclohexane
iACME	Interspecies Aggregation Combinatorics, Microfluidics, and Entrapment
DyMMM	Dynamic multi-species metabolic modeling
WGS	Whole Genome sequencing
IAA	Indole acetic acid
PGP	Plant growth promoting
MAGs	Metagenome-Assembled Genomes
FBA	Flux Balance Analysis

1 Introduction

1.1 Definition and purpose

Genome-scale metabolic models (GEMs) are described as mathematical representations that capture the metabolic functions of a cell at a systems level (Orth et al., 2010). They allow for sophisticated mathematical analysis of metabolism, typically using a stoichiometric matrix to represent the metabolic network (Orth et al., 2010). This matrix facilitates the computation and prediction of metabolic flux states under various constraints, making GEMs a powerful tool for studying genotype-phenotype relationships and predicting the effects of genetic and environmental perturbations on metabolic phenotypes (Bernstein et al., 2021). In the context of the BIOSYSMO project, these GEMs facilitate the identification of crucial metabolic pathways involved in the transformation or degradation of pollutants by targeted microbes. This allows selection of the most suitable microbes or engineering of new strains for bioremediation. Using GEMs to simulate how these microorganisms interact with various environmental pollutants will expedite the design of effective biodegradation and bioremediation systems by facilitating the selection of the most optimal microbial strains to establish microbial communities or multi-species biosystems.

Currently in the BIOSYSMO project, 14 contaminated sites, three of which were discarded, have been selected for performing a thorough physiochemical analysis of the contaminants in WP1, as well as the isolation and enrichment of individual microbes and communities with biodegradation or bioremediation capacities in WP3. Their full characterization, soil analysis and screening results has been collected along the project course. Based on their performance during screening tests, several candidates were chosen for their *de-novo* or whole-genome sequencing (WGS). In this document, IDENER will cover the construction of their GEMs to make quantitative predictions of their bioremediation capacities at a metabolic level. To develop these models, IDENER has implemented a dual approach: 1) a refined bottom-up methodology for individual strains, and later combining them to study community-level interactions or multi-species models and 2) a top-down approach for constructing GEMs of natural or synthetic microbial communities directly from their genomic or metagenomic data. Exploiting the network-like structure of GEMs, IDENER uses flux balance analysis (FBA) to study the metabolite exchanges and community-level dynamics essential for effective biodegradation.

Additionally, this deliverable will also demonstrate the detection of key proteins and genes that are involved in metal interaction and resistance mechanisms within the GEMs to map out specific biodegradation pathways. This comprehensive analysis aids the associated partners in further optimising the design and selection of candidates for the biosystems for enhanced bioremediation, ensuring that each member of the microbial community contributes optimally to the degradation processes, thereby maximizing the overall efficacy of the bioremediation strategy.

1.1.1 Contribution of the BIOSYSMO database platform

To contribute to one of the core objectives of the project related to the "*Development and application of a computational pipeline to improve the design and construction of synergistic bioremediation systems",* IDENER has made substantial progress creating a multifaceted and integrated database platform referred to as the BIOSYSMOdb by collecting extensive information on the biodegradation and bioremediation pathways of target contaminants and pollutants that are dispersed across several public databases, along with the enzymes and genes associated with them. One of the pivotal applications of this database is to computationally examine candidate genomes that are isolated from the contaminated sites and to detect their mechanism to break down complex pollutants. Amongst the list of relevant

contaminants, genes, and enzyme information that are consolidated within the BIOSYSMOdb, IDENER has also compiled specialised datasets of proteins known for metal resistance mechanisms in bacteria and fungi, as well as plant growth-promoting (PGP) factors. These protein sequences are classified according to their amino acid composition and functionality to generate Hidden Markov Models (HMMs) (Yoon, 2009). To achieve this, similar protein sequences were first clustered to reduce redundancies and improve computational efficiency (Steinegger & Söding, 2017). Each cluster was then aligned using multiple sequence alignment tools[1], following which the HMMer tool[2] was used to generate the HMMs for each aligned cluster to identify the conserved region. These HMMs help predict the presence of specific functional characteristics and associated genes in the genomes studied under the BIOSYSMO project.

In the context of this deliverable, the BIOSYSMOdb also serves as a repository of microbial-assisted bioremediation pathways for the expansion of the constructed GEMs. GEM generation protocols heavily rely on the completeness of the reference database that is associated with the reconstruction tool (Medlock & Papin, 2020; Thiele & Palsson, 2010). Consequently, GEMs may lack key reactions if the database is missing the relevant enzyme, reaction, or pathway. Once a GEM is constructed from any genome of interest, the BIOSYSMOdb can also be utilized to include traits that are missed in the GEM generation protocol and further enrich the standalone models.

1.1.2 Progress towards the application of GEMs

So far, IDENER has taken an existing GEM of *Pseudomonas putida* KT2440[3] and optimized it to guide metabolic engineering activities for ICL in relation to the generation of mutant strains containing lindane and atrazine degradation capacities. This optimisation consisted of introducing membrane constraints to transport reactions to further improve estimation of their growth rates while carrying out FBA analysis in the COBRA Toolbox[4]. These improvements enhanced model accuracy in estimating cellular energetics and growth rates. Concurrently, GEMs were also generated for selected fungal strains after their isolation, characterisation, and screening for remediation competencies. Preliminary analysis of these GEMs has shown biosynthetic pathways for plant growth-promoting (PGP) factors like indole acetic acid, jasmonic acid, and gibberellins, with ongoing studies to measure their extracellular concentrations using FBA. Furthermore, traits for phosphate solubilization were identified by matching genes in the GEMs with those in the PLaBase[5] database, highlighting genes linked to the production of organic acids that enhance phosphate availability to plants. Several metal interaction genes from the BIOSYSMOdb have been localised within the fungal GEMs, pinpointing to crucial metal resistance or interaction mechanisms for potential bioremediation applications.

[1] https://mafft.cbrc.jp/alignment/server/index.html

[2] https://www.ebi.ac.uk/Tools/hmmer/

[3] Taken from Nogales, J., Mueller, J., Gudmundsson, S., Canalejo, F. J., Duque, E., Monk, J., & Palsson, B. O. (2020). High-quality genome-scale metabolic modelling of Pseudomonas putida highlights its broad metabolic capabilities. Environmental microbiology, 22(1), 255-269.

[4] https://opencobra.github.io/cobratoolbox/stable/index.html

[5] https://plabase.cs.uni-tuebingen.de/pb/plabase.php

2 Methodologies

2.1 Bottom-up approach: Generation of standalone models

In this project, IDENER has defined the bottom-up approach for the generation of the first set of refined GEMs for selected strains. For some of the synergistic biosystems, it is required to perform metabolic engineering of certain strains or take a closer look at the biodegradation mechanisms of individual strains. In those cases, it is pivotal to focus on some key elements in the GEMs of these individual strains and to simulate them to a higher degree of accuracy than others. Additionally, this approach also encompasses the initial stage in constructing refined, standalone models that can later be integrated to simulate more complex systems, such as microbial communities or plant-microbe interactions. This subsequent progression from singular entities to collective systems allows for a focused exploration of biological interactions and processes, according to the specific requirements of the biosystem design.

2.1.1 Workflow with an existing model:

P. putida KT2440 is a soil bacterium and is certified to be safe by the Recombinant DNA Advisory Committee (Zhao et al., 2021). ICL employed already established genome-editing methods for *P. putida* based on plasmid transfer to convert it into a microbial chassis for an isoform of lindane, γ-hexachlorocyclohexane (γ-HCH). The genes responsible for γ-HCH degradation have been derived from *S. japonicum*(Zhao et al., 2021). In the assembled pathway, the substrate, that is γ-HCH will be converted into 2,5-Dichlorohydroquinone and then to Hydroquinone (Figure 1). These reactions will use the LinA, LinB, LinC, LinD, LinE, LinF, LinGH and LinJ genes to carry out the subsequent reactions (Figure 1). Ultimately, Hydroquinone will be converted into β-ketoadipate, an intermediate of the tricarboxylic cycle, thereby entering the central metabolic pathway for biomass synthesis.

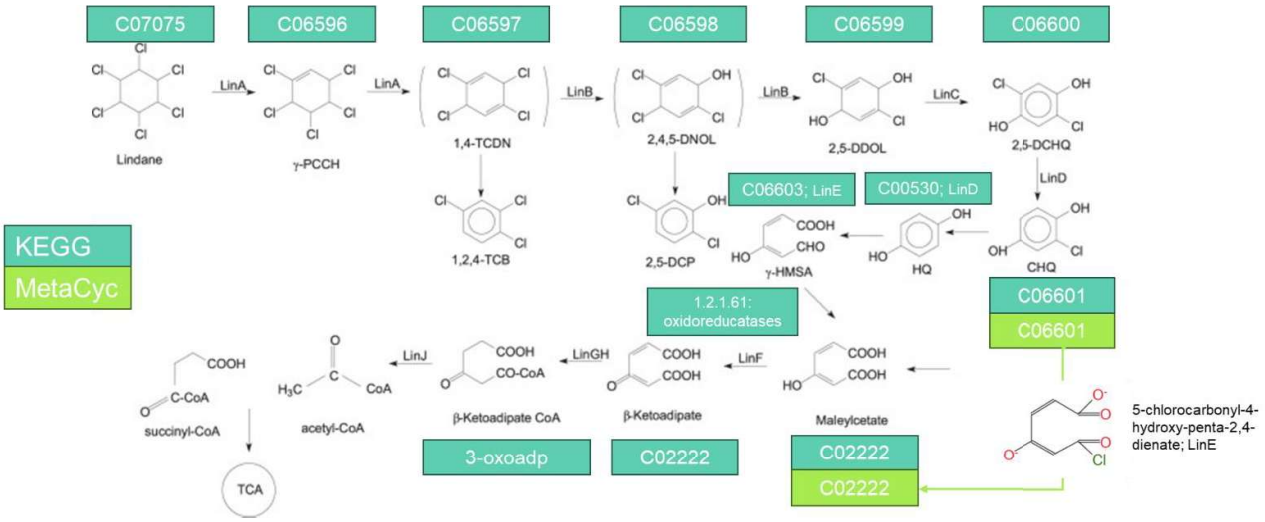


Figure 1: The degradation pathway for lindane, specifically for its γ-HCH isomer as proposed in Zhang *et. al*. (2020) with the corresponding identifiers of each compound from the KEGG and the MetaCyc databases.

For the integration of the pathway into the genomic-scale metabolic model, IDENER selected the most recent genomic-scale model of *P. putida* KT2440 generated by Nogales *et. al.* (2020). Compared to the previous GEM for this strain, this model has been updated with 255 new carbohydrate reactions and

367 transport reactions. A scheme of the γ-HCH obtained from Zhang et. al. (2020) is shown in Figure 1 and was used as the basis of generating the degradation reactions. IDENER converted this pathway into a series of reactions while maintaining the information on the corresponding genes as shown in Figure 2. This information was obtained and verified from two different databases: MetaCyc[6] and KEGG[7]. Most of the reactions constituting the lindane degradation pathway were identical in the two databases, except for the degradation of chlorohydroquinone (CHQ) and formation of maleylacetate (Figure 1). According to the data extracted from KEGG database, this transformation took place through three reactions, resulting in the formation of two intermediates, while in MetaCyc only two spontaneous reactions were required resulting in the formation of a single intermediate (Figure 1).

LIN3	step1	C07075[e] -> C06596[c] + cl[c] + h[c]	linA
LIN4	step2	C06596[c] -> C06597[c] + h[c] + cl[c]	linA
LIN5	step3	C06597[c] + h2o[c] -> C06598[c] + h[c] + cl[c]	linB
LIN6	step4	C06598[c] + h2o[c] -> C06599[c] + cl[c] + h[c]	linB
LIN7	step5	C06599[c] + nad[c] -> C06600[c] + nadh[c] + h[c]	linC
LIN8	step6	C06600[c] + 2 gthrd[c] -> C06601[c] + 2 gthox[c] + cl[c] + h[c]	linD
LIN9	step7	C06601[c] + o2[c] -> CPD10821[c] + 2 h[c]	linE
LIN10	step8	CPD10821[c] + h2o[c] -> C02222[c] + cl[c] + h[c]	spontaneous
LIN11	step9	C02222[c] + nadph[c] + h[c] -> 3oxoadp[c] + nadp[c]	linF
Dead end reactions	Dead end reactions	C06597[c] -> C06594[c] + cl[c] + h[c]	spontaneous
R_C06594_sink	R_C06594_sink	C06594[c] -> C06594[e]	dead end
Dead end reactions	Dead end reactions	C06598[c] -> C06602[c] + cl[c] + h[c]	spontaneous
R_C06602_sink	R_C06602_sink	C06602[c] -> C06602[e]	dead end
EX_C06594	EX_C06594	C06594[e] <=>	metabolite exchange
EX_C06602	EX_C06602	C06602[e] <=>	metabolite exchange

Figure 2: The γ-HCH degradation pathway converted into a series of equations for performing simulations in the existing model of *P. putida* KT2440. For most of the compounds participating in the pathway, their identifiers from the original databases were maintained.

IDENER followed the same protocol for the inclusion of the atrazine degradation into the GEM of *P. putida*. In the assembled pathway, atrazine will be converted into Cyanuric Acid using reactions catalysed by AtzA, AtzB and AtzC (Figure 3). A recently discovered set of Atz genes, AtzEG and AtzH, catalyse the subsequent reactions that convert Cyanuric Acid into 1-carboxybiuret, Urea-1,3-dicarboxylate and, finally, Allophanate (Urea-1-carboxylate) (Esquirol L. et al. (2018). This metabolite will then be taken by AtzF and transform it to CO_2 and ammonia (NH_3/NH_4^+), which enters in the central metabolic pathway leading up to biomass synthesis (Figure 3).

[6] https://metacyc.org/

[7] https://www.genome.jp/kegg/genes.html

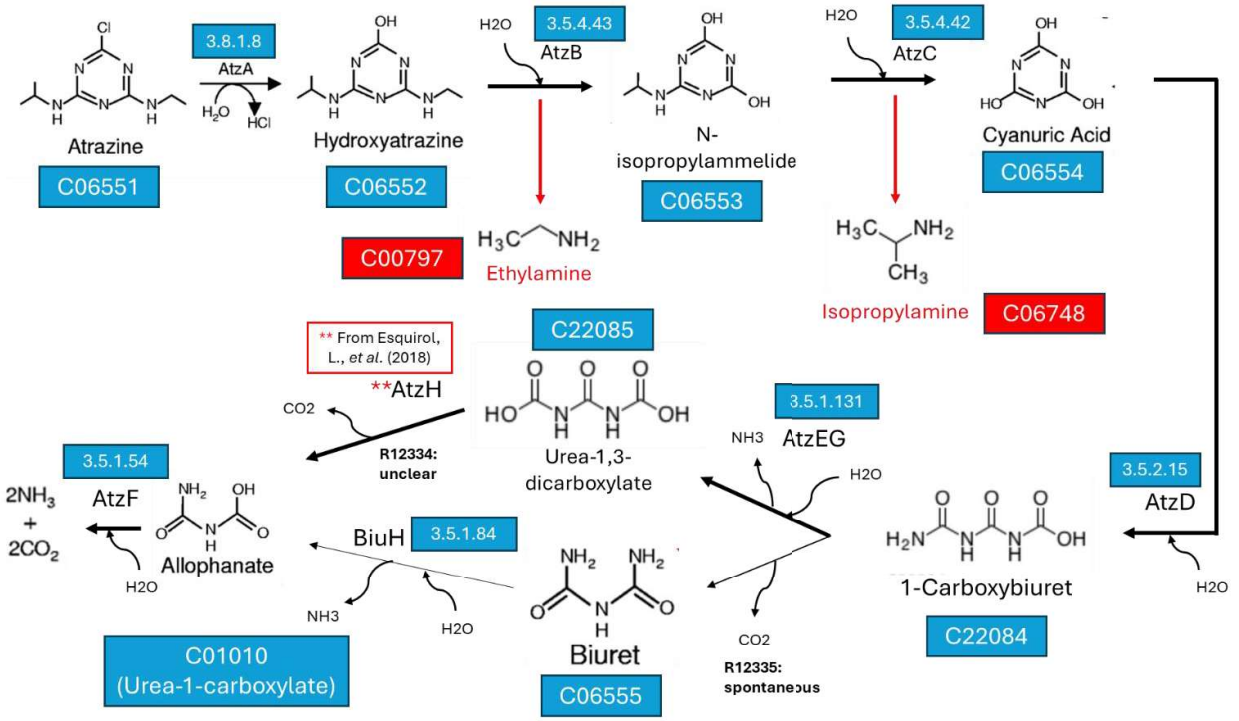


Figure 3: Atrazine degradation pathway, with the corresponding identifiers of each compound from the KEGG database (own adaptation from Liu and Parales, (2009); Wackett et al. (2002); Esquirol et al. (2018)).

Following the protocol as previously explained, each transformation reaction was systematically added into the model, preserving gene-related information as depicted in Figure 4. Like the protocol followed for lindane degradation, these reactions were also validated to those present in the KEGG database. Initially, certain intermediate compounds such as ethylamine and isopropylamine were considered as dead-end metabolites and were incorporated into the model as such (Figure 4) (Esquirol et al., 2018). Such dead ends could be excreted intermediates that should be monitored *in-silico* and *in-vitro* because of their potential toxicity or any interference they may have with the intermediates from the degradation of other pollutants. Therefore, further investigation will be performed to determine if: (1) *P. putida* possesses the genomic capacity for their biotransformation, or (2) the fate of these intermediates once its excreted. Additionally, according to KEGG database entries, allophanate can be generated from 1-carboxybiuret via two distinct pathways: one involves the spontaneous conversion to biuret, followed by its transformation to allophanate catalyzed by the BiuH enzyme; the latter involves the enzyme AtzEG. IDENER selected the reaction catalysed through the ATzEG enzyme, that is also encoded within the genome of *P. putida*.

AtrzHydro	Atrazine hydrolisation	C06551[c] + h2o[c] <=> C06552[c] + h[c] + cl[c]	AtzA
HAtrzHydro	Hydroxyatrazine hydrolisation	C06552[c] + h2o[c] <=> C06553[c] + C00797[c]	AtzB
N-IsopropylamHydro	N-Isopropylammelide hydrolisation	C06553[c] + h2o[c] <=> C06554[c] + C06748[c]	AtzC
CyanAcidHydro	Cyanuric Acid hydrolisation	C06554[c] + h2o[c] <=> C22084[c]	AtzD
CarboxybiuHydro	Carboxybiuret hydrolisation	C22084[c] + h2o[c] <=> C22085[c] + nh4[c]	AtzEG
Urea13decarboxy	Urea-1,3-dicarboxylate decarboxylation	C22085[c] <=> C01010[c] + co2[c]	AtzH
AllophanateHydro	Allophanate hydrolisation	C01010[c] + h2o[c] <=> 10nh4[c] + 2 co2[c]	AtzF
EX_atrazine	Atrazine exchange	C06551[e] <=>	Exchange
Atrz_p	Atrazine exchange	C06551[e] -> C06551[p]	Exchange
Atrz_c	Atrazine exchange	C06551[p] -> C06551[c]	Exchange
EX_ethylamine	exit of ethylamine	C00797[e] <=>	Unknown
R_sink_ethyl	exit of ethylamine	C00797[c] -> C00797[p]	Unknown
R_sink_ethyl_2	exit of ethylamine	C00797[p] -> C00797[e]	Unknown
EX_isopropylamine	exit of isopropylamine	C06748[e] <=>	Unknown
R_sink_isopropyl	exit of isopropylamine	C06748[c] -> C06748[p]	Unknown
R_sink_isopropyl_2	exit of isopropylamine	C06748[p] -> C06748[e]	Unknown

Figure 4: The atrazine degradation pathway converted into a series of equations for performing simulations in the existing model of P. putida KT2440. For most of the compounds participating in the pathway, their identifiers from the original databases were maintained.

Yield estimation through FBA

FBA operates on the idea that metabolic networks in microbial cells maintain a steady-state condition, where the balance of metabolite production and consumption remains stable (Orth et al., 2010). The approach involves representing all reactions and pathways that are encompassed within this metabolic network as a system of linear equations, where each reaction may have an estimated flux through it (Orth et al., 2010). The network is represented as a stoichiometric matrix (S-matrix)[8], where rows correspond to metabolites and columns correspond to reactions. During FBA, the main goal is to optimize this flux through a specific biological function, such as growth or product formation. While running FBA using the Cobra Toolbox on MATLAB[8], the software typically follows a structured process that works with the S-matrix described above. Each reaction in the network has upper and lower bounds that define the possible range of flux values. These bounds can reflect physical, thermodynamic, or experimental constraints. MATLAB's optimization solvers are used to solve the linear programming problem (Heirendt et al., 2019). The objective is to maximize or minimize the chosen objective function, subject to the constraints defined by the S-matrix and the flux bounds.

FBA is used to estimate the theoretical yield of growth or specific products. By setting the objective function to either maximize biomass production or the synthesis of a target metabolite, FBA can identify the maximum potential yield under given conditions (Orth et al., 2010). The result of the optimization provides a flux distribution that shows how metabolites flow through the network under the given conditions. This involves interpreting the fluxes associated with the objective reaction and related pathways. This helps in understanding the limits of metabolic processes and guides efforts to optimize these processes in actual experiments or during metabolic engineering. Constraints based on nutrient availability and environmental factors are applied to ensure the predictions are realistic and achievable (Heirendt et al., 2019).

[8] https://opencobra.github.io/cobratoolbox/stable/tutorials/tutorialFBA.html

Adding Membrane Constraints to Transport Reactions

In Zhuang *et. al*. (2011) it was proposed that there is only a finite surface area available for transport proteins in the cytoplasm. To maximise substrate consumption, microorganisms must regulate the expression of membrane proteins. To understand this better conceptually, let us consider the occupancy of (1) respiratory membrane proteins and (2) substrate transporters. During faster catabolic rates, the increase in the substrate uptake is performed by the substrate transporters, thereby, reducing the activity as well as the occupancy of the respiratory membrane proteins and the respiration rate (Zhuang, Vemuri, et al., 2011). This additional substrate is then further catalysed by the fermentation pathways that usually not use any transmembrane protein. Moreover, since the occupancy of the enzyme on the membrane is inversely proportional to its efficiency and turnover, this influences the overall respiratory stoichiometry. To capture this change in the stoichiometry of our genomic scale metabolic models of *P. putida* as well as other models that will be generated in the future, IDENER introduced a cytoplasmic membrane constraint, or stated as the variable "CMC" within the genome-scale metabolic model. The incorporation of this variable into the different reactions has been shown in Figure 5 (left). It will reflect the cost paid by the transmembrane proteins to achieve a certain reaction turnover based in the enzyme efficiency and membrane occupancy. As illustrated in Figure 5 (right), the highly efficient ATP synthase enzyme will have the highest occupancy cost with a value of 0.1 CMC while the relatively less efficient quinones and ubiquinone will have a lower stoichiometric cost with a value of 0.025 CMC.

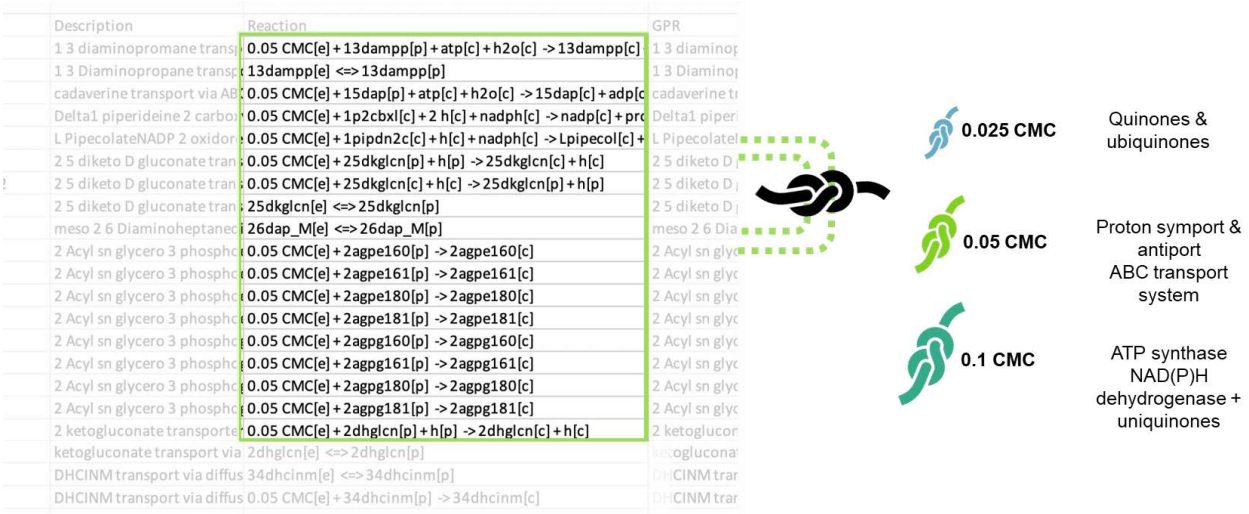


Figure 5: (Left) Adding an additional "CMC" variable specifically to transport reactions in the model. (Right) Assigning coefficients to the membrane constraint based on their membrane occupancy and enzyme efficiency.

2.1.2 Workflow for resequenced and *de-novo* genome sequences

In WP1, UBFC/UFC carried sampling campaigns for comprehensive data collection and analysis in one of targeted sites of the project that was polluted with a high concentration of metals. Following the initial assessment for metal tolerance and Plant Growth Promotion (PGP) traits, 80 isolates from the first sampling campaign were narrowed down to 6 fungal strains. These selected strains were then subjected to WGS (resequencing) for molecular analysis. As some genomes showed poor homologies with reference genomes for four of these strains, these 4 genomes were submitted to *de-novo* sequencing, as detailed in deliverable D1.2 (confidential). The genomic data (resequencing data for 2

strains and de novo sequencing data for 4 strains) obtained were subsequently utilized as inputs for the GEM construction protocol.

In constructing the GEMs for these six fungal species, the methodology adhered to the RAVEN protocol executed within MATLAB (Wang et al., 2018). This protocol utilizes the genomic sequences of each species as a starting point. RAVEN automates the initial reconstruction of metabolic networks by leveraging gene annotations to predict enzymatic functions and corresponding metabolic reactions. This automated pipeline efficiently maps out metabolic pathways directly from genomic data, streamlining the process of model generation. This entirely automated process requires from the user only a FASTA file of the protein sequences. It is important to note here that when building GEMs using the RAVEN toolbox, the genome sequence data should be complete, accurately assembled, well-annotated, and free of sequencing errors or gaps, providing a reliable foundation for metabolic network reconstructions.

There are several approaches to construct GEMs with RAVEN; for the *de-novo* sequences RAVEN, homologous protein sequences were used to build metabolic models without the need for template models. This protocol utilizes protein sequence data and metabolic reactions from the KEGG and the MetaCyc databases. A crucial component of this method is the use of KEGG Orthology (KO) IDs, which are manually annotated gene sets that contain similar metabolic functions and are subsequently associated with numerous metabolic reactions. The process begins by downloading relevant sections of the KEGG database to a local directory, parsing this data to generate a comprehensive GEM. This overarching model includes 7,029 metabolites, 8,398 reactions, and 843,369 genes, as per the latest KEGG release. For the two strains that highly matched to their reference sequences, a template model of the reference genome is selected from the KEGG database, connecting the proteins in the query sequences to its corresponding proteins in the reference model, and therefore, collecting the genes and reactions.

2.2 Future strategies for the top-down approach

In this project, IDENER has also defined a top-down approach as outlined creation of the initial set of GEMs for microbial community models. This strategy starts with a broad analysis of the genomic data from complex systems such as microbial communities or plant-microbe interactions. By examining these systems through the structure of a GEM, overarching patterns and key interactions that exist within the communities can be identified. These community models are then detailed by progressively incorporating genomic datasets from the individual microbes. This systematic progression of taking genomic data from complex systems and studying encompassed metabolic interactions between singular entities enables a comprehensive and targeted exploration of pollutant transformation or resistance processes, that can be manipulated to meet the unique demands of biosystem design.

2.2.1 Workflow for a community-based GEMs

JSI has developed the iACME approach to derive microbial consortia capable of utilising and tolerating organic and inorganic pollutants. All the aggregated consortia went through several screening steps in WP1, and their performance was evaluated based on their estimated growth and respiration parameters. This information is elaborated in Deliverable 1.2 (confidential). Currently, 9 consortia have been selected to undergo metagenomic analysis and WGS to determine dependencies and interactions between strains and potentially coupled metabolic pathways of pollutant transformation such as PAHs.

For the next reporting period, IDENER will collect metagenomic sequencing data from the consortia of interest based on partner feedback. This sequencing data should be processed with appropriate tools

to for screen for sequence composition, coverage, and read linkage to accurately bin contigs into metagenome-assembled genomes (MAGs) (Lapidus & Korobeynikov, 2021). The resulting MAGs can then be further studied to understand the diversity, taxonomy, and functional potential of the individual microbial genomes present in the original metagenomic sample. This step is pivotal for converting the sequences into an appropriate input sequence for the GEM reconstruction protocol. Once the genomes of the constitutive members are assembled, the rest of the protocol that will be followed will be identical to Section 2.1.2.

The combination of the constitutive GEMs from the metagenomic dataset can be simulated through FBA to gather knowledge on metabolite levels, nutrient exchange between community members and other community-level properties (Figure 6). This will be achieved by using the DyMMM tool also available through MATLAB (Zhuang, Izallalen, et al., 2011). This tool enables the grouping of individual metabolic models of the microbes into a compartmentalized community-scale metabolic model that simulates the metabolic interactions and nutrient exchange between community members over time (Figure 6Figure 6). This tool's capacity to predict changes in the community's composition is beneficial for determining its stability. Moreover, parameters such as, the interactions between microbes, including substrate exchange, product secretion, and growth rates can be fixed to adapt the simulations to the experimental conditions. At later stages the model findings can be compared to experimental data on growth and metabolic activities that will be generated by JSI for these extracted microbial communities to capture growth dynamics in the presence of the different pollutants and various selective pressures. This will help us iteratively refine model parameters to improve its predictive accuracy and help in the identification of key members/interactions in the communities for enhancing the biodegradation efficiency of the entire consortium as the project progresses.

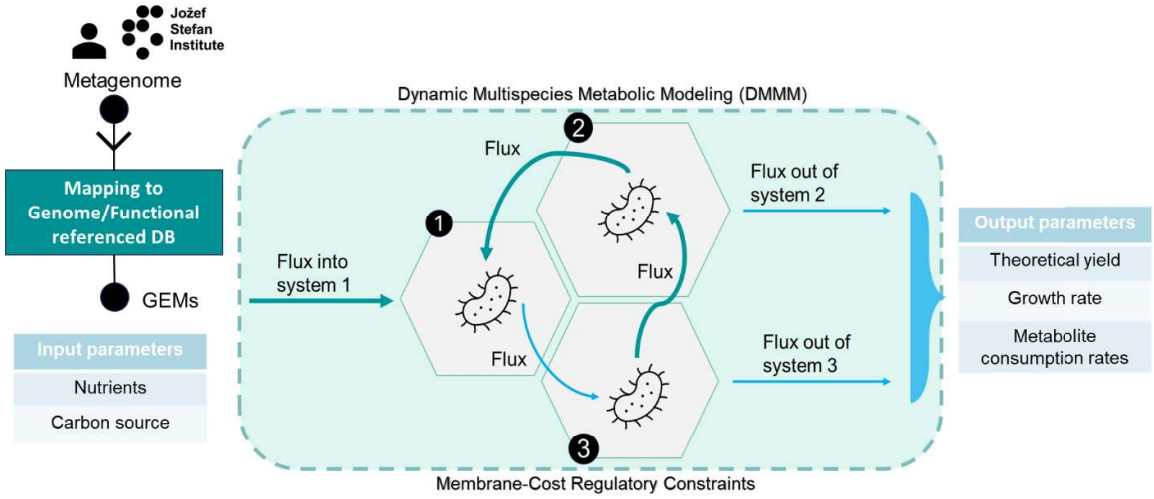


Figure 6. Utilisation of metagenomic datasets obtained from the partners which represent different microbial communities, and their development in community-level GEMs for generating quantitative data on substrate exchange, theoretical yields for secondary metabolites and growth parameters.

3 Exploration of GEM models in the bottom-up approach

This section will discuss the generation, exploration and refinement of individual models that have been constructed using the bottom-up approach. In the case of *P. putida*, the GEMs are intended for assisting the genetic modification and optimisation of the two mutant strains through *in-silico* predictions of growth parameters in presence of the lindane and atrazine degradation pathways. In the case of the six fungal

strains isolated by UBFC/UFC, the main objective is the exploratory analysis of these strains to support the selection of the most ideal candidates for the soil microcosm experiments planned by UBU in WP4 dedicated to testing/generating eligible candidates for the biosystems. In both the cases, the feedback from the experimental data, will lay the foundation for new validation strategies of the generated GEMs.

3.1 Growth rate predictions with *P. putida* KT2660

Growth on lindane

After adding the pathways of lindane degradation into the model, IDENER ran the first round of simulations by allowing lindane to act as the sole carbon source, and the synthesis of biomass as the objective function. The complete conversion of the lindane, into its subsequent intermediates, followed by its integration into the tricarboxylic acid (TCA) cycle through the formation of acetyl-CoA was observed (Figure 1). The flux through the different reactions leading up to the formation of succinyl-CoA and acetyl-CoA was 0.5 mmol gDW^{-1} h^{-1}. From that step onwards, the flux varied for all the reactions participating in the TCA cycle resulting in biomass production at a rate of 0.9 h^{-1}. Given the range of expected values from the literature and the pathway observed for lindane, this value was regarded as an unusually high growth rate for a microorganism growing with lindane as only carbon source. A similar issue was encountered when the simulations were performed with atrazine. The atrazine transformation reactions were effectively integrated into the model; although the reaction fluxes leading to the formation of ammonium were remarkably low, at 0.004 mmol gDW^{-1} h^{-1}, with a growth rate of 0.94 h^{-1}. This initiated a model refinement strategy through the introduction of another layer of constraints to optimise the growth rate and model performance.

After adding Membrane Constraints to Transport Reactions

After adding the CMC, IDENER observed a substantial reduction in the overall growth rate as well as the substrate uptake rate of the microorganism with lindane as the carbon source (Table 1). Similar observations were made for the GEM of *P. putida* with the atrazine degradation pathway, however, contrary to the reduction in the consumption of substrate, an effective integration of the pathway into the central metabolism of biomass synthesis was demonstrated through a higher uptake of atrazine (Table 1).

Table 1: Growth rate estimation for the two GEMs *P. putida* KT2440 with and without membrane constraints for lindane and atrazine degradation pathways.

		Without membrane constraint	With membrane constraint
Lindane	Growth rate [h^{-1}]	1.1	0.042
	Substrate uptake rate [mmol gDW^{-1} h^{-1}]	3.5	0.5
Atrazine	Growth rate [h^{-1}]	0.95	0.071
	Substrate uptake rate [mmol gDW^{-1} h^{-1}]	0.004	0.24

3.2 In-depth analysis of fungal genomes

By using the two whole resequenced genomes sequenced and the four *de novo* sequenced genomes from UBFC/UFC as an input for the RAVEN toolbox on MATLAB, IDENER successfully generated six GEMs. As stated in Section 2.2, the collection of the different reactions for each GEM was done based on the protein homology between the query sequences and the proteins sequences present within the MetaCyc and KEGG databases. The contribution of each database is depicted in Figure 7 for the four *de-novo* sequences. Each network in the Figure 7 represents a single GEM consisting of reactions that are made up of metabolites. IDENER utilises these networks to identify three key attributes of the fungal model that can contribute to the selection of the most ideal candidate for the planned experiments in the following work packages with contaminated soils. The attributes that have been explored are:

1) Metal resistance/interaction capacity
2) Plant growth promoting traits
3) Phosphate-solubilisation mechanisms

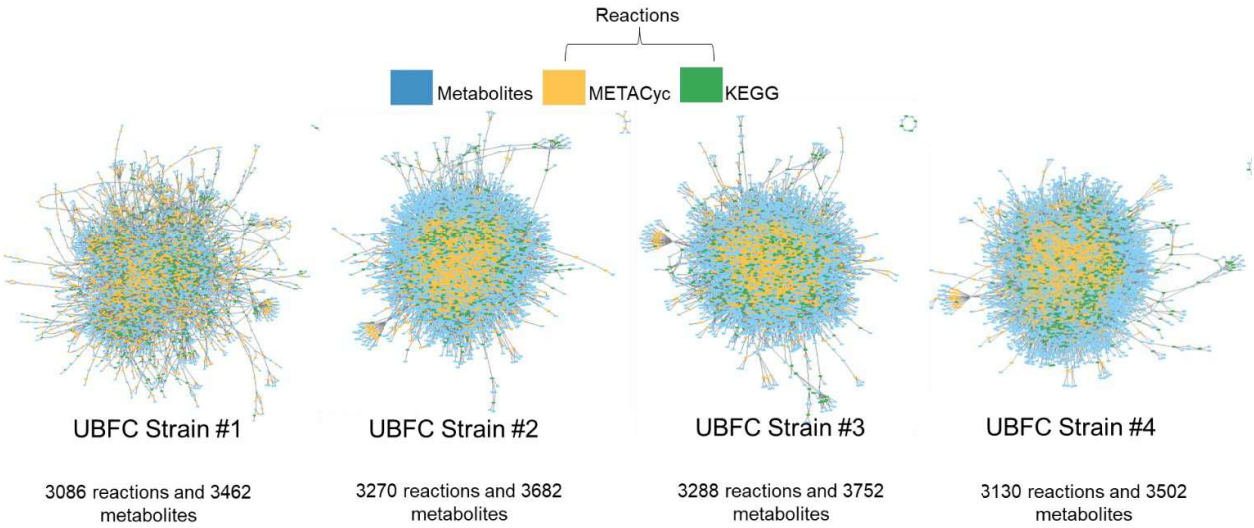


Figure 7: Four different network-like representations of GEMs generated for the *de-novo* genomic sequences. Each GEM is composed of certain number metabolites and reactions depending on its protein composition.

All four fungal models exhibit a high degree of complexity and metabolic diversity, with a significant overlap in reactions and metabolites, indicating a shared core of metabolic functions across different fungal species (Figure 8). An analysis of the subsystems elucidated during the automatic reconstruction of these GEMs revealed that these genes that are in common to all six models are mostly pertaining to biomass synthesis, development of cell wall, and the synthesis of compounds and substances crucial to the generic primary metabolic pathways usually present in a fungal GEM. However, a striking observation that was consistent with the results observed during the screening of these isolates, is that all the six models share the indole-3-acetic acid (IAA) biosynthesis pathway. In the context of plant-microbe interactions, IAA produced by symbiotic microbes can significantly influence plant physiology, promoting root growth and establishment, which enhances nutrient uptake (Syed et al., 2023a). As it aids in stress tolerance and defence mechanisms for plants against metals, biosynthesis of IAA exhibits a vital interplay between microbial communities and plant health, causing immense implications for the planned biosystems within the scope of the BIOSYSMO project.

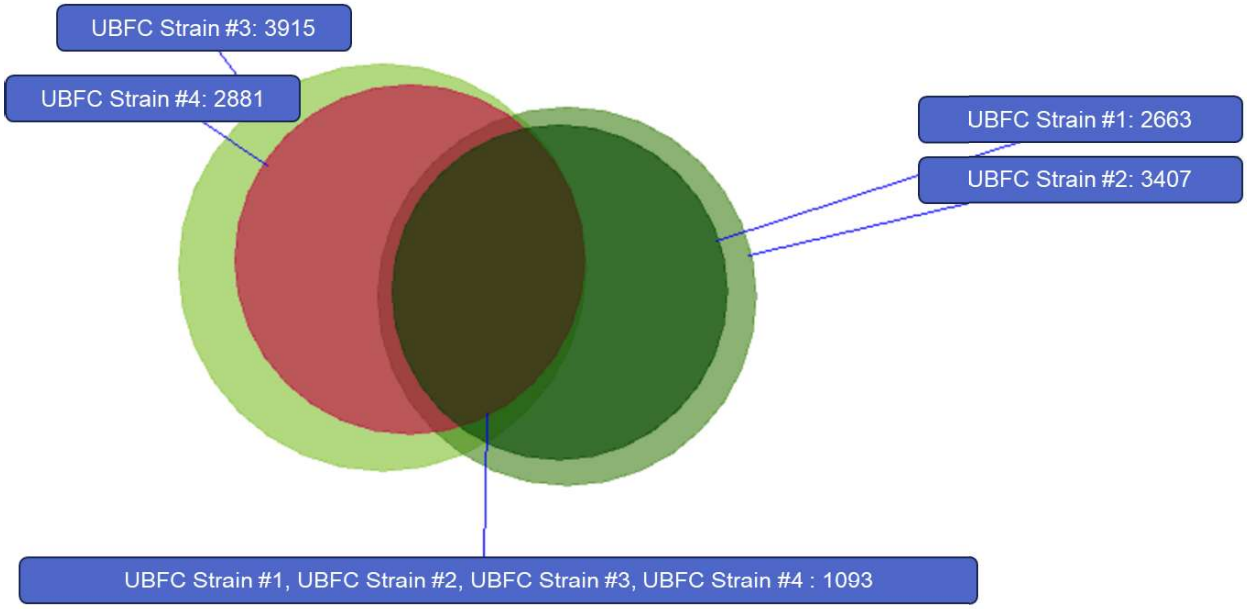


Figure 8: A Euler representation of the all the genes that are present in the generated GEMs of the fungal strains were sequenced de-novo. The area common to all four sets represents approximately 27.92% of the largest set (UBFC Strain #3) and 41.04% of the smallest set (UBFC Strain #1) in the Euler diagram.

Metal resistance

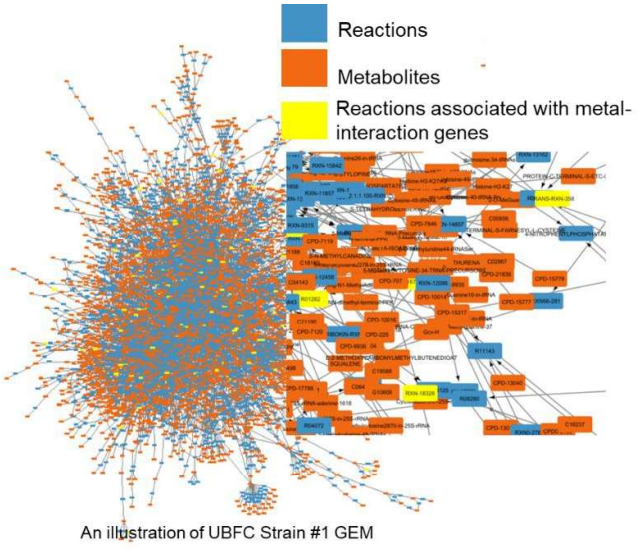
One of the metal contaminated sites within the BIOSYSMO project (Site 10, Vieux Charmont) and exhibits high risks for human health due to the presence of arsenic, lead and zinc in excess (Collot et al., 2023). The microbiome from this highly contaminated soil has been isolated and characterized during WP1 and WP3, respectively. By using the generated GEMs, IDENER evaluated the metal-resistance or interaction capacity of the six fungal strains that were isolated from this site. The ability to tolerate these metals is an important adaptation for fungi in contaminated environments. A high proportion (71.88%) of the fungal isolates from a Hg-contaminated site displayed moderate to very high resistance to mercury(Văcar et al., 2021). Many of the fungal isolates (53.13%) showed resistance to lead, in addition to Hg resistance. Some isolates exhibited very high Pb resistance (Văcar et al., 2021).

Fungi employ several mechanisms to utilize and resist metals for their growth and survival (Bazzi et al., 2020):

1. **Mechanism#1**: Storing excess metal ions, such as zinc, copper, iron, and manganese, in vacuoles or cell walls to reduce toxic concentrations in the cytoplasm.
2. **Mechanism#2**: Pumping out metal ions like zinc, copper, and silver from the cell via membrane transporters, which is a key mechanism for metal resistance.
3. **Mechanism#3**: Binding metals using metallothioneins and other chelating/precipitating compounds to reduce metal toxicity. This is important for resistance to copper, zinc, and cadmium.
4. **Mechanism#4**: Transformation of metals into less toxic forms and absorb them onto their biomass, aiding in bioremediation of metal-contaminated environments.

A GEM obtained from the *de novo* sequencing of any fungal strain with experimentally demonstrated metal interaction or resistance capabilities, can be a useful tool to identify its corresponding reactions and even derive the potential mechanism under different growth conditions. Examples of reactions within such a model typically involve metal ion transporters, such as metallothioneins, and enzymes catalysing redox reactions influenced by metal ions. By simulating these reactions, the model can quantitatively predict the rates of metal uptake, accumulation, and the metabolic adjustments under varying metal concentrations. These simulations can provide insights into the fungal species' resilience or toxicity in metal-rich environments, aiding in bioremediation strategies or in understanding ecological impacts of metal resistance.

Exploiting the BIOSYSMOdb, a cross-reference process was performed with the genes present in the GEMs, identifying several reactions contributing directly and indirectly to metal interaction mechanisms. Localisation of such reactions within the GEMs will aid in the quantification of their activities both during the *in-silico* or *in-vitro* evaluation of these strains. Such estimations of reaction activities can be determined during FBA simulations. Figure 9 shows an illustration of all metal-resistant or metal-interaction reactions (in yellow) that were derived from the BIOSYSMOdb and are present within UBFC Strain #1 GEM. A detailed description of the number of reactions localised within each of the six genomes are provided in the table in Figure 9.



GEM	Reactions Associated
UBFC Strain #1	237
UBFC Strain #2	379
UBFC Strain #3	640
UBFC Strain #4	520
UBFC Strain #5	409
UBFC Strain #6	291

Figure 9: Using the BIOSYSMOdb to identify genes within the GEMs that are known to be associated with metal resistance or metal interaction mechanisms. The table on the right shows the total number of reactions that are associated with metal resistance gene for each of the six fungal strains.

In Table 2, a few examples some genes from the UBFC Strain #1 GEM match with a high similarity score to the metal interaction genes from the BIOSYSMOdb. The reaction pertaining to these genes are also provided in Table 2. For instance, the gene g2516.t1 aligns with the Q8NQC8 protein[9] which is also known as the Arsenical-resistance protein Acr3. Based on the gene-protein-reaction information associated within the GEM of UBFC Strain #1, IDENER detected this strain's capacity to perform a protein-mediated transport of arsenate (Table 2). Further confirmation of the occurrence of such

[9] https://www.uniprot.org/uniprotkb/Q8NQC8/entry

reactions will be made by running simulations in appropriate growth conditions, accompanied by physiological and molecular tests conducted on the strain.

Table 2: Some examples of metal interaction reactions associated with the genes taken from the BIOSYSMOdb that match with those present in the GEM of fungal strain UBFC Strain #1.

Reaction	Gene	Aligned protein	Reactions from the GEM
RXN-21035	G2459.t1	sp|P21441|MDR_LEITA	ATP + H2O + Cd (cytosol) = ADP + phosphate + Cd (vacuole)
RXN-11187	G2516.t1	sp|Q8NQC8|ACR3_CORGL	ARSENATE + MYCOTHIOL -> MYCOTHIOL-ARSENATE CONJUGATE+ WATER
TRANS-RXN0-200	g3855.t1	sp|P75757|ZITB_ECOLI	Zinc (Zn+2) transport out via proton antiport (periplasm)
RXN-21018	g9239.t1	tr|E8PS81|E8PS81_YERPE	METHYLARSONITE + NADPH + OXYGEN-MOLECULE -> METHYLARSONATE + NADP + WATER
TRANS-RXN-178	G101.t1	tr|A7ISW5|A7ISW5_SOYBN	ATP + WATER -> ADP + CU+2 + PROTON + Pi

In addition to the arsenate transport reaction RXN-11187 shown in Table 2 , IDENER was able to confirm the presence of other metal-resistant or metal tolerance mechanisms that play crucial roles in heavy metal homeostasis and detoxification:

1. **ABC-type ATPase:** As demonstrated in reactions RXN-21035 and TRANS-RXN-178, these proteins are notable for their ATP-binding domain but does not undergo phosphorylation during its transport process. It is crucial for energy-dependent transport of various substances across cellular membranes without altering the transported substrate. (Evidence for **Mechanism#2**)
2. **Yeast Heavy Metal Export Enzyme:** As demonstrated in RXN-11187, this enzyme specializes in exporting metals like cadmium (Cd2+) from the cytosol into the vacuole. It also participates in the reduction of toxic arsenate, converting it into arseno-mycothiol. This product is subsequently reduced to arsenite, coupled with the formation of mycothiol-mycoredoxin disulfide, with mycoredoxin being recycled by mycothiol to form mycothione resulting in a defense system against arsenate toxicity (Evidence for **Mechanism#1** and **Mechanism#3**).
3. **FMN-containing Enzyme:** As demonstrated in TRANS-RXN0-200, this protein was originally identified in bacteria. This enzyme detoxifies trivalent methylated and aromatic arsenicals by oxidizing them to less harmful pentavalent species. The presence of FMN (flavin

mononucleotide) indicates its role in redox reactions essential for the detoxification pathway. (Evidence for **Mechanism#4**)

4. **P-type ATPase from *Archaeoglobus fulgidus*:** As demonstrated for the reaction RXN-21018, this enzyme is characterized by covalent phosphorylation during its transport cycle.A known example could be the extrusion of copper from cells, aiding in maintaining copper homeostasis under thermal stress conditions. (Evidence for **Mechanism#3**)

Identification of plant growth promoting factors

A GEM of fungal species with plant growth-promoting capacities, such as the production of hormones like indole acetic acid (IAA), gibberellins, and abscisic acid, can aid in the mapping of the specific biosynthetic pathways for these compounds. All six models include reactions for the synthesis of tryptophan for IAA production, terpenoid backbone synthesis for gibberellins, and carotenoid oxidation for abscisic acid. The terminal reactions for the synthesis of indole acetic acid, abscisic acid and jasmonic acid from their respective pathways in the GEMs are shown in Figure 10Figure 10. By simulating these reactions, the model can quantitatively evaluate the fungal species' capacity to produce these hormones under different environmental and physiological conditions. Such simulations can help in predicting how fungal interactions with plants can enhance growth, stress responses, or development, thereby assisting in designing of the biosystems in the succeeding work packages.

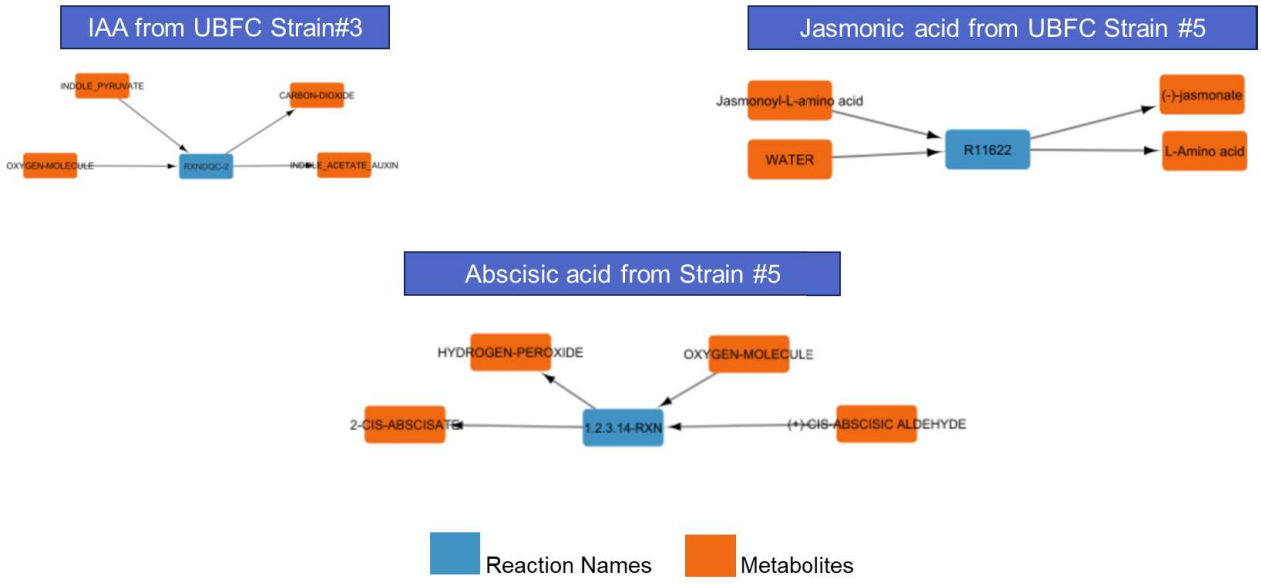


Figure 10: Terminal reactions of the biosynthetic pathways of indole acetic acid, abscisic acid and jasmonic acid taken from specific strains of the fungal GEMs.

Additionally, in Trp-dependent IAA biosynthesis, four pathways have been proposed:

(1) Pathway Type-I(a&b): the tryptamine (TAM) pathway

(2) Pathway Type-II: the indole-3-pyruvic acid (IPA) pathway

(3) Pathway Type-III: the indole-3-acetamide (IAM) pathway

While these pathways are typically associated with plants, fungal strains can exhibit variations in their IAA biosynthetic pathways (Bunsangiam et al., 2019). Research suggests that fungal strains, such as

Rhodotorula fluvialis DMKU-CP293, can utilize multiple pathways for IAA biosynthesis simultaneously, including the IAM, IPA, and TAM pathways (Bunsangiam et al., 2019). This indicates that fungal strains may have the capacity to engage in more than one Trp-dependent IAA biosynthetic pathway concurrently depending on the environmental growth conditions and activation of enzymes. For instance, oxygen is key for the conversion reactions in the redox reactions and decarboxylation events of Pathway Type-I and Pathway Type-II. The enzymatic steps, such as those catalysed by indole-3-acetaldehyde:oxygen oxidoreductase and indole-3-pyruvate NADPH-oxo-acid-lyase, are crucial for these conversion processes. A quantitative measure of the overall theoretical production yield of IAA will be estimated by performing FBA for each strain in relevant growth conditions.

Detection of phosphate solubilisation mechanisms

Phosphate-solubilizing fungi, such as the Trichoderma spp., can efficiently solubilize and mobilize Zn, Pb, and Cd in contaminated soils by increasing the bioavailability of these metals for plant uptake and phytoremediation (Syed et al., 2023a). In some cases, this is achieved through the production of organic acids, which lower the soil pH and form complexes with the metal cations, making them more bioavailable for plant uptake (Syed et al., 2023a). The organic acids secreted by these fungi, such as oxalic, citric, and gluconic acids, can dissolve insoluble phosphate minerals and release phosphate ions. This increased phosphate availability is beneficial for plant growth and development in metal-contaminated soils. This microbial-assisted approach has been shown to significantly increase the mobility and accumulation of Zn, Pb, and Cd in plants grown in contaminated soils. Thus, the phosphate-solubilization mechanisms of fungi play a crucial role in remediating rhizosphere at some of the sites within BIOSYSMO that are known to be contaminated with these metals.

Phosphate solubilization capacities can be detected within the GEMs to simulate the pathways and reactions involved in the production and secretion of organic acids such as citric, oxalic, and gluconic acids, which are known to solubilize bound phosphate (Figure 11). GEMs can be used to visualize the enzymatic reactions leading to these acids, such as, the conversion of glucose to gluconate via glucose dehydrogenase or citrate synthesis via the TCA cycle. By running simulations, these models can estimate the rates of organic acid production and their efficacy in phosphate solubilization under various environmental conditions. This insight can aid in designing plant-microbe biosystems for bioremediation purposes, enhancing phosphate recovery from polluted environments.

For detecting the possible phosphate solubilising traits within the genomes and for identifying the reactions, IDENER referred to an extensive list of such traits present within the PLaBase database, and extracted the orthologous gene groups from KEGG that they are associated with. This was followed by the localisation of the reactions within that are associated with these specific KEGG orthologous groups within the GEMs. This aided in the identification of all the potential[10] reaction routes and the organic acids that could be released during the phosphate solubilisation process by each fungal strain. Upon closely investigating these reactions, it was found that most of them were associated with the release organic acids.In Figure 11, the different organic acids that correspond to each phosphate-solubilisation mechanism that were obtained from PLaBase has been illustrated. It showcases phosphate solubilisation mechanisms for UBFC Strain #1, wherein 11 different phosphate-solubilisation traits were localised. The total number of reactions corresponding to each trait in the UBFC Strain #1 GEM is shown in the histogram on the right. Additionally, Figure 12 demonstrates the connectivity between the

[10] This hypothesis mandates further investigation and simulations to confirm whether these reactions by the virtue of pertaining to certain orthologous groups within KEGG are truly associated with the microbe's phosphate solubilizing mechanisms

reactions belonging to different phosphate-solubilisation categories (or the secretion of different organic acids) and their respective metabolites. For instance, several metabolites are shared between the synthesis of propionic and succinic acid (Figure 12).

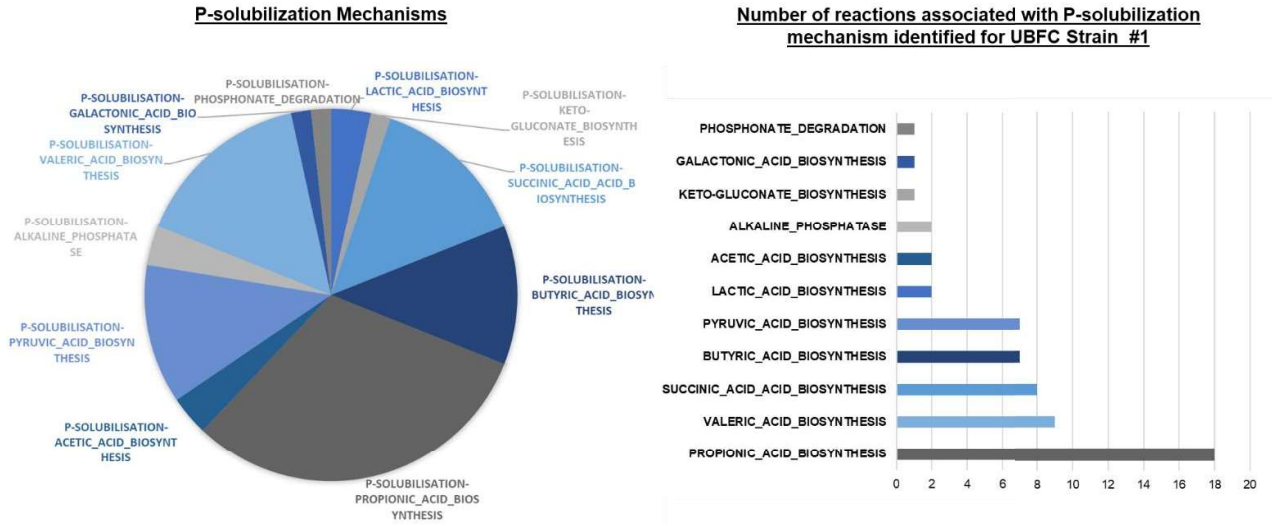


Figure 11: A pie-chart representation of the several phosphate-solubilising mechanisms detected within UBFC Strain #1 GEM alongside the details of the number of reactions associated. In most cases, the mechanism consists of the secretion of an organic acid.

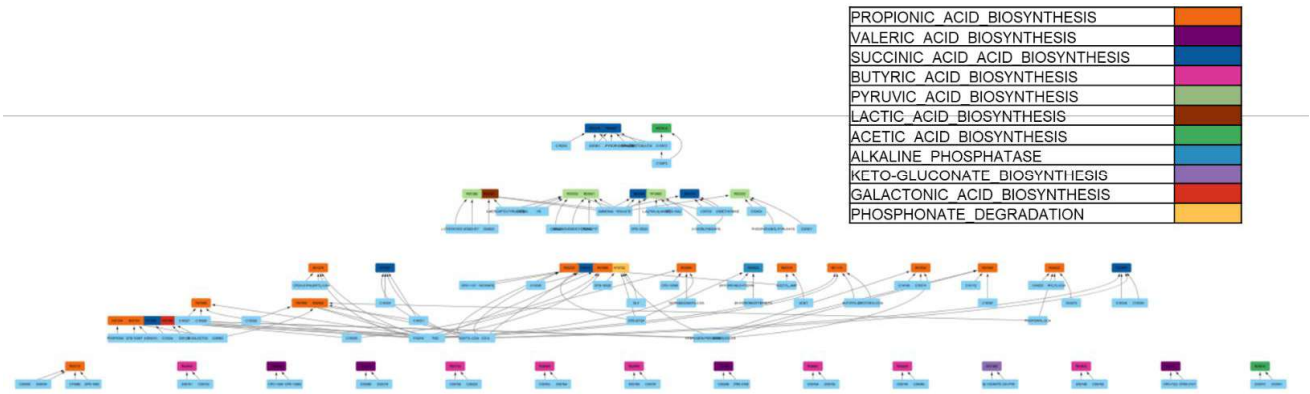


Figure 12: Representation of the connectivity of the different phosphate solubilisation reactions and their metabolites in the UBFC Strain #1 GEM. Each coloured box shows a reaction categorized into a specific mechanism and the boxes in blue are the metabolites associated.

4 Conclusions from the GEM models in the bottom-up approach

4.1 GEM performance with lindane and atrazine degradation

Through the optimisation of the GEM for *P. putida*, IDENER demonstrated a maximum specific growth rate of 0.042 h^{-1} with lindane as the substrate and the sole carbon source. Zhao et al. (2021) generated a genetically modified mutant strain of *P. putida* KT2440, referred to as *P. putida* KTLTG with

biodegradation pathways for 1,2,3-trichloropropane and lindane. The maximum specific growth rate observed for *P. putida* KTLTG in a complex medium was 0.3 h^{-1}. This inconsistency could be attributed to the variations in the introduced biodegradation pathways and the presence of additional carbon source through the utilisation of a rich complex media (Zhao et al., 2021). Therefore, the next set of model optimisations or corrections will be based on the experimental estimation of the growth rates from the *P. putida* KT2440_pSEVA-lindane mutant strain engineered for lindane degradation at ICL. Additionally, there are two toxic metabolites that are known to be produced during the degradation of lindane: a) 1,2,4-trichlorobenzene (1,2,4-TCB) and b) 2,5-dichlorophenol (2,5-DCP). They are regarded as dead-end products in the original γ-HCH degrader, *S. japonicum* UT26 (Lal et al., 2010; Nagasawa et al., 1993), and therefore, might be generated in *P. putida* through spontaneous dehydrochlorination reactions (Figure 1). However, these metabolites were absent in the genetically modified *P. putida* KTLTG strain developed in Zhao *et al.* (2021) for lindane degradation. This finding was also consistent with the output of the model where the degradation of γ-HCH did not result in the flux through the reactions representing the dead-end metabolites.

Following a similar protocol for atrazine degradation, IDENER estimated a growth rate of 0.071 h^{-1} accompanied by an atrazine degradation rate of 0.24 mmol gDW^{-1} h^{-1}. As stated earlier in Section 2.1.1, to IDENER will investigate the fate of the dead-end metabolites, by considering the inclusion of ethylamine transformation reactions into the GEM. According to results shown by Benninghaus et al. (2021), *P. putida* KT2440 strain includes a specific enzyme capable of catalysing the reaction between L-glutamate and ethylamine to form Theanine (also known as N5-ethyl-L-glutamine), an analog of amino acids, accompanied by the consumption of ATP. This enzyme is identified as EC 6.3.1.6, glutamate---ethylamine ligase. However, further discussion will be initiated with ICL to confirm this and add the corresponding reactions into the GEM. For the second dead-end metabolite, isopropylamine, current databases lack evidence to confirm the presence of appropriate enzymes within the genome of *P. putida* for its metabolism. In fact, in a recent study by Benninghaus et al. (2019), *P. putida* had to be genetically modified to contain the enzyme γ-glutamylmethylamide synthetase (GMAS) for consuming isopropylamine and producing the amino acid analog γ-Glutamyl−isopropylamide, further confirming the strain's inability to consume this dead-end metabolite.

4.2 In-depth analysis of fungal models

Based on the results and the detection of the different metal resistance, PGP traits and capacity of the microbes to solubilise phosphate, IDENER has provided its support to the design of the biosystems by using the bottom-up approach, in which the synthesis and optimisation of individual models have the capacity to predict these different traits will show us how they perform whenever they are combined into communities or paired with plants that are intended to be used in the biosystems. A summary of the different traits or reactions that have been identified in the respective GEM of each strain is given below:

Metal resistance and interactions

At sites contaminated by metals, the microbiome was initially characterized for its high metal content, particularly arsenic, lead, and zinc, posing significant health risks. The six fungal strains were selected and analysed using GEMs to assess their metal-resistance capabilities revealed that these strains possess various mechanisms for metal detoxification, including sequestration in vacuoles, efflux through membrane transporters, and binding by chelating or precipitating compounds. In the future, simulation of these processes within the GEMs will, therefore, facilitate a detailed understanding of the metabolic pathways engaged in metal resistance, and give quantitative estimations of their

performance. This insight is crucial for forthcoming experiments with contaminated soil and selection of candidates for the biosystems in WP3.

Plant growth promoting traits

The GEMs developed for the fungal strains also mapped the biosynthetic pathways for several plant growth-promoting compounds such as IAA, jasmonic acid, and abscisic acid. These pathways are instrumental for the synthesis of crucial plant hormones that enhance growth, stress response, and development (Syed et al., 2023b). Simulation results from these models will provide quantitative insights into the capability of fungi to produce IAA and the simultaneous activation of different pathways to do so. The effective synthesis of these compounds suggests potential benefits for multi-species biosystems where enhancement of plant growth and resistance traits in the presence of endophytic fungi will be tested. The representation of these pathways in the GEMs helps in understanding the extracellular concentration of metabolites released by the fungal strain and their uptake by the plants, facilitating the design of effective biosystems.

Phosphate solubilization mechanisms

The capacity for phosphate solubilization was another significant trait that could be investigated through the GEMs constructed. Within the context of BIOSYSMO, fungi play a pivotal role in mobilizing essential nutrients like zinc, lead, and cadmium in contaminated soils, enhancing their availability for plant uptake. The GEMs generated for six fungal strains that were screened *in-vitro* and sequenced by UBFC, contain the capacity to produce and secrete organic acids, which solubilize bound phosphate, thereby aiding plant growth and development in metal-polluted environments. In the future, the simulation of these biochemical pathways within the GEMs would allow for a quantitative prediction of the rates at which these acids are produced and their effectiveness in solubilizing phosphate under variable growth conditions. Such models are valuable for selecting suitable candidates for biosystems and their subsequent inoculation with contaminated soil having high concentrations of metals.

Enrichment of GEMs with bioremediation/biodegradation reactions from the BIOSYSMOdb and other databases

As has been stated in Section 1.1.1, it is important to state again that the richness and the diversity of the reactions in a GEM depends on the reference database (Thiele & Palsson, 2010). In the case that some key reactions pertaining to bioremediation and biodegradation are absent in the reference database associated with the model reconstruction tool, those reactions/genes will also remain undetected in the GEM. Therefore, in the next reporting period, IDENER will take advantage of the BIOSYSMOdb, to expand and enrich the generated GEMs to contain key reactions. Moreover, efforts will also be made to enrich the GEMs during the process of their construction by attempting to change the associated reference database inherently present within the algorithm to a more extensive one such as MetaNetX[11] (Moretti et al., 2021).

5 References

Bazzi, W., Abou Fayad, A. G., Nasser, A., Haraoui, L.-P., Dewachi, O., Abou-Sitta, G., Nguyen, V.-K., Abara, A., Karah, N., Landecker, H., Knapp, C., McEvoy, M. M., Zaman, M. H., Higgins, P. G., & Matar, G. M. (2020). Heavy Metal Toxicity in Armed Conflicts Potentiates AMR in A. baumannii by

[11] https://www.metanetx.org/

Selecting for Antibiotic and Heavy Metal Co-resistance Mechanisms. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00068

Benninghaus, L., Walter, T., Mindt, M., Risse, J. M., & Wendisch, V. F. (2021). Metabolic Engineering of *Pseudomonas putida* for Fermentative Production of -Theanine. *Journal of Agricultural and Food Chemistry*, *69*(34), 9849–9858. https://doi.org/10.1021/acs.jafc.1c03240

Bernstein, D. B., Sulheim, S., Almaas, E., & Segrè, D. (2021). Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. *Genome Biology*, *22*(1), 64. https://doi.org/10.1186/s13059-021-02289-z

Bunsangiam, S., Sakpuntoon, V., Srisuk, N., Ohashi, T., Fujiyama, K., & Limtong, S. (2019). Biosynthetic Pathway of Indole-3-Acetic Acid in Basidiomycetous Yeast *Rhodosporidiobolus fluvialis*. *Mycobiology*, *47*(3), 292–300. https://doi.org/10.1080/12298093.2019.1638672

Collot, J., Binet, P., Malabad, A. M., Pauget, B., Toussaint, M.-L., & Chalot, M. (2023). Floristic survey, trace element transfers between soil and vegetation and human health risk at an urban industrial wasteland. *Journal of Hazardous Materials*, *459*, 132169. https://doi.org/10.1016/j.jhazmat.2023.132169

Esquirol, L., Peat, T. S., Wilding, M., Liu, J.-W., French, N. G., Hartley, C. J., Onagi, H., Nebl, T., Easton, C. J., Newman, J., & Scott, C. (2018). An unexpected vestigial protein complex reveals the evolutionary origins of an s-triazine catabolic enzyme. *Journal of Biological Chemistry*, *293*(20), 7880–7891. https://doi.org/10.1074/jbc.RA118.001996

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir, H. S., Wachowiak, J., Keating, S. M., Vlasov, V., Magnusdóttir, S., Ng, C. Y., Preciat, G., Žagare, A., Chan, S. H. J., Aurich, M. K., Clancy, C. M., Modamio, J., Sauls, J. T., Fleming, R. M. T. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, *14*(3), 639–702. https://doi.org/10.1038/s41596-018-0098-2

Lal, R., Pandey, G., Sharma, P., Kumari, K., Malhotra, S., Pandey, R., Raina, V., Kohler, H.-P. E., Holliger, C., Jackson, C., & Oakeshott, J. G. (2010). Biochemistry of Microbial Degradation of Hexachlorocyclohexane and Prospects for Bioremediation. *Microbiology and Molecular Biology Reviews*, *74*(1), 58–80. https://doi.org/10.1128/MMBR.00029-09

Lapidus, A. L., & Korobeynikov, A. I. (2021). Metagenomic Data Assembly – The Way of Decoding Unknown Microorganisms. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.613791

Medlock, G. L., & Papin, J. A. (2020). Guiding the Refinement of Biochemical Knowledgebases with Ensembles of Metabolic Networks and Machine Learning. *Cell Systems*, *10*(1), 109-119.e3. https://doi.org/10.1016/j.cels.2019.11.006

Moretti, S., Tran, V. D. T., Mehl, F., Ibberson, M., & Pagni, M. (2021). MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models. *Nucleic Acids Research*, *49*(D1), D570–D574. https://doi.org/10.1093/nar/gkaa992

Nagasawa, S., Kikuchi, R., & Matsuo, M. (1993). Indirect identification of an unstable intermediate in γ-HCH degradation by Pseudomonas paucimobilis UT26. *Chemosphere*, *26*(12), 2279–2288. https://doi.org/10.1016/0045-6535(93)90353-7

Nogales, J., Mueller, J., Gudmundsson, S., Canalejo, F. J., Duque, E., Monk, J., Feist, A. M., Ramos, J. L., Niu, W., & Palsson, B. O. (2020). High-quality genome-scale metabolic modelling of *Pseudomonas putida* highlights its broad metabolic capabilities. *Environmental Microbiology*, *22*(1), 255–269. https://doi.org/10.1111/1462-2920.14843

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248. https://doi.org/10.1038/nbt.1614

Steinegger, 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

Syed, A., Elgorban, A. M., Bahkali, A. H., Eswaramoorthy, R., Iqbal, R. K., & Danish, S. (2023a). Metal-tolerant and siderophore producing Pseudomonas fluorescence and Trichoderma spp. improved the growth, biochemical features and yield attributes of chickpea by lowering Cd uptake. *Scientific Reports*, *13*(1), 4471. https://doi.org/10.1038/s41598-023-31330-3

Syed, A., Elgorban, A. M., Bahkali, A. H., Eswaramoorthy, R., Iqbal, R. K., & Danish, S. (2023b). Metal-tolerant and siderophore producing Pseudomonas fluorescence and Trichoderma spp. improved the growth, biochemical features and yield attributes of chickpea by lowering Cd uptake. *Scientific Reports*, *13*(1), 4471. https://doi.org/10.1038/s41598-023-31330-3

Thiele, I., & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, *5*(1), 93–121. https://doi.org/10.1038/nprot.2009.203

Văcar, C. L., Covaci, E., Chakraborty, S., Li, B., Weindorf, D. C., Frențiu, T., Pârvu, M., & Podar, D. (2021). Heavy Metal-Resistant Filamentous Fungi as Potential Mercury Bioremediators. *Journal of Fungi*, *7*(5), 386. https://doi.org/10.3390/jof7050386

Wang, H., Marcišauskas, S., Sánchez, B. J., Domenzain, I., Hermansson, D., Agren, R., Nielsen, J., & Kerkhoven, E. J. (2018). RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor. *PLOS Computational Biology*, *14*(10), e1006541. https://doi.org/10.1371/journal.pcbi.1006541

Yoon, B.-J. (2009). Hidden Markov Models and their Applications in Biological Sequence Analysis. *Current Genomics*, *10*(6), 402–415. https://doi.org/10.2174/138920209789177575

Zhang, W., Lin, Z., Pang, S., Bhatt, P., & Chen, S. (2020). Insights Into the Biodegradation of Lindane (γ-Hexachlorocyclohexane) Using a Microbial System. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00522

Zhao, Y., Che, Y., Zhang, F., Wang, J., Gao, W., Zhang, T., & Yang, C. (2021). Development of an efficient pathway construction strategy for rapid evolution of the biodegradation capacity of Pseudomonas putida KT2440 and its application in bioremediation. *Science of The Total Environment*, *761*, 143239. https://doi.org/10.1016/j.scitotenv.2020.143239

Zhuang, K., Izallalen, M., Mouser, P., Richter, H., Risso, C., Mahadevan, R., & Lovley, D. R. (2011). Genome-scale dynamic modeling of the competition between *Rhodoferax* and *Geobacter* in anoxic subsurface environments. *The ISME Journal*, *5*(2), 305–316. https://doi.org/10.1038/ismej.2010.117

Zhuang, K., Vemuri, G. N., & Mahadevan, R. (2011). Economics of membrane occupancy and respiro-fermentation. *Molecular Systems Biology*, *7*(1). https://doi.org/10.1038/msb.2011.34